Big-Data Informed Citizen Participatory Urban Identity Design

Mei-Chih Chang[1]*, Peter Buš*[2]*, Ayça Tartar*[3]*, Artem Chirkin*[4]*,*
Gerhard Schmitt[5]
[1,2,3,4,5]*Chair of Information Architecture ETH Zürich*
[1,2,3,4,5]*{chang|bus|tartar|chirkin|schmitt}@arch.ethz.ch*

The identity of an urban environment is important because it contributes to self-identity, a sense of community, and a sense of place. However, under present-day conditions, the identities of expanding cities are rapidly deteriorating and vanishing, especially in the case of Asian cities. Therefore, cities need to build their urban identity, which includes the past and points to the future. At the same time, cities need to add new features to improve their livability, sustainability, and resilience. In this paper, using data mining technologies for various types of geo-referenced big data and combine them with the space syntax analysis for observing and learning about the socioeconomic behavior and the quality of space. The observed and learned features are identified as the urban identity. The numeric features obtained from data mining are transformed into catalogued levels for designers to understand, which will allow them to propose proper designs that will complement or improve the local traditional features. A workshop in Taiwan, which focuses on a traditional area, demonstrates the result of the proposed methodology and how to transform a traditional area into a livable area. At the same time, we introduce a website platform, Quick Urban Analysis Kit (qua-kit), as a tool for citizens to participate in designs. After the workshop, citizens can view, comment, and vote on different design proposals to provide city authorities and stakeholders with their ideas in a more convenient and responsive way. Therefore, the citizens may deliver their opinions, knowledge, and suggestions for improvements to the investigated neighborhood from their own design perspective.

Keywords: *Urban identity, unsupervised machine learning, Principal Component Analysis (PCA), citizen participated design, space syntax*

INTRODUCTION

East-Southeast Asia is currently one of the fastest-urbanizing regions in the world, with countries such as China climbing from 20% to 50% urbanization in just a few decades. The identity of an urban environment is important because it contributes to self-identity, a sense of community, and a sense of place. However, under present-day conditions, the identities of expanding cities are rapidly deteriorating and vanishing, especially in the case of Asian cities. The elements and definition of urban identity have been widely discussed by K. Lynch (1960), E. Relph (1976), C. Norberg-Schultz (1979), J. Dixon and K. Durrheim (2000), and H. M. Proshansky (2007). Nevertheless, the principle was advocated significantly by K. Lynch.

One recent study (Chang et al. 2017) uses data mining technologies, including both unsupervised machine learning and feature selection methodologies, to explore the important features of urban identities from the quality of open public space. The quantification of this quality is based on K. Lynch's theory (1960) and POI combined space syntax (Hillier and Hanson, 1984). It demonstrates how public space can be used to identify an urban identity by quantization. However, this study does not demonstrate how designers can use this information to create urban designs according to these features. In fact, these features could contribute to exploring the traditional features for historical areas, which could help designers more clearly identify an urban identity.

Nowadays, cities are reconsidering their respective city models, trying to increase the strategic value of their territory to transform into more attractive and cohesive cities that offer greater quality of life and better citizen coexistence. In this way, a change in their traditional identity is taking place. In other words, cities need to build an urban identity that extends to the past. At the same time, cities need to add new features to improve urban life. Therefore, while data mining methodologies can help the designer consider the urban identity, the city also needs new technology to enable its citizens to lead modern lives . In this paper, a novel framework integrated with data mining technologies and the citizen participatory tool is proposed for building creative urban identities for contemporary and future life.

The data mining sources and the proposed methodology will be discussed in section - Data Source And Methodology. Urban identity designs in the Taiwanese workshop are discussed in section - Demonstration and Result. Discussion and future works are discussed in the final section.

DATA SOURCE AND METHODOLOGY

OpenStreetMap & POI Information

OpenStreetMap (OSM) is often chosen to form the backbone of urban layouts because of its universal coverage. It includes buildings, roads, transportation stations, parks, fields, playgrounds, and POI. Different types of POI are available, including restaurants, shops, banks, and cinemas. After being collected from OSM, the data are imported to 3D geometry tools (Rhinoceros 5, Grasshopper, and Elk) and converted into 2D geometry data. The POI data obtained from OSM including parks, shops, and train stations.

POI Integrated Space Syntax Analysis Data

DeCodingSpaces (Abdulmawla , et al., 2018) is a plugin software tool for the Grasshopper graphical algorithm editor in the Rhinoceros 5 environment. It implements accessibility functions in space syntax, including betweenness centrality, closeness centrality, gravity, weighted betweenness, and degree centrality. For this paper, DecodingSpaces was used to analyze the different features of the public space in Taipei City. These features, integrated with attributed weight (Chang, Buš, & Schmitt, 2017), could correspond with what K. Lynch identified as the five elements that bring out the quality in cities that can be considered as urban identity. Attributed weight closeness centrality (AWCC) and attributed weight betweenness centrality (AWBC) were used to analyze K. Lynch's five elements of the quality of the city.

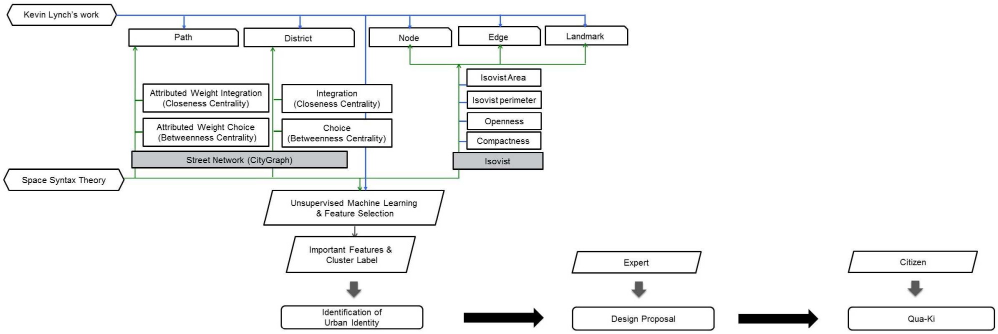


Figure 1
The system architecture of overall framework

Principal Component Analysis and K-means clustering

Clustering is an unsupervised machine learning method that groups similar objects together based on their natural similar characteristics. The K-means algorithm is the most famous algorithm for clustering data by trying to separate samples into n groups of equal variances, minimizing a criterion known as inertia, or within-cluster sum-of-squares. For this paper, the clustering algorithm was used to help the designer to identify the features of urban identity from public space. These features could be the characteristics of urban identity. However, the dimension of the featured space is quite high. Therefore, dimensionality reduction is used to remove noisy and redundant attributes (Dash and Liu, 1997). Principal component analysis (PCA) is a popular algorithm for reducing dimensionality and was used in this paper to conduct feature extraction for helping the urban designer to identify which features are important. After these two algorithms were applied, the resulting numeric output data were catalogued to simplify the understanding of the features for designers.

Qua-kit and Overall Framework

Qua-kit is a web platform for viewing and manipulating simple urban geometry (A. & Chair of Information Architecture ETH Zürich, 2018). While users are logged into this platform, they can work on a single design problem, share ideas and design proposals, and view and discuss the works of others. Therefore, after proposed designs are uploaded by experts, involved citizens can add comments, view and modify designs, and vote for their favorite designs. The overall framework of our proposed system (diagrammed in Figure 1) is that the observed and learned features will be identified as the urban identity through data mining technologies, which would help designers understand the contemporary citizen lifestyle. These expert-based design proposals are then shared via a web platform to involve citizens in the citizen participatory design, which can make the city adapt, learn, and intelligently respond to citizens' lives.

Figure 2
The Nan Ji Chang Urban Redevelopment Area

DEMONSTRATION AND RESULT

SITE

The Nan Ji Chang Urban Redevelopment Area is located in Taipei City, Taiwan, and encompasses a traditional Taiwanese street night market as illustrated in Figure 2. At this site, there are many challenges; including improving quality of life, preserving the traditional Taiwanese night market, expanding public transportation, and adding open public spaces and public utilities. The Taipei City Government initiated

Label color	Important Feature Index & Value					
	5	11	12	19	20	16
Red	4.145369	92.23764	0.706854	0.062287	182.8178	10393.69
Orange	6.425341	51.29905	0.43621	0.768914	80.82609	10953.99
Yellow	13.7944	165.3083	0.344122	0.114332	343.5471	4781.872
Green	3.533283	6902.45	-0.64644	0.063714	562.3333	1730.776
Blue	3.847212	150.7482	0.872468	0.120596	408.5116	3709.837
Baby Pink	8.294647	110.0754	1.116783	0.510371	103.5455	408104.4
Pink	6.594564	1939.793	-0.74627	0.149124	562.0965	5741.549
Purple	7.912001	0.811928	12.93768	0.133651	378.2162	21207.54

Table 1
List of important feature index and value

an urban redevelopment plan for this area and invited local citizens to participate in the plan through a workshop. This workshop adopted the methodology of this paper to explore the urban identity for this area and help urban planners create proper designs. After the workshop, these design proposals could be viewed and referenced as part of the citizen participatory design.

Result of K-means Clustering and Principle Component Analysis

This research proposes a total of 18 features for K-means clustering. These 18 features include six statistics features (mean, variance, standard deviation, median, skewness, kurtosis) from AWBC and AWCC, as well as four isovist features (isovist area, isovist perimeter, openness, and compactness). The building density and the size of the open public spaces are entirely included in these 18 features. According to OSM, there are 1,049 open public spaces (parks, gardens, shopping streets), 562 train stops, and 18,099 POI locations in Taipei City. The implementation of our overall system architecture was done with Rhinoceros 5, Grasshopper, Grasshopper plugins (DeCodingSpaces & Elk), and Scikit-learn software library codes for unsupervised machine learning.

There were four to nine components of PCA and four to nine cluster numbers of K-means to be tested. The test cases with different PCA component numbers and the cluster number of K-means were selected for the best one , with the bigger parks having the same cluster. A PCA component value of 8 and K-means cluster number of 6 were the best, as displayed in Figure 3. Figure 3 illustrates the clustering label for the main part of Taipei City. Colored circles indicate open public spaces with the clustering label, and all open public spaces with the same clustering label have been assigned the same color.

Urban Identity Analysis

Lynch's path, node, and district elements were applied to analyze the eight types of public spaces in Taipei City to identify the urban identity by using clustering labels and important features from PCA as indicated in Table 1.

Index 5 relates to the kurtosis value of AWBC. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution-that is, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers (Freund & Peries, 2006). The kurtosis for a standard normal distribution is 3; therefore, when the value of kurtosis is higher than 3, it means the data are heavy-tailed and have similar data values. AWBC describes the frequencies of nodes in the shortest paths between two indirectly connected nodes (Hai & Junsheng, 2010). The higher kurtosis value of AWBC indicates this path like a bridge . Therefore, when the kurtosis value of AWBC is high, it means most of the AWBC values are similar-that is, this area has very equal accessibility. Conversely, when the kurtosis value of AWBC is low, it means most of the AWBC values are different-that

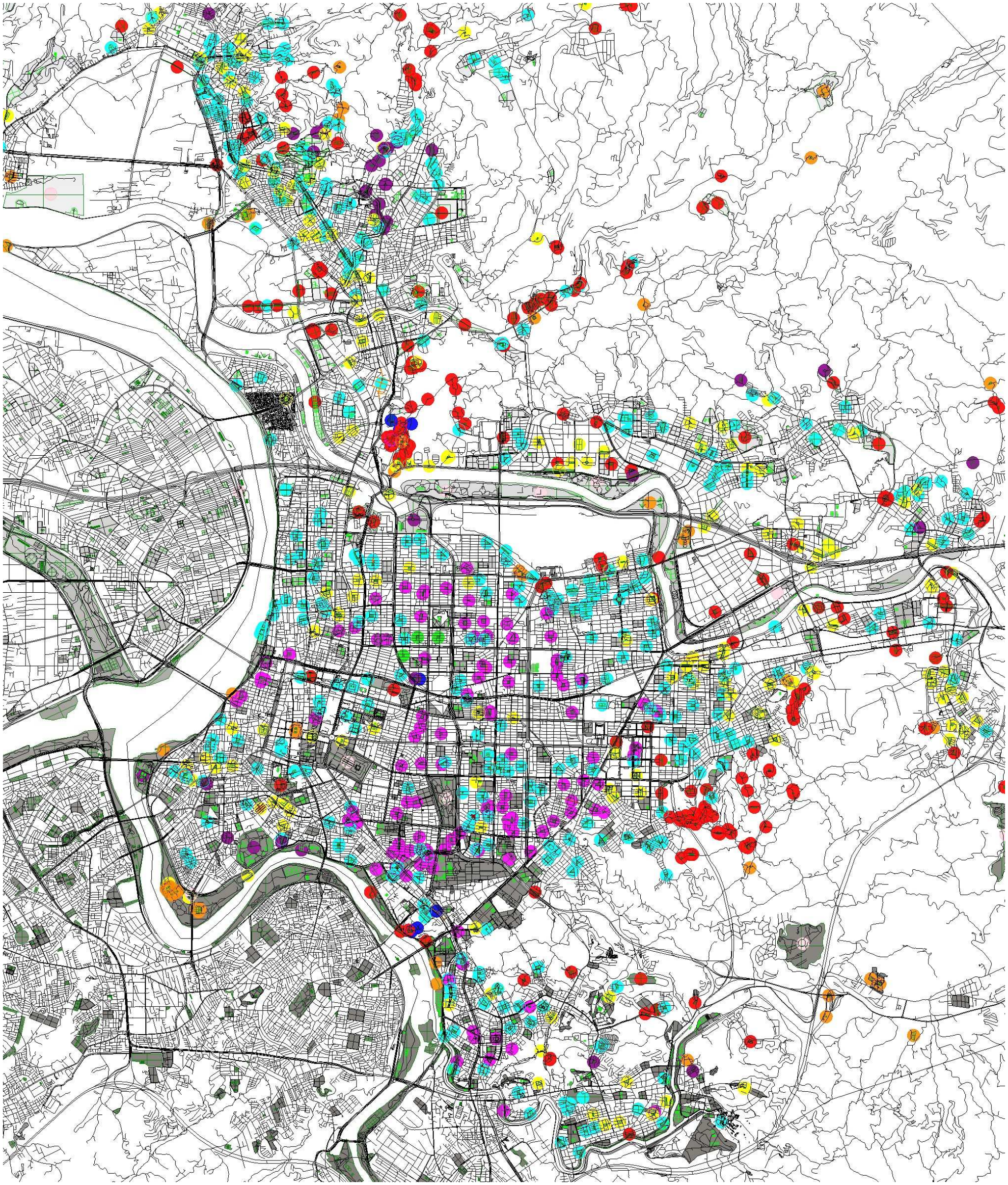


Figure 3
Clustering label of
Taipei city

is, this area has some specific roads with higher traffic volume like a bridge.

Indexes 11 and 12 relate to the median value and the skewness value of AWCC. The median is the value separating the higher half of a data sample. Skewness is a measure of symmetry, or more precisely, lack of symmetry (Freund & Peries, 2006). Skewness for a normal distribution is 0, and any symmetric data should have a skewness near 0. Negative values for skewness indicate data that are skewed left, and positive values for skewness indicate data that are skewed right. By skewed left, we mean that the left tail is long relative to the right tail. Similarly, skewed right means that the right tail is long relative to the left tail. Therefore, there are four major conditions for skewness. The first condition is when the median value is low and the skewness value is positive: most of the data samples are quite small. The second condition is when the median value is high and the skewness value is negative: most of the data samples are quite large. The third condition is when the median value is high and the skewness value is positive: most of the data samples are large. The fourth condition is when the median value is low and the skewness value is negative: most of the data samples are small. However, there are other conditions in between these four conditions, which will be discussed on a case-by-case basis.

AWCC describes the efficiency of information propagation from one node to the others (Hai and Junsheng 2010). Therefore, a higher median value of AWCC indicates that the area is well connected. Conversely, a lower median value of AWCC indicates that the area is not well connected. Therefore, the first condition of the median value and the skewness value of AWCC indicate that the area is not well connected. The remaining conditions may be deduced by analogy.

For index 19, compactness measures the ratio of average to farthest (or maximum) distance from each vantage point (Batty 2001). This measure thus varies from 0 to 1, with low values for stringy, noncompact linear shapes to high values for compact circular shapes. For index 20, building density measures the total number of buildings in a circle with a radius of 500 meters . For index 16, area measures the area size.

To help the designer understand the meaning of these numeric data, the data were sorted into different levels: very low (L *), low (L), median (M), high (H), and very high (H *), as depicted in Table 3. Every label is summarized in feature descriptions by referencing Table 2 and sorted by locations in Table 3. From Table 3, it can be seen that the site of the workshop belongs to the yellow label, which means this area is not located at the center of the city, but it is easy to access with high pedestrian flow. From the third column of Table 3, it can also be seen that most of the open space belongs to the blue label, which means that most of the open space in Taipei City is located outside the center of the city and has a median size. Big parks or open spaces constitute only 13% of the total open space, and most of them are located around the river as illustrated in Figure 4.

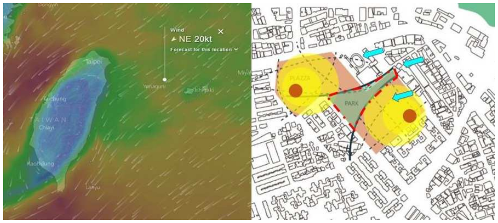


Figure 4
Concept design proposal from stream of food design proposal

Urban Identity Design and Analysis

Urban Identity Design. There were eight design proposals for this site, each reflecting a different approach for preserving or rethinking the local identity and livability for the traditional night market in this Nan Ji Chang Urban Redevelopment Area. During this Taiwan workshop, the parametric design was used for finding different approaches with space syntax analysis, the generation of plots and buildings by DeCodingSpaces (Abdulmawla , et al., 2018), and the features of the livability of the city. Four of these design proposals were discussed in detail as follows.

Table 2
List of feature description, location and ratio

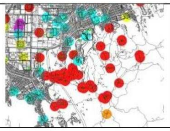
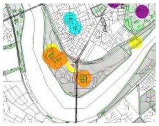
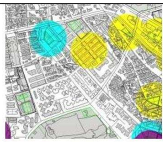
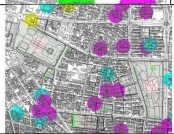
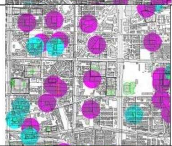
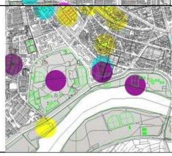
Label color	*Feature Description, Location & Ratio*		
	Feature Description	**Location**	**Ratio**
Red	Not well connected, fewer buildings and big size of the open space		22%
Orange	Not well connected, fewer buildings and big size of the open space with the shape similar to square		9%
Yellow	Median connected and most of areas behave like the bridge with high pedestrian flow. (Our site belongs to this label)		16%
Green	Very well connected, high building density and small area size of open space		1%
Blue	Median connected median building density and median size of the open space.		37%
Baby Pink	Not well connected, fewer buildings and huge size of the open space		1%
Pink	Well connected, high building density and median area size of open space		11%
Purple	Not well connected, median building density and big size of the open space		3.5%

Table 3
List of important feature property and level

Label color	*Important Feature Property & Level*					
	AWBC Kurtosis	**AWCC Median**	**AWCC Skewness**	**Compactness**	**Building Density**	**Area**
Red	M	L	+ H	L	L	H
Orange	H	L	+ M	H	L	H
Yellow	H*	M	+ M	L	M	M
Green	M	H*	- M	L	H	L
Blue	M	M	+ H	L	M	M
Baby Pink	H	L	+ H	M	L	H*
Pink	H	H	- H	L	H	M
Purple	H	L*	+ H*	L	M	H

(1) Stream of Food. The idea of the stream of food design proposal is to keep the concept of the pedestrian flow in this traditional market area. To ensure the coexistence of the residential area with the commercial area, it proposes to make use of the seasonal wind and green spaces to connect and separate them, thereby preserving the quiet and safety of the residential area as illustrated in Figure 4. The analyses of AWBC, AWCC, and isovist are illustrated in Figure 5.

(2) Fusion Double Layer. The idea of the fusion double layer design proposal is to reduce the impact of the commercial area on the residential area by creating a new pedestrian path. Therefore, the traditional market remains on the ground floor, while a new upper level includes a pedestrian path for protecting local residents, as illustrated in Figure 6. The analyses of AWBC, AWCC, and isovist are illustrated in Figure 7.

(3) Daily Axis of Urban Context. The idea of the daily axis of urban context design proposal is to integrate different activities, including culture for school, the commercial area, the traditional night market, and the residential area, by using nodes and the axis to connect them. Therefore, the traditional market remains on the ground floor, while a new upper level includes a pedestrian path for protecting local residents , as illustrated in Figure 8. The analyses of AWBC, AWCC, and isovist are illustrated in Figure 9.

(4) The Green Belt. The idea of the green belt design proposal is to create a new style for the night market with fluid circulation, mixed-use commercial and residential areas, and more green spaces. With fluid circulation, this proposal also creates fluid buildings that result in a new outlook and design for the traditional night market, as illustrated in Figure 10. The analyses of AWBC, AWCC, and isovist are illustrated in Figure 11.

Figure 5
Result of Space Syntax analysis for stream of food design proposal

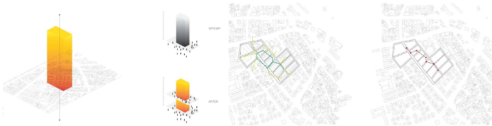


Figure 6
Concept design proposal from fusion double layer design proposal

Figure 7
Result of Space Syntax analysis for fusion double layer design proposal

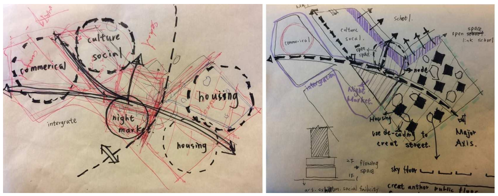


Figure 8
Concept design proposal from daily axis of urban context design proposal

Figure 9
Result of Space Syntax analysis for daily axis of urban context design

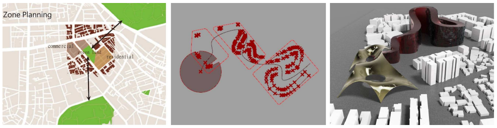


Figure 10
Concept design proposal from the green belt design proposal

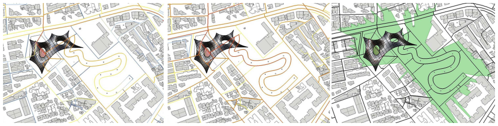


Figure 11
Result of Space Syntax analysis for the green belt design proposal

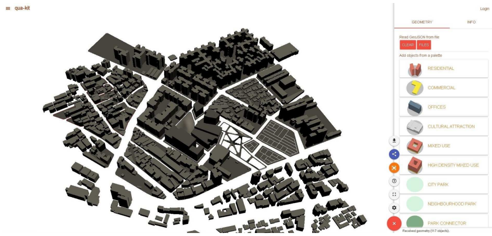


Figure 12
The interface of Qua-kit with the design proposal

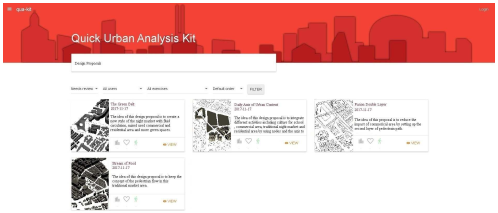


Figure 13
List view of all design proposals by Qua-kit

Analysis of Urban Identity Design. All calculated features of the different design proposals were analyzed by K-means clustering to predict their label and determine the urban identity as indicated in Table 4 and Table 5. From Table 5, it can be seen that the first three design proposals keep the same features of the traditional market. Although these proposals increase the size of the open space, the major features are still preserved. The fourth design proposal reduces too many paths for pedestrians ; therefore, it lacks closeness . However, it creates an interesting path for different views.

Quick Urban Analysis Kit (qua-kit)

After workshop, every design proposals are uploaded to Qua-kit for public viewing, voting and making comments, as illustrated in Figure 9. All design proposals can be viewed together as shown in Figure 13, which provide users to review all possible ideas for the site. At same time, citizens can discuss and compare these design then they also make their proposals from Qua-kit. Therefore, by this open and transparent web-platform, citizens can inform authorities about their ideas, which enables them to participate the urban planning to make the city responsive and livable.

In summary, data-mining technology can help designers identify the urban identity, which can fit the local urban features to help local residents to keep the sense of place, self-identity and a sense of community. An open, transparent and public citizen participatory web platform can help local residents to join the urban planning to make their live fitting their need.

CONCLUSION

The loss of identity that plays a vital role in urban structure is one of the problems of modern urban spaces. This paper reveals how data mining technologies can be used to provide designers with quantified data relating to the quality of spaces, which helps designers explore the important features of the urban identities of Asian cities. The Taiwan workshop demonstrates that with the aid of these quantified features, the designer can identify whether their designs preserve the local traditional urban structure. At the same time, these proposed designs are publicly accessible by citizens via a web platform, which

Design Proposal	Feature Value					
	AWBC Kurtosis	AWCC Median	AWCC Skewness	Compactness	Building Density	Area
1	12.58	542.67	-1.4236	0.08	614	5670
2	16.1331	553.85	-1.51123	0.1	633	11814
3	11.95	486.2	-1.7587	0.09	745	4280
4	7.88	98.87	-1.7678	0.1	666	9327

Table 4
List of feature value

Design Proposal	Feature Level and Label						
	AWBC Kurtosis	AWCC Median	AWCC Skewness	Compactness	Building Density	Area	Predicted Label
1	H*	M	-M	L	H	M	Yellow
2	H*	M	-M	L	H	H	Yellow
3	H*	M	-M	L	H	M	Yellow
4	H	L	-M	L	H	H	Purple

Table 5
List of feature level and label

allows them to participate in urban planning to make their cities more responsive and livable.

Future research should focus on exploring different cultural aspects that are unique to Asian or European cities in identifying their urban identity. These aspects can be integrated into guidelines for urban planning to develop suitable and customized rules for cities with different cultures, which will contribute to building sustainable cities.

ACKNOWLEDGEMENT
The authors would like to thank Asst. Prof. Shi-Yen Wu and all students who participated in the workshop BIG DATA-INFORMED URBAN DESIGN FOR THE SMART CITY: WORKSHOP 2017 for their contribution to this paper by using our methodology to demonstrate how it works. The authors would also like to thank Mr. Liu and the Taipei Urban Redevelopment Center for their kind help in providing information on the site.

REFERENCES
Batty, M. 2001, 'Exploring Isovist Fields: Space and Shape in Architectural and Urban Morphology', *Environment and Planning B: Planning and Design*, 28, pp. 123-150

Chang, M.-C., Buš, P. and Schmitt, G. 2017 'Feature Extraction and K-means Clustering Approach to Explore Important Features of Urban Identity', *Proceedings of the 16th IEEE International Conference on Machine Learning and Applications*, Cancun, Mexico, pp. 1139-1144

Dash, M. and Liu, H. 1997, 'Feature selection for classification', *Intelligent Data Analysis*, 1, pp. 131-156

Freund, J. E. and Peries, B. M. 2006, *Modern Elementary Statistics (12th Edition)*, Prentice Hall

Hai, Z. and Junsheng, Z. 2010, 'Topological Centrality and its e-Science Applications', *Journal of the Association for Information Science and Technology*, 61, pp. 1824-1841

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

Lynch, K. 1960, *The image of the city*, MIT Press, USA

Turner, Alasdair and Penn, A. 1999 'Making isovists syntactic: isovist integration analysis', *Proceedings of of Second International Symposium on Space Syntax*, Universidad de Brasilia, Brazi, pp. 01.1-01.9

[1] A.,C.,&ChairofInformationArchitectureETHZurich.(2018).QuickUrbanAnalysis.Retrievedfromhttps://qua-kit.ethz-ch

[2] Abdulmawla,A.,Bielik,M.,Bus,P.,Chang,M.-C.,Dennemark,M.,Fuchkina,E.,etal.(2018).DeCodingSpaces.(BauhausUniversit□¤tWeimar)Retrievedfromwww.decodingspaces-toolbox.org