Chang, Mei-Chih, Buš, Peter and Schmitt, Gerhard (2018) *Feature Extraction and K-means Clustering Approach to Explore Important Features of Urban Identity.* In: Proceedings of the 16th IEEE International Conference on Machine Learning and Applications (ICMLA 2017). . pp. 1139-1144. IEEE ISBN 978-1-5386-1419-8. E-ISBN 978-1-5386-1418-1.

2017 16th IEEE International Conference on Machine Learning and Applications

Feature Extraction and K-means Clustering Approach to Explore Important Features of Urban Identity

Mei-Chih Chang, Peter Buš, Gerhard Schmitt
Chair of Information Architecture
Swiss Federal Institute of Technology in Zurich
Zürich, Switzerland
{chang, bus, schmitt}@arch.ethz.ch

Abstract—**Public spaces play an important role in the processes of formation, generation and change of urban identity. Under present day conditions, the identities of cities are rapidly deteriorating and vanishing. Therefore, the importance of urban design, which is a means of designing urban spaces and their physical and social aspects, is ever growing. This paper proposes a novel methodology by using Principle Component Analysis (PCA) and K-means clustering approach to find important features of the urban identity from public space. K. Lynch's work and Space Syntax theory are reconstructed and integrated with POI (Points of Interest) to quantify the quality of the public space. A case study of Zürich city is used to test of these redefinitions and features of urban identity. The results show that PCA and K-means clustering approach can identify the urban identity and explore important features. This strategy could help to improve planning and design processes and generation of new urban patterns with more appropriate features and qualities.**

Keywords—Urban identity; Space syntax; K-means clustering; Feature extraction

I. INTRODUCTION

There is growing evidence that urban generation within traditional settings has diminished the place meanings and attachment embedded in the existing social and cultural setting, which happened in transformed and or newly constructed places [1-3]. Identity of the urban elements is important because they contribute to self-identity, sense of community and sense of place [4]. The elements and definition of urban identity are widely discussed as K. Lynch (1960), E. Relph (1976), C. Norberg-Schultz (1979), J. Dixon and K. Durrheim (2000), H.M. Proshansky (2007). Nevertheless, the principle was advocated significantly by K. Lynch.

Urbanists have long held the view that the physical and social dynamics of public space play a central role in the formation of publics and public culture [5]. A city's streets, parks, squares, and other shared spaces have been seen as symbols of collective well-being and possibility. Therefore, public space is used as a means of creating urban identity. The physical or spatial qualities of a space play a large part in creating a space's identity. People recognize and use spaces based on these qualities, and these qualities also help to form and reinforce emotional connections as the cultural root. K. Lynch, in The Image of the City [6] argued for legibility being a significant quality of the city. According to him, the legibility of the city, or "the ease with which a city's parts can be recognized and can be organized into a coherent pattern", is significant not only for aiding practical tasks such as way-finding, but also that it is central to the emotional and physical well-being of the inhabitant population, personally as well as socially. He continues by equating the legible environment with an "imageable" one. Imageability, according to him, is "that qualities a physical object which gives it a high probability of evoking a strong image in any given observer." Lynch identifies five elements that bring out this quality in cities: paths, edges, districts, nodes, and landmarks. When these elements are strong, the image and memorability of the city is strong, but when these elements are not detectable, the city becomes dull and forgettable.

Recently, two researches (R. C. Dalton et al. 2003, S. Bafna et al. 2012) re-interpret Lynch's five environmental features in terms of the basic spatial descriptors commonly used in space syntax research. Space syntax is a theory, method, tool for analysis, and interpretation, which enables to analyze the configuration of space and social variable [7, 8]. Space syntax has been regarded as a new computational language for the study of the urban structure in investigating relationships between spatial layout and a range of social, economic and environmental phenomena. Most applications of space syntax are related to way finding, crime and traffic flow. This paper proposes a novel methodology by using K-means clustering to find important features of the urban identity from public spaces with Kevin Lynch's work and space syntax theory. This could help urban planners to design the city with more appropriate characteristics of an urban identity. The quantified data from space syntax will be the feature data source for K-means clustering. Compared with recent urban identity researches, they focused on using computer vision technologies with machine learning methodologies to identify the perceptual and cultural aspects from urban images (C. Doersch et al. 2012, L. Lu et al. 2016). However, these methods needed a quite amount of image data and were limited in the domain of visual geographies of cities. Our methodology focuses on the spatial layout and quality of place, which could contribute to the generation of building and street patterns considering the features of urban identity.

For the time being, space syntax only deals with a 'space' not a 'place' [16]. There is a limitation of space syntax, which capabilities of analysis cover the socio-spatial information of the settlement patterns and on the ground level only. It cannot infer at all about the society. Place syntax proposed to use places, not spaces, to evaluate the accessibility with giving a function of a space as a place [16]. This paper will reconstruct the equations of the space syntax to integrate it with a place from POI (Points of Interest).

Section II discusses the data source for machine learning, including 2D geometric data of streets, buildings and POIs. The integration of K. Lynch's work and space syntax with POI is discussed in Section III. K-means clustering with PCA based feature extraction is discussed in Section IV. Results are shown in Section V. Discussion and future works are discussed in Section VI.

II. Data Source

A. OpenStreet Map & Geometry Information

Lately the road network and buildings provided by OpenStreetMap (OSM) is often chosen to form the backbone of urban layouts because of its universal coverage and standard defined for all modes of transport and city data. Due to its open access nature and volunteered contributions, OSM can have a very good level of complement. It includes buildings, roads, transportation stations, parks, fields, playground and POI. Different types of POI are available including restaurant, shops, banks, cinemas and so on. After data collected from OpenStreetMap, the data are imported to 3D geometry tools (Rhinoceres 5, Grassshopper and Elk) and transferred to 2D Geometry data.

B. Space Syntax Analysis Data – CityGraph

The DeCodingSpaces [9] is a plugin software tool for Grasshopper graphical algorithm editor in Rhinoceres 5 environment. It implements accessibility function in space syntax including Betweenness Centrality, Closeness Centrality, Gravity, Weighted Betweenness and Degree Centrality. In this paper, we use DecodingSpace to analysis the different features of the public space in Zurich city as Fig. 1.

Fig. 1. Bewteenness Centrality analysis of public spaces in Zürich city by CityGraph

III. Methodology

A. Spatial Qualities in the Identity of Public Space

K. Lynch considers the urban identity as "imageability," which he defines as "that quality in a physical object which gives it a high probability of evoking a strong image in a given observer."

Lynch identifies five elements that bring out this quality in cities: paths, edges, districts, nodes, and landmarks. When these elements are strong, the image and memorability of the city is strong, but when these elements are not detectable, the city becomes dull and forgettable. Imageability is a quality that can also be found in a public space and can make that space an important part of the city, one that reflects it.

1) Path: Paths create movements and flows in the city. Lynch found in his study that paths, namely streets, were the predominant features in people's image of the city. Paths are important because they lead users to a space, and an identifiable path lends an expectation to the approach to a space. On a site scale, a space that is located on a prominent path will not only attract more users, but is likely to figure prominently in the overall image of the city due to its relation to the primary path.

Therefore, R. C. Dalton and S. Bafna proposed that the lines of high integration coincide with the significant paths in Lynch's maps. However, integration (closeness) and choice (betweenness) are often used together in space syntax to measure how nodes are close to each segment. Futhermore, how much movement is likely needed to pass through each segment [10]. These two equations are listed as follows:

Closeness Centrality [11]:

$$CC(s) = \frac{\phi_{st}}{\sum_{t \neq s \in V} d_{st}} \tag{1}$$

where d_{st} is a distance of shortest path from node s to node t and ϕ_{st} is total number of shortest paths which is reachable from s.

Betweenness Centrality [12]:

$$BC(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

where σ_{st} is total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

At the same time, Kevin mentioned that Paths within a space are also important to that space's identity. Paths determine where people go and often provide a sequential pattern of movement through different areas of the space, often increasing the memorability of a particular place. This means that, when people move along paths including starting and end places, the scenes or images along paths are important. However, equation (1) and (2) doesn't mention this. In this paper, we propose the function of place should be calculated, too. Therefore, POI (point of interest) is proposed to be included in these two equations as equation (3) and (4).

Attributed Weight Closeness Centrality:

$$AWCC(s) = \frac{\phi_{st}}{\sum_{t \neq s \in V, s \in \lambda \& t \in \theta} d_{st} \times W_{\lambda\theta}} \tag{3}$$

where d_{st} is a distance of the shortest path from node s to node t, ϕ_{st} is the total number of shortest paths which is reachable from node s and w_{θ} is the weight value where t belong to place θ.

Attributed Weight Betweenness Centrality:

$$AWBC(v) = \sum_{\substack{s \neq v \neq t \in V \\ s \in \lambda, t \in \theta}} \frac{\rho_{st\lambda\theta}(v)}{\rho_{st\lambda\theta}}, \quad \rho_{st\lambda\theta} = \rho(\sigma_{st}, \lambda, \theta) \text{ and } s \in \lambda, t \in \theta \tag{4}$$

where $\rho_{st\lambda\theta}$ is the total number of shortest paths with attribute λ, θ from node s to node t and $\rho_{st\lambda\theta}(v)$ is the number of those paths with attribute λ, θ that pass through v.

In this paper, to provide the space with meaning, the meaning of space is given by the place, which is the function of the space. Therefore, gardens, squares, fields, playgrounds, transportations, shops are assigned with attributed weight. The same function of space belongs to the same attributed weight. There are total four types of the place – leisure places (gardens, playgrounds, squares), fields, transportations and shops.

2) Node

Nodes are the strategic foci into which the observer can enter, typically either junctions of paths or concentrations of some characteristic." (Lynch 1960, 72) When applied to the scale of the public space, nodes are the places within the space where activities are positioned. These could be the entrances to a site or specific areas within it where people go for a particular activity: a court to play basketball, a bench in the sun. Nodes bring people to the site for the activities that can be done there. Nodes also bring a sense of identity when a space becomes known for those activities.

R. C. Dalton and S. Bafna proposed that isovist mapping of the nodes could help characterize further the significant nodes. In this paper, the properties of isovist include isovist area, isovist perimeter, openness and compactness as Fig. 6 These properties are considered in both researches - R. C. Dalton and M. Knöll (2015).

Fig. 2. Isovist area with radius 250 m

3) District

Districts are recognizable sections of a city that often have a distinct character. At the site scale, districts are those areas within a public space that have a particular consistent character. An entire space could have a specific character, often defined by the character of the surrounding neighborhood, or subspaces within a site can have specific characters or qualities, that often also have a clear coordinating spatial structure.

R.C. Dalton and S. Bafna proposed that the range of integration value could find the local structural features of district. In this paper, our proposed the attributed weight closeness centrality as equation (3) is adopted to evaluate the feature of districts.

4) Edge and Landmark

Edges mark boundaries. They separate one section of a city from another. R. C. Dalton and S. Bafna thought that it illustrates isovist with visually prominent edges constituting a major proportion of the visual boundary. Therefore, in this paper, the four properties of isovist are used for the features of edge.

Landmarks can help a space reflected the city in two ways: they can be physically within the public space, or they can be viewpoints from a space. The landmark that is physically on the site can become a connection from the space to the city—a point that may eventually become a viewpoint of the city. However, it is hard to define landmark by space syntax. The only feature of the landmark is having a distinctive isovist shape as isovist compactness. Therefore, this paper uses the four properties of isovist to relate the features of landmark. These four properties include isovist area, isovist perimeter, openness and compactness.

B. K-means clustering and Principle Component Analysis

Clustering is an un-supervised machine learning method grouping similar objects together based on their natural similar characteristics. Therefore, objects belonging to a group are more similar as compared to the objects belonging to other groups. There is a list of different clustering algorithms can be performed with Scikit-learn software library. K-means algorithm is the most famous algorithm to cluster data by trying to separate samples in n groups of equal variance, minimizing a criterion known as the inertia or within-cluster sum-of-squares. It aims to choose centroids that minimize the inertia, or within-cluster sum of squared criterion. It has been successfully used for image segmentation and multi-class classification [13].

In this paper, we propose that clustering algorithm can help the designer to identify the features of urban identity from the public space. Owing to the characteristics of clustering, it can help the designer to identify the similar features of the same cluster of the public space. These features could be the characteristics of urban identity. However, the dimension of the feature space is quite high. Important features can help in creating clusters while unimportant maybe not help in creating clusters and, in contrary, it may affect the clustering algorithms adversely by blurring the clusters. Dimensionality reduction is one popular technique to remove noisy and redundant attributes [14]. In this paper, Principle Component Analysis (PCA) is adopted to do feature extraction because it can help the urban designer to identify which features are more important than others. Fig. 3 shows the overall system architecture of this methodology.

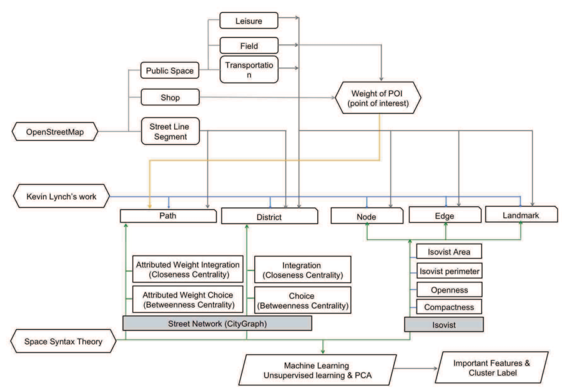


Fig. 3. System architecture

IV. RESULTS

A. Overall Result of K-means clustering and Principle Component Analysis

This research proposes total 22 features for K-means clustering. Eight statistics features (mean, variance, standard deviation, median, skewness, kurtosis, 3rd-moment, 4th -moment) from attributed weight betweenness centrality (AWBC) and attributed weight closeness centrality (AWCC). Two distance measurements from the public space to the segment line (road) with highest value of AWBC or AWCC. Four isovist features - isovist area, isovist perimeter, openness and compactness are included in these 22 features.

There are total 597 public spaces (parks, gardens, playgrounds and squares), 21 train stations, 84 tram stops, 309 fields and 8720 POI locations for the Zürich city from OpenStreetMap. The implementation of our overall system architecture is done by Rhino 5, Grasshopper, Grasshopper plugins (CityGraph & Elk) and Scikit-learn software library codes for unsupervised machine learning (as shown in Fig. 4).

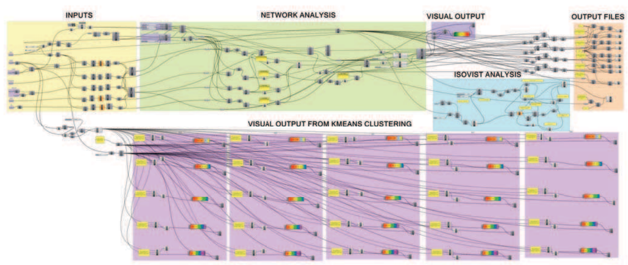


Fig. 4. Implemented system architecture with Grasshopper and plugins

In this research, there are two to six components of PCA and three to seven cluster numbers of K-means to be tested. The test cases with different PCA component number and the cluster number of K-means are selected for the best one with larger of variance of the area size (bigger than 30,000m^2) of public space and smaller of variance of clustering label. In Zürich city, there are many parks and gardens in urban city and they mix with building, this is one of urban identity. However, the bigger parks are more isolated from building. Therefore, we specify the clustering result for bigger park (area size bigger than 30,000m^2) should have the same cluster label, that is, the variance of clustering label in these public spaces should have bigger variance value. And the result of clustering label should be more equally distributed because each clustering label should keep some common features for most of pubic spaces. Table I shows these two values for different PCA component number and K-means cluster number.

In table I, we find PCA component value with 4 and K-means cluster number with 5 is the best. We check this case for eleven popular public spaces in Zürich city, whose public spaces are selected from a Copenhagen report (C. Bräm et al. 2014). However, the result is not so good beause the clustering labels are almost the same in these eleven public spaces then it is hard to identify the difference between them. The second best of case is PCA component value with 6 and K-means cluster number with 5, which can better identify the differences between these eleven public spaces. Fig. 5 shows the clustering label for the whole Zürich city. Colored circles show the public space with clustering label and the same clustering label has the same color.

TABLE I. LIST OF ZURICH CLUSTERING LABEL

PCA num.	Cluster num.	List of Zurich Clustering Label	
		Clustering label %	***Variance of area > 30,000 m^2 / Variance of Label***
2	3	C0:50.42% C1:46.06% C2: 3.52%	-0.166667 -0.177457
	4	C0:44.89% C1:3.51% C2: 48.41% C3: 3.18%	-0.135842 -0.140472
	5	C0:43.38%C1:47.4%C2:1.51% C3: 4.19%C4: 3.52%	-0.108673 -0.116771
	6	C0:32.8%C1:21.6%C2: 1.5%C3:4.2%C4:37.0%C5: 1.5%	-0.106718 -0.117616
	7	C0:17.76% C1:30.65% C2:1.51% C3: 12.73% C4: 4.19% C5:2.35% C6: 30.82%	-0.107143 -0.108683
3	3	C0:71.86% C1:24.62% C2: 3.52%	-0.133503 -0.140586
	4	C0:61.81% C1:3.52% C2: 22.78% C3:11.89%	-0.151051 -0.137672
	5	C0:33.84%C1:39.03%C2:3.52% C3: 17.25%C4:6.37%	-0.129082 -0.139627
	6	C0:27.64%C1:12.4%C2:3.52% C3: 30.82%C4:3.35%C5:22.3%	-0.127551 -0.126878
	7	C0:4.02% C1:33.5% C2:11.56% C3: 19.26% C4:3.35% C5:1.51% C6: 26.8%	-0.094023 -0.10893
4	3	C0:54.94% C1:41.54% C2: 3.52%	-0.184524 -0.17478
	4	C0:16.42% C1:36.86% C2: 43.22% C3:3.52%	-0.160714 -0.12314
	5	C0:6.7%C1:46.73%C2:27.14% C3: 3.52%C4:15.91%	-0.136224 -0.135383
	6	C0:28.64%C1:15.41%C2:14.07% C3: 3.52%C4:3.35%C5:35.0%	-0.121173 -0.124915
	7	C0:19.6% C1:26.47% C2:3.52% C3: 3.52% C4:14.9% C5:19.93% C6: 12.06%	-0.115525 -0.116082
5	3	C0:46.4.94% C1:50.08% C2: 3.52%	-0.16667 -0.177547
	4	C0:37.02% C1:43.55% C2: 3.52% C3:15.91%	-0.148597 -0.161684
	5	C0:51.76%C1:3.52%C2:6.2% C3: 22.78%C4:15.75%	-0.134694 -0.130067
	6	C0:5.36%C1:15.91%C2:14.57% C3: 32.33%C4:3.52%C5:28.31%	-0.113946 -0.127447
	7	C0:28.98% C1:12.47% C2:13.1% C3:23.79% C4:15.24% C5:3.52% C6: 3.02%	-0.108236 -0.114519
6	3	C0:32.29% C1:67.05% C2: 0.66%	-0.142007 -0.148703
	4	C0:13.57% C1:24.29% C2: 58.63% C3:3.52%	-0.139668 -0.144414
	5	C0:12.73%C1:19.77%C2:21.27% C3: 42.71%C4:3.52%	-0.12857 -0.130067
	6	C0:11.73%C1:29.65%C2:19.77% C3: 31.83%C4:3.52%C5:3.52%	-0.121173 -0.12592
	7	C0:26.63% C1:3.52% C2:11.56% C3:24.96% C4:12.4% C5:3.69% C6: 17.25%	-0.110058 -0.115099

B. Urban Idenity Analysis

Lynch's path, node and district elements are applied to analyze the eleven public spaces in Zürich to identify the urban identity by using clustering labels (shown in the left part of Fig. 6), important features from PCA (shown in table Ⅱ). There are five clustering labels with green, red and yellow color circle shown in the left part of Fig. 6 and Fig. 7. In the left part of Fig. 6 and Fig. 7, there are four colored areas: green areas show parks, orange areas show squares, pink areas show train stations, pink dots show tram stops and small white spots show POI (including shops, cinemas, banks, pharmacies, schools and restaurant). The images of the right part of Fig. 6 are from the Google map, which show real conditions of these public spaces.

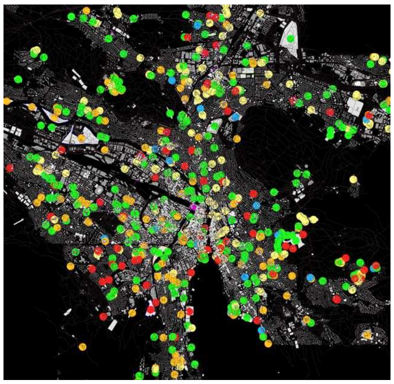


Fig. 5. Clustering label of Zürich city

Lynch's path element can be identified by the feature index 9 and 18, which are the distance to the center of traffic and the distance to the center of POIs. That is, it means how far away the fork flow and how far away the street shops are located.

Lynch's node element can be identified by the feature index 21 and 22, which are the openness and the sense of the orientation of the space. Mostly, a higher openness value means that this public is easy to be seen from other locations, which could be a place as landmark and a stronger sense of the orientation also has the same meaning.

Lynch's district element can be identified by the feature index 1 and 5, which are the mean and skewness of AWBC. The higher value of AWBC shows stronger relation in this district.

The values and levels of all features for five clustering label are listed in table Ⅱ and Ⅲ. According to these two table, the urban identity from the five clustering label are described as follows.

TABLE II. LIST OF IMPORTANT FEATURE INDEX & VALUE

Label color	***Important Feature Index & Value***					
	1	**5**	**21**	**18**	**9**	**22**
Orange	435.05	4.1656	0.0881	230.89	248.49	0.2863
Red	857.05	5.01	0.6424	220.34	242.3	0.1742
Green	568.58	4.2887	0.7813	225.87	246.97	0.3411
Blue	383.02	3.853	0.7067	243.44	249.65	0.1635
Yellow	13382	5.6288	0.5734	217.53	241.25	0.2131

Index 1: mean of AWBC, index 5: skewness of AWBC, index 21: isovist openness, index 18: distance between the public space to the road with highest AWCC value, index 9: distance between the public space to the road with highest AWBC value, index 22: isovist compactnes

TABLE III. LIST OF IMPORTANT FEATURE PROPERTY & LEVEL

Label color	***Important Feature Indexes***					
	Relation (1)	**Relation (2)**	**Openness**	**Distance to center of POI**	**Distance to center of traffic**	**Sense of ort.**
Orange	Med.	Med.	Very Low	Med.	High	Med.
Red	High	High	Med.	Med.	Low	Low
Green	Med.	Med.	High	Med.	Med.	High
Blue	Low	Low	High	High	High	Low
Yellow	Very High	Very High	Med.	Low	Low	Med.

1) Green clustering label:

Oerlikon park, Bucheggplatz, Bürkliplatz and Zürichhorn park belong to the green clustering label. These public spaces are similar to landmarks of the city with a higher value of the feature index 21 and 22. The surrounding environment also has shops to support services. These parks or squares mix with city well.

2) Red clustering label:

Schwamendingerplatz and Rigiplatz belong to the red clustering label. These public spaces are located at a well-connected community with a higher value of the feature index 1 and 5. They provide the local community with a safer space by a medium value of the feature index 21. They are also easy to access by people with a lower value of the feature index 9.

3) Yellow clustering label:

MFO-Park, Lindenplatz, Turbinenplatz, Backeranlage Park and Central belong to the yellow clustering label. They are located at the popular locations of the city with a very high value of the feature index 1 and 5. These public spaces provide the city places to relax. Therefore, they are easy to access with a higher value of the feature index 9. And the surrounding is full of services with a higher value of the feature index 18.

4) Orange and blue clustering label:

These two clustering label are not listed in the popular public space. Most of them are located at the courtyard of a group of buildings (the orange clustering label) or a place far from the city (the blue clustering label). The public spaces of the orange label are easy to access with a higher value of the feature index 9.

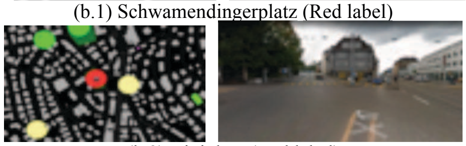
(a.1) Oerlikon park (Green label)

(a.2) Bucheggplatz (Green label)

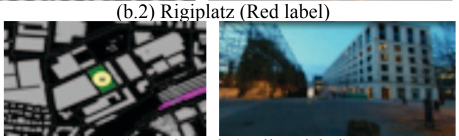
(a.3) Bürkliplatz (Green label)

(a.4) Zürichhorn Park (Green label)

(b.1) Schwamendingerplatz (Red label)



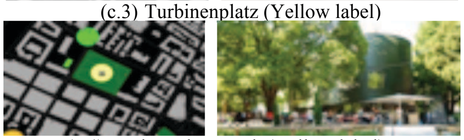
(b.2) Rigiplatz (Red label)



(c.1) MFO-Park (Yellow label)

(c.2) Lindenplatz (Yellow label)

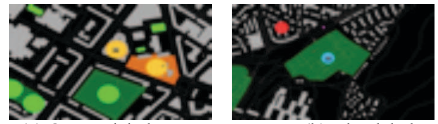
(c.3) Turbinenplatz (Yellow label)



(c.4) Bäckeranlage Park (Yellow label)

(c.5) Central (Yellow label)

Fig. 6. Labeling condition and images for 11 popular public spaces in Zürich



(a) Orange label (b) Blue label

Fig. 7. Labeling condition

V. CONCLUSION

This paper shows that PCA and K-means clustering algorithm could help the urban designer to identify the characteristic of urban identity for the public space. The same clustering label shows the same features of public spaces. These identify the important features for it. These features related to the spatial layout and quality of place will inform urban planners to create more appropriate urban patterns. The future work will focus on different cities in the world, especially comparing eastern and western countries. And the results will be validated with other methodologies.

REFERENCES

[1] M. A. E. Saleh, "The Integration of Tradition and Modernity: A Search for an Urban and Architectural Identity in Arriyadh, The Capital of Saudi Arabia," *Habitat International,* vol. 22, no. 4, pp. 571-589, 1998.

[2] N. Ujang, "Place Attachment and Continuity of Urban Place Identity," *Procedia - Social and Behavioral Sciences,* vol. 49, pp. 156-167, 2012.

[3] A. Abdel-Hadi, "Culture, Quality of life, Globalization and Beyond," *Procedia - Social and Behavioral Sciences,* vol. 50, pp. 11-19, 2012.

[4] R. B. Hull, M. Lam, and G. Vigo, "Place identity: symbols of self in the urban fabric," *Landscape and Urban Planning,* vol. 28, no. 2-3, pp. 109-120.

[5] A. Amin, "Collective culture and urban public space," *City,* vol. 12, no. 1, pp. 5-24, 2008.

[6] K. Lynch, *The image of the city*. USA: MIT Press, 1960.

[7] B. Hillier and J. Hanson, *The Social Logic of Space*. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore: Cambridge University Press, 1984.

[8] B. Hillier, *Space is the machine*. London, UK: Cambridge University Press, 2007.

[9] A. Abdulmawla *et al.*, "DeCodingSpaces," ed: Bauhaus-Universität Weimar, 2016.

[10] P. Nourian, F. v. d. Hoeven, S. Rezvani, and S. Sariylidiz, "Easiest paths for walking and cycling: Combining syntactic and geographic analyses in studying walking and cycling mobility," in *SSS10 Proceedings of the 10th International Space Syntax Symposium*, London, UK, 2015, pp. 78.1-78.10.

[11] Wikipedia. (2017). *Closeness Centrality*. Available: https://en.wikipedia.org/wiki/Closeness_centrality.

[12] Wikipedia. (2017). *Betweenness centrality*. Available: https://en.wikipedia.org/wiki/Centrality - Betweenness_centrality

[13] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys,* vol. 31, no. 3, pp. 264-323, 1999.

[14] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," in *Data Classification: Algorithms and Applications*, C. C. Aggarwal, Ed. New York, USA: Chapman and Hall/CRC, 2014, pp. 37-64.