

Improving less constrained iris recognition

A Thesis Submitted to
The University of Kent
for The Degree of
Doctor of Philosophy
in Electronic Engineering

Yang Hu

Jan 2017

Abstract

The iris has been one of the most reliable biometric traits for automatic human authentication due to its highly stable and distinctive patterns. Traditional iris recognition algorithms have achieved remarkable performance in strictly constrained environments, with the subject standing still and with the iris captured at a close distance. This enables the wide deployment of iris recognition systems in applications such as border control and access control. However, in less constrained environments with the subject at-a-distance and on-the-move, the iris recognition performance is significantly deteriorated, since such environments induce noise and degradations in iris captures. This restricts the applicability and practicality of iris recognition technology for some real-world applications with more open capturing conditions, such as surveillance, forensic and mobile device security applications. Therefore, robust algorithms for less constrained iris recognition are desirable for the wider deployment of iris recognition systems.

This thesis focuses on improving less constrained iris recognition. Five methods are proposed to improve the performance of different stages in less constrained iris recognition. First, a robust iris segmentation algorithm is developed using ℓ_1 -norm regression and model selection. This algorithm formulates iris segmentation as robust ℓ_1 -norm regression problems. To further enhance the robustness, multiple segmentation results are produced by applying ℓ_1 -norm regression to different models, and a model selection technique is used to select the most reliable result. Second, an iris liveness detection method using regional features is investigated. This method seeks not only low level features, but also high level feature distributions for more accurate and robust iris liveness detection. Third, a signal-level information fusion algorithm is presented to mitigate the noise in less constrained iris captures. With multiple noisy iris captures, this algorithm proposes a sparse-error low rank matrix factorization model to separate noiseless iris structures and noise. The noiseless structures are preserved and emphasised during the fusion process, while the noise is suppressed, in order to obtain more reliable signals for recognition. Fourth, a method to generate optimal iris codes is proposed. This method considers iris code generation from the perspective of optimization. It formulates traditional iris code

generation method as an optimization problem; an additional objective term modelling the spatial correlations in iris codes is applied to this optimization problem to produce more effective iris codes. Fifth, an iris weight map method is studied for robust iris matching. This method considers both intra-class bit stability and inter-class bit discriminability in iris codes. It emphasises highly stable and discriminative bits for iris matching, enhancing the robustness of iris matching.

Comprehensive experimental analysis are performed on benchmark datasets for each of the above methods. The results indicate that the presented methods are effective for less constrained iris recognition, generally improving state-of-the-art performance.

Acknowledgements

Pursuing a PhD is a challenging journey. It would not have been possible for me to write this doctoral thesis without the professional and personal support received from many kind people.

I wish to express my deepest gratitude to my supervisors, Dr. Gareth Howells and Dr. Konstantinos Sirlantzis. Dr. Howells and Dr. Sirlantzis gave me this precious opportunity to pursue PhD under their supervision. They have profound expertise and broad research vision which have guided me to several fruitful directions for my PhD research. I am grateful for the advices, comments, encouragements and project opportunities received from them during this study.

University of Kent and its School of Engineering and Digital Arts deserve grateful thanks for many things. University of Kent and its School of Engineering and Digital Arts have an open and inclusive multicultural environment which creates excellent study and research atmosphere for international students like me. The advanced education ideology and administration methods are greatly beneficial for my PhD study. The technical support is solid and the devices are state-of-the-art, significantly contributing to the completion of my PhD study. I would like to especially thank the administration team of the school led by Dr. Farzin Deravi and Mr. Giles Polglase for their professional support during my PhD study.

I would like to express appreciation to Dr. Nicolas Ragot and Dr. Paul Rodriguez. As academic collaborators, Dr. Ragot and Dr. Rodriguez have kindly given me tremendous help and guidance for my research. I have enormously benefited from our constructive and productive collaboration. It is my great luck to have the opportunity to collaborate with them.

As long as I am writing names down, I cannot neglect that of Chunwei Tan. Tan has given me crucial help by sharing with me the data used in his experiment. These data have been a critical part of the datasets used by me to perform experiments not only to produce my first few publications, but also to obtain some experimental results for this thesis.

Last but not the least, I am also deeply appreciated other collaborators, colleagues and persons who have directly or indirectly helped me academically and personally.

TO MY PARENTS

for every effort of their support

for every minute of their consideration

for every word of their encouragement

for every tear that they have wiped away from my face

for every smile that they have beamed for my achievements

for every thing that they have done for me

for their love that will be forever engraved in my mind

Contents

Abstract	i
Acknowledgements	iii
List of Tables	xv
List of Figures	xx
Acronyms	xxiv
List of Publications	xxix
1 Introduction	1
1.1 Biometric recognition systems	1
1.2 Iris recognition	3
1.3 Motivation and objectives	6
1.4 Contributions	7
1.5 Thesis structure	9

2	Previous works on iris recognition	11
2.1	Iris segmentation	11
2.1.1	Integro-differential operator	11
2.1.2	Hough Transform	13
2.1.3	Fourier active contour	13
2.1.4	Geodesic active contour	14
2.1.5	Pulling and pushing method	15
2.1.6	Integro-differential constellation method	16
2.1.7	Colour and sclera features based classification	17
2.1.8	Localised Zernike moments based classification	17
2.1.9	Random walker algorithm and post-processing	18
2.2	Iris liveness detection	19
2.2.1	Brightness variation based thresholding	19
2.2.2	Gray level co-occurrence matrix (gray level co-occurrence matrix (GLCM)) based classification	20
2.2.3	Textural features based classification	20
2.2.4	Wavelet packets transform (wavelet packets transform (WPT)) based classification	22
2.2.5	Quality measures based classification	22
2.2.6	Multi-resolution local binary pattern (local binary pattern (LBP)) and Adaboost based classification	22

2.2.7	Gaze features based classification	23
2.2.8	Quality and texture based features for colour captures	23
2.2.9	Local descriptors based classification	24
2.2.10	Pupil Dynamic features based classification	24
2.2.11	Deep learning based method	25
2.3	Feature extraction	25
2.3.1	Binary features	25
	Gabor features	25
	Logarithm Gabor (log-Gabor) features	26
	Ordinal features	26
	Geometric key (geoKey) based features.	27
2.3.2	Non-binary features	28
	Ma <i>et al.</i> features	28
	Tan <i>et al.</i> features	28
	Weighted co-occurrence phase histogram (WCPH)	28
	Zernike moments phase features	29
	Deep representation	29
2.4	Iris matching	30
2.4.1	General distance metrics based methods	30
	Hamming distance	30

ℓ_2 distance	30
ℓ_1 distance	31
Cosine distance	31
Diffusion distance	31
χ^2 distance	31
Bhattacharyya distance	31
2.4.2 Sparse representation based method	32
2.4.3 Iris weight map based methods	34
Fragile bits	34
Personalised weight map	35
Power law based weight map	35
Bit discriminability based map	36
3 Datasets	37
3.1 CASIA.v4 thousand (CASIAT)	37
3.2 ND-iris-0405 (ND0405)	38
3.3 CASIA.v4 distance (CASIAD)	39
3.4 UBIRIS.v2 (UBIRIS2)	40
3.5 Face Recognition Grand Challenge (FRGC)	40
3.6 Clarkson iris liveness detection 2013 (LivDet-Clarkson)	41
3.7 Warsaw iris liveness detection 2013 (LivDet-Warsaw)	41

3.8	Notre Dame iris liveness detection 2013 (LivDet-ND)	42
3.9	MobBIOfake	42
3.10	MICHE	42
3.11	CSIR	43
4	A robust iris segmentation algorithm using ℓ_1-norm regression and model selection technique	44
4.1	Introduction	44
4.2	Iris segmentation using ℓ_1 -norm regression	47
4.2.1	Coarse iris localisation	47
4.2.2	Limbic boundary segmentation	50
4.2.3	Pupillary boundary segmentation	56
4.2.4	Upper eyelid fitting	57
4.2.5	Lower eyelid fitting	58
4.2.6	Reflection removal	58
4.2.7	Silhouette removal	59
4.2.8	Analysis of the three models	60
4.3	Model selection	61
4.4	Experiments	63
4.4.1	Experimental setting	64
4.4.2	Analysis for each individual iris segmentation model	66

4.4.3	Analysis for model selection	70
4.4.4	Comparison with other methods	73
4.5	Summary	75
5	Iris liveness detection using regional features	76
5.1	Introduction	76
5.2	Methodology	78
5.2.1	Iris segmentation	78
5.2.2	Spatial pyramid model	78
5.2.3	Relational measure model	81
5.2.4	Low-level features	84
5.2.5	Classifier	87
5.2.6	Score-level fusion	87
5.3	Experiments	88
5.3.1	Datasets and evaluation	88
5.3.2	Performance of single regional features	88
5.3.3	Comparison between regional features and low-level features	92
5.3.4	Effect of score-level fusion	96
5.3.5	Comparison with existing methods	97
5.4	Summary	101

6	Signal-level information fusion using sparse-error low rank matrix factorisation	103
6.1	Introduction	103
6.2	Related work	105
6.2.1	Information fusion in iris recognition	105
6.2.2	Low rank approximation (low-rank approximation (LRA))	107
6.3	Methodology	109
6.3.1	Sparse-error low rank matrix factorisation (sparse-error low rank matrix factorization (SE-LRMF))	109
6.3.2	Signal-level fusion	113
6.3.3	Code level combination	117
6.4	Experiment	118
6.4.1	Experimental setting	119
6.4.2	Analysis of each signal-level fusion step	122
6.4.3	Comparison between SE-LRMF and existing low rank approximation algorithms	124
6.4.4	Comparison with no fusion methods	129
6.4.5	Comparison with existing signal-level fusion methods	132
6.4.6	Comparison with existing score-level fusion methods	132
6.4.7	Comparison with other recent iris recognition algorithms	138
6.4.8	Summary on the performance of SE-LRMF and existing algorithms	139

6.4.9	The influence of noise on the performance of SE-LRMF	141
6.5	Summary	146
7	Optimal generation of iris codes	148
7.1	Introduction	148
7.2	Producing iris codes by optimisation	150
7.2.1	Illustration of the traditional iris code production from a perspective of optimisation	150
7.2.2	Iris code production using additional objective term	152
7.3	Experiment	155
7.3.1	Experimental setting	155
7.3.2	Comparison with iris codes produced by an implementation of traditional iris code generation method	159
7.3.3	Analysis of the spatial correlation in iris textures	163
7.3.4	Analysis on the computational cost	167
7.4	Summary	169
8	An iris weight map method exploiting bit stability and discriminability	170
8.1	Introduction	170
8.2	Methodology	171
8.2.1	Computation of meta-images and meta-codes	172
8.2.2	The stability map	173

8.2.3	The discriminability map	180
8.2.4	The combined map	182
8.2.5	Iris matching	183
8.3	Experiment	184
8.3.1	Datasets	184
8.3.2	Experimental configuration	185
8.3.3	Parameter tuning and performance evaluation	186
8.3.4	Analysis of the proposed iris weight map	188
8.3.5	Comparison with other methods	198
8.3.6	Analysis of computational cost	204
8.4	Summary	206
9	Conclusion	208
9.1	Summary of thesis achievements	208
9.2	Summary of findings	211
9.3	Future work	212
	Bibliography	214

List of Tables

3.1	Information of all the datasets used in this thesis.	39
4.1	Significance analysis of each component in individual iris segmentation models.	67
4.2	Comparison between ℓ_1 -norm and ℓ_2 -norm regression on the proposed circle model.	69
4.3	Comparison between ℓ_1 -norm and ℓ_2 -norm regression on the proposed parametric ellipse model.	69
4.4	Comparison between ℓ_1 -norm and ℓ_2 -norm regression on the proposed ellipse model.	69
4.5	Results of the proposed model selection method.	70
4.6	Mean E1-score of proposed model selection method using different training data for colour datasets.	72
4.7	Mean E1-score on MICHE dataset with device-specific training and test data.	73
4.8	Comparison of the performance with other methods.	74
5.1	Information of the datasets used in this experiment	88

5.2	Results of score-level fusion (error rate)	97
5.3	Results of comparisons with other methods (error rate)	99
6.1	Parameter setting on each dataset.	122
6.2	Percentage of inconsistent bits between low rank mean (LRM) and error weighted mean (EWM) fusion results	128
6.3	The rank 1 recognition accuracy (R1RA) of SE-LRMF and existing iris recognition algorithms.	142
6.4	The equal error rate (EER) of SE-LRMF and existing iris recognition algorithms.	142
6.5	The genuine acceptance rate (GAR)@false acceptance rate (FAR)= 0.01% of SE-LRMF and existing iris recognition algorithms.	142
6.6	The discriminability index (DI) of SE-LRMF and existing iris recog- nition algorithms.	143
7.1	Information and parameter setting of the datasets used in this exper- iment.	156
7.2	The R1RA for comparison between the proposed iris code and the iris code produced by Masek's implementation of traditional iris code generation method, with the result of significance test.	160
7.3	The EER for comparison between the proposed iris code and the iris code produced by Masek's implementation of traditional iris code generation method, with the result of significance test.	161

7.4	The GAR@FAR= 0.01% for comparison between the proposed iris code and the iris code produced by Masek's implementation of traditional iris code generation method, with the result of significance test.	161
7.5	The DI for comparison between the proposed iris code and the iris code produced by Masek's implementation of traditional iris code generation method, with the result of significance test.	162
7.6	The R1RA for the analysis of spatial correlation in iris textures, together with the results of significance test.	165
7.7	The EER for the analysis of spatial correlation in iris textures, together with the results of significance test.	166
7.8	The GAR@FAR= 0.01% for the analysis of spatial correlation in iris textures, together with the results of significance test.	167
7.9	The DI for the analysis of spatial correlation in iris textures, together with the results of significance test.	167
7.10	Computational cost of each method to generate iris codes (ms). . . .	168
8.1	Summarisation on the datasets for experiments in this chapter. . . .	185
8.2	Parameter setting.	187
8.3	The R1RA to analyse the proposed iris weight map in the single-sensor experiment, with the result of significance test.	190
8.4	The EER to analyse the proposed iris weight map in the single-sensor experiment, with the result of significance test.	191
8.5	The GAR@FAR= 0.01% to analyse the proposed iris weight map in the single-sensor experiment, with the result of significance test. . . .	192

8.6	The DI to analyse the proposed iris weight map in the single-sensor experiment, with the result of significance test.	192
8.7	The R1RA to analyse the proposed iris weight map in the cross-sensor experiment, with the result of significance test.	196
8.8	The EER to analyse the proposed iris weight map in the cross-sensor experiment, with the result of significance test.	196
8.9	The GAR@FAR= 0.01% to analyse the proposed iris weight map in the cross-sensor experiment, with the result of significance test. . . .	196
8.10	The DI to analyse the proposed iris weight map in the cross-sensor experiment, with the result of significance test.	197
8.11	The R1RA for comparison in the single-sensor experiment, with the result of significance test.	199
8.12	The EER for comparison in the single-sensor experiment, with the result of significance test.	200
8.13	The GAR@FAR= 0.01% for comparison in the single-sensor experiment, with the result of significance test.	201
8.14	The DI for comparison in the single-sensor experiment, with the result of significance test.	201
8.15	The R1RA for comparison in the cross-sensor experiment, with the result of significance test.	202
8.16	The EER for comparison in the cross-sensor experiment, with the result of significance test.	203
8.17	The GAR@FAR= 0.01% for comparison in the cross-sensor experiment, with the result of significance test.	203

8.18 The DI for comparison in the cross-sensor experiment, with the result of significance test.	203
8.19 Analysis of computational cost (ms).	205

List of Figures

1.1	Examples of cumulative match characteristic (CMC) and receiver operating characteristics (ROC) curves.	4
1.2	Illustrations of a general iris recognition system.	5
3.1	Examples of datasets used in this thesis.	38
4.1	Flowchart and illustrations of the proposed ℓ_1 -norm regression based iris segmentation.	45
4.2	Flowchart and illustrations of the proposed model selection method.	46
4.3	Examples of super-pixels and coarse iris map.	48
4.4	Examples of correlation histogram corresponding to an iris super-pixel and a non-iris super-pixel.	49
4.5	An illustration of the proposed limbic boundary segmentation algorithm.	51
4.6	An illustration of the effect of the size of iris centre candidate region.	52
4.7	Example results of eyelids fitting.	59
4.8	The mean estimation error as a function of noise level on the synthetic data.	62

4.9	An illustration of the effect of the model selection.	71
5.1	Flowchart of the proposed method for iris liveness detection.	77
5.2	An illustration of the spatial pyramid-based regional feature.	80
5.3	An illustration of the relational measure model on an iris region with soft contact lens.	82
5.4	An example of LBP feature.	85
5.5	Performance of single regional features using spatial pyramid model. . .	89
5.6	Performance of single regional features using relational measure model.	90
5.7	Performance of single regional features on Notre Dame iris liveness detection 2013 (LivDet-ND) dataset down-sampled by a factor of 0.75.	91
5.8	Comparison of the performance between low-level features and re- gional features constructed based on each low-level feature on all the datasets.	93
6.1	Flowchart of the proposed method for iris liveness detection.	104
6.2	Examples of the results of SE-LRMF based signal-level fusion.	114
6.3	The CMC and ROC curves to analyse each signal-level fusion step. . .	123
6.4	The CMC and ROC curves for the comparison between SE-LRMF and existing low rank approximation models.	125
6.5	The distribution of ℓ_1 distances between the raw feature values ex- tracted from LRM and EWM results.	130
6.6	The CMC and ROC curves for the comparison between SE-LRMF and no fusion methods.	133

6.7	The CMC and ROC curves for the comparison between SE-LRMF and existing signal-level fusion methods.	134
6.8	The CMC and ROC curves for the comparison between SE-LRMF and existing score-level fusion methods.	136
6.9	The CMC and ROC curves for the comparison between SE-LRMF and other recent iris recognition algorithms.	140
6.10	The performance of SE-LRMF and score-level fusion methods on varying levels of different quality measures.	145
7.1	Examples of the iris codes produced by different methods.	154
7.2	Comparison of the CMC curves between the proposed iris code and the iris codes produced by Masek's implementation of traditional iris code generation method.	159
7.3	Comparison of the ROC curves between the proposed iris code and the iris code produced by Masek's implementation of traditional iris code generation method.	160
7.4	CMC curves of the proposed iris code for the analysis of spatial correlation in iris textures.	165
7.5	ROC curves of the proposed iris code for the analysis of spatial correlation in iris textures.	166
8.1	Examples of meta-images.	173
8.2	The function defined by Eqn. 8.6 with different choices of β_1 and β_2	177
8.3	Examples of stability map with linear model, discriminability map and combined map.	183

8.4	The CMC curves to analyse the proposed iris weight map in the single-sensor experiment.	190
8.5	The ROC curves to analyse the proposed iris weight map in the single-sensor experiment.	191
8.6	The CMC curves to analyse the proposed iris weight map in the cross-sensor experiment.	195
8.7	The ROC curves to analyse the proposed iris weight map in the cross-sensor experiment.	195
8.8	The CMC curves for comparison between proposed iris weight maps and existing iris weight map methods in the single-sensor experiment.	199
8.9	The ROC curves for comparison between proposed iris weight maps and existing iris weight map methods in the single-sensor experiment.	200
8.10	The CMC curves for comparison between proposed iris weight maps and existing iris weight map methods in the cross-sensor experiment.	202
8.11	The ROC curves for comparison between proposed iris weight maps and existing iris weight map methods in the cross-sensor experiment.	202

List of Abbreviations

l_1 -C l_1 -norm circle. 54, 61, 67–69

l_1 -E l_1 -norm ellipse. 55, 61, 67–69

l_1 -PE l_1 -norm parametric ellipse. 55, 61, 67–69, 71

l_2 -C l_2 -norm circle. 68, 69

l_2 -E l_2 -norm ellipse. 68, 69

l_2 -PE l_2 -norm parametric ellipse. 68, 69

1-D 1-dimensional. 27, 29, 120, 121, 157, 158, 185, 187

2-D 2-dimensional. 25, 27, 28

3-D 3-dimensional. 40

AO architecture optimization. 25

AUC area under curve. 3

BSIF binarized statistical image features. 24

CASIAD CASIA.v4 distance. 37–39, 46, 53, 56, 57, 64–71, 73, 74, 119, 121–123, 125–128, 130, 133–137, 140–143, 145, 155–162, 165–168, 184–187, 190–192, 199–201, 205

- CASIAT** CASIA.v4 thousand. 37–39, 155–162, 165–168
- CH** correlation histogram. 49, 50
- CLC** code level combination. 122, 124, 126–129
- CMC** cumulative match characteristic. xx–xxiii, 3, 4, 122, 123, 131–138, 140, 158, 159, 162, 164, 165, 187, 189, 190, 194, 195, 199, 202
- CoA-LBP** co-occurrence of adjacent LBP. 24
- CSCI** cumulative sparsity concentration index (SCI). 33
- DA** discriminant analysis. 24
- DC** direct current. 26
- DI** discriminability index. xvi–xix, 139, 143, 161–164, 167, 187, 189, 192, 194, 197, 198, 201, 203
- DTFT** discrete-time fourier transform. 86
- EER** equal error rate. xvi–xviii, 3, 139, 142, 159, 161, 163, 164, 166, 187, 189, 191, 194, 196, 198, 200, 203
- EF** eyelids fitting. 67
- EWM** error weighted mean. xvi, xxi, 109, 113–118, 122, 124, 126–130, 146, 209
- FAR** false acceptance rate. xvi–xviii, 3, 139, 142, 161, 163, 164, 167, 187, 189, 192, 194, 196, 198, 201, 203
- FB** fragile bits. 198–205
- FFS** fisher feature selection. 198–205
- FO** filter optimization. 25

- FRGC** Face Recognition Grand Challenge. 37–40, 46, 47, 53, 56, 57, 64–74
- GAR** genuine acceptance rate. xvi–xviii, 3, 139, 142, 159, 161, 163, 164, 167, 187, 189, 192, 194, 196, 198, 201, 203
- geoKey** geometric key. 27
- GLCM** gray level co-occurrence matrix. vii, 20, 21, 23
- HOG** histogram of oriented gradients. 46, 63
- KNN** k-nearest neighbour. 24
- LBC** limbic boundary candidate. 53–56, 60
- LBP** local binary pattern. vii, xxv, xxvii, 22–24, 84–86, 94
- LCPD** local contrast-phase descriptor. 24
- LD** local descriptor. 98–100
- LED** light-emitting diode. 19
- LivDet-Clarkson** Clarkson iris liveness detection 2013. 37–39, 41, 88–91, 93, 95–101
- LivDet-ND** Notre Dame iris liveness detection 2013. xxi, 37–39, 42, 88–91, 93, 95–99, 101
- LivDet-Warsaw** Warsaw iris liveness detection 2013. 37–39, 41, 88–90, 93–99, 101
- log-Gabor** logarithm Gabor. 26, 27, 29, 120, 121, 157, 158, 185, 187
- LPQ** local phase quantization. 24, 86, 98
- LPS** limbic and pupillary boundary segmentation. 67
- LRA** low-rank approximation. xii, 104, 107, 108, 124, 126, 128, 129

- LRM** low rank mean. xvi, xxi, 109, 113–115, 117, 118, 122, 124, 126–130, 146, 209
- MF** mean fusion. 122, 132, 142, 143
- MLDF** multilobe differential filter. 26, 27
- MRF** Markov Random Field. 149
- mRMR** minimum redundancy - maximum relevance. 36
- ND0405** ND-iris-0405. 37–39, 119, 121–123, 125–128, 130, 133–137, 140, 142–145, 155–162, 165–168
- NIR** near-infrared. 4, 6, 8, 17, 18, 37–39, 41–43, 47, 65, 66, 68, 85, 88, 119, 127, 135, 138, 156, 162–164
- PCA** principal component analysis. 106
- PL** power law based weight map. 198–205
- PW** personalized weight map. 198–205
- QWM** quality weighted mean fusion. 132, 142, 143
- r-score** reference score. 71
- R1RA** rank 1 recognition accuracy. xvi–xviii, 139, 142, 159–161, 163–165, 187, 189, 190, 194, 196, 198, 199, 202
- RBF** radial basis function. 20, 22
- RF** refional features. 98–100
- Ric-LBP** ratation invariant co-occurrence of adjacent LBP. 24
- ROC** receiver operating characteristics. xx–xxiii, 3, 4, 122, 123, 127, 131–138, 140, 158–160, 162, 164, 166, 187, 189, 191, 194, 195, 200, 202

RPCA robust principal component analysis. 107, 108, 124, 126–129

RR reflection removal. 67

SCI sparsity concentration index. xxv, 33

SE-LRMF sparse-error low rank matrix factorization. xii, xvi, xxi, xxii, 8, 104, 109–116, 118, 121, 122, 124, 126–129, 131–147, 209

SFFS sequential floating feature selection. 22, 24, 36

SID shift-invariant descriptor. 24, 86, 98

SIFT scale-invariant feature transform. 24, 28

SLIC simple linear iterative clustering. 48

SR silhouette removal. 67

SVM support vector machine. 20–24, 63, 65, 66, 74, 87

UBIRIS2 UBIRIS.v2. 37–40, 46, 47, 53, 56, 57, 64–74, 119, 121–123, 125–128, 130, 133–137, 140–145, 155–168, 184–187, 190–192, 199–201, 205

UEC upper eyelid candidates. 57

WCPH weighted co-occurrence phase histogram. 28, 29

WLD Weber local descriptor. 24

WPT wavelet packets transform. vii, 22

List of Publications

The research in this thesis have produced the following publications:

Journal:

- [J10] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells. Optimal generation of iris codes. *IEEE Transactions on Information Forensics and Security*, vol. 12, page 157-171, 2017.
- [J9] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells. A novel iris weight map method for less constrained iris recognition based on bit stability and discriminability. *Image and Vision Computing*, 2017. (Accepted, in Press)
- [J8] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells. Signal-level information fusion for less constrained iris recognition using sparse-error low rank matrix factorization. *IEEE Transactions on Information Forensics and Security*, vol. 11, page 1549-1564, 2016.
- [J7] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells. Iris liveness detection using regional features. *Pattern Recognition Letters*, vol. 82, page 242-250, 2016.
- [J6] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells. Improving colour iris segmentation using a model selection technique. *Pattern Recognition Letters*, vol. 57, page 24-32, 2015.

Conference:

- [C5] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells. Exploiting stable and discriminative iris weight map for iris recognition under less constrained environment. *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2015.

- [C4] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells. A robust algorithm for colour iris segmentation based on 1-norm regression. *International Joint Conference on Biometrics (IJCB)*, 2014.

Also, some other publications produced during this research period with less relevance to this thesis are as follows:

Journal:

- [J3] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells, Nicolas Ragot, Paul Rodriguez. An online background subtraction algorithm deployed on a NAO humanoid robot based monitoring system. *Robotics and Autonomous Systems*, vol. 85, page 37-47, 2016.

Conference:

- [C2] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells. A study on iris textural correlation. *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2016.
- [C1] **Yang Hu**, Konstantinos Sirlantzis, Gareth Howells, Nicolas Ragot, Paul Rodriguez. An online background subtraction algorithm using a contiguously weighted linear regression model. *European Signal Processing Conference (EU-SIPCO)*, 2015.

Chapter 1

Introduction

1.1 Biometric recognition systems

Biometric recognition systems aim to automatically authenticate humans based on their behavioural and physiological traits, instead of traditional ID cards or passwords [1–3]. Behavioural traits are the attributes related to the behavioural pattern of a person, for example, gait, voice, keystroke, signature, *etc.* Physiological traits are the characteristics related to human body, for example, fingerprint, iris, face, palm print, finger vein, DNA, retina, *etc.* For a biometric recognition system, the selection of biometric traits should consider the following characteristics [1–3]:

- **Universality.** Each individual should have the trait.
- **Distinctiveness.** The trait should be sufficiently discriminating among different individuals.
- **Permanence.** The trait should be sufficiently stable, with no or little change over time.
- **Collectability.** It should be possible to capture and quantise the trait with proper devices.

- **Performance.** The trait should achieve satisfactory accuracy with allowable cost of resources specified by applications.
- **Acceptability.** The users should be willing to present the trait to the system.
- **Circumvention.** The trait should be robust to the fraudulent methods like fake copies.

A biometric recognition system usually consists of two phases: enrolment and recognition. In enrolment phase, the system captures and stores the biometric trait from individuals as gallery templates. In recognition phase, the system establishes the identity of a presented individual by comparing the similarity between the gallery templates and probe capture.

Depending on the application scenarios, the recognition phase can be operated in two modes: identification and verification. In the identification mode, the system aims to determine the identity of a presented subject. This mode performs a one-to-many comparison. The probe capture is compared with all the gallery templates stored in the biometric system to generate a set of similarity scores indicating the similarity between the probe and each gallery template; the identity of the presented subject is determined as the identity of the gallery that corresponds to the highest similarity score. An ‘un-enrolled’ decision may be made if the highest similarity score is lower than a threshold. This mode is usually used in surveillance and forensic applications like criminal investigation [4].

In the verification mode, the system attempts to validate the identity claimed by a presented subject. This mode performs a one-to-one matching. The probe capture of the presented subject is compared with the gallery templates of the claimed identity to produce similarity scores. A decision of match or non-match is made based on a threshold applied to the similarity score. This mode is often used in access control and mobile security applications such as mobile banking [5].

The performance of a biometric recognition system is usually evaluated by cumulative match characteristic (CMC) and receiver operating characteristic (ROC). CMC is used to evaluate the performance of identification mode. It calculates the recognition rate in top k ranked similarity scores. CMC is usually plotted as a curve showing the change of recognition accuracy with respect to rank. An example of CMC curve is shown in Fig. 1.1(a). Rank 1 recognition accuracy is one of the most widely used measures to evaluate the performance of CMC curves; it indicates the recognition accuracy of the best ranked similarity score (*i.e.* $k = 1$).

ROC is used to evaluate the performance of verification mode. Given a threshold to make the decision of match/non-match, two measures can be obtained: false acceptance rate (FAR) and genuine acceptance rate (GAR). FAR calculates the percentage of false matching in inter-class comparisons. GAR calculates the percentage of genuine matching in intra-class comparisons. ROC reflects the trade-off between GAR and FAR when the threshold is varied. It is usually plotted as a curve of GAR vs. FAR. An example of ROC curve is shown in Fig. 1.1(b). Some widely used measures to evaluate the performance of ROC curves are equal error rate (EER), area under curve (AUC) and GAR when FAR is 0.1% or 0.01% (GAR@FAR= 0.1%, GAR@FAR= 0.01%). EER is the rate where FAR equals to $1 - \text{GAR}$ (false non-matching rate). AUC measures the area between a ROC curve and x-axis. The meaning of GAR@FAR= 0.1% and GAR@FAR= 0.01% are straightforward: they are the GAR when FAR is 0.1% and 0.01%, respectively.

1.2 Iris recognition

Iris has been proven to be one of the most reliable biometric traits [6–12]. This is due to some desirable characteristics of human iris. First, the iris pattern is highly unique. For example, it is well known that the left and right iris of a single individual have different patterns [8, 9, 13], and identical twins also have distinctive

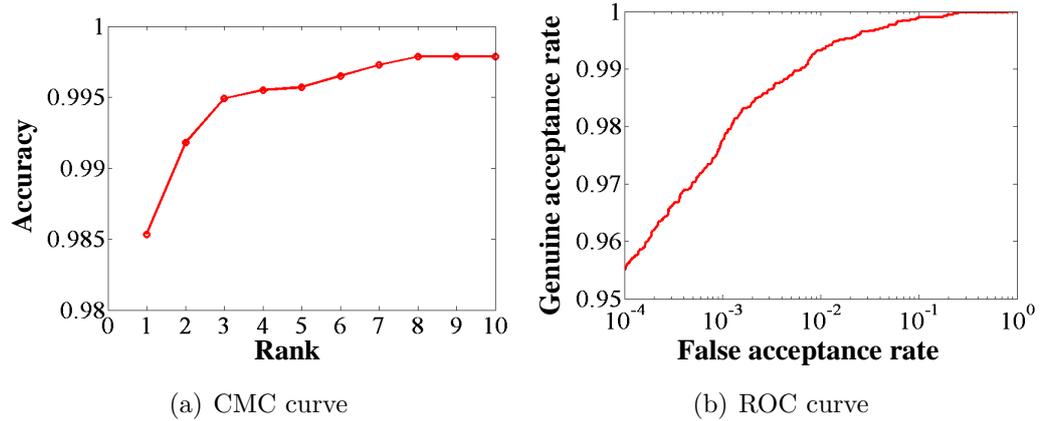


Figure 1.1: Examples of CMC and ROC curves.

iris patterns [1, 13]. Second, the iris pattern is highly stable. It is generally believed that the iris pattern is formed during fetal development and it is stabilised within the first two years of life, then, the pattern of healthy iris changes little during the rest lifespan [1, 8, 9, 13]. Third, iris is an internal organ. This characteristic protects the iris pattern from the influence of environment, enhancing the stability. Fourth, iris is visible externally. Therefore, it is possible to capture the iris texture using proper imaging devices.

The first automatic iris recognition method is proposed by Daugman [6]. Following this work, state-of-the-art iris recognition algorithms have achieved remarkable performance [7–12]. In real applications, current nationwide deployments of iris recognition systems in UAE and India are considered successful, with millions of subjects enrolled [14].

Generally, an iris recognition system usually consists of the following components: image acquisition, liveness detection, iris segmentation, feature extraction and iris matching. Fig. 1.2(a) shows the flowchart of a general iris recognition system.

The image acquisition stage captures the raw sample of eyes, and it quantises the raw sample into digital form for further processing. Fig. 1.2(b) shows an example digital captures of an eye under near-infrared (NIR) wavelength.

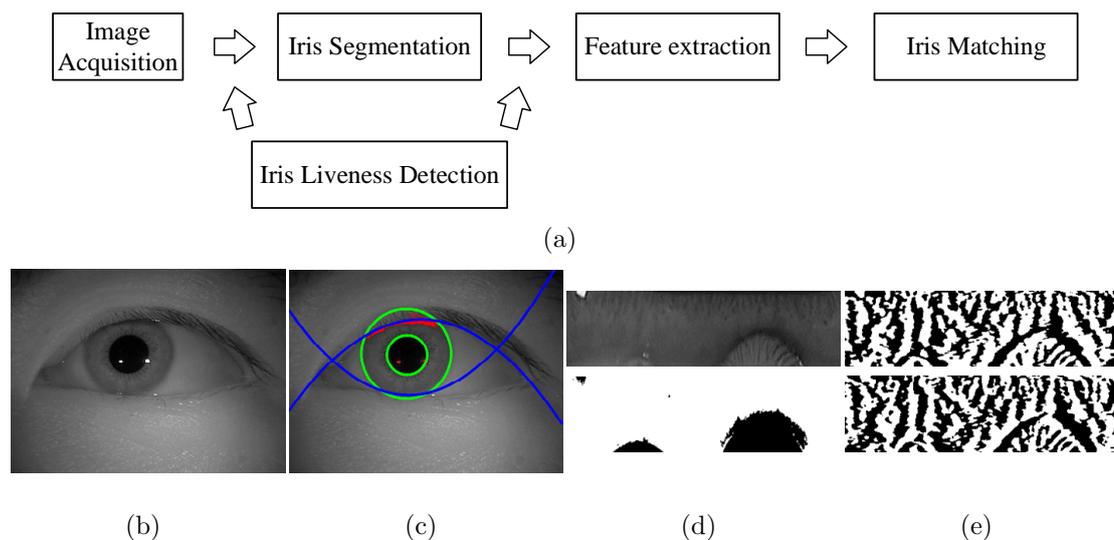


Figure 1.2: Illustrations of a general iris recognition system; (a) the flowchart of the system; (b) example of a captured iris image; (c) the detected iris region in iris segmentation stage; green circles mark the detected limbic and pupillary boundaries; blue curves are the detected eyelids; red regions are the detected reflection and silhouette; (d) the normalised iris region produced after iris segmentation stage; the top image shows the normalised iris region; the bottom image is a mask indicating iris (white) and non-iris (black) pixels in the normalised iris region; (e) the iris code extracted in feature extraction stage using 1-D log-Gabor filter [15] for iris matching; the top image is the real part of the iris code; the bottom image is the imaginary part.

The liveness detection stage protects the iris recognition system from spoofing attacks using synthetically generated iris textures, such as printed photos, textured contact lenses and artificial eyes. This stage is usually considered as a two-class classification problem. Features are extracted from iris captures to distinguish between real and fake iris patterns.

The iris segmentation detects and normalises the iris region for the following feature extraction. It aims to separate the iris region from the other eye regions. The detection of iris region is usually performed by identifying a set of boundaries and regions, including limbic boundary, pupillary boundary, upper and lower eyelids, specular reflection and silhouette. Fig. 1.2(c) shows an example result of iris segmentation. In Fig. 1.2(c), green circles mark the detected limbic and pupillary boundaries; blue curves are the detected eyelids; red regions are the detected reflection and silhouette.

After the detection of iris region, the iris region is usually normalised to improve the invariance of the obtained region to factors like gaze orientation, eye position, eye size and pupil dilation. This normalisation is usually performed using Daugman's rubber sheet model [6, 7]. It maps the segmented iris region from Cartesian coordinates to nonconcentric polar coordinates. Fig. 1.2(d) shows an example of normalised iris region.

The feature extraction stage constructs feature representations from normalised iris region. Currently, binary iris codes [6, 7, 15] are the most widely used features for iris recognition. The binary nature of iris codes brings significant advantage in memory and computational cost, enabling the large scale deployment of iris recognition systems. Fig. 1.2(e) shows an example of iris code extracted using 1-D log-Gabor filter [15].

The iris matching stage compares the features extracted from different captures to produce matching scores for recognition. In identification mode, this stage performs a one-to-many comparison. The feature extracted from a probe capture is compared with all the gallery features stored by the system during the enrolment phase. In verification mode, this stage performs a one-to-one matching. The feature extracted from a probe capture is only compared with the gallery features from the claimed identity.

1.3 Motivation and objectives

To achieve satisfactory performance, most of current iris recognition systems need to apply several constraints to the capturing environment. For example, since iris is a small organ, a close distance is required to obtain captures with sufficient resolution; the subject is expected to stand still to eliminate the noise in the captures caused by motion; the images are captured in near-infrared (NIR) wavelength to

mitigate the influence of specular reflection. These constraints usually require full cooperation of the user. It restricts the usability and practicality of iris recognition technology for some applications with more open capturing environment, such as forensic, surveillance and mobile device security applications.

Recently, in order to further improve the usability and practicality of iris recognition technology for wider applications, significant research efforts have been devoted to allow the iris images to be captured in less constrained environment, with the subject at-a-distance and on-the-move [16–31]. However, such less constrained environment induces noise and degradations in iris captures, and this can significantly deteriorate iris recognition performance [32]. Currently, how to overcome the noise and degradations to improve the performance of iris recognition under less constrained environment remains an open problem.

Therefore, this thesis focuses on developing robust algorithms to improve less constrained iris recognition. The objectives are summarised as follows:

- Investigating a robust iris segmentation algorithm for less constrained iris captures.
- Designing effective features for iris liveness detection.
- Fusing multiple less constrained iris captures to mitigate the noise in such captures.
- Studying the method to produce more effective iris codes for iris recognition.
- Exploiting a robust iris matching algorithm for the features extracted from less constrained iris captures.

1.4 Contributions

This thesis have five main contributions which are summarised as follows.

First, a robust algorithm is proposed for iris segmentation using less constrained captures, based on ℓ_1 -norm regression and model selection technique (Chapter 4). This algorithm formulates the segmentation of iris boundaries as ℓ_1 -norm regression problems. It is shown that the robustness induced by ℓ_1 -norm is effective to mitigate the influence of noise and degradations in less constrained iris captures. Furthermore, the robustness of algorithm is enhanced by a model selection technique. Specifically, multiple segmentation results of iris boundaries are produced by applying ℓ_1 -norm regression to different models, and a model selection method is proposed to select the most reliable result. The experimental results demonstrate that the proposed algorithm is highly effective on both NIR and colour datasets captured in less constrained environments.

Second, regional features are exploited for iris liveness detection (Chapter 5). Regional features are constructed not only using low-level features, but also using high-level feature distribution information. Two models are employed to seek the high-level feature distribution information: spatial pyramid and relational measure. The experimental results on benchmark datasets show that regional features are able to achieve an improved performance in comparison to low-level features which do not include the high-level feature distribution information. The experimental results also demonstrate that, regional feature based method is able to achieve an improved performance compared to state-of-the-art iris liveness detection methods on a dataset with heavier noise, while its performance is comparable to state-of-the-art methods on other datasets.

Third, a signal-level information fusion method is investigated to mitigate the noise in less constrained iris captures (Chapter 6). A sparse-error low-rank matrix factorisation (SE-LRMF) model is proposed to separate noiseless iris structures and noise pixels for information fusion. It is demonstrated that not only the noiseless structure can be straightforwardly fused, but also noise pixels are able to contribute to the signal-level fusion in a weighted fusion scheme. The experimental results on

benchmark datasets show that this signal-level fusion method is able to achieve a generally improved iris recognition performance for less constrained iris captures, in comparison to existing iris recognition algorithms, especially for the captures with heavy noise and low quality.

Fourth, the optimal generation of iris codes is studied (Chapter 7). The traditional method to generate iris codes is formulated as an optimisation problem. Therefore, more effective iris codes can be generated by designing effective objective terms to add to this optimisation problem. This thesis designs an additional objective term modelling the spatial correlation in iris codes, using the prior of iris textural correlations. Experimental results on benchmark datasets show that the iris codes generated with this additional objective term achieve a generally improved iris recognition performance in comparison to Masek's implementation of traditional iris code generation method [15]. The observation of spatial correlation of iris textures in previous research is also verified in this experiment.

Fifth, an iris weight map method is exploited for robust iris matching (Chapter 8). This iris weight map is the combination of a stability map modelling intra-class bit stability and a discriminability map expressing inter-class bit discriminability in iris codes. It emphasises the bits with both high stability and high discriminability for iris matching. Experimental results show that this iris weight map is highly effective for both single-sensor and cross-sensor iris captures in less constrained environments, achieving a generally improved performance compared to existing iris weight map methods.

1.5 Thesis structure

This thesis is organised as follows. Chapter 2 reviews the previous works on iris recognition. Chapter 3 introduces the datasets used for experimental analysis in this

thesis. Chapter 4 describes the proposed algorithm for iris segmentation. Chapter 5 presents the proposed method for iris liveness detection. Chapter 6 details the proposed method for signal-level information fusion. Chapter 7 presents the proposed method to generate optimal iris codes. Chapter 8 describes the proposed iris weight map for robust iris matching. Chapter 9 summarises the concluding remarks and discusses the future work.

Chapter 2

Previous works on iris recognition

This chapter revisits the previous works on iris recognition. Since this thesis focuses on algorithms rather than hardware design, this chapter concentrates on the following four stages of iris recognition: iris segmentation, iris liveness detection, feature extraction and iris matching.

2.1 Iris segmentation

Iris segmentation detects and normalises the iris region. It is one of the fundamental steps for iris recognition. This section reviews some remarkable methods developed for iris segmentation in the literature.

2.1.1 Integro-differential operator

The first algorithm for automatic iris segmentation is proposed by Daugman in [6]. This algorithm models the limbic boundary using a circle with three parameters x_c , y_c , and r . x_c and y_c denote the coordinate of circle centre, while r denotes the radius of circle. Let $I(x, y)$ be an iris capture. The three parameters x_c , y_c , and r

are sought by maximizing an integro-differential operator:

$$\max_{r, x_c, y_c} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r, x_c, y_c} \frac{I(x, y)}{2\pi r} ds \right| \quad (2.1)$$

where $G_\sigma(r)$ is a Gaussian smoothing function with a standard deviation of σ , ds is an element of circular arc, and $*$ is the convolution operator. Eqn. 2.1 is essentially a circular edge detector. It seeks the limbic boundary as the circle with the highest inner-to-outer contrast at the boundary. Within the limbic boundary, the same method is used to seek the pupillary boundary.

After limbic and pupillary boundaries are found, [6] performs iris normalisation using a rubber sheet model. The aim is to improve the invariance of obtained iris region to factors like pupil dilation, gaze direction and eye size. This normalisation is a mapping from Cartesian coordinate system (x, y) to non-concentric polar coordinates (r, θ) where $r \in [0, 1]$ and $\theta \in [0, 2\pi]$. The mapping can be represented as follows:

$$I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta) \quad (2.2)$$

and

$$\begin{cases} x(r, \theta) = (1 - r)x_p(\theta) + rx_s(\theta) \\ y(r, \theta) = (1 - r)y_p(\theta) + ry_s(\theta) \end{cases} \quad (2.3)$$

where $(x_p(\theta), y_p(\theta))$ denotes the Cartesian coordinate of the point on pupillary boundary along the direction θ , and $(x_s(\theta), y_s(\theta))$ denotes the Cartesian coordinate of the point on limbic boundary along the direction θ . An example of iris normalisation with rubber sheet model is shown in Fig. 1.2(d).

In the works following [6], the iris normalisation is usually performed using the above rubber sheet model, and the variation of methodology mainly exists in the detection of iris region.

2.1.2 Hough Transform

In [8], Wildes proposes to perform iris segmentation using Hough transform [33]. This method firstly detects edge points in iris captures, then the detected edge points vote to find the boundary parameters. For example, for the limbic boundary which is modelled using a circle, the Hough transform is formulated as follows:

$$H(x_c, y_c, r) = \sum_{i=1}^n h(x_i, y_i, x_c, y_c, r) \quad (2.4)$$

where x_c , y_c and r are circle parameters as in Eqn. 2.1, and (x_i, y_i) , $i = 1, 2, \dots, n$ is the set of edge points. $h(x_i, y_i, x_c, y_c, r)$ in Eqn. 2.4 is defined as follows:

$$h(x_i, y_i, x_c, y_c, r) = \begin{cases} 1, & \text{if } (x_i - x_c)^2 + (y_i - y_c)^2 - r^2 = 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.5)$$

The combination of x_c , y_c and r that maximizes $H(x_c, y_c, r)$ is determined as the parameters for the limbic boundary. In other words, this method seeks the limbic boundary as the circle that passes the largest number of edge points. The pupillary boundary is sought using a similar method. Also, upper and lower eyelids are fitted as parabolic arcs in similar ways.

2.1.3 Fourier active contour

In [34], Daugman uses an active contour model based on Fourier series to improve the previous algorithm in [6]. To determine limbic and pupillary boundaries, this method samples radial edge points from N uniformly spaced angular directions. It leads to N edge points denoted by $\{r_\theta\}_{\theta=0}^{N-1}$. The M Fourier coefficients $\{C_k\}_{k=0}^{M-1}$

for $\{r_\theta\}_{\theta=0}^{N-1}$ can be calculated by:

$$C_k = \sum_{\theta=0}^{N-1} r_\theta e^{-2\pi i k \theta / N} \quad (2.6)$$

Therefore, the potential boundary can be approximated by inverse discrete Fourier transform as follows:

$$R_\theta = \frac{1}{N} \sum_{k=0}^{M-1} C_k e^{2\pi i k \theta / N} \quad (2.7)$$

where $\{R_\theta\}_{\theta=0}^{N-1}$ are the approximated boundary points. This method detects eyelashes by testing if the lower tail of the intensity histogram in iris region supports a hypothesis of multimodal mixing. It calculates a difference histogram between the histograms of upper and lower iris regions. A threshold to eliminate the eyelashes is determined such that the hypothesis tests support that the lower tail up to this threshold is separable and sufficiently different between the difference histogram and histogram obtained from all the iris pixels.

2.1.4 Geodesic active contour

In [35], the limbic boundary is sought using a geodesic active contour model. This method gravitates a curvature inside the iris towards possible limbic boundary. Let $\gamma(t)$ be the curvature at time t . Let $I(x, y)$ be an iris capture. Let $\psi(x, y)$ be the signed distance between the pixel at (x, y) and $\gamma(t)$. $\psi(x, y)$ satisfies that:

$$\psi(x, y) = \begin{cases} 0, & \text{if } (x, y) \text{ is on } \gamma(t) \\ > 0, & \text{if } (x, y) \text{ is inside } \gamma(t) \\ < 0, & \text{if } (x, y) \text{ is outside } \gamma(t) \end{cases} \quad (2.8)$$

In other words, $\gamma(t)$ is the zero level set of $\psi(x, y)$. According to the derivation in [35], $\psi(x, y)$ can be evolved as follows to gravitate its zero level set (*i.e.* $\gamma(t)$)

towards the potential boundary:

$$\psi^t = -K (c + \epsilon\kappa) \|\nabla\psi\| + \nabla\psi\nabla K \quad (2.9)$$

where c is the velocity of the evolution and ϵ is the smoothness degree of the level sets. κ is the curvature of the level sets computed by:

$$\kappa = -\frac{\psi_{xx}\psi_y^2 - 2\psi_x\psi_y\psi_{xy} + \psi_{yy}\psi_x^2}{(\psi_x^2 + \psi_y^2)^{\frac{3}{2}}} \quad (2.10)$$

where ψ_x and ψ_y are first-order gradient of the image along x and y direction respectively, and ψ_{xx} , ψ_{yy} and ψ_{xy} are second-order gradients based on x and y directions. K is the stopping term which decelerates the evolution near the boundaries:

$$K(x, y) = \left[1 + \left(\frac{\|\nabla(G(x, y) * I(x, y))\|}{k} \right)^\alpha \right]^{-1} \quad (2.11)$$

where k and α are constants, and $G(x, y)$ is a smoothing function. After convergence, the zero level set of $\psi(x, y)$ is considered as the limbic boundary.

2.1.5 Pulling and pushing method

In [36], a pulling and pushing method is proposed to find limbic and pupillary boundaries. This method is based on Hooke's law on the force of spring [37]. Firstly, it uses a Adaboost-cascade iris detector to locate an initial iris centre O . Then, edge detection is performed along N radial directions originated from O , obtaining N boundary points $\{P_i\}_{i=0}^N$. Finally, each segment OP_i is viewed as a spring. The composition force of all the N springs directs O towards the equilibrium position which is viewed as potential iris centre. The composition force is obtained by:

$$\vec{F} = \frac{1}{N} \sum_{i=0}^{N-1} \vec{f}_i = \frac{1}{N} \sum_{i=0}^{N-1} -k(R - r_i) \vec{e}_i \quad (2.12)$$

where k is a constant; R is the equilibrium length of all the springs; r_i is the current length of the i^{th} spring; \vec{e}_i is the direction of the i^{th} spring initiated from O . R is estimated as follows:

$$\vec{R} = \frac{1}{N} \sum_{i=0}^{N-1} \overline{OP}_i \quad (2.13)$$

As a result, O is pulled and pushed towards the equilibrium position. The algorithm iterates between pulling/pushing O and detecting boundary points until convergence. The eyelids are detected by fitting parabola curves to candidate points. The candidate points are selected by performing vertical edge detection on the image processed by 1-D rank filtering, and noisy candidate points are eliminated by comparing the similarity between the detected candidate points and three generic eyelids models learned from training data. A statistically learned prediction model is used to eliminate silhouette. This model calculates the dissimilarity between the histogram of candidate silhouette region and iris region, and it uses a cubic polynomial curve to model the relationship between the dissimilarity and the percentage of silhouette.

2.1.6 Integro-differential constellation method

In [38], Tan *et al.* propose an integro-differential constellation method to determine the limbic boundary of iris images captured in visible wavelength which induces heavier noise in captures. Beginning with an initial point, this algorithm iteratively seeks the path to maximize the score of integro-differential operator (Eqn. 2.1) along three rings (constellations). Although it is possible that this method may be trapped in local optima, the experimental analysis shows excellent convergence.

2.1.7 Colour and sclera features based classification

In [39], Proença proposes a classification based method to perform iris segmentation using less constrained iris captures. This method extracts features from position, colour and sclera information to distinguish between iris and non-iris pixels. In terms of position, the coordinate of each pixel is directly used as features. Also, colour features are extracted from hue, blue and red chroma colour components. The mean and standard deviation of each colour component in rectangular regions centred at each pixel with different radius are used as features. On the other hand, sclera features are calculated as the portion of sclera pixels along a given direction with respect to the reference pixel. Sclera region is detected by training a neural network classifier using the above described position and colour features. These features are concatenated to train a neural network to categorize iris and non-iris pixels

Given the detected iris pixels, the limbic and pupillary boundaries are parameterized using polynomial fitting in polar coordinate system. Although some features in this method are extracted from colour channels, this method experimentally shows good adaptability to less constrained NIR captures.

2.1.8 Localised Zernike moments based classification

In [21], localised Zernike moments [40] are exploited as features to distinguish between iris and non-iris pixels for less constrained iris captures in both NIR and visible wavelength. The Zernike moments with order $m \in \mathbb{N}$ and repetition $n \in \mathbb{Z}$ in a local region l can be calculated as follows:

$$Z_{mn}^l = \frac{m+1}{\pi} \sum_{r \in l} \sum_{\theta \in l} f(r, \theta) [V_{mn}(r, \theta)]^* \quad (2.14)$$

where $[V_{mn}(r, \theta)]^*$ is the complex conjugate of $V_{mn}(r, \theta)$, $f(r, \theta)$ is the extracted local region mapped to a unit circle in polar coordinate system. $V_{mn}(r, \theta)$ is Zernike polynomial defined as follows in polar coordinate system:

$$V_{mn}(r, \theta) = R_{mn}(r) e^{in\theta} \quad (2.15)$$

where $R_{mn}(r)$ is the orthogonal radial polynomial defined as follows:

$$R_{mn}(r) = \sum_{s=0}^{(m-|n|)/2} (-1)^s \frac{(m-s)!}{s! \left(\frac{m+|n|}{2} - s\right)! \left(\frac{m-|n|}{2} - s\right)!} r^{m-2s} \quad (2.16)$$

This method constructs features using Zernike moments together with position and intensity of pixels. Two types of classifier are trained to distinguish iris and non-iris pixels: neural network and support vector machine.

In this method, eyelids are detected by fitting second-order polynomial curves to candidate points; the candidate points are determined as vertical edge points located in low intensity regions. The reflection region is detected by propagating from some seeds determined as the pixels within the iris region and with high intensity. The silhouette is detected based on a threshold calculated using the location of peak of two histograms corresponding to candidate iris region and candidate silhouette region.

2.1.9 Random walker algorithm and post-processing

In [41], a fast iris segmentation algorithm is proposed for less constraint captures in both NIR and visible wavelength. Firstly, a random walker algorithm [42] is used to roughly locate the iris region. Then, based on the rough iris location, rough segmentation is performed using Hough transform based on the Edge map. Next, the rough segmentation is refined by thresholding to distinguish between sclera and iris region

at the boundary of rough segmentation. The threshold is derived from the statistical intensity information (mean and standard deviation of intensity) in regions around the boundary sought by rough segmentation. After that, eyelids are fitted. Eyelids candidates are detected using edge detection; noisy candidates are eliminated by comparing all candidate points to three pre-defined models and removing the candidates deviating too much from the best fitted model; the rest candidate are used to fit second-order polynomial curves as eyelids. Finally, silhouette is eliminated using an adaptive threshold derived from the statistical intensity information (mean and standard deviation of intensity) of the lower half of iris region which is expected to be more noise-free.

2.2 Iris liveness detection

Iris liveness detection can be considered as a two-class classification problem. The key of this problem is to construct effective features to distinguish between real and fake iris captures. This section reviews some promising approaches for iris liveness detection in the literature.

2.2.1 Brightness variation based thresholding

In [43], brightness variation is used as features for iris liveness detection tested on attacks using printed fake iris patterns. The variation of brightness in iris region is induced by weak flashlight generated by two white light-emitting diodes (LEDs). After the flashlight illumination, the variation rate of average brightness (denoted by V) in two regions at both sides of iris is used as the feature:

$$V = \frac{|P(t_0) - \min(P(t))|}{P(t_0)} \quad (2.17)$$

where $P(t_0)$ is the average brightness in the two regions immediately after the flash light illumination; $P(t)$ is the average brightness in the two regions at time t ; 10 frames are considered after the the flash light illumination. A threshold is applied to V to distinguish between real and fake printed iris patterns.

2.2.2 Gray level co-occurrence matrix (GLCM) based classification

In [44], GLCM [45] is exploited for iris liveness detection tested against fake patterns induced by colour contact lenses. GLCM calculates the distribution of co-occurring intensity values at a given offset in an image. Let $p(i, j)$ be an element of a GLCM. $p(i, j)$ calculates the frequency that two pixels separated by a distance occur, with one pixel having a gray level of i and the other pixel having a gray level of j . [44] uses the contrast (denoted by con) and angular second moment (denoted by asm) of GLCM extracted from the lower half of normalised iris images as GLCM based features:

$$con = \sum_{i=1}^N \sum_{j=1}^N (i - j)^2 p(i, j) \quad (2.18)$$

$$asm = \sum_{i=1}^N \sum_{j=1}^N p(i, j)^2 \quad (2.19)$$

where N is the number of elements in a GLCM. These GLCM based features together with the mean and standard deviation of the lower half of normalised iris images are used to trained a support vector machine (SVM) with radial basis function (RBF) kernel to classify real and fake iris patterns.

2.2.3 Textural features based classification

In [46], three features are investigated for iris liveness detection: iris edge sharpness, iris-texton histogram and Gray level co-occurrence matrix (GLCM). Iris edge

sharpness (denoted by IES) computes the contrast at the iris edge (transition from sclera to iris region):

$$IES = \sum_{\theta=0}^{2\pi} (I(r_{i+\xi}, \theta) - I(r_{i-\xi}, \theta)) \quad (2.20)$$

where $I(r, \theta)$ denotes an iris image in polar coordinate system, r_i is the radius of iris and θ is the angle of a point in the limbic boundary. Therefore, $I(r_{i+\xi}, \theta)$ and $I(r_{i-\xi}, \theta)$ are pixels in sclera and iris regions, respectively. A threshold is applied to IES to classify real and fake iris patterns.

As for iris-texton histogram, iris-texton vocabularies are firstly produced by performing K-means clustering on features obtained by passing some training images through Gabor filter banks [6]. Then, given the iris-texton vocabularies and a new image, histogram representations are obtained by assigning the Gabor features of the new image to the nearest texton vocabulary and counting the occurrence of each vocabulary. For this feature, a SVM is trained as the classifier.

As for GLCM, three features are extracted from GLCM: inverse difference moment (f_{idm}), sum average (f_{sa}) and sum entropy (f_{se}). The definitions of the features are as follows:

$$f_{idm} = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j) \quad (2.21)$$

$$f_{sa} = \sum_{i=2}^{2N_g} i p_{x+y}(i) \quad (2.22)$$

$$f_{se} = - \sum_{i=2}^{2N_g} p_{x+y}(i) \log(p_{x+y}(i)) \quad (2.23)$$

where $p(i, j)$ is the GLCM of an image; N_g is the number of gray levels; $p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)$, $k = i + j = 2, 3, \dots, 2N_g$. For these features, a SVM is trained as the classifier.

Experimentally, the performance of this method is tested against fake iris patterns

induced by contact lenses.

2.2.4 Wavelet packets transform (WPT) based classification

In [47], WPT [48] is used to construct features for iris liveness detection. WPT decomposes an image into 4 output subimages, including approximation, horizontal detail, vertical detail and diagonal detail coefficients respectively. These subimages can be further decomposed into the same 4 components, constructing a multilevel representation. [47] uses the standard deviation of multilevel horizontal detail, vertical detail and diagonal detail coefficients as features. A SVM with RBF kernel is used as the classifier. This method is tested against fake iris printouts.

2.2.5 Quality measures based classification

In [49], iris quality measures are exploited as features for iris liveness detection. The investigated quality measures include blur [7, 50, 51], motion [9, 50, 52, 53], occlusion [10, 50, 53, 54], global and local contrast [49], pupil dilation [55]. Sequential floating feature selection (SFFS) [56] is used to select the best combination of these quality measures. The classification is performed using a standard quadratic classifier fitting the training data using multivariate normal densities with diagonal covariance estimates stratified by group. This method is tested on the data simulating the spoofing attack using high quality iris printouts.

2.2.6 Multi-resolution local binary pattern (LBP) and Adaboost based classification

In [57], multi-resolution LBP [58] and Adaboost are used to produce effective features for iris liveness detection [59]. This method produces a feature pool including

multi-resolution LBP features. Then, Adaboost learning is adopted to select the most discriminative subset from the feature pool and to classify the selected features. The Adaboost learning is performed by three components: the weak learner, the component classifiers and the re-weighting function. The weak learner selects the best features from the feature pool; the component classifier is learned from the current selected features to distinguish real and fake samples; the re-weighting function weights the training samples to emphasize hard negative samples for the subsequent classifier training. This work focuses on the detection of spoofing attacks using colour contact lenses, although it includes some iris printouts and glass eyes in the test data.

2.2.7 Gaze features based classification

In [60], features are extracted from the gaze information in the process of eye tracking for iris liveness detection. This method decomposes the eye movement signal into a set of elementary units reflecting eye micro-movements. For each elementary unit, the centroid of its horizontal and vertical components is used to represent the corresponding local position characteristics. The final features are formed using the mean and standard deviation of the centroid of horizontal and vertical components calculated from all the elementary units. A SVM is used as the classifier. The performance of this work is examined using fake iris printouts.

2.2.8 Quality and texture based features for colour captures

In [61], several features are exploited for iris liveness detection on mobile devices with colour captures. Several traditional features used in iris liveness detection are implemented and their effectiveness on mobile iris captures is tested. The implemented features include high frequency power in [7], local and global contrast in [49], frequency distribution rates in [10], and GLCM based features in [44]. The optimal

subset of these features are selected using SFFS [56]. Three classifiers are tested: discriminant analysis (DA), k-nearest neighbour (KNN) and SVM. This method is tested on data simulating two types of spoofing attack: iris printouts and contact lenses.

2.2.9 Local descriptors based classification

In [62], a variety of local descriptors are investigated as features for iris liveness detection. Based on the feature coding step, the investigated local descriptors are categorized into independently quantized features and jointly quantized features. The independently quantized features studied in this work include LBP [58], co-occurrence of adjacent LBP (CoA-LBP) [63], rotation invariant co-occurrence of adjacent LBP (Ric-LBP) [64], local phase quantization (LPQ) [65], Weber local descriptor (WLD) [66], local contrast-phase descriptor (LCPD) [67] and binarized statistical image features (BSIF) [68]. The investigated jointly quantized features consist of scale-invariant feature transform (SIFT) [69], DAISY [70] and shift-invariant descriptor (SID) [71, 72]. The classification is performed using a linear SVM. In this work, experiments are conducted for spoofing attacks on three modalities: fingerprint, iris and face. For the case of iris, this method is tested on several datasets using fake iris printouts and cosmetic contact lenses for spoofing attack.

2.2.10 Pupil Dynamic features based classification

In [73], dynamics of the pupil under visible light stimuli is exploited for iris liveness detection. The model in [74] is adopted to describe the pupil constriction and dilation during the stimuli. The parameters of the model are determined by least-squares curve fitting. The fitted parameters are used as features. Linear and non-linear SVMs are used as classifiers. This work views odd pupil dynamics as unlive cases with possible spoofing attack. To construct dataset for test, it use videos of

spontaneous oscillations of pupil as the case of odd pupil dynamics with possible spoofing attack, and it uses videos of pupil reaction under the stimulate of a visible light as the case of normal pupil reaction which means there is no spoofing attack.

2.2.11 Deep learning based method

In [75], iris liveness detection is performed based on deep learning technique [76]. Two deep learning approaches are investigated: architecture optimization (AO) and filter optimization (FO). AO focuses on seeking a suitable architecture for the convolutional neural network, while FO aims to learn optimized weights for filters in the convolutional neural network. In [75], AO and FO are not only evaluated separately, but also combined to build the liveness detection systems. This work is tested on fake iris capture produced by printing. This method is tested for spoofing attacks on 3 modalities: fingerprint, iris and face. For the case of iris, this method is tested on several datasets using fake iris printouts for spoofing attack.

2.3 Feature extraction

The features used for iris recognition can be categorized into binary and non-binary features. This section reviews some representative binary and non-binary features in the literature.

2.3.1 Binary features

Gabor feature. In [6, 7], Daugman proposes to use iris codes produced based on 2-dimensional (2-D) Gabor filters [77, 78] for iris recognition. A 2-D Gabor filter

over the image domain (x, y) is formulated as follows:

$$G(x, y) = \exp \left\{ -\pi \left[\frac{(x - x_0)^2}{\alpha^2} + \frac{(y - y_0)^2}{\beta^2} \right] \right\} \cdot \exp \{ -2\pi i [\mu_0 (x - x_0) + \nu_0 (y - y_0)] \} \quad (2.24)$$

where (x_0, y_0) is the centre of Gabor filter; α and β specify the effective width and height of the filter, respectively; μ_0 and ν_0 determine the modulation with the spatial frequency $\omega_0 = \sqrt{\mu_0^2 + \nu_0^2}$ and direction $\theta_0 = \arctan(\nu_0/\mu_0)$. Given a normalised iris image, Gabor filters are applied to this image, and binary iris codes are produced based on the phase of filtering results. Specifically, a phase is encoded into 11 if this phase falls in the first quadrant, and the code is 01, 00, 10 if the phase falls into the second, third and fourth quadrants, respectively. In other words, the iris codes are generated based on the sign of real and imaginary parts of Gabor filtering results.

Logarithm Gabor (log-Gabor) features. In [15], the log-Gabor filter is used as an alternative of Gabor filters to produce iris codes. The frequency response of a log-Gabor filter is a Gaussian on a logarithm frequency axis; the formulation is as follows:

$$G(f) = \exp \left\{ \frac{-[\log(f/f_0)]^2}{2[\log(\sigma/f_0)]^2} \right\} \quad (2.25)$$

where f_0 gives the centre frequency, and σ is the bandwidth of the filter. Log-Gabor filters have some desirable characteristics. For example, as studied in [79], natural images can be better coded by filters that have Gaussian transfer functions when viewed on the logarithmic frequency scale; log-Gabor filters have no direct current (DC) components *etc.* The procedure of iris code generation with log-Gabor filters is similar to that with Gabor filters. The log-Gabor filter is applied to normalised iris images, and the phase of filtering results is quantized based on the quadrant they fall in.

Ordinal features. In [20], Sun and Tan propose to use ordinal features to generate iris codes. Multilobe differential filters (MLDFs) are used to extract ordinal features.

MLDF is defined as follows:

$$MLDF = C_p \sum_{i=1}^{N_p} \frac{1}{\sqrt{2\pi\delta_{pi}}} \exp \left[\frac{-(X - \mu_{pi})^2}{2\delta_{pi}^2} \right] - C_n \sum_{j=1}^{N_n} \frac{1}{\sqrt{2\pi\delta_{nj}}} \exp \left[\frac{-(X - \mu_{nj})^2}{2\delta_{nj}^2} \right] \quad (2.26)$$

where μ and σ are the mean and standard deviation of 2-D Gaussian filters; N_p denotes the number of positive filters; N_n denotes the number of negative filters; C_p and C_n are constants which ensure the zeros sum of MLDF by satisfying $C_p N_p = C_n N_n$. To produce iris codes, MLDF is applied to normalised iris images, and the filtering results are quantized based on their signs.

Geometric key (geoKey) based features. In [80], iris images are encoded using geoKey for less constrained iris recognition. The geoKey $\mathbf{K} = \{\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_d\}$ is a set of random coordinate pairs generated as follows:

$$\mathbf{K}_i = \left\{ (\mathbf{x}_1, \mathbf{x}_2) \sim \text{i.i.d. } G \left(0, \frac{1}{5}B \right) \right\} \quad (2.27)$$

where \mathbf{x}_1 and \mathbf{x}_2 are a pair of coordinates; B is the size of image patch; i.i.d. $G \left(0, \frac{1}{5}B \right)$ is independent and identical Gaussian distribution with a mean of 0 and a standard deviation of $\frac{1}{5}B$. Given the geoKey, binary codes are produced by:

$$f(\mathbf{w}; \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{K}) = \begin{cases} 1 & \text{if } L(\mathbf{w}, \mathbf{x}_1) < L(\mathbf{w}, \mathbf{x}_2) \\ 0 & \text{otherwise} \end{cases} \quad (2.28)$$

where \mathbf{w} is a local window; $L(\mathbf{w}, \mathbf{x}_i)$ is the average value of the log-Gabor filter response within the window \mathbf{w} at position \mathbf{x}_i . The geoKey is personalised, *i.e.* it is generated specifically for each enrolled subject. To perform iris recognition, the geoKey based binary features are combined with 1-dimensional (1-D) log-Gabor based iris codes [15] at the score level.

2.3.2 Non-binary features

Ma *et al.* features. In [10], iris features are extracted using spatial filters defined as follows:

$$G(x, y, f) = \frac{1}{2\pi\delta_x\delta_y} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\delta_x^2} + \frac{y^2}{\delta_y^2} \right) \right] \cos \left[2\pi f \left(\sqrt{x^2 + y^2} \right) \right] \quad (2.29)$$

where f is the frequency of modulation function; δ_x and δ_y specify the effective width and height of the filter, respectively. Differently from the 2-D Gabor filters defined in Eqn. 2.24 which has a fixed direction, the spatial filter in Eqn. 2.29 considers the information in all the directions. To extract features, this filter is applied to normalised iris image; the filtering result is partitioned into small blocks; the mean and standard deviation of the magnitude in each block are concatenated to construct features.

Tan *et al.* features. In [81], multiple non-binary features are extracted from iris and eye regions, and these non-binary features are fused with binary ordinal codes [20] (Eqn. 2.26) and semantic information for colour iris recognition. The used non-binary features include colour histogram from iris regions and texton histogram from eye regions. Colour histogram is the concatenation of histograms obtained from RGB, HSI and $l\alpha\beta$ [82] colour spaces. The texton histogram is obtained based on densely extracted SIFT descriptors [69], using similar method as described in section 2.2.3. The semantic information is extracted based on two difference filters applied to the regions around upper eyelid at both sides of iris.

Weighted co-occurrence phase histogram (WCPH). In [83], WCPH is proposed as features for colour iris recognition. Let g_z and θ_z be the gradient magnitude and phase angle at position z in a normalised iris image. Let $p_d(\mu, \nu)$ be the bin (μ, ν) of a WCPH which considers the positions with a distance of d . $p_d(\mu, \nu)$ is

defined as follows:

$$p_d(\mu, \nu) = \frac{1}{C_d} \sum_{\|z-z'\|=d} q(\mu, \theta_z) q(\nu, \theta_{z'}) \quad (2.30)$$

where $q(\mu, \theta_z)$ is calculated by:

$$q(\mu, \theta_z) = \begin{cases} g_z \cos^n(\theta_z - \alpha_\mu) & \text{if } |\theta_z - \alpha_\mu| < \pi/2 \\ 0 & \text{otherwise} \end{cases} \quad (2.31)$$

where $n > 1$ is a parameter and α_μ is the centre of bin μ . Essentially, in a WCPH, the gradient phase angle at a pixel contributes to multiple neighbouring bins, and the quantity of contribution is determined using the gradient magnitude and the angular difference between the gradient phase angle and each bin.

Zernike moments phase features. In [23], the phase of localised Zernike moments [40] are extracted as features for iris recognition under less constrained environment. The method to calculate Zernike moments is the same as that described in section 2.1.8. The extracted features are fused at the score level with the matching score of 1-D log-Gabor based iris codes [15] with an iris weight map [22].

Deep representation. In [84], deep convolutional neural network is used to learn features instead of extracting features based on manually designed filters. The iris recognition is performed based on a deep convolutional neural network using iris image pairs as input. This network begins with a pairwise filter layer, followed by several pooling layers, normalisation layers and local layers. It ends up with two fully connected layers served as the classifier. The parameters of this network are trained using back-propagation [85]. This method shows excellent performance in the task of heterogeneous iris recognition.

2.4 Iris matching

Given two features from two iris captures, iris matching compares the two features to produce an iris matching score indicating their distance or similarity. This section reviews several representative methods for iris matching in the literature.

2.4.1 General distance metrics based methods

Depending on the type of feature, various distance metrics have been employed to perform iris matching. Some representative methods are as follows.

Hamming distance. For binary features like [6, 7, 15, 20, 80], Hamming distance is widely used for iris matching [6, 7]. Given two binary features with n bits denoted by $\mathbf{C}^i \in \{0, 1\}^n$ and $\mathbf{C}^j \in \{0, 1\}^n$ (note that, in here and the rest of this thesis, it is assumed the bits or features corresponding to the noise like reflections have been eliminated from iris codes for iris matching), Hamming distance is defined as follows:

$$\text{HD} = \frac{1}{n} \sum_{k=1}^n \mathbf{C}_k^i \oplus \mathbf{C}_k^j \quad (2.32)$$

where \oplus is element-wise exclusive-or operator; \mathbf{C}_k^i and \mathbf{C}_k^j are the k^{th} bits in the two iris codes.

ℓ_2 distance. ℓ_2 distance is one of the most widely used measures to represent the distance between two non-binary feature vectors. It is adopted for iris matching in methods such as [10, 23]. Given two feature vectors $\mathbf{f}_i \in \mathbb{R}^n$ and $\mathbf{f}_j \in \mathbb{R}^n$, ℓ_2 distance calculates the ℓ_2 -norm of the difference between the two feature vectors:

$$D_{\ell_2} = \|\mathbf{f}_i - \mathbf{f}_j\|_2 \quad (2.33)$$

ℓ_1 distance. Some research have shown that ℓ_1 -norm is able to achieve better robustness [86–90]. Therefore, ℓ_1 distance is employed for iris matching in literatures like [10]. The ℓ_1 distance between two feature vectors $\mathbf{f}_i \in \mathbb{R}^n$ and $\mathbf{f}_j \in \mathbb{R}^n$ is defined as follows:

$$D_{\ell_1} = \|\mathbf{f}_i - \mathbf{f}_j\|_1 \quad (2.34)$$

where $\|\bullet\|_1$ is the ℓ_1 -norm.

Cosine distance. Cosine distance is an angular similarity measure. It is used for iris matching in research such as [10]. The cosine distance between two feature vectors \mathbf{f}_i and \mathbf{f}_j is calculated as follows:

$$D_{\cos} = 1 - \frac{\mathbf{f}_i^T \mathbf{f}_j}{\|\mathbf{f}_i\|_2 \|\mathbf{f}_j\|_2} \quad (2.35)$$

Diffusion distance. In [81], diffusion distance [91] is used to measure the similarity of two colour histogram features \mathbf{h}_1 and \mathbf{h}_2 . The diffusion distance is defined as follows:

$$D_{\text{diffusion}} = \sum_{i=0}^L k(|\mathbf{d}_i(\mathbf{h}_1, \mathbf{h}_2)|) \quad (2.36)$$

where $l = 1, 2, \dots, L$ is the number of pyramid layers; $\mathbf{d}_0(x) = \mathbf{h}_1 - \mathbf{h}_2$; $\mathbf{d}_l(x) = [\mathbf{d}_{l-1}(x) * \phi(x, \delta)] \downarrow_2$; δ is the standard deviation of Gaussian filter ϕ ; \downarrow_2 denotes half-size downsampling; $k(\bullet)$ is a distance metric which is selected to be ℓ_1 -norm in [81].

χ^2 distance. In [81], χ^2 distance is employed to match texon histogram features. Let \mathbf{h}_1 and \mathbf{h}_2 be two histogram features, χ^2 distance is formulated as follows:

$$D_{\chi^2} = \sum_i \frac{[\mathbf{h}_1(i) - \mathbf{h}_2(i)]^2}{\mathbf{h}_1(i) + \mathbf{h}_2(i)} \quad (2.37)$$

where $\mathbf{h}_1(i)$ and $\mathbf{h}_2(i)$ are the i^{th} bin of the two histograms.

Bhattacharyya distance. In [83], Bhattacharyya distance is employed to match co-occurrence histogram features due to its desirable characteristics in Bayes error, symmetry and stability [92]. The Bhattacharyya distance between two co-occurrence histogram features $p(\mu, \nu)$ and $q(\mu, \nu)$ is formulated as follows:

$$D_b = \left(1 - \sum_{\mu} \sum_{\nu} \sqrt{p(\mu, \nu) q(\mu, \nu)} \right)^{\frac{1}{2}} \quad (2.38)$$

2.4.2 Sparse representation based method

In [17], a robust iris matching method is proposed based on sparse representation. Firstly, this method constructs a dictionary including all the gallery features. Suppose that there are L classes and n gallery captures for each class. The sub-dictionary for the i^{th} class, denoted by $\mathbf{D}_i \in \mathbb{R}^{N \times n}$, is constructed by:

$$\mathbf{D}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{in}] \quad (2.39)$$

where $\mathbf{x}_{ij} \in \mathbb{R}^N$ is the Gabor feature extracted from the j^{th} gallery captures of the i^{th} class. Given the sub-dictionaries for all the classes, the final dictionary \mathbf{D} is constructed by $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_L] \in \mathbb{R}^{N \times (n \times L)}$.

Then, given the Gabor feature $\mathbf{y} \in \mathbb{R}^N$ extracted from a probe capture, this method seeks the sparse representation of \mathbf{y} on \mathbf{D} using basis pursuit [93–95]:

$$\arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{D}\boldsymbol{\alpha} \quad (2.40)$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{(n \times L)}$ is the vector of coefficients of sparse representation. With the coefficients in $\boldsymbol{\alpha}$, the class label d of the gallery capture is determined by:

$$\arg \min_k \|\mathbf{y} - \mathbf{D}_k \Pi_k(\boldsymbol{\alpha})\|_2 \quad (2.41)$$

where $\Pi_k(\boldsymbol{\alpha})$ extracts the coefficients in $\boldsymbol{\alpha}$ corresponding to \mathbf{D}_k . Essentially, Eqn. 2.41 determines the class label as the class whose sub-dictionary has the minimum representation residue for \mathbf{y} in sparse representation.

Also, SCI [96] is used to measure the recognition confidence:

$$SCI = \frac{L (\max \|\Pi_k(\boldsymbol{\alpha})\|_1 / \|\boldsymbol{\alpha}\|_1) - 1}{L - 1} \quad (2.42)$$

The value of SCI is between 0 and 1. SCI= 1 means that the features from the probe capture can be represented by only 1 image in the dictionary (*i.e.* the representation coefficients concentrate on one gallery feature), hence it corresponds to a high confidence. In contrast, SCI=0 means that the representation coefficients are spread evenly across all the gallery features in all classes, so it corresponds to a low confidence.

Finally, [17] divides the iris region into multiple sectors. Each sector is individually classified using sparse representation based on Eqn. 2.39, Eqn. 2.40, Eqn. 2.41 and Eqn. 2.42 to produce a SCI. The SCIs of all the sectors are fused by a Bayesian fusion scheme to produce a cumulative SCI (CSCI) as the final matching score. Let d_1, d_2, \dots, d_M be the class labels of all M sectors determined by sparse representation (Eqn. 2.41). Let $\{c_1, c_2, \dots, c_L\}$ be the set of all class labels. CSCI derived from the Bayesian fusion scheme is as follows [17]:

$$CSCI(c_l) = \frac{\sum_{j=1}^M SCI(d_j) \delta(d_j = c_l)}{\sum_{j=1}^M SCI(d_j)} \quad (2.43)$$

where $\delta(\bullet)$ is Kronecker delta function; $CSCI(c_l)$ is the confidence score that the probe capture belongs to class c_l ; $SCI(d_j)$ is the SCI of the classification result for the j^{th} sector. The class with the highest CSCI is selected as classification result and the corresponding CSCI is used as the final matching score.

2.4.3 Iris weight map based methods

Iris weight map is a method for the robust matching of binary features (iris codes). Differently from the Hamming distance (Eqn. 2.32) which views each bit equally, the iris weight map method assigns different weights to each bit to emphasize more valuable bits in the iris code. Let $\mathbf{C}_{\text{gallery}}^j$ be a gallery feature of the j^{th} class. Let $\mathbf{C}_{\text{probe}}$ be a probe feature. The iris matching using iris weight map is based on weighted Hamming distance as follows (recall that it is assumed that the bits from noise pixels like reflections have been eliminated from iris codes for iris matching):

$$\text{WHD} = \frac{\|(\mathbf{C}_{\text{probe}} \oplus \mathbf{C}_{\text{gallery}}^j) \otimes \mathbf{w}^j\|}{\|\mathbf{w}^j\|} \quad (2.44)$$

where \otimes is element-wise multiplication operator; \mathbf{w}^j is the iris weight map that assigns weights to each bit.

The key of iris weight map methods is the calculation of weight map \mathbf{w}^j . Some representative methods to calculate \mathbf{w}^j are reviewed as follows.

Fragile bits. In [97], Hollingsworth *et al.* illustrate the existence of fragile bits. Given an iris capture of a specific eye, after Gabor filtering, some feature values are closer to the axis of complex plane than the others; therefore, the bits corresponding to those feature values closer to the axis are more fragile: they are more likely to flip between 0 and 1 in different captures of the same eye. [97] shows that the iris recognition performance can be improved by eliminating too fragile bits¹.

The bit fragility studied in [97] can be used to calculate weight maps for iris matching [98]. Suppose the i^{th} bit is 1 for m_1 times and 0 for m_0 times in the gallery iris codes of the j^{th} class. The stability of the i^{th} bit, s_i , can be calculated as

¹Before the work in [97], similar technique has been used in the implementation of Daugman's algorithm [6, 7].

follows [97, 98]:

$$s_i = \frac{|m_1 - m_0|}{m_1 + m_0} \quad (2.45)$$

s_i is higher when the i^{th} bit is more stable in all the gallery iris codes. The weight assigned to the i^{th} bit is determined as follows:

$$w_i = \begin{cases} 1, & s_i \geq 0.3 \\ 0, & s_i < 0.3 \end{cases} \quad (2.46)$$

Essentially, this method uses 0.3 as the threshold to eliminate too fragile bits for iris matching.

Personalised weight map. In [98], a personalised weight map is proposed based on the concept of bit fragility for iris matching. Suppose the i^{th} bit is 1 for m_1 times and 0 for m_0 times in the gallery iris codes of the j^{th} class. This method calculates the weight for the i^{th} bit as follows:

$$w_i = 2 \times \frac{m_1^2 + m_0^2}{(m_1 + m_0)^2} - 1 \quad (2.47)$$

Actually, $P = \frac{m_1^2 + m_0^2}{(m_1 + m_0)^2}$ is the percentage of successful match of the i^{th} bit in the exhaustive matching of the iris codes of the j^{th} class. A bit is more fragile if P is lower. Since $P \in [0.5, 1]$, Eqn. 2.47 normalises P to between 0 and 1 as the weight in the weight map.

Power law based weight map. In [22], the weight is calculated by applying power law to the stability s_i obtained by Eqn. 2.45:

$$w_i = s_i^c \quad (2.48)$$

where c is adaptive crest factor (peak-to-average ratio) measuring the quality of the

weight map. c is calculated as follows:

$$c = \begin{cases} \frac{1}{\mu}, & \text{if } \mu \neq 0 \\ 1, & \text{if } \mu = 0 \end{cases} \quad (2.49)$$

where $\mu = \frac{1}{N} \sum_{i=1}^N s_i$; N is the number of bits in an iris code.

Bit discriminability based map. In [99], Proença extends the concept of bit stability to bit discriminability. Three feature selection algorithms are used to select bits with high discriminability: Fisher-score [100], minimum redundancy - maximum relevance (mRMR) [101] and SFFS [56]. These methods are equivalent to binary iris weight maps which assign the weight of 1 to the selected bits by feature selection, and the weights of the other bits are set to 0.

Chapter 3

Datasets

This chapter introduces the datasets used for the experimental analysis in this thesis. There are totally 11 datasets used in this thesis: CASIA.v4 thousand (CASIAT) [102], ND-iris-0405 (ND0405) [103], CASIA.v4 distance (CASIAD) [102], UBIRIS.v2 (UBIRIS2) [19], Face Recognition Grand Challenge (FRGC) [104], Clarkson iris liveness detection 2013 (LivDet-Clarkson) [105], Warsaw iris liveness detection 2013 (LivDet-Warsaw) [105], Notre Dame iris liveness detection 2013 (LivDet-ND) [105], MobBIOfake [61], MICHE [5], CSIR [106]. These datasets have been widely used in the research related to iris recognition like [5, 16–18, 21–23, 38, 39, 61, 62, 75, 80, 81, 83, 97, 99, 103, 105–112]. Fig. 3.1 shows some example images in each dataset. Tab. 3.1 summarises the information of each dataset.

3.1 CASIA.v4 thousand (CASIAT)

CASIAT [102] is a large scale dataset consisting of 2000 eyes from 1000 subjects, captured in NIR wavelength and at a close distance. The dataset includes 20,000 images captured by a commercial IKEMB-100 camera. The images in this dataset have very high overall quality, despite of the influence of glasses and specular reflections in some images.

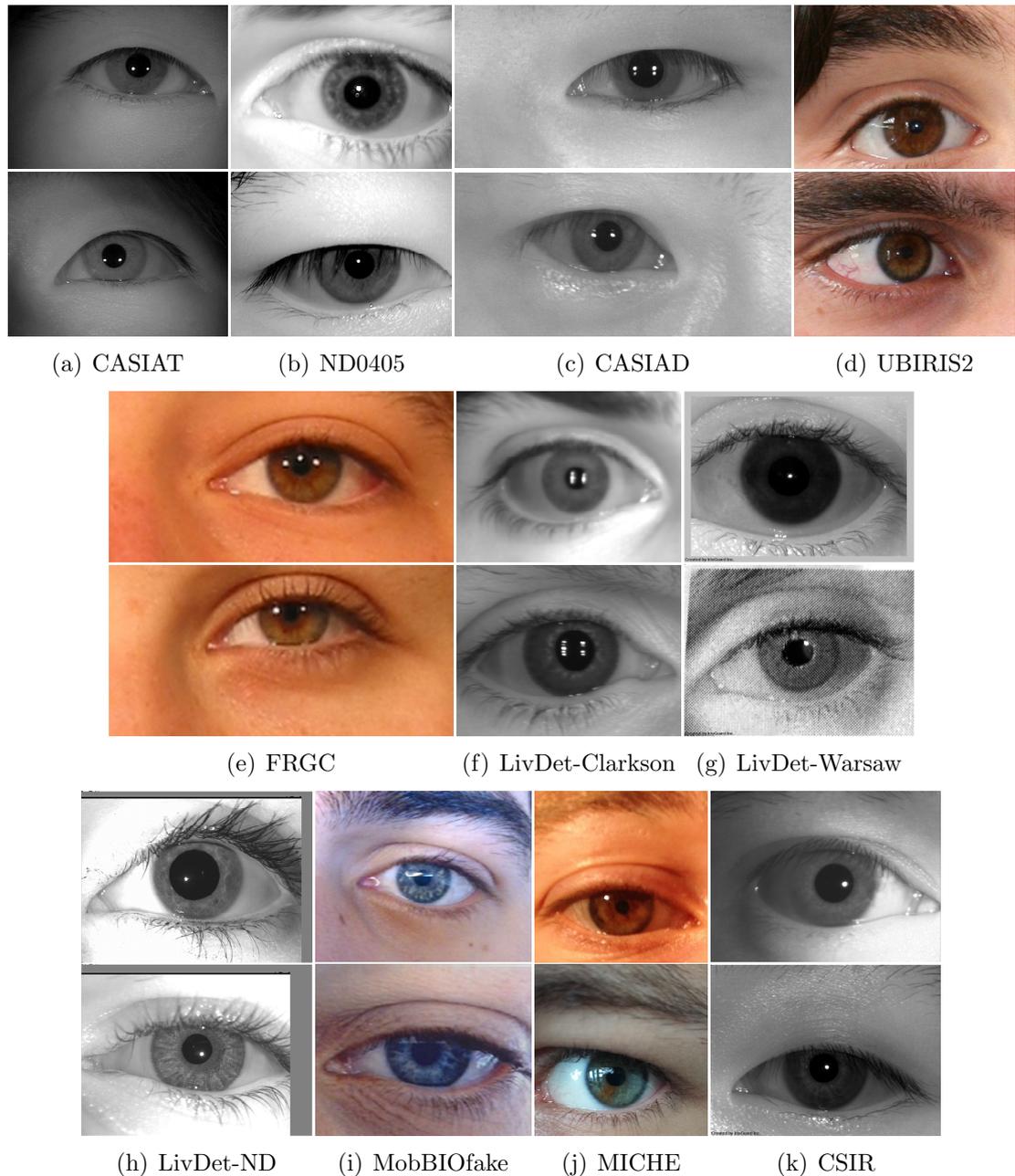


Figure 3.1: Examples of datasets used in this thesis.

3.2 ND-iris-0405 (ND0405)

ND0405 [103] is a large-scale dataset captured in NIR wavelength and at a close distance. The current version of this dataset includes 64,982 iris images from 712 eyes. The images in this dataset have relatively higher overall quality. They are captured at a wavelength inducing less noise (NIR) and a close distance, hence with

Table 3.1: Information of all the datasets used in this thesis.

Dataset	Eyes	Images	Wavelength	Sensor type	Capture environment
CASIAT	1000	20000	NIR	Single	Constrained
ND0405	712	64982	NIR	Single	Less constrained
CASIAD	284	5134	NIR	Single	Less constrained
UBIRIS2	522	11102	Visible*	Single	Less constrained
FRGC	150	500	Visible*	Single	Less constrained
LivDet-Clarkson	64	1356	NIR	Single	Less constrained
LivDet-Warsaw	284	1667	NIR	Single	Constrained
LivDet-ND	287	4200	NIR	Single	Constrained
MobBIOfake	200	1600	Visible*	Single	Less constrained
MICHE	75	2559	Visible*	Cross	Less constrained
CSIR	200	8000	NIR	Cross	Less constrained

*In this thesis, the captures in visible wavelength are also referred to as colour captures.

good resolution and clear iris texture. However, some images suffer from noise and degradations due to real-world conditions, including blurring, specular reflection, off-angle, *etc.* Also, some subjects wear soft contact lenses which cause distortion on iris textures.

3.3 CASIA.v4 distance (CASIAD)

CASIAD [102] is a dataset captured at a distance under NIR wavelength. The stand-off distance is around 3 meters. It consists of 142 subjects. The images include most of facial features and patterns. In this thesis, the left and right eye regions are detected using classical Viola-Jones object detector [113] from the images in this dataset. The false detections are manually corrected. It results 5134 iris images from 284 eyes used for experiments. These images have relatively lower overall quality. They are captured at NIR wavelength inducing less noise but at a distance. As a result, the iris region has a low resolution; the iris texture is visually less clear; the noise in this dataset is relatively heavier. The eye region of most images suffers from low resolution, blinking, eyelids occlusion, specular reflection and motion blur.

3.4 UBIRIS.v2 (UBIRIS2)

UBIRIS2 [19] is a dataset captured under visible (colour) wavelength with dynamic lighting conditions. The stand-off distance is between 3 and 8 meters. This dataset consists of 11,102 images from 522 eyes of 261 subjects. The images have very low overall quality. They suffer from both heavy noise and information loss due to the combination of distant capture and colour (visible) wavelength. Typical noise and degradations in this dataset include specular reflection across the eye region, off-angle and blurring.

3.5 Face Recognition Grand Challenge (FRGC)

FRGC [104] is a face recognition dataset. It consists of three types of captures: high-resolution still images, 3-dimensional (3-D) images, and multi-images of a person. In [21], Tan and Kumar construct a subset of this dataset to examine the performance of iris segmentation algorithms. The subset includes 500 images from 150 eyes. The images in the subset are selected from the sessions **2002-269** to **2002-269** of “**Fall 2002**” academic year. As described in [21], the eye regions are detected from the selected images using Viola-Jones object detector [113], and the ground truth segmentations are manually produced. This thesis employs this subset for the experiments on iris segmentation. The captures in this subset have very low quality, due to the heavy noise induced by visible wavelength.

3.6 Clarkson iris liveness detection 2013 (LivDet-Clarkson)

LivDet-Clarkson [105] is an iris liveness detection dataset captured in NIR wavelength. It contains two illumination types. One is a stimulated night environment and the other one is with normal lights in the room. This dataset includes 516 live images and 840 fake images. The spoofing type of this dataset is patterned contact lenses. The fake images are from 6 subjects each wearing 19 patterned contact lenses. The live images are from 64 eyes (including all the eyes in the fake collection) with up to 5 images per eye at each illumination type. The images in this dataset are captured as video sequences with the subjects moving through all focus ranges. Thus, this dataset includes images with varying focus (*i.e.* heavier noise).

3.7 Warsaw iris liveness detection 2013 (LivDet-Warsaw)

LivDet-Warsaw [105] is an iris liveness detection dataset captured in NIR wavelength. It consists of 284 eyes with 852 live images and 815 fake images. This dataset uses printouts for spoofing. The fake images are created by laser printing on matt papers. Two printers are used to produce the fake images. One is a black and white laser printer, representing the printed images of a low cost device with a low resolution. The other one is a semi-professional laser printer which is used to produce high resolution fake images. A commercial iris recognition system is further used to select among the fake images. Only the fake images successfully spoofing the system are preserved.

3.8 Notre Dame iris liveness detection 2013 (LivDet-ND)

LivDet-ND [105] is an iris liveness detection dataset captured in NIR wavelength. It has 287 eyes with 2800 live images and 1400 fake images. The spoofing type is textured contact lenses. The textured contact lenses in this dataset are from three different suppliers and with varying colours. Furthermore, the live images in Notre Dame dataset have two types. One is “clear” image with no contact lenses, while the other one is from the eyes wearing soft contact lenses without texture.

3.9 MobBIOfake

MobBIOfake [61] is an iris liveness detection dataset captured in visible (colour) wavelength. It includes 800 real and 800 fake iris images from 200 eyes. The fake samples are printed images of real samples. All the images are captured by an Asus Transformer Pad TF 300T, and the capturing distance is within 50cm.

3.10 MICHE

MICHE [5] is a cross-sensor iris dataset captured in visible wavelength by mobile devices. The images are mainly captured by a iPhone5 and a Samsung Galaxy S4, in both indoor and outdoor environments. It consists of more than 2500 colour images from 75 eyes. The images in this dataset mainly suffer from heavy specular reflection, especially for the outdoor captures. Additionally, the images are also influenced by eyelids occlusion and blurring.

3.11 CSIR

CSIR [106] is a cross-sensor iris dataset captured in NIR wavelength released for a competition on cross-sensor iris recognition [106]. The images are captured using two iris cameras: AD100 produced by irisGuard and IKEMB-220 produced by irisKing. Besides cross-sensor capturing, some images in this dataset are influenced by eyelids occlusion, illumination variation, specular reflection and blurring. CSIR consists of a training set and a test set. In this thesis, the CSIR training set is employed. It consists of 8000 images from 200 eyes.

Chapter 4

A robust iris segmentation algorithm using ℓ_1 -norm regression and model selection technique

Part of this chapter has been adapted from the journal paper [J6] (published) and the conference paper [C4] (published) as in the List of Publications.

4.1 Introduction

In this chapter, a robust iris segmentation algorithm is proposed for the task of less constrained iris recognition. This algorithm overcomes the noise and degradations in less constrained iris captures by the combination of (1) robust ℓ_1 -norm regression based segmentation and (2) a model selection technique. Specifically, this chapter demonstrates that limbic and pupillary boundaries, together with eyelids, can be fitted robustly via ℓ_1 -norm regression [114,115] on a set of estimated boundary points (boundary candidates), even if part of these boundary candidates are inaccurate due to noise and degradations. Furthermore, multiple segmentation results of iris boundaries are produced by applying ℓ_1 -norm regression to different models, and a

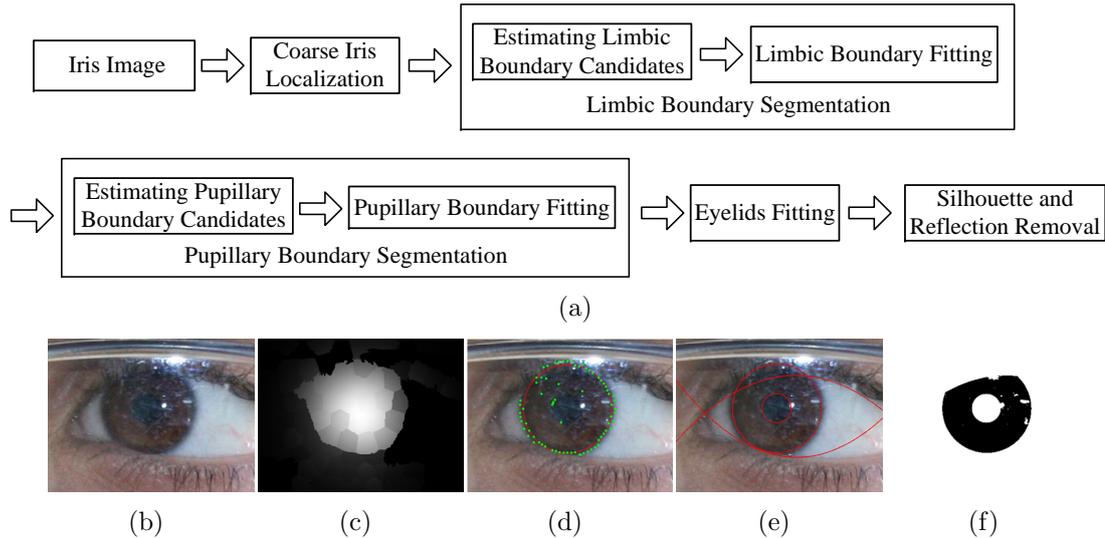
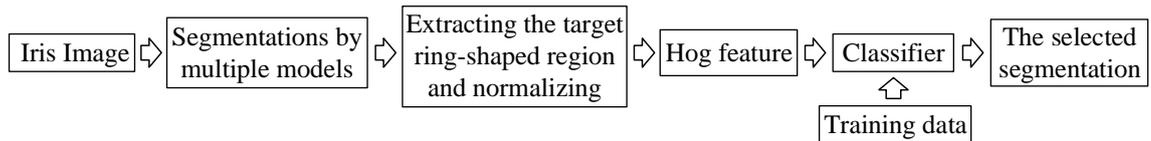


Figure 4.1: Flowchart and illustrations of the proposed ℓ_1 -norm regression based iris segmentation. (a) The flowchart of the proposed algorithm; (b)-(f) illustrations of results using the proposed algorithm; (b) less constrained iris captures in visible wavelength suffered from specular reflection, glass occlusion and slight blur; (c) result of coarse iris localisation; iris region is highlighted; (d) limbic boundary segmentation; note that although part of the estimated limbic boundary candidates (green points) are inaccurate due to reflection, pupil and eyelids, the fitted limbic boundary (red curve) is accurate; (e) results after limbic and pupillary boundary segmentation, eyelids fitting; (f) final iris mask after silhouette and reflection removal.

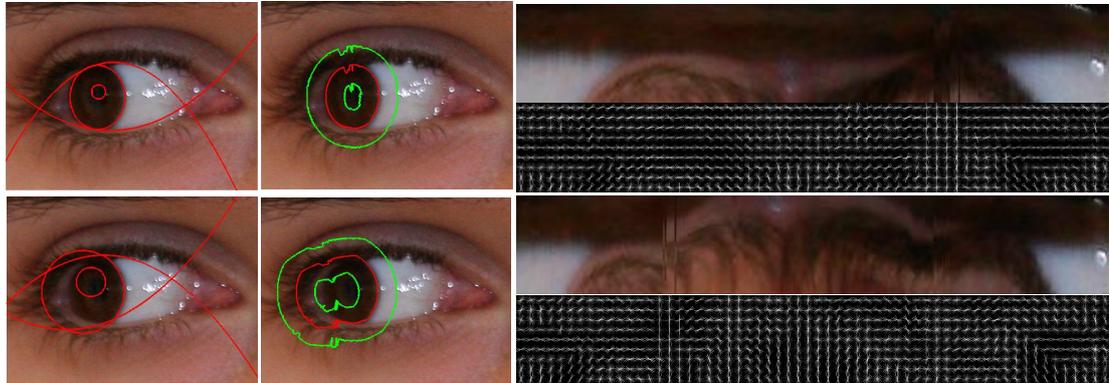
model selection method is proposed to select the most reliable result.

Fig. 4.1 shows the flowchart and illustrations of the proposed ℓ_1 -norm regression based iris segmentation. Firstly, coarse iris localisation is performed. Coarse iris region is located via super-pixel [116] based correlation histograms. Then, within coarse iris region, limbic boundary segmentation is performed. Limbic boundary is fitted by ℓ_1 -norm regression on limbic boundary candidates, which are obtained by seeking the local maxima of directional derivatives. After that, pupillary boundary segmentation is performed via a similar procedure to limbic boundary segmentation. Finally, eyelids are fitted, followed by reflection and silhouette removal.

Fig. 4.2 shows the flowchart and illustrations of the proposed model selection procedure. Firstly, for the input iris image, multiple segmentations are obtained by performing ℓ_1 -norm regression with different models. Then, a ring-shaped region is



(a)



(b)

(c)

(d)

Figure 4.2: Flowchart and illustrations of the proposed model selection method. (a) The flowchart of the proposed model selection method; (b)-(d) some example illustrations of the proposed model selection method; (b) examples of multiple segmentations for the same capture; (c) examples of the target ring-shaped region (the region between the two green curves) extracted around the outer segmentation boundary (the red curve); (d) examples of normalised ring-shaped region and the corresponding HOG features.

defined around the outer segmentation boundary and it is used to evaluate the quality of segmentations produced by each model. This ring-shaped region is normalised using Daugman's rubber sheet model [6], and histogram of oriented gradients (HOG) features [117] are extracted to distinguish good and poor segmentations. Finally, the classification decision of a pre-trained classifier using data from good and bad segmentations is used to select the model with optimal segmentation result.

The experiments are performed on four less constrained iris datasets introduced in Chapter 3: MICHE, UBIRIS2, FRGC and CASIAD. The results show that the proposed method achieves top performance compared to state-of-the-art iris segmentation algorithms.

The rest of this chapter is organised as follows. Section 4.2 describes and analyses the proposed iris segmentation algorithm based on ℓ_1 -norm regression with multiple

models. Section 4.3 presents the proposed method for model selection. Section 4.4 reports the experimental results on less constrained iris datasets. Section 4.5 summarises this chapter.

4.2 Iris segmentation using ℓ_1 -norm regression

This section presents the iris segmentation algorithm using ℓ_1 -norm regression. The algorithm may be divided into the following steps: coarse iris localisation, limbic boundary segmentation, pupillary boundary segmentation, eyelids fitting, silhouette and reflection removal. The algorithm consists of three models: one circle model and two ellipse models. Furthermore, synthetic data is used to study the characteristics of the three models and the result demonstrates that a model selection is able to achieve an improved performance over each individual model.

4.2.1 Coarse iris localisation

The iris region is located by a coarse iris map based on the correlation histogram of super-pixels [116]. For colour datasets (MICHE, UBIRIS2 and FRGC), the algorithm operates on red channel, since the observation in [118] suggests that red channel is the most informative channel for colour iris images. As a pre-processing step, contrast adjustment is performed to make iris region more distinguishable. The intensity values between a low threshold and a high threshold are mapped to $[0, 1]$, and the remaining intensity values are clipped. Let m_r be the mean intensity value in a iris capture. The low threshold is set to $m_r - \sigma$, and the high threshold is set to $m_r + \sigma$, where σ is a parameter experimentally set to 0.2 for all datasets (all the parameters in this chapter are set based on two sets of training data, one for colour captures and the other one for NIR captures; please see Section 4.4.1 for details; please also note that, in the rest of this chapter, the parameter setting is

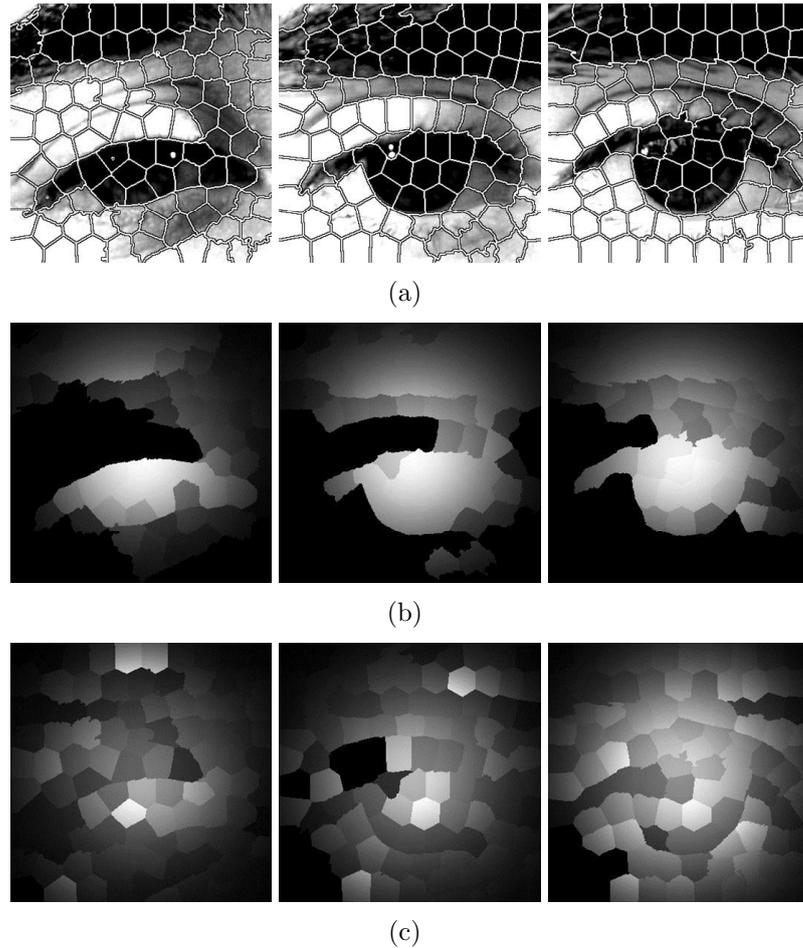


Figure 4.3: Examples of super-pixels and coarse iris map. The images in the same column are from the same capture. (a) Super-pixels on contrast-adjusted captures; (b) the coarse iris maps computed using contrast-adjusted captures; (c) the coarse iris maps computed using the captures without contrast adjustment.

the same for all the datasets if not specified).

To illustrate the effect of contract adjustment, some example results of coarse iris localisation with and without contrast adjustment are presented in Fig. 4.3(b) and Fig. 4.3(c), respectively. It can be seen that with contrast adjustment, the iris region in coarse iris maps is more distinguishable and consistent.

Super-pixels are computed after contrast adjustment. Super-pixels are essentially over-segmentation of an image. It is perceptually meaningful because it preserves local structure of an image. Examples of super-pixel are shown in Fig. 4.3(a). Simple linear iterative clustering (SLIC) algorithm [116] is employed to compute

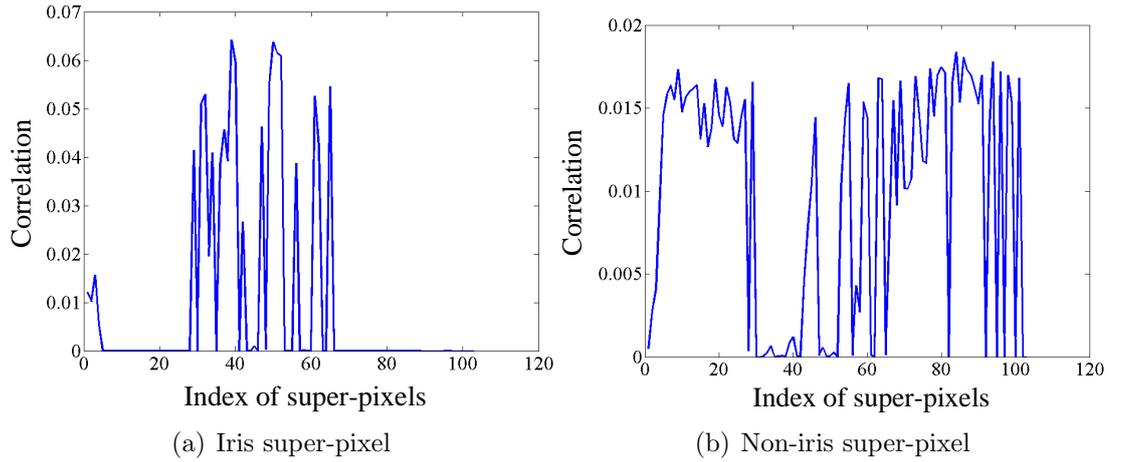


Figure 4.4: Examples of correlation histogram corresponding to an iris super-pixel and a non-iris super-pixel.

super-pixels.

Correlation histogram (CH) is used as a feature to distinguish between iris and non-iris super-pixels. Denote the total number of super-pixels in the contrast-adjusted capture by n , the CH of each super-pixel consists of $n - 1$ bins. For the i^{th} super-pixel, a n -bin histogram \mathbf{C}^i is firstly calculated, with the j^{th} bin, denoted by C_j^i , calculated by:

$$C_j^i = \frac{\langle \mathbf{h}_i, \mathbf{h}_j \rangle}{\sum_{j=1, j \neq i}^{j=n} \langle \mathbf{h}_i, \mathbf{h}_j \rangle} \quad (4.1)$$

where \mathbf{h}_i and \mathbf{h}_j denote respectively the normalised histogram of the i^{th} and j^{th} super-pixel, and $\langle \bullet, \bullet \rangle$ calculates the inner product of two vectors. The CH is obtained by removing the i^{th} bin of \mathbf{C}^i . The reason of the removal is that the i^{th} bin of \mathbf{C}^i reveals the self-similarity of the i^{th} super-pixel and it is less informative for the aim to roughly locate the iris region.

The intuition behind CH is that the iris region is visually more unique in an iris image. As a result, the CH of iris super-pixels has few peaks, since few super-pixels are similar with iris super-pixels. In contrast, the CH of non-iris super-pixels tends to distribute across most of the bins due to a broad similarity. Fig. 4.4 shows examples of CH corresponding to an iris super-pixel and a non-iris super-pixel.

Therefore, based on the above characteristic of CHs of iris and non-iris super-pixels, entropy is used to distinguish between iris and non-iris super-pixels. Let E^i be the entropy of the CH of the i^{th} super-pixel, E^i is calculated by:

$$E^i = - \sum_{j=1}^{n-1} C_j^i \log_2 C_j^i \quad (4.2)$$

An iris super-pixel tends to have a lower entropy value, because its CH concentrates on few bins. In contrast, a non-iris super-pixel is likely to correspond to a higher entropy value due to the broad distribution of its CH. To highlight iris region and build the coarse iris map, each super-pixel is filled with an intensity based on the entropy of CH. Let S^i be the intensity of the i^{th} super-pixel, S^i is calculated by:

$$S^i = e^{-E^i} \quad (4.3)$$

The examples of coarse iris maps are shown in Fig. 4.3(b). This map is called a ‘coarse’ map because it is not very accurate – it can be seen that some other parts of the image such as the eyelids and eyebrows are also highlighted. The reason is that these parts also show some uniqueness in iris images. However, the coarse iris map is able to roughly locate iris region in images. It gives important information for the following iris segmentation steps.

4.2.2 Limbic boundary segmentation

To segment limbic boundary, an initial iris centre is firstly located based on the coarse iris map. Points around the initial iris centre are selected as iris centre candidates. Then, a limbic boundary is fitted for each iris centre candidate. Among the limbic boundaries corresponding to different iris centre candidates, the one with the highest score of Daugman’s integro-differential operator [6, 7] is selected. The limbic boundary fitting is performed as robust ℓ_1 -norm regression on a set of limbic

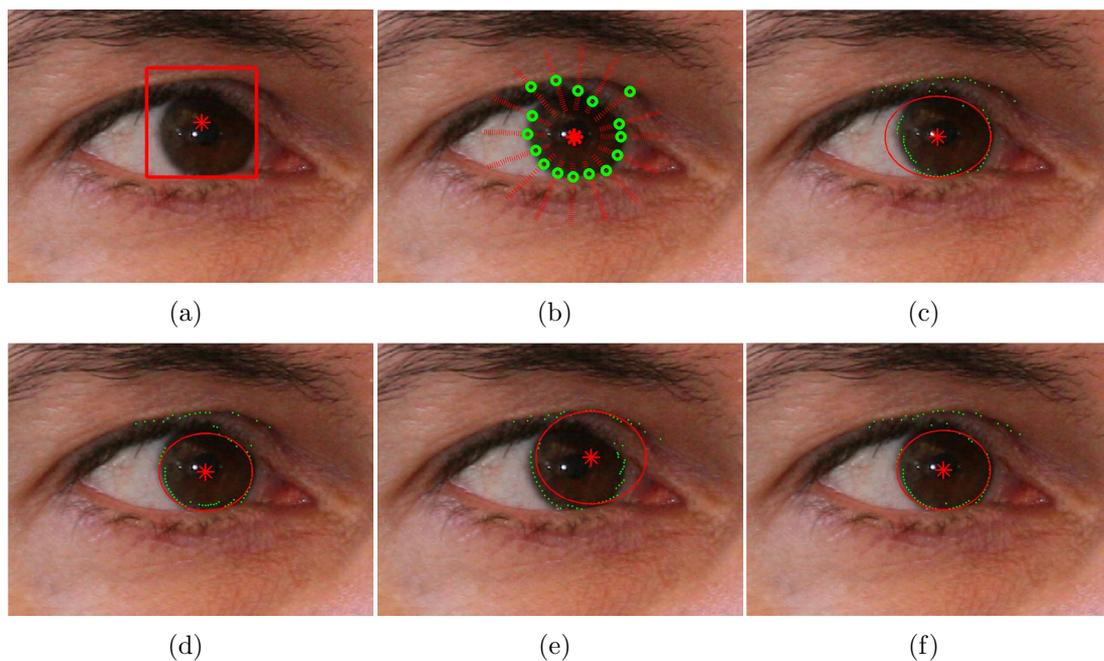


Figure 4.5: An illustration of the proposed limbic boundary segmentation algorithm. (a) The initial iris centre (red star) and region of iris centre candidates (red square); (b) the sought limbic boundary candidates (LBCs, green circles) corresponding to one centre candidate (red star); note only LBCs in a few directions (red dashed line) are drawn for clarity; (c)-(e): limbic boundary fitting results with different iris centre candidate (red star) and the corresponding LBCs (green points); (f) the selected optimal limbic boundary with the highest score of Daugman's integro-differential operator.

boundary points (limbic boundary candidates, LBCs), which are sought individually for each iris centre candidate. Fig. 4.5 illustrates the procedure of limbic boundary segmentation.

To locate the initial iris centre, the coarse iris map is binarized via Otsu threshold [119]. Then, the centre of the largest connected component is selected as the initial iris centre. The reason of selecting the largest connected component is to eliminate the influence of the eyebrow – it is observed experimentally that, although parts of the eyebrow may remain after binarisation, the area of the remaining eyebrow region is usually smaller than the iris region. The exceptional case where the eyebrow region is larger than the iris region is very rare. One such exceptional case is shown in Fig. 4.6(a), Fig. 4.6(b) and Fig. 4.6(c). It can be seen from the three figures that this exceptional case is due to the combination of glass frame, eyebrow,

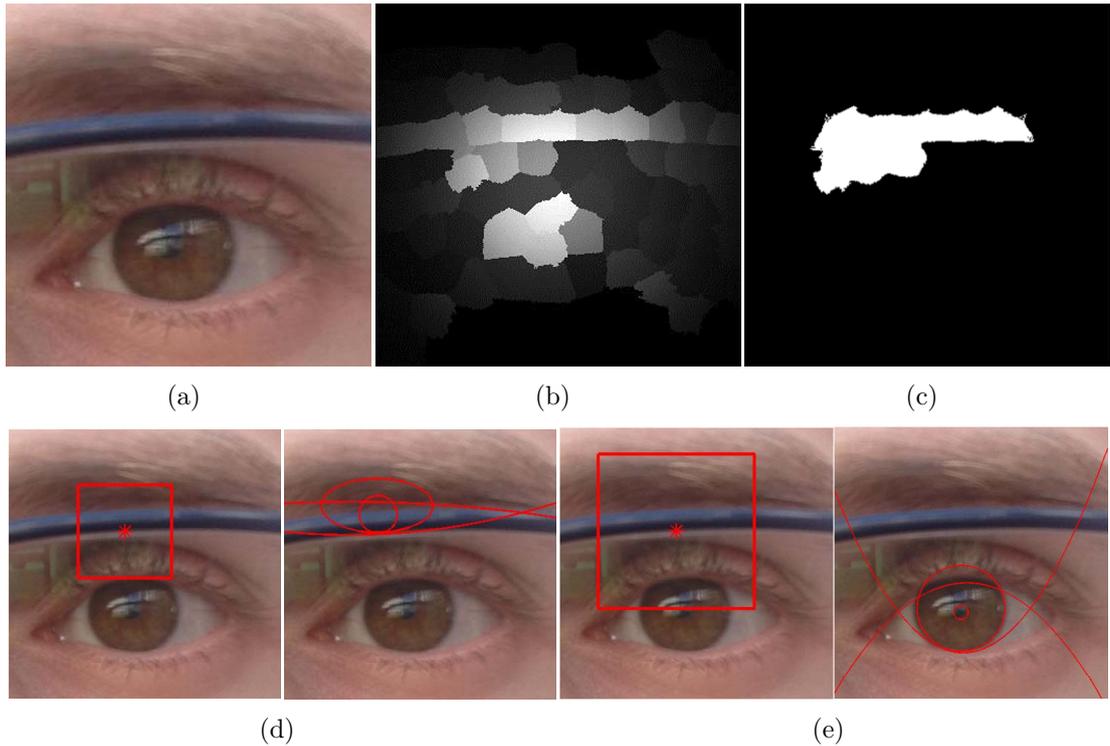


Figure 4.6: An illustration of the effect of the size of iris centre candidate region. (a) An iris image from MICHE dataset; (b) the result of coarse iris localisation; (c) the largest connected component after binarisation; note that this is a special case where the eyebrow region is larger than the iris region after binarisation; (d) the example of a small region of iris centre candidate (red rectangle, left) and the corresponding iris segmentation result (right); (e) the example of a large region of iris centre candidate (red rectangle, left) and the corresponding iris segmentation result (right). Note that the initial iris centre (the red star in the left image of (d) and (e)) is still near to the true iris centre, since the bias of iris centre is mainly caused by the eyebrow and glass frame. Therefore, a large iris centre candidate region (the red rectangle in the left image in (e)) is able to mitigate the influence of the biased initial iris centre and give an accurate segmentation (shown in the right image in (e)).

and inaccurate intensity of iris region in the coarse iris map. It will be show later that the influence of such failure in rough localisation can be overcome by a large iris centre candidate region.

The iris centre candidates are selected as the points in a square region centred at the initial iris centre. The width of the square region is set to 120 pixels for all the datasets. A relatively large width is selected to mitigate the influence of inaccurate coarse iris map. An illustration is shown in Fig. 4.6(d) and Fig. 4.6(e). It can be

seen from Fig. 4.6(d) and Fig. 4.6(e) that the inaccurate coarse iris map (Fig. 4.6(b) and Fig. 4.6(c)) leads to a biased initial iris centre (red star in the left image of Fig. 4.6(d) and Fig. 4.6(e)). However, this biased initial iris centre is still near to the true iris centre, since the error is mainly caused by eyebrow and glass frame. Therefore, a relatively larger iris centre candidate region is still able to reach the true iris centre and give accurate segmentation (the right image in Fig. 4.6(e)).

For each iris centre candidate, the corresponding LBCs are estimated by seeking the local maxima of directional derivatives along multiple rays initiated from the iris centre candidate. Denote the coordinate of an arbitrary iris centre candidate by (x_i, y_i) , denote the ray initiated from (x_i, y_i) with incline angle θ by l_i^θ , and denote the LBC along l_i^θ by (x_i^θ, y_i^θ) , (x_i^θ, y_i^θ) is estimated by:

$$\begin{aligned} (x_i^\theta, y_i^\theta) &= \arg \max_{x,y} \{I(x + \Delta r \cos \theta, y + \Delta r \sin \theta) \\ &\quad - I(x, y)\} \\ \text{s.t. } (x, y) &\in l_i^\theta \\ r_{\min} &\leq \sqrt{(x - x_i)^2 + (y - y_i)^2} \leq r_{\max} \end{aligned} \quad (4.4)$$

where Δr is the step length for derivative estimation, r_{\min} and r_{\max} restrict the search region, I is the processed image – the potential reflection regions are filled by the mean intensity of pixels around it; the potential reflection regions are found by high-pass filtering (since reflection leads to sudden intensity change and it induces high frequency components) and then binarizing via Otsu threshold. The regularisation terms in Eqn. 4.4 restrict the search on l_i^θ with a range of $[r_{\min}, r_{\max}]$. For each iris centre candidate, θ is varied to seek LBCs at multiple directions. The range of θ depends on the limbic boundary model and will be given later. For the other parameters, for colour datasets (MICHE, UBIRIS2 and FRGC), Δr is set to 7, $r_{\min} = (\text{width} + \text{height})/20$ and $r_{\max} = (\text{width} + \text{height})/7$ (width and height are size of image); for CASIAD dataset. Δr is set to 3, $r_{\min} = (\text{width} + \text{height})/20$ and

$r_{\max} = (\text{width} + \text{height})/10$. An example of an iris centre candidate and LBCs is shown in Fig. 4.5(b), respectively. Note that only a few directions are drawn for clarity in Fig. 4.5(b).

From Fig. 4.5(b), it can be seen that there are false LBCs due to eyelids. In such cases, if the limbic boundary is fitted based on traditional least square fitting, the fitting error of the false LBCs will bias the fitting result. Therefore, in this section, a limbic boundary fitting algorithm is proposed using robust ℓ_1 -norm regression [88, 89, 114, 115, 120].

Given an arbitrary iris centre candidate (x_i, y_i) , denote the LBCs corresponding to (x_i, y_i) along directions $\theta_1, \theta_2, \dots, \theta_t$ by $(x_i^{\theta_1}, y_i^{\theta_1}), (x_i^{\theta_2}, y_i^{\theta_2}), \dots, (x_i^{\theta_t}, y_i^{\theta_t})$, respectively, three individual models are proposed for limbic boundary fitting based on ℓ_1 -norm regression: a circle model, a parametric ellipse model and an ellipse model.

ℓ_1 -norm circle (ℓ_1 -C) model. This circle model fits limbic boundary as a circle.

Given the iris centre candidate and the corresponding LBCs, the only parameter of the circle model is the radius r . Let d_j be the distance between iris centre candidate (x_i, y_i) and LBC $(x_i^{\theta_j}, y_i^{\theta_j})$, d_j is calculated by $d_j = \sqrt{(x_i - x_i^{\theta_j})^2 + (y_i - y_i^{\theta_j})^2}$. Define $\mathbf{d} = [d_1, d_2, \dots, d_t]^T \in \mathbb{R}^{t \times 1}$, the optimal radius is sought robustly by

$$\arg \min_r \|\mathbf{d} - \mathbf{1}r\|_1 \quad (4.5)$$

where $\mathbf{1} \in \mathbb{R}^{t \times 1}$ is a column vector of ones. The solution of Eqn. 4.5 is the median value of \mathbf{d} . Note that in Eqn. 4.5, if the ℓ_1 -norm is replaced by ℓ_2 -norm, this equation becomes traditional least square fitting. In the circle model, for LBCs seeking, θ is varied in the range $[-\pi/6, \pi/3]$ and $[2\pi/3, 7\pi/6]$. The range of θ is designed to seek LBCs at both sides. The reason is two-fold. Firstly, circle is a less flexible model to eyelids occlusion. The range is designed to avoid false candidates caused by eyelids. Secondly, points at both sides are sufficient to compute the radius of circle.

ℓ_1 -norm parametric ellipse (ℓ_1 -PE) model. Circle is a less flexible model, especially in cases with eyelids occlusion. Alternatively, a more flexible parametric ellipse model is proposed. Given an iris centre candidate and the corresponding LBCs, the aim is to fit the x -axis (a) and y -axis (b) of an ellipse. Let $d_j^2 = (x_i - x_i^{\theta_j})^2 + (y_i - y_i^{\theta_j})^2$. Let $\mathbf{d}^{pe} = [d_1^2, d_2^2, \dots, d_t^2]^T \in \mathbb{R}^{t \times 1}$. According to the parametric equation of ellipse:

$$\begin{cases} x_i^{\theta_j} = x_i + a \cos \varphi_j \\ y_i^{\theta_j} = y_i + b \sin \varphi_j \end{cases} \quad (4.6)$$

it can be derived that $d_j^2 = a^2 \cos^2 \varphi_j + b^2 \sin^2 \varphi_j$. Define $\mathbf{s}_i = [\cos^2 \varphi_i, \sin^2 \varphi_i]^T \in \mathbb{R}^{2 \times 1}$ and $\mathbf{s} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_t]^T \in \mathbb{R}^{t \times 2}$. a and b can be computed by:

$$\arg \min_{a,b} \left\| \mathbf{d}^{pe} - \mathbf{s} [a^2, b^2]^T \right\|_1 \quad (4.7)$$

To calculate \mathbf{s} , $\varphi_1, \varphi_2, \dots, \varphi_t$ need to be estimated. To obtain $\varphi_1, \varphi_2, \dots, \varphi_t$, an assumption is made: despite of an ellipse, the limbic boundary is near to a circle. With this assumption, the following approximation is used: $\cos \varphi_j = (x_i^{\theta_j} - x_i) / d_j$ and $\sin \varphi_j = (y_i^{\theta_j} - y_i) / d_j$. Eqn. 4.7 is also a ℓ_1 -norm regression problem. It can be solved employing the primal-dual method implemented in ℓ_1 -magic toolbox [121].

In ℓ_1 -PE model, the range of θ is set differently from ℓ_1 -C model. θ is varied from 0 to 2π to seek LBCs for each iris centre candidate. The LBCs are sought around 360 degrees for ℓ_1 -PE model, since the points at both sides are insufficient to accurately estimate the radius along y axis.

ℓ_1 -norm ellipse (ℓ_1 -E) model. The limbic boundary can be also fitted by performing ℓ_1 -norm regression based on an ellipse model directly using the ellipse equation. Let $\mathbf{d}_j^e = \left[(x_i^{\theta_j} - x_i)^2, (y_i^{\theta_j} - y_i)^2 \right]^T \in \mathbb{R}^{2 \times 1}$. Let $\mathbf{d}^e = [\mathbf{d}_1^e, \mathbf{d}_2^e, \dots, \mathbf{d}_t^e]^T \in \mathbb{R}^{t \times 2}$. In

ℓ_1 -E model, a and b are estimated by:

$$\arg \min_{a,b} \left\| \mathbf{1} - \mathbf{d}^e [1/a^2, 1/b^2]^T \right\|_1 \quad (4.8)$$

where $\mathbf{1}$ is a column vector of ones. In ℓ_1 -E model, the range of θ is the same as that in ℓ_1 -PE model.

4.2.3 Pupillary boundary segmentation

The procedure of pupillary boundary segmentation is similar to limbic boundary segmentation. The initial pupil centre is determined as the centre of the region within the limbic boundary. Pupil centre candidates are defined as a square region around the initial pupil centre. The width of the square region is set to $\frac{1}{2}r$ for ℓ_1 -C model and $\frac{1}{4}(a+b)$ for ℓ_1 -PE model and ℓ_1 -E model (note that here and after some parameters are related to the limbic boundary model).

For each pupil centre candidate, Eqn. 4.4 is used to seek pupillary boundary candidates. The pupillary boundary is fitted using ℓ_1 -C model described in Section 4.2.2 (Eqn. 4.5). The circle with the highest score of Daugman's integro-differential operator is selected amongst all fitting results. At this stage, the I in Eqn. 4.4 is different from LBC seeking. More processing effort is required because the pupil region is less distinguishable and more noisy in less constrained captures, especially in some colour captures. For each pupil centre candidate, a square patch centred at that point is extracted. For colour datasets (MICHE, UBIRIS2 and FRGC), the width of the square patch is set to $\frac{4}{3}r$ for ℓ_1 -C model and $\frac{2}{3}(a+b)$ for ℓ_1 -PE model and ℓ_1 -E model; for CASIAD dataset, the width of the square patch is set to $\frac{8}{3}r$ for ℓ_1 -C model and $\frac{4}{3}(a+b)$ for ℓ_1 -PE model and ℓ_1 -E model. Contrast adjustment is performed on the square patch using the method in Section 4.2.1. Finally, the square patch is up-scaled 3 times and binarized. The aim of up-scaling is to reduce the noise effect marginally. The binarisation threshold is set to the lower quartile pixel value of the

contrast-adjusted square patch. The other parameters in Eqn. 4.4 are set as follows: for colour datasets (MICHE, UBIRIS2 and FRGC), Δr is set to 4, θ is varied in the range $[-\pi/6, \pi/3]$ and $[2\pi/3, 7\pi/6]$; $r_{\min} = 3 \times \frac{1}{6}r$ and $r_{\max} = 3 \times \frac{1}{2}r$ for ℓ_1 -C model (3 is the factor of up-scaling), $r_{\min} = 3 \times \frac{1}{12}(a+b)$ and $r_{\max} = 3 \times \frac{1}{4}(a+b)$ for ℓ_1 -PE and ℓ_1 -E models; for CASIAD dataset, Δr , θ and r_{\min} are the same as that on colour datasets, while $r_{\max} = 3 \times \frac{3}{4}r$ for ℓ_1 -C model, and $r_{\max} = 3 \times \frac{3}{8}(a+b)$ for ℓ_1 -PE and ℓ_1 -E models.

4.2.4 Upper eyelid fitting

Similar with limbic boundary fitting, an upper eyelid model is fitted based on some upper eyelid candidatess (UECs). Inspired by [107], UECs are sought based on the local maxima of vertical derivatives. Denote the UECs along vertical line $x = v$ by (v, y^v) , y^v is estimated by:

$$y^v = \arg \max_y \{I(v, y + \Delta y) - I(v, y)\} \quad (4.9)$$

$$s.t. \quad \Delta_1 \leq y \leq \Delta_2$$

where Δy is the step length, and Δ_1, Δ_2 restrict the search region. Let (x_c, y_c) be the iris centre. Δ_1 is set to $y_c - r$ and Δ_2 is set to $y_c + 0.5r$ for ℓ_1 -C model; Δ_1 is set to $y_c - b$ and Δ_2 is set to $y_c + 0.5b$ for ℓ_1 -PE model and ℓ_1 -E model. To seek a set of UECs, v is varied in the range $[x_c - 1.5r, x_c - 1.2r] \cup [x_c + 1.2r, x_c + 1.5r]$ for ℓ_1 -C model, and $[x_c - 1.5a, x_c - 1.2a] \cup [x_c + 1.2a, x_c + 1.5a]$ for ℓ_1 -PE model and ℓ_1 -E model. Δy is set to 4.

The upper eyelid is modeled by parabola $y = \alpha(x - \beta)^2 + \gamma$. To overcome false UECs, upper eyelid fitting is formulated as a ℓ_1 -norm regression problem. Assume there are T UECs $(v_1, y_1^{v_1}), (v_2, y_2^{v_2}), \dots, (v_T, y_T^{v_T})$. Let $\mathbf{y}^u = [y_1^{v_1}, y_2^{v_2}, \dots, y_T^{v_T}]^T \in \mathbb{R}^{T \times 1}$, $\boldsymbol{\psi}_i^u = [v_i^2, -2v_i, 1]^T \in \mathbb{R}^{3 \times 1}$ and $\boldsymbol{\psi}^u = [\boldsymbol{\psi}_1^u, \boldsymbol{\psi}_2^u, \dots, \boldsymbol{\psi}_T^u]^T \in \mathbb{R}^{T \times 3}$, (α, β, γ) are

computed using:

$$\arg \min_{\alpha, \beta, \gamma} \left\| \mathbf{y}^u - \boldsymbol{\psi}^u [\alpha, \alpha\beta, \beta^2 + \gamma]^T \right\|_1 \quad (4.10)$$

Borrowing an idea from [107], α , β and γ are further restricted as follows: $\alpha \in [0, \frac{1}{2r}]$, $\beta \in [x_c - 0.5r, x_c + 0.5r]$ and $\gamma \in [y_c - 1.5r, y_c + 1.5r]$ for ℓ_1 -C model; $\alpha \in [0, \frac{1}{2 \times \min(a,b)}]$, $\beta \in [x_c - 0.5a, x_c + 0.5a]$ and $\gamma \in [y_c - 1.5b, y_c + 1.5b]$ for ℓ_1 -PE model and ℓ_1 -E model.

4.2.5 Lower eyelid fitting

The procedure of lower eyelid fitting is the same as upper eyelid fitting: some lower eyelid candidates are sought and used to fit a parabola. The only difference is the parameter setting in Eqn. 4.9 and Eqn. 4.10. In Eqn. 4.9, Δ_1 is set to y_c and Δ_2 is set to $y_c + 1.5r$ for ℓ_1 -C model; Δ_1 is set to y_c and Δ_2 is set to $y_c + 1.5b$ for ℓ_1 -PE model and ℓ_1 -E model. Δy is still set to 4. v is varied in $[x_c - 0.5r, x_c + 0.5r]$ for ℓ_1 -C model, and $[x_c - 0.5a, x_c + 0.5a]$ for ℓ_1 -PE model and ℓ_1 -E model. In Eqn. 4.10, the range of α , β and γ is restricted as follows: $\alpha \in [-\frac{1}{2r}, 0]$, $\beta \in [x_c - 0.5r, x_c + 0.5r]$ and $\gamma \in [y_c - 1.5r, y_c + 1.5r]$ for ℓ_1 -C model; $\alpha \in [-\frac{1}{2 \times \min(a,b)}, 0]$, $\beta \in [x_c - 0.5a, x_c + 0.5a]$ and $\gamma \in [y_c - 1.5b, y_c + 1.5b]$ for ℓ_1 -PE model and ℓ_1 -E model. Some example results of upper and lower eyelid fitting are shown in Fig. 4.7.

4.2.6 Reflection removal

Generally, reflection regions have very high intensity value. To further distinguish a reflection region from the iris region, contrast adjustment is performed using the method in Section 4.2.1. At this stage, only iris region is considered. The mean intensity m_r is calculated using pixels in the iris region only. After contrast adjustment, the pixel value of red channel is normalised to $[0, 255]$; pixels with an intensity value higher than 215 are marked as reflection. Please note that, although

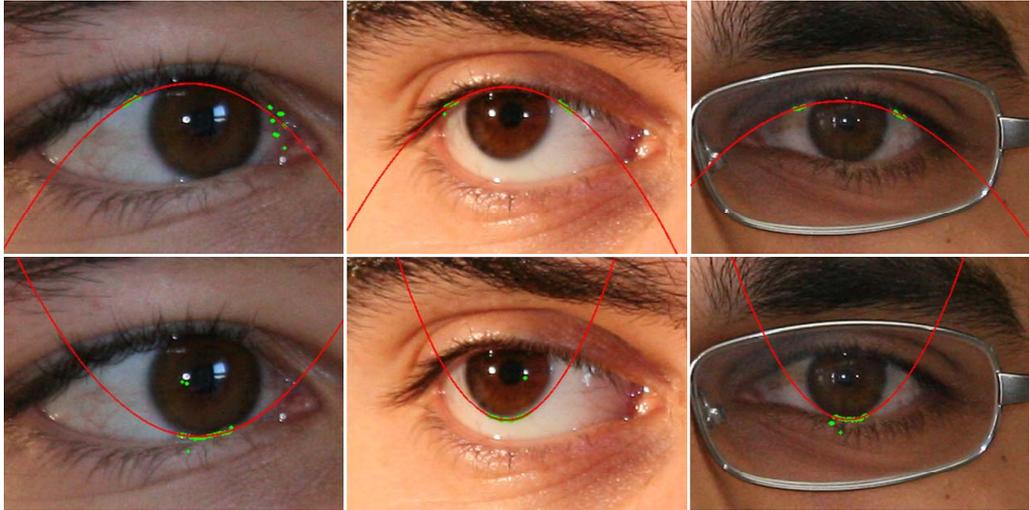


Figure 4.7: Example results of eyelids fitting. Top row: example results of upper eyelid fitting. Bottom row: lower eyelid fitting. The green points in the images are eyelids candidates, and the red curves are the fitted eyelids.

this method uses a simple fixed threshold, it is experimentally effective, since the previous step of contrast adjustment generally maps reflection regions to very high intensities.

4.2.7 Silhouette removal

There are two experimental observations on silhouette: (1) silhouette is one of the darkest regions in an iris and (2) silhouette locates near to eyelids and pupillary boundary. Therefore, the darkest pixels near to the eyelids and pupillary boundary are marked as silhouette. Firstly, contrast adjustment is performed using the method in Section 4.2.6. Then, the darkest $1/8$ pixels within iris region is selected as silhouette candidates. Otsu threshold is applied to select the darkest pixels among silhouette candidates. The selected darkest pixels are grouped into connected components; the connected components intersected with eyelids or pupillary boundary are marked as silhouette.

4.2.8 Analysis of the three models

This subsection analyses the ℓ_1 -C, ℓ_1 -PE and ℓ_1 -E models based on synthetic data. The analysis focuses on the accuracy of limbic boundary segmentation only. The reason is two fold. First, the performance of the three iris segmentation models largely depends on the accuracy of limbic boundary segmentation, as the limbic boundary is the basis for the following segmentation steps. Second, since the other components of the algorithm depend on the result of limbic boundary segmentation, it is not easy to synthesize appropriate data for them.

In the three models, the limbic boundary segmentation is performed as circle (or ellipse) fitting on a set of LBCs. The accuracy of limbic boundary segmentation is directly related to the accuracy of LBCs. Although the iris image may suffer from multiple noise factors such as motion blur and illumination variance, the influence of all these noise factors on the limbic boundary segmentation can be concluded as inducing false LBCs. Therefore, the performance of the three iris segmentation models can be analysed using corrupted circle (or ellipse) data. Firstly, some points are sampled from a circle (or ellipse) to stimulate the correct LBCs. Then, some of the sample points are corrupted by random noise to stimulate the influence of real noise on the LBCs. Finally, circle (or ellipse) is fitted using the corrupted sample points to estimate the influence of noise on the three limbic segmentation models.

Since the proposed algorithm has one circle model and two ellipse models, two sets of synthetic data are used: circle data and near-to-circle ellipse data (recall that the iris shape is assumed to be near to a circle in the description of ℓ_1 -PE model in Section 4.2.2). The experiment is designed as follows. For circle data, as the first step, a circle is randomly generated. Some points are sampled on this circle. For each of the three models, the position of sampling is consistent with the setting of θ in Section 4.2.2. Then, part of the sample points (randomly selected) are corrupted by zero mean noise (two kinds of noise are tested: Gaussian noise and uniform noise).

Finally, the radius is estimated using the ℓ_1 -C, ℓ_1 -PE and ℓ_1 -E model individually and the estimation error is calculated. The standard deviation (sigma) of noise is varied from 1 to 50. For each standard deviation level, the above procedure is repeated 100 times. The mean estimation error of the three models is drawn as the function of standard deviation level of the noise in Fig. 4.8(a) and Fig. 4.8(c).

For near-to-circle ellipse data, the whole procedure is the same as circle data. The only difference is that in the first step an ellipse is randomly generated. The ratio of short radius to long radius is randomly selected between 0.8 and 1. The mean estimation error of the three models is drawn as the function of the standard deviation level of noise in Fig. 4.8(b) and Fig. 4.8(d).

It can be seen that the result is similar for both Gaussian noise and uniform noise. From Fig. 4.8(a) and Fig. 4.8(c), it is observed that the circle model generally achieves the best performance on circle data even with large noise. It means that if the shape of the iris is exactly or very near to a circle, the ℓ_1 -C model will fit it with very high accuracy. From Fig. 4.8(b) and Fig. 4.8(d), it is shown that for near-to-circle ellipse data, the fitting error depends on the noise level. When the noise level is very low, the ℓ_1 -E model has the lowest fitting error. When the noise level is high, the ℓ_1 -PE model gives the best fitting. In conclusion, the performance of ℓ_1 -C, ℓ_1 -PE and ℓ_1 -E model depend on the iris shape and noise level. Therefore, a model selection is possible to improve the accuracy of iris segmentation over each individual model.

4.3 Model selection

The model selection is performed using a ring-shaped region around the outer segmentation boundary defined by limbic boundary and eyelids boundaries. Let $f(\theta, r)$ define the outer segmentation boundary in the polar coordinate. Let Δr_w be the

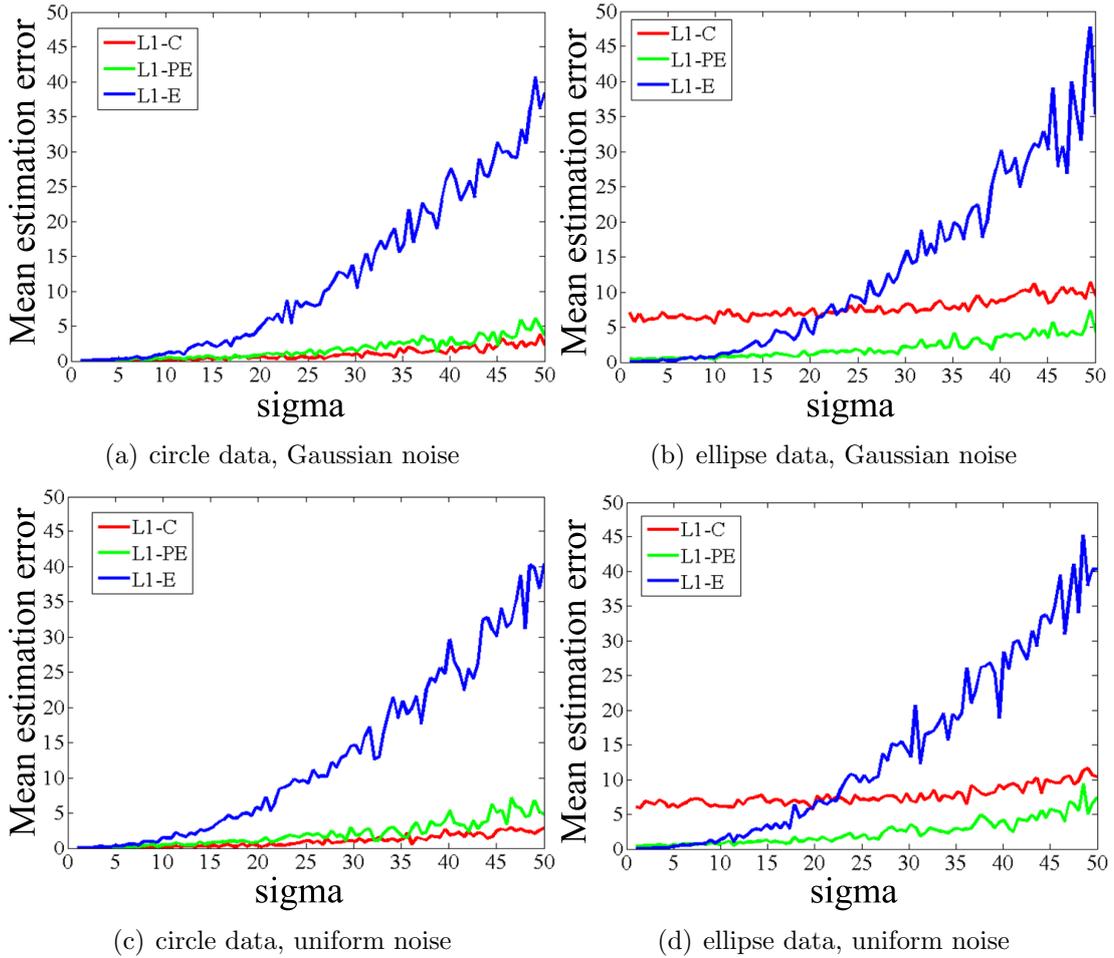


Figure 4.8: The mean estimation error as a function of noise level on the synthetic data.

width of the ring-shaped region along the radius direction. The inner boundary and the outer boundary of the ring-shaped region can be represented by $f(\theta, r + \Delta r_w)$ and $f(\theta, r - \Delta r_w)$, respectively. The ring-shaped region between the inner and outer boundary is normalised by the rubber sheet model in [6]. In this chapter, Δr_w is set to 30.

Fig. 4.2(b) shows the example of the ring-shaped regions of a good segmentation and a poor segmentation (note that the segmentation boundary is marked red and the boundaries of the ring-shaped region are marked green). The normalised ring-shaped region of the good segmentation is apparently divided into two halves. The top half is iris region, because the inner side of a good segmentation boundary should be all iris region. The bottom half of the good segmentation consists of sclera, skin and

silhouette. These parts correspond to the regions outside the segmentation boundary at the two sides, lower and upper eyelids. In contrast, in the poor segmentation, the skin and silhouette present in the top half due to the inaccurate segmentation boundary. In some other cases of poor segmentation, the segmentation boundary may be within the true limbic boundary. Correspondingly, the iris region may present in the bottom half of the normalised ring-shaped region.

HOG [117] is used as the feature to distinguish between the normalised ring-shaped region of good and poor segmentations. HOG feature calculates the histogram of gradient orientations in local cells, and all cells are normalised with a block-wise manner. HOG is a representation of local shape and appearance information. It has excellent invariance to local geometric and photometric transformations, as well as illumination variance. Therefore, it is chosen to describe the local shape and appearance of the unwrapped ring-shaped region while overcoming the noise and degradations in less constrained iris captures.

Some good and poor segmentations are selected as the training samples (please refer to Section 4.4.1 for details). A SVM with quadric hinge loss function as used in [122] is trained using the training samples, due to its reported effectiveness for image classification as in [122]. For the test iris images, the three ring-shaped regions corresponding to the segmentation of ℓ_1 -C, ℓ_1 -PE and ℓ_1 -E models are extracted and normalised. HOG features are extracted on the three normalised ring-shaped regions and are fed to the trained SVM. The segmentation result of the model with the highest classifier response is selected.

4.4 Experiments

This section performs experimental analysis for the proposed iris segmentation algorithm. Firstly, the experimental setting is introduced, including datasets, the

setting to analyse each individual iris segmentation model, the setting to analyse model selection, and the method for performance evaluation. Then, analysis is performed for each individual iris segmentation model described in Section 4.2. This includes: (1) the effect of components in each individual iris segmentation model; (2) the performance comparison of each individual model; (3) a comparison between ℓ_1 -norm regression and traditional least square (ℓ_2 -norm) regression to demonstrate the advantage of using ℓ_1 -norm regression in proposed method for iris segmentation. Next, analysis is performed for the model selection method described in Section 4.3. It consists of three analysis: (1) the effect of model selection; (2) the effect of training data on the performance of model selection on colour captures; (3) the impact of capturing device on the performance of model selection on mobile data. Finally, the overall performance of proposed iris segmentation algorithm is compared with state-of-the-art algorithms.

4.4.1 Experimental setting

Datasets. The experimental analysis are performed on four datasets described in Chapter 3: MICHE, UBIRIS2, FRGC and CASIAD. For MICHE dataset, 300 images from 'iPhone5' folder and 300 images from 'SamsungGalaxyS4' folder are randomly selected. Classical Viola-Jones detector [113] is used to extract the eye region in the selected images. The false positives are manually eliminated since they do not include eye region, and the remaining 569 eye regions are used for the experiment. The ground truth is segmented manually for the 569 images. Also, since the images in MICHE dataset have generally high resolutions, to reduce the computational cost, the segmentation is performed on downsampled eye regions for this dataset, and the segmented boundaries are mapped back to the original resolution. The size of downsampled eye region is chosen to be 300×400 . The size is designed to be the same as the images in UBIRIS2 [19], which has been a bench-mark for colour iris segmentation.

For UBIRIS2 dataset, two subsets released for NICE contest [18] are adopted. The first subset is the combination of training and test sets released for NICE I contest. It includes 945 images with ground truth. 500 images in this subset are from NICE I training set, and the rest images are from NICE I test set. The second subset is the released training set for NICE II contest. This subsets consist of 1000 images with ground truth.

For FRGC dataset, the subset constructed in [21] is adopted. This subset includes 500 images with manually segmented ground truth.

For CASIAD dataset, the subset used in [41] is adopted. This subset includes 581 images with manually segmented ground truth. These images are partitioned into a training set with 79 images and a test set with 502 images as in [41].

The setting to analyse individual iris segmentation models. For colour captures in colour datasets (MICHE, UBIRIS2 and FRGC), the parameters in the proposed method are determined based on the NICE I training set with 500 images in the first subset of UBIRIS2 as described above. The following data is used to test the performance: the MICHE subset with 569 images, the FRGC subset with 500 images, the second UBIRIS2 subset with 1000 images (the training set of NICE II contest).

For NIR captures in the NIR dataset (CASIAD), the training and test partition in [41] are adopted for parameter tuning and test, respectively.

The setting to analyse model selection. To perform model selection, a SVM needs to be trained using example good and poor segmentations. Different training data are used for mobile captures (MICHE), static camera captures (FRGC, UBIRIS2) and NIR captures (CASIAD). Note that although mobile captures and static camera captures are obtained in the same wavelength (colour), it is experimentally found that they require more specific training data to achieve good performance (please see Section 4.4.3 for details).

For MICHE subset, 100 good segmentations and 100 bad segmentations of 140 images are selected for training. The remaining 429 images are used for the test of performance.

For static camera captures (FRGC, UBIRIS2), the SVM is trained using 950 good segmentations and 950 poor segmentations selected from the first subset of UBIRIS2 (the combination of training and test sets for NICE I contest). The test is performed using the images in FRGC dataset and the second UBIRIS2 subset (the training set of NICE II contest).

For CASIAD dataset with NIR captures, the model selection is analysed using leave-one-out test on the test partition of the subset in [41] with 502 images. On CASIAD dataset, leave-one-out test is adopted instead of fixed training/test sets, because it is experimentally found that the data in this dataset is insufficient for both SVM training and performance test. For each training partition of leave-one-out test, 230 good segmentations and 230 bad segmentations are selected to train the SVM.

Performance evaluation. The performance of iris segmentation is evaluated by E1-score which is widely used in a number of research to evaluate the accuracy of iris segmentation such as [18, 21, 38, 41]. E1-score is calculated as follows:

$$E1 = \frac{1}{\text{width} \times \text{height}} \sum_c \sum_r O(c, r) \oplus C(c, r) \quad (4.11)$$

where width and height are the size of an iris image; O and C denote the ground truth and the iris mask generated by an iris segmentation algorithm, respectively; \oplus is the exclusive-or operator. A lower E1-score means a more accurate segmentation.

4.4.2 Analysis for each individual iris segmentation model

This subsection includes three analysis: (1) the effect of components in each individual iris segmentation model; (2) the performance comparison of three individual

Table 4.1: Significance analysis of each component in individual iris segmentation models.

Database	Evaluation	LPS	EF	RR	SR
MICHE	E1(ℓ_1 -C)	5.29%	3.89%	3.42%	3.36%
	p(ℓ_1 -C)	N/A	$6.9567e-74$	$1.0226e-36$	$2.8819e-10$
	E1(ℓ_1 -PE)	2.85%	2.27%	2.17%	2.15%
	p(ℓ_1 -PE)	N/A	$5.8372e-55$	$9.5779e-09$	0.0186
	E1(ℓ_1 -E)	4.29%	2.94%	2.60%	2.60%
	p(ℓ_1 -E)	N/A	$1.2501e-73$	$2.1517e-19$	0.5553
FRGC	E1(ℓ_1 -C)	2.88%	2.19%	2.01%	1.99%
	p(ℓ_1 -C)	N/A	$1.8089e-69$	$1.1248e-17$	$8.6030e-10$
	E1(ℓ_1 -PE)	2.06%	1.59%	1.48%	1.46%
	p(ℓ_1 -PE)	N/A	$9.7993e-45$	$8.1117e-19$	$5.4211e-11$
	E1(ℓ_1 -E)	3.47%	2.44%	2.11%	2.08%
	p(ℓ_1 -E)	N/A	$1.4905e-48$	$1.9633e-32$	$2.2447e-06$
UBIRIS2	E1(ℓ_1 -C)	2.69%	1.85%	1.69%	1.63%
	p(ℓ_1 -C)	N/A	$1.1128e-96$	$1.9087e-73$	$2.6788e-32$
	E1(ℓ_1 -PE)	2.17%	1.71%	1.57%	1.50%
	p(ℓ_1 -PE)	N/A	$1.3684e-62$	$6.4429e-84$	$1.6122e-32$
	E1(ℓ_1 -E)	2.75%	1.98%	1.76%	1.69%
	p(ℓ_1 -E)	N/A	$4.3336e-102$	$5.7369e-84$	$1.8251e-29$
CASIAD	E1(ℓ_1 -C)	2.72%	1.11%	1.01%	0.91%
	p(ℓ_1 -C)	N/A	$6.0363e-172$	$1.5317e-15$	$5.3839e-45$
	E1(ℓ_1 -PE)	2.20%	1.31%	1.19%	1.09%
	p(ℓ_1 -PE)	N/A	$2.0403e-100$	$2.4634e-19$	$4.4462e-68$
	E1(ℓ_1 -E)	2.72%	1.38%	1.24%	1.12%
	p(ℓ_1 -E)	N/A	$4.8596e-120$	$6.4164e-16$	$3.6213e-63$

models; (3) a comparison between ℓ_1 -norm regression and traditional least square (ℓ_2 -norm) regression to demonstrate the advantage of using ℓ_1 -norm regression in the proposed method. Note that the results in this subsection are based on the setting to analyse individual iris segmentation models described in Section 4.4.1.

The effect of components in the three individual models. The main components of the three models in Section 4.2 are: limbic and pupillary boundary segmentation (LPS), eyelids fitting (EF), reflection removal (RR) and silhouette removal (SR). To study the effect of each component, Tab. 4.1 reports the mean E1-score of the output after each component. Also, to analyse the statistical significance, paired student-t test is performed between each component and its previous component,

and the p value is reported in Tab. 4.1. The results in Tab. 4.1 show that each component in the proposed method leads to an improvement with generally good statistical significance (the p value is lower than 0.05 level).

The performance comparison of three individual models. From Tab. 4.1, it can be seen that, for colour datasets (MICHE, FRGC, UBIRIS2), ℓ_1 -PE model is the best individual model. For NIR dataset (CASIAD), ℓ_1 -C is the best individual model. An explanation of this observation is that the colour datasets are more noisy, so the shape of iris in these datasets is more variant and such shape can be better modelled by ℓ_1 -PE model; as for NIR dataset (CASIAD), since this dataset includes relatively lower noise compared to colour datasets, the shape of iris in this dataset is closer to a circle and such shape can be better described by the circle model. It is similar to the analysis in Section 4.2.8.

ℓ_1 -norm regression vs. ℓ_2 -norm regression. To validate the advantage of using ℓ_1 -norm regression in iris segmentation, a comparison is performed to study the performance between ℓ_1 -norm regression and traditional least square regression (ℓ_2 -norm regression). Specifically, the following methods are compared: (1) the three models proposed in Section 4.2.1; (2) the three models proposed in Section 4.2.1 with the ℓ_1 -norm in Eqn. 4.5, Eqn. 4.7, Eqn. 4.8 and Eqn. 4.10 replaced by ℓ_2 -norm (*i.e.* ℓ_2 -norm regression). To eliminate the influence of reflection and silhouette removal, the comparisons are made on the E1-scores after limbic and pupillary boundary segmentation (referred to as ℓ_1 -C*, ℓ_2 -C*, ℓ_1 -PE*, ℓ_2 -PE*, ℓ_1 -E* and ℓ_2 -E*), and the E1-scores after eyelids fitting (referred to as ℓ_1 -C**, ℓ_2 -C**, ℓ_1 -PE**, ℓ_2 -PE**, ℓ_1 -E** and ℓ_2 -E**).

The E1-scores of circle, parametric ellipse and ellipse models are shown in Tab. 4.2, Tab. 4.3 and Tab. 4.4, respectively. It is also reported in the three tables that the p value of paired student-t test between the E1-score of ℓ_1 -norm and ℓ_2 -norm regression of the same model. It can be seen that the E1-score of ℓ_1 -norm regression is consistently lower than ℓ_2 -norm regression on all the test data. With parametric

Table 4.2: Comparison between ℓ_1 -norm and ℓ_2 -norm regression on the proposed circle model.

Database	Evaluation	ℓ_1 -C*	ℓ_2 -C*	ℓ_1 -C**	ℓ_2 -C**
MICHE	E1	5.29%	6.43%	3.89%	4.09%
	p	$6.9815e-40$	N/A	0.0180	N/A
FRGC	E1	2.88%	3.44%	2.19%	2.27%
	p	0.0036	N/A	0.6674	N/A
UBIRIS2	E1	2.69%	2.88%	1.85%	1.98%
	p	$8.3212e-13$	N/A	$8.2319e-07$	N/A
CASIAD	E1	2.72%	2.99%	1.11%	1.44%
	p	$7.5981e-12$	N/A	$2.0818e-16$	N/A

Table 4.3: Comparison between ℓ_1 -norm and ℓ_2 -norm regression on the proposed parametric ellipse model.

Database	Evaluation	ℓ_1 -PE*	ℓ_2 -PE*	ℓ_1 -PE**	ℓ_2 -PE**
MICHE	E1	2.85%	3.72%	2.27%	2.74%
	p	$9.7956e-28$	N/A	$5.1516e-13$	N/A
FRGC	E1	2.06%	2.72%	1.59%	1.97%
	p	$5.2249e-17$	N/A	$2.7401e-09$	N/A
UBIRIS2	E1	2.17%	3.04%	1.71%	2.33%
	p	$2.9606e-60$	N/A	$1.0080e-44$	N/A
CASIAD	E1	2.20%	2.68%	1.31%	1.73%
	p	$1.5809e-15$	N/A	$4.8459e-17$	N/A

Table 4.4: Comparison between ℓ_1 -norm and ℓ_2 -norm regression on the proposed ellipse model.

Database	Evaluation	ℓ_1 -E*	ℓ_2 -E*	ℓ_1 -E**	ℓ_2 -E**
MICHE	E1	4.29%	5.47%	2.94%	3.69%
	p	$4.0023e-27$	N/A	$3.1258e-15$	N/A
FRGC	E1	3.47%	5.62%	2.44%	3.83%
	p	$2.7166e-09$	N/A	$4.6631e-05$	N/A
UBIRIS2	E1	2.75%	4.63%	1.98%	4.57%
	p	$2.1183e-52$	N/A	$9.4885e-81$	N/A
CASIAD	E1	2.72%	3.49%	1.38%	1.92%
	p	$1.0275e-16$	N/A	$2.1702e-09$	N/A

ellipse and ellipse models, ℓ_1 -norm regression shows excellent statistical significance over ℓ_2 -norm regression. However, the student-t test between ℓ_1 -C** and ℓ_2 -C** fails on FRGC, although the p value of circle model is good on the other datasets.

Table 4.5: Results of the proposed model selection method.

Dataset	Evaluation	ℓ_1 -C	ℓ_1 -PE	ℓ_1 -E	MS*	r-score
MICHE	E1	3.14%	2.01%	2.25%	1.93%	1.75%
	p	$3.9348e-63$	0.0051	$4.9048e-12$	N/A	N/A
FRGC	E1	1.99%	1.46%	2.08%	1.37%	1.26%
	p	$6.0200e-04$	$9.1162e-05$	$2.1817e-11$	N/A	N/A
UBIRIS2	E1	1.63%	1.50%	1.69%	1.43%	1.30%
	p	$6.7869e-17$	$5.1991e-06$	$1.4276e-17$	N/A	N/A
CASIAD	E1	0.91%	1.09%	1.12%	0.86%	0.74%
	p	0.0165	$3.0947e-09$	$9.4443e-05$	N/A	N/A

*MS-model selection

For this observation, a possible reason other than lack of data is that the eyelids fitting is influenced by inaccurate limbic boundary fitting. Many images in FRGC suffer from heavy eyelids occlusion, but the proposed circle model is less flexible to the heavy eyelids occlusion. The circle model seeks limbic boundary candidates at both sides, and heavy eyelids occlusion results that there are little limbic boundary even at both sides. In such case, circle model finds too many false limbic boundary candidates and hence the fitted limbic boundary is inaccurate. The following eyelids fitting is influenced, because it depends on the result of limbic boundary fitting. The error of limbic boundary fitting is too large, so ℓ_1 -norm and ℓ_2 -norm regression make little difference for eyelids fitting on FRGC dataset.

4.4.3 Analysis for model selection

This subsection consists of three analysis: (1) the effect of model selection; (2) the effect of training data on the performance of model selection on colour captures; (3) the impact of capturing device on the performance of model selection on mobile data. Note that the results in this subsection are based on the setting to analyse model selection described in Section 4.4.1.

Effect of model selection. This experiment investigates the effect of model selection. Tab. 4.5 reports the E1-score of each individual model and model selection on

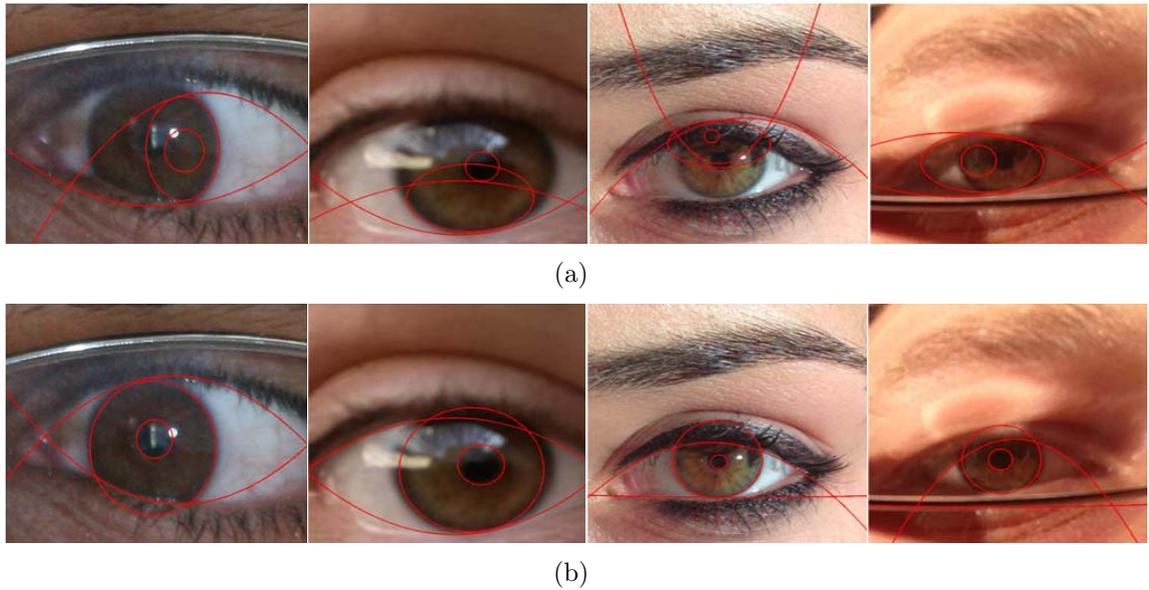


Figure 4.9: An illustration of the effect of the model selection. (a) Some examples where the best individual iris segmentation model as reported in Tab. 4.5 fails; (b) the iris segmentation result given by the model selection. It can be seen that the model selection is able to give a more reliable segmentation when the best individual model fails.

all the datasets. It is also reported in Tab. 4.5 that the p value of paired student-t test between the E1-scores of each individual model and model selection. Finally, a reference score (r-score) is reported in Tab. 4.5. The r-score is the mean of the minimum E1-score of each image's three segmentations obtained by the three models. It gives the highest performance that can be achieved by model selection.

It can be seen that the proposed model selection leads to a generally improved performance compared to each individual model with good significance (the p value is lower than 0.05 level). On the other hand, on MICHE dataset, the model selection leads to an improvement of $\frac{(2.01-1.93)}{(2.01-1.75)} = 30.77\%$ from the best individual model (ℓ_1 -PE for this dataset) towards the r-score. This percentage of improvement on UBIRIS2, FRGC and CASIAD are 35%, 45% and 29.41%, respectively.

Examples are shown in Fig. 4.9 to further illustrate the effect of proposed model selection. Fig. 4.9(a) shows some images where the best individual model as reported in Tab. 4.5 fails. Fig. 4.9(b) shows the result of selected model by the proposed model

Table 4.6: Mean E1-score of proposed model selection method using different training data for colour datasets.

Dataset	MICHE (M [*])	FRGC (SC [#])	UBIRIS2 (SC [#])
training set 1 (M [*])	1.93%	1.58%	1.57%
training set 2 (SC [#])	2.23%	1.37%	1.43%

*M–mobile data # SC–static camera data

selection for the same images in Fig. 4.9(a). It can be seen that model selection is able to seek a more reliable segmentation.

Effect of training data for model selection on colour captures. In Section 4.4.1, it is mentioned that, although the images in MICHE, FRGC and UBIRIS2 datasets are captured in the same wavelength (colour), different training data are required for mobile captures (MICHE) and static camera captures (FRGC and UBIRIS2). This part of experiment performs detailed analysis on the effect of different training data for colour iris captures. The training set used for the MICHE dataset is referred to as 'training set 1', and the training set for FRGC and UBIRIS2 datasets is referred to as 'training set 2'. Tab. 4.6 reports the mean E1-score of the proposed model selection method with different training sets on MICHE, FRGC and UBIRIS2 datasets.

It can be seen from Tab. 4.6 that the performance drops if static camera data is used for training and mobile data is used for test, and vice versa. A possible reason is that there exists inherent characteristic differences between the static camera data and mobile data.

Impact of capturing device for model selection on mobile data. In real applications, the mobile data can be captured by varying devices with different sensors and settings. Therefore, this part of experiment investigates the impact of capturing device on the overall performance of model selection on mobile data. This investigation is performed by studying the performance of model selection on MICHE dataset using device-specific training and test data. The training set of MICHE for model selection is split based on the capturing device. It results

Table 4.7: Mean E1-score on MICHE dataset with device-specific training and test data.

training \ test	iPhone5	Samsung Galaxy S4	Both devices
iPhone5	1.95%	1.94%	1.94%
Samsung Galaxy S4	1.93%	1.95%	1.94%
Both devices	1.93%	1.92%	1.93%

a iPhone5 training set with 59 good segmentations and 52 poor segmentations, and a Samsung Galaxy S4 training set with 41 good segmentations and 48 poor segmentations. The similar splitting is performed on the test set of MICHE for model selection. It leads to a iPhone5 test set with 266 images, and a Samsung Galaxy S4 test set with 163 images. Tab. 4.7 reports the mean E1-score of the model selection results with different combinations of these device-specific training and test sets. It can be seen from Tab. 4.7 that the mean E1-score changes little with varying data from different devices. This observation suggests that for the images captured by iPhone5 and Samsung Galaxy S4, the device variance has little influence on the overall performance of model selection.

4.4.4 Comparison with other methods

In this subsection, the proposed method is compared with 5 recently proposed methods [21, 38, 39, 41, 108]. These methods have achieved excellent iris segmentation performance for less constrained iris captures. [38] has achieved the best performance in NICE I contest [18]. The other four methods have reported a further improvement over [38]. Note that the results in this subsection are based on the test data for model selection described in Section 4.4.1, since the performance of model selection represents the final performance of the proposed algorithm. The five methods are not reimplemented for this comparison. Instead, for FRGC and UBIRIS2 datasets, the performance reported in [21, 41, 108] is cited. For CASIAD dataset, the performance reported in [41] is cited.

Table 4.8: Comparison of the performance with other methods.

Method	MICHE	FRGC	UBIRIS2	CASIAD
Tan <i>et al.</i> 2010 [38]	-	3.30%	-	-
Proença 2010 [39]	-	2.42%	-	-
Tan and Kumar 2012a [21]	-	1.84%	1.90%	-
Tan and Kumar 2012b [108]	-	1.63%	1.81%	1.09%
Tan and Kumar 2013 [41]	-	1.76%	1.72%	0.81%
Proposed	1.93%	1.37%	1.43%	0.86%

The mean E1-scores are reported in Tab. 4.8. For MICHE and FRGC, the mean E1-score on all the test data is reported. For UBIRIS2, to be consistent with the results in [21,41,108], the mean E1-score on the last 904 images in the test data is reported. For CASIAD, since it does not have enough data to train SVM for model selection, a leave-one-out test is performed to estimated the final performance of the proposed algorithm on the test data.

It can be seen that the proposed model selection method achieves top performance. It performs better than all the comparison methods on FRGC and UBIRIS2 datasets, while its performance is close to the top method on CASIAD dataset. It is also observed that the mean E1-score of the proposed method drops on MICHE dataset compared with the other three datasets. The reason is that the iris images captured by mobile device are more challenging, especially for the images captured outdoor. These images generally suffer from heavy reflection, resulting in drop of performance.

In terms of speed, the proposed method needs more than 20 seconds to segment 1 image on a desktop with Intel i5-3470 quad-core 3.20GHz CPU, 16GB RAM, Windows 7 64bit system and Matlab 2013a 64bit. This speed is more than 600 times slower than current commercial implementations of Daugmans iris recognition algorithm [6,7] which are able to perform tasks from iris segmentation to generating the IrisCode within 30 milliseconds. However, the proposed algorithm is designed to achieve robust iris segmentation for noisy iris captures, and its speed is not optimised. Several methods can be explored in the future research to improve the

speed. For example, for captures with less noise, the search region for iris centre candidate may be reduced; parallel implementations can be considered to improve the speed; the code can be optimised and run with a faster implementation language.

4.5 Summary

In this chapter, a robust algorithm is proposed for less constrained iris segmentation. This algorithm consists of a ℓ_1 -norm regression based iris segmentation algorithm with three models, and a model selection method to select the best model as the final segmentation result. Comprehensive experimental analysis is made for each individual model as well as the model selection method. The influence of training data and capturing device on the proposed algorithm are also investigated. The comparison with state-of-the-art algorithms demonstrates the top performance of the proposed algorithm. Future work may explore methods like parallel implementation to improve the speed.

Chapter 5

Iris liveness detection using regional features

Part of this chapter has been adapted from the journal paper [J7] (published) as in the List of Publications.

5.1 Introduction

Iris liveness detection aims to protect iris recognition systems from spoofing attacks using fake iris patterns, such as iris printouts, textured contact lenses and artificial eyes. Iris liveness detection is usually considered as a two-class classification problem. The key of this problem is to seek effective features to distinguish between real and fake iris patterns. Research on iris liveness detection have proposed numerous local and global features [43, 44, 46, 47, 49, 57, 60–62, 73, 75]. However, most of these research focus on the low-level features directly extracted from image pixels. Little recent research investigate to build features based on the high-level relationship among the features extracted from different regions. Compared with low-level features, the features based on the high level feature relationship not only include the information extracted from pixels, but also provide a deeper insight into the feature

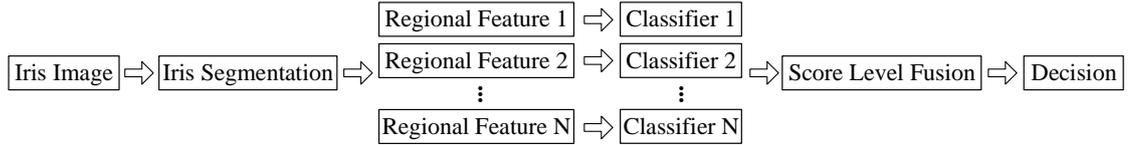


Figure 5.1: Flowchart of the proposed method for iris liveness detection.

distribution in different regions. Thus, it is reasonable to expect that the features based on the high-level feature relationship are able to achieve a better performance in comparison to using low-level features only.

In this chapter, iris liveness detection is performed using regional features which model the high-level distribution information of low-level features extracted from neighbouring regions. Two models are investigated to construct regional features: spatial pyramid and relational measure which seek the feature distributions in regions with varying size and shape respectively. The spatial pyramid model extracts features from coarse to fine grid regions. It models a local to global feature distribution. The local distribution captures the local feature variations, while the global distribution includes the information that is more robust to translational transform. The relational measure is based on a feature-level convolution operation defined in this chapter. By varying the shape of the convolution kernel, the feature distribution in regions with different shapes is obtained. To combine the feature distribution information in regions with varying size and shape, the results based on the two models are fused at the score level.

Based on regional features, a method is proposed for iris liveness detection. The flowchart of this method is shown in Fig. 5.1. Firstly, for an input iris image, iris segmentation is performed to find the iris region in the image. Then, different regional features are extracted from the iris region. These regional features are fed to suitably trained classifiers to produce a set of responses. Next, the classifier responses are fused at the score level. Finally, the fused response is used to distinguish between real and fake iris patterns.

Experiments are performed on four datasets. The results show that: (1) regional features are able to achieve an improved performance compared to low-level features; (2) on three of the four datasets, regional feature based iris liveness detection method is able to achieve comparable performance to state-of-the-art methods; on a dataset with heavier noise, regional feature based method is able to achieve better performance than state-of-the-art methods.

The remainder of this chapter is organised as follows. Section 5.2 presents the proposed method for iris liveness detection based on regional features. Section 5.3 reports the experimental results. Section 5.4 summarises this chapter.

5.2 Methodology

5.2.1 Iris segmentation

Given an iris image, iris segmentation is performed to find the iris region. The algorithm in Chapter 4 is employed to seek the limbic boundary, lower and upper eyelids. Then, a bounding rectangle is determined around these boundaries. The region within this rectangle is extracted as the iris region. All the features in the proposed method are extracted from the iris region.

5.2.2 Spatial pyramid model

Spatial pyramid is used as the first model to construct the regional features. Spatial pyramid model extracts low-level features from increasingly finer regions [123], and it concatenates the features at each region to form the feature vector. Since the features are calculated from regions with different size, spatial pyramid structure is able to capture both local and global information of an image. Also, a pooling operation [122] can be performed based on the spatial pyramid model (referred as spatial

pyramid pooling in this chapter). This operation builds the feature relationship in regions with different size by calculating the local maxima of the features.

An illustration of spatial pyramid based regional features is shown in Fig. 5.2, where Fig. 5.2(a) illustrates the procedure of spatial pyramid model based feature extraction and Fig. 5.2(b) describes the spatial pyramid pooling operation. From Fig. 5.2(a), it can be seen that low-level features are extracted from 3 levels. Each level gives a specific partition of the iris image. In level 0, the feature is computed globally on the whole image. In level 1 and level 2, features are extracted from 4 sub-regions (red rectangles) and 16 sub-regions, respectively. It is clear that level 0 corresponds to the coarsest partition and level 2 the finest partition. For each level, spatial pyramid pooling is performed to construct a feature vector, and the 3 feature vectors from the 3 levels are concatenated to form the final feature vector.

In Fig. 5.2(b), the level 2 in Fig. 5.2(a) is used as an example to explain the procedure of spatial pyramid pooling. For the 16 sub-regions in level 2, there are three pooling levels. Each pooling level defines a specific partition of the 16 sub-regions and produces a number of feature vectors. The feature vectors in all pooling levels are concatenated to form the final feature. In pooling level 2 as shown in Fig. 5.2(b), the partition is the original 16 sub-regions (red rectangles), and the produced features are those extracted from the 16 red rectangles. Meanwhile, it can be seen that a coarser partition with 4 larger sub-regions (marked as blue, green, yellow and purple rectangles) is also defined in pooling level 2. The 4 larger sub-regions correspond to the partition in pooling level 1. However, in pooling level 1, the produced features are not directly extracted from the 4 larger sub-regions. Taking the green rectangle for an example, the feature in the green rectangle is formed by pooling (see later for details) the features in the red rectangles (*i.e.* the partition of the previous pooling level) within the green rectangle. Similarly, in pooling level 1, a coarser partition (the black rectangle) corresponding to pooling level 0 is defined, and the produced feature for the black rectangle is formed by pooling the features of the 4 sub-regions

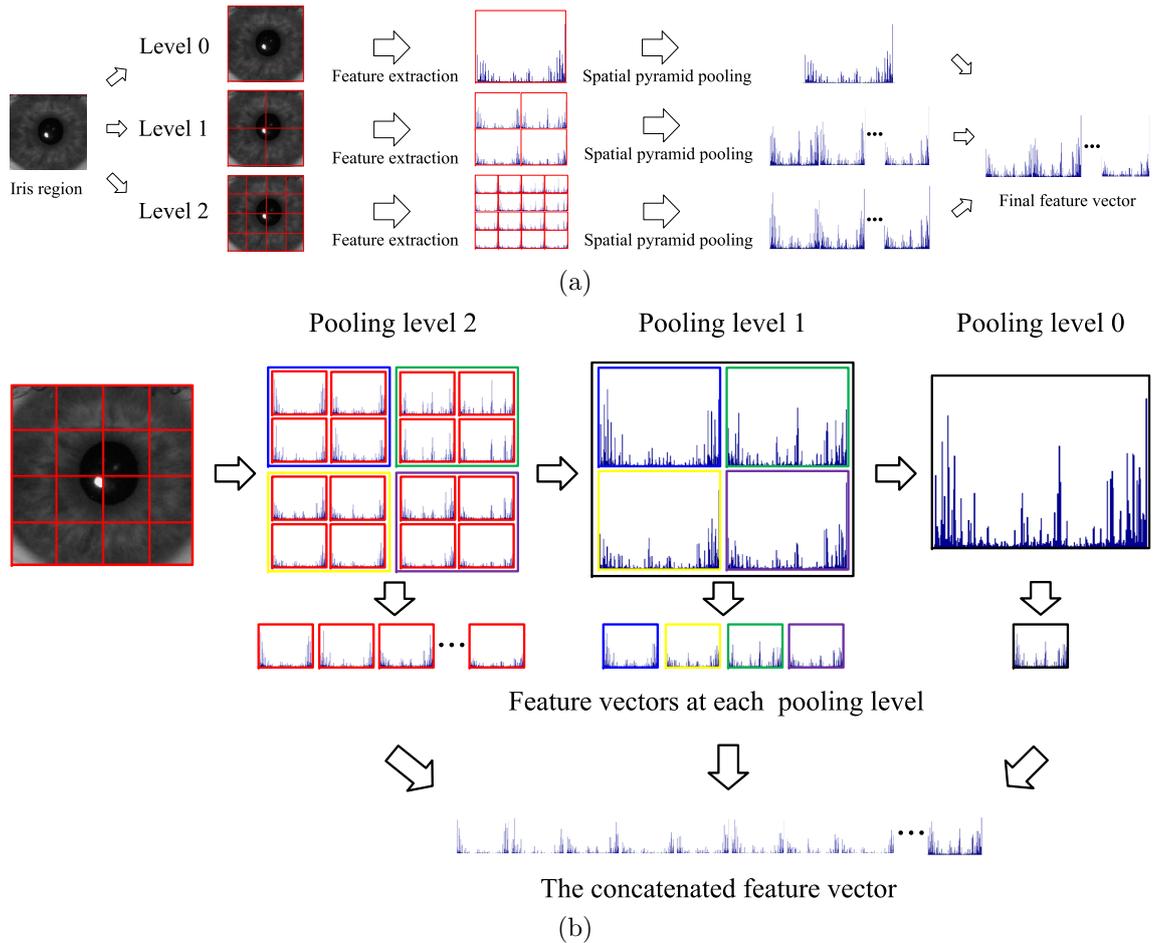


Figure 5.2: An illustration of the spatial pyramid-based regional feature. (a) Illustration of the whole procedure of spatial pyramid structure-based feature extraction; (b) illustration of spatial pyramid pooling step using pooling level 2 in (a) as an example.

in pooling level 1 within the black rectangle.

The procedure of spatial pyramid pooling is similar for level 1 and level 0 in Fig. 5.2(a). The only difference is the number of pooling levels considered each time. For level 1, there are two pooling levels with 4 and 1 sub-regions, respectively. For level 0, there is only one pooling level. In other words, the image is regarded as a whole at this level.

There are mainly two kinds of pooling: the average pooling and the max pooling [122, 124]. Given multiple feature vectors, the average pooling forms a new feature vector by calculating the mean of all the feature vectors; the max pooling uses the maximum

value at each dimension to form the new feature vector. In this chapter, the max pooling is adopted following [122]. At each pooling level shown in Fig. 5.2(b), the local maxima of the features are computed within regions defined by each specific partition. Therefore, the max pooling gives the information on the local to global feature distribution. Moreover, max pooling enhances the translational invariance of the resulting features. In the coarsest pooling level (an example is the pooling level 0 in Fig. 5.2(b)), the output feature is essentially the maxima of all the feature vectors in the whole image, and this changes only little with translational transform.

By extracting features from multiple levels with coarse to fine partitions (as shown in Fig. 5.2(a)), the final feature vector includes both global and local information. By performing spatial pyramid pooling in each level (as illustrated in Fig. 5.2(b)), the final feature vector also reveals the local to global feature distributions and the translational invariance is enhanced, since the maxima of all the feature vectors changes only little with translational transform.

5.2.3 Relational measure model

Recent research have shown the effectiveness of order relationship of neighbouring pixels for iris recognition [20, 125]. Inspired by these research, in this chapter, a relational measure model is proposed at the feature-level for iris liveness detection. This model aims to seek the feature distribution information in neighbouring regions with varying shapes.

A feature-level convolution operation is defined for the relational measure model. Let $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n]$ be the features extracted from n neighbouring regions. Let $\mathbf{k} = [k_1, k_2, \dots, k_n]$ be a convolution kernel. The convolution between \mathbf{F} and \mathbf{k} is defined as follows:

$$\mathbf{F} * \mathbf{k} = \sum_i k_i \mathbf{f}_i \quad (5.1)$$

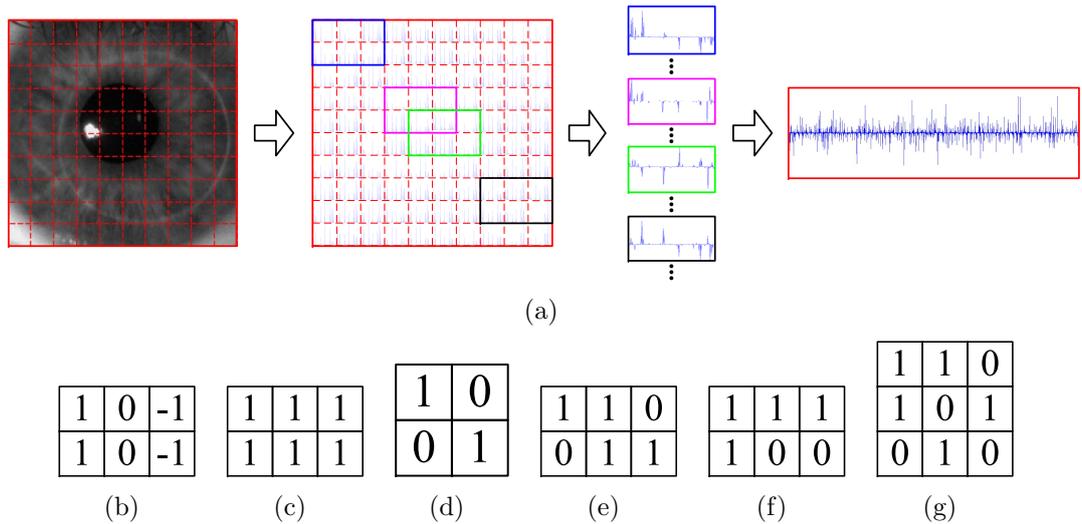


Figure 5.3: An illustration of the relational measure model on an iris region with soft contact lens. (a) A flowchart of relational measure-based regional feature extraction; firstly, an input iris region is divided into regular grid regions; then, low-level features are extracted from each region; a kernel slides across all the positions and feature-level convolution operation is performed at each position; example positions of the kernel are shown as solid colour rectangles in the flowchart; note that, in this flowchart, it is assumed the size of kernel is 2×3 ; finally, the final feature is formed by concatenating the results of feature-level convolution at each position; (b)-(g) examples of the kernels used in this chapter.

Based on the feature-level convolution operation, the relational measure is illustrated as follows. An input image is partitioned into multiple regions using regular grids, and low-level features are extracted from each grid. Then, a convolution kernel slides across all the grids and feature-level convolution operation is performed at each position. This procedure is similar to image filtering with a fixed kernel. The convolution results at all the positions are concatenated to form the regional feature. An illustration of this procedure is shown in Fig. 5.3(a).

The feature distribution information in regions with different shapes can be obtained by varying the shape of convolution kernel. In this chapter, the iris image is partitioned into 10×10 grids, and 24 convolution kernels are designed. Some examples of the convolution kernels are shown in Fig. 5.3(b) to Fig. 5.3(g). Note that these kernels are normalised in the experiment. That is, the positive weights are normalised to have a sum of 1, and the negative weights are normalised to have

a sum of -1 .

These kernels are chosen by experimentally testing the performance of a set of different kernels designed based on different intuitions or randomly generated. For example, the kernel in Fig. 5.3(b) calculates the difference between the features in neighbouring regions; it can be considered as a horizontal gradient operator and it extracts the ‘texture’ of local features; the kernel in Fig. 5.3(c) calculates the centroid of local features in a ‘fat’ window; it is able to mitigate the noise in each individual regions, while it captures the feature distribution with more weight on the horizontal direction; the kernels in Fig. 5.3(d) and Fig. 5.3(e) estimate the centroid of local features in diagonal and diagonal-like regions, respectively; these two kernels and their rotated variations are able to capture the information in some particular structures, like the light textures around the the soft contact lens in the eye image in Fig. 5.3(a). On the other hand, the kernels in Fig. 5.3(f) and Fig. 5.3(g) are not manually designed for a specific reason. They are selected from a set of randomly generated kernels by cross-validation on the training data in the experiment (see Section 5.3 for details of training data). It can be seen that, although they are randomly generated, both of the kernels have regular shape. Similar to the kernels in Fig. 5.3(d) and Fig. 5.3(e), such regular shape is able to capture the information in some particular structures. Taking the iris region in Fig. 5.3(a) as an example, it can be found that the kernel in Fig. 5.3(f) and its rotated variations can roughly cover some regions with the boundary of the soft contact lens, while the kernel in Fig. 5.3(g) can be viewed as a kind of diagonal-like kernels similar to the one in Fig. 5.3(e). Please note that, current selection method is based on experimental test; the principal and high level knowledge about the choice of kernels need to be studied in the future research.

5.2.4 Low-level features

To build regional features, some low-level features need to be extracted for the previously described spatial pyramid and relational measure models (*i.e.* low-level features need to be extracted from the red rectangles in Fig. 5.2 and the dashed red rectangles in Fig. 5.3(a)). The regional feature extraction is very flexible. Any low-level features with a fixed dimension can be utilised to construct regional features using the method described in Section 5.2.2 and Section 5.2.3. Such flexibility makes it possible to incorporate some state-of-the-art low-level features for iris liveness detection into the proposed method to build powerful regional features, such as the local descriptors studied in [62, 109]. In this chapter, 6 low-level features are extracted. It leads to 6 spatial pyramid-based regional features and $6 \times 24 = 144$ relational measure-based regional features.

F1: LBP histogram. LBP¹ [58, 126] is a texture feature. It encodes each pixel by a binary code. The binary code is calculated by comparing the intensity between a pixel and its neighbouring pixels. An example of LBP coding scheme is shown in Fig. 5.4. It can be seen that, for the centre pixel, its neighbouring pixels with intensity values higher than the centre pixel are assigned the binary value 1. The remaining neighbouring pixels are assigned the binary value 0. All the binary values of the neighbouring pixels are concatenated to obtain a binary code, and the decimal value of this binary code is assigned to the centre pixel as its feature value. LBP histogram computes the histogram of LBP feature values to enhance the robustness. The LBP histograms extracted from each partitioned region of spatial pyramid and relational measure models are normalised to have a sum of 1. To mitigate the influence of pupil, bin 256 is eliminated before normalisation, since this bin corresponds to a region with uniform feature value and it is very likely to be pupil.

¹code: <http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab>

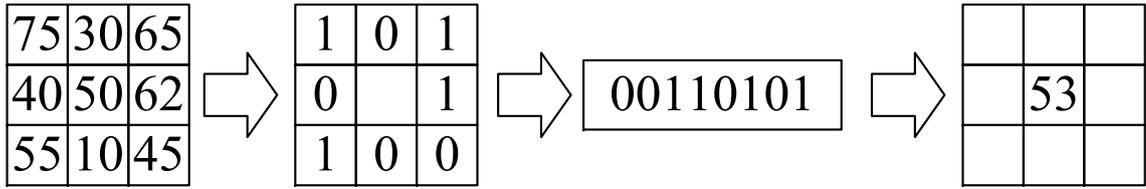


Figure 5.4: An example of LBP feature.

F2: LBP correlogram. The correlogram represents the frequency distribution of two feature values at a fixed distance [127]. It provides the pairwise distribution of feature values. Let $\mathbf{I}(\mathbf{x})$ be the LBP feature value at the position \mathbf{x} . Let $cor_d(m, n)$ be the bin (m, n) of the LBP correlogram with a fixed distance parameter d . $cor_d(m, n)$ is calculated by:

$$cor_d(m, n) = P(\mathbf{I}(\mathbf{x}_1) = m, \mathbf{I}(\mathbf{x}_2) = n \mid \|\mathbf{x}_1 - \mathbf{x}_2\| = d) \quad (5.2)$$

where $P(\bullet)$ calculates the probability of \bullet ; $\|\bullet\|$ calculates the distance between two positions. In this chapter, the distance at horizontal and vertical directions are considered. d is set to 2 (the parameters in this chapter are determined by cross-validation on the training data; see Section 5.3 for details of training data). Theoretically, the range of LBP feature value is between 0 and 255. It leads to a correlogram with $256^2 = 65536$ bins which is computationally expensive. To reduce the computational cost, the LBP feature values are quantized to k levels, and $k \ll 256$. The value of k is different for NIR and colour iris images and it is determined by cross-validation. Similar to LBP histogram, the influence of pupil is mitigated by eliminating $cor_d(k, k)$, and the LBP correlogram in each region is normalised to have a sum of 1.

F3: intensity histogram. The intensity histogram reflects the distribution of pixels amongst the 256 intensity levels. It consists of 256 bins and each bin counts the occurrence of a specific intensity level. The normalisation is the same as that of LBP histogram.

F4: intensity correlogram. Intensity correlogram calculates the correlogram of intensity. It is calculated using Eqn. 5.2 considering the intensity as the feature values. Similar to LBP correlogram, only the distance at horizontal and vertical directions is considered, and d in Eqn. 5.2 is set to 2. Also, the computational cost is reduced by quantizing the intensity values to fewer levels than 256. To mitigate the influence of pupil, $cor_d(0, 0)$ is eliminated in the intensity correlogram.

F5: LPQ histogram. LPQ² [65] is a texture descriptor. It extracts directional information in frequency domain. For a pixel, firstly, a short-time Fourier transform is performed on a local patch centred at this pixel. Then, in the frequency domain obtained by short-time Fourier transform, 4 coefficients corresponding to 4 directions are considered (see [65] for details). The real and imaginary parts of the 4 coefficients are quantized to 0 and 1 based on their signs. Finally, the quantized real and imaginary parts are concatenated to form a binary code, and the decimal value of this binary code is used as the feature value of this pixel. LPQ histogram calculates the histogram of LPQ feature values. It is normalised to have a sum of 1.

F6: SID histogram. SID³ [71, 72] is a scale and rotational invariant descriptor. It achieves such invariance by re-sampling the image in a log-polar grid and computing the discrete-time fourier transform (DTFT) of the re-sampled image. Since scaling and rotating the image only change the phase of DTFT of the re-sampled image, the magnitude of the DTFT is a scale and rotational invariant representation. Furthermore, directional derivative is used to enhance illumination invariance. In this chapter, the parameter setting in [72] is adopted. The log-polar grid includes 32 rays and 28 rings, and 4 orientations are considered for the computation of directional derivative. Also, only the low frequency Fourier coefficients are considered to construct the descriptor for each pixel. To obtain a histogram representation, a codebook is learned from SID descriptors extracted from training images, and the histogram representation is constructed using vector quantisation [123].

²code: <http://www.cse.oulu.fi/CMV/Downloads/LPQMatlab>

³code: <http://vision.mas.ecp.fr/Personnel/iasonas/descriptors.html>

5.2.5 Classifier

Support vector machine (SVM) is used for classification. Given a set of training features $\{\mathbf{f}_i\}_{i=1}^n$ and labels $\{y_i\}_{i=1}^n$, a hyperplane $\boldsymbol{\omega}$ is learned by:

$$\arg \min_{\boldsymbol{\omega}} \|\boldsymbol{\omega}\|_2^2 + \lambda \sum_i l(\boldsymbol{\omega}, \mathbf{f}_i, y_i) \quad (5.3)$$

where λ is a penalty parameter and $l(\boldsymbol{\omega}, \mathbf{f}_i, y_i)$ is the loss function. In this paper, the quadratic hinge loss function as used in [122] is adopted due to its reported effectiveness for image classification:

$$l(\boldsymbol{\omega}, \mathbf{f}_i, y_i) = [\max(0, \boldsymbol{\omega}^T \mathbf{f}_i y_i - 1)]^2 \quad (5.4)$$

where $y_i = 1$ for real samples and $y_i = -1$ for fake samples. After $\boldsymbol{\omega}$ is learned, given a test feature \mathbf{f}^{test} , the classification is based on the sign of $\boldsymbol{\omega}^T \mathbf{f}^{test}$. λ is set to 0.1 in this paper following [122].

5.2.6 Score-level fusion

Let r_1, r_2, \dots, r_t be the responses of t classifiers. Score-level fusion aims to find a set of weights $\alpha_1, \alpha_2, \dots, \alpha_t$ satisfying $\alpha_1 + \alpha_2 + \dots + \alpha_t = 1$, such that the fused score, denoted by r_s , is calculated by:

$$r_s = \alpha_1 r_1 + \alpha_2 r_2 + \dots + \alpha_t r_t \quad (5.5)$$

In this chapter, instead of fusing the classifier response of all the regional features, the top 3 features are used for score-level fusion. The top 3 features are chosen by cross validation on the training set. The optimal weights are computed by exhausting search on the training partition of the experiments.

Table 5.1: Information of the datasets used in this experiment

Dataset	Wavelength	Spoofing type	Image number
LivDet-Clarkson	NIR	Contact lenses	1356
LivDet-Warsaw	NIR	Printouts	1667
LivDet-ND	NIR	Contact lenses	4200
MobBIOfake	Visible	Printouts	1600

5.3 Experiments

5.3.1 Datasets and evaluation

The experiments are performed using four datasets for iris liveness detection introduced in Chapter 3: LivDet-Clarkson, LivDet-Warsaw, LivDet-ND and MobBIOfake. Tab. 5.1 revisits the information of each dataset. In this section, the partition of training and test sets varies for different experiments and will be introduced in each individual subsection.

The performance is evaluated using classification error rate. Classification error rate is defined as the ratio of the number of misclassified test images to the total number of test images. The lower the classification error rate, the better the performance.

5.3.2 Performance of single regional features

This subsection studies the performance of single regional features. For the three NIR datasets, the training and test partitions in LivDet-Iris 2013 competition [105] are adopted. For the colour dataset, the training and test partitions in MobILive 2014 competition [110] are used. For better visualisation, error rates of regional features are sorted in ascending order for each dataset. The sorted error rates on the four datasets are reported in Fig. 5.5 and Fig. 5.6, where Fig. 5.5 corresponds to spatial pyramid-based regional features, and Fig. 5.6 corresponds to relational measure-based regional features. In these figures, six colours are used to represent

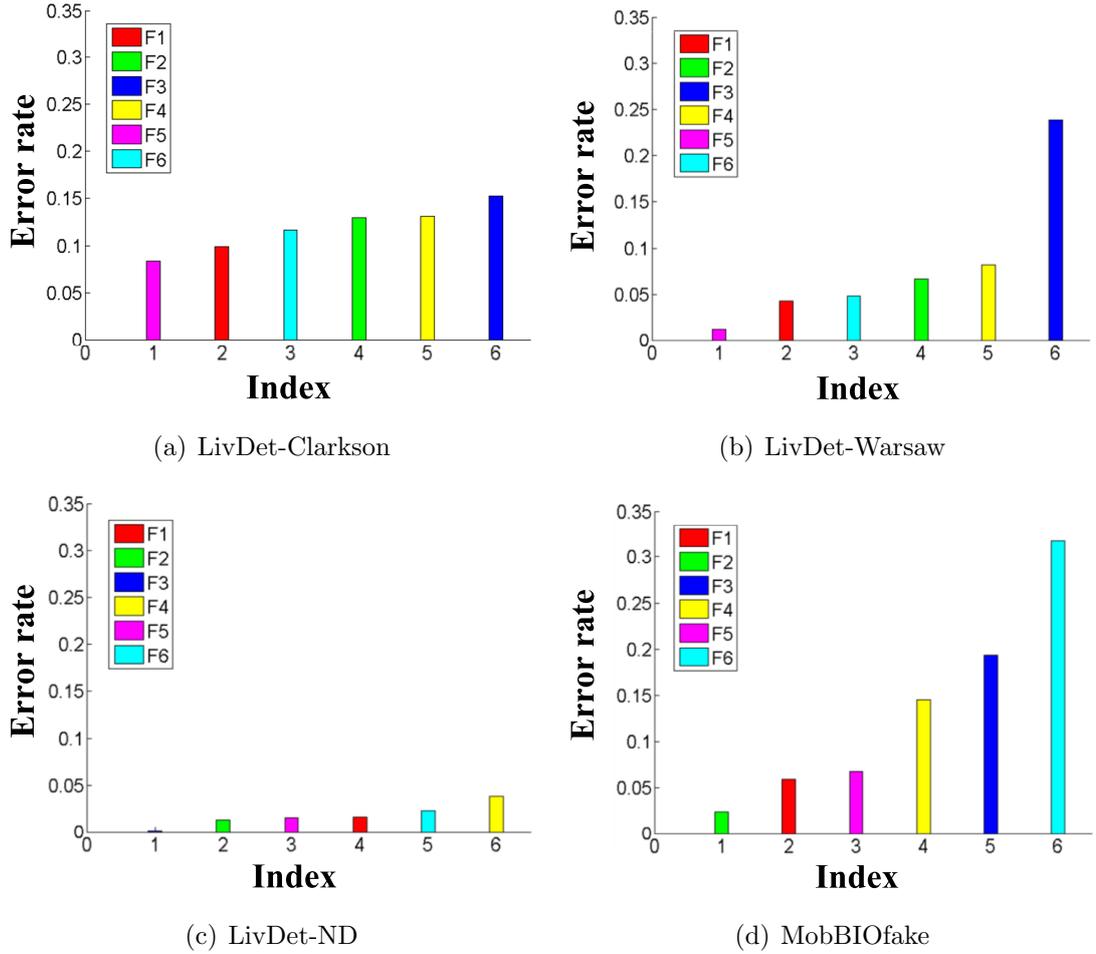


Figure 5.5: Performance of single regional features using spatial pyramid model.

the regional features based on six low-level features (F1 to F6), respectively.

Several conclusions can be drawn from the results in Fig. 5.5 and Fig. 5.6. First, generally speaking, local descriptor based regional features (F1, F2, F5, F6) perform better than intensity based regional features (F3, F4) on LivDet-Clarkson, LivDet-Warsaw and MobBIOfake datasets. On LivDet-ND dataset, intensity histogram (F3) based regional features have the best performance, but local descriptor based regional features still achieve low error rates. This observation suggests that local descriptors generally include more distinctive information than intensity values. Therefore, future feature design may rely more on local descriptors. It is consistent with the results in recent researches [62, 109].

Second, it is very interesting that, on LivDet-ND dataset, the order of performance

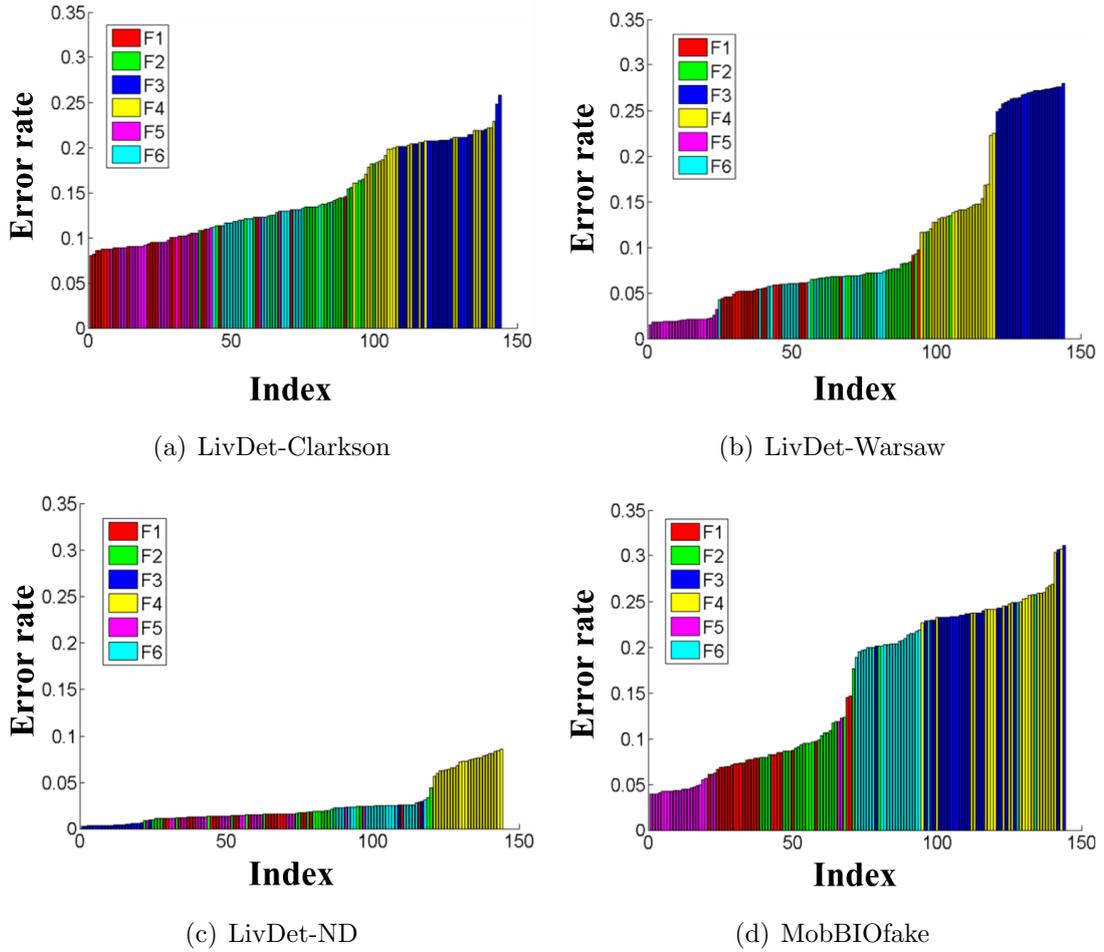


Figure 5.6: Performance of single regional features using relational measure model.

is significantly different from the other three datasets. It can be seen that intensity histogram (F3) based regional features achieve the best performance on LivDet-ND, while the performance of this feature is low on the other datasets. A possible reason is that intensity histograms are able to capture more valuable information at the specific resolution of LivDet-ND dataset given the sensor setting of this dataset, while on the other three datasets, it is possible that this feature requires a higher resolution to obtain more usable information under the sensor setting of these datasets. To examine the validity of this explanation, an additional experiment is performed on LivDet-ND dataset with all the iris regions down-sampled by a factor of 0.75. The error rates of all the individual features on this down-sampled LivDet-ND dataset are shown in Fig. 5.7. It can be seen that the error rates of intensity histogram based

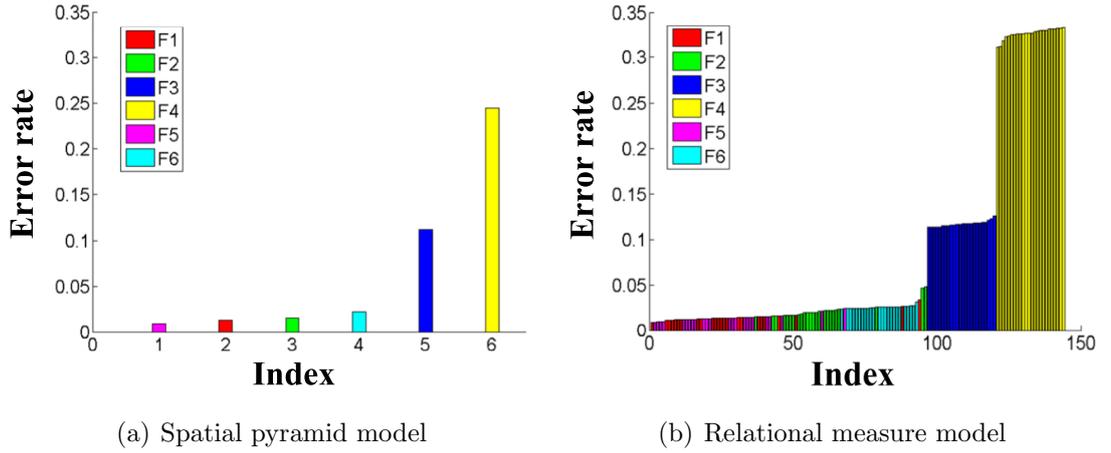


Figure 5.7: Performance of single regional features on LivDet-ND dataset down-sampled by a factor of 0.75.

regional features increase significantly compared to that on the original LivDet-ND dataset, and the order of the error rates becomes similar with the other three datasets. Therefore, it can be concluded that intensity histogram based regional features may be more effective if the iris region has a higher resolution.

Third, comparing the error rates among all the datasets, it can be found that LivDet-ND dataset has the lowest error rate. The reason of this is the number of training images. Compared to the other three datasets, LivDet-ND dataset has the highest number of training images. Due to the high dimensionality of the proposed regional features, a sufficient number of training images will lead to better trained classifiers and hence better performance.

Finally, it can be seen that the error rates on LivDet-Clarkson dataset are higher than the others. A possible reason is that the images in LivDet-Clarkson dataset are more noisy. The images in this dataset have varying focus levels, and it introduces more variation into extracted features. For LivDet-Clarkson dataset, more training samples may be needed to train suitable classifiers to improve the performance.

5.3.3 Comparison between regional features and low-level features

The proposed regional features are constructed based on low-level features. Additional to the information in low-level features, regional features seek the high-level feature distribution information. Therefore, it is necessary to investigate the effect of such high-level feature distribution information in iris liveness detection. In this subsection, this investigation is made by comparing between the performance of regional features and low-level features.

Specifically, given a low-level feature, comparisons are made among the following 4 features: (1) the original low-level feature (referred to as Original feature) extracted from the whole iris region without using the method described in Section 5.2; (2) spatial pyramid-based regional feature (referred to as Spatial pyramid) constructed using this low-level feature, using the spatial pyramid model described in Section 5.2.2; (3) relational measure-based regional feature with top 1 lowest classification error rate in 10-fold cross-validation on training data (referred to as Relational measure top 1); the proposed relational measure model has 24 designed kernels as described in Section 5.2.3; it leads to 24 relational measure-based regional features given a particular low-level feature; cross-validation is performed on the training data, and the feature with the lowest error rate among the 24 regional features is selected and its classification error rate on the test data is used for comparison; it represents the performance of the best relational measure-based regional feature that can be selected using training data only; (4) relational measure-based regional feature with minimum error rate on the test set (referred to as Relational measure min); due to the difference between training and test data, the feature selected in (3) may not be optimal on the test set; therefore, Relational measure min is used as a reference for the performance of the best relational measure-based regional feature.

The error rates of the above described features on all the datasets are reported in

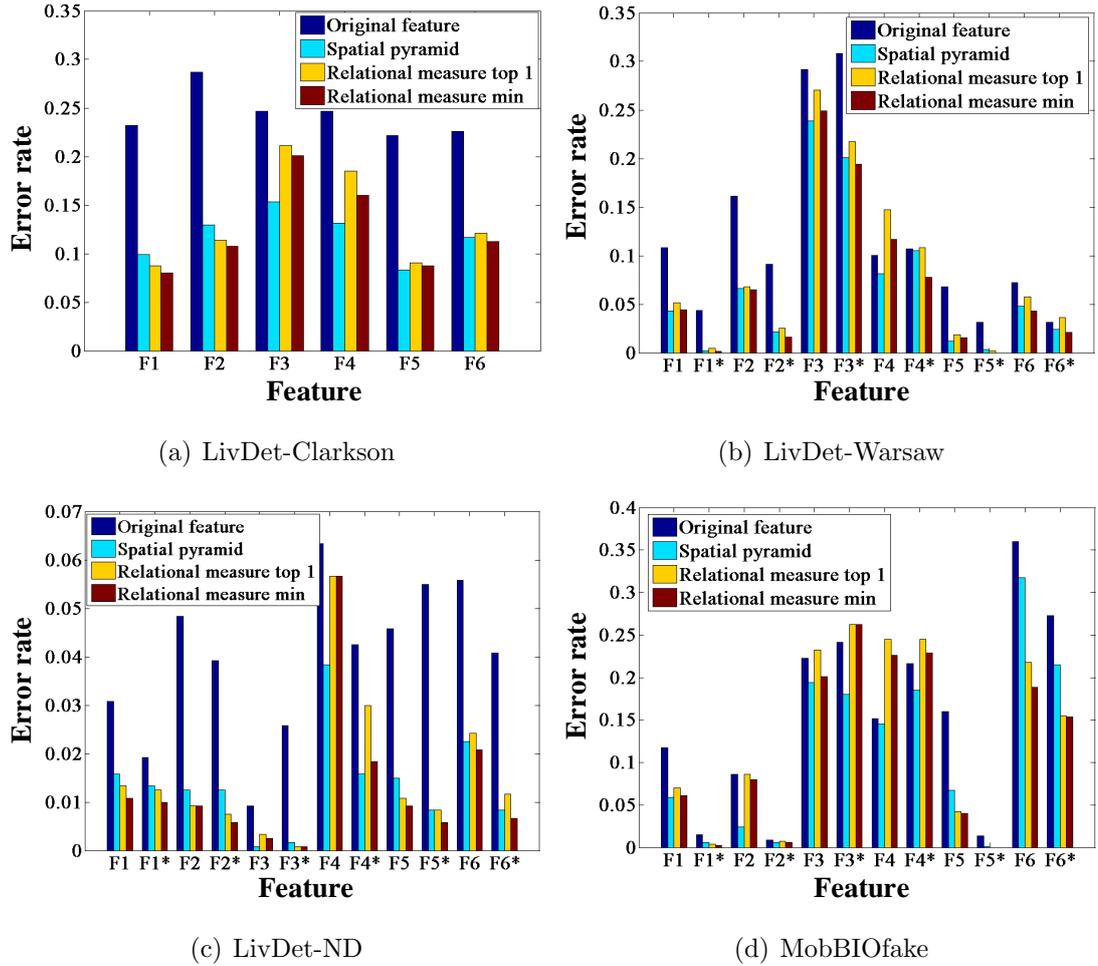


Figure 5.8: Comparison of the performance between low-level features and regional features constructed based on each low-level feature on all the datasets.

Fig. 5.8. In each subfigure, the six low-level features are referred to as F1 to F6, and the error rates are grouped based on the low-level feature type. For the three NIR datasets, the training and test partitions in LivDet-Iris 2013 competition are used. For the colour dataset, the training and test partitions in MobILive 2014 competition are used.

It can be seen from Fig. 5.8 that the spatial pyramid-based regional features have better performance compared to the original low-level features on all the datasets. Also, relational measure-based regional features generally lead to an improved performance in comparison to the original low-level features. The exceptions are F4 (intensity correlogram) on LivDet-Warsaw and MobBIOfake datasets. Possible rea-

son are: (1) some images in LivDet-Warsaw dataset have a low contrast in iris region and (2) the iris region in MobBIOfake dataset has a low resolution. Because of these factors, the intensity correlograms in local regions include little information: for the region with low contrast, the corresponding intensity correlogram will concentrate on only few bins; for the region with a low resolution, there are too few pixels with a fixed distance. Little information in low-level features further leads to degraded information of high-level feature distribution in the constructed regional features based on relation measure model, thus the performance of the corresponding regional features suffers.

Also, it is found that on all the datasets, Relational measure top 1 generally has a higher error rate than Relational measure min. It means that the regional feature selected by cross-validation on the training data are often not the optimal choice for the test data. The possible reason is insufficient training data. Since regional features usually have a high dimension, more training data are required for a more accurate feature selection.

Finally, the performance of some original low-level features (like F1, F5 and F6) in this experiment is lower than that reported in [62,109]. Possible reasons are analysed as follows.

First, the method proposed in this chapter does not concentrate on optimizing the parameters. For example, SVM has a lot of types, and the proposed method uses the one in [122] due to its reported effectiveness for image classification, rather than test all the SVMs and select the best one; LBP has a lot of variations as studied in [58,126], but the proposed method simply uses the most general 8-neighbour form. The method in this chapter does not target optimal parameter setting, because it focuses on investigating the effect of regional features exploiting the high-level feature distribution information of low-level features for iris liveness detection, rather than finding the best performance of low-level features which has been well studied in [62,109], or seeking the best classifier.

Second, the method proposed in this chapter does not preprocess the image. According to [109], preprocessing is able to improve the performance on MobBIOfake dataset. In the experiment of this chapter, the preprocessing method described in [109] is tested. It is found that this preprocessing is effective for MobBIOfake dataset but not for the other three NIR datasets. Although some other preprocessing methods are possible to be able to improve the performance on NIR datasets, the proposed method directly operate on the original images considering the generality of the method.

Third, iris segmentation error will lead to a higher error rate. It can be seen in Fig. 5.1 that, in this chapter, feature extraction is based on the output of iris segmentation stage. A less accurate iris segmentation results that the features are extracted from inaccurate regions. It induces outliers in the feature space, hence leading to a drop of accuracy. Although the method in [62] performs no segmentation and the result is promising, it is found that, on LivDet-Clarkson dataset, iris segmentation contributes significantly to the performance of regional features. A possible reason is that [62] assumes that the background is generally static and it has little influence on the relevant statistics in extracted features, but this assumption does not hold for LivDet-Clarkson dataset, due to the varying focus of the images in this dataset.

To validate the influence of the preprocessing and segmentation error, experiments are conducted on LivDet-Warsaw, LivDet-ND and MobBIOfake datasets using some meta-information included in these datasets. For LivDet-Warsaw and LivDet-ND, to eliminate iris segmentation error, features are extracted from the iris region indicated in the meta-information provided in the two datasets, instead of using the iris segmentation step in Fig. 5.1. For MobBIOfake, since the fake images are produced by printing which influences the whole image, features are extracted from the whole iris image to eliminate segmentation error. Also, the preprocessing in [109] is adopted for MobBIOfake, since this preprocessing is reported to lead to an im-

proved performance on this dataset. The results of these experiments are reported in Fig. 5.8, and the features extracted with meta-information are referred to as F1* to F6*. Similar experiment is not conducted on LivDet-Clarkson dataset, since this dataset does not include relevant meta-information.

In Fig. 5.8, by comparing the error rates between the methods with and without using meta-information, it can be seen that using meta-information generally leads to an improved performance for most regional features. It means that the above mentioned factors are able to influence the performance. Also, regional features still have a generally better performance in comparison to the original low-level features.

In summary, it is shown that, (1) when all other factors like iris segmentation, preprocessing and classifier type fixed, regional features lead to a generally better performance in comparison to the corresponding low-level features; (2) by using meta-information, the performance can be significantly improved, and the regional features still lead to an generally improved performance in comparison to the original low-level features.

5.3.4 Effect of score-level fusion

This subsection investigates the effect of score-level fusion. Due to the high dimensionality of proposed regional features, different partitions of training and test samples are used for this analysis. More samples are assigned for training to better train the classifiers to produce more reliable classification responses. Specifically, the following partitions are used: 1150 training and 206 test for LivDet-Clarkson; 1172 training and 495 test for LivDet-Warsaw; 3000 training and 1200 test for LivDet-ND; 1000 training and 600 test for MobBIOfake. Each dataset is randomly partitioned for 100 times using the above number of training and test samples. For each specific partition, the top 3 features are found by 10-fold cross validation on the training samples. Then, the classifier responses of the top 3 features on the test samples are

Table 5.2: Results of score-level fusion (error rate)

Dataset	top 1	top 2	top 3	fusion
LivDet-Clarkson	2.71%	2.62%	2.85%	2.43%
LivDet-Warsaw	1.08%	1.04%	1.07%	1.05%
LivDet-ND	0.41%	0.81%	0.80%	0.41%
MobBIOfake	4.48%	4.90%	4.78%	2.40%

fused using Eqn. 5.5. The first 50 random partitions are used to learn the weights in Eqn. 5.5, and the last 50 random partitions are used to test the performance of score-level fusion. Tab. 5.2 reports the mean classification error rates of the top 3 features and score-level fusion on the last 50 random partitions of each dataset.

It can be seen from Tab. 5.2 that the effect of score-level fusion varies on different datasets. Compared to the top 1 feature found by cross-validation, the score-level fusion achieves the largest improvement on MobBIOfake dataset especially. This observation demonstrates that score-level fusion is effective for colour printouts. Also, an improvement is achieved on LivDet-Clarkson dataset. On the other hand, for LivDet-Warsaw dataset, the classification error rate is similar between top 1 feature and the fusion result, and no improvement is observed for LivDet-ND dataset. Based on the above observations, it can be concluded that score-level fusion is effective as a general method to improve the performance, since it leads to a performance either better than or near to the top 1 selected feature on all the datasets.

5.3.5 Comparison with existing methods

The analysis in previous sections are based on the experimental setting in this chapter, *i.e.* the features are extracted from the segmented iris region and the classifier in Section 5.2.5 is used for classification. In this subsection, the proposed iris liveness detection method based on regional features is compared to existing methods with their own setting on feature extraction and classifier.

This comparison adopts the partition of training and test samples in LivDet-Iris

2013 [105] and MobILive2014 [110] competitions. For the proposed method, the top 3 features and the corresponding weights for score-level fusion are obtained by cross-validation on the training samples.

The first comparison is made with the results reported in the above competitions. It includes the performance of all the methods (ATVS, Federico, Porto) in LivDet-Iris 2013 competition on LivDet-Clarkson, LivDet-Warsaw and LivDet-ND datasets, and the performance of the winning method (IIT-Indore) in MobILive2014 competition on MobBIOfake dataset.

Note that the comparison to the competition results is not fully fair, since, in these competitions, the test data is inaccessible. However, it is reasonable to make the comparison, because, in this comparison, the experimental configuration is very close to the competitions. The algorithms take an input image and produce a decision with pre-learned parameters using the same training data to the competitions.

The proposed method is also compared with state-of-the-art local descriptor (LD) in [62]. Two local descriptors are selected, LPQ and SID. The reason of choosing LPQ and SID is that, as reported in [62], they are top features with independent quantisation and joint quantisation, respectively. On each dataset, the lowest error rate between LPQ and SID is reported to represent the best performance of local descriptors. For MobBIOfake dataset, the preprocessing in [62] is applied before feature extraction.

Finally, two error rates are reported for the proposed regional features. The first error rate (referred to as RF) is the result of the proposed method based on regional features as shown in Fig. 5.1. RF does not use the preprocessing, and it extracts features within the segmented iris region. Therefore, this error rate represents the error of the whole system in Fig. 5.1 for iris liveness detection. The second error rate (referred to as RF*) is the lowest error rate achieved by regional features. It reflects the error of the proposed regional features only, with the other errors

Table 5.3: Results of comparisons with other methods (error rate)

Method	LivDet-Clarkson	LivDet-Warsaw	LivDet-ND	MobBIOfake
ATVS	43.74%	17.07%	—	—
Federico	24.49%	11.00%	19.08%	—
Porto	15.30%	8.55%	—	—
IIT Indore	—	—	—	0.25%
LD	10.79%	0.00%	0.00%	0.00%
RF	7.87%	6.15%	0.08%	1.50%
RF*	8.02%	0.00%	0.08%	0.00%

induced by other components like iris segmentation and feature selection minimized. Specifically, to eliminate iris segmentation error, for LivDet-Warsaw and LivDet-ND datasets, regional features are extracted within the iris region indicated in the meta-information; for MobBIOfake dataset, regional features are extracted from the whole image (it is reasonable since the spoofing type of printout will influence the whole image). In terms of preprocessing, the method in [109] is applied to MobBIOfake. To minimize the error of feature selection, among the error rates achieved by all the regional features, the lowest one is used as the second error rate. Note that, for LivDet-Clarkson dataset, since it does not have relevant meta-information, the second error rate of this dataset is simply the minimum error rate of regional features extracted from the iris region segmented by the method in 5.2.1.

Tab. 5.3 shows the classification error rates of all the methods on all the datasets. Investigating the performance of RF*, it can be seen that RF* performs better than the methods in LivDet-Iris 2013 (ATVS, Federico, Porto) and MobILive2014 (IIT Indore) competitions. Also, the performance of RF* is better than LD on LivDet-Clarkson dataset, and it is comparable to LD on the other datasets. This result shows that, on three of the four datasets used in the experiment (LivDet-Warsaw, LivDet-ND, MobBIOfake), the proposed regional feature based iris liveness detection method is able to achieve a comparable performance to state-of-the-art methods, if the error induced by iris localisation, preprocessing and feature selection is minimized; on a more noisy dataset (LivDet-Clarkson), the proposed method is

able to achieve a better performance compared to state-of-the-art methods, even with some error induced by the above three components (note that LivDet-Clarkson is a more noisy dataset where the images have varying focus, and this dataset does not include meta-information on iris position).

However, it is observed that RF has generally higher error rates than RF*. This is because of the combination of three factors: iris segmentation error, preprocessing, and feature selection error. As shown in Section 5.3.3, iris segmentation error and preprocessing influence the performance of regional features. Furthermore, in the proposed method, the top 3 features selected for fusion based on the training data may not include the best regional feature for the test data. Therefore, although the fusion leads to a lower error rate than each selected feature, the performance may be still lower than the best feature on the test data.

Finally, it is found that, compared to LD which achieves the best performance among the comparison methods other than the proposed one, the proposed regional feature based method achieves the largest improvement on LivDet-Clarkson dataset, although this dataset does not include the meta-information about the iris position. A possible reason is that LivDet-Clarkson dataset includes images with multiple focus levels and hence it is more noisy; in this dataset, the proposed regional features are able to capture more distinguishing information than the other methods, even with segmentation error in a few images, thus achieving better performance. In other words, the experimental results show that the proposed method is preferable for the dataset with heavier noise.

In summary, the following conclusions can be drawn based on all the experimental observations in Section 5.3. First, local descriptor based regional features perform generally better than intensity based regional features. Second, regional features are able to achieve a generally better performance in comparison to low-level features. Third, the score-level fusion of regional features achieves a generally improved performance compared to the best regional feature sought by the feature selection.

Four, in comparison to state-of-the-art methods, regional feature based method is able to achieve a comparable performance on three of the four datasets used in the experiment (LivDet-Warsaw, LivDet-ND and MobBIOfake), with accurate iris localisation, proper pre-processing and reliable feature selection; in a dataset with heavier noise (LivDet-Clarkson), regional feature based method is able to achieve better performance than state-of-the-art methods, even with some error induced by the above three components.

5.4 Summary

In this chapter, regional features are proposed for iris liveness detection. Regional features exploit the information not only from low-level features, but also from high-level feature distribution. Two models are exploited to express the high-level feature distributions: a spatial pyramid model capturing the feature distribution in regions with varying size, and a relational measure model calculating the feature distribution in regions with varying shape. The constructed regional features based on the two models are fused at the score-level to make the final decision. Experiments are conducted on four benchmark datasets for iris liveness detection. The results show that regional features are able to achieve a generally improved performance compared to low-level features. Also, it is demonstrated that, in three of the four datasets used in the experiment, regional feature based iris liveness detection method is able to achieve a comparable performance to state-of-the-art methods with accurate iris localisation, proper preprocessing and reliable feature selection; in a dataset with heavier noise, regional feature based method is able to achieve better performance than state-of-the-art methods, even with some error induced by the above three components. Future work may consider to develop more accurate iris localisation and feature selection methods, to enable regional features to achieve the best performance.

Also, another limitation of the proposed method is that the trained classifier remains a “black box”: it is difficult to study what feature the classifier has learned. This is a common problem in research for image classification, and it can be addressed in future research effort.

Chapter 6

Signal-level information fusion using sparse-error low rank matrix factorisation

Part of this chapter has been adapted from the journal paper [J8] (published) as in the List of Publications.

6.1 Introduction

The main challenge for less constrained iris recognition is the noise and degradations in iris captures. In less constrained environment with the subject at-a-distance and on-the-move, the captures usually suffer from a low resolution, together with some noise like motion blur, specular reflection, off-angle, *etc* (in the rest of this chapter, the term ‘noise’ is used to refer to all these noise and degradations, following most literature on less constrained iris recognition, like [18, 19, 21–23, 51, 98]). These noise significantly deteriorate iris recognition performance [32]. Therefore, a critical problem of less constrained iris recognition is how to overcome the influence of noise in the iris captures.

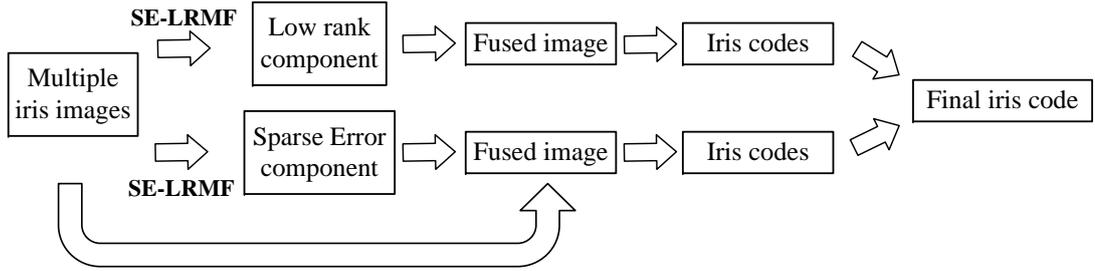


Figure 6.1: Flowchart of the proposed method for iris liveness detection.

In this chapter, a novel signal-level information fusion method is proposed for less constrained iris recognition to mitigate the noise in less constrained iris captures. This method is based on low rank approximation (LRA). Given multiple noisy captures of the same eye, the method assumes that: (1) the potential noiseless images lie in a low rank subspace; (2) the noise is spatially sparse. Based on these assumptions, a LRA of noisy captures is sought to separate the noiseless images and noise for information fusion.

This chapter proposes a novel sparse-error low rank matrix factorisation (SE-LRMF) model to perform LRA. The flowchart of the proposed method based on SE-LRMF is shown in Fig. 6.1. Firstly, given multiple iris captures of the same eye, a low rank component and a sparse error component are computed using SE-LRMF. The low rank component estimates the clear iris pixels, and the sparse error component is viewed as noise pixels. Then, the low rank component (clear iris pixels) and sparse error component (noise pixels) are considered individually for signal-level fusion. The low rank component is fused straightforwardly, while the sparse error component is utilised to perform a weighted fusion of the original iris captures. Such a scheme leads to two fused images: one using the low rank component, the other one using the sparse error component and original iris captures. Finally, two iris codes are extracted from the two fused images, and the two iris codes are combined to produce a single iris code as the final fusion result.

Experiments are conducted on three benchmark datasets. The results demonstrate that the proposed signal-level fusion method is able to achieve a generally improved

iris recognition performance in less constrained **environments**, in comparison to existing iris recognition algorithms, especially for the iris captures with heavy noise and low quality.

The remainder of this chapter is organised as follows. Section 6.2 revisits the related works to information fusion in iris recognition, and low rank approximation. Section 6.3 presents the proposed signal-level fusion method. Section 6.4 reports the results of experimental analysis and comparisons. Section 6.5 summarises this chapter.

6.2 Related work

6.2.1 Information fusion in iris recognition

Existing information fusion methods for iris recognition can be categorized into score-level fusion, signal-level fusion and feature-level fusion. Given a set of gallery iris codes and a set of probe iris codes, score-level fusion produces one matching score by fusing the matching scores between all the combinations of gallery and probe iris codes. [26] reports the performance of minimum fusion and mean fusion of matching scores on NIR iris videos.

Signal-level fusion combines multiple iris captures to be one image, and iris codes are extracted from the fused image for matching. In [26], Hollingsworth *et al.* fuse iris images in video captures by mean and median fusion. As reported in [26], signal-level fusion leads to better performance than no fusion, and mean fusion performs better than median fusion. The mean fusion method in [26] essentially assigns equal importance to all the pixels in all the iris captures. Different from this, weighted mean fusion methods are proposed in [27,28]. The weights are calculated based on image quality measures. The iris captures with higher quality are assigned higher

weights. In [27], Nguyen *et al.* use focus score as quality measure. In [28], a better performance is achieved by combining multiple quality scores to produce a quality measure. The considered quality scores include focus, off-angle, illumination variance and motion. Compared to score-level fusion, signal-level fusion has significant advantage in speed and memory cost. However, existing signal-level fusion methods directly use the original iris captures for fusion. In less constrained environment, the fusion results are likely to be influenced by the noise in original captures, although the noise level can be mitigated by fusion.

Feature-level fusion aims to estimate the iris code of a clear and high-resolution iris image using the iris codes extracted from multiple noisy and low-resolution iris captures. The key step of feature-level fusion is to learn the relationship between the iris codes of noiseless high-resolution iris images and noisy low-resolution iris images. Nguyen *et al.* [29] formulate feature-level fusion in a Bayesian framework using eigen-iris extracted by principal component analysis (PCA) as features, and it can be solved by maximum a posteriori estimation approach. In [30,128], Nguyen *et al.* demonstrate that the similar Bayesian framework can be adapted to Gabor feature domain which is the most widely used feature in iris recognition. Liu *et al.* [31] propose to use a Markov network model for feature-level fusion. This model considers the high resolution iris code as the hidden node of a Markov network, and it is computed using multiple low resolution iris codes served as the observations. Feature-level fusion performs generally better than score-level and signal-level fusion, since the features (iris codes) are more directly related to the performance of iris recognition. However, most feature-level fusion methods require a learning stage to build the relationship between high resolution and low resolution features, and such relationship can be data dependent. It influences the generality of the algorithm in real applications. Also, similarly to the existing signal-level fusion methods, most of current feature-level fusion methods use the iris codes extracted from the original noisy iris captures for fusion. As a result, the fusion result is still influenced by the

noise embedded in the feature space.

Note that the algorithms in [27, 28, 30, 128] are called super-resolution in these papers. However, differently from other iris super-resolution methods like [129, 130], the key step in [27, 28, 30, 128] is the fusion of information in multiple iris images, rather than the estimation of high-resolution texture from low-resolution observations. Therefore, this chapter considers them as information fusion methods.

6.2.2 Low rank approximation (LRA)

LRA is a powerful method to seek the potential noiseless structures of highly-correlated data from noisy observations. Given multiple noisy observations, a basic assumption of LRA is that the potential noiseless data lie in a low rank subspace. Based on this assumption, the LRA of noisy observations is sought as the estimation of noiseless data. Existing LRA algorithms can be categorized into low rank decomposition [131–136] and low rank matrix factorisation [87, 137–143]. Low rank decomposition represents the original noisy observations as the sum of a low rank component and an error component. The low rank component estimates the potential noiseless data, while the error component models the noise which is usually assumed to be spatially sparse. The representative algorithm of low rank decomposition is robust principal component analysis (RPCA) [136]. Let $\mathbf{Y} \in \mathbb{R}^{m \times n}$ be a data matrix with n observations (each column is an observation). RPCA seeks the low rank approximation of \mathbf{Y} as follows:

$$\arg \min_{\mathbf{A}, \mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \quad s.t. \quad \mathbf{Y} = \mathbf{A} + \mathbf{E} \quad (6.1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the low rank component and $\mathbf{E} \in \mathbb{R}^{m \times n}$ is the error component. $\|\bullet\|_*$ is the nuclear norm which calculates the sum of the singular values of \bullet . Minimizing $\|\mathbf{A}\|_*$ leads to a minimisation on the rank of \mathbf{A} . $\|\bullet\|_1$ calculates the sum of the absolute value of all the elements in \bullet . Minimizing $\|\mathbf{E}\|_1$ induces sparsity in \mathbf{E} . The

main advantage of RPCA is its convexity. The convexity enables to solve Eqn. 6.1 by tractable convex optimisation techniques like interior point method [134], proximal gradient method [131, 133], augmented Lagrange multiplier [132, 136] and random projection [135]. However, most of the algorithms to solve Eqn. 6.1 require to compute a singular value decomposition (SVD) in each iteration. It is computationally expensive.

Differently from low rank decomposition minimizing the nuclear norm to seek a low rank solution, low rank matrix factorisation approximates the observation matrix \mathbf{Y} by the product of two low rank matrixes. It is formulated as follows:

$$\arg \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{W} \otimes (\mathbf{Y} - \mathbf{UV}^T)\| \quad (6.2)$$

where $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ are two low rank matrix ($r \ll m, n$), \otimes is element-wise multiplication operator, and $\mathbf{W} \in \{0, 1\}^{m \times n}$ is a binary weight matrix indicating missing entries in \mathbf{Y} . An element in \mathbf{W} is assigned 0 if the corresponding entry in \mathbf{Y} is missing, otherwise it is assigned 1. Since both \mathbf{U} and \mathbf{V}^T are low rank matrix, their product is still a low rank matrix and it forms the LRA of \mathbf{Y} .

Although the solution of Eqn. 6.2 is not unique and it is also dependent on the structure of \mathbf{W} , many research have shown that effective solutions can be computed with the norm in Eqn. 6.2 being Frobenious norm [138–141], ℓ_1 -norm [137, 142], rotational invariant R_1 -norm [143], or maximizing the ℓ_1 dispersion of \mathbf{Y} (*i.e.* seeking \mathbf{U} that maximizes $\|\mathbf{U}^T \mathbf{Y}\|_1$) [87]. Especially, in recent research, Meng *et al.* [137] demonstrate that with ℓ_1 -norm, Eqn. 6.2 can be efficiently solved by a cyclic weighted median filtering algorithm. As reported in [137], this method achieves a more robust reconstruction of noiseless signals in comparison to RPCA and some other representative algorithms of low rank matrix factorisation.

However, compared to low rank decomposition, low rank matrix factorisation does not explicitly formulate the noise component (*i.e.* \mathbf{E} in Eqn. 6.1). The noise ele-

ments can be indicated in \mathbf{W} , but such weighting scheme heavily relies on the prior of noise positions which are sometimes unavailable in real applications. As a result, the noise in the observations always influences the estimation of \mathbf{U} and \mathbf{V} in the procedure of solving Eqn. 6.2, although such influence can be mitigated by using some more robust norms, like the above mentioned ℓ_1 -norm, rotational invariant R_1 -norm, and ℓ_1 dispersion of \mathbf{Y} .

6.3 Methodology

This section presents the proposed signal-level fusion method for less constrained iris recognition. Firstly, the proposed SE-LRMF model to seek a low rank approximation is described. Given multiple iris captures, SE-LRMF decomposes them into a low rank component expressing the potential noiseless images and an error component estimating the noise. Then, the signal-level fusion method based on the result of SE-LRMF is detailed. It consists of a low rank mean (LRM) fusion method using the low rank component and an error weighted mean (EWM) fusion method employing the error component. Finally, the code level combination method of the results of LRM and EWM is presented.

6.3.1 Sparse-error low rank matrix factorisation (SE-LRMF)

Given multiple noisy captures of the same eye, the proposed SE-LRMF assumes that the potential noiseless images lie in a low rank subspace, due to the inherent stability of iris patterns. Similar to the notations in Section 6.2.2, $\mathbf{Y} \in \mathbb{R}^{m \times n}$ is used to denote a data matrix with each of its column being an unwrapped iris image. $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ are used to denote the two matrix forming the factorisation of \mathbf{Y} , and $\mathbf{E} \in \mathbb{R}^{m \times n}$ is used to denote the error component which

models the noise in the iris images. The SE-LRMF model is formulated as follows:

$$\arg \min_{\mathbf{U}, \mathbf{V}, \mathbf{E}} \frac{1}{2} \|\mathbf{Y} - \mathbf{UV}^T - \mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_1 \quad (6.3)$$

where λ is a penalty parameter. Eqn. 6.3 includes two terms. The first term is a reconstruction fidelity term. It represents the noisy iris captures by the sum of the noiseless images (low rank component \mathbf{UV}^T) and noise (error component \mathbf{E}). The second term is a sparsity regularisation on the error component. Following RPCA [136], SE-LRMF assumes that the noise is spatially sparse, and ℓ_1 -norm is used to induce sparsity. The model has two parameters: $r \ll m, n$ is one of the dimensions of \mathbf{U} and \mathbf{V} , controlling the rank of approximation; λ is a scalar balancing between the reconstruction error and the sparsity.

Differently from traditional low rank matrix factorisation models in Eqn. 6.2, SE-LRMF explicitly formulates the noise to suppress its influence. Also, compared to Eqn. 6.2, Eqn. 6.3 does not include \mathbf{W} which is the weight matrix indicating missing entries. The reason is three-fold. First, although \mathbf{W} can be constructed based on the results of iris segmentation, the segmentation errors will lead to inaccurate entries. Even with an accurate localisation of iris, pupil and eyelids, false detections on reflection and silhouette still induce inaccuracy. Second, including \mathbf{W} increases the complexity of the objective function, and it is possible to induce more computational load. Third, experimentally, it is found that the error component \mathbf{E} and the sparsity regularisation on it are powerful enough to model the noise and mitigate its influence on the estimation of \mathbf{U} and \mathbf{V} .

Eqn. 6.3 is not convex in \mathbf{U} , \mathbf{V} and \mathbf{E} , but it is convex in one of them with the other two fixed. Therefore, a local minimum can be sought via an alternating scheme. It leads to a \mathbf{U} , \mathbf{V} -subproblem and an \mathbf{E} -subproblem.

\mathbf{U} , \mathbf{V} -subproblem. Let $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{E}$. With error component \mathbf{E} fixed, minimizing

Eqn. 6.3 leads to the following problem:

$$\arg \min_{\mathbf{U}, \mathbf{V}} \left\| \tilde{\mathbf{Y}} - \mathbf{U}\mathbf{V}^T \right\|_F^2 \quad (6.4)$$

The solution of Eqn. 6.4 is not unique. By fixing one of \mathbf{U} and \mathbf{V} and setting the partial derivative of the objective function with respect to the other one to zero, the objective function can be descended by updating \mathbf{U} and \mathbf{V} as follows:

$$\mathbf{V} = \tilde{\mathbf{Y}}^T \mathbf{U} (\mathbf{U}^T \mathbf{U})^{-1} \quad (6.5)$$

$$\mathbf{U} = \tilde{\mathbf{Y}} \mathbf{V} (\mathbf{V}^T \mathbf{V})^{-1} \quad (6.6)$$

E-subproblem. Let $\hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{U}\mathbf{V}^T$. With \mathbf{U} and \mathbf{V} fixed, minimizing Eqn. 6.3 leads to the following problem:

$$\arg \min_{\mathbf{E}} \frac{1}{2} \left\| \hat{\mathbf{Y}} - \mathbf{E} \right\|_F^2 + \lambda \|\mathbf{E}\|_1 \quad (6.7)$$

According to [144], the solution of Eqn. 6.7 can be obtained by performing soft-thresholding operation on $\hat{\mathbf{Y}}$:

$$\mathbf{E} = \text{sign}(\hat{\mathbf{Y}}) \max\left(\left|\hat{\mathbf{Y}}\right| - \lambda, 0\right) \quad (6.8)$$

where $\text{sign}(\bullet)$ is the sign function. All the operations in Eqn. 6.8 are element-wise.

It can be seen that, in such an alternating scheme, \mathbf{U} and \mathbf{V} forming the low rank component are computed based on noise-eliminated signals $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{E}$. In other words, with the explicitly formulated error component \mathbf{E} , the proposed SE-LRMF is able to suppress the influence of noise on low rank matrix factorisation without relying on any specific robust norms or weight matrix based on the priors of noise.

In the implementation, the order of alternation is Eqn. 6.5, Eqn. 6.6, Eqn. 6.8.

Algorithm 1: Solving SE-LRMF

Input : matrix of unwrapped iris images \mathbf{Y} ; parameter λ

- 1 Initialisation: random \mathbf{U}_0 , $\mathbf{E}_0 = \mathbf{0}$, $i = 0$;
- 2 **while** *not converge* **do**
- 3 compute \mathbf{V}^{i+1} using \mathbf{Y} , \mathbf{E}^i , \mathbf{U}^i based on Eqn. 6.5;
- 4 compute \mathbf{U}^{i+1} using \mathbf{Y} , \mathbf{E}^i , \mathbf{V}^{i+1} based on Eqn. 6.6;
- 5 compute \mathbf{E}^{i+1} using \mathbf{Y} , \mathbf{U}^{i+1} , \mathbf{V}^{i+1} based on Eqn. 6.8;
- 6 $i \leftarrow i + 1$;
- 7 **end**

Output: $\mathbf{U}, \mathbf{V}, \mathbf{E}$

Accordingly, \mathbf{U} and \mathbf{E} need to be initialised in the first iteration. Experimentally, it is found that although Eqn. 6.3 has multiple local minimum and the solution depends on the starting point, a simple random initialisation of \mathbf{U} as used in [137] is able to achieve a generally good performance. \mathbf{E} is initialised to be a matrix of zeros. The whole algorithm to solve Eqn. 6.3 is summarised in Algorithm 1.

Convergence analysis. The algorithm in Algorithm 1 is guaranteed to converge to a local minimum. This is demonstrated as follows. Let $f(\mathbf{U}, \mathbf{V}, \mathbf{E})$ be the objective function in Eqn. 6.3. Let \mathbf{V}^i be the \mathbf{V} obtained in the i^{th} iteration. Let \mathbf{U}^i be the \mathbf{U} obtained in the i^{th} iteration. Let \mathbf{E}^i be the \mathbf{E} obtained in the i^{th} iteration. It can be obtained that:

(1) alternating between \mathbf{U} , \mathbf{V} -subproblem and \mathbf{E} -subproblem leads to a sequence of monotonically decreasing objective function values. As a proof, $f(\mathbf{U}^i, \mathbf{V}^i, \mathbf{E}^i) \leq f(\mathbf{U}^i, \mathbf{V}^i, \mathbf{E}^{i-1}) \leq f(\mathbf{U}^{i-1}, \mathbf{V}^i, \mathbf{E}^{i-1}) \leq f(\mathbf{U}^{i-1}, \mathbf{V}^{i-1}, \mathbf{E}^{i-1})$ can be obtained as follows:

$$f(\mathbf{U}^{i-1}, \mathbf{V}^i, \mathbf{E}^{i-1}) = \arg \min_{\mathbf{V}} f(\mathbf{U}^{i-1}, \mathbf{V}, \mathbf{E}^{i-1}) \leq f(\mathbf{U}^{i-1}, \mathbf{V}^{i-1}, \mathbf{E}^{i-1}) \quad (6.9)$$

$$f(\mathbf{U}^i, \mathbf{V}^i, \mathbf{E}^{i-1}) = \arg \min_{\mathbf{U}} f(\mathbf{U}, \mathbf{V}^i, \mathbf{E}^{i-1}) \leq f(\mathbf{U}^{i-1}, \mathbf{V}^i, \mathbf{E}^{i-1}) \quad (6.10)$$

$$f(\mathbf{U}^i, \mathbf{V}^i, \mathbf{E}^i) = \arg \min_{\mathbf{E}} f(\mathbf{U}^i, \mathbf{V}^i, \mathbf{E}) \leq f(\mathbf{U}^i, \mathbf{V}^i, \mathbf{E}^{i-1}) \quad (6.11)$$

(2) $f(\mathbf{U}, \mathbf{V}, \mathbf{E})$ is lower bounded ($f(\mathbf{U}, \mathbf{V}, \mathbf{E}) \geq 0$).

Therefore, according to above (1) and (2), the convergence of Algorithm 1 is guaranteed. Furthermore, since $f(\mathbf{U}, \mathbf{V}, \mathbf{E})$ is not convex in \mathbf{U} , \mathbf{V} and \mathbf{E} , it can be obtained that the Algorithm 1 converges to a local minimum.

After obtaining \mathbf{U} , \mathbf{V} and \mathbf{E} , a matrix $\mathbf{A} = \mathbf{UV}^T$ can be constructed as the low rank component considered as noiseless iris images, and the error component \mathbf{E} is directly used as the estimated noise. Some example results of SE-LRMF are shown in Fig. 6.2(a) to Fig. 6.2(c). It can be seen that given the noisy iris captures in Fig. 6.2(a), the low rank components in Fig. 6.2(b) mainly include the potential stable structure of the iris, while the error component in Fig. 6.2(c) accurately models noise like reflections. Note that part of eyelids and silhouette are still contained in the low rank component. This is because the pixels in some regions are always occupied by eyelids and silhouette, and there lacks clear iris information for these regions. Also, it can be found that some clear iris texture is possible to be detected as noise. An example is the second image from the top. It can be seen that, in this image, some clear iris textures in the middle part at the horizontal direction are included in the error component. The reason is that these local textures in this image are visually different from the textures in other images at the same position. It leads to information loss. This problem of information loss will be handled at the signal-level fusion stage.

6.3.2 Signal-level fusion

Given multiple noisy iris captures of the same eye, the SE-LRMF model proposed in Section 6.3.1 can be used to seek a low rank component \mathbf{A} as the noiseless images and an error component \mathbf{E} as the noise. In this subsection, \mathbf{A} and \mathbf{E} are utilised individually for signal-level fusion. It leads to a low rank mean (LRM) fusion and an error weighted mean (EWM) fusion.

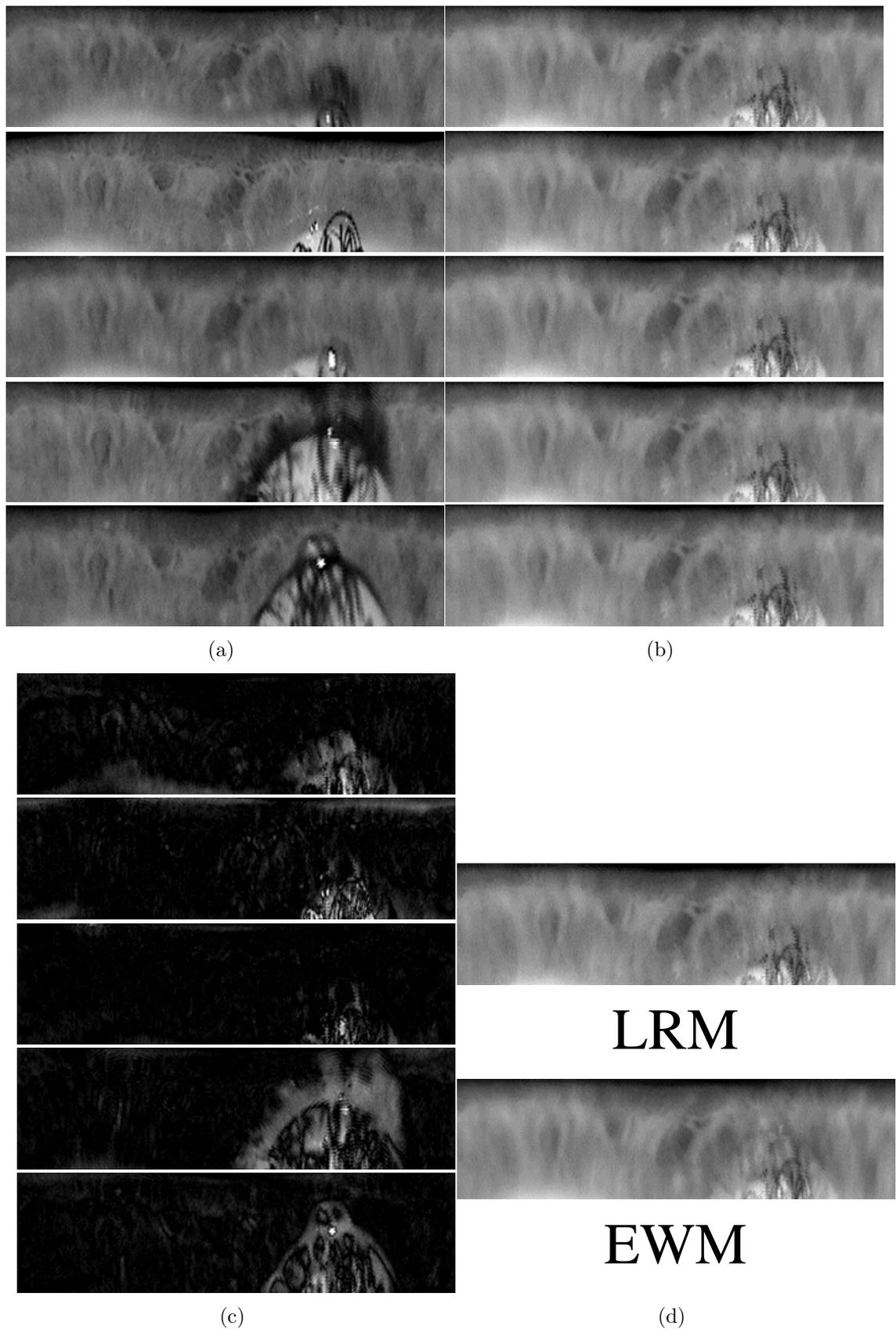


Figure 6.2: Examples of the results of SE-LRMF based signal-level fusion. (a) The original unwrapped iris captures from the same eye; (b) the low rank component (noiseless iris images) of the images in (a) sought by SE-LRMF; (c) the error component (noise) of the images in (a) sought by SE-LRMF; (d) the fusion result using low rank mean (LRM) fusion and error weighted mean (EWM) fusion.

Low rank mean (LRM) fusion. Since \mathbf{A} is viewed as a matrix of noiseless iris images, the images in \mathbf{A} can be fused directly by mean fusion. That is, a fused image is produced by calculating the mean of the images corresponding to each column of \mathbf{A} . This fusion method is named low rank mean, since it is the mean fusion on the low rank component sought by SE-LRMF.

However, simple mean fusion on noiseless component is not enough. It is found that although the influence of noise is mitigated in the low rank component \mathbf{A} , some images in \mathbf{A} suffer from texture loss. The reason is that, in some cases, some local textures in one image are visually different from the textures in the other images at the same position. Consequently, these textures are possible to be excluded from the low rank component \mathbf{A} and considered as noise (see the example discussed at the end of Section 6.3.1). It results in information loss and influences the performance of iris recognition using LRM fusion. Therefore, an error weighted mean (EWM) fusion method is proposed to mitigate this problem. EWM fusion utilises all the information in the original images. It mitigates the influence of noise by a weighting scheme based on the error component \mathbf{E} sought by SE-LRMF.

Error weighted mean (EWM) fusion. Recall that the matrix of original iris captures is denoted by $\mathbf{Y} \in \mathbb{R}^{m \times n}$ with each of its column being an observation. Let $\mathbf{y}_i \in \mathbb{R}^m$ be the i^{th} column of \mathbf{Y} . The proposed EWM fusion is formulated as follows:

$$\mathbf{y}^f = \omega^1 \otimes \mathbf{y}_1 + \omega^2 \otimes \mathbf{y}_2 + \cdots + \omega^n \otimes \mathbf{y}_n \quad (6.12)$$

where $\mathbf{y}^f \in \mathbb{R}^m$ denotes the fused image, \otimes is element-wise multiplication operator, $\omega^i = [\omega_1^i, \omega_2^i, \dots, \omega_m^i]^T \in \mathbb{R}^m$ is a vector of weights corresponding to \mathbf{y}_i . The weights in $\omega^1, \omega^2, \dots, \omega^n$ satisfy that, for any arbitrary $j \in [1, m]$, $\sum_{i=1}^n \omega_j^i = 1$. It means that the weights on the same position of all the images have a sum of 1.

Differently from the global weighting scheme in the existing weighted mean fusion methods [27,28], EWM uses a pixel-wise weighting scheme. The weights are assigned

based on the individual reliability of each pixel; the weights of different pixels in the same image can be different. Compared to the existing global weighting method, EWM has a more flexible and finer assignment of the weights. Therefore, it is expected that EWM is able to better preserve usable information while suppressing noise.

Given Eqn. 6.12, the key problem is how to compute the weights in $\omega^1, \omega^2, \dots, \omega^n$. In this chapter, these weights are calculated based on the error component \mathbf{E} obtained by SE-LRMF model. The idea is straightforward: \mathbf{E} is considered as the estimation of noise; for a pixel, the higher the noise level, the lower the reliability of the information in this pixel; hence, the weights should be inversely proportional to $|\mathbf{E}| \in \mathbb{R}^{m \times n}$. Based on this idea, the weights are calculated as follows.

Firstly, the elements in $|\mathbf{E}|$ are normalised to between 0 and 1 by:

$$\mathbf{E}' = \frac{|\mathbf{E}| - f_{\min}(|\mathbf{E}|)}{f_{\max}(|\mathbf{E}|) - f_{\min}(|\mathbf{E}|)} \quad (6.13)$$

where \mathbf{E}' denote the normalised matrix of $|\mathbf{E}|$, $f_{\min}(\bullet)$ and $f_{\max}(\bullet)$ find the minimum and maximum value in a matrix, respectively.

Then, a weight matrix $\mathbf{W}' \in \mathbb{R}^{m \times n}$ is calculated by:

$$\mathbf{W}' = \mathbf{1} - \mathbf{E}' \quad (6.14)$$

Thus, the elements in \mathbf{W}' are inversely proportional to $|\mathbf{E}|$.

Finally, each row of \mathbf{W}' is normalised to have a sum of 1. The columns of the normalised \mathbf{W}' are used as $\omega^1, \omega^2, \dots, \omega^n$. The reason is that each column of \mathbf{E} corresponds to the estimated noise in an iris capture, so each column of the normalised \mathbf{W}' can be considered as the weights corresponding to each image.

Iris mask production. The final problem in signal-level fusion is how to produce

an iris mask for the fused image. This iris mask indicates the detected noise like specular reflection. In this paper, a simple AND rule is adopted: given a set of iris captures, a pixel is masked as noise if it is detected as noise in the iris segmentation phase of any individual iris captures. Note that such a scheme is possible to lead to information loss, since it is possible to eliminate the pixels successfully reconstructed in the low rank component. A better solution is to seek a threshold based on the times a position is detected as noise in all the iris captures. However, it is experimentally found that this AND rule is sufficient to obtain a well performance. Also, it avoids the selection of a threshold which might be data dependent.

Example results of LRM and EWM are shown in Fig. 6.2(d). Note that, despite of a similar visual appearance, the results of these two methods contain the information sought from different aspects: LRM eliminates all the estimated noise at a cost of information loss; EWM utilises all the information in the original iris captures while suppressing the estimated noise.

6.3.3 Code level combination

The result of either LRM or EWM fusion is not fully noise-free. For LRM, it is possible to fail to estimate the noiseless information when a position is consistently occupied by noise in most of the images. As for EWM, it utilises the original iris captures for fusion. The weighting scheme of EWM is able to mitigate the noise, but not fully eliminate it. Therefore, a code level combination method is proposed to further suppress the noise in the results of LRM and EWM.

This code level combination method is based on the inherent stability of iris patterns: assuming that there are some fully noise-free iris captures of the same eye, the iris codes extracted from these captures should be highly consistent. Based on this, more reliable bits in iris codes can be sought via the bit consistency between the iris codes extracted from the LRM and EWM results. In other words, a bit is considered more

reliable if it is consistent between the iris codes extracted from LRM and EWM results. By only preserving these more reliable bits for iris matching stage, it is expected that an improved performance can be achieved. Essentially, the idea is similar to the iris weight map methods [22, 97, 98].

Specifically, given multiple noisy iris captures of the same eye, two images are obtained via the proposed signal-level fusion method, one using LRM and the other one using EWM. Two iris codes are extracted from the two images. The two iris codes are fused by only preserving the bits that are consistent between them. This can be simply implemented by using either one of the two iris codes as the final iris code, and eliminating the inconsistent bits between the two iris codes in the corresponding iris mask.

6.4 Experiment

This section presents the results of experimental analysis for the proposed signal-level fusion method. First, the experiment setting is introduced, including datasets, fusion setting, parameter setting and performance evaluation (Section 6.4.1). Then, with the low rank approximation method fixed, the effect of each individual signal-level fusion step is analysed, including low rank mean (LRM) fusion, error weighted mean (EWM) fusion and code level combination (Section 6.4.2). Next, with the signal-level fusion steps fixed, the performance of the propose SE-LRMF model for low rank approximation is studied (Section 6.4.3). The performance of SE-LRMF is compared with representative low rank approximation algorithms. After that, the proposed SE-LRMF based signal-level fusion is compared with existing methods, including no-fusion (Section 6.4.4), signal-level fusion (Section 6.4.5), score-level fusion (Section 6.4.6) and other recent iris recognition algorithms (Section 6.4.7). Finally, the influence of noise level on the performance of SE-LRMF is analysed (Section 6.4.9).

6.4.1 Experimental setting

Datasets. The experiments are conducted on three benchmark datasets as introduced in Chapter 3: ND0405 [103], CASIAD [102] and UBIRIS2 [19]. The three datasets cover the data captured in varying less constrained environments. ND0405 represents the iris data with a relatively lower noise level. The iris images in this subset are captured at a wavelength inducing less noise (NIR) and a close distance, hence with good resolution and clear iris texture, but suffering from the noise due to real-world conditions. In this experiment, a subset with the first 15 images of each eye is used. For the eyes with less than 15 images, all the iris images are used. Furthermore, 82 images are eliminated from the subset due to too poor quality (these images have too small iris region due to factors like heavy eyelids occlusion and large off-angle). It leads to a ND-iris-0405 subset consisting of 9811 images from 712 eyes to use in the experiment.

CASIAD represents the iris data with a medium noise level. The images are captured at NIR wavelength but at a distance. As a result, the iris region has a low resolution; the iris texture is visually less clear; the noise in this dataset is heavier than that in ND0405. In this experiment, all captures in CASIAD are adopted, and 97 captures are eliminated from CASIAD because of too poor quality (most of the eye regions in these images are completely covered by specular reflection). It results a subset with 5037 iris images from 284 eyes to use in the experiment.

UBIRIS2 represents the iris data with a relatively higher noise level. The images are captured not only at a distance, but also in a wavelength inducing heavy noise (visible). In the experiment, a subset with 1000 images from 171 eyes is used. This subset was released for NICE II contest [18]. It is reasonable to use this subset as a representative set for the data in the whole UBIRIS2 dataset, because of its good generality. This subset was released for a public contest, and it is widely used in many following research like [21–23, 80]. Also, due to its good generality, using

this subset enables to roughly compare the performance of algorithms in relevant literature. This experiment operates on the red channel of RGB colour space.

For all the datasets, the iris region is segmented using the algorithm in Chapter 4. The failed segmentations are manually corrected. The segmented iris regions are unwrapped using Daugman’s rubber sheet model [6, 7]. The size of unwrapped iris image is set to 100×360 . The unwrapped iris image is divided by a factor of 255 to normalise the intensity to $[0, 1]$. 1-D log-Gabor filter [15] is used to extract iris codes.

Fusion setting. Given the above datasets, gallery and probe sets need to be constructed for experimental analysis of information fusion methods. This chapter simulates the scenario where both gallery and probe images are captured in less constrained environment, so both of them suffer from the influence of noise. Accordingly, both gallery and probe sets are constructed by fusing the information in multiple noisy captures. Note that some research like [27, 28] directly use high quality still iris images as gallery without fusion. This scenario is not adopted in this chapter, because it is difficult to obtain high quality still iris images to construct the gallery in some applications like surveillance or mobile device-based applications. This experiment simulates a scenario which has less requirements on the iris image quality in both registration and iris matching phases. Under such scenario, most existing feature-level fusion methods [30, 31, 128] are less applicable, since these methods need high quality iris images to learn the relationship between low resolution and high resolution features. Thus, signal-level fusion and score-level fusion are considered in the experiment.

Based on the above simulated scenario, for signal-level fusion, the gallery and probe sets are constructed as follows. For a specific eye, the first 5 images are fused to produce a single gallery iris code. Then, from the remaining images of this eye, 5 images are randomly selected and fused to produce a probe iris code. To produce multiple probe iris codes, the above random selection and fusion based on

the remaining images are performed multiple times. It simulates the case that, in the iris matching phase, the captured images can be different at each time of matching. If an eye has less than 5 images, all the images are fused to produce the gallery iris code, and there are no probe iris codes. If an eye has less than 10 images, the first 5 images are fused to produce the gallery iris code, and all the remaining images are fused to produce a probe iris code without random selection. The number of images for fusion (5) is chosen so that the information included in images is enough for fusion, while a sufficient number of probe iris codes can be produced as well.

For score-level fusion, the fusion setting is similar. The difference is that, given a set of iris captures, instead of fusing them and producing one iris code, iris codes are extracted from each individual iris capture to produce an iris code set. In the matching phase, given a gallery iris code set and a probe iris code set, the final matching score is calculated by fusing the matching scores between the iris codes in the probe and gallery iris code sets.

Parameter setting and performance evaluation. Exclusive data are used for parameter setting and performance evaluation. For ND0405 dataset, the first 20 eyes are used to tune the parameters, and the remaining 692 eyes are used to evaluate the performance. For CASIAD dataset, the first 20 eyes are used to tune the parameters, and the remaining 264 eyes are used to evaluate the performance. For UBIRIS2 dataset, the first 19 eyes are used to tune the parameters, and the remaining 152 eyes are used to evaluate the performance.

Considering parameter setting, there are 4 parameters to be determined: r is the rank of low rank approximation in SE-LRMF model; λ in Eqn. 6.3 is a parameter controlling the trade-off between the reconstruction error and the sparsity of noise; $1/f_0$ and σ/f_0 are 1-D log-Gabor parameters (see Eqn. 2.25). Based on the above described data for parameter tuning, the setting of these parameters are reported in Tab. 6.1. Note that the 1-D log-Gabor parameters are learned based on the noiseless estimate of iris images obtained by the proposed SE-LRMF model.

Table 6.1: Parameter setting on each dataset.

Dataset	r	λ	$1/f_0$	σ/f_0
ND0405	1	0.01	35	0.49
CASIA4	1	0.01	19	0.48
UBIRIS2	1	0.01	40	0.38

For performance evaluation of signal-level fusion, the above settings of fusion and performance evaluation data lead to the following gallery and probe sets: 692 gallery iris codes and 6200 probe iris codes for ND0405; 264 gallery iris codes and 5114 probe iris codes for CASIAD; 152 gallery iris codes and 606 probe iris codes for UBIRIS2. As for score-level fusion, the number of fused intra-class and inter-class matching scores is exactly the same as signal-level fusion.

The performance of algorithms is examined in two tasks: identification and verification. The identification performance is evaluated by cumulative match characteristic (CMC), while the verification performance is evaluated by receiver operating characteristic (ROC).

6.4.2 Analysis of each signal-level fusion step

As described in Section 6.3, given the low rank and error components produced by SE-LRMF, the proposed method includes three fusion steps: low rank mean (LRM) fusion, error weighted mean (EWM) fusion and code level combination (CLC). In this subsection, the effect of each individual fusion step is analysed. Mean fusion (MF) is used as the baseline, because EWM is essentially an improved mean fusion. Fig. 6.3 shows the CMC and ROC curves of MF, LRM, EWM, CLC.

There are two main observations from Fig. 6.3. First, EWM achieves a generally better performance than MF. This result illustrates that the proposed EWM can better mitigate the noise in the iris captures compared to MF. Second, it can be seen that, although the performance of LRM and EWM vary on different datasets, CLC consistently leads to improved CMC and ROC curves compared to either LRM or EWM

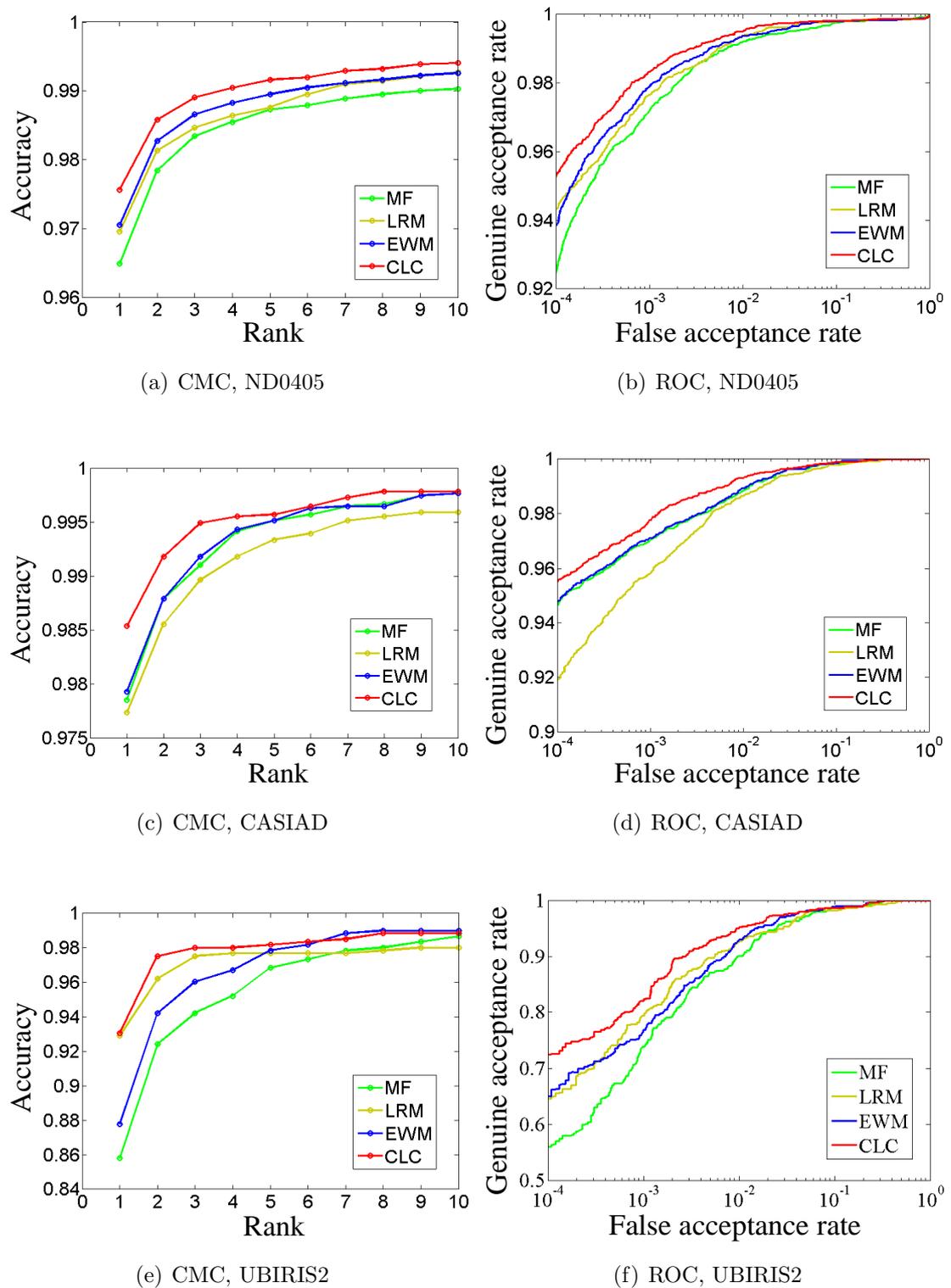


Figure 6.3: The CMC and ROC curves to analyse each signal-level fusion step.

on all the datasets. This observation suggests that with SE-LRMF for low rank approximation, CLC is able to achieve an improved and more stable performance in comparison to LRM and EWM. It validates the effectiveness of the proposed code level combination.

6.4.3 Comparison between SE-LRMF and existing low rank approximation algorithms

This subsection makes comparisons between the proposed SE-LRMF model and existing low rank approximation (LRA) algorithms to study the effectiveness of SE-LRMF in the proposed framework of signal-level fusion. The comparisons are made by: (1) fixing the fusion steps (LRM, EWM, CLC); (2) varying the algorithm of LRA to produce the low rank and error components used by the fusion steps.

As described in Section 6.2.2, existing LRA algorithms can be categorized into low rank decomposition and low rank matrix factorisation. Accordingly, the comparisons are made with a representative low rank decomposition algorithm and a representative low rank matrix factorisation algorithm. In terms of low rank decomposition, the comparisons are made with robust principal component analysis (RPCA) in [131], since RPCA is currently one of the most widely used algorithms for low rank decomposition. As for low rank matrix factorisation, the comparisons are made with the cyclic weight median algorithm with ℓ_1 error term proposed in [137] (referred as L1-LRMF). As reported in [137], this algorithm outperforms most existing LRA algorithms in the task of reconstructing noiseless images from multiple corrupted observations. To eliminate the influence of CLC and have an insight on the effect of the low rank and error components produced by different LRA algorithms, the performance of LRM, EWM and CLC with different LRA algorithms are reported in Fig. 6.4.

There are several observations from the results in Fig. 6.4. First, considering the

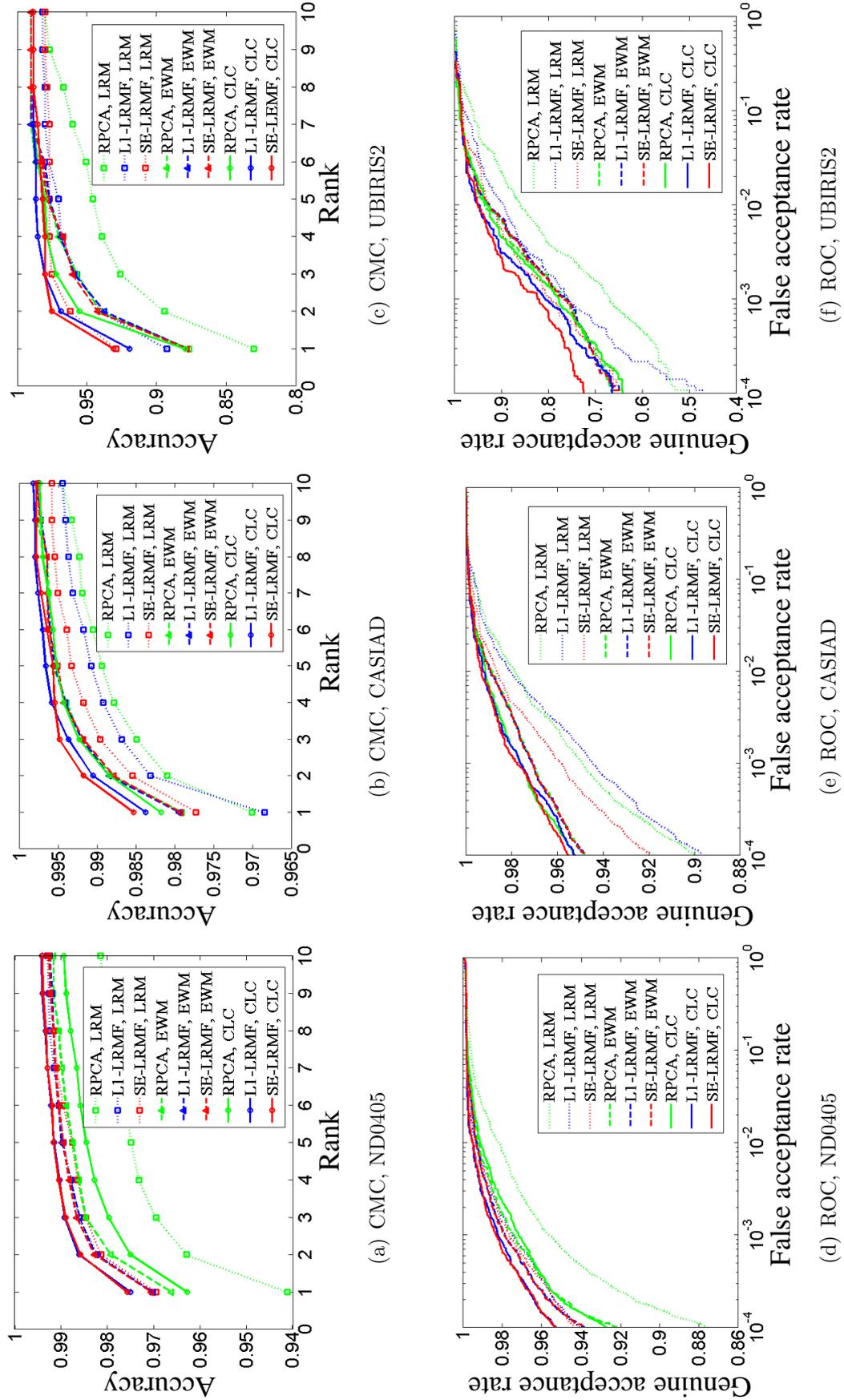


Figure 6.4: The CMC and ROC curves for the comparison between SE-LRMF and existing low rank approximation models.

results of LRM, it is found that the proposed SE-LRMF achieves a generally better performance in comparison to RPCA and L1-LRMF. Specifically, on ND0405 dataset, LRM with SE-LRMF achieves comparable identification and verification performance to LRM with L1-LRMF, and the performance is better than RPCA. On CASIAD and UBIRIS2 datasets, LRM with SE-LRMF outperforms LRM with RPCA and L1-LRMF in both identification and verification. Moreover, the performance of LRM with SE-LRMF is more stable. It can be seen that LRM with L1-LRMF performs better than LRM with RPCA on ND0405 and UBIRIS2 datasets, while LRM with RPCA is better than LRM with L1-LRMF on CASIAD dataset. In contrast, LRM with SE-LRMF achieves a more stable performance at a top level on all the datasets. This observation suggests that, for iris recognition using LRM fusion, the low rank components sought by SE-LRMF are able to achieve a more stable and top-level performance, compared to the low rank components sought by RPCA and L1-LRMF.

Second, generally, it can be found that EWM with different LRA algorithms have similar performance. It means that varying the LRA algorithm has little influence on the result of EWM. The possible reason is that, although different LRA algorithms lead to varying estimations on low rank components, these LRA algorithms have similar results on error components. Therefore, EWM with different LRA algorithms has similar performance. The only exception is ND0405 where EWM with SE-LRMF and L1-LRMF perform better than EWM with RPCA. Since that ND0405 is captured at a close distance, this observation may suggest that EWM with SE-LRMF and L1-LRMF are preferable for this type of iris data. When the capturing distance grows, EWM with different LRA algorithms have little difference.

Third, given the similar performance of EWM, the performance variation of CLC with different LRA algorithms is not as large as that of LRM with different LRA algorithms. This observation validates the effectiveness of CLC: it further eliminates the unreliable bits in LRM results, hence different LRA algorithms have similar

performance after CLC. However, it can be found that CLC with SE-LRMF still has a top performance in comparison to CLC with RPCA and L1-LRMF. In terms of identification, on ND0405 dataset, CLC with SE-LRMF achieves a comparable rank 1 recognition accuracy to CLC with L1-LRMF, and this performance is better than CLC with RPCA. On CASIAD and UBIRIS2 datasets, the rank 1 recognition accuracy of CLC with SE-LRMF is higher than that of CLC with RPCA and L1-LRMF. As for verification, on ND0405 dataset, CLC with SE-LRMF achieves a comparable ROC curve to CLC with L1-LRMF, and this performance is better than CLC with RPCA. On CASIAD dataset, the ROC curve of CLC with SE-LRMF is slightly better than that of CLC with RPCA and L1-LRMF. On UBIRIS2 dataset, CLC with SE-LRMF also leads a better ROC curve. Considering the noise level of all the datasets (ND0405 (close + NIR) < CASIAD (distant + NIR) < UBIRIS2 (distant + visible)), it can be concluded that CLC with SE-LRMF performs better for the data with relatively heavier noise, while it is still able to achieve a top performance on the data with less noise.

Four, it can be observed that CLC does not always lead to an improved performance over LRM and EWM. On ND0405 dataset, CLC with RPCA has a lower rank 1 recognition accuracy than EWM with RPCA. Also, on ND0405 and UBIRIS2 datasets, the ROC curves of CLC with RPCA and EWM with RPCA are very similar. A possible reason is that, on these two datasets, the result of LRM with RPCA contains a high level of noise. The noise level is so high that it influences the result of CLC: too many noiseless bits in EWM results are excluded due to their inconsistency with the noisy bits in LRM results. In other words, the result of CLC is biased due to noise.

An experiment is conducted to gain a deeper insight into the above bias effect. In this experiment, the noise level in LRM and EWM results are measured by the percentage of inconsistent bits between the iris codes extracted from LRM and EWM results. A higher percentage of inconsistent bits means that at least one of the LRM and EWM

Table 6.2: Percentage of inconsistent bits between LRM and EWM fusion results

Dataset	RPCA	L1-LRMF	SE-LRMF
ND0405	9.74%	8.88%	8.29%
CASIAD	18.02%	17.36%	14.04%
UBIRIS2	23.47%	20.71%	17.61%

results has a high noise level, since the iris codes extracted from LRM and EWM results should be highly consistent if both LRM and EWM results are low-noise or noise-free, due to the inherent stability of iris pattern. The detailed experiment is as follows. Given the fusion setting in Section 6.4.1, there are multiple image sets for fusion to produce gallery and probe iris codes. Given a specific image set and a LRA algorithm, a LRM fused image and an EWM fused image are produced, and the percentage of inconsistent bits between the iris codes extracted from the two LRM and EWM fused images are measured. For a specific LRA algorithm, the mean percentage of inconsistent bits calculated using all the image sets is used as the estimation of the noise level in the fusion results of this LRA algorithm (note that this experiment does not distinguish gallery and probe). Tab. 6.2 reports the percentage of inconsistent bits measured using RPCA, L1-LRMF and SE-LRMF on the three datasets.

By investigating the results in Tab. 6.2, it can be found that the percentage of inconsistent bits increases with the noise level of datasets ($ND0405 < CASIAD < UBIRIS2$). It is consistent with the above analysis to use the percentage of inconsistent bits to measure the noise level. Also, it is very clear that RPCA has the highest percentage of inconsistent bits among the three LRA algorithms. It means that the fusion results with RPCA contain relatively heavier noise. Consequently, the result of CLC with RPCA is more likely to be biased due to the noise in the LRM and EWM fused images. This explains the above experimental observation four. Moreover, it can be found that SE-LRMF has the lowest percentage of inconsistent bits among the three LRA algorithms on all the three datasets. This observation demonstrates that the LRM and EWM fused images with SE-LRMF have the lowest

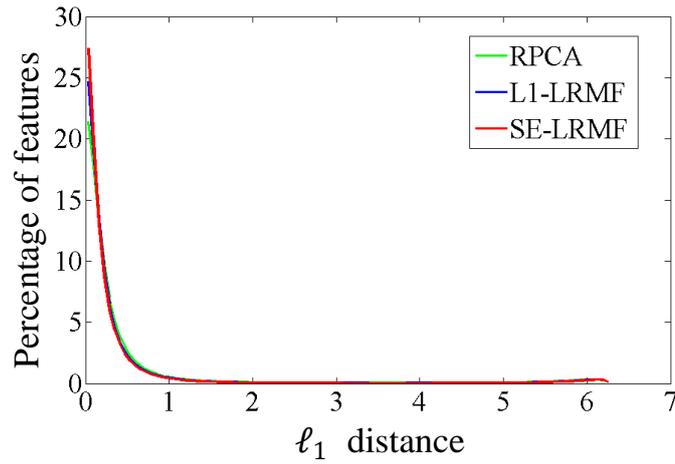
noise level. Thus, CLC with SE-LRMF is more likely to avoid the bias due to noise. In other words, compared to RPCA and L1-LRMF, SE-LRMF has the potential to achieve a more robust performance in the proposed framework for signal-level fusion.

Moreover, another similar analysis is performed. In this analysis, the noise level in LRM and EWM results is measured using the ℓ_1 distance between the raw feature values (feature values before quantisation) extracted from these two results. The smaller the ℓ_1 distance, the lower the noise level. The reason is similar to the one stated in the above analysis via bit inconsistency. Fig. 6.5 shows the distribution of ℓ_1 distances calculated from all the image sets for fusion on the three datasets, with RPCA, L1-LRMF and SE-LRMF for LRA.

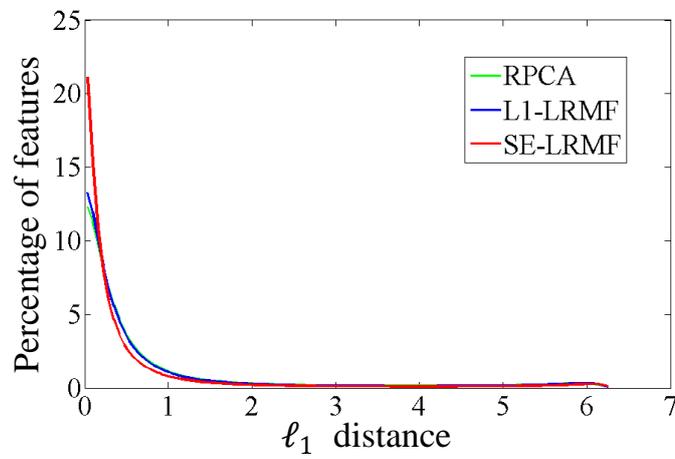
It can be seen that the result is similar to the analysis based on bit inconsistency. The distribution is consistent with the noise level in each dataset: for the dataset with a lower noise level, there are more raw feature values with a smaller ℓ_1 distance between LRM and EWM results. Also, among the three LRA algorithms, RPCA has the smallest number of small ℓ_1 distances on all the datasets. It means the fusion result of RPCA has heavier noise, and this explains the observation four above. Finally, compared to RPCA and L1-LRMF, the proposed SE-LRMF leads to more small ℓ_1 distances and fewer large ℓ_1 distances on the three datasets. This means that, compared to RPCA and L1-LRMF, SE-LRMF leads to less noise in fusion results, so it is more likely to avoid the bias due to noise, achieving a more robust performance.

6.4.4 Comparison with no fusion methods

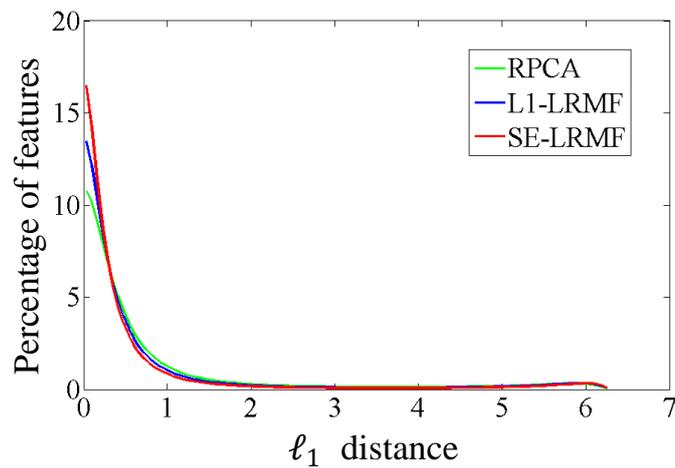
The iris recognition results without fusion are used as the baseline. The comparisons are made between the proposed SE-LRMF based signal-level fusion (referred to as SE-LRMF) and no fusion baseline methods. Two no fusion methods are designed.



(a) ND0405



(b) CASIAD



(c) UBIRIS2

Figure 6.5: The distribution of ℓ_1 distances between the raw feature values extracted from LRM and EWM results.

No fusion method 1 (referred to as NF1) is a single gallery single probe method. It simulates the case that one image per eye is used as the gallery in registration phase, and a single probe image is used for matching in iris matching stage. The comparison between NF1 and SE-LRMF explores the performance between: (1) only using the information in single capture as gallery and probe and (2) fusing the information in multiple captures to produce gallery and probe. The gallery and probe sets for NF1 are constructed as follows. Recall that, in the fusion setting in Section 6.4.1, there are some image sets for fusion to construct gallery and probe sets for information fusion methods. To construct the gallery of NF1, all the image sets used to construct the gallery in the fusion setting are retrieved. From each retrieved image set, 1 image is randomly selected, and the iris code is extracted from the selected image as one NF1 gallery iris code. In other words, given a set of images from the same eye, SE-LRMF fuses them to produce one gallery iris code, while NF1 randomly selects one of them to produce one gallery iris code. It leads to a NF1 gallery set with the same iris code number and label to the SE-LRMF gallery set. The same method is used to produce the probe set of NF1. Given a pair of gallery and probe sets of NF1, CMC and ROC curves are calculated to evaluate the performance. The above random selection of gallery and probe sets of NF1 is repeated for 50 times. The mean of the CMC and ROC curves obtained from the 50 runs are used as the performance of NF1 for comparison. Note that the above procedure may lead to repeated probe iris codes in the NF1 probe set, since in the fusion setting, the image sets used to produce the probe set can overlap. However, the random selection and multiple runs make NF1 a reasonable estimator of the performance of a single probe single gallery no fusion setting.

No fusion method 2 (referred to as NF2) is a multiple gallery single probe method without fusion. It simulates the case that the information in gallery is enriched by capturing multiple iris images per eye but without fusing them. In matching phase, a single probe is captured and compared with the gallery. The probe set of NF2

is constructed using exactly the same method as NF1. As for the gallery set, all the image sets used to produce the gallery in the fusion setting are retrieved, and the iris codes extracted from all the retrieved images are used as gallery. Similar to NF1, the above procedure is repeated 50 times, and mean CMC and ROC curves are used as the estimation of the performance of NF2 for comparison. Note that for NF2, the gallery set is exactly the same in each run, and only the probe set is different due to random selection.

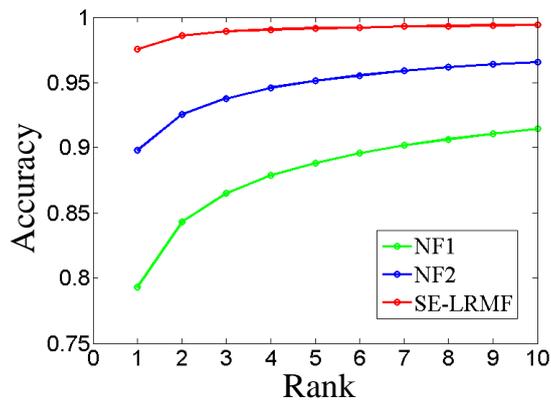
Fig. 6.6 reports the CMC and ROC curves of NF1, NF2 and SE-LRMF on all the datasets. It can be seen that: (1) NF2 has a better performance compared to NF1; (2) SE-LRMF leads to a significantly improved performance compared to the two no fusion methods. The results demonstrate that: (1) an improved performance can be achieved by using more gallery images; (2) signal-level fusion of the information in multiple captures performs better than using the information in single captures without fusion.

6.4.5 Comparison with existing signal-level fusion methods

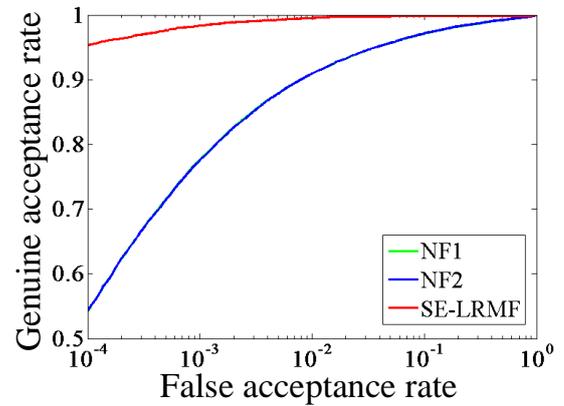
This subsection compares the proposed SE-LRMF based signal-level fusion method with existing signal-level fusion methods. The comparisons are made with mean fusion (MF) proposed in [26] and quality weighted mean fusion (QWM) fusion proposed in [28]. Fig. 6.7 reports CMC and ROC curves of MF, QWM and SE-LRMF. It can be seen that SE-LRMF based signal-level fusion outperforms existing signal-level fusion methods.

6.4.6 Comparison with existing score-level fusion methods

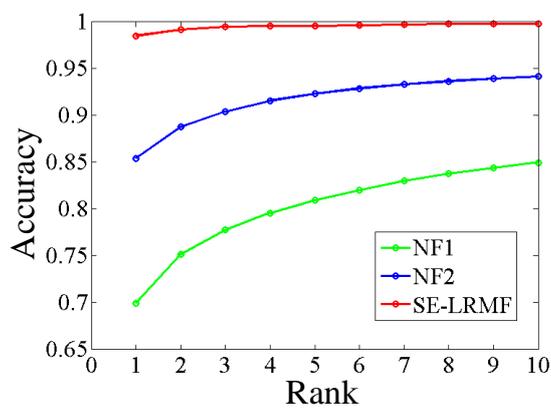
This subsection makes comparisons between the proposed SE-LRMF based signal-level fusion method and existing score-level fusion methods. The comparisons are



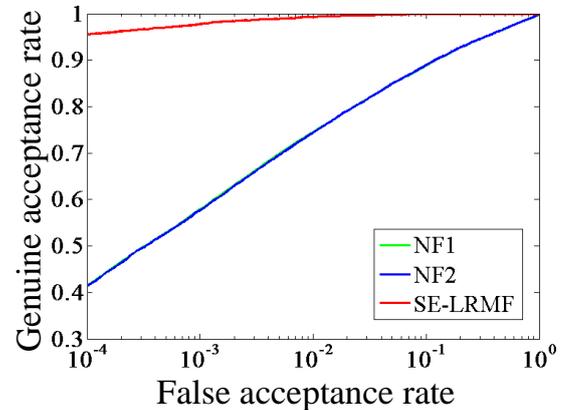
(a) CMC, ND0405



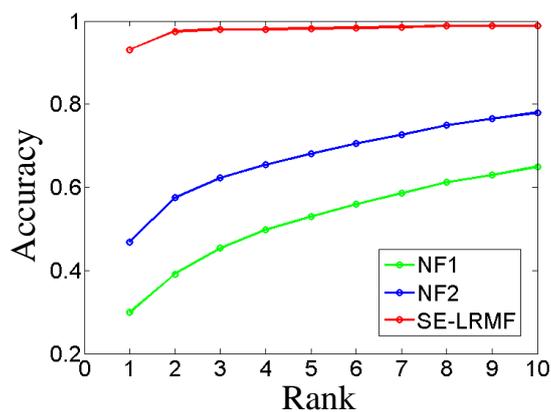
(b) ROC, ND0405



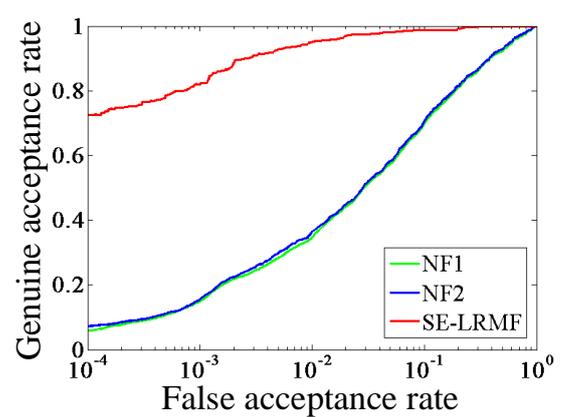
(c) CMC, CASIAD



(d) ROC, CASIAD

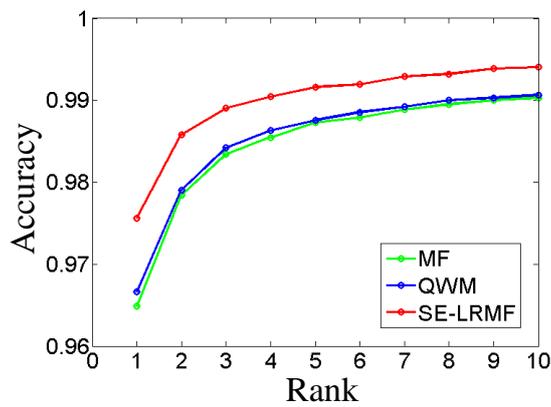


(e) CMC, UBIRIS2

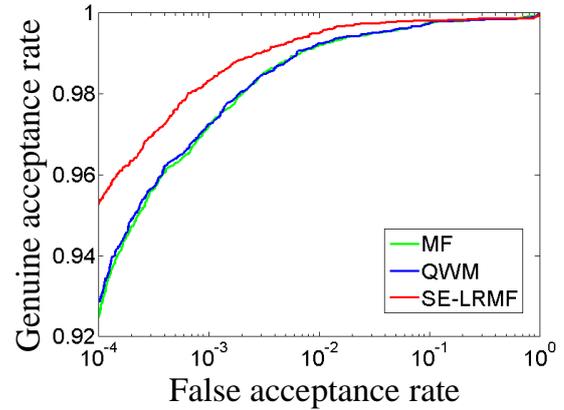


(f) ROC, UBIRIS2

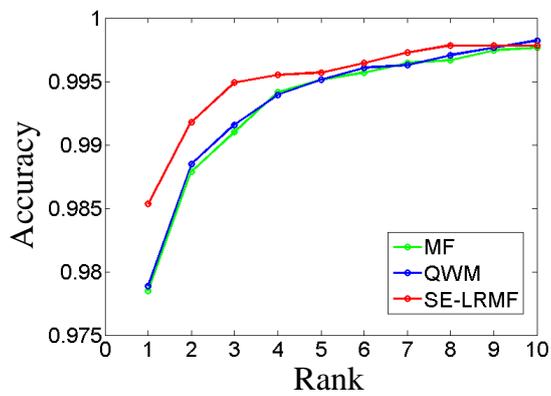
Figure 6.6: The CMC and ROC curves for the comparison between SE-LRMF and no fusion methods.



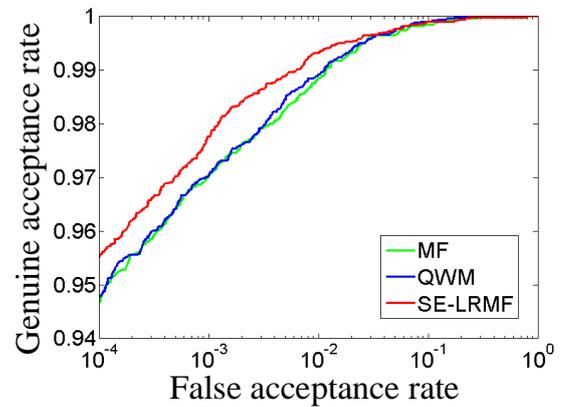
(a) CMC, ND0405



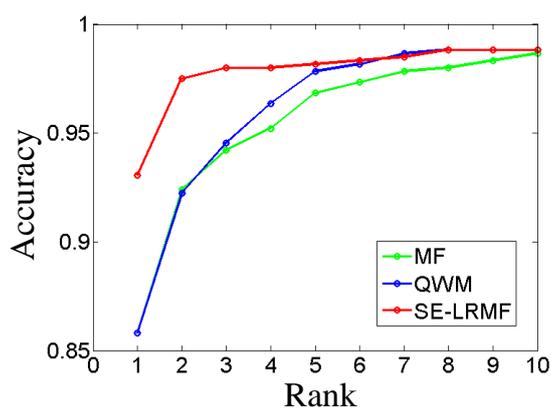
(b) ROC, ND0405



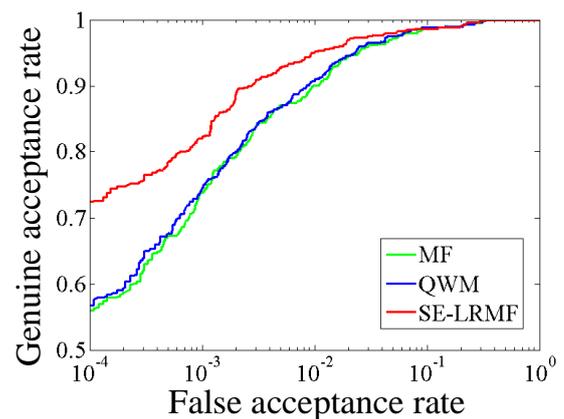
(c) CMC, CASIAD



(d) ROC, CASIAD



(e) CMC, UBIRIS2



(f) ROC, UBIRIS2

Figure 6.7: The CMC and ROC curves for the comparison between SE-LRMF and existing signal-level fusion methods.

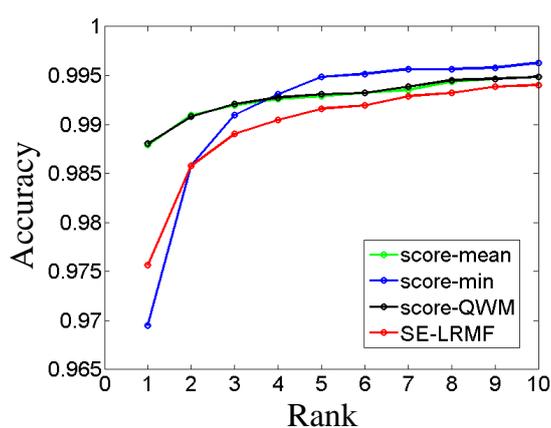
made with three score-level fusion methods: fusion using the mean of scores (referred to as score-mean) [26, 31], fusion using the minimum of scores (referred to as score-min) [26, 31], fusion using quality weighted mean of scores (referred to as score-QWM) [28, 128]. Fig. 6.8 reports the CMC and ROC curves of score-mean, score-min, score-QWM and SE-LRMF.

Considering the identification performance (CMC) shown in Fig. 6.8, it can be found that the performance of different algorithms differs according to the datasets. On ND0405, mean score based fusion methods (score-mean, score-QWM) achieve higher rank 1 recognition accuracies than SE-LRMF, while SE-LRMF performs better than score-min. On CASIAD, the performance of SE-LRMF is comparable to score-min, and it is better than mean score based methods. On UBIRIS2, SE-LRMF outperforms all the other methods.

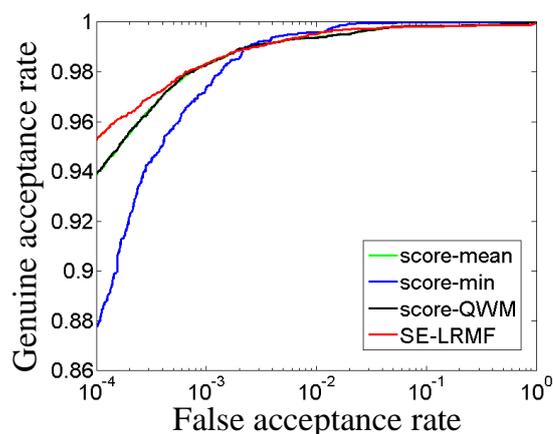
However, despite of the varying performance, it can be seen that the performance of SE-LRMF is close to the top score-level method on less noisy NIR datasets, and it achieves a significantly improved performance on more noisy colour dataset. The rank 1 recognition accuracy of SE-LRMF is 1.25% lower than the best method on ND0405, and it is 0.12% to the best method on CASIAD distance. On UBIRIS2, the rank 1 recognition accuracy of SE-LRMF is 10.40% higher than the second best method.

Also, considering the stability of identification performance, it can be found that SE-LRMF is more stable compared to each individual score-level fusion method. It can be seen from the CMC curves that the performance of score-level fusion methods is less stable. score-mean and score-QWM perform better than score-min on ND0405, but score-min performs better than the two methods on CASIAD and better than score-mean on UBIRIS2. In contrast, SE-LRMF achieves mid-level performance on the NIR datasets, and it performs best on the colour dataset.

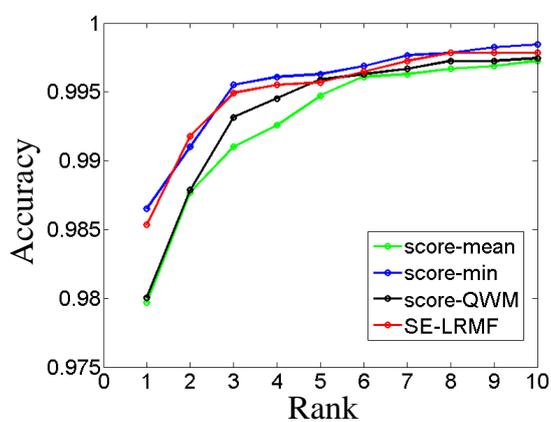
The above identification performance is closely-related to the noise level in the iris



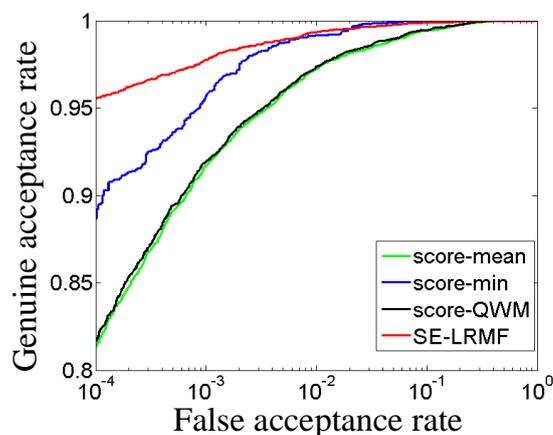
(a) CMC, ND0405



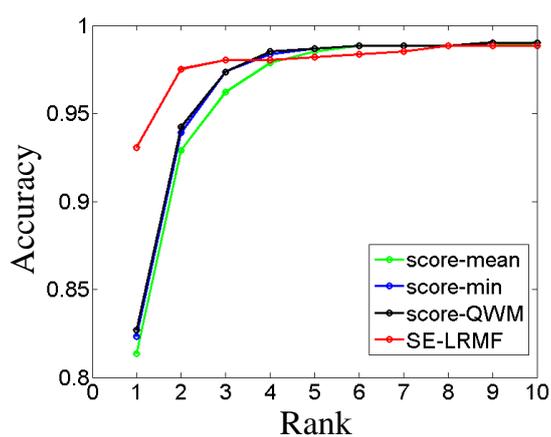
(b) ROC, ND0405



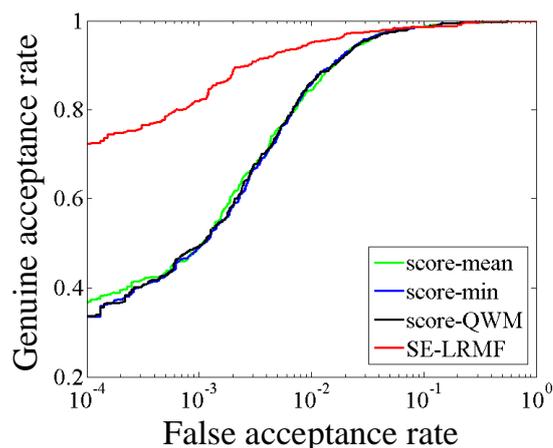
(c) CMC, CASIAD



(d) ROC, CASIAD



(e) CMC, UBIRIS2



(f) ROC, UBIRIS2

Figure 6.8: The CMC and ROC curves for the comparison between SE-LRMF and existing score-level fusion methods.

images in each dataset. Score-level fusion is sensitive to the noise level in iris captures. Lower noise level leads to higher reliability in the matching scores between individual gallery and probe iris codes, hence it results more reliable fused scores. On the contrary, if the noise level is too high, the matching scores between individual gallery and probe iris codes become highly unreliable. As a result, fusing these matching scores brings less advantage. Compared to score-level fusion, signal-level fusion is less sensitive to the noise level of iris captures. The reason is that the iris codes of signal-level fusion are extracted from the fused images, and the fusion process is able to suppress noise and enhance the stable structure in iris images.

The above analysis is consistent to the CMC curves in Fig. 6.8. ND0405 dataset has a relatively lower noise level. Therefore, on this dataset, the best score-level fusion method (score-QWM) performs better than the signal-level SE-LRMF. The noise level of the images in CASIAD dataset is higher than that in ND0405 dataset. Correspondingly, on this dataset, signal-level SE-LRMF performs similar to the best score-level fusion method (score-min). The images in UBIRIS2 dataset contain relatively heavier noise due to a combination of visible wavelength and distant capturing. Consequently, signal-level SE-LRMF still performs well, while the performance of score-level fusion methods heavily degrades.

In terms of verification performance, SE-LRMF achieves better ROC curves compared to score-level fusion methods on all the datasets. It is interesting that score-level fusion performs good in CMC curves but not in ROC curves. Such observation can be explained as follows. Score-level fusion is more accurate in intra-class matching. It leads to more accurate lowest matching scores and hence good CMC curves. However, score-level fusion is less effective in inter-class comparisons. It results in less accurate inter-class matching scores which influence ROC curves. In contrast, SE-LRMF performs well in both intra-class and inter-class matching, and it leads to good performance in both CMC and ROC curves.

In conclusion, compared to score-level fusion, (1) SE-LRMF based signal-level fusion

is able to achieve a better identification performance on the colour data with relatively heavier noise, while it also performs well on the NIR captures with relatively lower noise; (2) SE-LRMF based signal-level fusion is more stable in identification; (3) SE-LRMF based signal-level fusion achieves better verification performance on the iris data with varying noise levels.

6.4.7 Comparison with other recent iris recognition algorithms

This subsection makes comparisons with other recently proposed iris recognition algorithms. These algorithms generally focus on improving the feature extraction and iris matching stages, instead of performing information fusion using multiple captures. This comparison studies the effectiveness of information fusion in less constrained iris recognition.

The compares are made with two recent algorithms: (1) geometric key-based iris encoding [80] (referred to as GeoKey) and (2) stabilised iris encoding and Zernike moments phase features [23] (referred to as SZM). These two algorithms are chosen for comparison, since they are recently proposed algorithms achieving top performance compared to other state-of-the-art algorithms in less constrained iris recognition, as reported in [80] and [23]. Also, comparisons are made with the result of the fusion of the two algorithms by combining their matching scores (referred to as GeoKey+SZM).

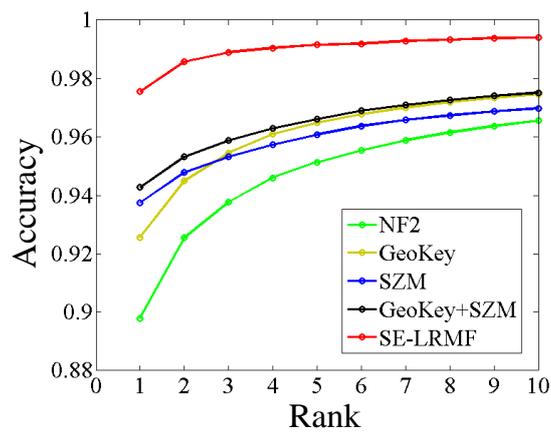
Since the above algorithms are not information fusion algorithms, in order to make a reasonable comparison, the performance of these algorithms are evaluated using the same experimental setting to the no fusion method 2 (NF2) in Section 6.4.4. Also, NF2 is used as a reference of baseline performance.

Fig. 6.9 shows the CMC and ROC curves for all the comparison algorithms on

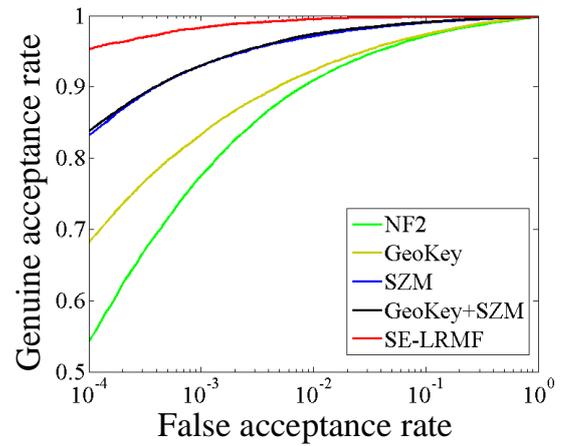
all the datasets. It can be seen that GeoKey, SZM and GeoKey+SZM perform significantly better than NF2, while SE-LRMF achieves top performance among all the comparison methods. Actually, nearly all the information fusion algorithms used in the comparisons in previous subsections achieve better performance than GeoKey, SZM and GeoKey+SZM. This result shows that information fusion is able to significantly improve the performance of less constrained iris recognition, even with less robust features (note that the results of SE-LRMF and the information fusion methods used in the comparisons in previous subsections are based on the same feature as NL2). A possible reason is that the robustness of the algorithm is significantly enhanced in the fusion process of signals (scores) from multiple captures.

6.4.8 Summary on the performance of SE-LRMF and existing algorithms

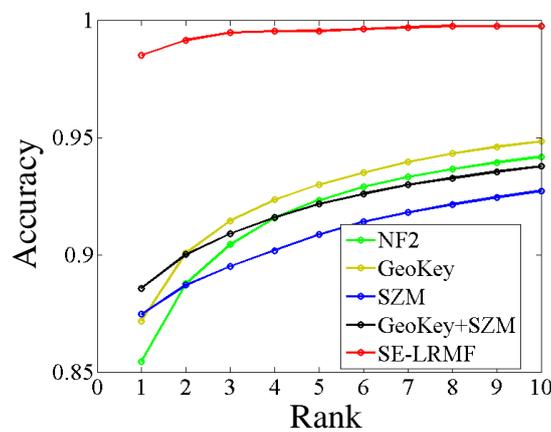
This subsection summarises the performance of SE-LRMF and existing algorithms. Specifically, rank 1 recognition accuracy (R1RA) is calculated to represent the identification performance; equal error rate (EER) and genuine acceptance rate (GAR) when false acceptance rate (FAR) is 0.01% ($\text{GAR@FAR}=0.01\%$) are calculated to represent the verification performance; decidability index (DI) [6] is calculated to represent the intra-class and inter-class separability of matching scores. Furthermore, a significance analysis is performed based on the above measures using paired student-t test. For each dataset, 100 subsets are randomly selected from the performance evaluation set; the subjects included in each subset are randomly selected; R1RA, EER, $\text{GAR@FAR}=0.01\%$ and DI are calculated for each subset, and paired student-t test is performed between each method and the best method in terms of R1RA, EER, $\text{GAR@FAR}=0.01\%$ and DI. The mean and standard deviations of R1RA, EER, $\text{GAR@FAR}=0.01\%$ and DI calculated using all the subsets, and the results of corresponding paired student-t test, are reported in Tab. 6.3, Tab. 6.4,



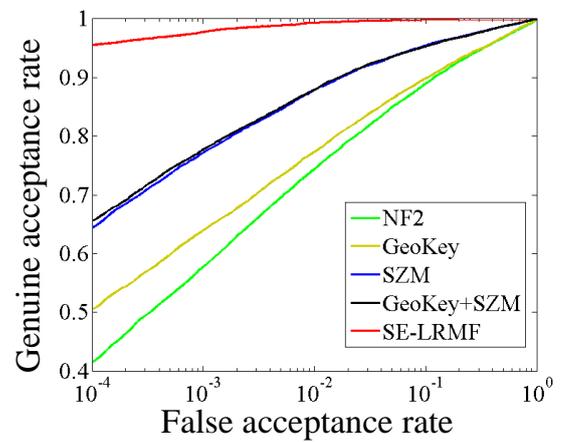
(a) CMC, ND0405



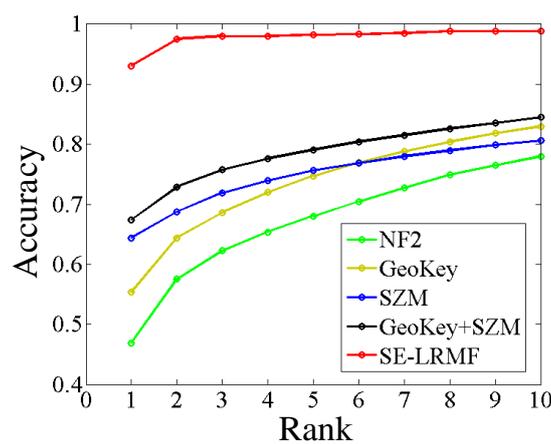
(b) ROC, ND0405



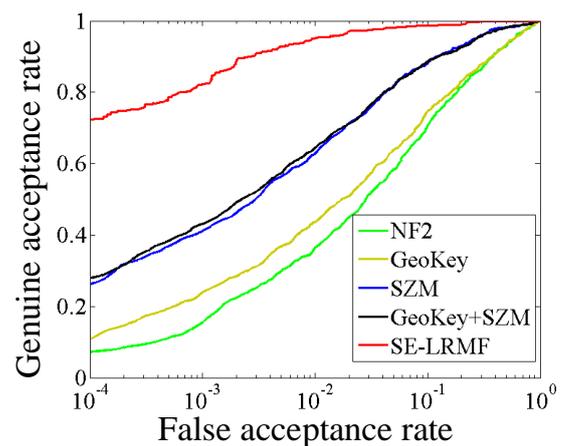
(c) CMC, CASIAD



(d) ROC, CASIAD



(e) CMC, UBIRIS2



(f) ROC, UBIRIS2

Figure 6.9: The CMC and ROC curves for the comparison between SE-LRMF and other recent iris recognition algorithms.

Tab. 6.5 and Tab. 6.6, respectively. It can be seen that (1) SE-LRMF achieves the highest overall ranking in all the measures, considering the performance on all the datasets; (2) SE-LRMF performs especially good for the datasets with heavier noise (*i.e.* CASIAD and UBIRIS2), in comparison to the existing methods.

6.4.9 The influence of noise on the performance of SE-LRMF

As shown in the comparison in Section 6.4.6 as well as Tab. 6.3, Tab. 6.4, Tab. 6.5, Tab. 6.6, the performance of SE-LRMF based fusion depends on the noise level of captures. Therefore, this subsection conducts experiments to gain a deeper insight into the influence of noise types and noise levels on the performance of SE-LRMF.

This experiment groups the images sets for fusion in the fusion setting into several groups based on the noise type and noise level, and it investigates the performance of the fusion result corresponding to each group of image sets. Some quality measures (see below) are used to compute the level of different types of noise in the data (*e.g.* blur, off-angle, *etc.*). Given an image set for fusion and given a quality measure, the quality of this image set is computed as the mean quality of each image in it. Given the image sets in the fusion setting evaluated by a specific quality measure, these image sets are grouped into 5 quality levels for each dataset. The first level includes the image sets with top 20% best quality; the second level includes the image sets with 21% to 40% quality; similarly for the other levels. Therefore, the group with a larger index for quality level has a higher noise level. In this chapter, the grouping is performed for the image sets used to produce the probe set in the fusion setting. To evaluate the performance of each group of image sets, the probes produced by each group are matched to the full gallery set produced by the fusion setting. It simulates the scenario that, in real applications, the gallery set usually remains unchanged after registration phase, and the variation of noise types and noise levels mainly exist in the probes.

Table 6.3: The R1RA of SE-LRMF and existing iris recognition algorithms (in percentage, format: mean \pm standard deviation (rank), best performance bolded).

Method	ND0405	CASIAD	UBIRIS2	MR*
NF2	92.49 \pm 1.20 (10) [#]	87.70 \pm 0.90 (10) [#]	52.15 \pm 4.97 (10) [#]	10.0
GeoKey	94.57 \pm 0.85 (9) [#]	89.10 \pm 0.82 (8) [#]	59.96 \pm 5.23 (9) [#]	8.7
SZM	95.27 \pm 0.81 (8) [#]	89.10 \pm 0.82 (8) [#]	67.32 \pm 4.43 (8) [#]	8.0
GeoKey+SZM	95.75 \pm 0.76 (7) [#]	90.10 \pm 0.77 (7) [#]	70.40 \pm 4.53 (7) [#]	7.0
socre-min	98.22 \pm 0.70 (4) [#]	98.99 \pm 0.39 (1)	86.69 \pm 5.04 (5) [#]	3.3
socre-mean	99.05 \pm 0.34 (2) [#]	98.45 \pm 0.53 (4) [#]	85.61 \pm 4.98 (6) [#]	4.0
socre-QWM	99.06 \pm 0.35 (1)	98.49 \pm 0.50 (3) [#]	87.00 \pm 4.80 (4) [#]	2.7
MF	97.57 \pm 0.88 (6) [#]	98.36 \pm 0.41 (6) [#]	88.23 \pm 4.64 (3) [#]	5.0
QWM	97.69 \pm 0.85 (5) [#]	98.39 \pm 0.40 (5) [#]	88.30 \pm 4.74 (2) [#]	4.0
SE-LRMF	98.40 \pm 0.63 (3) [#]	98.88 \pm 0.32 (2) [#]	94.68 \pm 2.26 (1)	2.0

[#]The p value of paired student-t test to the best method is lower than 0.05 level.

*MR—mean rank (best rank bolded)

Table 6.4: The EER of SE-LRMF and existing iris recognition algorithms (in percentage, format: mean \pm standard deviation (rank), best performance bolded).

Method	ND0405	CASIAD	UBIRIS2	MR*
NF2	4.45 \pm 0.35 (10) [#]	10.58 \pm 0.47 (10) [#]	19.35 \pm 1.24 (10) [#]	10.0
GeoKey	4.02 \pm 0.32 (9) [#]	10.02 \pm 0.48 (9) [#]	18.62 \pm 1.27 (9) [#]	9.0
SZM	2.00 \pm 0.24 (8) [#]	5.88 \pm 0.45 (7) [#]	11.11 \pm 1.05 (8) [#]	7.7
GeoKey+SZM	1.85 \pm 0.23 (7) [#]	5.89 \pm 0.43 (8) [#]	11.09 \pm 1.12 (7) [#]	7.3
socre-min	0.52 \pm 0.14 (1)	0.88 \pm 0.23 (2) [#]	3.58 \pm 0.58 (3) [#]	2.0
socre-mean	0.64 \pm 0.23 (3) [#]	1.89 \pm 0.40 (6) [#]	3.62 \pm 0.59 (4) [#]	4.3
socre-QWM	0.64 \pm 0.23 (3) [#]	1.80 \pm 0.37 (5) [#]	3.70 \pm 0.60 (5) [#]	4.3
MF	0.85 \pm 0.23 (6) [#]	1.04 \pm 0.19 (4) [#]	3.71 \pm 0.59 (6) [#]	5.3
QWM	0.83 \pm 0.23 (5) [#]	0.99 \pm 0.19 (3) [#]	3.45 \pm 0.51 (2) [#]	3.3
SE-LRMF	0.60 \pm 0.19 (2) [#]	0.77 \pm 0.16 (1)	2.71 \pm 0.52 (1)	1.3

[#]The p value of paired student-t test to the best method is lower than 0.05 level.

*MR—mean rank (best rank bolded)

Table 6.5: The GAR@FAR= 0.01% of SE-LRMF and existing iris recognition algorithms (in percentage, format: mean \pm standard deviation (rank), best performance bolded).

Method	ND0405	CASIAD	UBIRIS2	MR*
NF2	55.28 \pm 4.77 (10) [#]	41.58 \pm 2.24 (10) [#]	6.62 \pm 1.73 (10) [#]	10.0
GeoKey	68.79 \pm 3.17 (9) [#]	50.70 \pm 1.58 (9) [#]	10.86 \pm 2.65 (9) [#]	9.0
SZM	83.59 \pm 2.78 (8) [#]	64.33 \pm 2.22 (8) [#]	26.34 \pm 5.27 (8) [#]	8.0
GeoKey+SZM	84.18 \pm 2.51 (7) [#]	65.65 \pm 1.85 (7) [#]	27.38 \pm 5.05 (7) [#]	7.0
socre-min	87.16 \pm 6.54 (6) [#]	88.28 \pm 3.49 (4) [#]	32.73 \pm 9.69 (5) [#]	5.0
socre-mean	94.11 \pm 1.88 (3) [#]	81.56 \pm 3.11 (6) [#]	34.90 \pm 9.19 (4) [#]	4.3
socre-QWM	94.15 \pm 1.86 (2) [#]	82.00 \pm 3.10 (5) [#]	32.45 \pm 9.85 (6) [#]	4.3
MF	92.31 \pm 3.26 (5) [#]	94.79 \pm 1.00 (3) [#]	56.10 \pm 10.00 (3) [#]	3.7
QWM	92.55 \pm 3.21 (4) [#]	94.93 \pm 0.96 (2) [#]	57.71 \pm 9.59 (2) [#]	2.7
SE-LRMF	95.24 \pm 1.77 (1)	95.60 \pm 0.88 (1)	71.22 \pm 7.44 (1)	1.0

[#]The p value of paired student-t test to the best method is lower than 0.05 level.

*MR—mean rank (best rank bolded)

Table 6.6: The DI of SE-LRMF and existing iris recognition algorithms (format: mean \pm standard deviation (rank), best performance bolded).

Method	ND0405	CASIAD	UBIRIS2	MR*
NF2	3.221 \pm 0.070 (9) [#]	2.276 \pm 0.044 (10) [#]	1.647 \pm 0.079 (10) [#]	9.7
GeoKey	3.198 \pm 0.067 (10) [#]	2.328 \pm 0.043 (9) [#]	1.695 \pm 0.081 (9) [#]	9.3
SZM	4.267 \pm 0.116 (7) [#]	3.046 \pm 0.068 (7) [#]	2.389 \pm 0.103 (7) [#]	7.0
GeoKey+SZM	4.154 \pm 0.107 (8) [#]	2.990 \pm 0.063 (8) [#]	2.328 \pm 0.098 (8) [#]	8.0
socre-min	5.833 \pm 0.129 (2) [#]	4.322 \pm 0.116 (4) [#]	3.098 \pm 0.151 (5) [#]	3.7
socre-mean	5.001 \pm 0.163 (6) [#]	3.763 \pm 0.108 (6) [#]	3.127 \pm 0.170 (4) [#]	4.7
socre-QWM	5.005 \pm 0.163 (5) [#]	3.778 \pm 0.108 (5) [#]	3.081 \pm 0.152 (6) [#]	5.3
MF	5.481 \pm 0.218 (4) [#]	4.701 \pm 0.131 (3) [#]	3.729 \pm 0.222 (3) [#]	3.3
QWM	5.498 \pm 0.225 (3) [#]	4.712 \pm 0.129 (2) [#]	3.739 \pm 0.219 (2) [#]	2.3
SE-LRMF	6.108 \pm 0.278 (1)	5.158 \pm 0.143 (1)	4.087 \pm 0.205 (1)	1.0

[#]The p value of paired student-t test to the best method is lower than 0.05 level.

*MR—mean rank (best rank bolded)

This subsection investigates four representative noise types: blur, off-angle, occlusion and motion. The corresponding quality measures are calculated using the method in [28, 111]. The comparison is focused between the identification performance of SE-LRMF and score-level fusion methods. For the other cases (comparison with signal-level fusion and other recent iris recognition algorithms in identification and verification; comparison with score-level fusion in verification), it is found that SE-LRMF performs generally better than the compared methods across all the levels of different noise types, on the three datasets. It is consistent with the results in previous subsections where only score-level fusion performs better than SE-LRMF on some datasets in identification.

Rank 1 accuracy is used to represent the identification performance. Fig. 6.10 shows the rank 1 accuracy of all the comparison methods on different levels of each quality measure. It can be seen that, on ND0405 which has the lowest noise level, SE-LRMF performs generally worse than the best score-level fusion method on all the levels of each quality measure; on CASIAD with a higher noise level, SE-LRMF performs worse than the best score-level fusion method on the first few quality levels, but it performs generally better on the last few quality levels; on UBIRIS2 with the highest noise level, SE-LRMF performs generally better than the other methods across all

the quality levels. The result shows that SE-LRMF performs better for the data with a higher noise level.

However, it can be found that, in Fig. 6.10, the trend of performance change is not fully consistent with the noise level. In some cases, the performance has obvious perturbation or increasing when the noise level increases (for example, the focus measure on ND0405 and UBIRIS2). The possible reason is the mixture of noise types. In the three datasets, a capture usually suffers from multiple types of noise. Consequently, the final performance is influenced by the combination of all these noise types, rather than a specific one. Therefore, given that the image sets are grouped based on one quality measure (*i.e.* the level of one noise type), the final performance is possible to be influenced by the variation in the level of other noise types which may also affect each group. Also, more complexity exists since noise types may not be independent with each other. This problem of mixture of noise types can be addressed in future research, either by performing data capturing with the target noise type varied and the others restricted at low levels, or by combining multiple individual quality measures to obtain a quality measure that is better related to the final performance.

Furthermore, it is observed that the performance change is more consistent with the change of motion level. This observation means that the influence of motion on the final performance is more dominant than the other noise types. The possible reason is that motion is also source of other noise types, like blur and off-angle. Therefore, the corresponding quality measure summarises more relevant noise types.

In all, SE-LRMF performs better than the other methods on the data with a higher level in any of the four noise types, despite of the problem of mixture of noise types.

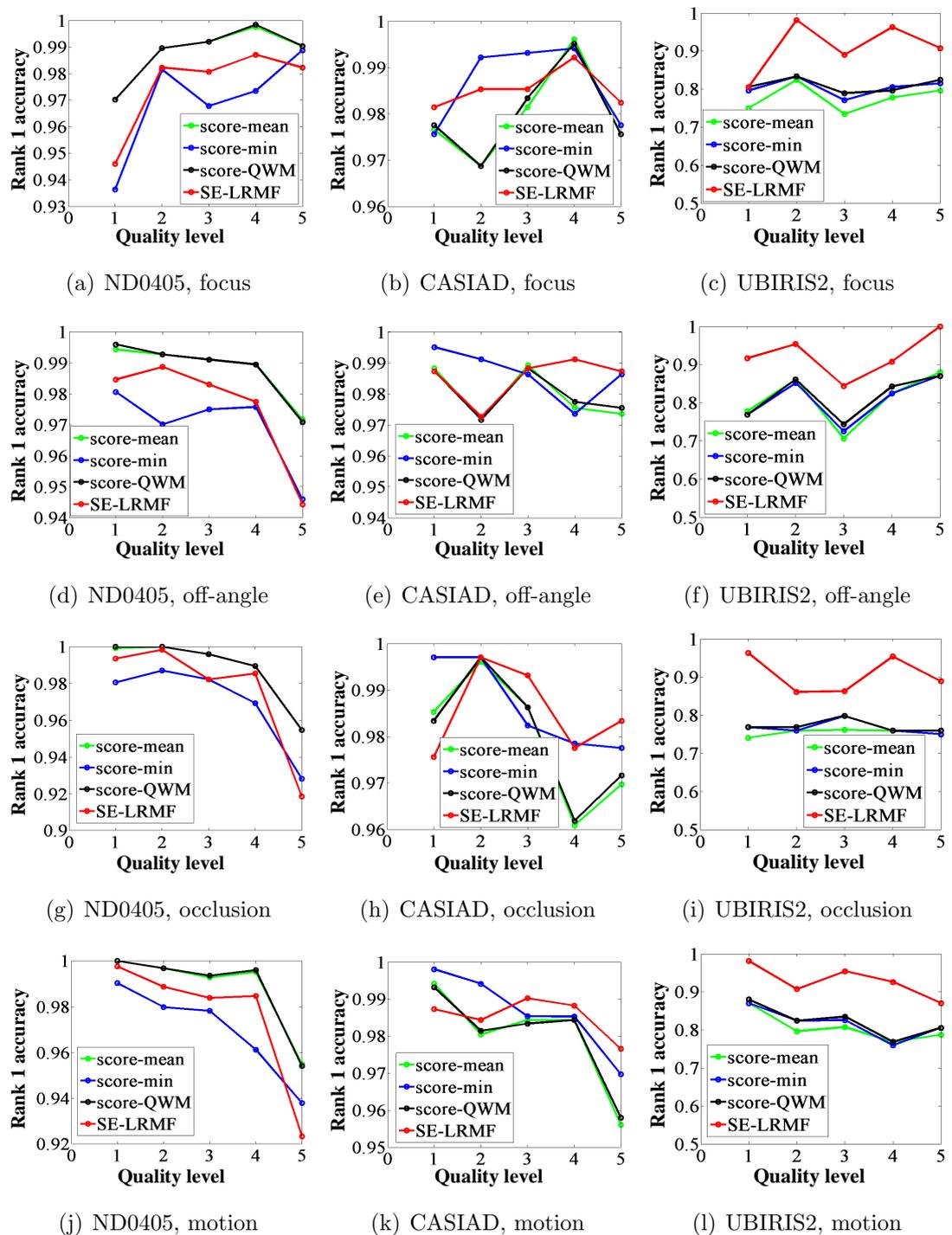


Figure 6.10: The performance of SE-LRMF and score-level fusion methods on varying levels of different quality measures.

6.5 Summary

This chapter proposes a signal-level information fusion method based on low rank approximation for less constrained iris recognition, in order to mitigate the noise in captures induced by less constrained environments. Firstly, a sparse-error low rank matrix factorisation (SE-LRMF) model is proposed to separate noiseless iris pixels and noise pixels in multiple iris captures. Then, the iris pixels and noise pixels are utilised to perform signal-level fusion individually, leading to a low rank mean (LRM) fusion method and an error weighted mean (EWM) fusion method. Finally, the results of LRM and EWM are combined at the code level to produce a final iris code. Experiments are conducted on benchmark iris datasets captured under different wavelengths and distances, with varying noise types and noise levels. The main conclusions are summarised as follows:

- Using SE-LRMF for low rank approximation, the code level combination of LRM and EWM results leads to an improved performance over individual LRM and EWM results.
- In comparison to representative low rank approximation algorithms, SE-LRMF leads to a better and more stable performance in the proposed framework of signal-level fusion. Also, SE-LRMF has the potential to achieve a more robust performance.
- The SE-LRMF based signal-level fusion leads to a significantly improved performance compared to the baseline algorithms without fusion. It also performs better than recently proposed iris recognition algorithms that do not fuse the information from multiple captures.
- The SE-LRMF based signal-level fusion outperforms existing signal-level fusion methods.

-
- Compared to existing score-level fusion methods, the SE-LRMF based signal-level fusion has a competitive and more stable identification performance, and a better verification performance.
 - The SE-LRMF based signal-level fusion performs better for the captures with relatively heavier noise.

Chapter 7

Optimal generation of iris codes

Part of this chapter has been adapted from the journal paper [J10] (published) as in the List of Publications.

7.1 Introduction

Iris recognition is usually performed based on binary features known as iris codes [6, 7, 15, 20, 80]. The binary nature of iris codes brings significant advantage in memory and computational cost, enabling the large scale deployment of iris recognition systems. Therefore, the calculation of binary iris codes from feature values (*e.g.* the result of Gabor transform) is a key step in iris recognition systems.

Most state-of-the-art methods generate iris codes using simple binarisation based on the sign of feature values [6, 7, 15, 20, 80]. Experimentally, this binarisation method achieves a generally promising performance, but, currently, there lacks a deeper insight into this method. An important question one may ask is this: “Is it optimal to produce the iris code by binarisation based on the sign of feature values?”. A similar question is: “How could we find more effective iris codes given feature values?”. Currently, although a large number of feature selection algorithms are investigated to obtain more effective iris codes [145–150], these methods focus on selecting more

valuable bits from iris codes, rather than directly calculating optimal iris codes from feature values. Therefore, the above questions remain open.

This chapter investigates the issue of iris code production from the perspective of optimisation. It is illustrated that the traditional iris code based on the sign of feature values is the solution of an optimisation problem. This optimisation problem seeks iris codes by minimizing the distance between the feature values and iris codes. Based on this illustration, additional objective terms can be applied to this optimisation problem, in order to obtain more effective iris codes. This chapter designs an additional objective term based on the spatial correlation in iris codes. The idea is that iris pixels have inherent spatial correlations [6, 31, 151]; therefore, as a feature representation of iris textures, the bits at different positions of an iris code should also be spatially correlated. Accordingly, this additional objective term explicitly formulates the spatial correlation in iris codes, and it leads to a Markov Random Field (MRF) model to compute iris codes.

Experimental analysis are performed on benchmark datasets captured with varying environments. The experimental results demonstrate that the iris code produced by solving the proposed optimisation problem with the additional objective term achieves a generally improved performance in comparison to an implementation of traditional iris code generation method [15] binarizing feature values based on their signs. The direction of spatial correlation in iris textures as observed in [6, 31, 151] is also experimentally verified in the experiment of this chapter.

The remainder of this chapter is organised as follows. Section 7.2 presents the proposed method to produce iris codes based on optimisation. Section 7.3 reports the result of experimental analysis. Section 7.4 summarises the paper.

7.2 Producing iris codes by optimisation

7.2.1 Illustration of the traditional iris code production from a perspective of optimisation

Let $\mathbf{f} = [f_1, f_2, \dots, f_n]^T \in \mathbb{R}^n$ be a vector consisting of all feature values extracted from an unwrapped iris image. For example, for the most widely used Gabor feature [6, 7], \mathbf{f} is constructed by concatenating the real and imaginary parts in all positions into a vector after applying Gabor transform to the unwrapped iris region; each element in \mathbf{f} is either the real or imaginary part of the Gabor transform result. Let $\mathbf{b} = [b_1, b_2, \dots, b_n]^T$ be a binary vector of the iris code corresponding to \mathbf{f} with n bits. For the i^{th} bit b_i , traditional binarisation method to calculate it is as follows:

$$b_i = \begin{cases} 1 & \text{if } f_i \geq 0 \\ 0 & \text{if } f_i < 0 \end{cases} \quad (7.1)$$

In other words, traditional iris codes are produced based on the sign of feature values.

This subsection demonstrates that the traditional binarisation method can be illustrated as the solution of the following optimisation problem:

$$\arg \min_{\mathbf{b} \in \{-1, 1\}^n} \|\mathbf{b} - \mathbf{f}\|_2^2 \quad (7.2)$$

Although $\mathbf{b} \in \{-1, 1\}^n$ in Eqn. 7.2 is different from the traditional iris code where $\mathbf{b} \in \{0, 1\}^n$, it will be shown that the solution of Eqn. 7.2 is equivalent to the traditional iris code in binary feature domain, and it makes no difference in the Hamming distance given two iris codes.

The objective function in Eqn. 7.2 can be expanded as follows:

$$\|\mathbf{b} - \mathbf{f}\|_2^2 = \|\mathbf{b}\|_2^2 + \|\mathbf{f}\|_2^2 - 2\mathbf{f}^T\mathbf{b} \quad (7.3)$$

Note that $\|\mathbf{f}\|_2^2$ is a constant, and $\|\mathbf{b}\|_2^2 = n$ due to $\mathbf{b} \in \{-1, 1\}^n$. Therefore, Eqn. 7.2 is equivalent to:

$$\arg \min_{\mathbf{b} \in \{-1, 1\}^n} -\mathbf{f}^T\mathbf{b} \quad (7.4)$$

Since $\mathbf{f}^T\mathbf{b} = \sum_i f_i b_i$, the optimisation problem with respect to b_i reduces to:

$$\arg \min_{b_i \in \{-1, 1\}} -f_i b_i \quad (7.5)$$

To minimize Eqn. 7.5, the sign of b_i should be consistent with the sign of f_i , thus the solution is as follows:

$$b_i = \begin{cases} 1 & \text{if } f_i \geq 0 \\ -1 & \text{if } f_i < 0 \end{cases} \quad (7.6)$$

It can be seen that the solution is equivalent to the traditional iris code (Eqn. 7.1) if replacing all the -1 by 0 in the obtained iris code \mathbf{b} , and it does not change the Hamming distance given two iris codes. On the other hand, from the perspective of optimisation, an explanation is that the solution of Eqn. 7.5 is equivalent to the solution of the following problem in binary feature space:

$$\arg \min_{b_i \in \{0, 1\}} -f_i b_i \quad (7.7)$$

Based on the above illustration, the traditional method of iris code production can be explained as follows. Given an unwrapped iris image, the extracted feature values construct a n -dimensional feature vector. In each dimension of the feature vector, two anchor points are set at -1 and 1 . The traditional method of iris code production binarizes a feature value by assigning it to the nearest anchor point at

the corresponding dimension (*i.e.* Eqn. 7.2).

7.2.2 Iris code production using additional objective term

Based on the derivation from Eqn. 7.2 to Eqn. 7.7 in Section 7.2.1, Eqn. 7.2 formulating the traditional iris code production can be rewritten into the following equivalent problem:

$$\arg \min_{\substack{b_i \in \{0,1\} \\ i=1,2,\dots,n}} - \sum_i f_i b_i \quad (7.8)$$

Based on Eqn. 7.8, more complex models can be used to obtain iris codes by adding terms to the objective function of this optimisation problem. It is expected that, by adding proper objective term, more effective iris codes can be obtained.

This subsection exploits spatial correlation in iris codes as the additional objective term. It models the spatial correlation of the bits in different positions of an iris code. Specifically, traditional iris code production method binarizes the feature values in different positions individually (as shown in Eqn. 7.1 and Eqn. 7.6). In other words, it considers each feature value separately. However, as studied in [6, 31, 151], the iris texture has inherent correlations along the radial direction. For example, as pointed out in [6, 31, 151], a furrow or ciliary pattern tends to propagate in the radial direction. Therefore, as a feature representation of iris patterns, the bits in an iris code should be dependent along the vertical direction. Due to the binary nature of iris codes, it is reasonable to assume that such a vertical dependency will lead to a vertical bit-adjacency in an iris code. Accordingly, this subsection designs the additional objective term to exploit the vertical bit-adjacency in an iris code. The additional objective term is defined as follows:

$$\sum_i \sum_{j \in N^i} |b_i - b_j| \quad (7.9)$$

where N^i denotes two immediate vertical neighbors of bit i : one at the top of i and one at the bottom of i . Minimizing Eqn. 7.9 with respect to $b_i \in \{0, 1\}$, $i = 1, 2, \dots, n$ prompts each bit in an iris code to have the same value as its vertical neighbours. Incorporating Eqn. 7.9 into Eqn. 7.8 leads to the following optimisation problem:

$$\arg \min_{\substack{b_i \in \{0,1\} \\ i=1,2,\dots,n}} \sum_i (-f_i)b_i + \alpha \sum_i \sum_{j \in N^i} |b_i - b_j| \quad (7.10)$$

where α is a parameter controlling the trade-off between the influence of the sign of feature values and the vertical adjacency of iris codes. Eqn. 7.10 is a first order Markov Random Field with binary labels. It can be solved via graph cuts [152,153].

Fig. 7.1 shows examples of (1) the iris code produced by Masek's implementation of traditional iris code generation method [15] and (2) the iris code produced using the additional objective term. All the iris codes are obtained from the same unwrapped iris image. In each subfigure, the bottom image which corresponds to the iris code (2) also marks the bits that are different from the iris code produced by Masek's implementation (*i.e.* the changed bits due to the additional objective term). The red colour is used to mark the bits being 1 in the iris code produced by Masek's implementation but being 0 in the current iris code, and the blue colour is used to mark the bits being 0 in the iris code produced by Masek's implementation but being 1 in the current iris code.

It can be seen that, in the iris code produced with the additional objective term (the bottom image in each subfigure), the regions of 1 and 0 are more adjacent, compared to the iris code produced by Masek's implementation of the traditional binarisation method. Some small regions of 1 or 0 surrounded by large regions with a different bit value in the iris code produced by Masek's implementation are assigned the same bit value as their surrounding regions in the iris code produced using the additional objective term, due to the exploited spatial dependency.

Also, it is noted that the visual appearance of iris codes with/without considering

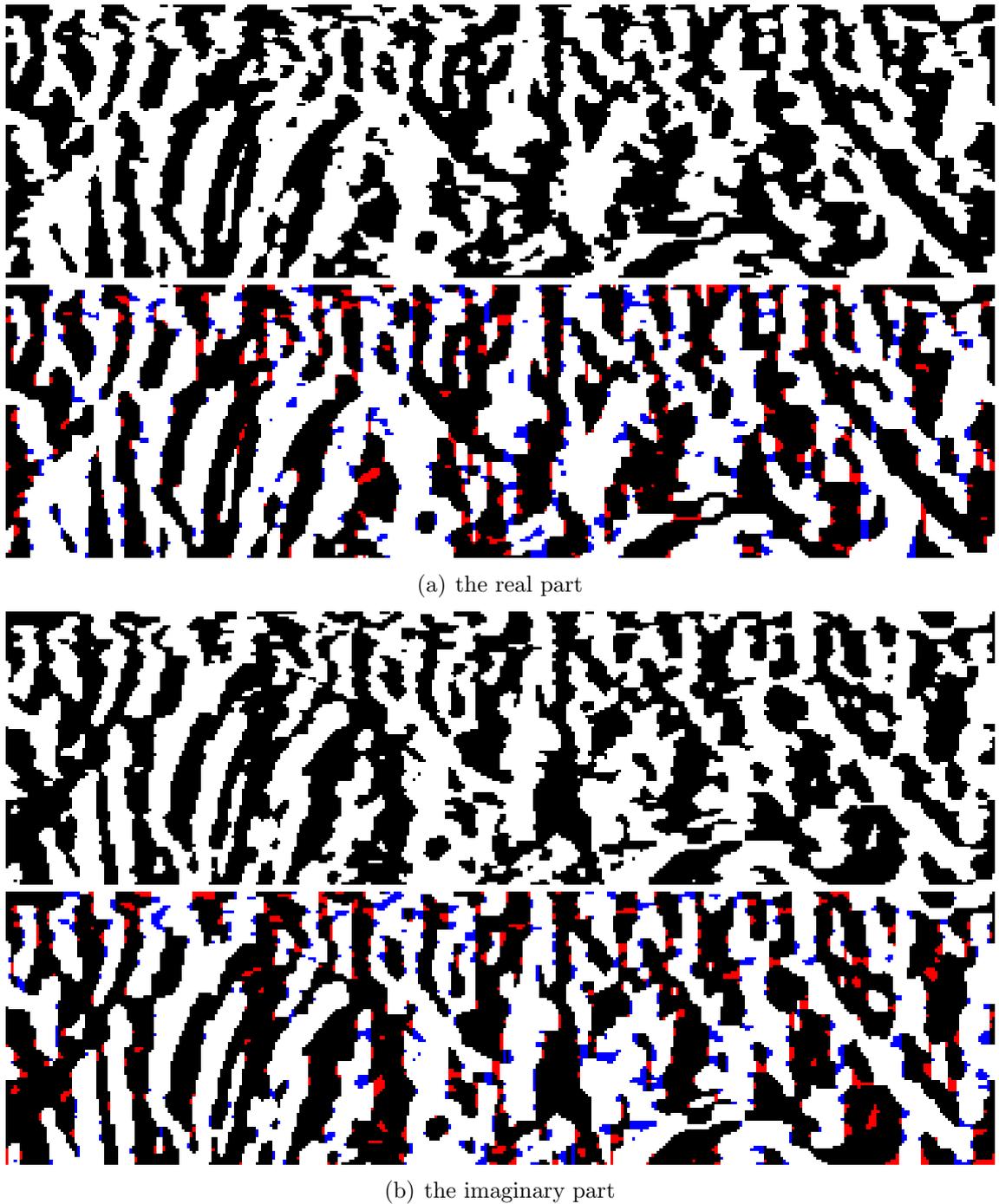


Figure 7.1: Examples of the iris codes produced by different methods. For each sub-figure, the top image shows the iris code produced by Masek’s implementation of the traditional binarisation method, and the bottom image shows the iris code produced by the proposed method. The red colour marks the bits being 1 in the iris code produced by Masek’s implementation but being 0 in the proposed iris code; the blue colour marks the bits being 0 in iris code produced by Masek’s implementation but being 1 in the proposed iris code.

radial correlation shown in Fig. 7.1 are very similar. This is due to the choice of parameter α in Eqn. 7.10. For the example in Fig. 7.1, a relatively small α is chosen (0.3). If a larger α is chosen, there will be larger difference in the visual appearance between the iris code with and without considering radial correlations. However, in this thesis, α is chosen based on the experimental performance (please see Section 7.3.1 for more details). Therefore, with this choice of α , although the iris codes with/without considering radial correlation have similar visual appearance, the iris code with considering radial correlation experimentally achieves better performance experimentally.

7.3 Experiment

This section conducts experimental analysis on the proposed optimisation method to produce iris codes. Firstly, the experimental setting is introduced, including datasets, parameter setting and performance evaluation (Section 7.3.1). Then, to study the effect of the proposed method for iris code production, comparisons are made between the performance of the iris codes produced by an implementation of traditional iris code generation method [15] and the proposed iris codes produced using the additional objective term (Section 7.3.2). After that, an analysis of the spatial correlation of the bits in iris codes is performed based on the proposed additional objective term (Section 7.3.3), aiming to gain a deeper insight into the correlation in iris texture. Finally, the computational load of the proposed method is studied (Section 7.3.4).

7.3.1 Experimental setting

Datasets. The experiments are conducted on four benchmark datasets as introduced in Chapter 3: CASIAT [102], ND0405 [103], CASIAD [102] and UBIRIS2 [19].

Table 7.1: Information and parameter setting of the datasets used in this experiment.

Dataset	Information			Parameter setting		
	Eyes	Images	Wavelength	α	$1/f_0$	σ/f_0
CASIAT	1790	12108	NIR	1.2	22	0.48
ND0405	712	5613	NIR	1.2	35	0.49
CASIAD	284	5037	NIR	1.2	19	0.48
UBIRIS2	171	1000	Visible	0.3	40	0.38

These datasets cover the data captured in varying environments. The information of the data used in this experiment is introduced in Tab. 7.1.

CASIAT is used to represent the iris captures with a satisfactory quality for iris recognition. The images in this dataset have a high overall quality, despite the influence of glasses and specular reflections in some images. To keep CASIAT representing the captures with a satisfactory quality, the images where the iris segmentation fails are eliminated (see the details on iris segmentation below), since such images usually have a higher noise level which causes the iris segmentation failure.

ND0405 represents the iris data with a relatively higher quality. The iris images in this dataset are captured at a wavelength inducing less noise (NIR) and a close distance, hence with good resolution and clear iris texture, but suffering from the noise due to real-world conditions, for example, blurring, specular reflection, rotation, off-angle, *etc.* In this experiment, a subset with the first 8 images of each eye is used, and some images are eliminated from the subset since they have too small eye region.

CASIAD represents the iris data with a relatively lower quality. The images are captured at NIR wavelength but at a distance. As a result, the iris region has a low resolution; the iris texture is visually less clear; the noise in this dataset is heavier than that in ND0405 and CASIAT. This experiment uses the images from both left and right eyes of all the subjects in CASIAD dataset, and some images with too small iris regions are eliminated.

UBIRIS2 represents the iris data with a very low quality. The images are captured

not only at a distance, but also in a wavelength inducing heavy noise (visible). In the experiment, a subset with 1000 images from 171 eyes is used. This subset was released for NICE II contest [18]. This experiment reports the result on both the illumination channel (Y) of YCbCr colour space and the red channel of RGB colour space, since it is experimentally observed that the performance has some variations on these two channels (see the experimental results in Section 7.3.2).

The iris segmentation is performed as follows. For CASIAT dataset which is used to represent the captures with a generally satisfactory quality, Hough transform based segmentation as in [15] is applied to seek two circular boundaries for limbus and pupil. For this dataset, the images with the segmentation failure in limbic or pupillary boundary are manually eliminated. The reason is that the failures of the method in [15] are mainly due to the influence of noise factors, so the images with segmentation failure are eliminated to keep the data representing high quality iris captures. The eyelids, silhouette and reflections are detected using the algorithm in Chapter 4 for the remaining images in CASIAT dataset. For the other three iris datasets (ND0405, CASIAD and UBIRIS2), the algorithm in Chapter 4 is employed for iris segmentation, including the segmentation of limbus and pupil, and the detection of eyelids, reflection and silhouette. The segmentation failures for the other three datasets are corrected manually, so that most noisy captures are preserved in these three datasets, enabling them to represent iris captures with varying noise levels. The size of unwrapped iris image is set to 100×360 . 1-D log-Gabor filter [15] is employed to produce the feature values given an unwrapped iris image.

The gallery and probe images are set for each dataset as follows. For CASIAT, the first image of each eye is used as the gallery image, and the remaining images are used as probe images. For ND0405 dataset, the first 2 images of each eye are used as gallery images, and the rest of the images are used as probes. For CASIAD and UBIRIS2 datasets, the first 5 images of each eye are used as gallery images, and the rest of the images are used as probe images. The number of gallery images is chosen

based on the image quality of each dataset. More gallery images are used for the dataset with lower quality.

Parameter tuning and performance evaluation. Exclusively separated data are used for parameter setting and performance evaluation. For CASIAT dataset, the first 50 eyes are used to tune the parameters, and the remaining 1740 eyes are used for performance evaluation. For ND0405 dataset, the first 20 eyes are used to tune the parameters, and the remaining 692 eyes are used to evaluate the performance. For CASIAD dataset, the first 20 eyes are used to tune the parameters, the remaining 264 eyes are used to evaluate the performance. For UBIRIS2 dataset, the first 19 eyes are used to tune the parameters, and the remaining 152 eyes are used to evaluate the performance.

There are 3 parameters to be determined in the proposed method of iris code production: α in Eqn. 7.10 is the parameter controlling the trade-off between the influence of the sign of feature values and the vertical adjacency of iris codes, $1/f_0$ and σ/f_0 are 1-D log-Gabor parameters (see Eqn. 2.25). Based on the above described data for parameter setting, the setting of these parameters is reported in Tab. 7.1. Note that, for 1-D log-Gabor parameters on ND0405, CASIAD and UBIRIS2 datasets, the setting in Section 6.4.1 is adopted. It is found that the α on colour dataset (UBIRIS2) is lower than that on NIR datasets. A possible reason is that colour iris images contain a relatively higher amount of noise, and using a small α on colour data is able to prevent the bit-adjacency term incorrectly spreading the influence of noise bits in an iris code.

The performance is evaluated in two tasks: identification and verification. The identification performance is evaluated by cumulative match characteristic (CMC), while the verification performance is evaluated by receiver operating characteristic (ROC).

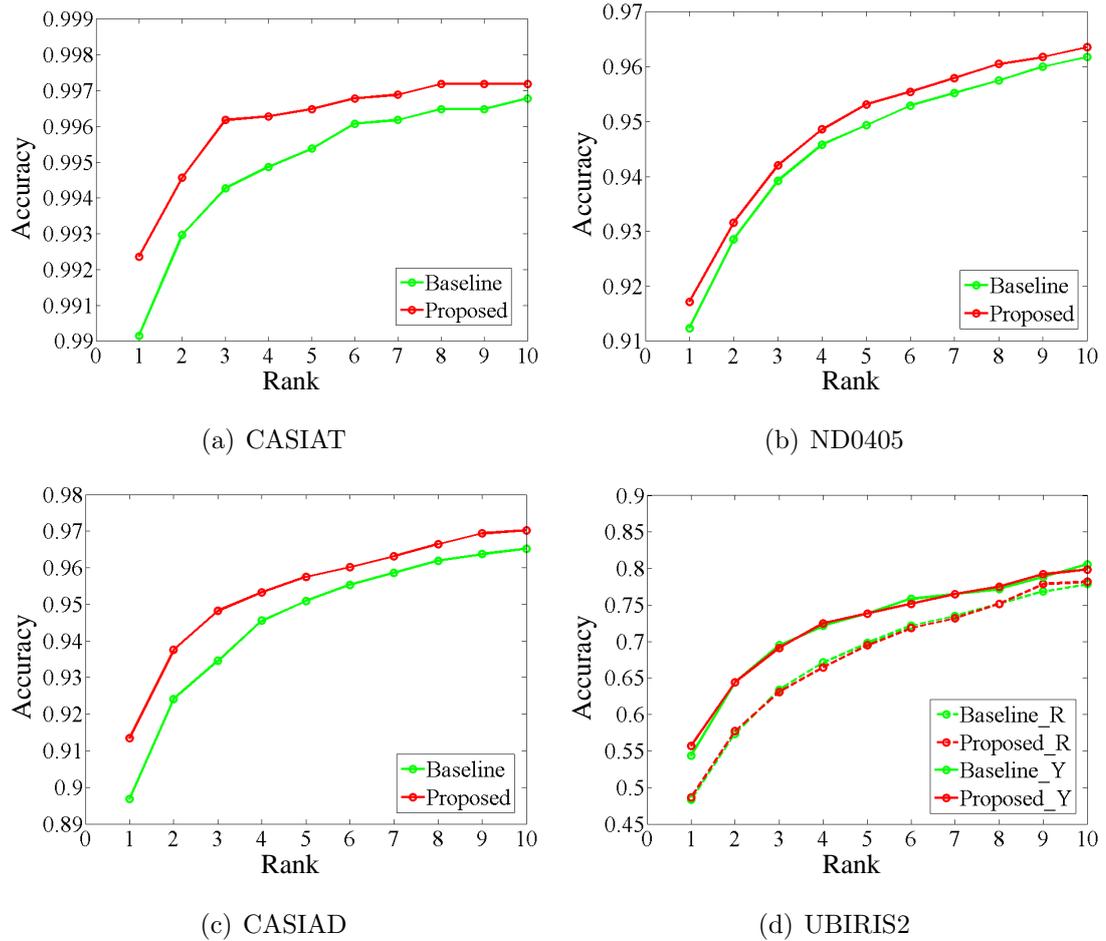


Figure 7.2: Comparison of the CMC curves between the proposed iris code and the iris codes produced by Masek’s implementation of traditional iris code generation method.

7.3.2 Comparison with iris codes produced by an implementation of traditional iris code generation method

This subsection compares between the iris codes produced by Masek’s implementation of traditional iris code generation method [15] (referred to as Baseline) and the proposed iris code produced with the additional objective term (referred to as Proposed). Fig. 7.2 and Fig. 7.3 show the CMC and ROC curves of all the comparison methods on all the datasets, respectively.

Moreover, rank 1 recognition accuracy (R1RA) is calculated to represent the identification performance; equal error rate (EER) and genuine acceptance rate (GAR)

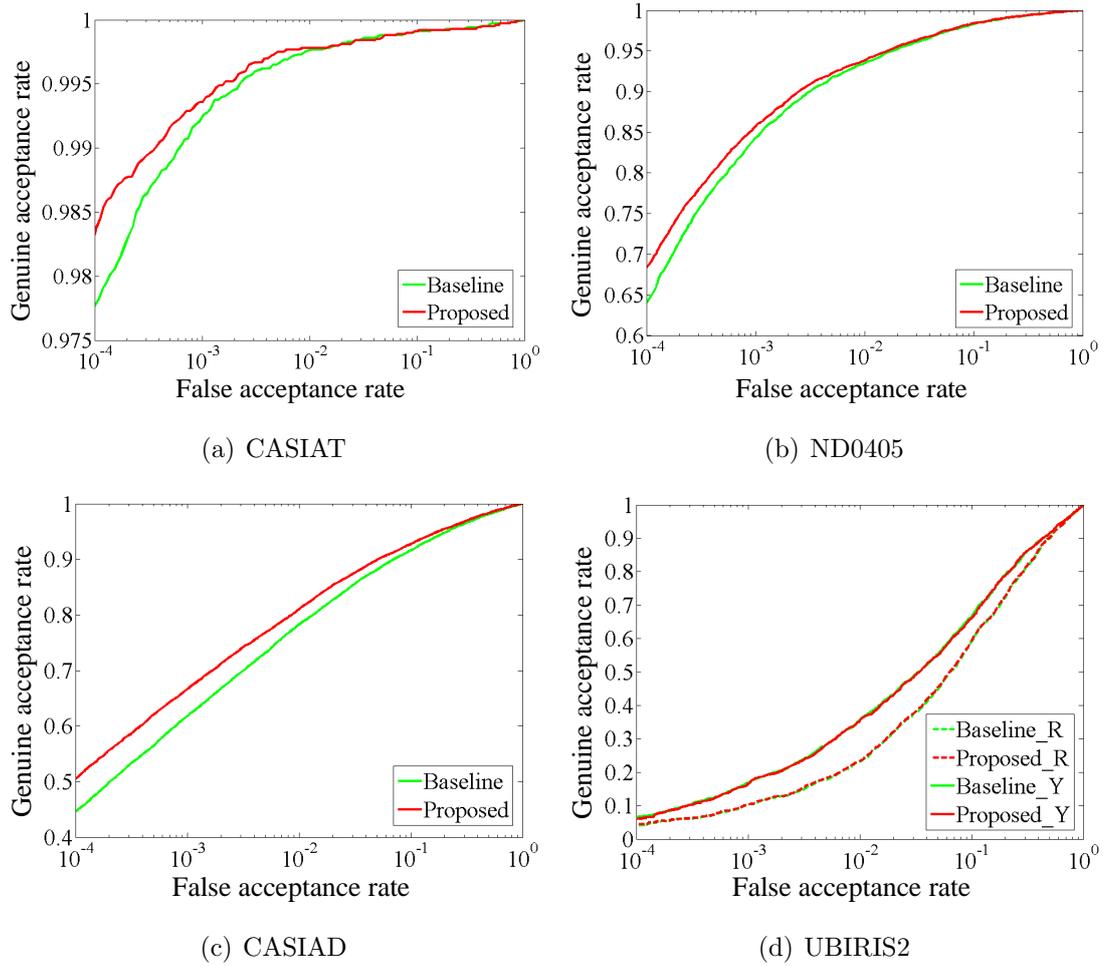


Figure 7.3: Comparison of the ROC curves between the proposed iris code and the iris code produced by Masek's implementation of traditional iris code generation method.

Table 7.2: The R1RA for comparison between the proposed iris code and the iris code produced by Masek's implementation of traditional iris code generation method, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Baseline	99.18 \pm 0.14	92.74 \pm 0.90	91.45 \pm 0.72	-
Proposed	99.37 \pm 0.12[#]	93.14 \pm 0.86[#]	92.91 \pm 0.66[#]	-
Baseline_R	-	-	-	50.83 \pm 2.59
Proposed_R	-	-	-	51.18 \pm 2.55[#]
Baseline_Y	-	-	-	56.90 \pm 2.60
Proposed_Y	-	-	-	58.03 \pm 2.55[#]

[#]The p value of paired student-t test is lower than 0.05 level.

Table 7.3: The EER for comparison between the proposed iris code and the iris code produced by Masek’s implementation of traditional iris code generation method, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Baseline	0.38 \pm 0.06	3.65 \pm 0.32	8.86 \pm 0.42	-
Proposed	0.33 \pm 0.05[#]	3.47 \pm 0.31[#]	8.07 \pm 0.40[#]	-
Baseline_R	-	-	-	23.61 \pm 0.75
Proposed_R	-	-	-	23.42 \pm 0.72[#]
Baseline_Y	-	-	-	20.52 \pm 0.90
Proposed_Y	-	-	-	20.67 \pm 0.87 [#]

[#]The p value of paired student-t test is lower than 0.05 level.

Table 7.4: The GAR@FAR= 0.01% for comparison between the proposed iris code and the iris code produced by Masek’s implementation of traditional iris code generation method, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Baseline	97.76 \pm 0.22	64.00 \pm 5.47	44.64 \pm 2.54	-
Proposed	98.34 \pm 0.22[#]	68.16 \pm 4.80[#]	50.66 \pm 2.22[#]	-
Baseline_R	-	-	-	4.15 \pm 0.69
Proposed_R	-	-	-	4.29 \pm 0.74[#]
Baseline_Y	-	-	-	6.39 \pm 0.97
Proposed_Y	-	-	-	5.93 \pm 0.93 [#]

[#]The p value of paired student-t test is lower than 0.05 level.

when false acceptance rate (FAR) is 0.01% (GAR@FAR= 0.01%) are calculated to represent the verification performance; decidability index (DI) [6] is calculated to represent the intra-class and inter-class separability of matching scores. A significance analysis is performed based on the above measures using paired student-t test. For each dataset, 100 subsets are randomly selected from the performance evaluation set; the subjects included in each subset are randomly selected; R1RA, EER, GAR@FAR= 0.01% and DI are calculated for each subset, and paired student-t test is performed between the performance of the iris code produced by Masek’s implementation of traditional iris code generation method and the proposed iris code to study the statistical significance. The mean and standard deviations of R1RA, EER, GAR@FAR= 0.01% and DI, together with the result of significance analysis, are reported in Tab. 7.2, Tab. 7.3, Tab. 7.4 and Tab. 7.5, respectively.

Table 7.5: The DI for comparison between the proposed iris code and the iris code produced by Masek’s implementation of traditional iris code generation method, with the result of significance test (best performance bolded, format: mean \pm standard deviation).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Baseline	4.334 \pm 0.047	3.435 \pm 0.066	2.439 \pm 0.044	-
Proposed	4.639 \pm 0.050[#]	3.504 \pm 0.069[#]	2.601 \pm 0.048[#]	-
Baseline_R	-	-	-	1.367 \pm 0.034
Proposed_R	-	-	-	1.373 \pm 0.034[#]
Baseline_Y	-	-	-	1.531 \pm 0.050
Proposed_Y	-	-	-	1.530 \pm 0.050 [#]

[#]The p value of paired student-t test is lower than 0.05 level.

Note that, in Fig. 7.2(d), Fig. 7.3(d), Tab. 7.2, Tab. 7.3, Tab. 7.4 and Tab. 7.5, Baseline_R and Proposed_R are used to refer to the performance of iris codes on the red channel of RGB colour space of UBIRIS2 dataset, and Baseline_Y and Proposed_Y correspond to the illumination channel of YCbCr colour space of UBIRIS2 dataset.

It can be seen that the performance varies on NIR datasets (CASIAT, ND0405, CASIAD) and colour dataset (UBIRIS2). Considering the result on NIR datasets, it is found that the proposed iris code generally achieve better CMC and ROC performance, in comparison to the iris code produced by Masek’s implementation of traditional iris code generation method. This result demonstrates that more effective iris codes can be obtained by solving an optimisation problem with an additional objective term modeling the spatial correlation of the bits in iris codes.

On the other hand, for the colour dataset (UBIRIS2), it is found that the proposed method leads to little improvement compared to the iris codes produced by Masek’s implementation of traditional iris code generation method. A possible reason is that the colour captures in this dataset include too heavy noise. The heavy noise induces a large amount of highly unreliable feature values. Consequently, the spatial correlation exploited based on the proposed additional objective term is less reliable.

Finally, it can be seen that, for UBIRIS2 dataset, performance variations are observed on different channels. Specifically, as shown in Tab. 7.2, Tab. 7.3, Tab. 7.4

and Tab. 7.5, the proposed method leads to very slight improvement on R1RA, EER, GAR@FAR= 0.01% and DI on the red channel of RGB colour space, while it only improves R1RA on the illumination channel of YCbCr colour space. A possible reason of this observation is that different channels include different noise components; since the proposed method does not explicitly model the noise, the influence of different noise components on the proposed method is less controllable; thus it leads to performance variation on different channels on UBIRIS2 dataset.

In all, the experimental results in this subsection show that: (1) when the iris data is less noisy, the proposed iris code generated by solving an optimisation problem is able to achieve generally improved identification and verification performance, compared to the iris code produced by Masek's implementation of traditional iris code generation method; (2) the highest noise level where (1) applies is distant NIR captures in this experiment; (3) for the iris data with heavier noise (colour captures in the experiment), the iris code produced by the proposed method performs similarly to the iris code produced by Masek's implementation of traditional iris code generation method.

7.3.3 Analysis of the spatial correlation in iris textures

The radial correlation of iris textures has been demonstrated in several research such as [6, 31, 151]. This subsection studies this correlation from the perspective of iris code optimisation, using the proposed iris code generation method with the additional objective term. The analysis is performed by varying the neighbour type in the second term of Eqn. 7.10, and investigating the performance of obtained iris codes. Intuitively, the best performance will be achieved by a neighbour type that is closest to the inherent spatial correlation of iris textures. The reason is that such neighbour type correctly models the inherent spatial dependency in iris textures. In contrast, a neighbour type that is different from the inherent spatial correlation

will introduce incorrect dependency between the bits that should be independent; it will harm the randomness of iris codes, leading to degraded performance. In other words, the iris code with the best performance can be obtained only by using a neighbour type closest to the inherent spatial dependency of iris textures, since it expresses the inherent spatial correlation in iris textures without influencing the inherent randomness part of iris codes. This experiment tests three basic neighbour types, vertical neighbour (including 1 immediate top pixel and 1 immediate bottom pixel to the centre pixel, referred to as Proposed_V), horizontal neighbour (including 1 immediate left pixel and 1 immediate right pixel to the centre pixel, referred to as Proposed_H) and cross neighbour (combination of vertical and horizontal neighbours, referred to as Proposed_C).

Fig. 7.4 and Fig. 7.5 show the CMC and ROC curves of the iris codes produced by the proposed method using the three neighbour types on all the datasets. Also, similarly to Section 7.3.2, Tab. 7.6, Tab. 7.7, Tab. 7.8 and Tab. 7.9 report respectively the R1RA, EER, GAR@FAR= 0.01% and DI of all the methods on all the datasets, together with the result of paired student-t test between the best neighbour type and other neighbour types. Note that, in Fig. 7.4(d), Fig. 7.5(d), Tab. 7.6, Tab. 7.7, Tab. 7.8 and Tab. 7.9, Proposed_R_H, Proposed_R_C and Proposed_R_V are used to refer to the performance of proposed iris codes on the red channel of RGB colour space of UBIRIS2 dataset; Proposed_Y_H, Proposed_Y_C and Proposed_Y_V correspond to the illumination channel of YCbCr colour space of UBIRIS2 dataset.

It is found that, on NIR datasets, the iris code obtained with the vertical neighbour consistently outperforms the iris code obtained using the other two neighbour types. As for the colour dataset, it can be found that the vertical neighbour has slightly better performance than the other two neighbour types on the illumination channel of YCbCr colour space, while, on the red channel of RGB colour space, the performance of vertical neighbour is the best in EER, GAR@FAR= 0.01%. In all, it can be concluded that the iris code produced using the vertical neighbour generally

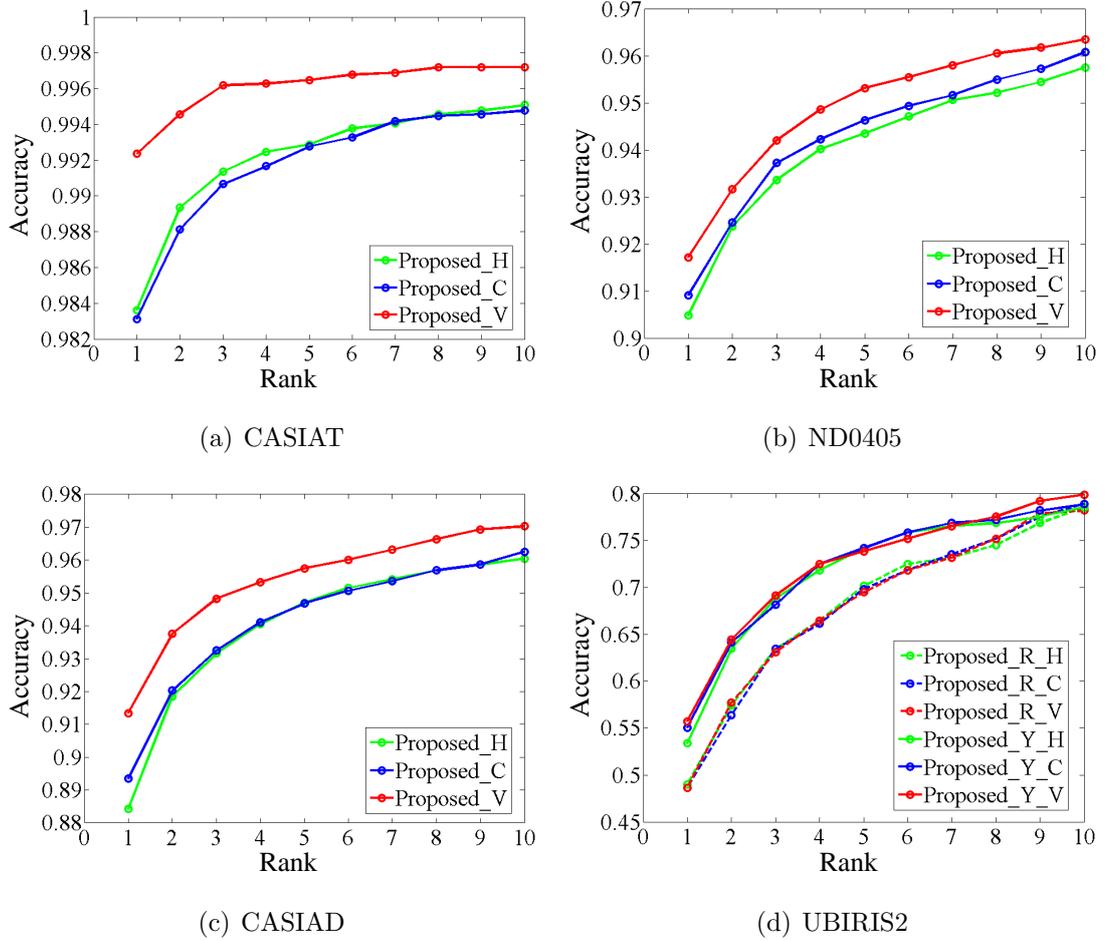


Figure 7.4: CMC curves of the proposed iris code for the analysis of spatial correlation in iris textures.

Table 7.6: The R1RA for the analysis of spatial correlation in iris textures, together with the results of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Proposed_H	98.65 \pm 0.15 [#]	92.10 \pm 0.90 [#]	90.60 \pm 0.72 [#]	-
Proposed_C	98.59 \pm 0.17 [#]	92.42 \pm 0.93 [#]	91.16 \pm 0.75 [#]	-
Proposed_V	99.37 \pm 0.12	93.14 \pm 0.86	92.91 \pm 0.66	-
Proposed_R_H	-	-	-	51.34 \pm 2.52
Proposed_R_C	-	-	-	50.98 \pm 2.47 [#]
Proposed_R_V	-	-	-	51.18 \pm 2.55 [#]
Proposed_Y_H	-	-	-	55.91 \pm 2.60 [#]
Proposed_Y_C	-	-	-	57.30 \pm 2.38 [#]
Proposed_Y_V	-	-	-	58.03 \pm 2.55

[#]The p value of paired student-t test is lower than 0.05 level.

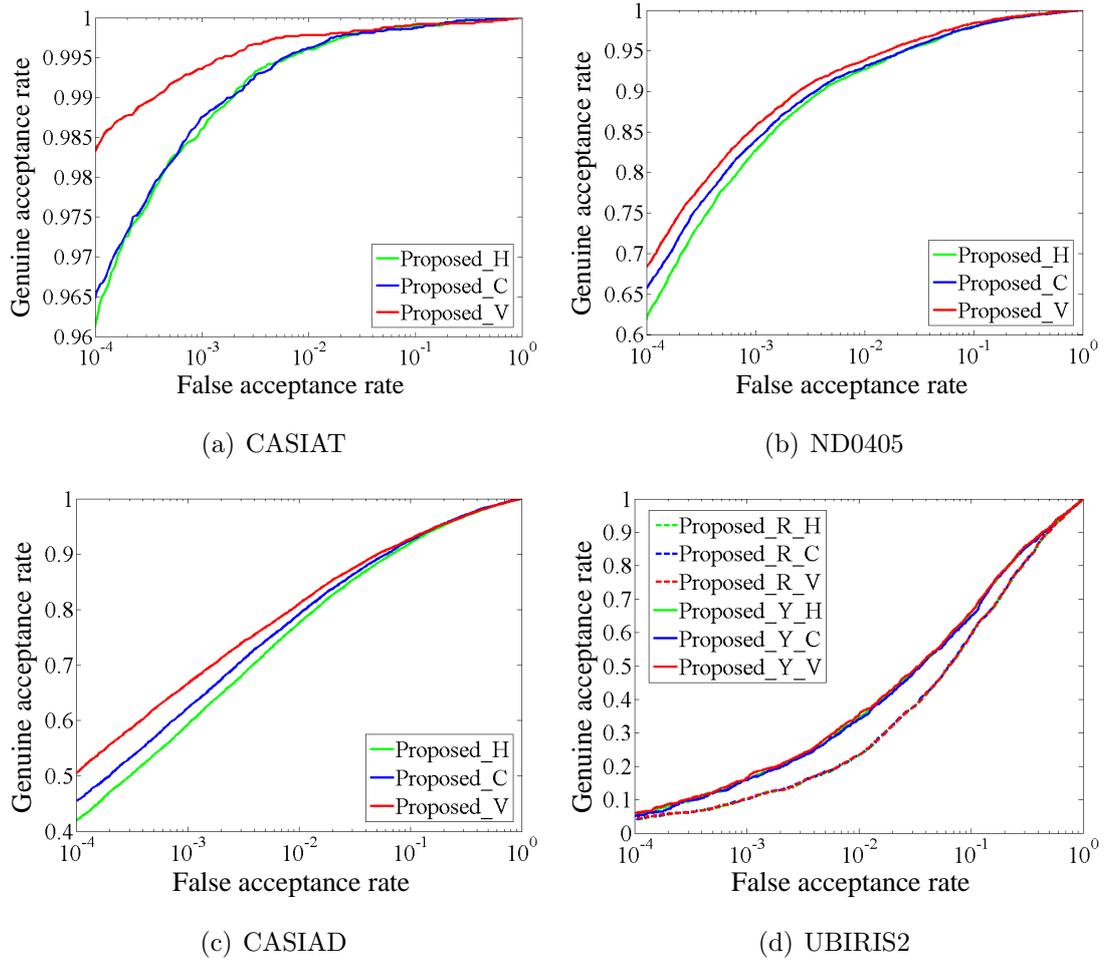


Figure 7.5: ROC curves of the proposed iris code for the analysis of spatial correlation in iris textures.

Table 7.7: The EER for the analysis of spatial correlation in iris textures, together with the results of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Proposed_H	0.54 \pm 0.07 [#]	4.02 \pm 0.33 [#]	8.69 \pm 0.43 [#]	-
Proposed_C	0.54 \pm 0.06 [#]	3.89 \pm 0.31 [#]	8.28 \pm 0.42 [#]	-
Proposed_V	0.33 \pm 0.05	3.47 \pm 0.31	8.07 \pm 0.40	-
Proposed_R_H	-	-	-	23.64 \pm 0.69 [#]
Proposed_R_C	-	-	-	23.50 \pm 0.71 [#]
Proposed_R_V	-	-	-	23.42 \pm 0.72
Proposed_Y_H	-	-	-	20.71 \pm 0.83 [#]
Proposed_Y_C	-	-	-	20.76 \pm 0.80 [#]
Proposed_Y_V	-	-	-	20.67 \pm 0.87

[#]The p value of paired student-t test is lower than 0.05 level.

Table 7.8: The GAR@FAR= 0.01% for the analysis of spatial correlation in iris textures, together with the results of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Proposed_H	96.15 \pm 0.38 [#]	61.93 \pm 5.30 [#]	42.02 \pm 2.34 [#]	-
Proposed_C	96.47 \pm 0.31 [#]	65.35 \pm 4.74 [#]	45.51 \pm 1.89 [#]	-
Proposed_V	98.34 \pm 0.22	68.16 \pm 4.80	50.66 \pm 2.22	-
Proposed_R_H	-	-	-	4.15 \pm 0.69 [#]
Proposed_R_C	-	-	-	4.17 \pm 0.76 [#]
Proposed_R_V	-	-	-	4.29 \pm 0.74
Proposed_Y_H	-	-	-	5.39 \pm 0.87 [#]
Proposed_Y_C	-	-	-	5.20 \pm 0.73 [#]
Proposed_Y_V	-	-	-	5.93 \pm 0.93

[#]The p value of paired student-t test is lower than 0.05 level.

Table 7.9: The DI for the analysis of spatial correlation in iris textures, together with the results of significance test (best performance bolded, format: mean \pm standard deviation).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Proposed_H	4.243 \pm 0.044 [#]	3.397 \pm 0.067 [#]	2.534 \pm 0.047 [#]	-
Proposed_C	4.426 \pm 0.044 [#]	3.451 \pm 0.070 [#]	2.651 \pm 0.052	-
Proposed_V	4.639 \pm 0.050	3.504 \pm 0.069	2.601 \pm 0.048 [#]	-
Proposed_R_H	-	-	-	1.370 \pm 0.035 [#]
Proposed_R_C	-	-	-	1.375 \pm 0.035
Proposed_R_V	-	-	-	1.373 \pm 0.034 [#]
Proposed_Y_H	-	-	-	1.523 \pm 0.050 [#]
Proposed_Y_C	-	-	-	1.515 \pm 0.050 [#]
Proposed_Y_V	-	-	-	1.530 \pm 0.050

[#]The p value of paired student-t test is lower than 0.05 level.

achieves the best performance among the three neighbour types. This observation experimentally illustrates that the iris texture has a general spatial correlation along the radial direction, from the perspective of iris code optimisation. It is consistent with the demonstrations in previous research such as [6, 31, 151].

7.3.4 Analysis on the computational cost

The computational cost is a critical factor in real applications. It is expected that the proposed method has higher computational cost than the traditional binarisation method. This is because adding the objective term leads to a more complex

Table 7.10: Computational cost of each method to generate iris codes (ms).

Method	CASIAT	ND0405	CASIAD	UBIRIS2
Baseline	0.60	0.56	0.56	-
Proposed	47.8	46.0	47.9	-
Baseline_R	-	-	-	0.35
Proposed_R	-	-	-	42.1
Baseline_Y	-	-	-	0.53
Proposed_Y	-	-	-	39.0

#The p value of paired student-t test is lower than 0.05 level.

algorithm than simple binarisation, and it induces more computational load.

This subsection studies the additional computational cost of the proposed method. It reports the computational cost of Masek’s implementation of traditional iris code generation method [15] using simple binarisation (referred to as Baseline) and the proposed iris code generation method by solving an optimisation problem (referred to as Proposed). The computational cost is estimated by measuring the elapsed time to produce an iris code given a vector of feature values. The measure is based on a Matlab implementation of both methods, running on a desktop with Intel i5-3470 quad-core 3.20GHz CPU, 16GB RAM, Windows 7 64bit system and Matlab 2013a 64bit. Tab. 7.10 reports the mean elapsed time of all the methods to produce an iris code on all four datasets. Similarly to Section 7.3.2 and Section 7.3.3, in Tab. 7.10, Baseline_R and Proposed_R are used to refer to the performance on the red channel of RGB colour space of UBIRIS2 dataset, and Baseline_Y and Proposed_Y correspond to the illumination channel of YCbCr colour space of UBIRIS2 dataset.

It can be seen that, as expected, the computational cost of the proposed method is higher than the traditional binarisation method. However, the highest elapsed time to compute an iris code using the proposed method is 47.8ms in the experimental setting of this subsection. This speed is considered acceptable in real applications, because the iris code production is performed only once for each iris capture, and a delay of around 47ms is practically very small.

7.4 Summary

This chapter investigates the issue of iris code generation from the perspective of optimisation. It is demonstrated that the traditional iris code generation method can be expressed as the solution of an optimisation problem. Furthermore, this chapter proposes to apply additional term to the objective function of this optimisation problem, in order to produce more effective iris codes. An additional objective term is designed to model the spatial correlation of the bits in iris codes. The experimental results on benchmark datasets demonstrate that the iris code produced by the proposed method leads to a generally improved performance in comparison to the iris code produced by an implementation of traditional iris code generation method [15], and the computational cost is acceptable in real applications. The experimental analysis also provides deeper insights into the characteristics of iris textures and iris codes. Future work may focus on: (1) designing an objective term to model noise, in order to achieve more robust performance on heavily degraded captures; (2) examining the proposed method on more datasets covering more data variations, in order to investigate the stability and robustness of the proposed method on more varying data.

Chapter 8

An iris weight map method exploiting bit stability and discriminability

Part of this chapter has been adapted from the journal paper [J9] (accepted, in press) and the conference paper [C5] (published) as in List of Publications.

8.1 Introduction

This chapter proposes and investigates a novel iris weight map method for iris matching stage to improve less constrained iris recognition. The proposed iris weight map considers both intra-class bit stability and inter-class bit discriminability of iris codes. It is the combination of a stability map and a discriminability map. The stability map models the intra-class bit stability in iris codes to improve the intra-class matching. It assigns more weight to the bits that have values more consistent with their noiseless and stable estimates obtained using a low rank approximation from a set of noisy training images. The idea is that a bit is more robust if its value is more consistent with its noiseless and stable estimate. Specifically, to calculate the stabil-

ity map, low rank approximation is firstly used to estimate the potentially noiseless and stable iris structures (named meta-image) for each user from the training data; then, iris codes are extracted from the meta-images (*i.e.* meta-codes). Meta-codes are viewed as the estimate of the iris code extracted from noiseless and stable iris structures. Finally, the stability map is computed based on the consistency between the iris codes extracted from the training data and meta-codes. It emphasises the bits which are highly consistent with meta-codes.

On the other hand, the discriminability map expresses the inter-class bit discriminability to enhance the inter-class separation. It emphasises the bits with more discriminative power in iris codes. The discriminability map is computed using a 1-to-N strategy. For each user, a discriminability map is obtained by comparing the training iris codes between this user and all the other users.

The final iris weight map is the combination of stability map and discriminability map. It values the bits that are both highly stable within each class and highly discriminative between classes for iris matching.

Experiments are conducted on both single-sensor and cross-sensor iris datasets. The experimental results show that the proposed iris weight map is applicable to both single-sensor and cross-sensor iris recognition scenarios. It achieves a generally improved performance compared to existing iris weight map methods in both scenarios.

The remainder of this chapter is organised as follows. Section 8.2 describes the proposed method to calculate the iris weight map. Section 8.3 reports the experimental results. Section 8.4 summarises this chapter.

8.2 Methodology

This section presents the proposed method to compute the new iris weight map. It begins with the construction of meta-images and meta-codes. Then, it describes the

method to compute the stability map, the discriminability map, and the final iris weight map. Finally, it introduces the iris matching strategy given the iris weight map.

8.2.1 Computation of meta-images and meta-codes

Challenges for iris recognition under less constrained environment are the noise and degradations, and cross-sensor variations in iris captures. To alleviate these problems, this subsection computes potentially noiseless and stable structures as meta-images, and uses the iris codes extracted from these meta-images as meta-codes to provide an estimate of potentially noiseless and stable iris codes.

The meta-images are computed using low rank approximation [136, 137]. The basic assumption is that, given multiple noisy iris captures of one user, the potentially noiseless and stable iris structures lie in a low rank subspace, due to the inherent stability of iris patterns. With this assumption, the meta-image is computed for each user by seeking a low rank approximation of the training iris images.

Assume there are M training iris images for an arbitrary user. Let $\mathbf{I} = [\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_M] \in \mathbb{R}^{N \times M}$ be the matrix of all the training images of this user with $\mathbf{I}_i \in \mathbb{R}^N$ being a column vector of the i^{th} training image. The low rank approximation of \mathbf{I} is sought via ℓ_1 low rank matrix factorisation model [137]:

$$\arg \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{I} - \mathbf{UV}^T\|_1 \quad (8.1)$$

where $\mathbf{U} \in \mathbb{R}^{N \times r}$ and $\mathbf{V} \in \mathbb{R}^{M \times r}$ form the factorisation of \mathbf{I} ; $r \ll m, n$ is a parameter controlling the rank of approximation. Eqn. 8.1 can be solved by cyclic weight median filtering algorithm [137]. After \mathbf{U} and \mathbf{V} are obtained, the low rank approximation of \mathbf{I} can be calculated by \mathbf{UV}^T . Each column in \mathbf{UV}^T corresponds to the low rank approximation of one training image in \mathbf{I} . The meta-image for this

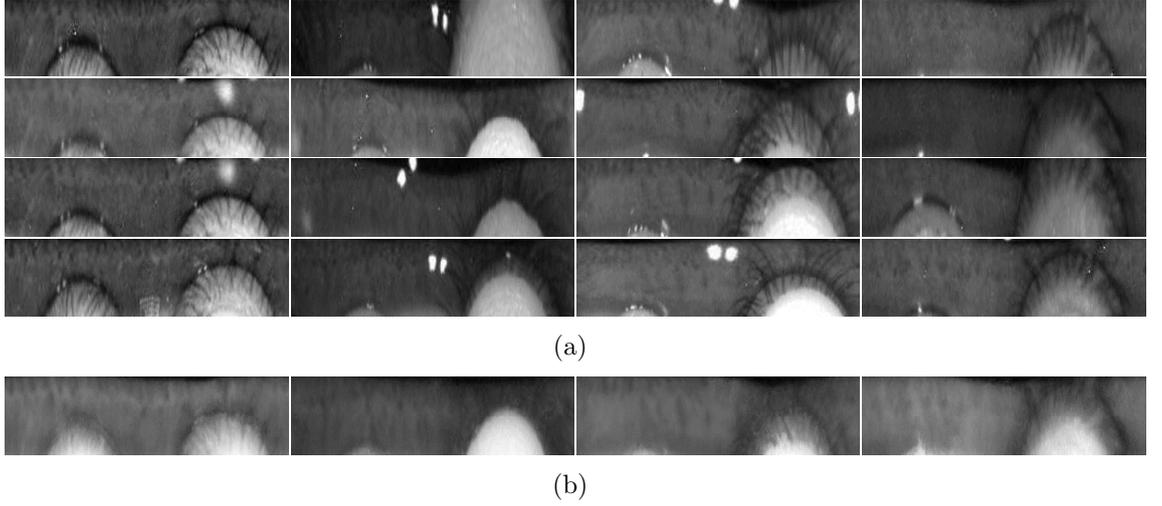


Figure 8.1: Examples of meta-images: (a) example training images; images in the same column are from the same eye; (b) the estimated meta-images correspond to the eye in each column of (a).

user is computed as the mean of the low rank approximation of all the training images in \mathbf{I} . Iris codes are extracted from the meta-image as meta-code. Some examples of meta-images are shown in Fig. 8.1. It can be seen that in meta-images, most of the noise (like specular reflections) in the training images are suppressed, and the stable structure is preserved. Note that some eyelids remain in the meta-images. It is because some regions in training images are always occupied by eyelids and it lacks the noiseless information for these regions.

8.2.2 The stability map

Suppose the iris code of each user has n bits. Denote the stability map of an arbitrary user by $\mathbf{s} = [s_1, s_2, \dots, s_n]$. The stability of the i^{th} bit, s_i , is modelled by:

$$s_i = f_s(p_i^2 + q_i^2) \quad (8.2)$$

In Eqn. 8.2, p_i and q_i denote the probability that the i^{th} bit of this user's iris code equals to 1 and 0, respectively; p_i and q_i satisfy $p_i + q_i = 1$; $f_s : \mathbb{R} \rightarrow \mathbb{R}$ is a function modelling the relationship between $p_i^2 + q_i^2$ and s_i . Eqn. 8.2 is a probability based

model. $p_i^2 + q_i^2$ calculates the probability that the i^{th} bit matches given two iris codes of one user (assuming the two iris codes are independent), and s_i is modelled as a function of $p_i^2 + q_i^2$.

To design $f_s(x)$, the relationship between $p_i^2 + q_i^2$ and s_i needs to be analysed. Intuitively, s_i should be monotonically increasing with $p_i^2 + q_i^2$. Considering that $p_i^2 + q_i^2 = 1 - \sigma_i$ where σ_i is the variance, a higher value of $p_i^2 + q_i^2$ means that the i^{th} bit has a lower variance in the iris codes of the user, hence this bit is more stable. In contrast, a lower $p_i^2 + q_i^2$ means that the i^{th} bit has a higher variance, so it is less stable. Based on the above analysis, two models are investigated to build the relationship between $p_i^2 + q_i^2$ and s_i : a linear model and a non-linear model.

The linear model is motivated by [98]. It is formulated as follows:

$$f_s(x) = 2x - 1 \quad (8.3)$$

It is obvious that the value of $p_i^2 + q_i^2$ is between 0.5 and 1. Eqn. 8.3 uses a linear mapping to normalise this value to between 0 and 1 as the bit stability.

The non-linear model is inspired by [22]. It is formulated as follows:

$$f_s(x) = (2x - 1)^c \quad (8.4)$$

where scalar c is the adaptive crest factor used as a measure of the quality of stability maps [22]. For an arbitrary user, let $\tilde{p}_i = p_i - 0.5$, let $t_i = 2(p_i^2 + q_i^2) - 1 = 4\tilde{p}_i^2$, and let $\mu = \frac{1}{n} \sum_{j=1}^n t_j$. c is determined as follows:

$$c = \begin{cases} \frac{1}{\mu} & \text{if } \mu > 0 \\ 1 & \text{if } \mu = 0 \end{cases} \quad (8.5)$$

Therefore, from Eqn. 8.4, it follows that $s_i = 4\tilde{p}_i^{2c}$. The non-linear model calculates

the bit stability by applying the adaptive crest factor c as the power to the result of the linear model. The power c is reciprocal of the mean stability obtained by the linear model when $\mu \neq 0$. Compared to the linear model, the non-linear model essentially imposes a larger penalty on the highly unstable bits.

For both linear and non-linear models, the key issue is the computation of p_i and q_i . Accordingly, it needs to analyse the influence of p_i and q_i on the final stability. The influence of p_i and q_i on the final stability estimate can be illustrated based on the level of certainty. It is obvious that $p_i^2 + q_i^2$ reaches its minimum (0.5) if and only if $p_i = q_i = 0.5$ (*i.e.* the lowest certainty), and it leads to a minimum stability of 0 in both linear and non-linear models. On the other hand, $p_i^2 + q_i^2$ reaches its maximum (1) when either p_i or q_i is 1 (*i.e.* the highest certainty), and it leads to a maximum stability of 1 in both models. The two models assign a higher stability to the i^{th} bit, if p_i and q_i are more distant from 0.5 (*i.e.* the higher certainty).

Based on the above analysis, a two-step scheme is proposed to estimate p_i and q_i using the consistency between training iris codes and meta-codes (noiseless and stable information). The first step calculates the intra-class bit distribution. The second step uses the intra-class bit distribution and the meta-code to calculate p_i and q_i . The idea is to improve the obtained stability if the intra-class distribution of a bit is consistent with the noiseless and stable estimate of this bit (*i.e.* meta-code), otherwise a penalty is applied to the stability. This scheme aims to achieve the improvement/penalty by shifting the probabilities of intra-class distribution away/towards 0.5, according to the above analysis based on certainty.

For one specific user, let m_i be the times the i^{th} bit is 1 in this user's training iris code; let n_i be the times of 0. The intra-class bit distribution is modelled as a binomial distribution: $p'_i = \frac{m_i}{m_i+n_i}$ and $q'_i = \frac{n_i}{m_i+n_i}$, where p'_i denotes the probability that the i^{th} bit equals to 1, and q'_i denotes the probability of 0.

After the intra-class bit distribution is obtained, p_i and q_i are estimated as follows.

For the case that the i^{th} bit of the meta-code is 1, q_i is calculated by:

$$q_i = \begin{cases} \frac{0.5 + \beta_1 (q'_i - 0.5)}{0.5 + \beta_1 (q'_i - 0.5) + p'_i} & \text{if } q'_i \geq 0.5 \\ \frac{\beta_2 q'_i}{\beta_2 q'_i + p'_i} & \text{if } q'_i < 0.5 \end{cases} \quad (8.6)$$

where $\beta_1, \beta_2 \in [0, 1]$ are parameters and their effect will be illustrated later. After obtaining q_i , p_i is set by $p_i = 1 - q_i$. The above method to calculate p_i and q_i can be decomposed into the following process. Firstly, p_i is set by $p_i = p'_i$. Then, if $q'_i \geq 0.5$, q_i is obtained by $q_i = 0.5 + \beta_1 (q'_i - 0.5)$; if $q'_i < 0.5$, q_i is obtained by $q_i = \beta_2 q'_i$. Finally, p_i and q_i are normalised to have a sum of 1. The reason of the above process can be explained as follows:

- Given that the i^{th} bit of the meta-code is 1, $q'_i \geq 0.5$ means that the i^{th} bit in most of training iris codes is different from the i^{th} bit in the meta-code which is noiseless and extracted from stable iris structures. Due to such inconsistency between the intra-class distribution of the i^{th} bit and the noiseless and stable counterpart estimate of this bit, the i^{th} bit is considered to be more vulnerable to noise and variations in iris captures, *i.e.* less reliable. Therefore, a penalty is applied to the obtained stability: q_i is calculated by shifting q'_i towards 0.5. This leads to a lower certainty after normalisation and hence reducing the stability value (*i.e.* q_i corresponds to a lower stability compared to q'_i).
- On the contrary, $q'_i < 0.5$ means that the intra-class distribution of the i^{th} bit is consistent with the meta-code. Hence, this bit is considered to be more robust to noise and capturing variations, *i.e.* more reliable. Thus, the obtained stability is improved: q_i is determined by reducing q'_i . This reduction shifts q'_i away from 0.5. It results in higher certainty after normalisation and hence improving the stability value (*i.e.* q_i corresponds to a higher stability compared to q'_i).

To illustrate the effect of β_1 and β_2 , Fig. 8.2 plots the function defined by Eqn. 8.6

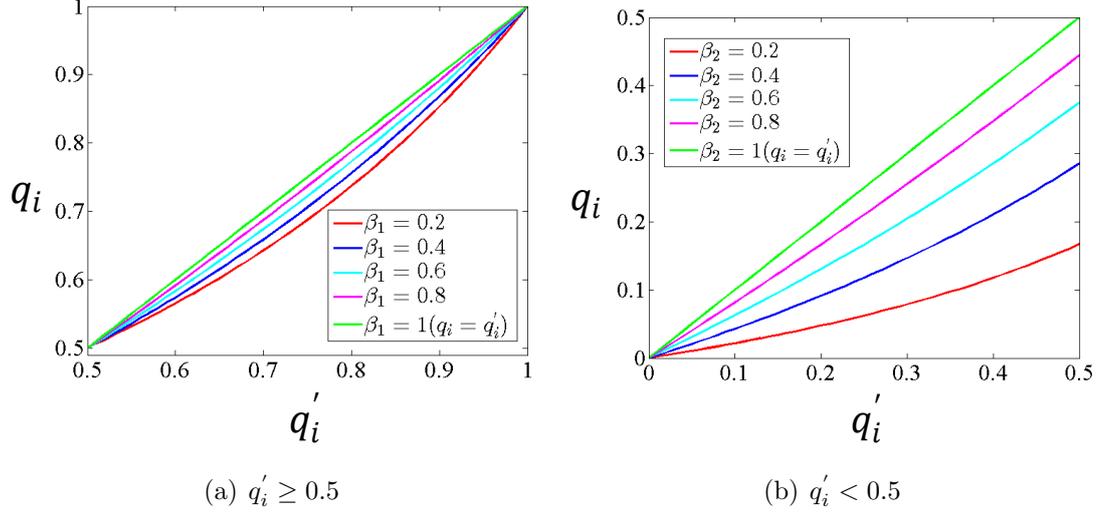


Figure 8.2: The function defined by Eqn. 8.6 with different choices of β_1 and β_2 .

with different choices of β_1 and β_2 . From Fig. 8.2(a), it can be seen that β_1 controls the strength of penalty on bit stability when the intra-class distribution is inconsistent with meta code ($q'_i \geq 0.5$). When $q'_i \geq 0.5$, a smaller β_1 leads to a q_i closer to 0.5, and it results a lower stability value as discussed above. In other words, a smaller β_1 imposes a heavier penalty to the stability of the bits whose intra-class distribution is inconsistent with meta-code. As a special case, $\beta_1=1$ means that no penalty is applied.

On the other hand, from Fig. 8.2(b), it can be seen that β_2 determines the level of improvement of bit stability when the intra-class distribution is consistent with meta-code ($q'_i < 0.5$). When $q'_i < 0.5$, a smaller β_2 makes the obtained q_i more distant from 0.5, leading to a higher stability value as discussed above. That is, a smaller β_2 applies more improvement to the stability of the bits whose intra-class distribution is consistent with meta-code. As a special case, $\beta_2=1$ means that no improvement is applied.

Finally, Eqn. 8.6 has discontinuity: since $\lim_{q'_i \rightarrow 0.5^+} q_i = 0.5$ and $\lim_{q'_i \rightarrow 0.5^-} q_i = \frac{\beta_2}{\beta_2+1} \leq 0.5$ ($0 \leq \beta_2 \leq 1$), Eqn. 8.6 is discontinuous at $q'_i = 0.5$ when $\beta_2 \neq 1$. This is reasonable. In the proposed method, $q'_i > 0.5$ and $q'_i < 0.5$ correspond to two different cases (*i.e.* the intra-class distribution is inconsistent/consistent with meta-

code); since the proposed method applies different strategies to these two cases, it leads to different right and left limits of q_i at $q'_i = 0.5$. As a detailed discussion, the left and right limit of q_i at $q'_i = 0.5$ is firstly analysed as follows:

- The right limit of q_i at $q'_i = 0.5$ is 0.5. Note that $q'_i \rightarrow 0.5^+$ means $q'_i > 0.5$. For Eqn. 8.6 (the meta-code is 1), it corresponds to the case that the intra-class distribution is inconsistent with meta-code. For this case, the proposed method applies a penalty to the obtained stability by shifting q'_i towards 0.5 (see the explanations following Eqn. 8.6). Therefore, the right limit of q_i at $q'_i = 0.5$ is essentially the probability corresponding to the lowest stability that can be obtained after applying this penalty. This lowest stability that the right limit of q_i corresponds to is equal to the minimum stability ($q_i = 0.5$ corresponds to $s_i = 0$).
- The left limit of q_i at $q'_i = 0.5$ is $\frac{\beta_2}{\beta_2+1} < 0.5$ (assuming $\beta_2 \neq 1$; $\beta_2 = 1$ will be discussed later). Note that $q'_i \rightarrow 0.5^-$ means $q'_i < 0.5$. For Eqn. 8.6, it corresponds to the case that the intra-class distribution is consistent with meta-code. For this case, the proposed method improves the obtained stability by reducing q'_i . Therefore, the left limit of q_i at $q'_i = 0.5$ is actually the probability corresponding to the lowest stability that can be obtained after this improvement. Due to the proposed improvement, this lowest stability that the left limit corresponds to is larger than the minimum stability ($q_i < 0.5$ corresponds to $s_i > 0$).
- As stated before, $\beta_2 = 1$ means that no improvement is applied when the intra-class distribution is consistent with meta-code. In this case, the left limit of q_i at $q'_i = 0.5$ is 0.5. This left limit means that the lowest stability that can be obtained without the improvement equals to the minimum stability ($q_i = 0.5$ corresponds to $s_i = 0$), when the intra-class distribution is consistent with meta-code.

From the above analysis, it can be seen that the discontinuity is due to the proposed improvement to the stability, when the intra-class distribution is consistent with meta-code. With this improvement (*i.e.* $\beta_2 \neq 1$), the left and right limit of q_i at $q'_i = 0.5$ are different, and they correspond to different minimum stabilities in two different cases in the proposed method. Without this improvement (*i.e.* set $\beta_2 = 1$), Eqn. 8.6 is continuous.

In Eqn. 8.6, $q'_i = 0.5$ is incorporated into the branch of $q'_i > 0.5$, because it leads to a more reasonable stability. $q'_i = 0.5$ means the highest instability in the intra-class distribution. Since, in this case, the intra-class distribution shows the highest instability and the consistency between the intra-class distribution and meta-code cannot be determined, $q'_i = 0.5$ is combined with the branch of $q'_i > 0.5$ to assign the minimum stability of 0 to the i^{th} bit (recall that the proposed method determines the stability based on both intra-class distribution and the consistency between the intra-class distribution and meta-code).

Similarly to Eqn. 8.6, when the i^{th} bit of the meta-code is 0, p_i is calculated by:

$$p_i = \begin{cases} \frac{0.5 + \beta_1 (p'_i - 0.5)}{0.5 + \beta_1 (p'_i - 0.5) + q_i} & \text{if } p'_i \geq 0.5 \\ \frac{\beta_2 p'_i}{\beta_2 p'_i + q_i} & \text{if } p'_i < 0.5 \end{cases} \quad (8.7)$$

After obtaining p_i , q_i is set by $q_i = 1 - p_i$. The effect of Eqn. 8.7 and parameters β_1, β_2 are the same as that in the above illustrations based on Eqn. 8.6.

Comparing with using intra-class bit distribution only to compute the stability map (*i.e.* set $p_i = p'_i$ and $q_i = q'_i$), the proposed stability map considers not only the intra-class bit distribution, but also the robustness of the bit to noise and capturing variations. It increases the weight if the intra-class distribution of a bit is consistent with its noiseless and stable estimate, otherwise the weight is reduced. Such a scheme is preferable for intra-class matching. It assigns higher weight to the bits that are not only stable within class, but also robust to noise and capturing variations. Such

bits are more likely to be consistent given two iris codes of the same eye under less constrained capture conditions, leading to a better intra-class matching.

However, the influence of the proposed stability map on the inter-class comparison is less controllable. Given two iris codes from different eyes, this stability map is possible to emphasise the local consistency in some regions and ignore the inconsistency in some other regions. In other words, the stability map is possible to impair inter-class separation. The main reason is that the stability map only models the intra-class matching which mainly relies on the bit consistency (stability), but it ignores the inter-class relations where the bit inconsistency (discriminability) plays a more important role. To address this problem, a discriminability map is proposed to model the inter-class comparison.

8.2.3 The discriminability map

The aim of discriminability map is to emphasise the bit with higher inter-class discriminative power. The discriminability map is calculated using a 1-to-N strategy. For an arbitrary user, let $\mathbf{d} = [d_1, d_2, \dots, d_n] \in \mathbb{R}^n$ be the discriminability map of this user; denote the probability that the i^{th} bit of all other users' iris code equals to 1 and 0 by p_i^* and q_i^* , respectively; $p_i^* + q_i^* = 1$. p_i and q_i are still used to denote the probability that the i^{th} bit of this user's iris code equals to 1 and 0, respectively. The discriminability of the i^{th} bit in the iris code, d_i , is calculated by:

$$d_i = f_d(p_i q_i^* + q_i p_i^*) \quad (8.8)$$

where $f_d : \mathbb{R} \rightarrow \mathbb{R}$ is a function modelling the relationship between $p_i q_i^* + q_i p_i^*$ and d_i . Similarly to the bit stability model (Eqn. 8.2), Eqn. 8.8 is also a probability based model. $p_i q_i^* + q_i p_i^*$ calculates the probability that the i^{th} bit does not match given one iris code of one user and one iris code of a different user, and d_i is modelled as a function of $p_i q_i^* + q_i p_i^*$. $p_i q_i^* + q_i p_i^*$ reaches its maximum (1) when $p_i = 1, p_i^* = 0$

or $p_i = 0$, $p_i^* = 1$, which means that the i^{th} bit of the iris code is most likely to be different between this user and other users. On the other hand, $p_i q_i^* + q_i p_i^*$ reaches the minimum (0) when $p_i = p_i^* = 1$ or $p_i = p_i^* = 0$.

Similarly to the model for bit stability in the Section 8.2.2, d_i should be monotonically increasing with $p_i q_i^* + q_i p_i^*$. Let $\mathbf{t}' = [t'_1, t'_2, \dots, t'_n] \in \mathbb{R}^n$ with $t'_i = p_i q_i^* + q_i p_i^*$. The relationship between $p_i q_i^* + q_i p_i^*$ and d_i is modelled as follows:

$$f_d(t'_i) = \frac{t'_i - \min\{\mathbf{t}'\}}{\max\{\mathbf{t}'\} - \min\{\mathbf{t}'\}} \quad (8.9)$$

where $\min\{\bullet\}$ and $\max\{\bullet\}$ calculate the minimum and maximum value in a vector, respectively.

For an arbitrary user, p_i and q_i can be estimated using the method in Section 8.2.2. As for p_i^* and q_i^* , assuming that in all the training iris codes of the other users, the i^{th} bit is 1 for m'_i times and is 0 for n'_i times, the following estimation is used: $p_i^* = \frac{m'_i}{m'_i + n'_i}$ and $q_i^* = \frac{n'_i}{m'_i + n'_i}$.

To investigate the influence of p_i , q_i , p_i^* and q_i^* on the discriminability map, $t'_i = p_i q_i^* + q_i p_i^*$ can be written as follows:

$$\begin{aligned} t'_i &= p_i q_i^* + q_i p_i^* \\ &= p_i (1 - p_i^*) + (1 - p_i) p_i^* \\ &= (1 - 2p_i) p_i^* + p_i \end{aligned} \quad (8.10)$$

Since d_i is monotonically increasing with t'_i as in Eq. 8.9, some interesting characteristics of the discriminability map can be found:

- if $p_i = 0.5$ ($p_i = q_i$), t'_i will be always 0.5 whatever the value of p_i^* . In other words, when the intra-class value of the i^{th} bit is highly uncertain, a moderate discriminability will be assigned. This is reasonable, since, in such case, the value of the i^{th} bit of this user's iris code is very likely to flip between 0 and

1, so this bit is always possible to be the same with some iris codes and to be different with other iris codes in intra-class comparisons. A similar case is $p_i^* = 0.5$ ($p_i^* = q_i^*$).

- if $p_i < 0.5$ ($p_i < q_i$), t'_i is monotonically increasing with p_i^* , and hence d_i is monotonically increasing with p_i^* . That is, if the i^{th} bit of an user's iris code is more likely to be 0, the d_i of this user will be higher when the i^{th} bit of other users' iris code is more likely to be 1.
- if $p_i > 0.5$ ($p_i > q_i$), t'_i is monotonically decreasing with p_i^* , so that d_i is monotonically decreasing with p_i^* . That is, if the i^{th} bit of an user's iris code is more likely to be 1, the d_i of this user will be lower when the i^{th} bit of other users' iris code is more likely to be 1.

It is obvious that these properties make possible to emphasise the bits with higher inter-class discriminative power.

8.2.4 The combined map

Given the stability map and the discriminability map of one user, the final iris weight map of this user, denoted by \mathbf{w} , is computed by combining the two maps as follows:

$$\mathbf{w} = \mathbf{s} \otimes \mathbf{d} \quad (8.11)$$

where \otimes is element-wise multiplication operator. This combination emphasises the bits that are both more stable within each class and more discriminative between classes.

Some examples of the proposed stability map, discriminability map and combined map are shown in Fig. 8.3. It can be seen that, the stability maps assign lower weights to the regions corresponding to top and bottom of iris, and it assigns higher

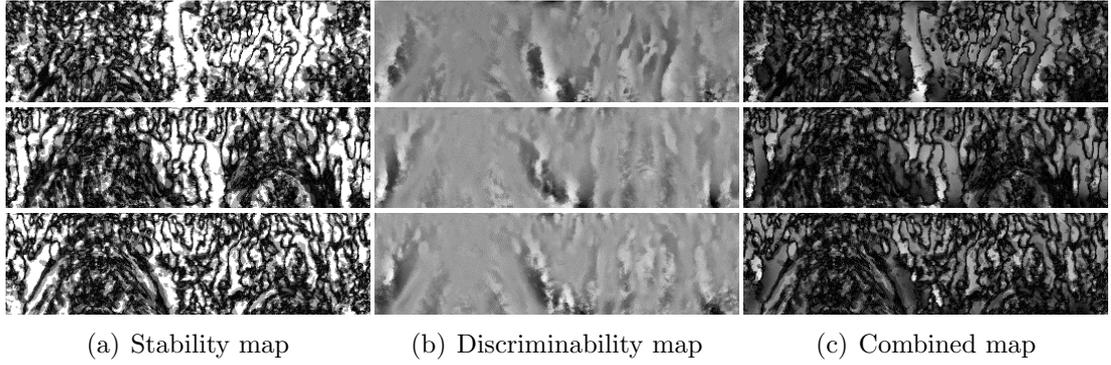


Figure 8.3: Examples of stability map with linear model, discriminability map and combined map; the images in the same row correspond to the same subject; the maps based on non-linear model are visually similar.

weights to the regions corresponding to both sides of iris. This is because top and bottom regions of iris are likely to be occluded by eyelids, hence the corresponding bits are less stable. On the other hand, the discriminability maps have a fog-like effect. This is because most of bits in iris codes have medium discriminability, and this results that most bits in the discriminability maps correspond to medium intensity values. However, it can be seen that light and dark regions exist in the discriminability maps, corresponding to the bits with higher and lower discriminability, respectively. Finally, it can be seen that the combined maps emphasise the bits with both high stability and discriminability.

8.2.5 Iris matching

Assume there are a probe iris code denoted by $\mathbf{C}_{\text{probe}}$, a gallery iris code of the j^{th} user denoted by $\mathbf{C}_{\text{gallery}}^j$, and an iris weight map for the j^{th} user denoted by \mathbf{w}^j . The distance between the gallery iris code and the probe iris code is calculated using a weighted Hamming distance [22, 98] (assuming the bits corresponding to noise like reflections have been eliminated for iris matching):

$$\text{HD} = \frac{\|(\mathbf{C}_{\text{probe}} \oplus \mathbf{C}_{\text{gallery}}^j) \otimes \mathbf{w}^j\|}{\|\mathbf{w}^j\|} \quad (8.12)$$

where \oplus is element-wise exclusive-or operator.

8.3 Experiment

This section presents the results of experimental analysis. Firstly, it introduces the datasets used in the experiment, followed by the experimental configuration, parameter setting and performance evaluation. Then, it analyses the proposed stability map, discriminability map and combined map individually to investigate the effect of each individual map. Next, it compares the proposed iris weight map to state-of-the-art iris weight map methods. It shows that the proposed iris weight map achieves a generally improved performance. Finally, it analyses the computational cost of the proposed method.

8.3.1 Datasets

The experiments are performed on four datasets introduced in Chapter 3: CASIAD [102], UBIRIS2 [19], CSIR [106] (recall that the images in CSIR dataset are captured by two devices with different sensors: AD100 and IKEMB-220) and MICHE [5] (recall that the images in MICHE dataset are captured by two mobile devices with different sensors: iPhone5 and Samsung Galaxy S4). The above datasets are chosen such that the experiments are performed on both single-sensor datasets (CASIAD, UBIRIS2) and cross-sensor datasets (CSIR, MICHE) captured in less constrained environments. Tab. 8.1 introduces the information of all the datasets used for the experiments in this chapter.

The images used for experiments are as follows. For CASIAD dataset, This experiment uses the images from both left and right eyes of all the subjects, and 97 images with too small iris regions are eliminated. For UBIRIS2, a subset with 1000 images from 171 subjects is adopted. This subset was released for NICE II contest [18]. For

Dataset	CASIA4	UBIRIS2	CSIR	MICHE
Eye number	284	171	200	75
Image number	5037	1000	7787	1479
Sensor type	Single-sensor	Single-sensor	Cross-sensor	Cross-sensor
Wavelength	NIR	Visible	NIR	Visible

Table 8.1: Summarisation on the datasets for experiments in this chapter.

CSIR dataset, all the subjects are adopted in this experiment. 213 images with too small iris region are eliminated. It results a subset with 7787 images from 200 eyes. For MICHE dataset, a subset is employed in this experiment. For each eye, at most 20 images are randomly selected, 10 from iPhone5 captures and 10 from Samsung Galaxy S4 captures. 19 images with too small iris region are eliminated. It leads to a subset with 1479 images from 75 eyes. For UBIRIS2 and MICHE datasets, this experiment operates on the red channel of RGB colour space.

For all the datasets, the iris region is segmented using the algorithm in Chapter 4. The failed segmentations are manually corrected. The size of unwrapped iris images is set to 100×360 . Iris codes are extracted from the unwrapped images using 1-D log-Gabor filter [15].

8.3.2 Experimental configuration

This subsection introduces the experimental configuration for the experiments in this chapter. Since the experiments are conducted on both single-sensor data and cross-sensor data, this leads to a configuration for single-sensor experiment and a configuration for cross-sensor experiment.

Configuration for single-sensor experiment. The following data are employed for single-sensor experiment. For CASIAD and UBIRIS2 datasets, all the images are used. For CSIR dataset, the images captured by IKEMB-220 are used. For MICHE dataset, the images captured by Samsung Galaxy S4 are used. On all the datasets,

for each eye, the first 5 images are used as gallery images, the rest images are used as probe images. If an eye has less than 5 images, all the images are used as gallery images, and there are no probe images for this eye. For each eye, the meta-image, meta-code and iris weight map are learned from the gallery images.

Configuration for cross-sensor experiment. This configuration is only applicable to CSIR and MICHE datasets. For each eye in CSIR dataset, the first 5 images captured by AD100 and the first 5 images captured by IKEMB-220 are used as the learning data. The meta-image, meta-code and iris weight map are obtained from the learning data. The 6th to 10th images captured by AD100 are used as gallery images, and probe images are the images captured by IKEMB-220 but not included in the learning data. The configuration on MICHE dataset is similar to that on CSIR dataset. For each eye, the learning data include 10 images, 5 from iPhone5 captures and 5 from Samsung Galaxy S4 captures. The 6th to 10th images captured by iPhone5 are used as the gallery images, and probe images are the images captured by Samsung Galaxy S4 but not included in the learning data.

8.3.3 Parameter tuning and performance evaluation

For CASIAD dataset, the first 20 eyes are used for parameter tuning, and the remaining eyes are used for performance evaluation. For UBIRIS2 dataset, the first 19 eyes are used for parameter tuning, and the remaining eyes are used for performance evaluation. For CSIR and MICHE datasets, the learning data in cross-sensor experiment are used for parameter tuning, and all the eyes are used for performance evaluation in both single-sensor and cross-sensor experiments. Note that, for CSIR and MICHE datasets, the gallery images in the single-sensor experiment are included in the data for parameter tuning. This is considered to be acceptable, since gallery images can be considered as known data, and the probe data used in performance evaluation remains excluded from parameter tuning in both single-sensor

Database	r	β_1	β_2	$1/f_0$	σ/f_0
CASIA4	1	0.3	0.8	19	0.48
UBIRIS2	1	0.3	0.8	40	0.38
CSIR	1	0.7	0.8	24	0.42
MICHE	1	0.7	0.8	17	0.26

Table 8.2: Parameter setting.

and cross-sensor experiments.

There are 5 parameters to be determined: r is the size of U and V in Eqn. 8.1, β_1 and β_2 in Eqn. 8.3 and Eqn. 8.4, $1/f_0$ and σ/f_0 are 1-D log-Gabor parameters (see Eqn. 2.25). Based on the above described data for parameter setting, the setting of these parameters are reported in Tab 8.2. Note that, for 1-D log-Gabor parameters ($1/f_0$ and σ/f_0) on CASIAD and UBIRIS2 datasets, the setting in Section 6.4.1 is adopted; for CSIR and MICHE datasets, uniform parameter setting is used for both single-sensor and cross-sensor experiments.

The performance is evaluated in two tasks: identification and verification. The identification performance is evaluated by cumulative match characteristic (CMC), while the verification performance is evaluated by receiver operating characteristic (ROC).

Also, paired student-t test are performed to analyse the statistical significance of the experimental results. Specifically, in all the datasets, 100 subsets are randomly selected from the performance evaluation set. The subjects included in each subset are randomly selected. For each subset, R1RA is calculated to represent the identification performance; equal error rate (EER) and genuine acceptance rate (GAR) when false acceptance rate (FAR) is 0.01% (GAR@FAR= 0.01%) are calculated to represent the verification performance; decidability index (DI) [6] is calculated to represent the intra-class and inter-class separability of matching scores. The p value of paired student-t test on R1RA, EER, GAR@FAR= 0.01% and DI between different methods are used to evaluate the statistical significance. The comparison

methods where the paired t-test is applied vary according to different experiments. This will be described specifically in each experiment in the following Section 8.3.4 and Section 8.3.5.

8.3.4 Analysis of the proposed iris weight map

The proposed iris weight map is the combination of a stability map and a discriminability map. Furthermore, the stability map can be produced using a linear model (Eqn. 8.3) or a non-linear model (Eqn. 8.4). This section makes detailed analysis on these maps (two stability maps corresponding to linear and non-linear models, discriminability map, two combined maps corresponding to combining the discriminability map with each of the two stability maps). To investigate the effect of using meta-code to compute the stability map, two baseline stability maps are designed corresponding to the linear and nonlinear models. The two baseline maps are computed using only the intra-class bit consistency, without using meta-codes. That is, based on Eqn. 8.2, two baseline stability maps are produced, one using the linear model in Eqn. 8.3 and the other using the nonlinear model in Eqn. 8.4, with p_i and q_i in Eqn. 8.2 calculated as follows: $p_i = \frac{m_i}{m_i+n_i}$ and $q_i = \frac{n_i}{m_i+n_i}$. Note that the baseline map using linear model is equivalent to the personalised weight map in [98].

The above maps are referred to as follows: the baseline stability map with the linear model is referred to as Baseline_L, the baseline stability map with the non-linear model is referred to as Baseline_NL, the proposed stability map with the linear model is referred to as Proposed_S_L, the proposed stability map with the non-linear model is referred to as Proposed_S_NL, the proposed discriminability map is referred to as Proposed_D, the combined map of the proposed stability map with the linear model and the proposed discriminability map is referred to as Proposed_C_L, the combined map of the proposed stability map with the non-linear model and the proposed discriminability map is referred to as Proposed_C_NL.

Results of single-sensor experiments. Fig. 8.4 and Fig. 8.5 report the CMC and ROC curves of the above iris weight maps in the single-sensor experiment, respectively. The paired student-t test is performed between: (1) Baseline_L and Proposed_S_L, Baseline_NL and Proposed_S_NL to analyse the effect of using meta-code; (2) Proposed_S_L and Proposed_C_L, Proposed_D and Proposed_C_L, Proposed_S_NL and Proposed_C_NL, Proposed_D and Proposed_C_NL, to study the effect of combining individual maps. Tab. 8.3, Tab. 8.4, Tab. 8.5 and Tab. 8.6 report the mean and standard deviation of R1RA, EER, GAR@FAR= 0.01% and DI, together with the result of corresponding significance test, respectively. In these tables, the results of significance test in the row ‘Baseline_L’ and ‘Baseline_NL’ correspond to the above two tests in paired t-test (1), and the results of significance test in the row ‘Proposed_S_L’, ‘Proposed_C_L’, ‘Proposed_S_NL’ and ‘Proposed_C_NL’ correspond to the above four tests in paired t-test (2).

The following observations are obtained from the results of single-sensor experiments shown in Fig. 8.4, Fig. 8.5, Tab. 8.3, Tab. 8.4, Tab. 8.5 and Tab. 8.6. First, compared to baselines, the propose stability map is able to achieve an improved identification performance with either linear or non-linear model. However, the verification performance of the propose stability map exhibits no improvement in comparison to baselines. The reason is that the proposed stability map focuses on intra-class matching, but it is possible to impair the inter-class separation, as illustrated in Section 8.2.2. Better intra-class matching improves the accuracy of the lowest matching scores, leading to a better R1RA. However, impaired inter-class separation influence the verification performance.

Second, individual discriminability map is insufficient to achieve a generally good performance in both identification and verification tasks. In terms of identification, the R1RA achieved by proposed discriminability map (Proposed_D) is generally lower than proposed iris weight maps with the linear model (Proposed_S_L and Proposed_C_L) in all datasets. The only exception is CSIR dataset where proposed dis-

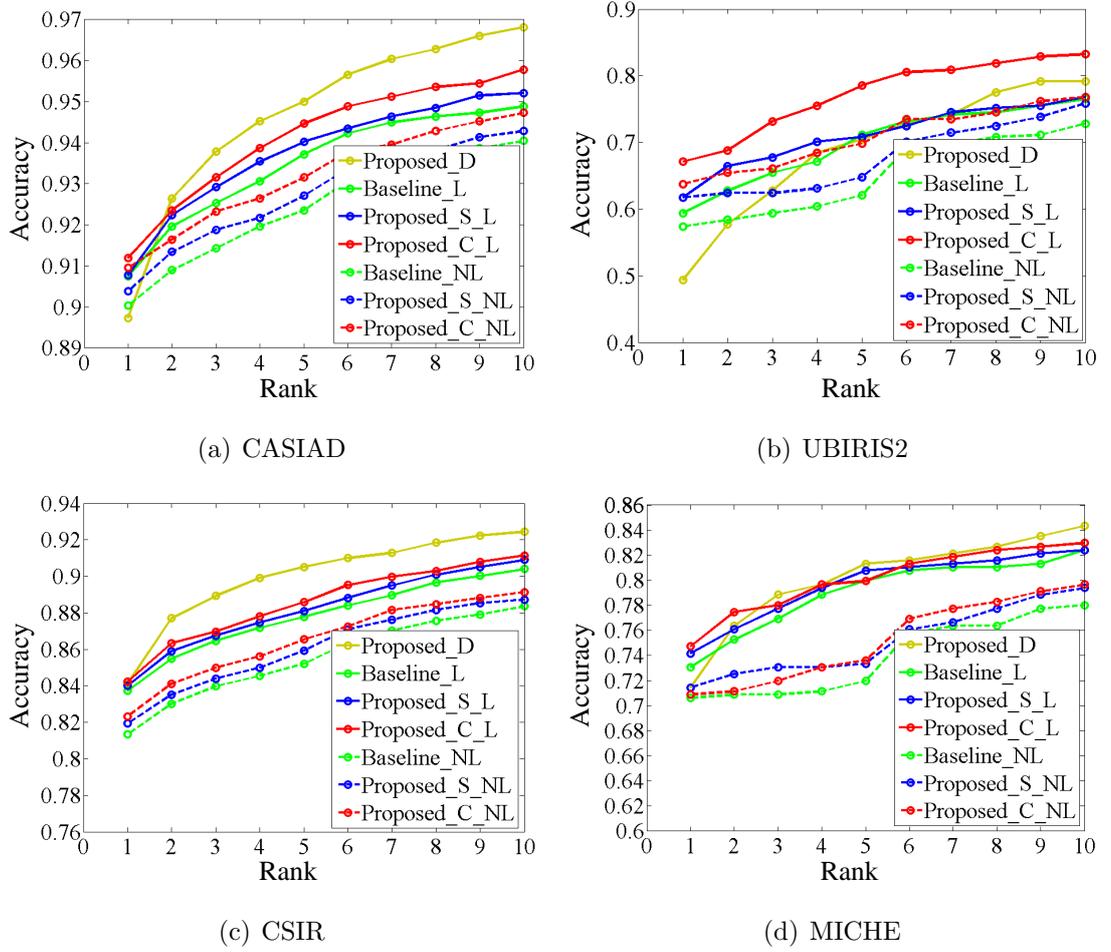


Figure 8.4: The CMC curves to analyse the proposed iris weight map in the single-sensor experiment.

Table 8.3: The R1RA to analyse the proposed iris weight map in the single-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAD	UBIRIS2	CSIR	MICHE
Proposed_D	90.57 \pm 0.48	51.43 \pm 2.11	85.41 \pm 0.85	73.31 \pm 2.02
Baseline_L	91.44 \pm 0.47 [#]	61.80 \pm 1.97 [#]	84.74 \pm 0.83 [#]	75.05 \pm 1.93 [#]
Proposed_S_L	91.53 \pm 0.50 [#]	63.24 \pm 2.03 [#]	85.00 \pm 0.82 [#]	75.57 \pm 1.83 [#]
Proposed_C_L	91.84 \pm 0.48[#]	68.84 \pm 2.05[#]	85.24 \pm 0.81 [#]	76.33 \pm 1.76[#]
Baseline_NL	90.72 \pm 0.54 [#]	60.20 \pm 1.95 [#]	82.59 \pm 0.86 [#]	73.06 \pm 2.06 [#]
Proposed_S_NL	91.07 \pm 0.52 [#]	63.89 \pm 1.92 [#]	83.11 \pm 0.86 [#]	73.72 \pm 1.98 [#]
Proposed_C_NL	91.52 \pm 0.50 [#]	66.09 \pm 1.78 [#]	83.51 \pm 0.83 [#]	73.38 \pm 2.02 [#]

[#]The p value of paired student-t test is lower than 0.05 level.

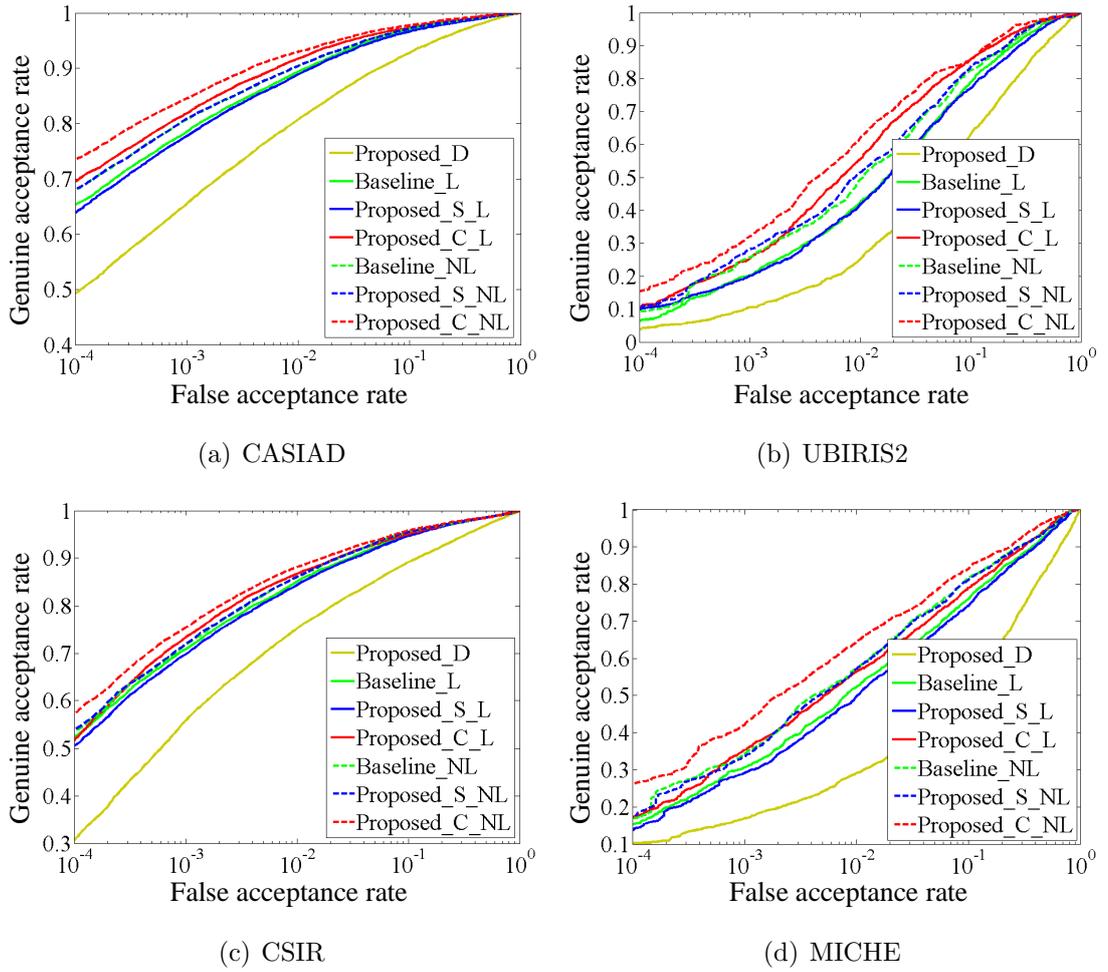


Figure 8.5: The ROC curves to analyse the proposed iris weight map in the single-sensor experiment.

Table 8.4: The EER to analyse the proposed iris weight map in the single-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAD	UBIRIS2	CSIR	MICHE
Proposed_D	8.11 \pm 0.24	22.38 \pm 0.51	10.54 \pm 0.40	28.16 \pm 0.92
Baseline_L	4.95 \pm 0.23 [#]	15.38 \pm 0.58 [#]	6.58 \pm 0.30 [#]	17.44 \pm 0.93 [#]
Proposed_S_L	5.07 \pm 0.22 [#]	16.45 \pm 0.67 [#]	6.76 \pm 0.32 [#]	18.27 \pm 0.94 [#]
Proposed_C_L	4.18 \pm 0.20 [#]	12.23 \pm 0.49 [#]	6.29 \pm 0.29 [#]	16.42 \pm 0.93 [#]
Baseline_NL	4.67 \pm 0.54 [#]	14.18 \pm 0.69 [#]	6.07 \pm 0.27 [#]	15.16 \pm 1.04 [#]
Proposed_S_NL	4.65 \pm 0.22 [#]	13.81 \pm 0.72 [#]	6.04 \pm 0.27 [#]	15.29 \pm 1.02 [#]
Proposed_C_NL	3.79 \pm 0.20[#]	12.02 \pm 0.56[#]	5.87 \pm 0.25[#]	13.34 \pm 0.98[#]

[#]The p value of paired student-t test is lower than 0.05 level.

Table 8.5: The GAR@FAR= 0.01% to analyse the proposed iris weight map in the single-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAD	UBIRIS2	CSIR	MICHE
Proposed_D	49.13 \pm 1.41	3.98 \pm 0.65	30.78 \pm 3.32	9.84 \pm 0.93
Baseline_L	65.16 \pm 1.53 [#]	6.97 \pm 1.78 [#]	52.41 \pm 2.84 [#]	15.11 \pm 4.12 [#]
Proposed_S_L	64.00 \pm 1.53 [#]	10.06 \pm 1.05 [*]	50.93 \pm 2.70 [#]	13.74 \pm 4.42 [#]
Proposed_C_L	69.42 \pm 1.35 [#]	9.85 \pm 2.50 [#]	51.64 \pm 4.11 [#]	16.06 \pm 5.51 [#]
Baseline_NL	67.70 \pm 1.46 [#]	8.11 \pm 2.92 [#]	53.59 \pm 2.04 [#]	17.70 \pm 6.38 [#]
Proposed_S_NL	68.07 \pm 1.40 [#]	10.24 \pm 2.29 [#]	53.78 \pm 2.10 [#]	17.17 \pm 5.95 [#]
Proposed_C_NL	73.41 \pm 1.23[#]	13.81 \pm 3.78[#]	56.91 \pm 2.74[#]	25.23 \pm 5.61[#]

[#]The p value of paired student-t test is lower than 0.05 level.

^{*}The p value of paired student-t test is higher than 0.05 level.

Table 8.6: The DI to analyse the proposed iris weight map in the single-sensor experiment, with the result of significance test (best performance bolded, format: mean \pm standard deviation).

Method	CASIAD	UBIRIS2	CSIR	MICHE
Proposed_D	2.499 \pm 0.029	1.439 \pm 0.029	2.334 \pm 0.037	1.081 \pm 0.034
Baseline_L	3.197 \pm 0.044 [#]	2.003 \pm 0.046 [#]	3.067 \pm 0.053 [#]	1.848 \pm 0.065 [#]
Proposed_S_L	3.266 \pm 0.043 [#]	1.969 \pm 0.047 [#]	3.016 \pm 0.052 [#]	1.782 \pm 0.062 [#]
Proposed_C_L	3.221 \pm 0.070 [#]	2.296 \pm 0.048 [#]	3.051 \pm 0.050 [#]	1.903 \pm 0.061 [#]
Baseline_NL	3.433 \pm 0.048 [#]	2.161 \pm 0.057 [#]	3.224 \pm 0.055 [#]	2.084 \pm 0.076 [#]
Proposed_S_NL	3.409 \pm 0.047 [#]	2.204 \pm 0.057 [#]	3.214 \pm 0.054 [#]	2.064 \pm 0.075 [#]
Proposed_C_NL	3.564 \pm 0.046[#]	2.458 \pm 0.061[#]	3.251 \pm 0.054[#]	2.241 \pm 0.073[#]

[#]The p value of paired student-t test is lower than 0.05 level.

criminability map has the best rank 1 recognition accuracy. As for verification, the performance of proposed discriminability map is the lowest among all the proposed iris weight maps. The possible reason is that, although the proposed discriminability map includes the information on inter-class bit discriminability, this map itself is insufficient to indicate the reliability of bits for iris matching. For example, as discussed in Section 8.2.3, the proposed discriminability map assigns a bit moderate discriminability if $p_i = q_i$; however, $p_i = q_i$ means that the i^{th} bit is highly unstable within class, so such bits are less reliable; although assigning such bits a moderate weight is consistent with the aim of the discriminability map, it is possible to reduce the overall performance using the weight individually.

Third, it can be seen that combining the proposed stability map and discriminabil-

ity map leads to generally improved identification and verification performance in both identification and verification tasks, with either linear or non-linear model. The only two exceptions are the identification performance of the non-linear model on CSIR and MICHE datasets. This observation demonstrates the effectiveness of proposed method which computes two individual weight maps based on intra-class stability and inter-class discriminability respectively and combines the two weight maps to produce the final iris weight map. The improvement is due to the advantage of considering both intra-class stability and inter-class discriminability. Such combination scheme only emphasises the bits with both high intra-class stability and high inter-class discriminability. The weight is suppressed if a bit is weak in either stability or discriminability. As an example, considering the bit with $p_i = q_i$ as discussed in the second observation above, although it has a moderate weight in the proposed discriminability map, the weight of such bits is 0 in proposed stability map (see Section 8.2.2); therefore, the final weight of such bit is 0 after combination.

Four, considering the performance of the final combined maps, it can be found that the performance of linear and non-linear models varies in identification and verification tasks. Specifically, the linear model (Proposed_C_L) performs better than non-linear model (Proposed_C_NL) in the identification task, but the non-linear model is better in the verification task. This can be explained by the different characteristics of the two models. For a less stable bit within class, the non-linear model (Eqn. 8.4 and Eqn. 8.5) essentially applies heavier penalty to this bit than the linear model (Eqn. 8.3). In the identification task with single-sensor data, such heavier penalty is possible to harm intra-class matching. In the iris code of single-sensor captures, a bit may still contain some useful similarity information, despite of a lower intra-class stability. Applying heavier penalty on such bits may cause the loss of similarity information, hence leading to a dropped rank 1 recognition accuracy. However, heavier penalty on less stable bits may be able to improve the verification performance by enhancing the inter-class separation. For the iris code

of a specific eye, the more likely a bit flips, the more likely it falsely matches with the bits in the iris codes of other eyes. A heavier penalty on the weight of such bits is able to suppress false inter-class matching, hence improving the verification performance.

Results of cross-sensor experiments. Fig. 8.6 and Fig. 8.7 report the CMC and ROC curves of different iris weight maps in the cross-sensor experiment, respectively. The significance analysis in the cross-sensor experiment is the same to that in the single-sensor experiment. The R1RA, EER, GAR@FAR= 0.01% and DI with the corresponding results of significant analysis are reported in Tab. 8.7, Tab. 8.8, Tab. 8.9 and Tab. 8.10, respectively.

Considering the result of cross-sensor experiments shown in Fig. 8.6, Fig. 8.7, Tab. 8.7, Tab. 8.8, Tab. 8.9 and Tab. 8.10, it is found that some conclusions are similar with single-sensor experiment, while others are not. Specifically, the second and third observations in the case of single-sensor experiments hold for the cross-sensor experiments as well. That is, individual discriminability map is insufficient to achieve a generally good performance, while combining the proposed stability and discriminability maps leads to a generally improved performance.

On the other hand, different observations are obtained for the performance between the proposed stability maps and the baselines (*i.e.* the first observation in single-sensor experiment), and the final performance of the linear and non-linear models (the fourth observation). In terms of the performance between the proposed stability maps and the baselines, it is found that only proposed stability map with the non-linear model (Proposed_S_NL) leads to an improved identification performance. The possible reason is explained as follows. The iris codes extracted from cross-sensor captures have a higher complexity than the iris codes extracted from single-sensor captures, due to the cross-sensor variation. Given such higher complexity, the linear model is too simple to express the relationship between $p_i^2 + q_i^2$ and s_i in Eqn. 8.2 (similar to under fitting). As a result, with the linear model, the estimated prob-

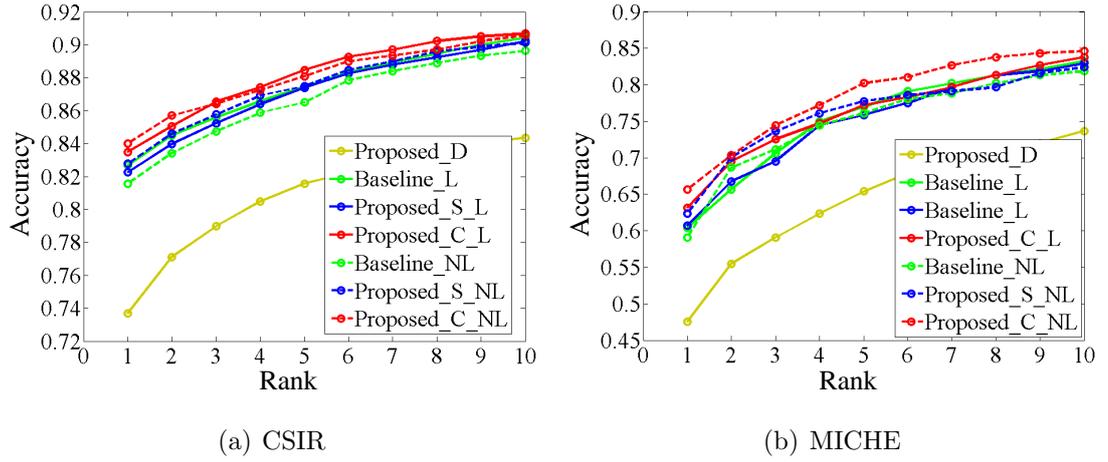


Figure 8.6: The CMC curves to analyse the proposed iris weight map in the cross-sensor experiment.

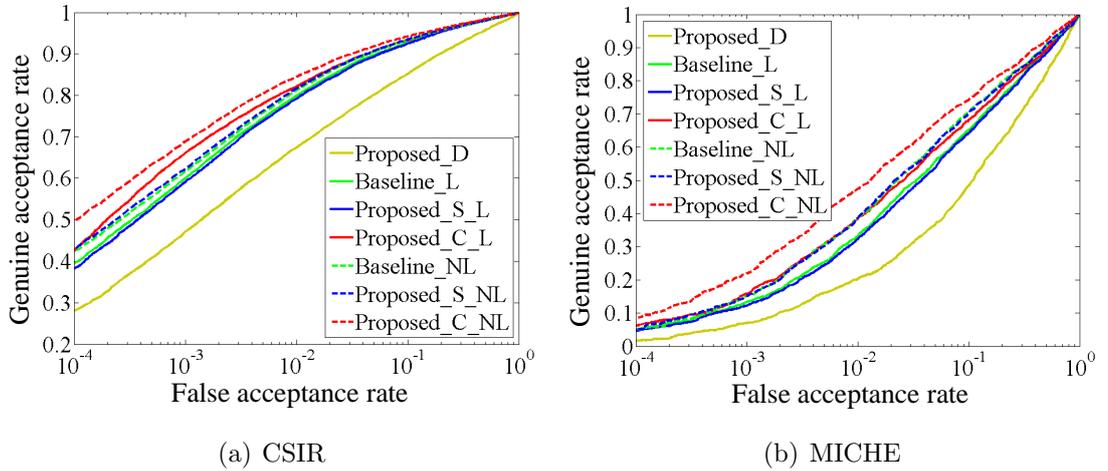


Figure 8.7: The ROC curves to analyse the proposed iris weight map in the cross-sensor experiment.

abilities p_i and q_i are not correctly mapped to a suitable stability value, hence it leads to little improvement of the performance. In contrast, the non-linear model, which has higher complexity than the linear one, is still able to accurately model the relationship between $p_i^2 + q_i^2$ and s_i , leading to an improved performance (note that the estimated p_i and q_i are exactly the same for the linear and non-linear models).

Also, comparing the final performance of the linear and non-linear models, Proposed_C_NL performs consistently better than Proposed_C_L in both identification and verification tasks. The possible reason is that, differently from single-sensor

Table 8.7: The R1RA to analyse the proposed iris weight map in the cross-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CSIR	MICHE
Proposed_D	74.83 \pm 1.08	49.07 \pm 1.50
Baseline_L	83.85 \pm 0.97 [#]	61.83 \pm 1.70 [#]
Proposed_S_L	83.43 \pm 0.95 [#]	62.02 \pm 1.68 [#]
Proposed_C_L	84.54 \pm 0.86 [#]	64.46 \pm 1.57 [#]
Baseline_NL	82.81 \pm 1.14 [#]	60.94 \pm 1.67 [#]
Proposed_S_NL	83.96 \pm 0.94 [#]	64.04 \pm 1.72 [#]
Proposed_C_NL	85.00 \pm 0.87[#]	66.82 \pm 1.50[#]

[#]The p value of paired student-t test is lower than 0.05 level.

Table 8.8: The EER to analyse the proposed iris weight map in the cross-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CSIR	MICHE
Proposed_D	12.93 \pm 0.39	29.55 \pm 0.53
Baseline_L	8.15 \pm 0.33 [#]	22.65 \pm 0.62 [#]
Proposed_S_L	8.33 \pm 0.34 [#]	23.20 \pm 0.58 [#]
Proposed_C_L	7.62 \pm 0.31 [#]	21.25 \pm 0.56 [#]
Baseline_NL	7.60 \pm 0.36 [#]	20.10 \pm 0.59 [#]
Proposed_S_NL	7.56 \pm 0.35 [#]	20.60 \pm 0.58 [#]
Proposed_C_NL	6.93 \pm 0.34[#]	18.39 \pm 0.61[#]

[#]The p value of paired student-t test is lower than 0.05 level.

Table 8.9: The GAR@FAR= 0.01% to analyse the proposed iris weight map in the cross-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CSIR	MICHE
Proposed_D	27.49 \pm 1.79	1.67 \pm 0.33
Baseline_L	39.87 \pm 2.89 [#]	4.94 \pm 0.64 [#]
Proposed_S_L	38.61 \pm 2.87 [#]	4.72 \pm 0.60 [#]
Proposed_C_L	42.41 \pm 3.47 [#]	5.96 \pm 0.93 [#]
Baseline_NL	42.51 \pm 2.49 [#]	4.87 \pm 0.88 [#]
Proposed_S_NL	43.36 \pm 2.34 [#]	5.20 \pm 1.06 [#]
Proposed_C_NL	49.33 \pm 2.26[#]	8.76 \pm 1.01[#]

[#]The p value of paired student-t test is lower than 0.05 level.

Table 8.10: The DI to analyse the proposed iris weight map in the cross-sensor experiment, with the result of significance test (best performance bolded, format: mean \pm standard deviation).

Method	CSIR	MICHE
Proposed_D	2.123 \pm 0.036	1.048 \pm 0.025
Baseline_L	2.738 \pm 0.049 [#]	1.462 \pm 0.031 [#]
Proposed_S_L	2.695 \pm 0.047 [#]	1.431 \pm 0.031 [#]
Proposed_C_L	2.753 \pm 0.047 [#]	1.535 \pm 0.033 [#]
Baseline_NL	2.898 \pm 0.057 [#]	1.608 \pm 0.037 [#]
Proposed_S_NL	2.889 \pm 0.054 [#]	1.604 \pm 0.037 [#]
Proposed_C_NL	2.962 \pm 0.055[#]	1.774 \pm 0.039[#]

[#]The p value of paired student-t test is lower than 0.05 level.

data where a bit with a lower intra-class stability may still contain some useful similarity information, the information in a less stable bit is less valuable for cross-sensor data, due to the higher data complexity. In this case, the heavier penalty assigned by the non-linear model on less stable bits is preferable for both identification and verification tasks.

Overall conclusions of the analysis. The above observations and analysis in both single-sensor and cross-sensor experiments lead to some conclusions as follows:

- The proposed stability map considering the bit consistency between the iris codes and meta-code is able to achieve an improved identification performance, in comparison to the baseline maps which do not use the meta-code. However, in the proposed stability model (Eqn. 8.2), besides the estimated probabilities p_i and q_i , the accuracy of f_s also plays a critical role in the performance. For example, for the identification task on single-sensor data, the linear model performs better, but, for the same task on cross-sensor data (with higher complexity), the non-linear model is preferable.
- Individual discriminability map is insufficient to achieve generally good performance alone.
- Combining proposed stability and discriminability maps leads to generally

improved identification and verification performance, with either linear or non-linear models, on both single-sensor and cross-sensor data.

- Considering the performance of the combined map using linear and non-linear models, it is found that the linear model performs better for the identification task on single-sensor data. For the other cases in the experiment, the non-linear model is preferable, especially for the cross-sensor iris recognition tasks.

8.3.5 Comparison with other methods

The proposed iris weight map is compared with 5 existing iris weight map methods: 1-D log-Gabor filter without any weight maps as the baseline [15], fisher feature selection (FFS) as used in [99], fragile bits (FB), personalized weight map (PW) [98] and power law based weight map (PL) [22]. The proposed combined map with linear model is referred to as Proposed_L, and the proposed combined map with non-linear model as Proposed_NL. Note that several feature selection strategies are used in [99]. The fisher feature selection is chosen as a reference for the performance of feature selection methods due to the following reasons: (1) as reported in [99], the performance of different feature selection strategies varies on different datasets using Gabor features; (2) it can be seen in [99] that the performance of FFS is stable at a mid-level and close to the top performance.

Fig. 8.8 and Fig. 8.9 show the CMC and ROC curves for all the comparison methods in the single-sensor experiment, respectively. Fig. 8.10 and Fig. 8.11 show the CMC and ROC curves in the cross-sensor experiment, respectively. The paired student-t tests are performed between the method with the best performance and all the other methods. Tab. 8.11, Tab. 8.12, Tab. 8.13 and Tab. 8.14 show respectively the mean and standard deviation of R1RA, EER, GAR@FAR= 0.01% and DI, together with the results of paired student-t tests, in the single-sensor experiment. Tab. 8.15, Tab. 8.16, Tab. 8.17 and Tab. 8.18 show these results in the cross-sensor experiment.

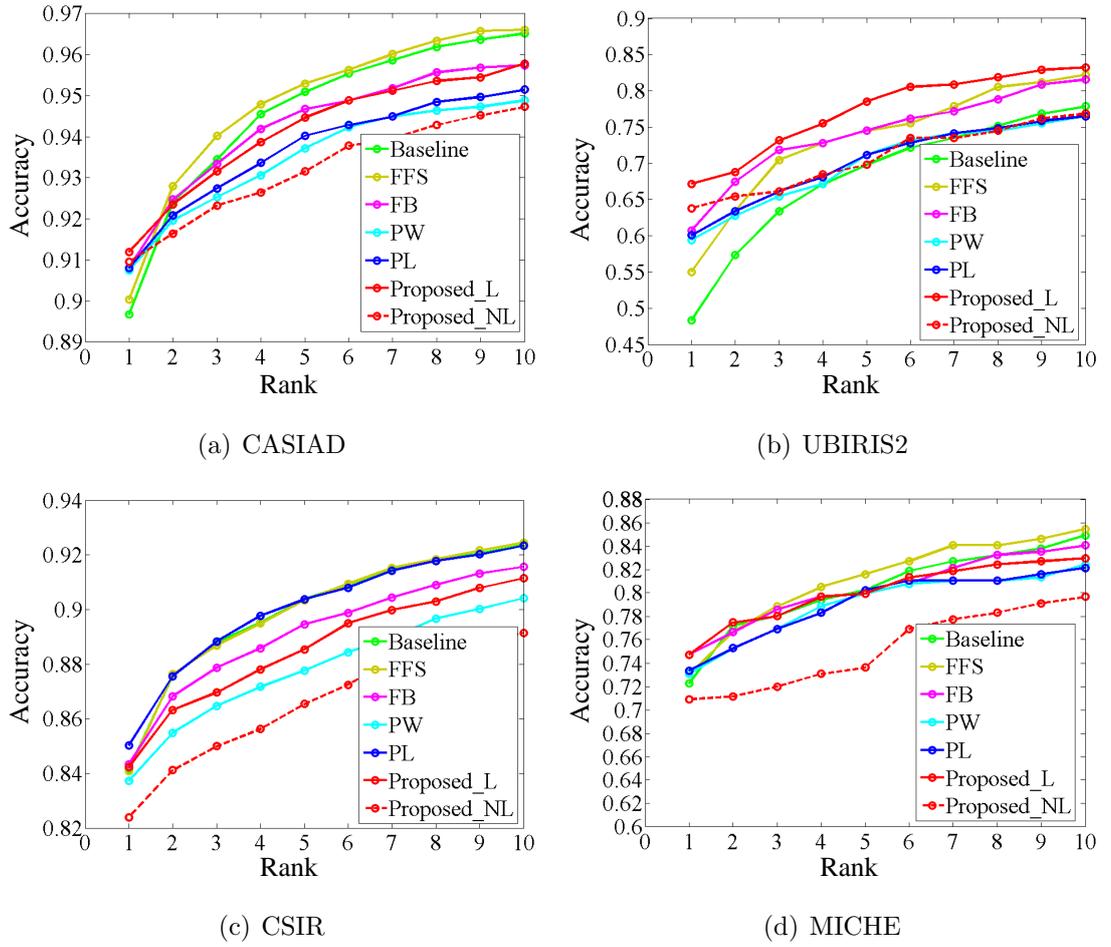


Figure 8.8: The CMC curves for comparison between proposed iris weight maps and existing iris weight map methods in the single-sensor experiment.

Table 8.11: The R1RA for comparison in the single-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAD	UBIRIS2	CSIR	MICHE
Baseline	90.49 \pm 0.48 [#]	50.54 \pm 2.02 [#]	85.41 \pm 0.84 [#]	73.99 \pm 1.95 [#]
FFS	90.86 \pm 0.52 [#]	57.08 \pm 1.96 [#]	85.36 \pm 0.86 [#]	74.47 \pm 1.90 [#]
FB	91.51 \pm 0.46 [#]	63.18 \pm 1.93 [#]	85.36 \pm 0.84 [#]	75.93 \pm 1.81 [#]
PW	91.44 \pm 0.47 [#]	61.80 \pm 1.97 [#]	84.74 \pm 0.83 [#]	75.05 \pm 1.93 [#]
PL	91.50 \pm 0.48 [#]	62.28 \pm 1.98 [#]	86.02 \pm 0.84	75.37 \pm 1.82 [#]
Proposed_L	91.84 \pm 0.48	68.84 \pm 2.05	85.24 \pm 0.81 [#]	76.33 \pm 1.76
Proposed_NL	91.52 \pm 0.50 [#]	66.09 \pm 1.78 [#]	83.51 \pm 0.83 [#]	73.38 \pm 2.02 [#]

[#]The p value of paired student-t test is lower than 0.05 level.

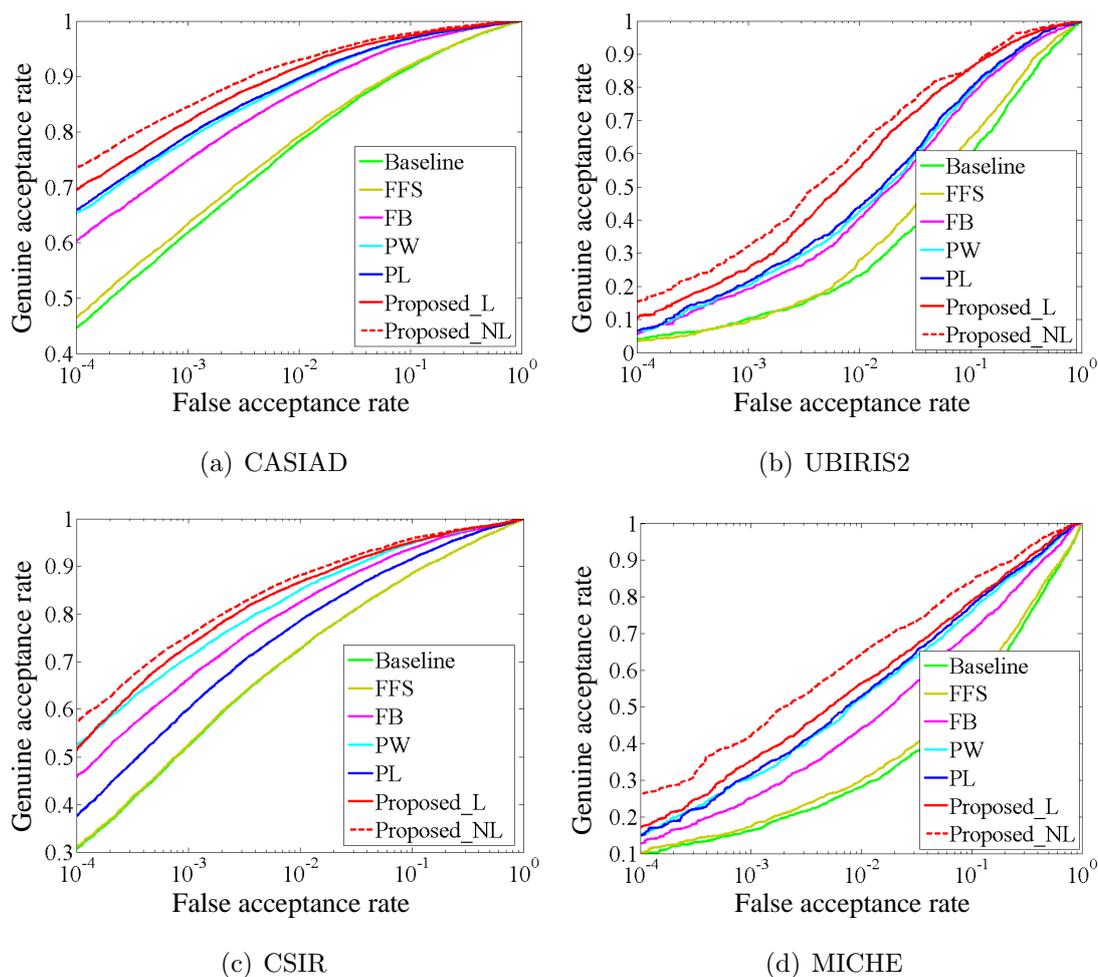


Figure 8.9: The ROC curves for comparison between proposed iris weight maps and existing iris weight map methods in the single-sensor experiment.

Table 8.12: The EER for comparison in the single-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAD	UBIRIS2	CSIR	MICHE
Baseline	8.92 \pm 0.27#	23.65 \pm 0.56#	10.95 \pm 0.40#	28.45 \pm 0.88#
FFS	8.62 \pm 0.26#	21.92 \pm 0.61#	10.90 \pm 0.40#	27.20 \pm 0.91#
FB	5.60 \pm 0.22#	16.00 \pm 0.55#	7.35 \pm 0.30#	20.56 \pm 0.93#
PW	4.95 \pm 0.23#	15.38 \pm 0.58#	6.58 \pm 0.30#	17.44 \pm 0.93#
PL	4.86 \pm 0.22#	15.08 \pm 0.63#	8.93 \pm 0.37#	16.77 \pm 0.93#
Proposed_L	4.18 \pm 0.20#	12.23 \pm 0.49#	6.29 \pm 0.29#	16.42 \pm 0.93#
Proposed_NL	3.79 \pm 0.20	12.02 \pm 0.56	5.87 \pm 0.25	13.34 \pm 0.98

#The p value of paired student-t test is lower than 0.05 level.

Table 8.13: The GAR@FAR= 0.01% for comparison in the single-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CASIAD	UBIRIS2	CSIR	MICHE
Baseline	44.63 \pm 1.63 [#]	4.09 \pm 0.61 [#]	31.10 \pm 2.07 [#]	9.77 \pm 1.13 [#]
FFS	46.62 \pm 1.57 [#]	3.52 \pm 0.59 [#]	31.37 \pm 2.09 [#]	10.38 \pm 1.47 [#]
FB	60.36 \pm 1.62 [#]	6.12 \pm 1.58 [#]	46.02 \pm 2.60 [#]	12.25 \pm 3.55 [#]
PW	65.16 \pm 1.53 [#]	6.97 \pm 1.78 [#]	52.41 \pm 2.84 [#]	15.11 \pm 4.12 [#]
PL	65.88 \pm 1.53 [#]	7.23 \pm 2.04 [#]	37.70 \pm 2.66 [#]	14.63 \pm 4.86 [#]
Proposed_L	69.42 \pm 1.35 [#]	9.85 \pm 2.50 [#]	51.64 \pm 4.11 [#]	16.06 \pm 5.51 [#]
Proposed_NL	73.41 \pm 1.23	13.81 \pm 3.78	56.91 \pm 2.74	25.23 \pm 5.61

[#]The p value of paired student-t test is lower than 0.05 level.

Table 8.14: The DI for comparison in the single-sensor experiment, with the result of significance test (best performance bolded, format: mean \pm standard deviation).

Method	CASIAD	UBIRIS2	CSIR	MICHE
Baseline	2.435 \pm 0.029 [#]	1.364 \pm 0.029 [#]	2.315 \pm 0.038 [#]	1.081 \pm 0.034 [#]
FFS	2.475 \pm 0.030 [#]	1.534 \pm 0.033 [#]	2.318 \pm 0.038 [#]	1.144 \pm 0.035 [#]
FB	2.992 \pm 0.038 [#]	1.916 \pm 0.040 [#]	2.863 \pm 0.045 [#]	1.602 \pm 0.053 [#]
PW	3.197 \pm 0.044 [#]	2.003 \pm 0.046 [#]	3.067 \pm 0.053 [#]	1.848 \pm 0.065 [#]
PL	3.247 \pm 0.044 [#]	2.037 \pm 0.048 [#]	2.603 \pm 0.042 [#]	1.898 \pm 0.066 [#]
Proposed_L	3.266 \pm 0.043 [#]	2.296 \pm 0.048 [#]	3.051 \pm 0.050 [#]	1.903 \pm 0.061 [#]
Proposed_NL	3.564 \pm 0.046	2.458 \pm 0.061	3.251 \pm 0.054	2.241 \pm 0.073

[#]The p value of paired student-t test is lower than 0.05 level.

Results of the single-sensor experiments. Considering the result shown in Fig. 8.8, Fig. 8.9, Tab. 8.11, Tab. 8.12, Tab. 8.13 and Tab. 8.14, it is found that the proposed weight map with linear model (Proposed_L) generally achieves the highest rank 1 recognition accuracy among all the methods compared in the identification task. In terms of verification performance, Proposed_NL outperforms all other methods. Also, the verification performance of Proposed_L is better than all the existing methods, although its performance is lower than Proposed_NL.

Results of the cross-sensor experiments. Considering the result shown in Fig. 8.10, Fig. 8.11, Tab. 8.15, Tab. 8.16, Tab. 8.17 and Tab. 8.18, it is found that the proposed weight map with non-linear model (Proposed_NL) outperforms all the comparison methods in both identification and verification tasks. Also, in verification task, Proposed_L performs generally better than existing methods, although the performance is lower than Proposed_NL.

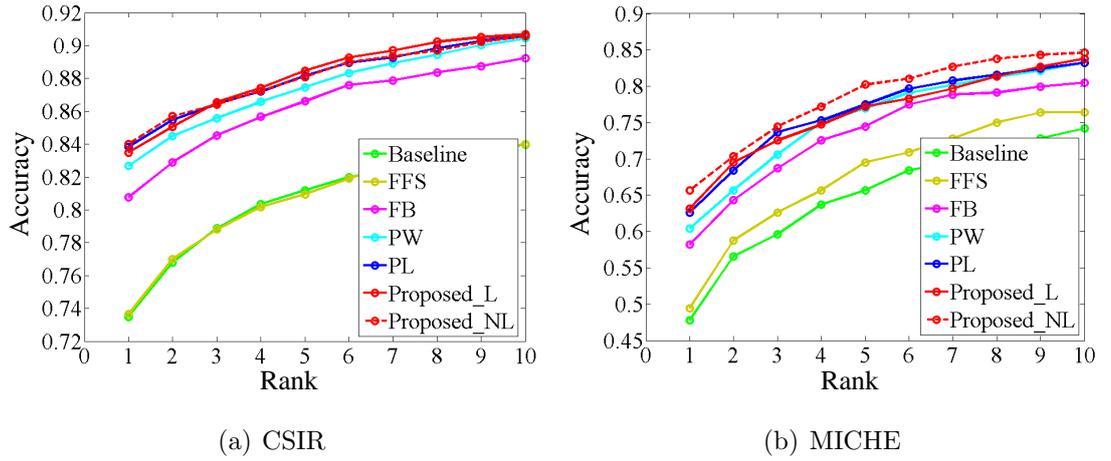


Figure 8.10: The CMC curves for comparison between proposed iris weight maps and existing iris weight map methods in the cross-sensor experiment.

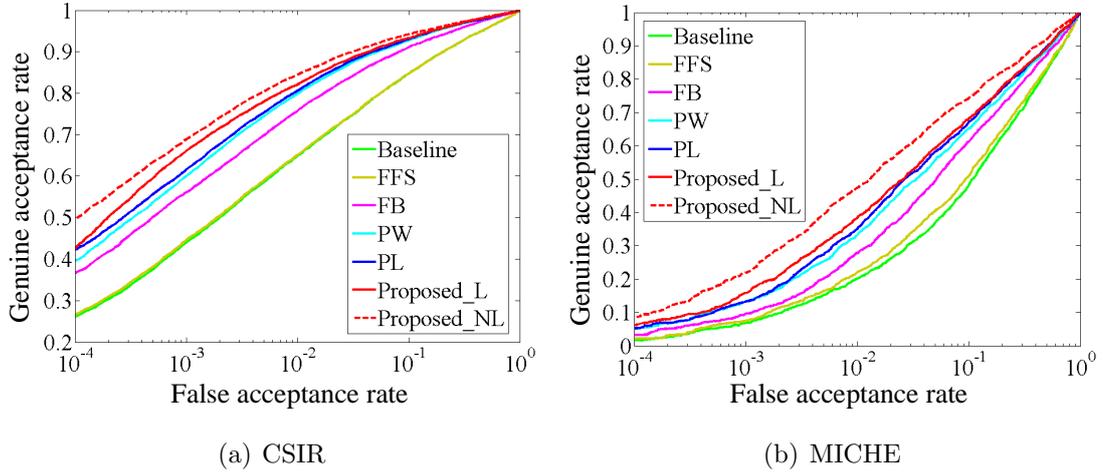


Figure 8.11: The ROC curves for comparison between proposed iris weight maps and existing iris weight map methods in the cross-sensor experiment.

Table 8.15: The R1RA for comparison in the cross-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CSIR	MICHE
Baseline	74.82 \pm 1.09 [#]	49.63 \pm 1.45 [#]
FFS	74.95 \pm 1.08 [#]	51.38 \pm 1.47 [#]
FB	81.93 \pm 0.92 [#]	59.60 \pm 1.47 [#]
PW	83.85 \pm 0.87 [#]	61.83 \pm 1.70 [#]
PL	84.84 \pm 0.86 [#]	64.21 \pm 1.65 [#]
Proposed_L	84.54 \pm 0.86 [#]	64.46 \pm 1.57 [#]
Proposed_NL	85.00 \pm 0.87	66.82 \pm 1.50

[#]The p value of paired student-t test is lower than 0.05 level.

Table 8.16: The EER for comparison in the cross-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CSIR	MICHE
Baseline	13.22 \pm 0.38 [#]	29.37 \pm 0.62 [#]
FFS	13.24 \pm 0.40 [#]	28.22 \pm 0.55 [#]
FB	9.33 \pm 0.32 [#]	24.52 \pm 0.56 [#]
PW	8.15 \pm 0.33 [#]	22.65 \pm 0.62 [#]
PL	7.93 \pm 0.33 [#]	21.63 \pm 0.58 [#]
Proposed_L	7.62 \pm 0.31 [#]	21.24 \pm 0.56 [#]
Proposed_NL	6.93 \pm 0.34	18.39 \pm 0.61

[#]The p value of paired student-t test is lower than 0.05 level.

Table 8.17: The GAR@FAR= 0.01% for comparison in the cross-sensor experiment, with the result of significance test (in percentage, best performance bolded, format: mean \pm standard deviation).

Method	CSIR	MICHE
Baseline	26.06 \pm 1.56 [#]	1.70 \pm 0.30 [#]
FFS	26.62 \pm 1.56 [#]	2.16 \pm 0.27 [#]
FB	36.64 \pm 2.28 [#]	3.31 \pm 0.81 [#]
PW	39.87 \pm 2.89 [#]	4.94 \pm 0.64 [#]
PL	42.40 \pm 2.30 [#]	5.20 \pm 0.72 [#]
Proposed_L	42.41 \pm 3.47 [#]	5.96 \pm 0.93 [#]
Proposed_NL	49.33 \pm 2.26	8.76 \pm 1.01

[#]The p value of paired student-t test is lower than 0.05 level.

Table 8.18: The DI for comparison in the cross-sensor experiment, with the result of significance test (best performance bolded, format: mean \pm standard deviation).

Method	CSIR	MICHE
Baseline	2.102 \pm 0.036 [#]	1.054 \pm 0.023 [#]
FFS	2.106 \pm 0.036 [#]	1.105 \pm 0.025 [#]
FB	2.516 \pm 0.040 [#]	1.328 \pm 0.028 [#]
PW	2.738 \pm 0.049 [#]	1.462 \pm 0.031 [#]
PL	2.768 \pm 0.049 [#]	1.523 \pm 0.033 [#]
Proposed_L	2.753 \pm 0.047 [#]	1.535 \pm 0.033 [#]
Proposed_NL	2.962 \pm 0.055	1.774 \pm 0.039

[#]The p value of paired student-t test is lower than 0.05 level.

Overall conclusions on comparison results. Considering the above results in both single-sensor and cross-sensor experiments, the concluding remarks are as follows.

- In the single-sensor experiment, the proposed linear model performs generally better than the existing iris weight map methods in both identification and verification tasks, while the proposed non-linear model is able to achieve even better verification performances than the linear model.
- In the cross-sensor experiment, the proposed non-linear model consistently outperforms the linear model and existing iris weight map methods.

8.3.6 Analysis of computational cost

This subsection analyses the computational cost of the proposed method. Comparisons of the computational cost are made between the proposed method and existing iris weight map methods, including FFS [99], FB [97], PW [98], PL [22].

The computational cost is evaluated by the mean elapsed time to calculate an iris weight map for an eye, given unwrapped iris images and the corresponding iris codes. For FB, PW and PL, this is the time to calculate the stability map (t_S); for FFS, this is the time to calculate the discriminability map (t_D); for the proposed method, this includes the time to calculate the stability map (t_S), discriminability map (t_D) and meta-code (t_M). Tab. 8.19 reports the computational cost of all the comparison methods. The results in Tab. 8.19 are measured based on a Matlab implementation of all the methods, running on a desktop with Intel i5-3470 quad-core 3.20GHz CPU, 16GB RAM, Windows 7 64bit system and Matlab 2013a 64bit. The results are based on the single-sensor configuration in Section 8.3.2, and the results are similar under the cross-sensor configuration.

Table 8.19: Analysis of computational cost (ms).

Dataset	FFS (t_D)	FB (t_S)	PW (t_S)	PL (t_S)	Proposed_L ($t_S+t_D+t_M$)	Proposed_NL ($t_S+t_D+t_M$)
CASIAD	4.7	3.0	2.9	5.3	7.7 + 3.9 + 148.1	10.1 + 3.9 + 148.6
UBIRIS2	3.7	2.5	2.4	4.2	6.2 + 3.4 + 128.6	8.0 + 3.4 + 123.8
CSIR	4.4	3.0	2.8	5.2	7.3 + 3.8 + 176.4	9.6 + 3.8 + 165.2
MICHE	4.5	3.1	2.9	5.3	7.6 + 3.9 + 222.8	10.0 + 4.0 + 221.7

It can be seen that the proposed method has higher computational cost than existing iris weight map methods. This is because the proposed method needs to compute more components than existing iris weight map methods. Specifically, existing iris weight map methods compute either a stability map (FB, PW, PL) or a discriminability map (FFS), while the proposed method computes a stability map, a discriminability map together with meta-codes. The computational cost of each component of the proposed method is analysed as follows. On the one hand, the proposed stability and discriminability maps do not induce too much computational cost. It can be seen that, in the proposed method, t_D is similar to the time to calculate existing iris weight maps, while t_S is slightly higher than the time to calculate existing iris weight maps. On the other hand, the calculation of meta-code accounts for the heaviest computational cost in the proposed method. It can be seen that t_M is significantly larger than t_S and t_D for Proposed_L and Proposed_NL. In all, it can be concluded that the proposed method has a higher computational cost than existing iris weight maps, and the additional computational cost is mainly induced by the calculation of meta-code.

However, despite of a higher computational cost compared to existing methods, the proposed method is still efficient. As shown in Tab. 8.19, the total time to compute the proposed iris weight map ($t_S + t_D + t_M$) is generally lower than 250ms. In other words, under the experimental configuration in this section, the proposed iris weight map can be computed in less than 1/4 seconds for each eye. Since the calculation of iris weight map only happens once for each eye, this speed is considered efficient.

8.4 Summary

This chapter proposes and investigates a novel iris weight map method for robust iris matching in less constrained environments. The proposed method models intra-class bit stability and inter-class bit discriminability individually to produce two weight maps: a stability map and a discriminability map. The final iris weight map is the combination of the two maps. Furthermore, the stability map is calculated using two models: a linear model and a non-linear model. It leads to two final iris weight maps: one using the linear stability model and the other using the non-linear stability model. Experiments are conducted using both single-sensor and cross-sensor data captured in less constrained environments, and the performance of the proposed iris weight map is evaluated in two tasks: identification and verification. The analysis focuses on: (1) the performance of each individual weight map in the proposed method; (2) the comparison of performance between the proposed iris weight map and existing iris weight map methods; (3) the computational cost of the proposed iris weight map method. The main findings are summarised as follows:

- For both proposed linear and non-linear models, the combination of stability map and discriminability map leads to a generally improved performance in comparison to each individual map.
- For single-sensor iris recognition, the proposed linear model generally outperforms the existing iris weight map methods in both identification and verification tasks, while the proposed non-linear model is able to achieve an even better verification performance than the linear model.
- For cross-sensor iris recognition, the proposed non-linear model consistently outperforms the linear model and existing iris weight map methods.
- The proposed iris weight map can be computed efficiently, although it requires higher computational cost than the existing iris weight map methods.

Finally, since the experimental observations show that the identification performance of proposed linear and non-linear models vary in single-sensor and cross-sensor experiments, future research may consider a model which is able to achieve a more consistent identification performance on both single-sensor and cross-sensor captures.

Chapter 9

Conclusion

9.1 Summary of thesis achievements

This thesis addresses the challenge of less constrained iris recognition with the subject at-a-distance and on-the-move. Five methods are proposed to improve the performance of different stages in less constrained iris recognition. Comprehensive experimental analysis are performed for each method. The results demonstrate that the proposed methods have achieved generally improved performance in comparison to state-of-the-art methods. Detailed achievements are as follows:

- A robust iris segmentation algorithm is proposed for less constrained iris captures. This algorithm formulates iris segmentation as robust ℓ_1 -norm regression using three models: a circle model, a parametric ellipse model and an ellipse model. To further improve the robustness, a model selection method is proposed to select the best model as the final segmentation result, based on a ring-shaped region around the segmentation boundary sought by each model. Comprehensive experimental analysis are made for each individual model as well as the model selection method. The comparison with state-of-the-art algorithms demonstrates that the proposed algorithm achieves a generally im-

proved performance.

- Regional features are investigated for iris liveness detection. Regional features exploit the information not only from low level features, but also from high level feature distribution. A spatial pyramid model and a relational measure model are employed to express the high level feature distribution. The constructed regional features based on the two models are fused at the score level to make the final decision. Experiments are conducted on four benchmark datasets for iris liveness detection. The results show that regional features are able to achieve a generally improved performance compared to traditional low level features. It is also demonstrated that, in three of the four datasets used in the experiment, regional feature based iris liveness detection method is able to achieve a comparable performance to state-of-the-art methods with accurate iris localisation, proper preprocessing and reliable feature selection; in a dataset with heavier noise, regional feature based method is able to achieve better performance than state-of-the-art methods, even with some error induced by the above three components.
- A signal level information fusion method is proposed to mitigate the noise in less constrained iris captures to improve the recognition performance. This method is based on a sparse-error low rank matrix factorisation (SE-LRMF) model. Given multiple noisy iris captures, SE-LRMF is used to separate noiseless iris structures and noise for information fusion. The noiseless iris structures are fused using a low rank mean (LRM) method, while the noise is used for an error weighted mean (EWM) fusion method. The results of LRM and EWM are combined at the code level to produce a final iris code. Experiments are conducted on benchmark iris datasets captured under varying less constrained environments. Experimentally, the proposed method demonstrates an improved performance compared to existing information fusion methods as well as recently proposed methods without information fusion, especially for

the datasets with heavier noise.

- The method to generate iris codes is investigated from the perspective of optimisation. It is demonstrated that the traditional iris code generation method can be expressed as the solution of an optimisation problem. Therefore, it is possible to apply additional objective terms to this optimisation problem to produce more effective iris codes. This thesis designs an additional objective term exploiting the spatial correlation of the bits in an iris code. The experimental results on benchmark datasets demonstrate that the iris code produced with this objective term leads to a generally improved performance in comparison to the iris code produced by an implementation of traditional iris code generation method [15], and the computational cost is acceptable in real applications. Also, based on this objective term, the spatial correlation of iris textures demonstrated in literature like [6, 31, 151] is experimentally studied and verified.
- An iris weight map method is investigated for robust iris matching. This weight map is the combination of two weight maps: a stability map modelling the intra-class bit stability and a discriminability map expressing the inter-class bit discriminability. Furthermore, the stability map is calculated using two models: a linear model and a non-linear model. This leads to two final iris weight maps: one using the linear stability model and the other one using the non-linear stability model. Experiments are conducted using both single-sensor and cross-sensor datasets. The results demonstrate that both models achieve generally improved performance in comparison to existing iris weight map methods for iris matching.

9.2 Summary of findings

The achievements in this thesis lead to several novel directions to improve the performance of less constrained iris recognition, summarised as follows:

- More robust models to overcome noise. This thesis finds that more robust models, such as the ℓ_1 -norm regression in Chapter 4 and low rank model in Chapter 6, are able to improve the performance of less constrained iris recognition.
- Explicitly modeling the noise to mitigate its influence. This thesis finds that, by explicitly modelling the noise, the influence of noise can be mitigated and the performance of less constrained iris recognition can be improved. This is demonstrated by the model and experimental results in Chapter 6.
- Exploiting high level information. Chapter 5 of this thesis finds that high level information is able to improve the performance of iris liveness detection, especially for a dataset with heavier noise. Accordingly, it may be possible to seek high level information to construct more effective features for the other stages of less constrained iris recognition.
- Seeking information from multiple captures. This thesis finds that seeking information from multiple captures is able to improve the performance of less constrained iris recognition, as demonstrated by the methods in Chapter 6 and Chapter 8.
- Incorporating more information into the model. This thesis finds that the performance of less constrained iris recognition can be improved by incorporating more information into the model. For example, the model in Chapter 7 incorporates the prior information of iris textural correlation; the model in

Chapter 8 incorporates the information of bit discriminability; both models achieve improved performance.

9.3 Future work

Future work may be conducted from the following perspectives to further improve the performance of less constrained iris recognition. First, better performance may be achieved by fusing iris with some other modalities. Although it is generally believed that combining a strong biometric like iris with a weak one leads to limited improvement, some recent research has shown that, in less constrained environment, an improved performance can be achieved by fusing iris with other modalities such as periocular [23, 154–156] and face [112, 157]. Therefore, it is possible to seek to fuse iris with other biometric modalities to improve the performance of iris recognition in less constrained environments.

Second, algorithms may be designed to explicitly model the noise to alleviate its influence. As shown in Chapter 6, a model for information fusion with a term explicitly modelling the noise is able to achieve a generally improved performance. Therefore, similar approaches may be considered for other stages of less constrained iris recognition to improve the robustness.

Third, more robust models may be considered to achieve improvements. As shown in Chapter 4, the performance of iris segmentation can be improved by formulating the problem using robust ℓ_1 -norm regression. Similarly, the model in Chapter 6 also relies on the robustness of ℓ_1 -norm. Thus, it is possible to introduce other robust functions such as ℓ_p -norm ($p < 1$) [158, 159], $\ell_{2,1}$ -norm [160, 161] and Huber loss [162] into the problem formulation to achieve improvements for different stages in less constrained iris recognition.

Four, high level information may be exploited to improve less constrained iris recog-

dition. As shown in Chapter 5, features including high level feature distribution information achieve generally improve performance for iris liveness detection, compared to low level features directly extracted from pixels. Since most existing features for iris recognition are still low level features [6, 7, 10, 15, 20, 23, 81, 83], it is expected that features with high level information will improve state-of-the-art performance. Especially, deep learning technique based on deep neural networks [76, 84, 163, 164] may have high potential to be used to exploit high level information for less constrained iris recognition.

Five, textural correlation in iris captures may be studied to improve iris recognition. As shown in Chapter 7, the performance of iris recognition is generally improved with a model considering the iris textural correlation to produce iris codes. Future research may consider a comprehensively quantitative analysis on iris textural correlation, in order to seek the direction with higher discriminating power to assist parameter selection and to design more powerful features.

Six, information may be exploited from multiple iris captures to improve the robustness of less constrained iris recognition. In the thesis, the methods in Chapter 6 and Chapter 8 seek the information that are noiseless or robust to noise from multiple noisy captures, leading to improved performance compared to state-of-the-art methods. Future research may focus on more effective models to seek information from multiple captures, in order to achieve a more robust performance.

Bibliography

- [1] A. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 4–20, 2004.
- [2] K. Delac and M. Grgic, “A survey of biometric recognition methods,” in *Proceedings of the 46th International Symposium on Electronics in Marine*, pp. 184–193, 2004.
- [3] A. Jain, P. Flynn, and A. Ross, eds., *Handbook of biometrics*. Springer, 2008.
- [4] W. Bennett and K. Hess, *Criminal investigation*. Cengage Learning, 8 ed., 2006.
- [5] M. Marsico, C. Galdi, M. Nappi, and D. Riccio, “FIRME: face and iris recognition for mobile engagement,” *Image and Vision Computing*, vol. 32, pp. 1161–1172, 2014.
- [6] J. Daugman, “High confidence visual recognition of persons by a test of statistical independence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1148–1161, 1993.
- [7] J. Daugman, “How iris recognition works,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 21–30, 2004.
- [8] R. Wildes, “Iris recognition: an emerging biometric technology,” *Proceedings of the IEEE*, vol. 85, pp. 1348–1363, 1997.

- [9] K. Bowyer, K. Hollingsworth, and P. Flynn, "Image understanding for iris biometrics: A survey," *Computer Vision and Image Understanding*, vol. 110, pp. 281–307, 2008.
- [10] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Personal identification based on iris texture analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1519–1533, 2003.
- [11] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Efficient iris recognition by characterizing key local variations," *IEEE Transactions on Image Processing*, vol. 13, pp. 739–750, 2004.
- [12] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," *Pattern Recognition*, vol. 43, pp. 1016–1026, 2010.
- [13] J. Daugman and C. Downing, "Epigenetic randomness, complexity and singularity of human iris patterns," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 268, pp. 1737–1740, 2001.
- [14] J. Daugman, "600 million citizens of India are now enrolled with biometric id," *SPIE Newsroom*, 2014.
- [15] L. Masek and P. Kovesi, "Matlab source code for a biometric identification system based on iris patterns," *The School of Computer Science and Software Engineering, The University of Western Australia*, 2003.
- [16] P. Li and G. Wu, "Iris recognition using ordinal encoding of log-euclidean covariance matrices," in *Proceedings of the 21st International Conference on Pattern Recognition*, pp. 2420 – 2423, 2012.
- [17] J. Pillai, V. Patel, R. Chellappa, and N. Ratha, "Secure and robust iris recognition using random projections and sparse representations," *IEEE Transac-*

- tions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 1877–1893, 2011.
- [18] H. Proença and L. Alexandre, “Toward convert iris biometric recognition: experimental results from the nice contests,” *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 798–808, 2012.
- [19] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. Alexandre, “The ubiris.v2: a database of visible wavelength images captured on-the-move and at-a-distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1529–1535, 2010.
- [20] Z. Sun and T. Tan, “Ordinal measures for iris recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2211–2226, 2009.
- [21] C. Tan and A. Kumar, “Unified framework for automated iris segmentation using distantly acquired face images,” *IEEE Transactions on Image Processing*, vol. 21, pp. 4068–4079, 2012.
- [22] C. Tan and A. Kumar, “Adaptive and localized iris weight map for accurate iris recognition under less constrained environment,” in *Proceedings of the 6th IEEE International Conference on Biometrics: Theory, Applications and Systems*, pp. 1–7, 2013.
- [23] C. Tan and A. Kumar, “Accurate iris recognition at a distance using stabilized iris encoding and zernike moments phase features,” *IEEE Transactions on Image Processing*, vol. 23, pp. 3962–3974, 2014.
- [24] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W. Zhao, “Iris on the move: acquisition of images for iris recognition in less constrained environments,” *Proceedings of the IEEE*, vol. 94, pp. 1936–1947, 2006.

- [25] S. Venugopalan, U. Prasad, K. Harun, K. Neblett, D. Toomey, J. Heyman, and M. Savvides, “Long range iris acquisition system for stationary and mobile subjects,” in *Proceedings of International Joint Conference on Biometrics*, pp. 1–8, 2011.
- [26] K. Hollingsworth, T. Peters, K. Bowyer, and P. Flynn, “Iris recognition using signal-level fusion of frames from video,” *IEEE Transactions on Information Forensics and Security*, vol. 4, pp. 837–848, 2009.
- [27] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman, “Focus-score weighted super-resolution for uncooperative iris recognition at a distance and on the move,” in *Proceedings of the 25th International Conference on Image and Vision Computing*, pp. 1–8, 2010.
- [28] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman, “Quality-driven super-resolution for less constrained iris recognition at a distance and on the move,” *IEEE Transactions on Information Forensics and Security*, vol. 6, pp. 1248–1258, 2011.
- [29] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman, “Feature-domain super-resolution for iris recognition,” in *Proceedings of the 18th IEEE International Conference on Image Processing*, pp. 3197–3200, 2011.
- [30] K. Nguyen, S. Sridharan, S. Denman, and C. Fookes, “Feature-domain super-resolution framework for gabor-based face and iris recognition,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2642–2649, 2012.
- [31] J. Liu, Z. Sun, and T. Tan, “Code-level information fusion of low-resolution iris image sequences for personal identification at a distance,” in *Proceedings of the 6th IEEE International Conference on Biometrics: Theory, Applications and Systems*, pp. 1–6, 2013.

- [32] N. Kalka, J. Zuo, N. Schmid, and B. Cukic, “Estimating and fusing quality factors for iris biometric images,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, pp. 509–524, 2010.
- [33] P. Hough, “Method and means for recognizing complex patterns.” U.S. Patent 3069654, 1962.
- [34] J. Daugman, “New methods in iris recognition,” *IEEE Transactions on Systems, Man and Cybernetics, part B*, vol. 37, pp. 1167–1175, 2007.
- [35] S. Shah and A. Ross, “Iris segmentation using geodesic active contours,” *IEEE Transactions on Information Forensics and Security*, vol. 4, pp. 824–836, 2009.
- [36] Z. He, T. Tan, Z. Sun, and X. Qiu, “Toward accurate and fast iris segmentation for iris biometrics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1670–1684, 2009.
- [37] H. Young and R. Freedman, *University Physics with Modern Physics*. Addison-Wesley, 2003.
- [38] T. Tan, Z. He, and Z. Sun, “Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition,” *Image Vision Computing*, vol. 28, pp. 223–230, 2010.
- [39] H. Proença, “Iris recognition: on the segmentation of degraded images acquired in the visible wavelength,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1502–1516, 2010.
- [40] C. Teh and R. Chin, “On image analysis by the methods of moments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 496–513, 1998.
- [41] C. Tan and A. Kumar, “Towards online iris and periocular recognition under relaxed imaging constraints,” *IEEE Transactions on Image Processing*, vol. 22, pp. 3751–3765, 2013.

- [42] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1768–1783, 2006.
- [43] M. Kanematsu, H. Takano, and K. Nakamura, "Highly reliable liveness detection method for iris recognition," in *Proceedings of SICE Annual Conference*, pp. 361–364, 2007.
- [44] X. He, S. An, and P. Shi, "Statistical texture analysis-based approach for fake iris detection using support vector machines," in *Proceedings of International Conference on Biometrics*, vol. 4642, pp. 540–546, 2007.
- [45] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, pp. 610–621, 1973.
- [46] Z. Wei, X. Qiu, Z. Sun, and T. Tan, "Counterfeit iris detection based on texture analysis," in *Proceedings of the 19th International Conference on Pattern Recognition*, pp. 1–4, 2008.
- [47] X. He, Y. Lu, and P. Shi, "A new fake iris detection method," in *Proceedings of International Conference on Biometrics*, vol. 5558, pp. 1132–1139, 2009.
- [48] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, pp. 909–996, 1988.
- [49] J. Galbally, J. Ortiz-Lopez, J. Fierrez, and J. Ortega-Garcia, "Iris liveness detection based on quality related features," in *Proceedings of the 5th IAPR International Conference on Biometrics*, pp. 271–276, 2012.
- [50] Z. Wei, T. Tan, Z. Sun, and J. Cui, "Robust and fast assessment of iris image quality," in *Proceedings of International Conference on Biometrics*, vol. 3832, pp. 464–471, 2006.

- [51] A. Abhyankar and S. Schukers, “Iris quality assessment and bi-orthogonal wavelet based encoding for recognition,” *Pattern Recognition*, vol. 42, pp. 1878–1894, 2009.
- [52] N. Kalka, J. Zuo, N. Schmid, and B. Cukic, “Image quality assessment for iris biometric,” in *Proceedings of SPIE 6202, Biometric Technology for Human Identification III*, 2006.
- [53] J. Zuo and N. Schmid, “Image quality assessment for iris biometric,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2009.
- [54] Y. Chen, S. Dass, and A. Jain, “Localized iris image quality using 2-d wavelets,” in *Proceedings of International Conference on Biometrics*, vol. 3832, pp. 373–381, 2006.
- [55] Y. Du, C. Belcher, Z. Zhou, and R. Ives, “Feature correlation evaluation approach for iris feature quality measure,” *Signal Processing*, vol. 90, pp. 1176–1187, 2010.
- [56] P. Pudil, J. Novovicova, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recognition Letters*, vol. 15, pp. 1119–1125, 1994.
- [57] Z. He, Z. Sun, T. Tan, , and Z. Wei, “Efficient iris spoof detection via boosted local binary patterns,” in *Proceedings of International Conference on Biometrics*, vol. 5558, pp. 1080–1090, 2009.
- [58] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multi resolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.
- [59] R. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” *Machine Learning*, vol. 37, pp. 297–336, 1999.

- [60] I. Rigas and O. Komogortsev, “Gaze estimation as a framework for iris liveness detection,” in *Proceedings of IEEE International Joint Conference on Biometrics*, 2014.
- [61] A. Sequeira, J. Murari, and J. Cardoso, “Iris liveness detection methods in mobile applications,” in *Proceedings of International Conference on Computer Vision Theory and Applications*, 2014.
- [62] D. Gagnaniello, G. Poggi, C. Sansone, and L. Verdoliva, “An investigation of local descriptors for biometric spoofing detection,” *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 849–863, 2015.
- [63] R. Nosaka, Y. Ohkawa, and K. Fukui, “Feature extraction based on co-occurrence of adjacent local binary patterns,” in *Proceedings of the 5th Pacific Rim Symposium on Image and Video Technology*, vol. 7088, pp. 82–91, 2011.
- [64] R. Nosaka, C. Suryanto, and K. Fukui, “Rotation invariant co-occurrence among adjacent LBPs,” in *Proceedings of ACCV International Workshops*, vol. 7728, pp. 15–25, 2012.
- [65] V. Ojansivu, E. Rahtu, and J. Heikkilä, “Rotation invariant local phase quantization for blur insensitive texture analysis,” in *Proceedings of the 19th International Conference on Pattern Recognition*, 2008.
- [66] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, “WLD: a robust local image descriptor,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1705–1720, 2010.
- [67] D. Gagnaniello, G. Poggi, C. Sansone, and L. Verdoliva, “Local contrast phase descriptor for fingerprint liveness detection,” *Pattern Recognition*, vol. 48, pp. 1050–1058, 2015.

- [68] J. Kannala and E. Rahtu, “BSIF: Binarized statistical image features,” in *Proceedings of the 21st International Conference on Pattern Recognition*, pp. 1363–1366, 2012.
- [69] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.
- [70] E. Tola, V. Lepetit, and P. Fua, “DAISY: An efficient dense descriptor applied to wide-baseline stereo,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 815–830, 2010.
- [71] I. Kokkinos and A. Yuille, “Scale invariance without scale selection,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2008.
- [72] I. Kokkinos, M. Bronstein, and A. Yuille, “Dense scale invariant descriptors for images and surfaces.” Technical Report RR-7914, INRIA, 2012.
- [73] A. Czajka, “Pupil dynamics for iris liveness detection,” *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 726–735, 2015.
- [74] M. Kohn and M. Clynes, “Color dynamics of the pupil,” *Annals of the New York Academy of Sciences*, vol. 156, pp. 931–950, 1969.
- [75] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcao, and A. Rocha, “Deep representations for iris, face, and fingerprint spoofing detection,” *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 864–879, 2015.
- [76] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [77] J. Daugman, “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters,” *Journal of the Optical Society of America A*, vol. 2, pp. 1160–1169, 1985.

- [78] J. Daugman, “Complete discrete 2d gabor transforms by neural networks for image analysis and compression,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 1169–1179, 1988.
- [79] D. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *Journal of the Optical Society of America A*, vol. 4, pp. 2379–2394, 1987.
- [80] C. Tan and A. Kumar, “Efficient and accurate at-a-distance iris recognition using geometric key-based iris encoding,” *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 1518–1526, 2014.
- [81] T. Tan, X. Zhang, Z. Sun, and H. Zhang, “Noisy iris image matching by using multiple cues,” *Pattern Recognition Letters*, vol. 33, pp. 970–977, 2012.
- [82] D. Ruderman, T. Cronin, and C. Chiao, “Statistics of cone responses to natural images: implications for visual coding,” *Journal of the Optical Society of America A*, vol. 15, pp. 2036–2045, 1998.
- [83] P. Li, X. Liu, and N. Zhao, “Weighted co-occurrence phase histogram for iris recognition,” *Pattern Recognition Letters*, vol. 33, pp. 1000–1005, 2012.
- [84] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan, “DeepIris: learning pairwise filter bank for heterogeneous iris verification,” *Pattern Recognition Letters*, vol. 82, pp. 154–161, 2016.
- [85] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998.
- [86] Q. Ke and T. Kanade, “Robust l_1 norm factorization in the presence of outliers and missing data by alternative convex programming,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 739–746, 2005.

- [87] N. Kwak, “Principal component analysis based on l_1 -norm maximization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1672–1680, 2008.
- [88] C. Lu, J. Shi, and J. Jia, “Online robust dictionary learning,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 415–422, 2013.
- [89] C. Zhao, X. Wang, and W. Cham, “Background subtraction via robust dictionary learning,” *EURASIP Journal on Image and Video Processing*, vol. 2011, pp. 1–12, 2011.
- [90] B. Li, Q. Yu, R. Wang, K. Xiang, M. Wang, and X. Li, “Block principal component analysis with nongreedy l_1 -norm maximization,” *IEEE Transactions on Cybernetics*, vol. 46, pp. 2543–2547, 2016.
- [91] H. Ling and K. Okada, “Diffusion distance for histogram comparison,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 246–253, 2006.
- [92] D. Comaniciu, V. Ramesh, and P. Meer, “Real-time tracking of non-rigid objects using mean shift,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 142–149, 2000.
- [93] D. Donoho and X. Huo, “Uncertainty principle and ideal atomic decomposition,” *IEEE Transactions on Information Theory*, vol. 47, pp. 2845–2862, 2001.
- [94] D. Donoho and M. Elad, “On the stability of the basis pursuit in the presence of noise,” *Signal Processing*, vol. 86, pp. 511–532, 2006.
- [95] E. Candès, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

- [96] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 210–227, 2008.
- [97] K. Hollingsworth, K. Bowyer, and P. Flynn, “The best bits in an iris code,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 964–973, 2008.
- [98] W. Dong, Z. Sun, and T. Tan, “Iris matching based on personalized weight map,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 1744–1757, 2010.
- [99] H. Proença, “Iris recognition: What’s beyond bit fragility,” *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 321–332, 2014.
- [100] R. Duda, P. Hart, and D. Stork, *Pattern classification*. Wiley-Blackwell, 2 ed., 2008.
- [101] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and minredundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.
- [102] “Biometrics ideal test.” Online Available: <http://biometrics.idealtest.org/dbDetailForUser.do?id=4>.
- [103] P. Phillips, W. Scruggs, A. O’Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, “FRVT 2006 and ICE 2006 large-scale experimental results,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 831–846, 2010.
- [104] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “Overview of the face recognition grand challenge,” in

- Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
- [105] D. Yambay, J. S. Doyle, K. W. Bowyer, and A. Czajka, “LivDet-iris 2013 - iris liveness detection competition 2013,” in *Proceedings of IEEE International Joint Conference on Biometrics*, 2014.
- [106] “The icb competition on cross-sensor iris recognition.” Online Available: <http://biometrics.idealtest.org/2015/csir2015.jsp>.
- [107] P. Li, X. Liu, L. Xiao, and Q. Song, “Robust and accurate iris segmentation in very noisy iris images,” *Image Vision Computing*, vol. 28, pp. 246–253, 2010.
- [108] C. Tan and A. Kumar, “Efficient iris segmentation using grow-cut algorithm for remotely acquired iris images,” in *Proceedings of the 5th IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2012.
- [109] D. Gagnaniello, C. Sansone, and L. Verdoliva, “Iris liveness detection for mobile devices based on local descriptors,” *Pattern Recognition Letters*, vol. 57, pp. 81–87, 2015.
- [110] A. Sequeira, H. Oliveira, J. Monteiro, J. Monteiro, and J. Cardoso, “MobILive 2014 - mobile iris liveness detection competition,” in *Proceedings of IEEE International Joint Conference on Biometrics*, 2014.
- [111] H. Proença, “Quality assessment of degraded iris images acquired in the visible wavelength,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 82–95, 2011.
- [112] D. Miao, M. Zhang, H. Li, Z. Sun, and T. Tan, “Bin-based weak classifier fusion of iris and face biometrics,” in *Proceedings of the 7th IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2015.

- [113] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of IEEE Computer Society Conference Computer Vision and Pattern Recognition*, 2001.
- [114] E. Tsonas, "Bayesian quantile inference," *Journal of Statistical Computation and Simulation*, vol. 73, pp. 659–674, 2003.
- [115] S. Elian, C. Andre, and S. Narula, "Influence measure for the L1 regression," *Communications in Statistics - Theory and Methods*, vol. 29, pp. 837–849, 2000.
- [116] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Sisstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 34, pp. 2274–2282, 2012.
- [117] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
- [118] P. Radu, K. Sirlantzis, W. Howells, F. Deravi, and S. Hoque, "Information fusion for unconstrained iris recognition," *International Journal of Hybrid Information Technology*, vol. 4, 2011.
- [119] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, 1979.
- [120] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 372–386, 2012.
- [121] E. Candes and J. Romberg. Online Available: <http://users.ece.gatech.edu/~justin/l1magic/>.

- [122] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [123] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.
- [124] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [125] A. Nigam, V. Krishna, A. Bendale, and P. Gupta, "Iris recognition using block local binary patterns and relational measures," in *Proceedings of IEEE International Joint Conference on Biometrics*, 2014.
- [126] T. Ojala, M. Pietikainen, and T. Maenpaa, "A generalized local binary pattern operator for multi-resolution gray scale and rotation invariant texture classification," in *Proceedings of the 2nd International Conference on Advances in Pattern Recognition*, pp. 397–406, 2001.
- [127] P. Chiranjeevi and S. Sengupta, "Detection of moving objects using multi-channel kernel fuzzy correlogram based background subtraction," *IEEE Transactions on Cybernetics*, vol. 44, pp. 870–881, 2014.
- [128] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman, "Feature-domain super-resolution for iris recognition," *Computer Vision and Image Understanding*, vol. 117, pp. 1526–1535, 2013.
- [129] K. Shin, K. Park, B. Kang, and S. Park, "Super-resolution method based on multiple multi-layer perceptrons for iris recognition," in *Proceedings of*

- the 4th International Conference on Ubiquitous Information Technologies & Applications*, 2009.
- [130] J. Huang, L. Ma, T. Tan, and Y. Wang, “Learning based resolution enhancement of iris images,” in *Proceedings of British Machine Vision Conference*, 2003.
- [131] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao, “Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization,” in *Proceedings of Neural Information Processing Systems*, 2009.
- [132] X. Yuan and J. Yang, “Sparse and low rank matrix decomposition via alternating direction method,” *Optimization Online*, 2009.
- [133] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, “Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix,” in *Proceedings of the 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2009.
- [134] V. Chandrasekharan, S. Sanghavi, P. Parillo, and A. Wilsky, “Rank-sparsity incoherence for matrix decomposition,” *SIAM Journal on Optimization*, vol. 21, pp. 572–596, 2011.
- [135] Y. Mu, J. Dong, X. Yuan, and S. Yan, “Accelerated low-rank visual recovery by random projection,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
- [136] E. Candes, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?,” *Journal of the ACM*, vol. 58, 2011.
- [137] D. Meng, Z. Xu, L. Zhang, and J. Zhao, “A cyclic weighted median method for l1 low-rank matrix factorization with missing entries,” in *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, 2013.

- [138] T. Okatani and K. Deguchi, “On the Wiberg algorithm for matrix factorization in the presence of missing components,” *International Journal of Computer Vision*, vol. 72, pp. 329–337, 2007.
- [139] N. Srebro and T. Jaakkola, “Weighted low-rank approximations,” in *Proceedings of the 20th International Conference on Machine Learning*, 2003.
- [140] A. Buchanan and A. Fitzgibbon, “Damped newton algorithms for matrix factorization with missing data,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 316–322, 2005.
- [141] T. Wiberg, “Computation of principal components when data are missing,” in *Proceedings of the 2nd Symposium on Computational Statistics*, pp. 229–236, 1976.
- [142] A. Eriksson and A. van den Hengel, “Efficient computation of robust low-rank matrix approximations in the presence of missing data using the l1 norm,” in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 771–778, 2010.
- [143] C. Ding, D. Zhou, X. He, and H. Zha, “R1-PCA: Rotational invariant l1-norm principal component analysis for robust subspace factorization,” in *Proceedings of the 23rd International Conference on Machine learning*, 2006.
- [144] E. Hale, W. Yin, and Y. Zhang, “Fixed-point continuation for l1-minimization: methodology and convergence,” *SIAM Journal on Optimization*, vol. 19, pp. 1107–1130, 2008.
- [145] K. Roy and P. Bhattacharya, “Optimal features subset selection and classification for iris recognition,” *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–20, 2008.

- [146] M. Pereira and A. Veiga, "Application of genetic algorithms to improve the reliability of an iris recognition system," in *Proceedings of IEEE Workshop Machine Learning for Signal Processing*, 2005.
- [147] K. Chen, C. Chou, S. Shih, W. Chen, and D. Chen, "Feature selection for iris recognition with adaboost," in *Proceedings of the 3rd International Conference Intelligent Information Hiding and Multimedia Signal Processing*, 2007.
- [148] Z. He, Z. Sun, T. Tan, X. Qiu, C. Zhong, and W. Dong, "Boosting ordinal features for accurate and fast iris recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [149] L. Wang, Z. Sun, and T. Tan, "Robust regularized feature selection from iris recognition via linear programming," in *Proceedings of the 21st International Conference on Pattern Recognition*, 2012.
- [150] Z. Sun, L. Wang, and T. Tan, "Ordinal feature selection for iris and palmprint recognition," *IEEE Transactions on Image Processing*, vol. 23, pp. 3922–3934, 2014.
- [151] J. Daugman, "Information theory and the IrisCode," *IEEE Transactions on Information Forensics and Security*, vol. 11, pp. 400–409, 2016.
- [152] Y. Boykov, O. Veksler, and R. Zabih, "Markov random fields with efficient approximations," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998.
- [153] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximation energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222–1239, 2001.
- [154] H. Proença, "Ocular biometrics by score-level fusion of disparate experts," *IEEE Transactions on Image Processing*, vol. 23, pp. 5082–5093, 2014.

- [155] G. Santos, E. Grancho, M. Bernardo, and P. Fiadeiro, “Fusing iris and periocular information for cross-sensor recognition,” *Pattern Recognition Letters*, vol. 57, pp. 52–59, 2015.
- [156] I. Nigam, M. Vatsa, and R. Singh, “Ocular biometrics: A survey of modalities and fusion approaches,” *Information Fusion*, vol. 26, pp. 1–35, 2015.
- [157] H. Sim, H. Asmuni, R. Hassan, and R. Othman, “Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images,” *Expert Systems with Applications*, vol. 41, pp. 5390–5404, 2014.
- [158] F. Seidel, C. Hage, and M. Kleinsteuber, “pROST: A smoothed lp-norm robust online subspace tracking method for background subtraction in video,” *Machine Vision and Applications*, vol. 25, pp. 1227–1240, 2014.
- [159] Y. Chen and H. Sun, “ ℓ_p -norm based iterative adaptive approach for robust spectral analysis,” *Machine Vision and Applications*, vol. 94, pp. 144–148, 2014.
- [160] F. Nie, H. Huang, X. Cai, and C. Ding, “Efficient and robust feature selection via joint $\ell_{2,1}$ -norms minimization,” in *Proceedings of Neural Information Processing Systems*, 2010.
- [161] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, “Robust recovery of subspace structures by low-rank representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 171–184, 2013.
- [162] P. Huber, “Robust estimation of a location parameter,” *The Annals of Mathematical Statistics*, vol. 35, pp. 73–101, 1964.
- [163] I. Goodfellow, Y. Bengio, and A. Courville, “Deep learning.” Book in preparation for MIT Press, 2016.
- [164] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.