Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 9033021, 10 pages
https://doi.org/10.1155/2021/9033021

Research Article

Construction of a Visual Saliency Model for Neighborhood Building Landmarks Based on K-Means Clustering

Chen Li [1,2] **and Zheng Qiao** [1]

[1] *School of Architecture, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China*
[2] *School of Design and Art, Xijing University, Xi'an, Shaanxi 710012, China*

Correspondence should be addressed to Chen Li; 20120066@xijing.edu.cn

Received 15 July 2021; Revised 9 August 2021; Accepted 13 August 2021; Published 24 August 2021

Academic Editor: Yi-Zhang Jiang

In this paper, firstly, based on the quantitative relationship between K-means clustering and visual saliency of neighborhood building landmarks, the weights occupied by each index of composite visual factors are obtained by using multiple statistical regression methods, and, finally, we try to construct a saliency model of multiple visual index composites and analyze and test the model. As regards decomposition and quantification of visual saliency influencing factors, to describe and quantify these visual significance factors of the landmarks, the significant factors are decomposed into several quantifiable secondary indicators. Considering that the visual saliency of the landmarks in the neighborhood is reflected by the variance of the influencing factors and that the scope of the landmarks is localized, the local outlier detection algorithm is used to solve the variance of the secondary indicators. Since the visual significance of neighborhood building landmarks is influenced by a combination of influencing factors, the overall difference degree of secondary indicators is calculated by K-means clustering. To facilitate the factor calculation, a factor-controlled virtual environment was built to carry out the experimental study of landmark perception and calculate the different degrees of each index of the building. The data of visual indicators of the neighborhood buildings for this experiment were also collected, and the significance values of the neighborhood buildings were calculated. The influence weights of the indicators were obtained by using multiple linear regression analysis, the visual significance model of the landmarks of the neighborhood buildings in the factor-controlled environment was constructed, and the model was analyzed and tested.

1. Introduction

The construction status of neighborhood buildings is one of the important symbols of urban development, and the density and trend of the distribution of neighborhood buildings in a city contain information about the development of a city, which can be a key object for studying the development status of a city [1]. The automatic extraction technology of neighborhood buildings has an important role in urban development planning, land-use change monitoring, disaster monitoring, early warning, and national defense. The rapid extraction of urban neighborhood building information can provide certain guidance to the management and is of great value in promoting the construction of a smart digital modern city. Along with the rapid development of space and aviation platforms as well as communication information, sensors, and other self-research and development technologies, it has been able to provide high spatial and temporal resolution as well as large-scale remote sensing observation of the Earth's surface [2]. Among them, the spatial resolution has been able to provide space observation up to the decimeter level, so that in addition to the traditional spectral information in the satellite remote sensing images can obtain more detailed characterization images of the features, such as fine texture, clear appearance, and shape information. The performance of the neighborhood buildings in the images can show detailed information such as the external outline of the neighborhood buildings more clearly and accurately, and its ability in the detailed description is nearly up to the level of aerial

images [3]. However, unlike aerial imagery, which is relatively expensive to acquire, remote sensing satellite imagery is relatively of low cost, but its data volume is extremely large. Clustering analysis only clusters similar things together, without caring too much about what a particular class is. Generally, some distance or similarity between samples is the basis for cluster analysis, so that similar (or close) samples are clustered together, and those not similar (or distant) samples are grouped into different classes [4]. Clustering is a way to mine the structure and features implied within a large amount of data by such a division method to further analyze the required knowledge and information. Clustering analysis is a common method in the preprocessing step of data mining and is the core of data mining.

At present, the traditional calculation mode of small data is continued in the data application and service system, resulting in the problems of human and material. And financial resources consumed in the information transfer between remote sensing data acquisition and distribution to the thematic applications for end-users, which urgently need to propose more efficient methods and technical systems systematically to give engineering solutions. It is a hot research topic in the field of high-resolution remote sensing image analysis and application, but there are still obvious limitations when existing methods or algorithms are applied to the object recognition of high-resolution remote sensing ground cover [5, 6]. The recognition tasks for different elements have different emphasis on feature selection and require sufficient industry expertise and rich a priori knowledge, which leads to the lack of universal application of the method. The object recognition of image targets in high-definition images is mainly realized by multiscale segmentation technology, but it is still difficult to fully perform the task of automated information extraction for complex feature targets, and there is no systematic and efficient method or a technical system to give engineering solutions. The random probability selection strategy and the positive and negative pheromone feedback mechanism in the algorithm ensure the diversity of solutions and the ability of the algorithm to jump out of the local optimal solution, and the update of pheromones along with the iteration of the algorithm ensures the continuous evolution of the algorithm [7]. The parallel distributed computation in the ant colony algorithm ensures that the ant colony algorithm has strong robustness.

The main idea and framework of the study are to qualitatively select the factors affecting the visual salience of neighborhood building landmarks from the perspective of spatial objects and navigators and to describe, quantify, and calculate the degree of difference of the factors by combining the quantitative methods of related fields. By building and carrying out the experiments of visual saliency assessment of neighborhood building landmarks, we construct the mathematical model of visual saliency and indicators of neighborhood building landmarks with the help of statistical methods, build a multifactor composite virtual environment to carry out the experiments of landmark cognition, calculate the variance value of each factor indicator and the visual saliency value of buildings based on the previous two parts of the study, and use the method of multiple regression analysis to get the weight of each factor. Finally, we construct a multifactor composite visual saliency model and test the validity of the model. Chapter one is dedicated to the introduction. The research background of the paper is discussed in terms of the demand for navigation applications and the development of K-means clustering technology, the purpose, and significance of the research are described, and the organization of the paper is given. Chapter two is dedicated to the related work. The current research status of visual saliency of neighborhood building landmarks and K-means clustering in landmark selection is specifically analyzed, and the research content of this paper is proposed by analyzing the shortage of the current status. Chapter three is dedicated to the study of the visual saliency model of neighborhood building landmarks based on K-means clustering. The visual indicators related to the visual saliency of neighborhood building landmarks are selected by statistical analysis, and the mathematical model of visual saliency of neighborhood building landmarks and K-means clustering is constructed. Chapter four is dedicated to research analysis. Multiple linear regression methods were used to obtain the influence weights of indicators, and, finally, an attempt was made to construct a model of visual saliency of building landmarks in a factor-controlled environment, and the model was analyzed and multidimensionally tested. Chapter five is dedicated to summary and prospect. The research work of this paper is summarized, and the innovation points and the problems that still need further research in the future are discussed.

2. Related Work

The K-means algorithm is an unsupervised clustering algorithm. It is relatively simple to implement and has a good clustering effect, so it is widely used. There are a large number of variants of the K-means algorithm, including initialization optimization K-means++, distance calculation optimization Elkan K-means algorithm, and optimization Mini Batch K-means algorithm in the case of big data. The deterministic algorithm converts the landmark visual saliency problem into an optimization problem and converts the local pattern matching in the landmark visual saliency of the video sequence into a cost function minimization problem, the most representative deterministic algorithm is the K-means clustering algorithm, and the advantages of the K-means clustering algorithm are fast convergence, being used in the landmark visual saliency of the high frame rate, and being very suitable for the landmark visual saliency analysis of real-time scenes when a considerable number of landmark visual saliency algorithms are based on the improvement of K-means clustering algorithm [8]. However, the K-means clustering method also has its drawbacks; for example, it is difficult to cope with the scale change and shape change of the target in the landmark visual saliency model acquisition process, easy to be influenced by the similar background and the interference of light change, and easy to occur in the building surface clustering in the

building surface process. Thyagharajan and Kalaiarasi added the scale estimation module in the framework of the classical K-means clustering algorithm, which can cope with the landmark visual. For the problem of insufficient feature description capability when the scale change of the target occurs in the process of significance model construction and the process of adding two prior values as regular terms in the feature extraction, which is used to cope with the case of small and drastic changes in the scale of the building surface target, for the case of the constant size of the building surface target, the reverse size consistency check is used to ensure the correctness of the target scale [9]. Eraslan et al. improved the mean clustering algorithm in combination with local three-value pattern texture features by introducing the least-square median in the local three-value pattern texture feature extraction process to achieve adaptive thresholding of local three-value pattern texture feature extraction while integrating particle filtering and mean clustering into the same building surface framework to cope with the occlusion problem in building surface scenes [10].

High-resolution remote sensing images contain rich feature detail information, and the recognition and extraction of buildings, as one of the important artificial features, play an important role in urban planning and disaster prediction [11]. In recent years, with the rise of artificial intelligence, deep neural networks have been widely used in the fields of computer vision and image processing, and significant progress has been made in remote sensing image classification and segmentation tasks, including building detection, boundary extraction, and regularization as well as change detection [12]. However, the complexity of the surrounding environment, the structural diversity of the buildings themselves, the data sources, and the shooting conditions all put the performance and accuracy of automatic building extraction to the test. Ji et al. proposed a K-means clustering network algorithm with a hierarchical densely connected nested network architecture to solve the problem that current DNN models lose local information in the downsampling operation and the upsampling method cannot correctly recover structural information [13]. Liu et al. proposed an effective end-to-end visual saliency model for K-means clustering which extracts both dense and multiscale features through dense spatial pyramidal pooling (DSPP), which helps to extract buildings at all scales [14]. Song et al. proposed a visual saliency model that uses separable factorization residual blocks as well as inflated convolution, aiming to guarantee a small accuracy loss with low computational cost and memory consumption [15]. Although high-resolution remote sensing images provide rich feature information, they also bring about a large amount of significant noise. In addition, the differences in spectral features and spatial features structural characteristics between different building clusters in remote sensing images all put a high test on the efficiency as well as the accuracy of accurately achieving building instance recognition [16, 17].

The method based on the feature rule set starts from the actual attributes of the features in the remote sensing image. It needs to manually observe the features of each feature and analyze the differences to establish a feature rule set. This method is more subjective. The results of classification extraction will also be greatly affected, and the accuracy may be high or the accuracy may be low. It is mainly determined by the choice of feature rule set [18]. At the level of urban planning, socioecological niche construction can be understood as determining the appropriate urban site and reasonable construction scope through reasonable master planning, urban design, and single-unit design methods, planning specific use functions according to the economic development goals, making the building groups actively adapt to society and improve the surrounding social environment, creating an urban ecological environment conducive to the progress of urban civilization, economic development, and regional cultural construction [19, 20]. The building is planned according to the economic development goals. Through the construction and improvement of the basic conditions of the area where the building is located, the building can make reasonable use of the existing urban ecological location while actively exploring the potential architectural ecological location, continuously improving the habitat conditions of the urban supertall building life body, and promoting the harmonious development of architecture and environment [21, 22].

3. Study on Visual Saliency Model of Neighborhood Building Landmarks Based on K-Means Clustering

3.1. Quantification of Visual Saliency Factors of Neighborhood Building Landmarks. Feature selection strategy inside the field of feature engineering is an important part of feature processing. Feature engineering usually includes feature application methods, feature extraction methods, and feature processing methods, and feature processing is the core application content of feature engineering. Feature processing is divided into single feature processing, multifeature processing, dimensionality reduction operation of feature space, data variation of feature space, and feature selection strategy [23]. In this paper, the filtering method is used as the feature selection method in the feature selection strategy. The feature selection method uses the basic criterion by variance comparison to select the building video image feature with the largest variance as the approximate representation of the target image.

The human visual system gives different attention to different regions in a scene, and a region containing a unique and well-defined target may be assigned more attention, while many similar regions may be assigned less attention. The process of generating saliency maps based on visual saliency algorithms treats salient regions in an image in a similar way, assigning saliency probabilities to each pixel location in the image. The saliency value of a pixel point Z in an image is defined as the probability that the pixel point belongs to a salient region, assuming that the location of the pixel point x in the image is L, the feature description is T, the local information is LI, the global information is GI, C is a constant, and the saliency value is defined as $f(x)$, where $f(x)$ is calculated as shown in the following equation:

$$f(x) = \ln P(T, LI, GI) + \ln P\left(T, LI, GI|\frac{C}{L}\right). \quad (1)$$

The features extracted from the building video images should be able to discriminate well between the block building targets and backgrounds in the video sequence images, which is the fundamental element of video target image feature selection. The predefined fixed set of features cannot always keep the block building target and background with high discriminative ability. Adaptive selection of feature combinations with the high discriminative ability for targets and backgrounds is especially important, and feature selection based on log-likelihood ratios helps to solve this problem, which can still ensure the high discriminative ability for targets and backgrounds of block buildings in video sequences when block building video sequences are switched or when block building scenes are changed, improve the feature description ability of target images, and ensure the stability of the video target block building algorithm. The log-likelihood ratio can be calculated by the histogram of the features in the target and background regions [24]. The log-likelihood ratio can associate the target image in the video sequence with positive values and the background image in the video sequence with negative values, and the frequency of each pixel value of foreground and background images in the feature histogram can be calculated as follows:

$$\begin{cases} f(\text{bin}) = \dfrac{S_f(\text{bin})}{\text{num}(\text{fg})}, \\ g(\text{bin}) = \dfrac{S_b(\text{bin})}{\text{num}(\text{bg})}, \\ \min(f(\text{bin}), g(\text{bin})). \end{cases} \quad (2)$$

In equation (2), S_f represents the feature histogram of the target image in the video sequence image, S_b represents the background feature histogram of the target image in the video sequence image, num(fg) is the number of pixels in the candidate target region in the video sequence image, num(bg) is the number of pixels in the background region in the video sequence image, and bin is the dimension of the normalized feature histogram. The log-likelihood ratio of the features of a certain video sequence image can be expressed as equation (3), where β is a relatively small penalty factor, f(bin) is the frequency of pixel values of the target image in the video sequence appearing in the feature histogram, and g(bin) is the frequency of pixel values of the background image in the video sequence appearing in the feature histogram, and then the combination of features with the largest variance is found among the set of all extracted features.

$$M(\text{bin}) = \max\left(\frac{\max(f(\text{bin}), \beta_M)}{\max(g(\text{bin}), \beta_M)}, \log \min\left(\frac{f(\text{bin})}{g(\text{bin})}, \beta_M, -1\right)\right). \quad (3)$$

By calculating the log-likelihood ratio variance of each feature in the building video image, the descriptive ability of each feature for the target image can be well evaluated. The descriptive ability of these features for the target image is ranked by the size of the log-likelihood ratio variance of the features, and the feature with the largest log-likelihood ratio variance is considered to be the feature with the strongest descriptive ability for the target image, which will eventually be used for the video feature description of the target image region.

The spectral brightness values of building roofs in the high-definition remote sensing images are more uniform compared with other background features such as vegetation, while the roofs of buildings are made of various materials and some buildings have solar panels on the roofs, which may lead to different spectral features on the top. Usually, the spectral features of buildings in high-resolution images are represented by five types of features: spectral mean, maximum difference, ratio, standard deviation, and brightness. Textures on the building surface in the high-resolution images are mainly artificial textural features with human involvement and different from the surface properties of naturally existing objects. It is mainly represented by four angles in mean, homogeneity, standard deviation, heterogeneity, information entropy, correlation, contrast, and angular second-order moments with a total of eight feature parameters [25]. Quantitative expressions are made in terms of the mean and gray size, local variations of gray details in the image, and their degree of variation and homogeneity. The size and geometry of buildings in the high-definition remote sensing images are very complex, but, in most cases, the shape of buildings can be regarded as regular and quantitative geometric shapes such as rectangles or combinations of rectangles and circles. The specific expression includes roundness, principal direction, area, aspect ratio, compactness, density, and shape index. The same light and scale of all surface features in the same remote sensing image make the main direction of shadows represented by the corresponding buildings in the image consistent when the buildings in the image have the same main direction among them, while the group of buildings in most cases will be perpendicular or parallel to the surrounding roads.

The cluster analysis method can divide the data into different categories according to the relationship between the data, so that the objects in the same cluster have greater similarity, but the objects between different clusters have great dissimilarities. The purpose of offline clustering is to classify the stability of buildings, and, through offline clustering, the offline clustering centroid is obtained. In this paper, the building environment of the block is divided into 4 hazard levels. The 4 cluster centroids are obtained, and each cluster centroid represents the state of the building (hazard level). The meanings of the 4 hazard levels are shown in Figure 1.

3.2. Construction of Visual Saliency Model Based on K-Means Clustering. With the increasing building height and density, super high-rise buildings are like trees in the forest competing for sunlight, and the sunlight problem is becoming

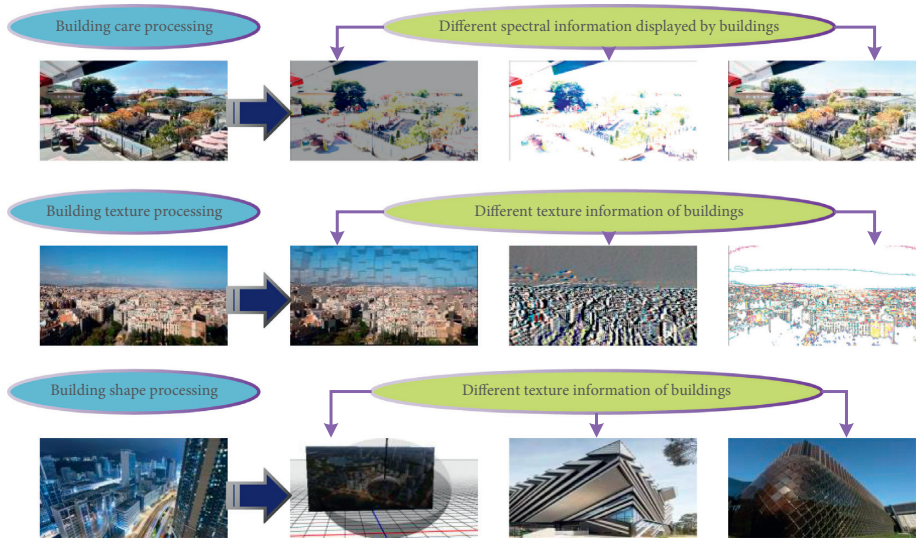

Figure 1: Information on the characteristics of the building.

more and more prominent. Sunlight is an inexhaustible renewable energy source, and the sunlight problem not only directly affects building lighting but also plays a very important role in indoor and outdoor hygiene. The full spectrum of natural light has an important effect on human physiology and psychology, and, in medicine, sunlight therapy can even effectively treat some people who suffer from depression due to the lack of sunlight in winter [26, 27]. The blocking of sunlight by buildings forms shadows at the same time. For a single building, the length of shadows formed in the effective sunlight time is relatively small due to the large sun height angle, and the shadows move faster and will not produce absolute blocking of sunlight in the same direction for a long time, but when they are arranged in groups, the overlapping of multiple shadows will produce a large area of sunlight dead space, which is not conducive to the recovery of ecological environment. To make all kinds of buildings and sites in the building group enjoy reasonably sufficient sunlight, the layout, height control, and orientation of super-tall buildings will be important elements in the construction of the natural ecological position of super-tall buildings.

Through computer simulation, architects can know in advance the impact of neighborhood buildings on the surrounding wind environment to prevent wind hazards through different methods, reduce the danger of pedestrians, build a wind vein for the high-density urban areas, and create a public environment that meets the requirements of pedestrian comfort. For the building monolith, when the wind encounters the blockage of the neighborhood building, it will also produce a high-speed downward wind field-headwind vortex, which will have a greater impact on the ground street. To solve these problems, we mainly need to make changes to the texture of the windward facade of the building, reduce the smooth surface from the bottom to the top, and reduce the impact of downward wind speed by increasing the friction.

The cluster analysis method can divide the data into different classes based on the relationship between the data so that there is a large similarity between objects in the same cluster and a large dissimilarity between objects in different clusters. The purpose of offline clustering is to classify the lateral stability of vehicles, and, by offline clustering, the offline clustering prime is obtained. In this paper, the neighborhood building environment is divided into 4 hazard classes, 4 clustering plenums are obtained, and each clustering plenum represents a kind of vehicle driving state (hazard class). The meanings represented by the 4 hazard classes are shown in Table 1.

K-means clustering algorithm is the most commonly used clustering algorithm. Suppose that two n-dimensional vectors are $M = (m_1, m_2, \ldots, m_n)$ and $N = (n_1, n_2, \ldots, n_n)$; then the Euclidean distance between two points A and B is as follows:

$$d(M, N) = \sqrt{\sum_{k=1}^{n} (m_k - n_k)^2}. \tag{4}$$

The ants use not only the heuristic function but also the pheromone between the sample points and the clustering center when choosing the path. The pheromone distribution between the sample and the clustering center is w_{kh}, and the pheromone distribution is between the sample and the clustering center. In the search process of the algorithm, the probability of sample points being assigned to each cluster center is calculated by the formula shown in equation (5), where α and β denote the relative importance of pheromone and heuristic factor, respectively, M is the total number of ants ($h \in [-1, M]$), $q \in [-1, 1]$ is a given parameter, randomly generated $R \in [-1, 1]$, and t is the number of iterations. $A_m(i)$ is the set of samples outside the taboo table, and k is the kth element of the taboo table, the kth sample traveled by ant m. The visual saliency of a landmark is determined by the observer, the environment, and the geographical object. It is produced by the interaction of the three. During the navigation process, the observer is located in the environment.

Table 1: Meanings of hazard classes.

Risk level	Risk	Meaning
PR1 level	Low risk	Safety
PR2 level	Medium risk	Existing security
PR3 level	Medium-to-high risk	Typical potential hazard
PR4 level	High risk	Danger

Based on the sensory input of the surrounding environment and the current task (sightseeing, walking to the destination), the navigator can quickly notice and distinguish salient spatial objects (those spatial images that are in sharp contrast with the surrounding environment) and use them as landmarks.

$$P_{kh} = \begin{cases} \dfrac{\min\left(w_{kh}(n)^{\alpha}, v_{kh}(n)^{\beta}\right)}{\sum_{h \in A_m(i)}^{M} w_{kh}(n)^{\alpha} * v_{kh}(n)^{\beta}}, & q < R, \\ \max\left(w_{kh}(n)^{\alpha}, v_{kh}(n)^{\beta}\right), & q \geq R. \end{cases} \tag{5}$$

After all ants in the algorithm finish constructing the solution, the clustering result obtained by the objective function is evaluated and only the ant with the best objective function value is retained each time, so the rough K-means clustering can be regarded as an optimization problem, and the process of optimization is the obtaining of the minimum value of the objective function, which is the optimal clustering result, as shown in equation (6). $D(j)$ denotes the intraclass distance, which is used to evaluate the degree of cohesion of the clusters. d (M, N) denotes the Euclidean distance; $w_l(k)$ and $w_b(k)$ are the lower approximation of the kth cluster and the weight value of the boundary region, respectively.

$$D(j) = \sum_{j=1}^{J} \left(w_l(k) + w_b(k) * \sum_{i=1}^{M} \left| x_i - m_k \right| \right). \tag{6}$$

The dynamic adjustment of the weights of the lower approximation and the boundary region in the clustering process can avoid the setting of empirical weights that leads to ignoring the variability of the data distribution. The number of elements in the lower approximation and the boundary region can measure the relative importance ratio, which is calculated by the following formula:

$$\begin{cases} w_l(k) + w_b(k) = 1, \\ \dfrac{w_l(k)}{w_b(k)} = \dfrac{M_k}{M_k - m_k}. \end{cases} \tag{7}$$

The positive feedback mechanism attracts more ants to the current optimal path and promotes the convergence of the clustering algorithm. However, when the ants search for a local optimal solution, the negative feedback mechanism eliminates the effect of the positive feedback mechanism to prevent more ants from being attracted to the path that leads to the final result of the local optimal solution. Through the positive and negative feedback mechanisms on the pheromones released by the ants, the ants are not limited to the local optimal solution in the process of searching for the best solution but can continuously find new solutions. The artificial ant in the ant colony algorithm uses the overall information of the ant colony, and the global update of the residual pheromone is performed only after the completion of an optimization search. The pheromone update formula on each path in the ant colony algorithm is (8), where $D(j)_{\min}$ is the intraclass distance when the objective function obtains the minimum value. $v(wkj)$ is the pheromone increment; M is the total amount of pheromone released by ants; and $D(j)$ $(0 < D(j) < 1)$ is the pheromone volatility coefficient.

$$\begin{cases} w_{kj}(n+1) = D(j) w_{kj}(n) + v\left(w_{kj}\right), \\ v\left(w_{kj}\right) = \dfrac{W}{D(j)_{\min}}. \end{cases} \tag{8}$$

Initially, the ant randomly selects M sample points as the starting point, calculates the probability of selecting each cluster center for that sample according to the random probability selection strategy concerning the heuristic function and the number of pheromones on the path, and then determines the class to which it should belong according to the probability. Then the ants randomly select another sample and repeat the above process until they have traversed all samples, and a solution is formed. After all the ants finish constructing the solution, the optimal value is evaluated and retained using the objective function; then the approximate region weights and the boundary region weights are calculated, and, finally, the center of mass of each cluster is recalculated. The pheromone is updated globally, and only the paths of the ants with the best clustering results are pheromone-increased during the iteration of the algorithm, while the paths of the remaining ants being pheromone-decayed.

To jump out of the local optimum of K-means, this paper proposes the MK-means algorithm, in which the algorithm adds a random exploratory vector to the class center of ordinary K-means in the following way as shown in equation (9), where β(k) is the current D-dimensional random explorer vector in the iteration.

$$M(x)_k^{*} = M(x)_k + \beta_k, \quad x \in [-1, 1]. \tag{9}$$

Experimental evaluation and tuning of the hyperparameters are required during the model training process. The hyperparameters of the model include learning rate, iteration number epoch, and Anchor Scales. The experimental results of hyperparameters are used to determine the optimal hyperparameters of the model. The Anchor Scales should be selected according to the size of the image and the target, and the size of the Anchor Scales corresponds to each feature map. When the number of samples selected for one training, the size of which is very important for the performance and speed of the network model, is too small, this will cause the network not to converge easily and affect the training speed of the model, and when it is too large it will

cause the data to lack randomness and fall into local optimum. Considering the computer load and operation efficiency, this experiment needs to divide the data into several smaller batches for input into the network. After inputting all training data into the network to complete the feature learning process once, the training data need to be learned iteratively several times to make the network model fit and converge. As an important super parameter in supervised learning and deep learning, the learning rate determines whether and when the objective function converges to a local minimum. Too large a learning rate can lead to missed optima, and the smaller the learning rate is, the slower the loss gradient decreases and the longer the convergence time is.

4. Results and Analysis

4.1. Visual Saliency Model Analysis. Clustering performance metrics are also called effectiveness indicators, which are divided into external indicators, which compare the clustering results with a reference model after clustering is completed, and internal indicators, which directly examine the clustering results without using any reference model. The external index is based on the prespecified structure to judge the results of the clustering algorithm. This structure reflects people's intuitive understanding or prior knowledge of the data structure. In the model verification of this paper, the clustering index chooses the visual saliency of landmarks as the external index, and the building saliency is used as the internal index. Combine the two for cluster analysis.

In terms of statistical tests, the more critical regression equation significance test and the regression coefficient significance test were looked at firstly. As mentioned earlier, the ANOVA of the final regression equation showed that the probability of F was 0.000, indicating that the equation was significant, the null hypothesis was rejected, and the variables included in the equation as a whole had a significant effect on the dependent variable. The test results are shown in Figure 2, and it is found that the probability of t corresponding to the 10 variables included in the final equation is less than 0.05; they all have a significant effect on the dependent variable; and then look at the goodness of fit, for the multiple regression equation needs to look at the adjusted goodness of fit, according to the description of the model summary information in the previous section, the adjusted goodness of fit of the final model is 0.307, although it is not as high as many models in economics. Although it does not reach the high fit of many models in economics, the model can be considered reasonable at the exploratory level if it can achieve a goodness of fit of 0.3 or higher as a result of the exploratory study. The test fully demonstrates the rationality and validity of the model of the relationship between the visual salience of landmarks and differences in building characteristics from practical, theoretical, and logical perspectives.

In this paper, the algorithm can make a good judgment of the degree of blurring and modify the blurred chunks in the blurred image frames, and it can still achieve good clustering of the target when serious blurring occurs. The experimental data are shown in Figure 3, from which it can

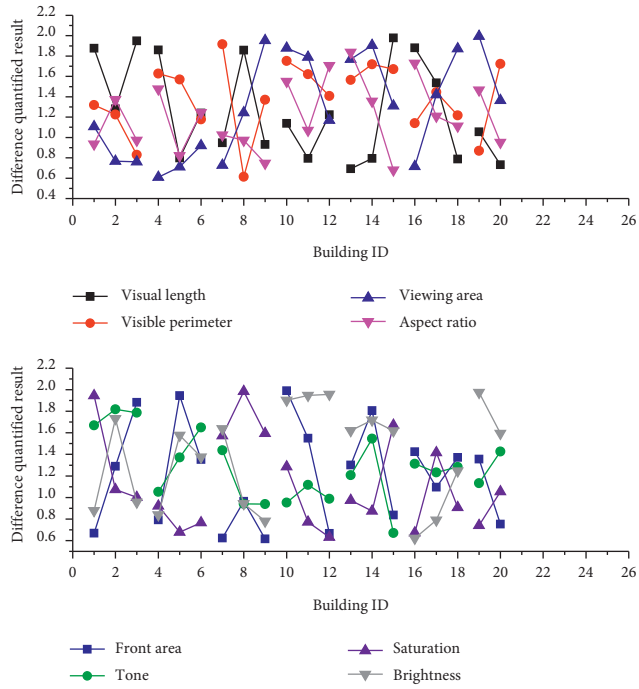

Figure 2: Differentiated metrics of building attributes.

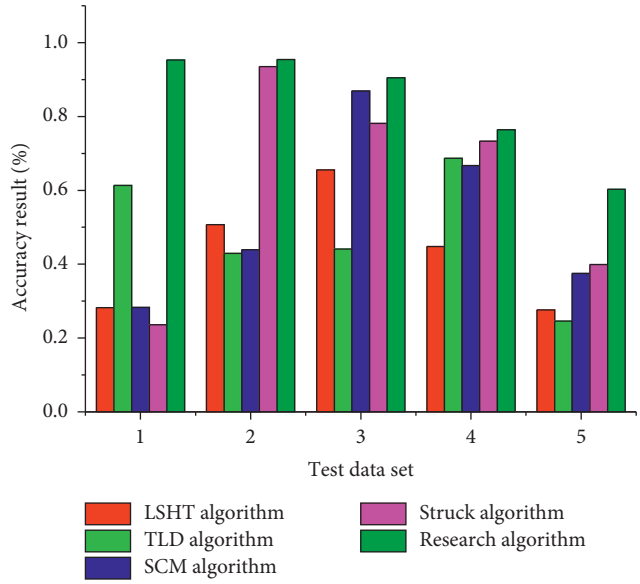

Figure 3: Accuracy comparison experimental results.

be seen that the accuracy of the algorithm in the fuzzy test sequence is higher than other algorithms, and the algorithm can achieve real-time target clustering and meet the real-time standard of target clustering. It can be concluded that, in the same complex neighborhood building scene, this paper's algorithm has higher accuracy and real-time performance compared with several other algorithms.

Struck, SCM, and ASLA algorithms have good stability when the target occlusion range is relatively small; however, localization failure occurs when the range of encountered

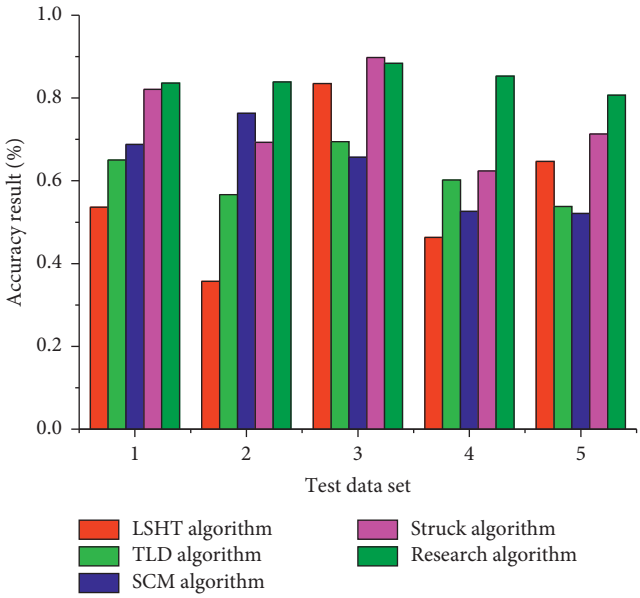

Figure 4: Accuracy results of each algorithm in the masking sequence.

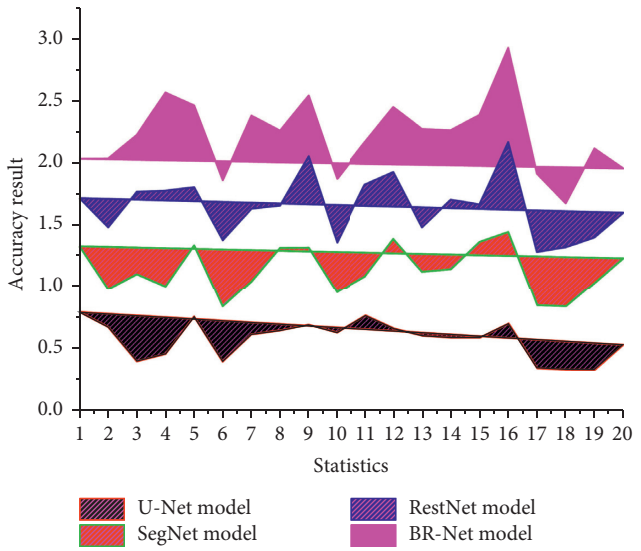

Figure 5: Accuracy of region extraction results.

target occlusion becomes large. The experimental results show that the visual saliency of neighborhood buildings based on K-means clustering has obvious advantages and achieves stable and robust clustering under the occlusion environment. The central position error is used to evaluate the clustering effect of the four algorithms, the central error of the algorithm in this paper is always kept low, and the experimental result data are shown in Figure 4. From Figure 4, it can be seen that the accuracy of this paper is higher than other algorithms in most cases, and the algorithm of this paper shows a good clustering effect.

According to the accuracy statistics in Figure 5, the accuracy distribution of the U-Net model is 35.3%–47.9%, with an average accuracy of 39.1%. The extraction accuracy distribution of the SegNet model is 47.9%–66.8%, with an average accuracy of 54.9%; the extraction accuracy distribution of the ResNet model is 57.1%–74.9%, and the average accuracy is 63.1%. The average precision of the BR–Net model is 76%, and the average precision of the BR-Net model is 71.2%–80.8%. In the extraction of neighborhood buildings in the four clustering models in the six categories of accuracy statistics, all have a significant decrease compared to the accuracy statistics of other buildings.

4.2. Visual Saliency Clustering Analysis. The buildings in the experimental area are staggered, and the background is mostly a vegetation area. The spectral characteristics of building roofs are similar, and they are clusters of small buildings, except for the buildings in region I, where the whole building is detected as multiple small buildings. The overall confidence level of the building examples in the experimental region IV reached 0.965. The result of the construction example confidence evaluation is shown in Figure 6.

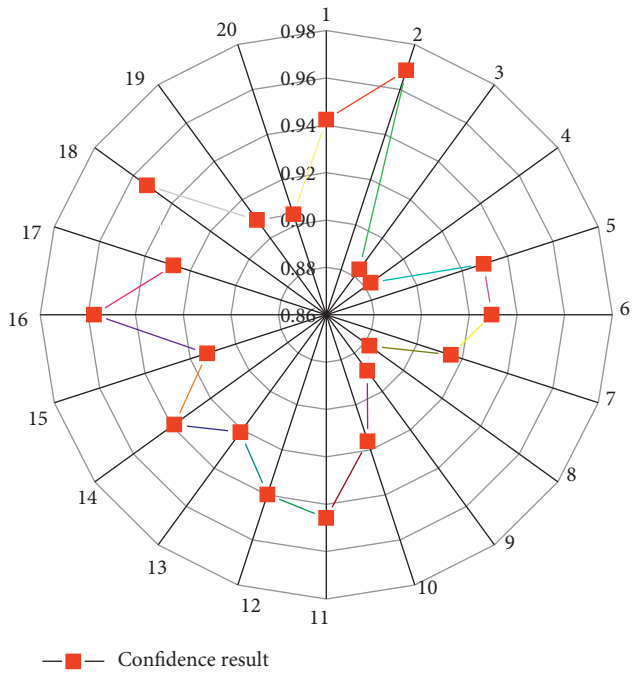

Figure 6: Construction example confidence evaluation.

The clustering results of the four areas in Figure 7 show that the clustering results of the single buildings and small building clusters are staggered in area A. The clustering effect of the single buildings is good, and the small building clusters can achieve the correct extraction of the building area targets, but there is a slight confusion of the building contours due to the irregular shape of the buildings and the prominent contiguity phenomenon. However, the clustering results show that the clustering profiles of buildings are very obvious and close to the labeled map and the original data, indicating that the model has excellent results in clustering

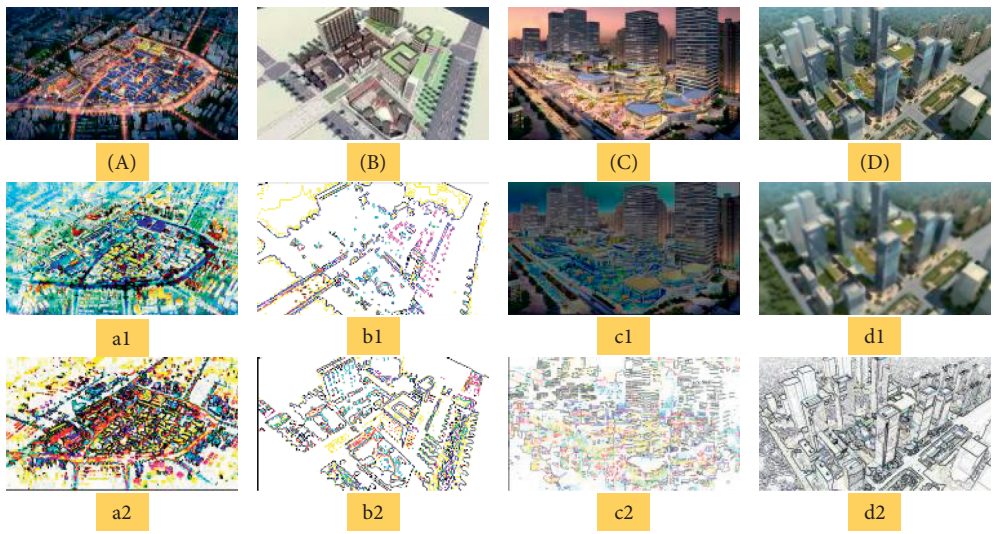

Figure 7: Results of building clustering in the test area.

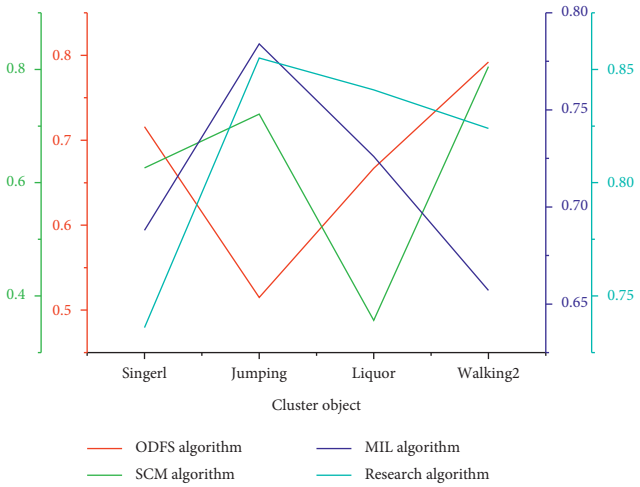

Figure 8: Success rate comparison results.

single buildings. The clustering results are very intuitive and close to the labeled map and the original image.

The experimental data of the clustering results of each video sequence are shown in Figure 8, from which the following can be seen, among the test sequences selected for this paper: Singer1 in the block building clustering scene where lighting changes, jumping in the block building clustering scene where blurring occurs, and Liquor and Walking2 in the two test sequences where serious background interference occurs. The success rate of this algorithm is higher than that of other algorithms; therefore, this algorithm has better clustering performance compared with the other algorithms.

5. Conclusion

In the extraction of building contours from high-resolution remote sensing images, K-means clustering has achieved good results in image segmentation with its advantage of being able to quickly cluster and analyze a large number of typical images. At the same time, due to the advantage of K-means clustering to automatically cluster a large number of features, it can reduce a large amount of manual analysis cost and image feature extraction cost in the calculation of massive high-resolution remote sensing images. However, due to the complexity of buildings in high-resolution images and between buildings and various background features, no algorithm can extract buildings in high-resolution remote sensing images with absolute accuracy. In this paper, we try to innovatively add the boundary information of buildings into the training of deep learning, enhance the network performance with boundary constraints, and extract buildings by using the efficient performance of K-means clustering, aiming to improve the accuracy of building information extraction in high-resolution images and provide rapid data support for modern urban construction and other aspects. A virtual experimental environment based on multiple visual factors is constructed to carry out visual saliency experiments, and the weights are obtained by regression analysis, and a visual saliency model based on multiple visual factor composites is constructed. Firstly, we build a factor-controlled virtual environment to carry out landmark cognition experiments and then calculate the different degrees of each index of buildings based on K-means clustering; meanwhile, we collect the visual index data of this experiment and calculate the significance value of buildings based on the mathematical relationship constructed by K-means clustering; finally, we use the method of multiple linear regression analysis to obtain the influence weights of the indexes and finally try to construct a factor-controlled environment under. Finally, we try to construct a visual significance model of building landmarks in a controlled environment and analyze and test the model in multiple dimensions.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The study was supported by "Xi'an Social Science Planning Fund Project, China (Grant no. 19S31)."

References

[1] Y. Shi, C. Otto, and A. K. Jain, "Face clustering: representation and pairwise constraints," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1626–1640, 2018.

[2] S. Lynen, B. Zeisl, D. Aiger et al., "Large-scale, real-time visual-inertial localization revisited," *The International Journal of Robotics Research*, vol. 39, no. 9, pp. 1061–1084, 2020.

[3] A. Kumar, G. S. Walia, and K. Sharma, "A novel approach for multi-cue feature fusion for robust object tracking," *Applied Intelligence*, vol. 50, no. 10, pp. 3201–3218, 2020.

[4] K. S. Arun, V. K. Govindan, and S. D. M. Kumar, "Enhanced bag of visual words representations for content based image retrieval: a comparative study," *Artificial Intelligence Review*, vol. 53, no. 3, pp. 1615–1653, 2020.

[5] M. Bouchakwa, Y. Ayadi, and I. Amous, "A review on visual content-based and users' tags-based image annotation: methods and techniques," *Multimedia Tools and Applications*, vol. 79, no. 29, pp. 21679–21741, 2020.

[6] L. G. Nonato and M. Aupetit, "Multidimensional projection for visual analytics: linking techniques with distortions, tasks, and layout enrichment," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 8, pp. 2650–2673, 2018.

[7] Z. Mehmood, M. Rashid, A. Rehman, T. Saba, H. Dawood, and H. Dawood, "Effect of complementary visual words versus complementary features on clustering for effective content-based image search," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 5, pp. 5421–5434, 2018.

[8] A. B. Vasudevan, D. Dai, and L. Van Gool, "Talk2nav: long-range vision-and-language navigation with dual attention and spatial memory," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 246–266, 2021.

[9] K. K. Thyagharajan and G. Kalaiarasi, "A review on near-duplicate detection of images using computer vision techniques," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 897–916, 2021.

[10] G. Eraslan, Ž. Avsec, J. Gagneur, and F. J. Theis, "Deep learning: new computational modelling techniques for genomics," *Nature Reviews Genetics*, vol. 20, no. 7, pp. 389–403, 2019.

[11] N. Allen, "Concepts of neighbourhood: a review of the literature," *Development*, vol. 8, no. 2, pp. 96–115, 2018.

[12] H. Wang, D. Zhao, and H. Ma, "Informative image selection for crowdsourcing-based mobile location recognition," *Multimedia Systems*, vol. 25, no. 5, pp. 513–523, 2019.

[13] Z. Ji, Y. Yang, F. Wang, L. Xu, and X. Hu, "Feature encoding with hybrid heterogeneous structure model for image classification," *IET Image Processing*, vol. 14, no. 10, pp. 2166–2174, 2020.

[14] L. Liu, F. Nie, A. Wiliem, Z. Li, T. Zhang, and B. C. Lovell, "Multi-modal joint clustering with application for unsupervised attribute discovery," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4345–4356, 2018.

[15] F. Song, T. Dan, R. Yu et al., "Small UAV-based multi-temporal change detection for monitoring cultivated land cover changes in mountainous terrain," *Remote Sensing Letters*, vol. 10, no. 6, pp. 573–582, 2019.

[16] H. Liu, Q. Zhao, J. T. Mbelwa, S. Tang, and J. Zhang, "Weighted two-step aggregated VLAD for image retrieval," *The Visual Computer*, vol. 35, no. 12, pp. 1783–1795, 2019.

[17] X. Ma, Y. Zhao, X. Qian, and Y. Y. Tang, "Multi-source fusion based geo-tagging for web images," *Multimedia Tools and Applications*, vol. 77, no. 13, pp. 16399–16417, 2018.

[18] L. E. Carvalho, A. C. Sobieranski, and A. von Wangenheim, "3D segmentation algorithms for computerized tomographic imaging: a systematic literature review," *Journal of Digital Imaging*, vol. 31, no. 6, pp. 799–850, 2018.

[19] D. Xu, Y. Shi, I. W. Tsang et al., "Survey on multi-output learning," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 7, pp. 2409–2429, 2019.

[20] Y. Xiao, Z. Tian, J. Yu et al., "A review of object detection based on deep learning," *Multimedia Tools and Applications*, vol. 79, no. 33, pp. 23729–23791, 2020.

[21] F. Wang, Y. Ma, Y. Jin, Y. Jiang, and Y. Wang, "Retracted article: discovering graphical visual features for abnormal semantic event detection," *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 3245–3260, 2018.

[22] R. Liu, Y. Zhao, and S. Wei, "Enhance neighbor reversibility in subspace learning for image retrieval," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1338–1350, 2018.

[23] C. Silberer, J. Uijlings, and M. Lapata, "Understanding visual scenes," *Natural Language Engineering*, vol. 24, no. 3, pp. 441–465, 2018.

[24] C. Sur, "Survey of deep learning and architectures for visual captioning-transitioning between media and natural languages," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 32187–32237, 2019.

[25] M. Backman, E. Lopez, and F. Rowe, "The occupational trajectories and outcomes of forced migrants in Sweden. Entrepreneurship, employment or persistent inactivity?" *Small Business Economics*, vol. 56, no. 3, pp. 963–983, 2021.

[26] L.-Y. Duan, Y. Wu, Y. Huang, Z. Wang, J. Yuan, and W. Gao, "Minimizing reconstruction bias hashing via joint projection learning and quantization," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3127–3141, 2018.

[27] A. Valls, K. Gibert, A. Orellana, and S. Antón-Clavé, "Using ontology-based clustering to understand the push and pull factors for British tourists visiting a Mediterranean coastal destination," *Information & Management*, vol. 55, no. 2, pp. 145–159, 2018.