Kent Academic Repository

Catley, Edward, Sirlantzis, Konstantinos, Kelly, Stephen W. and Howells, Gareth (2014) *Non-overlapping dual camera fall detection using the NAO humanoid robot.* **In: 5th International Conference on Emerging Security Technologies, 10-12 September 2014, Alcala de Henares, Spain.**

Non-overlapping dual camera fall detection using the NAO humanoid robot

E.Catley, K. Sirlantzis, G. Howells, S. Kelly
Image and Information Engineering Research Group
School of Engineering and Digital Arts, University of Kent, Kent CT2 7NT, UK.
Email: EC364@kent.ac.uk

Abstract. With an aging population and a greater desire for independence, the dangers of falling incidents in the elderly have become particularly pronounced. In light of this, several technologies have been developed with the aim of preventing or monitoring falls. Failing to strike the balance between several factors including reliability, complexity and invasion of privacy has seen prohibitive in the uptake of these systems. Some systems rely on cameras being mounted in all rooms of a users home while others require being worn 24 hours a day. This paper explores a system using a humanoid NAO robot with dual vertically mounted cameras to perform the task of fall detection.

I. INTRODUCTION

It has been reported that one in three over 65s fall each year and that falls are the leading cause of both fatal and non-fatal injuries [1]. In the UK alone there are currently 10 million of 65's with this number expected to double by 2050. As part of the COALAS (Interreg IVA) project, work has been undertaken into providing minimally intrusive in home care for the elderly and disabled with one of the aspects being fall detection. While other work undertaken in this field has had some success with wearable technology such as accelerometers and gyroscopes, multiple camera systems for house-wide coverage or single camera systems for localised coverage, these implementations generally suffer from certain drawbacks.

Wearable technology exists in the form of fall alarms and/or detectors. In the case of the former this is only capable of sending out a manually activated distress signal which after a fall, depending on the seriousness, the user may be incapable of activating it. In both cases one of the most limiting factors is the intrusiveness of the system. Studies have found that the device is not worn at all times either deliberately, due to the discomfort brought, or unintentionally if forgotten after temporarily removing [2].

More recently there has been research into systems using either one or more cameras and image processing algorithms. These systems have benefitted from being relatively non-invasive but have problems related to limited field of view, reliability and acceptance among the target demographic.

This paper proposes a system that, in conjunction with other functions, allows for reliable fall detection and verification using the RGB cameras of one of Aldebaran's NAO humanoid robot. This robot enjoys 25 degrees of freedom, two 960p 30fps non-overlapping cameras with parallel fields of view, 4 microphones for directional audio and wireless connectivity. The ability to walk and the two vertically mounted cameras make this robot ideal for research that involves human observation. Where other systems which utilise cameras have relied upon static cameras, which even in multi-camera systems can result in blind spots, occlusion or poor image quality at distance, this systems concept allows for a great number of camera positions and angles due to the mobile nature of the platform.

In this paper shall be discussed the implemented first half of the envisaged system. The NAO is used as a stationary observer with, relative to the literature, simplistic algorithms which make use of its vertically mounted dual camera configuration.

II. RELATED WORK

Of the most recent related literature the work generally falls into two categories, those that use RGB-D data taken from sensors such as the Microsoft Kinect [3] [4] and those that rely on RGB data alone. While RGB-D approaches generally provide very accurate fall detection and appear to be a promising line of research, computational requirements and issues with IR sunlight saturation remain a prohibitive factor for widespread uses.

Methods using RGB data alone are more widespread and are themselves clustered into those that monitor for abnormal events [5] [6] and those that explicitly attempt to detect falls [7] [8] [9]. The former generally have to be taught what "normal" behaviour looks like so as to identify that which is not. The limitation with this is that the system often needs to be retrained upon repositioning of the camera. The latter on the other hand present a more refined method for specifically detecting falls but by using fixed cameras can suffer from occlusion, poor image quality and gaps in field of view. Because the cameras are usually mounted in the top corners of rooms or in the centre of the ceiling with fish eye lenses, the image quality is often lacking making the detection of static falls difficult due to a lack of features. The methodology used in the single wide angled lens camera approach of [8] results in a false positive rate of 0.31 and false negative of 0.22. This method uses the difference in angle of the body axes between a standing person which, due to the nature and position of the camera, will be

pointing towards the centre of the camera and that of a fallen person. The limitations of this system are that the difference in angle between the body axes of standing and lying must be ≥ 28° and it cannot distinguish between people lying on a sofa for example, and the floor. The single camera approach of [9] conducted separate experiments in cluttered and uncluttered environments. A mean average precision (MAP) of 89% and 90% was achieved in cluttered and uncluttered environments respectively.

The most pertinent paper [7] details the use of multiple cameras to allow the capture of falls from several angles, allowing the camera with the best view of the fall and hence the most reliable result, to make the call as to whether a fall has occurred.

A stage of region of interest location is used by a process of foreground detection through background subtraction, shadow removal and region of interest extraction using the largest foreground object of a size of no less the 17500 pixels out of the captured 640*480 image. A combination of four features are used to make this decision; fall angle, head speed, aspect ratio and centre of gravity.

By using a reasonably fast dual core processor running up to four threads (one for a different video stream) a processing rate of 8 frames a second was achieved.

An incredibly detailed and thorough data set was used, comprising of over 14,000 hours of video recording a total of 24 falls taken from within the homes of participants aged 83 to 95 year olds. The thoroughness of the data sets this paper apart from the rest, which mostly use simulated fall data sets.

This system proved to be quite good at detecting falls but was also prone to false positives. Furthermore it produced three false negatives in the 24 falls, though two of these three were due to partial occlusion due to camera positioning. Using an AUC method for a plot of sensitivity and specificity yields a highest score of 0.91 ±0.06 using the features of head speed and aspect ratio.

To conclude, the biggest drawback of all these systems lies in the availability of useful images. Because all systems make use of a limited number of static cameras, to maximise field of view they are positioned high in the room. Due to this there is a limitation on the quality of image at greater distance and availability of features to work with. This results in limited reliability of the system or excessive processing time.

III. METHODOLOGY

The overall fall detection system being implemented falls into three main stages; Person detection, preliminary fall detection and fall confirmation. In this paper we present only the implemented person detection and preliminary fall detection. The positive fall detection criteria have been made purposefully large in an effort to minimise the number of false negatives which, once implemented, will be corrected in the confirmation stage. Figure 1 illustrates the structure of this system showing the sections covered in this paper (dark) and those left for future work (light).

FIGURE 1
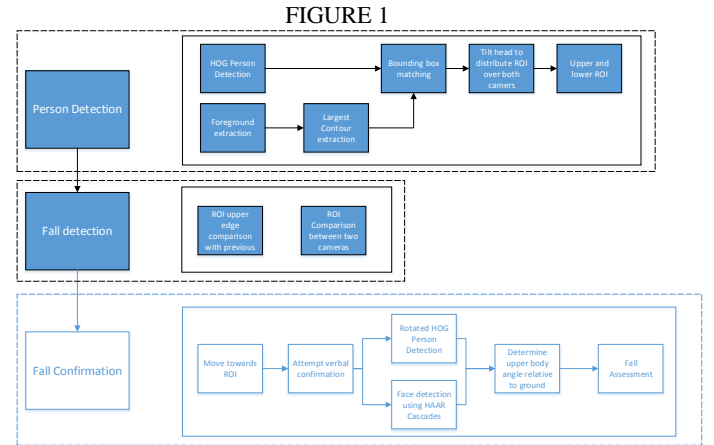
A. Person Detection

1) HOG Person Detection

Due to constraints on processing power and real time application the histogram of gradients (HOG) is used only at the initialisation of the system to confirm that the foreground and contour extraction has successfully identified a subject.

2) Foreground extraction

A method of extracting the foreground from the background is used which also ignores shadows. Due to varying light conditions there is a reasonable amount of noise introduced in the images an adaptive background is required, which does a very effective job of negating this problem. The drawback of this is that, if remaining static for long, the foreground is absorbed into the background meaning that the person being tracked can be lost. If this happens the system will re-initialise.

3) Largest contour extraction from foreground

To extract the person from this foreground image the largest contour is found, in almost all cases the largest contour is the subject that we wish to track. Save for other people or pets, there is unlikely to be any other foreground objects and thus few other contours.

4) Bounding box matching

To ensure that the contour extracted from the foreground is a person a simple method of comparing the bounding boxes from stages 1 and 3 to ensuring that they overlap occurs, with a certain margin of error allowed for slight variances in detection size. If this match is made then the system proceeds without needing to continue the resource and time draining person detection HOG. Should the tracked contour ever be lost (due to full occlusion or other system failure) this system will re-initialise.

B. Fall detection

By using both of the NAOs cameras, each with a region of interest focused around the upper and lower body, it is fair to assume that, in the case of an actual fall, the ROI height in the top will decrease rapidly over time and finally disappear while the ROI on the bottom will decrease over

time (with a slight delay after the top has disappeared) but not disappear entirely.

With the regions of interest found, a record of the value of the top edge is stored in a vector spanning the last 2 seconds, the longest period of time that we found a fall event to take.

For the preliminary fall detection phase to return a positive result four criteria must be met; The top of the upper ROI must decrease by more than 50% of the original in a second, the top ROI must disappear, the top of the bottom region of interest must change by 30% a second and must not disappear.

Should a positive result be returned by the preliminary fall detection stage then the system will continue on to the fall confirmation stage which lies outside the scope of this paper. The advantage of this method is that it is fast and not overly taxing on system resources making the chances of "missing" the fall due to slow ROI acquisition. Furthermore the use of two vertically mounted cameras allows for an extended field of view allowing the whole user to be tracked, even at short distances which tend to be the norm in the cluttered smaller houses of the elderly.

IV. RESULTS

To try and maintain a certain level of consistency and comparability with other papers the methodology for data collection will be modelled after [10] which has been adopted by others for evaluation of their own algorithms. Due to limited resources, the "syncope" and "rise from bed" tests had to be eliminated from this round of testing, and to keep the balance between positive and negative outcome tests equal the forward fall with rotation tests were removed. Those two in particular were chosen due to their similarity with other tests such as ending lying flat meaning the loss of their data would not be as significant.

Table 1 shows the complete list of tests which were performed along with their fall classification. Each test was repeated 5 times, with the NAO at a distance of 2.5 meter. The result is a data set of 80 actions, half with a positive and half with a negative classified fall outcome.

TABLE 1

Category	Name	Outcome
Backward fall (both legs straight or with knee flexion)	Ending sitting	Positive
	Ending lying	Positive
	Ending in lateral position	Positive
	With recovery	Negative
Forward fall	On the knees	Positive
	With forward arm protection	Positive
	Ending lying flat	Positive
	With rotation, ending in the lateral right position	Positive

	With rotation, ending in the lateral left position	Positive
	With recovery	Negative
Lateral fall to the right	Ending lying flat	Positive
	With recovery	Negative
Lateral fall to the left	Ending lying flat	Positive
	With recovery	Negative
Neutral	To sit down on a chair then to stand up (consider the height of the chair)	Negative
	Walk a few meters	Negative
	To bend down, catch something on the floor, then to rise up	Negative
	To cough or sneeze	Negative

Our results are outlined in the confusion matrix below indicating the number of classified events against the ground truth. Using our simplistic method we were able to attain a system sensitivity of 0.85, a specificity of 0.675 and a precision of 0.72. Due to the simple nature of the algorithm it was capable of running at ~10 frames per second with no GPU acceleration.

TABLE 2

		Results	
		Positive	Negative
Ground Truth	Positive	34	6
	Negative	13	27

V. DISCUSSION AND CONCLUSION

Comparing our results to those published in the literature, we found that this comparatively simplistic methodology has promise, yielding similar system sensitivity of 0.85, managing to identify 34 of the 40 simulated falls in our dataset. Our systems drawback is currently a very high false positive rate leading to a specificity of 0.675. While this initially looks discouraging it is in fact likely not. This is due to the two stage nature of our proposed system that capitalises on the mobile nature of the humanoid robot. This allows repositioning to confirm potential falls by gaining image data that would be inaccessible to a ceiling or corner

mounted camera. While more false positives may be flagged up than in the literature in this paper, many of these will be corrected during the confirmation stage which will result in more robust system overall.

We have demonstrated the preliminary research into the use of a two part identification and confirmation stage system to identify falls in the elderly in near real time on a humanoid robot. Our future work will focus extensively on the second half of this system, the confirmation of a fall, which will greatly increase the overall precision of the system. This paper introduced the concept of a two stage fall detection system using a humanoid robot but thus far has only taken advantage of its two vertically mounted cameras, not its ability to reposition. As shown in figure 1 the fall confirmation stage will focus on attempting to communicate with verbally and detecting the body orientation of the user visually.

VI. ACKNOWLEDGEMENTS

VII. REFERENCES

[1] WHO, "WHo Global report on falls Prevention in older Age," WHO, 2007.

[2] Senior Watch, "A market study about the specific IST needs of older and disabled people," 2001.

[3] C. Zhang, . Y. Tian and E. Capezuti, "Privacy Preserving Automatic Fall Detection for Elderly," in *ICCHP'12 Proceedings of the 13th international conference on Computers*, 2012.

[4] A. Dubois, "Automatic Fall Detection System with a RGB-D," in *ICOST*, 2013.

[5] F. Nater, H. Grabner and L. Van Gool, "Visual abnormal event detection for prolonged independent living," in *IEEE Healthcom Workshop on mHealth*, 2010.

[6] S. McKenna and H. Nait-Charif, "Activity Summarisation and Fall Detection in a Supportive Home Environment," in *Proceedings of the 17th International Conference on Pattern Recognition*, 2004.

[7] D. Glen , P. Karsmakers, M. Deschodt, E. Vlaeyen, E. Dejaeger, K. Milisen, T. Goedemé, B. Vanrumste and T. Tuytelaars, "Camera-based fall detection on real world data," in *15th International Workshop on Theoretical Foundations of Computer Vision*, Dagstuhl Castle, Germany, June 26 - July 1, 2011.

[8] M. Gövercin, J. Spehr, E. Steinhagen-Thiessen and F. M. Wahl, "Visual Fall Detection in Home Environments," in *International Conference for Gerotechnology*, 2008.

[9] J. Collomosse and E. Syngelakis, "A bag of features approach to ambient fall detection," in *Proc. Intl. Symp. on Ambient Technologies*, 2011.

[10] N. Noury, A. Fleury, P. Rumeau, A. K. Bourke, G. O. Laighin, V. Rialle and J. E. Lundy, "Fall detection – Principles and Methods," in *Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale*, Lyon, France, 2007.