

Research Article

SIFT Feature-Based Video Camera Boundary Detection Algorithm

Lingqiang Kong 

Media & Communication College, Shandong University of Arts, Shandong, Jinan 250014, China

Correspondence should be addressed to Lingqiang Kong; konglingqiang@sdca.edu.cn

Received 7 January 2021; Revised 22 March 2021; Accepted 29 March 2021; Published 12 April 2021

Academic Editor: Wei Wang

Copyright © 2021 Lingqiang Kong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problem of low accuracy of edge detection of the film and television lens, a new SIFT feature-based camera detection algorithm was proposed. Firstly, multiple frames of images are read in time sequence and converted into grayscale images. The frame image is further divided into blocks, and the average gradient of each block is calculated to construct the film dynamic texture. The correlation of the dynamic texture of adjacent frames and the matching degree of SIFT features of two frames were compared, and the predetection results were obtained according to the matching results. Next, compared with the next frame of the dynamic texture and SIFT feature whose step size is lower than the human eye refresh frequency, the final result is obtained. Through experiments on multiple groups of different types of film and television data, high recall rate and accuracy rate can be obtained. The algorithm in this paper can detect the gradual change lens with the complex structure and obtain high detection accuracy and recall rate. A lens boundary detection algorithm based on fuzzy clustering is realized. The algorithm can detect sudden changes/gradual changes of the lens at the same time without setting a threshold. It can effectively reduce the factors that affect lens detection, such as flash, movies, TV, and advertisements, and can reduce the influence of camera movement on the boundaries of movies and TVs. However, due to the complexity of film and television, there are still some missing and false detections in this algorithm, which need further study.

1. Introduction

As video is the most complex of all multimedia data types, it not only contains the content of the still image but also includes the motion information of the target in the scene and the information of the objective world changing with time. The huge amount of data and the unstructured characteristics of film and television make it extremely difficult to carry out effective film and television retrieval. The traditional film and television retrieval mainly relies on the manual definition of the key words of the film and television. Although this retrieval method is simple, it has many disadvantages. For example, the rich content contained in film and television is difficult to describe comprehensively with concise words; it is a subjective behavior to mark a complete film and television segment by means of employing people. In general, for small film and television units, such as a lens of the film frame, we can only use fast forward, fast back, and other methods to determine the accurate location of

information, and it will cause additional transmission bandwidth overhead.

Lens boundary detection is the basis of the content-based film and television retrieval system. Many scholars and institutions are conducting research related to content-based film and television retrieval and have developed a variety of film and television data retrieval systems, reflecting the main achievements of film and television retrieval system research [1, 2]. QBIC (Query By image Content) system is a typical representative of the content-based retrieval system. It allows users to query large image and film databases using sample images, sketching, color and texture mode selection, lens and target motion, and other information. The functions of the system for film and television include automatic segmentation of shots, key frame extraction of shots, static sample query, and text and title content query [3, 4]. It provides a set of tools for people on the Web to search and retrieve images and video, and it can be realized in Internet content-based video/image retrieval [5, 6]. Video allows the user to use

visual characteristics and to retrieve the relationship between time and space and film and television and has integrated text and visual search method, automatic video object segmentation and tracking, and rich visual feature library, including color, texture, shape, and motion, interactive query, and browsing on the Internet [7, 8]. According to the two-paragraph algorithm based on film and television content and audio content, the film and television are automatically divided into a large number of film and television clips with logical semantics, and the title decoder and word indicator are added to extract text information, and the film and television are edited by indexing [9, 10]. Due to the complexity of lens, the problem of lens boundary detection has not been completely solved [11]. Traditional lens boundary detection algorithms mainly include edge-based algorithm, histogram-based algorithm [45], and pixel difference method [12]. The edge-based lens boundary detection algorithm is not ideal for the complex frame image content of film and television lens boundary detection. The histogram-based lens detection algorithm, from the statistical point of view, statistics the pixel color distribution of the frame image or the gray distribution of the image, can better adapt to the low-speed motion of the camera equipment and the object in the lens. In addition, this method also has low computational complexity. The disadvantage is that when the light intensity changes and the lens moves rapidly, the histogram obtained will be distorted, which will lead to error detection. The pixel difference method has a low computational complexity, but it is very sensitive to pixel brightness changes and light changes caused by the motion of the camera equipment and objects in the film and television, which are likely to cause lens error detection [13, 14]. In recent years, texture features [15] and scale-invariant feature conversion features [16] are also often found in literatures related to lens edge detection. The texture feature is the global feature of image. In the detection of the lens edge, texture and other features form feature combination to realize the detection of the lens edge. The histogram of the gradient direction and color histogram of the texture are proposed to realize the detection of the lens edge [17, 18]. The color feature and texture feature are extracted by wavelet transform, and the difference of adjacent frames is defined according to the mutual information of the color feature and the mutual information of the texture feature, and the lens edge is further determined according to the dynamic threshold [19]. SIFT features are the local features of the image, which have the advantages of image size scaling, rotation, and brightness changes remaining unchanged. SIFT features can effectively reflect the local changes of moving objects. For the frame image of the same lens, the SIFT features of the same lens frame have a high degree of matching, and SIFT features are adopted to achieve lens detection [20], which can better distinguish the moving objects in different gradient lenses, so as to find the lens

boundary within the permissible error range. SIFT features are used to match adjacent frames. However, the use of SIFT features alone to realize the detection of lens edges is greatly affected by the rapid movement of the lens and the change of ambient light intensity [21, 22]. In addition, the SIFT feature was used only to detect the stacking lens, and the detection effect was not ideal. As an important feature of an image, the texture can better reflect the change of the underlying feature of the image, and the similarity between images can also be measured by the texture matching degree of the image. The texture concept of a static image can be extended to the temporal domain to form a dynamic texture. Dynamic textures are defined differently in different applications [23, 24]. The dynamic texture is described as an image sequence with the stable property of the moving scene on time sequence, such as the wave film lens and smoke film lens. [25, 26]. For two or more frames of images adjacent to the same lens, if the object in the frame image does local motion, the frame is divided into homogeneous blocks, and the average gradient of each block constitutes the average gradient matrix. There is a strong correlation between its gradient matrix, which is similar to the "dynamic texture" [27, 28]. Through the "movie dynamic texture," the uniform movie frame blocks are processed, and the average area of each gradient is calculated, which is composed of the average gradient matrix. According to the average gradient matrix of the "movie" dynamic texture and the dynamic change of the "texture" between adjacent frames [29, 30], it is determined whether the content of adjacent frames has changed dramatically. When the ambient light does not change dramatically [31, 32], the dynamic texture of adjacent frames in the same lens will not change fundamentally.

In the aspect of lens boundary detection, the optical flow method is used to extract the motion information in film and television, and a vector quantization method of the optical flow is proposed. According to the quantization method, the running amount of adjacent images is corrected and the frame difference is obtained. The frame difference is used to detect the edge of the candidate shot. A model matching method is proposed to detect the lens of mutation and gradient, so as to obtain the lens boundary information of film and television. In order to avoid the difficulty of threshold setting in lens detection by the threshold method and further improve the detection effect, a unified lens boundary detection strategy based on fuzzy clustering was realized in this paper on the basis of analyzing the characteristics of lens movement and combining the knowledge of fuzzy clustering. This algorithm can detect the abrupt/gradual lens at the same time without setting the min value. It can effectively reduce the impact of flash, subtitle insertion, advertising, and other factors on the lens detection and reduce the impact of the lens movement on the boundary detection of the film and television, thus further enhancing the robustness of the lens detection. The detection effect of the algorithm is verified by experiments.

2. Dynamic Texture Boundary Detection Based on SIFT Feature

Dynamic texture matching is to find out the similarity of the dynamic texture between two movie frames. The overall framework for the comparison of film and television dynamic textures is given first, as shown in Figure 1. Two frames of images for a movie clip. It is divided into $M * N$ subregions, and the average gradient of each region is calculated to form the average gradient matrix.

The similarity of the frame image space, the global change of the image space, and the local change of the space can be detected and judged by many methods, and the frame dynamic texture matching method is a better way. It can be seen from the definition of film and television dynamic texture that the film and television dynamic texture is defined by the average gradient of subregions. Two adjacent frames in the same area of the average gradient change reflects the region gray degree; in other words, the son of regional average gradient change reflects the local change of the frame because the whole image is implemented by the average gradient matrix; the change of the average gradient matrix is also reflected by the whole picture of adjacent frames, and the change of the frame for a given dynamic texture can be achieved by comparison:

$$\begin{aligned} |\text{Gra}_{i,k}^{j+1} - \text{Gra}_{i,k}^j| < \varepsilon, \\ \Delta\eta_{i,k+1} = |\eta_{i+1} - \eta_i|. \end{aligned} \quad (1)$$

SIFT features are the local features of the image, which has the characteristics of image size scaling, rotation, and brightness changes remaining unchanged. The SIFT feature is a feature commonly used in image object matching, and the SIFT feature of adjacent frames of the same lens has a high degree of matching. For shear lenses, if $\Delta\eta_{i,k+1}$ is dense and the SIFT feature of adjacent frames has a low matching degree, then adjacent frames are considered to belong to different lenses. For the gradual change lens with the complex structure (including flexible solution, light fading, and overlapping), if the $\Delta\eta_{i,k+1}$ generated by the frame is a sparse matrix, the SIFT feature of adjacent frames has a high degree of matching, and it cannot be completely determined that adjacent frames belong to different lenses, so continue to compare the frames. Here, $r < 24$, that is, the value is less than the limit resolution frame number of human eyes, and the lens boundary detection error is allowed to be less than the limit resolution frame number of human eyes. There are slight differences between human eyes in the limit resolution frame number of different types of films and television programs. In this paper, $r = 20$ is adopted:

$$T(t_i, t_{i+1}) = \frac{M_n}{\text{Min}(M_i, M_{i+1})}. \quad (2)$$

Since the detected extreme points are unstable, further processing of these extreme points is required, that is, to remove pixels with dog curvature asymmetry. SIFT employs three-dimensional quadratic function fitting, precise scale, and location of extremum points to improve antinoise

capability and enhance matching stability. The extreme points with low contrast are removed:

$$S(X) = S + \frac{\partial S^T}{2\partial X} \hat{X}. \quad (3)$$

SIFT features can reasonably describe the image, in the case of rotation, zoom, and translation, are not affected. Therefore, SIFT features are used to characterize film and television images. In the comparison of film and television pictures, it can be ensured that the diversity of the pictures is not caused by the rotation, contraction, and translation of a single picture, but by the diversity brought by the real difference of the pictures.

Lens boundary detection is mainly based on the similarity between adjacent frames inside the lens. When the lens is converted, the similarity is destroyed and the difference between frames is generally large. Therefore, the basic idea of lens detection is to compare the difference between film and television frames and to compare the difference between the frames and the min value. If the difference reaches a certain degree, it is judged that the lens conversion has occurred. When the shot is suddenly changed, the distance between the frames is often shown as a convex crest. In the process of lens gradient, there is a little bump in the whole waveform, but the difference is not as obvious as the shear.

After the lens boundary detection, the test indexes can be used to objectively evaluate the detection results, measure various lens detection algorithms, and assist to select the correct algorithm. In lens detection, recall and precision are the two most basic and commonly used evaluation parameters. According to different processing methods, the lens boundary can be divided into two types: the lens mutation boundary and the lens gradient boundary, as shown in Figure 2. On the basis of the lens boundary detection, the next step is to analyze the film and television lens by dividing it into two. Therefore, the segmentation accuracy of the video shot after the shot is particularly important for the extraction of key frames. The processing and data analysis of the key frame extraction related to the lens boundary detection and the back of the film affects the accuracy.

It is the main principle of the key frame extraction algorithm based on the image content analysis to measure the image similarity according to the change of some underlying features of the image, such as the image color, shape, texture, and other visual features. The specific steps are as follows:

- (1) Choose the first shot of the image as a key frame, and look at it as a comparison frame.
- (2) Calculate the degree of difference between the image frame and the comparison frame in the film and television successively. When a certain frame is found to have a great change, that is, the difference value between this and the comparison frame is greater than the preset threshold T , then the image frame is regarded as the key frame and regarded as the new comparison frame.

Continue the comparison between the subsequent image frames and the new comparison frames, and repeat the first

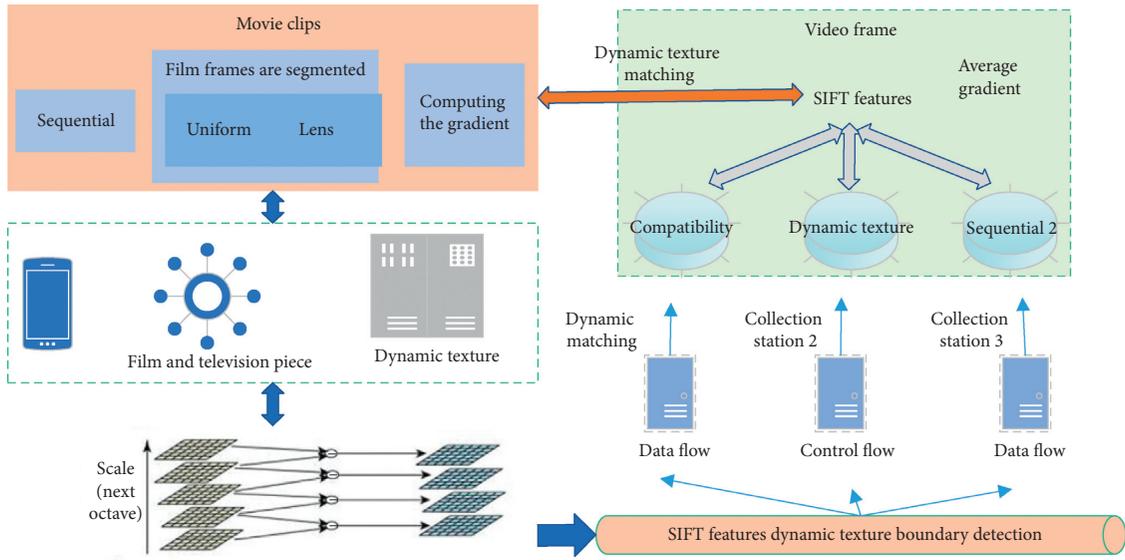


FIGURE 1: Dynamic texture comparison block diagram.

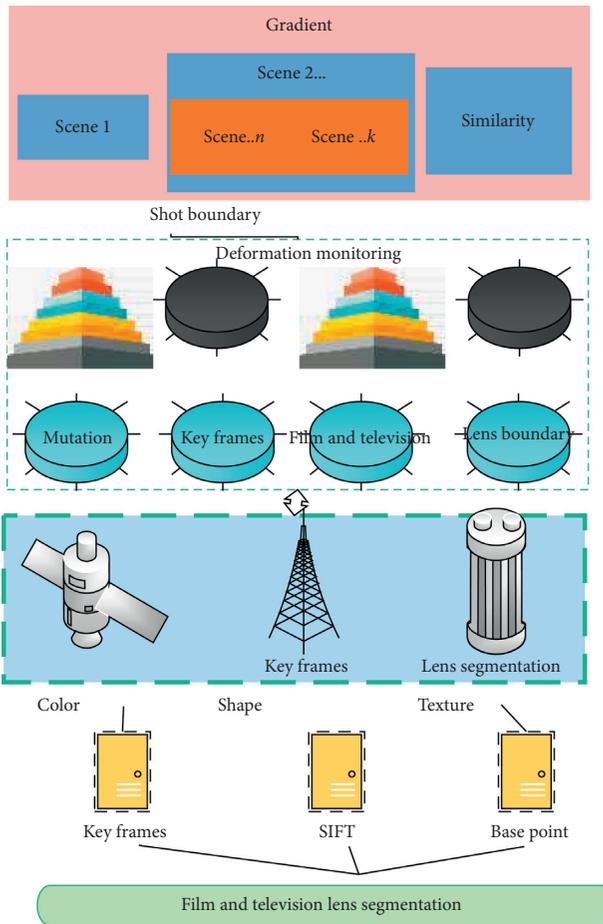


FIGURE 2: Schematic diagram of lens boundary.

two steps, until the detection of all the image frames in the film and television, and all the selected key frames are regarded as the final key frame set of the current film and television.

Figure 3 shows the outline diagram of the film and television editing algorithm based on the combination of SIFT features. The following three aspects will be introduced in detail: subjective features based on audience comments, objective features based on visual effects, and film and television editing generation. The following symbols are defined for expression. C represents a total set of highlights, I represents a highlight, and S represents a subset of a highlight set composed of multiple highlights, which is referred to as a fragment set. The goal of this chapter is to find the optimal set of fragments.

There are also a large number of flash cases in the film and television, and the general flash detection can only exclude the biggest change in the flash sequence of the frame on the impact of the lens detection, and it does not take into account the impact of flash sequence on the lens detection. In the analysis of the false detection, it is found that the mean value of the difference between frames tends to zero when there is a large number of small changes between frames.

3. Key Frame Extraction of Camera Lens Boundary Detection Based on SIFT Feature Fusion

The frames in the film are initially mapped to the points in the 2D image, and then, the image information is obtained by point cluster processing. The clustering algorithm based on the density value in the cluster classification process is similar to the k -Medoids algorithm, which only needs to analyze the distance between point pairs and density value. According to the distribution locations of point pairs and the density values of adjacent points, the whole point cluster is clustered to obtain various centers, and the key frames of film and television are selected by the clustering center to constitute the final abstract.

The clustering algorithm based on density analysis is judged and identified according to the distribution of

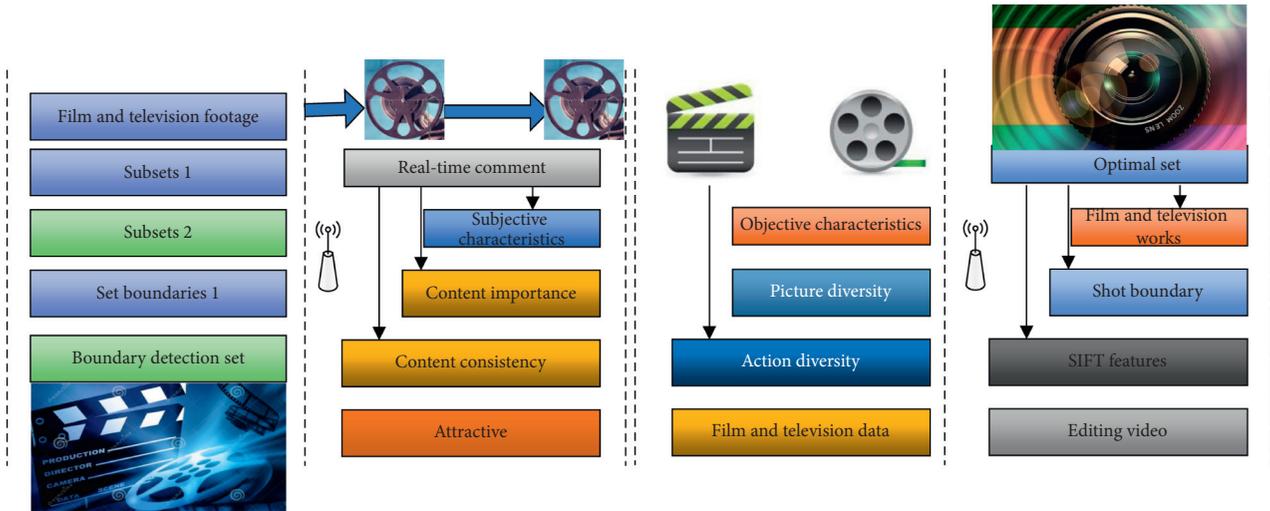


FIGURE 3: Outline of the film and television editing algorithm based on SIFT features.

experimental points and the corresponding relation. Due to the differences in the brightness, color, and other characteristics covered by each frame of film and television, it is possible to map each frame to the corresponding space based on these differences, and each frame is corresponding to the points in different coordinates. The classification method of the point cluster is adopted to judge the corresponding relation between images in the film and television. The division of the obtained point cluster is based on the distance between the density value of each point and the other corresponding points, instead of using specific coordinates in the two-dimensional space. In this section, the similarity between images is used to measure the corresponding distance. The smaller the similarity is, the greater the corresponding distance value will be. The corresponding relationship is as follows:

$$c_{ij} = 1 - t_{ij}. \quad (4)$$

SIFT was adopted to obtain the distance between various film and television frames, and then, each frame of the film and television was mapped to the relevant points in the two-dimensional space. Secondly, the frames were divided into clusters through clustering operation. The decision diagram is used to indicate that the selected cluster center points correspond to high values, indicating that the selected point is the center of each class, that is, each center point represents the characteristics of the corresponding group. The center of all categories corresponds to the selected frame to express the information of this class. Therefore, this image is selected as the key frame to describe the main information of the film and television. The process of selecting key frames using the SIFT image frame-mapping method is as follows:

Input: image sequence marking in the source video

Output: get the set of key frames corresponding to the source film and television

Step 1. Calculate the statistics and texture characteristics for each image

Step 2. Obtain the interval between any point pair, and the SIFT form is used to map the film frame to the points in the two-dimensional space

Step 3. Collect the region P value and corresponding value of each point

Step 4. Follow the local density value P ; the required decision graph is drawn with the function relation of distance, and the number of point groups contained in the decision graph is determined in an interactive form

Step 5. Define the image subsets contained in various clusters according to the relationship between film and television frames in spatial point mapping

In the process of expressing main information through film and television abstract, it is more important to obtain the number of key frames. If too many frames are used, the redundancy between frames will be caused. If the quantity selected is insufficient, it will affect the expression of complete film and television information. The selection of the number of frames is another objective evaluation element in the film summary. In the density peak clustering algorithm, the number of clusters is obtained from the decision graph in the form of human-computer interaction. In the case of clustering in the film and television abstract proposed in this section, the influence between them will be intensified when the length of the film and television is long and the number of frames is large. If the interactive form is adopted again, the number of categories cannot be determined automatically and quickly.

By selecting the path, it is obvious that the closer to the reference point, the more pixels are used for comparison. Choosing the sampling path in this way can not only express the illumination characteristics of the whole image but also reflect the main content of the image. Figure 4 shows the change curves of brightness values of mutant and fading lenses. It can be observed from Figure 4 that when fading into the lens like lens, the image brightness value curve has an obvious process of gradually increasing. However, for the

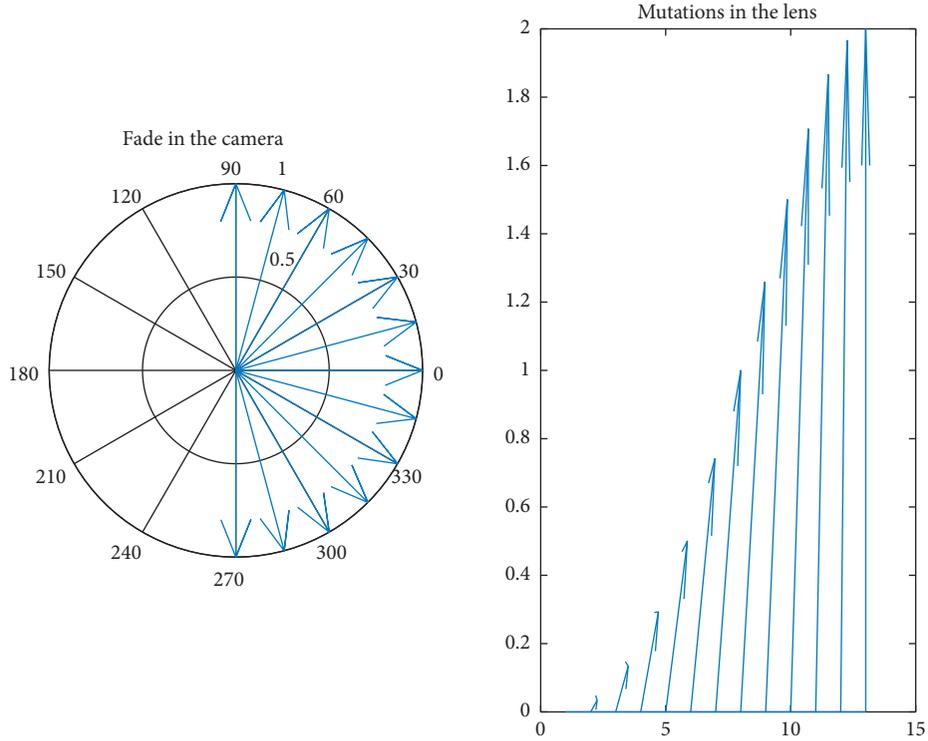


FIGURE 4: Brightness curves of different lenses.

mutant lens in Figure 4, the image brightness value curve has a sudden change process, with large difference before and after, which is conducive to rapid detection.

When significance values are measured at multiple scales, mean values are used to enhance the contrast between significant and nonsignificant areas. It is considered that the significance threshold is acquired from the local significance subregion in the significance graph, and the pixel significance value not located within the scope of the subregion is obtained by the Euclidean distance weighting between the adjacent significant pixels, so as to obtain the new significance value. In this way, the significance value of the vicinity of the significant target is increased and the significance value of the background part is weakened:

$$S_i = s_i(1 - d_{ij}) + \frac{1}{N} \sum s_i. \quad (5)$$

The acquired significant regions are of great significance for image analysis. We strengthened the analysis of the important content of the image, ignoring some minor parts, which became the key to improve the efficiency and optimize the effect. After obtaining the significant area of the image, the processing of the significant area can increase the difference between different lenses. Mutual information mainly represents the information correlation between the two systems, that is, a system covers the size of the information in the corresponding system. Image mutual information is a measure of how much information each image contains, and switching at the gradient lens is the process of merging the contents of front and rear lenses. Therefore, we use mutual

information to measure the similarity between images. Mutual information is defined as follows:

$$I(C, D) = \sum_i \sum_j \log q(c_j, d_i) q(c_i, d_j). \quad (6)$$

4. Example Verification

In order to verify the edge detection effect of the film and television lens proposed in this paper, two groups of films and television films were selected for algorithm effect verification. The first group tested 150 film and television clips from the Internet, which included shear lenses and a variety of gradient lenses with complex structures (light fading, dissolving, and overlapping). Film and television types include film clips, sports films, and news films. In the second group, the most authoritative international evaluation TRECVID2003 film and television collection was taken as the test film and television, and 6 classic segments were selected, each segment containing the shear lens and gradient lens. The film and television clips include the black and white lens and color lens, with specific parameters shown in Table 1.

In order to measure the detection effect, recoil, accuracy, precision, and F1, a comprehensive evaluation index combining the recall rate and accuracy was used to represent the lens recall rate. Among them, the higher the value of F1, the better the detection effect.

In this algorithm, the choice of parameters has a great influence on the experimental results. The value of δ reflects

TABLE 1: Experimental data set parameters.

Movie clips	Frames	Lens number	Lens type		Film and television data subset
			Shear	Gradient	
C1	11.345	68	39	25	N25
C2	16.587	75	43	30	MN26
C3	12.314	105	36	62	N26
C4	31.398	173	94	54	N26
C5	12.532	87	42	25	Air
C6	13.657	98	32	44	Global

the degree of gradient change in subregions, the value of δ_1 reflects the degree of sparsity of the gradient change matrix, and the value of δ_2 reflects the matching degree of SIFT features of adjacent frames.

4.1. Influence of δ Value on Experimental Results.

Without considering the contribution of frame SIFT features to lens edge detection, different δ values have different effects on detection results. Figure 5 shows the trend of change of recall rate, accuracy, and precision with TT when $\delta = 0.75$. It can be seen from Figure 5 that good recall rate and accuracy can be achieved when $\delta = 0.4$.

Without considering the contribution of frame SIFT features to lens edge detection, different ε values have different effects on detection results. Figure 6 shows the trend of recoil and accuracy precision ε changes of recall rate under $\delta = 0.75$.

In this paper, SIFT features and dynamic texture have different effects on lens detection. To verify SIFT features and the effect of the dynamic texture on lens detection, the algorithm in this paper is compared with the SIFT feature and dynamic texture for lens detection. The first group was used to test the three algorithms. In the experiment, the subregion size of the frame image was taken as 13×13 , and $\delta = 0.3$ and $\varepsilon = 0.6$ as an empirical decimal where the higher the AVF value is, the better the detection effect will be. It can be seen from Table 2 that the detection results of the algorithm in this paper are better than that of the SIFT feature or dynamic texture only.

In the experiment, the film and television were firstly segmented manually, and the segmentation results were taken as the reference lens boundary. There are some differences in the criterion of edge judgment for different types of lenses. For the gradient lens, the position of the lens edge is difficult to define accurately, and the error is allowed within 20 frames. The experimental results of the algorithm in this paper are shown in Table 3.

After verifying the relationship between the ratio and the accuracy of shear detection, the experiment was conducted again to verify the influence of the partitioning ratio on the fitting feature error rate at the gradient. Figure 7 shows the influence of the ratio on the fitting error rate for the extraction of the gradient region.

It can also be seen from Figure 7 that the error of gradual fitting features reaches a minimum when the block ratio is close to 0.6. The smaller the error rate of the fitting features

is, the closer the arch waveform generated by the feature description of the film and television is to the standard waveform shape. Such an approach is conducive to gradual detection and will improve its accuracy.

To compare robustness, the experimental data were used for several different types of video clips downloaded from the network in the MPG format, including advertisements, video clips, and MTV. A total of 7726 frames were used, including 109 lens boundary conversions, among which 85 were mutant shots and 23 were gradient shots. The algorithm presented in this paper is used to compare and experiment these movie sequences with the histogram-based double-min method and the pixel-based double-min method. The specific experimental results are shown in Table 4. Algorithm 1 is the algorithm presented in this paper, algorithm 2 is the histogram-based double threshold method, and algorithm 3 is the pixel-based double threshold method.

As can be seen from Table 4 above, for different types of film and television clips, algorithm 1 can achieve good detection results in overall recall rate and precision rate of more than 72%, while algorithm 2 and algorithm 3 have good and bad detection results for different films and television, which confirms the general adaptability of algorithm 1 to films and television. In addition, from the point of view of different types of testing, film, and television, the algorithm has a good test effect on video cameras of advertising and movie editing types. This is because the algorithm has made some improvements in robustness, so when commercials and movie clip-type video cameras are in more complicated situations such as flashing, camera movement, subtitle changes, and noise interference, better test results can still be obtained.

It can be clearly seen from Figure 8 that algorithm 1 has a high recall rate and precision rate, especially a high recall rate, because this algorithm reduces the influence of subtitle insertion and icon insertion on lens detection in the process of calculating and selecting clustering features. On the basis of clustering, preprocessing was carried out according to the characteristics of abrupt and gradual transition boundaries of film and television, and the results were further analyzed. In the process of lens detection, flash, lens movement, and noise detection were added to reduce their impact on lens detection, and finally, the lens boundary detection was realized. Therefore, it has better robustness to common interference situations and can obtain better detection effects for the same film and television clips.

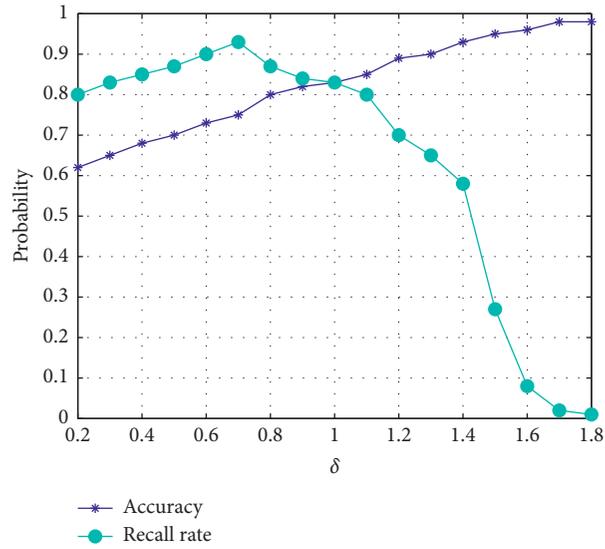


FIGURE 5: Influence on recall rate and accuracy of predicted results when $\delta = 0.75$.

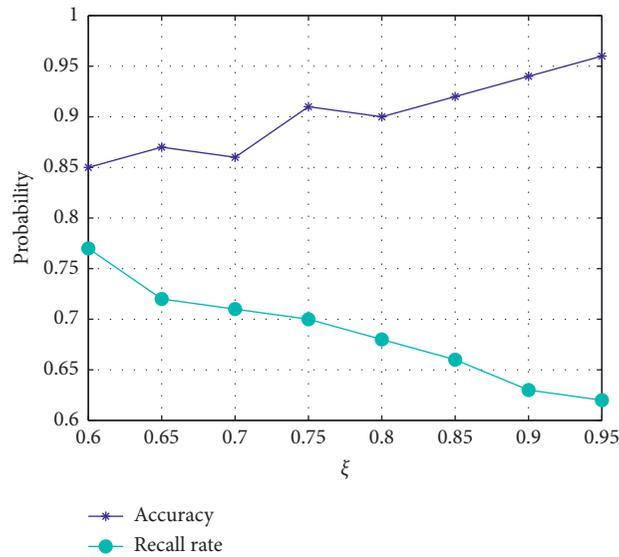


FIGURE 6: Influence of $\delta = 0.3$ and ϵ on recall rate and accuracy of predicted results.

TABLE 2: Results of the three methods on the test set of sports films and films.

Algorithm	Movement			Movie		
	R	P	AVF	R	P	AVF
SIFT	0.72	0.889	0.765	0.723	0.876	0.784
DT	0.754	0.893	0.812	0.751	0.867	0.82
SIFT + DT	0.832	0.891	0.853	0.772	0.845	0.823

TABLE 3: The results of experiment ($\delta = 0.3$ and $\varepsilon = 0.7$).

Film and television sets	Shear			Gradient		
	R	P	F1	R	P	F1
C1	0.945	0.942	0.942	0.921	0.863	0.873
C2	0.972	0.876	0.935	0.904	0.81	0.856
C3	0.892	0.95	0.926	0.952	0.89	0.963
C4	0.932	0.896	0.917	0.946	0.871	0.921
C5	0.95	0.934	0.948	0.817	0.823	0.817
C6	0.922	0.906	0.912	0.879	0.923	0.889

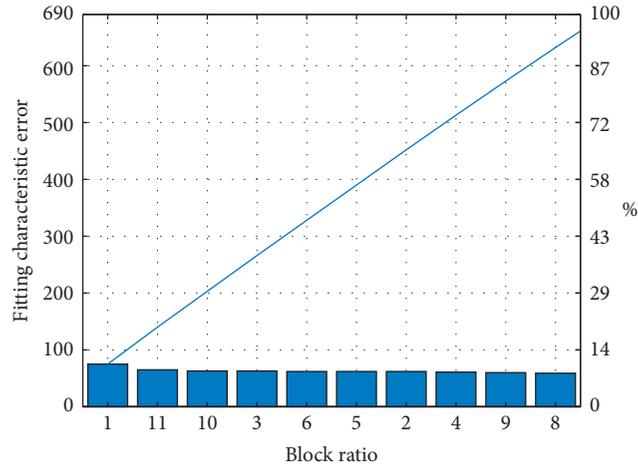


FIGURE 7: Block ratio and fitting characteristic error.

TABLE 4: Comparison of detection results of different types of film and television clips.

		Advertising	Video clip editing	MTV
		Algorithm 1	Precision (%)	99
	Recall (%)	84.5	84.5	98
Algorithm 2	Precision (%)	57.8	68.5	81.5
	Recall (%)	98	75.8	92.7
Algorithm 3	Precision (%)	57.7	85.8	55.8
	Recall (%)	57.4	82.4	71.4

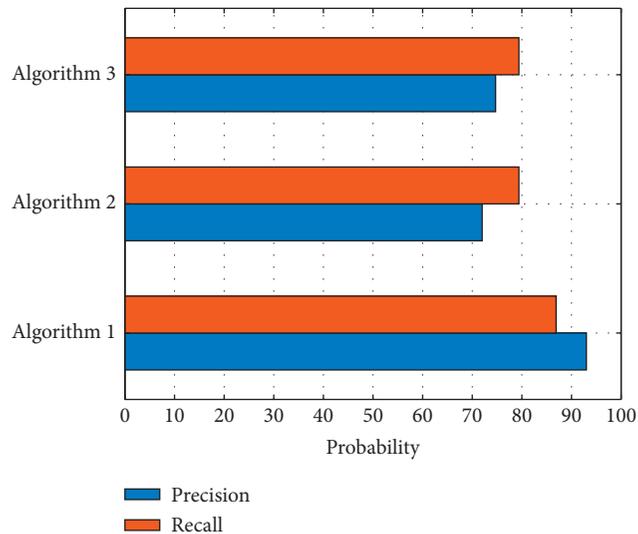


FIGURE 8: Comparison of recall and precision of the three algorithms.

5. Conclusion

A new method of lens edge detection based on dynamic texture and SIFT features is proposed for many kinds of film and television data. The algorithm mainly includes four aspects: movie and television dynamic texture construction, movie and television dynamic texture matching, frame image SIFT feature matching, and false detection processing. The dynamic texture of film and television takes into account the local and global changes of the frame image. The method presented in this paper is effective for edge detection of both shear and gradient lenses, especially for edge detection of laminated lenses, and also for reducing the influence of light on lens detection. However, the edge detection effect of the method in this paper on fluid objects (such as seawater in the film and television) needs to be improved, and the adaptive selection method in this paper needs to be further improved. In the detection process, in order to effectively reduce the influence of subtitle and other factors on lens detection, an improved histogram segmentation method is constructed, and the frame difference of each block is taken as the feature, and the weight of each block is taken as the feature weight for fuzzy clustering. On the basis of clustering, preprocessing is carried out on the sudden change and gradient boundary features of movies, and the results are further analyzed to exclude the influence of flash and lens motion lens detection and finally realize the lens boundary detection. However, due to the complexity of film and television, there are still some missing and false detections in this algorithm, which need further study. Film and television is the carrier of a variety of information, including images, text, and sound. Most of the current detection algorithms are limited to using the image features to detect the boundary through the change of the image content. In the future, algorithms can make use of more information, such as superimposed text and audio information.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] Z. El Khattabi, Y. Tabii, and A. Benkaddour, "Video shot boundary detection in sport video using the scale invariant feature transform," *Applied Mechanics and Materials*, vol. 850, pp. 152–158, 2016.
- [2] Z. E. Khattabi, Y. Tabii, and A. Benkaddour, "Video shot boundary detection using the scale invariant feature transform and RGB color channels," *International Journal of Electrical & Computer Engineering*, vol. 7, no. 5, pp. 2565–2578, 2017.
- [3] J. Fan, S. Zhou, and M. A. Siddique, "Fuzzy color distribution chart -based shot boundary detection," *Multimedia Tools and Applications*, vol. 76, no. 7, pp. 10169–10190, 2017.
- [4] T. Alfatawii and I. Hadi, "Video's cut transitions detection based on multiple features," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 3, pp. 1208–1216, 2019.
- [5] T. Steiner, R. Verborgh, J. Gabarro, E. Mannens, and R. Van de Walle, "Clustering media items stemming from multiple social networks," *The Computer Journal*, vol. 58, no. 9, pp. 1861–1875, 2015.
- [6] Y. Wang, J. Yue, Y. Dong, and Z. Hu, "Review on kernel based target tracking for autonomous driving," *Journal of Information Processing*, vol. 24, no. 1, pp. 49–63, 2016.
- [7] G. Ding, J. Wang, and Q. Wu, "Cellular-base-station assisted device-to-device communications in TV white space," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 107–121, 2015.
- [8] L. I. Zong-Yan, S. Li-Mei, and X. Jiang-Tao, "A stereo matching algorithm based on SIFT feature and homography matrix," *Optoelectronics Letters*, vol. 11, no. 5, pp. 390–394, 2015.
- [9] D. Haifeng and Y. Jun, "Research on robot binocular ranging based on SIFT feature extraction algorithm," *Journal of Physics: Conference Series*, vol. 1607, no. 1, pp. 012015–012021, 2020.
- [10] R. Hannane, A. Elboushaki, K. Afdel, P. Naghabhushan, and M. Javed, "An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram," *International Journal of Multimedia Information Retrieval*, vol. 5, no. 2, pp. 89–104, 2016.
- [11] I. P. Febin, K. Jayasree, and P. T. Joy, "Violence detection in videos for an intelligent surveillance system using MoBSIFT and movement filtering algorithm," *Pattern Analysis and Applications*, vol. 23, no. 3, pp. 89–98, 2019.
- [12] A. Oad, H. D. Jun, M. A. Bhutto, A. Oad, X. W. Cui, and W. Lei, "Object detection algorithm based on Gaussian mixture model and sift keypoint match," *Journal of Computational and Theoretical Nanoscience*, vol. 15, no. 6, pp. 1940–1947, 2018.
- [13] D. Jiang and J. Kim, "Video searching and fingerprint detection by using the image query and placenet-based shot boundary detection method," *Applied Sciences*, vol. 8, no. 10, pp. 1735–1751, 2018.
- [14] W. Cui, G. Wang, C. Feng, Y. Zheng, J. Li, and Y. Zhang, "Spmk and grabcut based target extraction from high resolution remote sensing images," *ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 7, pp. 195–203, 2016.
- [15] A. Yusefi, A. Durdu, and C. Sungur, "Performance and trade-off evaluation of SIFT, SURF, FAST, STAR and ORB feature detection algorithms in visual odometry," *European Journal of Science and Technology*, pp. 455–460, 2020.
- [16] H. Chang, G. Wu, and M. Chiang, "Remote sensing image registration based on modified SIFT and feature slope grouping," *Geoscience and Remote Sensing Letters, IEEE*, vol. 16, no. 9, pp. 1363–1367, 2019.
- [17] Y. Wang, Y. Yuan, and Z. Lei, "Fast SIFT feature matching algorithm based on geometric transformation," *IEEE Access*, vol. 8, pp. 88133–88140, 2020.
- [18] A. Wang, Y. Zhao, J. Zhao, Y. Iwahori, and X. Wang, "A novel feature detection algorithm based on improved 2DPCA-SIFT," *International Journal of U- and E-Service, Science and Technology*, vol. 8, no. 6, pp. 299–304, 2015.
- [19] Y. Wang, Y. Yuan, and Z. Lei, "Fast SIFT feature matching algorithm based on geometric transformation (January 2020)," *IEEE Access*, vol. 52, no. 9, pp. 278–298, 2020.

- [20] P. Wang, X. Li, and Z. Lv, "Vehicle outline detection based on improved SIFT algorithm," *International Journal of Earth Sciences and Engineering*, vol. 8, no. 1, pp. 396–400, 2015.
- [21] E. Wang, J. Jiao, J. Yang, D. Liang, and J. Tian, "Tri-SIFT: a triangulation-based detection and matching algorithm for fish-eye images," *Information*, vol. 9, no. 12, pp. 299–312, 2018.
- [22] Y. Lyu, X. Shen, and H. Chen, "Copy-paste detection based on a SIFT marked graph feature vector," *Chinese Journal of Electronics*, vol. 26, no. 2, pp. 345–350, 2017.
- [23] J.-H. Yeong, "Multi-camera based images through feature points algorithm for HDR panorama," *International Journal of Advanced Smart Convergence*, vol. 4, no. 2, pp. 6–13, 2015.
- [24] C. Loukas and E. Georgiou, "Performance comparison of various feature detector-descriptors and temporal models for video-based assessment of laparoscopic skills," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 12, no. 3, pp. 387–398, 2016.
- [25] S. A. Bakar, X. Jiang, X. Gui, G. Li, and Z. Li, "Image stitching for chest digital radiography using the SIFT and SURF feature extraction by RANSAC algorithm," *Journal of Physics: Conference Series*, vol. 1624, pp. 042023–042032, 2020.
- [26] Y. Cheng, "Image mosaic technology based on an optimized method of SIFT algorithm," *Journal of Image and Signal Processing*, vol. 4, no. 4, pp. 139–145, 2015.
- [27] W. JunBing, "Copy-move forgeries detection based on SIFT algorithm," *Computer Science and Application*, vol. 5, no. 7, pp. 255–263, 2015.
- [28] Z. Zhou, Q. M. J. Wu, S. Wan, W. Sun, and X. Sun, "Integrating SIFT and CNN feature matching for partial-duplicate image detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 5, pp. 593–604, 2020.
- [29] F. P. An and X. W. Zhou, "BEMD-SIFT feature extraction algorithm for image processing application," *Multimedia Tools and Applications*, vol. 76, no. 11, pp. 31–50, 2017.
- [30] N. Najva and K. E. Bijoy, "SIFT and tensor based object detection and classification in videos using deep neural networks," *Procedia Computer Science*, vol. 93, pp. 351–358, 2016.
- [31] S. Chen, X. Li, S. Chi, Z. Li, and M. Yuxing, "Ship target discrimination in SAR images based on BOW model with multiple features and spatial pyramid matching," *IEEE Access*, vol. 8, pp. 166071–166082, 2020.
- [32] J. D. Gavade and S. Chougule, "A SIFT with RANSAC based spatial tampering detection in digital video," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 3, pp. 1156–1163, 2019.