WILEY | Hindawi

Research Article

Interactive Multimodal Television Media Adaptive Visual Communication Based on Clustering Algorithm

Huayuan Yang [ID] **and Xin Zhang**

School of Culture and Media, Huanghuai University, Zhumadian City, Henan Province 463000, China

Correspondence should be addressed to Huayuan Yang; 20151507@huanghuai.edu.cn

This article starts with the environmental changes in human cognition, analyzes the virtual as the main feature of visual perception under digital technology, and explores the transition from passive to active human cognitive activities. With the diversified understanding of visual information, human contradiction of memory also began to become prominent. Aiming at the problem that the existing multimodal TV media recognition methods have low recognition rate of unknown application layer protocols, an adaptive clustering method for identifying unknown application layer protocols is proposed. This method clusters application layer protocols based on similarity of the load characteristics of network stream application layer protocol data. The method divides the similarity calculation in the clustering algorithm to improve the clustering efficiency of the algorithm. Experimental results show that the proposed method can efficiently and accurately recognize unknown visual communication. This article proposes that, in the interactive multimodal visual information transmission, human visual perception experience has changed, the diversity of visual information content expression makes the aesthetic subject more personalized and stylized.

1. Introduction

With the popularization of social informatization and digitization, the visual forms we face have also undergone changes [1]. The most important change is the shift from single modality to multimodality in the dissemination of visual information. Information dissemination under single mode is mainly performed by a single means or form, such as graphic mode, image mode, sound mode, video mode, and interactive mode; information dissemination under multimode can be two or more single modes combined with each other to form a means of communication and a form of carrying to complete information transmission [2]. The main reason for this change is that, with the development of science and technology, audiences have higher and higher requirements for information carriers, and the manifestations of information have become more complex and diverse. In the communication of this diversified form of visual information, it is necessary to consider not only the form's own performance capabilities, expression methods, form beauty, and the content contained in the form but also people's visual perception response, group psychological characteristics, and individual life experience and visual experience [3]. At the same time, in an environment based on the Internet and supported by digital technology, technical factors are also an important cause of its development. Multimodality is exactly the form of information dissemination that appears in this digital context. The Internet is an important part of the daily life of the audience. It is this change in lifestyle that has gradually changed the way of information dissemination [4]. The original single-modal, linear form of communication has changed to a multimodal, nonlinear form and digital technology. It also gives this form more room for development and expressiveness.

Visual communication is the main means of human cognition of things, and visual perception includes image perception, image recognition, and spatial perception, which is to distinguish a certain image from many stimuli in the

objective environment, recognize familiar images, and generate three-dimensional space [5]. Whether it is graphic perception, image recognition, or spatial perception, this is the human perception of cognitive objects. This perception can be graphic images, shapes and colors, or the movement of objects. If you think that this is the subject's visual perception of the object, you are wrong. As far as human visual cognition is concerned, in addition to cognizing the object itself, the environment in which the object is located is also an important component [6]. Different environments make humans have different cognitive processes. A black circle of the same size, located in a square environment of different sizes or surrounded by circles of different sizes, gives people the impression that they have different shapes and sizes. Straight lines of the same length have the illusion of different growths due to different environments. In recent years, researchers have introduced machine learning methods into the field of network traffic identification. At present, there are many researches on application layer protocol recognition methods based on supervised learning. Chen and Cheung [7] proposed a traffic recognition algorithm based on an adaptive BP neural network, which achieved a high protocol recognition accuracy. According to Vryzas et al. [8], the convolutional neural network is applied to the field of network traffic recognition and classification, the useless information that may affect feature extraction is eliminated through traffic cleaning, and the classification accuracy is improved. Dash et al. [9] proposed a recognition and classification method based on LeNet-5 deep convolutional neural network and obtained the optimal classification model by cyclically adjusting relevant parameters. Perveen [10] proposed a distance-based nearest neighbor recognition method, which can improve the low performance of other methods in the recognition of unbalanced network traffic. Jain studied the protocol recognition of convolutional neural networks trained by different optimizers. The experimental results show that the Stochastic Gradient Descent (SGD) optimizer produces the best recognition effect and proposed a protocol recognition method for wireless communication networks. First, a one-dimensional convolutional neural network is used for automatic feature extraction, and then, application layer protocols are classified based on SVM.

The above research uses classification models such as neural networks and uses labeled protocol data for model training. The trained model can more accurately identify the application layer protocol network traffic. However, if the protocol specification of the application layer protocol is unknown, it is difficult to identify the corresponding network traffic using this type of method.

Under the influence of digital technology, visual forms have developed (single modal to multimodal), and the environment in which humans perceive things has gradually changed. Under traditional circumstances, human visual perception is produced in a real environment of light and shadow, and light and shadow are the preconditions for visual perception. Wahl S said that light and shade are an attribute of individual and independent objects, and light exposure provides a common basis for the existence of all

objects [11]. With this foundation, objects and parts of objects can emerge from the dark abyss. Light forms a certain environmental atmosphere through brightness, color, and strength, which affects the subjective visual perception of the audience. In the process of visual communication, some environments are deliberately formed by humans, and the purpose is nothing more than strengthening the intensity of information expression and strengthening the intensity of visual perception. In other words, the environment also has a "meaningful form." The visual cognition of the environment is mainly realized by the contrast of colors, light and shadow symbols, subject, and background. In the new digital context, human visual perception environment has changed and virtuality is its main feature. The visual cognition environment under digital virtual reality relies more on computers, data helmets, ring screens, data gloves, holographic projection, and other technical equipment to achieve. The human factor of this virtual environment is much larger than that of the real environment. Although there are also human designs in the real environment, most of them are guided by the situation, making the information transmission more purposeful.

This paper proposes that human visual perception experience has changed when interactive multichannel visual information is transmitted. The diversified expression of visual information content also increases the interaction and information between people and, meanwhile, makes the aesthetic experience of the aesthetic subject become more and more personalized and stylized.

2. Visual Communication Analysis under Multimodal Information

2.1. Diversified Understanding of Visual Information. Traditional web page visual communication design methods have the shortcomings of one-way transmission of visual data, lack of the actual communication process between web pages and users, and poor visual communication, resulting in web page display monotonous, even user disgust, user page views, and other problems [12]. The user behavior tracking system can evaluate and extract user behaviors, recommend pages that users can generate interest, and achieve better human-computer interaction through the continuous feedback of users on the pages and the learning of the interest behavior tracking system, which solves the one-way transmission of visual data problem.

In order to solve the defects of one-way visual data transmission and visual monotony in traditional web page visual communication design methods, this paper proposes a web page visual communication design method based on user's personalized characteristics. First, analyze the user's excitement color characteristics, obtain web page graphic design solutions, and solve the monotonic problem of traditional web page visual design for different visual designs displayed by users with different aesthetics, effectively reducing the user's boringness of web pages, and then combine with the interest behavior tracking system. The user's visual excitement is combined and integrated with the computer vision adaptive module and adaptive

mechanism. Through the self-learning of the computer, the interactive behavior of the web page and the user is continuously enhanced [13]. Finally, the experiment verified the web visual communication design method based on the user's personalized characteristics, and the user's browsing willingness, time, and satisfaction increased in different age groups, different working backgrounds, and different living environments.

As people gradually participate in the construction of the visual information ontology, they regard their own understanding of visual objects as part of the object and no longer have people and clouds. What follows is that people put forward higher requirements on the way of expressing visual information, so the multimodal representation of visual information came into being. It should be said that there is a complementary and mutually reinforcing relationship between the two. The schematic diagram of diversification of visual information is shown in Figure 1. Multimodal visual information expression methods have enriched human language expression ability, and human pursuit of the essence of information has further promoted the development of expression means. As far as cognitive psychology is concerned, all human behavioral awareness can be understood as the problem-solving process of information objects. From the input of information, the encoding and processing of information, and the output of information, its understanding of information is seen.

Today, in the digital environment, people's cognition of information, especially visual information, has reached the active stage. That is to say, people are not only limited to the cognition of visual information itself but also require further design and processing under self-understanding to form vision, expression, and then enter the mass communication.

Cognitive psychology is called information processing theory because it uses information processing as the core to understand and explain the relationship between human high-level thinking activities and information processing. In this relationship, multimodal visual information is perceived by human visual perception, which is completed through perception, attention, understanding, thinking activities, visual language, and artificial intelligence.

2.2. Changes in Thinking Activities. People's thinking activities are not only affected by the cognitive environment but also by the means and methods of information transmission. When our neurons feel more information, the greater the intensity of the information, the greater the thinking response produced by the brain. In the information monomodal state, the brain receives relatively single information, relatively single thinking activities, and lower levels [14]. In the multimodal information mode, the brain receives more information channels and stronger means. It was originally only a graphic mode to transmit information, and it became a combination of multiple modalities such as sound, text, graphics, images, and video; the audience can understand information from perspectives, multiple levels, and multiple dimensions, the thinking activities have

become stronger and deeper, and the audience's thinking has become more affected. The visual perception activities under information multimodality are shown in Figure 2.

3. Adaptive Visual Communication Algorithm

3.1. Algorithm Overview. This article takes the application layer protocol data in network communication as the analysis object. The network data of the same protocol have a certain similarity, which can be used to distinguish different application layer protocols. The method in this paper first reorganizes the network stream from the collected original network data, extracts the application layer protocol data of the network stream, and calculates the similarity of the protocol data. The similarity between the application layer protocol data is used as the basis for protocol identification [15, 16]. Then, an improved hierarchical clustering algorithm is used to adaptively cluster the application layer protocol data of the network flow and automatically identify unknown application layer protocols.

Specifically, as shown in Figure 3, the proposed application layer protocol identification method includes the following processing steps:

(1) Data preprocessing: the collected network traffic data are processed, and the network traffic data are converted into byte streams through substeps such as data filtering and sorting, stream reorganization, and application layer protocol data extraction.

(2) Similarity calculation: intercept the fixed-length bytes at the front of the application layer protocol data, and calculate the similarity between different application layer protocol data.

(3) Unknown application layer protocol clustering: initialize the application layer protocol data, calculate the similarity between clusters through the similarity algorithm between clusters, and use the improved clustering algorithm to iterate repeatedly until the clustering stop condition is reached. The application layer protocol data are gathered in a cluster, and finally, a cluster set is output. Each cluster in the set is a set of network flow information corresponding to an application layer protocol [16].

3.2. Unknown Application Layer Protocol Clustering. This article has improved on the basis of the hierarchical clustering method, and the improvement is mainly reflected in the calculation of similarity. The traditional hierarchical clustering algorithm will repeatedly calculate the similarity between data objects when calculating the similarity between clusters. This paper divides the similarity calculation into the similarity calculation between the application layer protocol data before clustering and the similarity between clusters in the cluster [17]. Complete the similarity calculation between complex application layer protocol data before clustering, and then save the result in the array. When calculating the similarity between clusters, when the similarity between
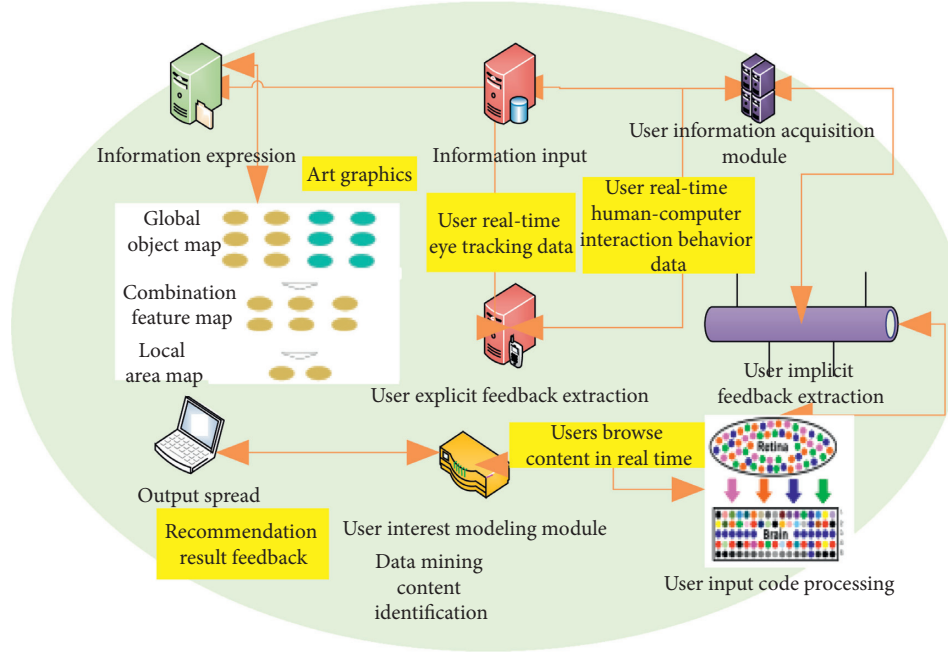
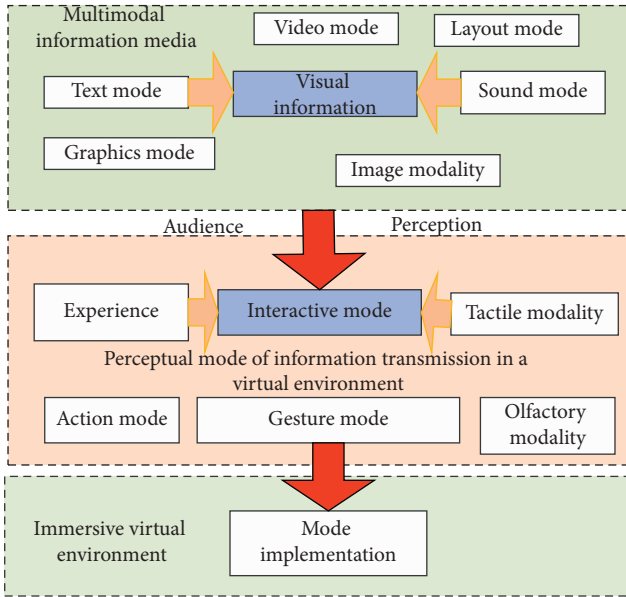Figure 1: Schematic diagram of diversification of visual information.

Figure 2: Visual perception activities under information multimodality.

application layer protocol data needs to be calculated, only the value can be extracted from the array, which simplifies the calculation of similarity in clustering and improves the efficiency of clustering. Figure 4 is a schematic flow diagram of the improved hierarchical clustering method used in this article.

The improved hierarchical clustering algorithm process used in this paper includes cluster initialization, calculation of similarity between clusters, comparison of similarity values, and cluster merging. The application layer protocol data obtained by data preprocessing are used as the input of

the algorithm to initialize the protocol data [18]. The specific operation is to save the protocol data of each network stream independently and add the initial cluster mark. The protocol data of each network stream belong to a different cluster after cluster initialization.

Clustering algorithms include hierarchical clustering algorithms, partitioned clustering algorithms, and clustering algorithms based on density and grid. At present, in the field of protocol identification, researchers mostly use partition clustering algorithms. The K-means algorithm was used to cluster the mixed data composed of a small number of labeled samples and a large number of unlabeled samples, and several clusters were obtained. Then, use the K-nearest neighbor algorithm combined with the labeled samples in the cluster to identify unlabeled samples. The experimental results show that this classifier can achieve better classification results for unbalanced network flows. Kang et al. [19] applied the EM algorithm to network traffic classification, but it can only perform protocol identification roughly, and the accuracy rate is low. Aimed at the shortcomings of EM algorithm's strong initial value sensitivity and easy convergence to a local optimal solution, a protocol identification method based on improved EM was proposed. This method narrows the search range and improves the accuracy of protocol identification. The problem of multipath partitioning is to construct a classifier based on the idea of graph theory and finally identify and classify the protocol based on the results of the graph partition. First, the constrained clustering algorithm is used to extract new patterns of unlabeled data, and these patterns do not exist in the tagged data and only represent unknown protocols; then, the new patterns based on the labeled data and unlabeled data are trained into a binary classifier, which determines the protocol type of the sample data according to the result of the classifier.
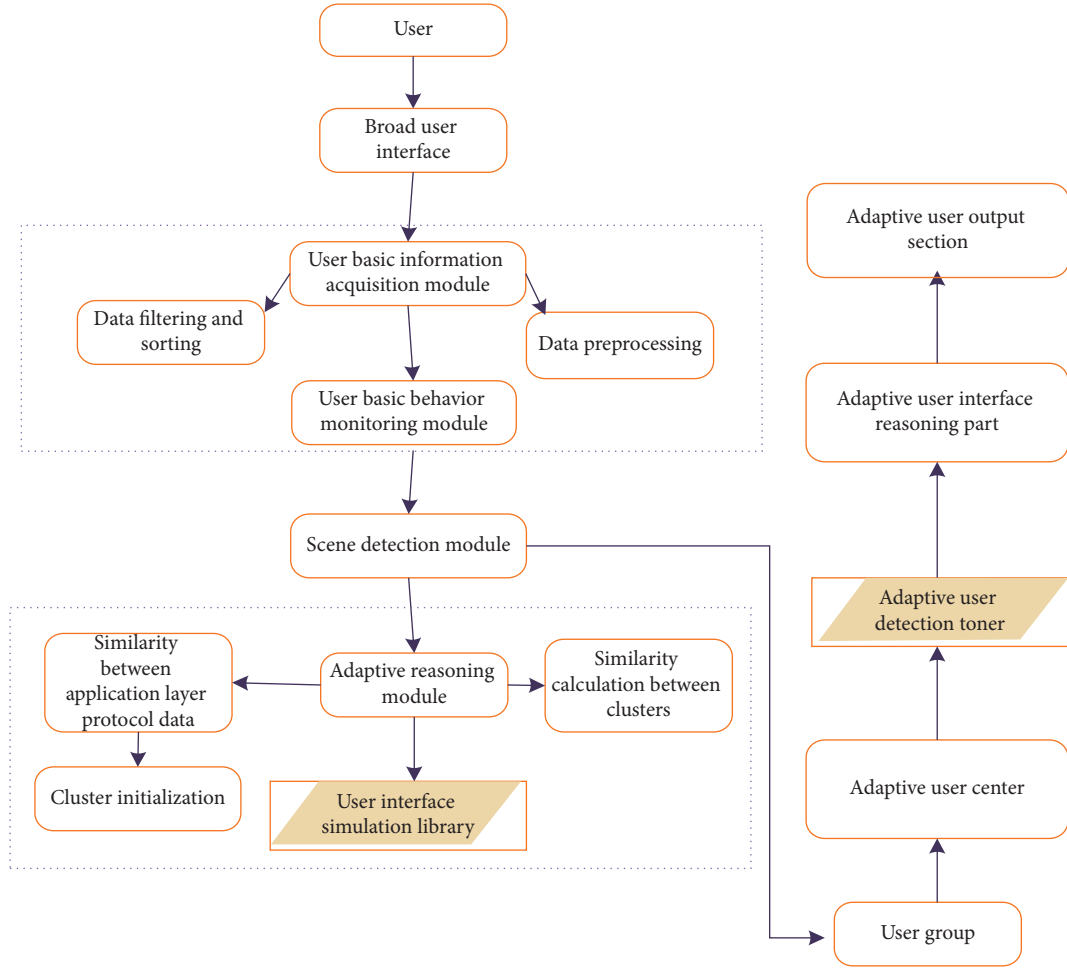
Figure 3: Processing flow of unknown application layer protocol identification based on adaptive clustering.

In general, although the existing clustering-based network traffic methods can classify unknown protocol network traffic, most methods need to input the number of target clusters, and the accuracy of protocol recognition greatly affected by the number of target clusters cannot automatically classify the unknown protocol traffic, and the practical application has greater limitations [20].

Take the cluster set as the input of the similarity algorithm between clusters, and choose two clusters. First, calculate the average of the similarity between each protocol data in the cluster and all the protocol data in the other cluster, as the difference between the protocol data and the other cluster. Then, calculate the mean value of the similarity between all protocol data in the cluster and another cluster as the relative similarity between clusters. Finally, the mean value of the relative similarity between the two clusters is calculated to obtain the similarity between the clusters [21, 22].

For example, choose two clusters when clustering:

$$
\begin{aligned}
A_1 &= \{a_1, a_2, \ldots, a_n\}, \\
B_1 &= \{b_1, b_2, \ldots, b_m\}.
\end{aligned} \tag{1}
$$

First, calculate the similarity ($\sum_{i=1}^{n}$ Similarity $(a_i, b_j)/n$) between the protocol data $\{a_1, a_2, \ldots, a_n\}$ in cluster A_1 and cluster A_2, denoted as Similarity (a_i, A_2). Among them, similarity (a_i, b_j) is the similarity of the application layer protocol data $\{a_1, a_2, \ldots, a_n\}$ to the application layer protocol data $\{b_1, b_2, \ldots, b_m\}$, and m is the total number of application layer protocol data contained in the cluster A_2.

Then, calculate the relative similarity between A_1 and A_2: ($\sum_{i=1}^{n}$ Similarity $(a_i, A_2)/n$) denoted as Similarity (A_1, A_2), and n is the total number of protocol data contained in cluster A_1. Repeat the above steps to calculate the relative similarity of cluster A_2 to A_1. Finally, the average similarity between cluster A_1 and cluster A_2 is obtained ((Similarity (A_1, A_2) + Similarity (A_2, A_1))/2).

If the similarity between clusters is greater than the similarity value, merge the two clusters and update the cluster set. The specific method is to select two similar clusters: $A_1 = \{a_1, a_2, \ldots, a_n\}$ and $B_1 = \{b_1, b_2, \ldots, b_m\}$. Based on any one of the clusters, add all the data in the other cluster.

For example, based on A_1, the combined result is $A_1 = \{a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m\}$. Then, delete cluster A_2 from the cluster set, which completes the operation of merging the cluster set.
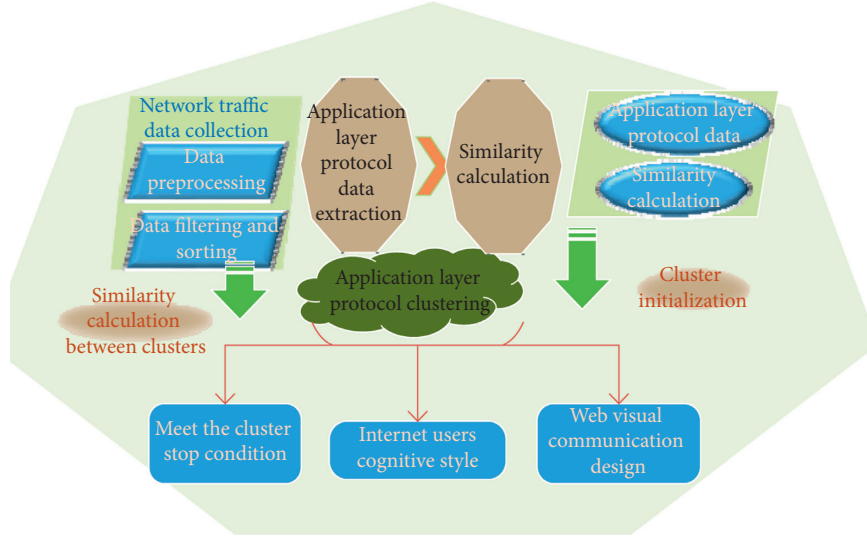
Figure 4: Improved hierarchical clustering algorithm flow.

Repeat the steps of calculating the similarity between clusters, comparing the similarity values, and merging clusters until the algorithm meets the cluster termination condition. The termination condition of clustering is generally that the similarity between clusters in the cluster set is less than the value of similarity, and clusters cannot be merged. Then, output the cluster set; each cluster contains all network flow information corresponding to an application layer protocol.

As we all know, the aesthetic style of visual art comes from the hands of professional graphic and image designers and artists, but with the popularity of pop art worldwide in the late 1960s, it broke the boundaries between life and art and broke the boundary between elegance and vulgarity and has made art move towards the direction of individualization and popularization [23]. This process has been infinitely accelerated and amplified by the Internet. The people's life, study, work, etc., are all affected by it. The pictures and videos taken by mobile phones at any time are quickly displayed. Information clusters are formed around things, and these information clusters also form a situation where visual expressions flourish because of the difference in the aesthetic background of the designers [24].

The multimodality of visual language has led to differentiation between audiences, forming a group of groups. These groups gather together because of different occupations, differences in beliefs, knowledge, education, backgrounds, ethnicities, and regions, and their aesthetic perceptions are like this. Because of individual differences between similar social groups, individualized characteristics are presented in the visual expression of information.

4. Results and Analysis

With the gradual changes in the human cognitive environment, in the virtual environment, the audience's expressions facing the object are presented in a variety of forms; just like the representation of graphic images, there can be several different art forms to express each kind of

aesthetic experience. It is not the same as the visual experience. With so many forms and combinations of forms, it brings a variety of choices to the audience.

4.1. Cluster Analysis of Adaptive Visual Communication.
The algorithm in this paper involves two important parameters: one is the similarity threshold between clusters, which is the minimum similarity between two clusters in the clustering algorithm that can be merged into one cluster; the other is the length of the intercepted message, which affects the algorithm. The completeness of the feature short sequence is automatically extracted. In this paper, the influence of two parameters on the clustering effect is tested by the controlled variable method [25].

4.1.1. The Influence of Similarity Threshold between Clusters on Clustering Accuracy.
This article uses protocol traffic extracted from the data set for testing. First, test the impact of similarity threshold between clusters on clustering accuracy. The intercluster similarity threshold is the minimum intercluster similarity required for two clusters in a cluster to be merged into one cluster. The size of the threshold will affect the clustering effect. The length of the intercepted application layer protocol data set in the experiment is 60 bytes. The test result is shown in Figure 5. It can be seen from the figure that when the similarity threshold between clusters is 0.3, the clustering accuracy is the highest and the clustering effect is the best. The clustering accuracy of these protocols reached 100%. After the similarity threshold between clusters is greater than 0.3, the clustering accuracy of the three protocols begins to decrease; when the threshold is close to 1, the clustering accuracy of HTTP and FTP protocols drops to close to 0%, while the clustering accuracy of FTP protocol is still low, which can be maintained above 50%.

After analysis, although the test data only intercepts the 60-byte data in the front of the protocol, it still contains some
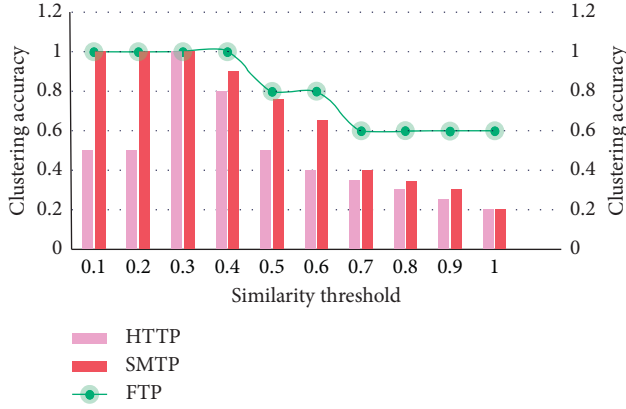
FIGURE 5: The effect of closed value of similarity between clusters on clustering accuracy.

FIGURE 6: The effect of intercepted message length on clustering accuracy.

user data. The effect of closed value of similarity between clusters on clustering accuracy is shown in Figure 5. When calculating the similarity between clusters, it will have a certain impact on the similarity results. The similarity of similar protocols is about 0.3. Therefore, as the similarity threshold between clusters approaches 0.3, the clustering effect becomes better and better. With the similarity threshold close to 1, the clustering effect becomes worse.

4.1.2. The Impact of Intercepted Application Layer Protocol Data Length on Clustering Accuracy. The length of the intercepted application layer protocol data determines the number of protocol feature short sequences and the number of user data contained in the intercepted data and affects the accuracy of calculation of similarity between application layer protocol data. In this paper, the similarity threshold between clusters is set to 0.3, and the clustering accuracy rate under each length is obtained by changing the length of the intercepted application layer protocol data. The selection range of the interception length is (10,100), and the stride length is 10. Figure 6 shows the test result. As can be seen from the figure, when the similarity threshold between clusters is 0.3, the clustering effect is best when the intercepted application layer protocol data length is equal to 10 bytes or greater than or equal to 60 bytes. The effect of intercepted message length on clustering accuracy is shown in Figure 6. The SMTP protocol is most affected by the intercepted data length. The FTP protocol is hardly affected by the length of the intercepted data.

After analysis, it is believed that the protocol data are an alternating mixture of the protocol feature short sequence and the user data. Although the feature short sequence is in the front of the protocol data, the front part still contains some user data, and the user data are less feature data than most protocols. When the length of the intercepted data is 10 bytes, all the intercepted data are the protocol feature short sequence or the protocol feature short sequence is obviously more than the user data, so the accuracy of the clustering result is very high. As the length of the intercepted data increases, the proportion of user data in the intercepted data
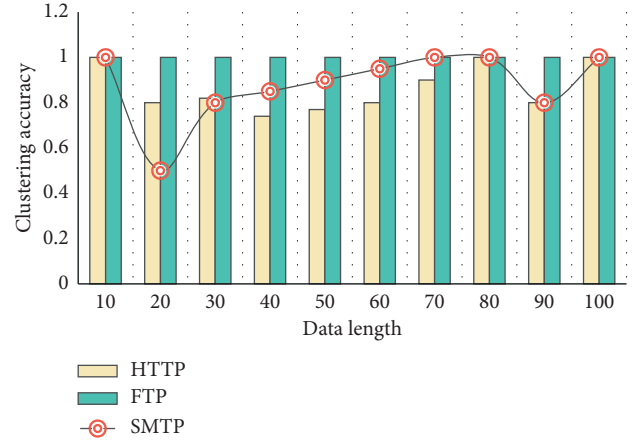
gradually exceeds the protocol feature short sequence, which causes the clustering accuracy to begin to decline. At the same time, because the feature short sequence and the user data are alternately arranged, when the intercepted data length is further increased, the proportion of the protocol feature short sequence will gradually increase, resulting in the clustering accuracy rate increasing again.

4.2. Clustering Accuracy Test. This article implements the algorithm through Python and then enters the test data set. The algorithm needs to manually set two parameters: the similarity threshold between clusters and the intercepted application layer protocol data length [26]. The test experiment selects the intercluster similarity threshold as 0.3, and the intercepted application layer protocol data length is 60 bytes. Finally, the clustering results are shown in Table 1.

The information dissemination of visual language has changed from one way to two ways with the support of digital technology. One-way communication refers to the process from encoding to decoding of visual information, such as the traditional reading and newspaper reading, listening to the radio, and watching TV. Two-way communication refers to the two-way reversibility of the process from visual information input to information processing to information dissemination. With the intervention of digital technology, this situation has become a process of two-way change in the subject and object of information. The subject changes its perspective due to differences in its life experience, knowledge, education, and cognition. Make use of their own feedback information to the object, so that the connotation of the object change more rich, and this change will act on the information object in time with the support of modern technology. Even some social media, the Internet, and chat platforms will increase the interaction between information and audiences to the multiparty interaction of audience-information-audience. This multimodal information communication method allows humans to understand with the maximization of promotion, and the essence of things is also explored from more angles by this way of interaction. The comparison results between traditional

TABLE 1: Clustering results.

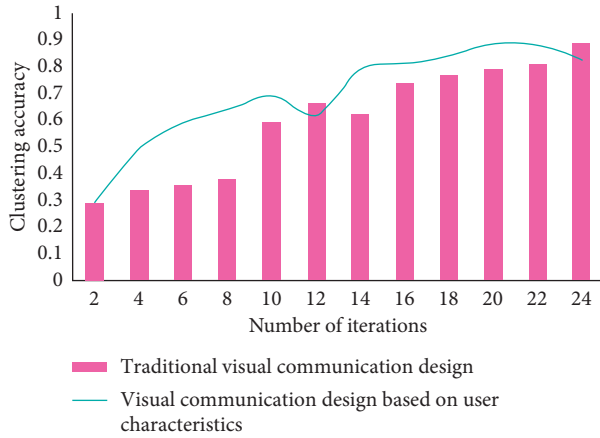| Cluster	Data similarity	Network flow number	Correct rate %	Error rate %
1	0.74	96	100.00	0.00
2	0.52	96	97.35	2.65
3	0.59	10	94.25	5.75
4	0.47	34	100.00	0.00
5	0.98	12	98.48	1.52
6	0.65	64	99.47	0.53
7	0.71	17	99.94	0.06
8	0.42	8	100.00	0.00
9	0.35	17	100.00	0.00
10	0.46	38	99.24	0.76

FIGURE 7: Experimental comparison results.

visual communication design and visual communication design based on user characteristics is shown in Figure 7.

All in all, in the era of picture reading, people's observation of things and their cognition of the essence of events have increasingly relied on more intuitive visual forms (graphics, images, and videos) and more and more through multimodality. The language model is realized through various modern digital media. People actively participate in the construction of information sources. They are both designers and disseminators as well as receivers. This active mentality also makes the connotation of information appear in a dynamic composition. Color is not a mode of expression under visual communication, and its information transmission is usually done with the help of a specific form. In the traditional single mode, the color is a subtractive mode, and the printing color is the main color. In the new media environment, in the process of multimodal communication mode, the performance of color has changed a lot, mainly in the application of this mode. This mode is a color mode in which red, green, and blue are superimposed and mixed, also called additive color mode. This mode is a screen display mode, which is mainly used for network multimedia display. Color shows almost all colors that the human eye can and cannot perceive. In addition to the number of colors, color intensity (brightness and saturation) is also a very important factor. In this mode, the color intensity has reached a new height, especially the combination of colors makes this

height range wider and larger. From mobile phones, computer displays to stereoscopic projections, and more and more color effects can be perceived by humans. Correspondingly, the visual perception under the multimodal visual language also has diversified characteristics. Humans can complete their personal aesthetic experience journey through human-computer interaction among a variety of choices.

In this paper, the accuracy of the algorithm clustering is obtained by comparing the result with the manual labeling result. It can be seen from the results that the algorithm in this paper successfully distinguishes the request network flow and response network flow of HTTP, FTP, SMTP, and custom protocols and clusters them into different clusters, with a clustering accuracy of 100%. The algorithm in this paper does not use terminal-day information for protocol differentiation but uses protocol application layer data load characteristics for protocol clustering. Therefore, this article can successfully distinguish between custom protocols that use terminal-day camouflage technology and FTP protocol to realize unknown protocol identification.

5. Conclusion

This paper proposes a method for identifying and classifying unknown application layer protocols based on adaptive clustering. First, the network stream is reorganized from the collected original network data, the application layer protocol data load characteristics of the network stream are extracted, and the application layer protocol data are calculated. Similarity is used as the basis for the identification and classification of application layer protocols, and the clustering algorithm is used to automatically cluster the application layer protocol data of network flows to efficiently and accurately realize the identification and classification of unknown application layer protocol network traffic. The method makes full use of the advantages of the clustering algorithm, avoids the training process, is efficient and accurate, and has high practical value.

All in all, in the era of multimodal television media, people's observation of things and their cognition of the essence of events have increasingly relied on more intuitive visual forms (graphics, images, and videos) and more and more through multimodality. The language model is realized through various modern digital media. People actively

participate in the construction of information sources. They are both designers and disseminators as well as receivers. This active mentality also makes the connotation of information appear in a dynamic composition. Correspondingly, the visual perception under the multimodal visual language also has diversified characteristics. Humans can complete their personal aesthetic experience journey through human-computer interaction among a variety of choices.

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 they have no conflicts of interest in publishing this paper.

References

[1] A. Belozertsev, K. Gusev, E. Grineko et al., "A multimodal quality assessment of compressed materials for adaptive bitrate video streaming," in *Proceedings of the 2018 International Symposium on Consumer Technologies (ISCT)*, vol. 1, no. 1, pp. 29–31, IEEE, St. Petersburg, Russia, May 2018.

[2] J. Coelho, F. Rito, and C. Duarte, ""You, me & TV"-fighting social isolation of older adults with facebook, TV and multimodality," *International Journal of Human-Computer Studies*, vol. 98, no. 8, pp. 38–50, 2017.

[3] A. Kohn, "Iconic situations: multimodality, witnessing and collective memory," *Visual Communication*, vol. 19, no. 1, pp. 73–97, 2020.

[4] F. van der Sluis, E. L. van den Broek, A. van Drunen et al., "Enhancing the quality of service of mobile video technology by increasing multimodal synergy," *Behaviour & Information Technology*, vol. 37, no. 9, pp. 874–883, 2018.

[5] T. E. Powell, H. G. Boomgaarden, and K. De Swert, "Video killed the news article? comparing multimodal framing effects in news videos and articles," *Journal of Broadcasting & Electronic Media*, vol. 62, no. 4, pp. 578–596, 2018.

[6] W. D. Feng, "Metonymy and visual representation: towards a social semiotic framework of visual metonymy," *Visual Communication*, vol. 16, no. 4, pp. 441–466, 2017.

[7] Z. T. Chen and M. Cheung, "Consumption as extended carnival on Tmall in contemporary China: a social semiotic multimodal analysis of interactive banner ads," *Social Semiotics*, vol. 2, no. 1, pp. 1–21, 2020.

[8] N. Vryzas, L. Vrysis, and R. Kotsakis, "Speech emotion recognition adapted to multimodal semantic repositories," in *Proceedings of the 2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, vol. 1, no. 2, pp. 31–35, IEEE, Zaragoza, Spain, January 2018.

[9] A. K. Dash, P. Patnaik, and D. Suar, "A multimodal discourse analysis of glocalization and cultural identity in three Indian TV commercials," *Discourse & Communication*, vol. 10, no. 3, pp. 209–234, 2016.

[10] A. Perveen, "Facilitating multiple intelligences through multimodal learning analytics," *Turkish Online Journal of Distance Education*, vol. 19, no. 1, pp. 18–30, 2018.

[11] S. Wahl, "Advertising travels via subject teleshopping–a linguistic and multimodal analysis," *Strategies of Adaptation in Tourist Communication*, pp. 221–240, Brill, Atlanta, Georgia, 2018.

[12] S. Tan, K. L. O'Halloran, P. Wignell et al., "A multimodal mixed methods approach for examining recontextualisation patterns of violent extremist images in online media," *Discourse, Context & Media*, vol. 21, pp. 18–35, 2018.

[13] W. Pearce, S. M. Özkula, A. K. Greene et al., "Visual cross-platform analysis: digital methods to research social media images," *Information, Communication & Society*, vol. 23, no. 2, pp. 161–180, 2020.

[14] J. Cox, R. Jones, and C. Northwood, "Object-based production: a personalised interactive cooking application," in *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, pp. 79-80, Hilversum, Netherlands, June 2017.

[15] T. Tehseem, M. Sibtain, and Z. Obaid, "Exploring gender stereotypes in media adverts: a multimodal analysis," *Journal of Social Sciences and Humanities*, vol. 57, no. 2, pp. 155–175, 2018.

[16] A. Abdel-Raheem, "Mental model theory as a model for analysing visual and multimodal discourse," *Journal of Pragmatics*, vol. 155, no. 5, pp. 303–320, 2020.

[17] J. Chepchirchir, "Determination of prominent representations of domestic products advertisements in kass TV on to kipsigis consumer in belgut sub-county, Kericho, Kenya," *East African Journal of Arts and Social Sciences*, vol. 2, no. 1, pp. 8–14, 2020.

[18] L. Vigoroso, F. Caffaro, and E. Cavallo, "Occupational safety and visual communication: user-centred design of safety training material for migrant farmworkers in Italy," *Safety Science*, vol. 121, no. 1, pp. 562–572, 2020.

[19] S. Kang, D. Kim, and Y. Kim, "A visual-physiology multimodal system for detecting outlier behavior of participants in a reality TV show," *International Journal of Distributed Sensor Networks*, vol. 15, no. 7, pp. 155-156, 2019.

[20] Y. Kautt, "Practice and structure as concepts of a sociology of visual communication: toward a methodology of visual and structural analysis," *Qualitative Inquiry*, vol. 25, no. 4, pp. 350–362, 2019.

[21] M. Vaishali, T. E. Powell, and M. Van Der, "A picture paints a thousand lies? the effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media," *Political Communication*, vol. 37, no. 2, pp. 281–301, 2020.

[22] K. M. Bos, M. A. Ismail, and N. M. Ghanem, "An enhancedvideo summarization using densityased spatial clustering," in Image Analysis and Processing ICIAP 2013.Lecture Notes inComputer Science, A. Petrosino, Eds., p. 8156, Springer, Berlin, Germany, 2013.

[23] J. Li and Y. Y. Lee, "Multimodal Interaction and Believability: how can we design and evaluate the next generation of IPA," in *Proceedings of the 31st International BCS Human Computer Interaction Conference*, pp. 1–4, Sunderland, UK, July 2017.

[24] K. M. Mahmoud, N. A. Ghanem, and M. A. Ismail, "Unsupervised video summarization via dynamic modeling-based hierarchical clustering," in *Proceedings of the 2013 12th International Conference on Machine Learning and Applications*, pp. 303–308, Miami, FL, USA, December 2013.

[25] P. Gang, J. Hui, and S. Stirenko, "User-driven intelligent interface on the basis of multimodal augmented reality and brain-computer interaction for people with functional disabilities," in *Proceedings of the Future of Information and Communication Conference*, pp. 612–631, Springer, Cham, Singapore, April 2018.

[26] N. S. Sreekanth, N. Varghese, and C. H. Pradeepkumar, "Multimodal interface for effective man machine interaction," *Media Convergence Handbook*, vol. 2, no. 1, pp. 261–281, 2016.