Research Article

Research on Postproduction of Film and Television Based on Computer Multimedia Technology

Xiao Xu (ID),[1] **Hao Yan** (ID),[2] and **Xiaolei Wang** (ID)[2]

[1]*Yangtze University College of Arts and Sciences, Jingzhou 434020, China*
[2]*Wuhan University of Engineering Science, WuHan 430200, China*

Correspondence should be addressed to Hao Yan; yanhaowh@163.com

In the postproduction of film and television, the fast movement of objects or camera shake will lead to an excessive selection of key frames, resulting in information redundancy and affecting the stability of postproduction videos and video-editing effects. Aiming at this problem, this paper proposes a method of film and television postproduction based on digital media technology. Firstly, the film and video shots are segmented to find out the changing scenes; secondly, the bottom feature vector of each image is formed according to the color space entropy, motion vector, and motion region, and the video key frame is extracted based on digital media technology; finally, the pixel label information is established in the key frame, and the feature points located in the foreground are removed, while the time feature matching points and spatial feature matching points in the background are left, so as to achieve dynamic video mosaic. The experimental results show that the film and television postproduction method based on digital media technology has strong advantages compared with other methods, which can improve the stability score of video and reduce the average stitching error.

1. Introduction

With the rapid development of digital media technology and computer graphics generation technology, film and television production has entered the digital era, and the whole process from image production to projection has undergone Earth-shaking changes. Since its inception, film and television production has taken a typical scientific and technological attribute. Each evolution and development of science and technology and media will promote the corresponding changes in the creation, storage, distribution, projection, and viewing methods of film and television and then promote the internal changes of film and television. The digital film can be shot with a digital camera to obtain high-definition and high-fidelity audio and video files or carry out digital modeling and digital adjustment through virtual camera, image and graphics processing software, graphics and video processing software. It realizes the correlation between real images and models and the

integration of dynamic image synthesis, seamless image editing, and other process links to obtain virtual images. Film and television postproduction based on digital media technology can be regarded as an art form relying on new media technology. It includes not only the organization and control of the formal language of traditional film and television postproduction and traditional media but also the film and television process of digital media technology, such as creation, design, application, and communication. Film and television post-production has become a comprehensive art category integrating art, technology, and humanities [1]. This is also the inevitable trend of the sustainable development and evolution of the information society in the Internet era. The profound impact of the development of digital media technology on film and television aesthetics has not only changed the traditional way of film and television production combining digital virtual technology with traditional realism, so as to push film and television from the original "juggling" scene to a

TABLE 1: Lens conversion types.

Type	Name	Characteristic
Mutation	Interrupt	The simplest form of shot conversion, the conversion between different shots is carried out on a single frame
	Fade in	Several frames at the beginning of the lens, and the image frames appear from the full black screen at a uniform speed
	Fade out	For several frames at the end of the lens, the image frames darken to full black at a uniform speed
Gradual change	Erase	The start frame of the next shot appears slowly and covers the end frame of the previous shot
	Rotate	The start frame of the next shot is rotated out of the screen while overwriting the end frame of the current shot
	Superposition	The frames at the end of the current shot slowly darken until they disappear, and the starting frame of the next shot gradually appears

more amazing spectacle for the audience. In the process of developing creative forms, the film also provides the audience with a new viewing experience and a new way of thinking. Digital movies can add digital scenes and carry out color matching treatment at will. The creation of digital images can be modified unlimited times. The picture generation and artistic effects such as composition perspective, color expression, space distance, texture, environment, and climate can be designed according to the needs of the plot to pursue audio-visual wonders that confuse the truth with the truth or stimulate shock. Digital media technology has changed the existing environment of film and television video to some extent, and its development has brought more simple creative tools and means for film and television postproduction. More and more ordinary people can carry out simple digital image creation and postproduction activities. Film and television production no longer exists as a unique media form or art type and has gradually become the unique spokesman of digital media technology in the twenty-first century. As a technical means of film, digital media technology continues to develop the interactive function of film images with the audience in the process of generation and dissemination and has fully penetrated into various processes of the film industry with high-tech means such as 3D, VR, AR, high-resolution, and high-frame rate.

Based on the comprehensive research content, it can be seen that the research of digital media technology in the postproduction of film and television is still in the development stage. Therefore, this paper divides the film and video footage, finds out the changing scenes, and forms each scene according to the color space entropy, motion vector, and motion area. The bottom feature vector of an image is extracted, and video key frames are extracted based on digital media technology, pixel label information is established in the key frames, the feature points in the foreground are removed, and the temporal feature matching points and spatial feature matching points in the background are retained to realize dynamic video stitching. Digital media technology in film and television postproduction is applied to improve the audience's viewing experience. This paper studies the application of digital media technology in film and television postproduction, in order to improve the audience's viewing experience.
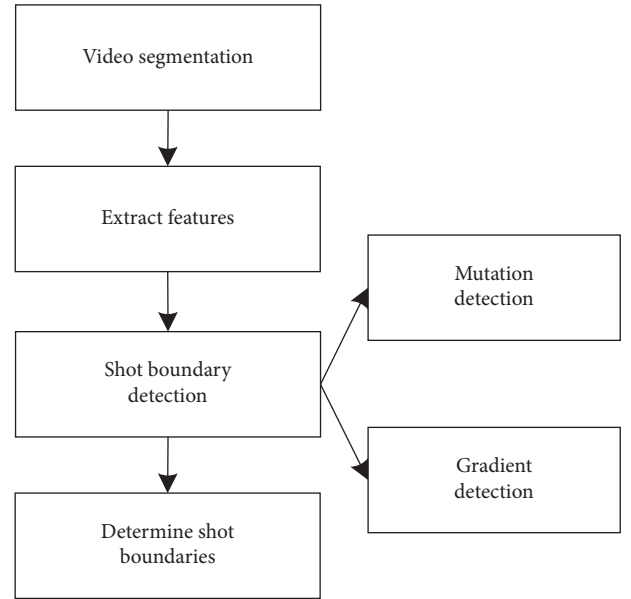

FIGURE 1: Video shot segmentation process.

2. Film and Television Postproduction Method Based on Digital Media Technology

2.1. Video Shot Segmentation. In film and television postproduction, the video shot needs to be segmented first. The boundary between two adjacent shots can be accurately found in the process of film and television postproduction. This process can be regarded as the basis of video processing, and the processed effect can pave the way for the follow-up. Before and after the video shot switching, the scene and the objects in the scene will generally change. Backgrounds and objects also change, and some are still in the same scene, but the objects will change or the position will shift [2]. Lens conversion can be roughly divided into two types, as shown in Table 1.

Abrupt lens refers to the direct splicing of two lenses at the junction without any transition. Gradual lens refers to the fusion of the last frames of the first lens and the front frames of the last lens at the junction through the processing of computer special effects, resulting in a slow transition [3]. The video shot segmentation process designed in this paper is shown in Figure 1.

Through the idea of divide and conquer, the video frame sequence is divided into several small segments (about 1-2s), and then, the first and last frames of each segment are compared to exclude the segments in the same shot and retain only the segments that may have shot switching. The feature in the video frame sequence is extracted by inception-v3, which is a 1024 dimensional vector. After extracting the feature vector, the difference between the two frames is calculated by calculating the cosine distance of the two feature vectors. The method is calculated as follows:

$$\beta(\alpha_m, \alpha_{m+1}) = \frac{1}{2} + \frac{\sum_n (p_i, q_i)}{2\sqrt{\sum_n {p_i}^2}\sqrt{\sum_n {q_i}^2}}, \qquad (1)$$

where α_m and α_{m+1} represent the feature vectors of frames m and $m + 1$; β represents the difference between α_m and α_{m+1}; (p_i, q_i) represents the pixel value of the pixel point; i and n represent pixel coordinates, sequence number, and total number. The number of transitions in the video far exceeds the number of transitions. Therefore, the current work focuses on the reliability detection of mutation in the video, but the detection of gradient mirror cannot be ignored [4]. The purpose of setting the global threshold is to avoid the situation that the current difference between frames is greater than the mutation threshold due to small changes in the two frames. After segmenting the video, the feature similarity of each head and tail frame is calculated first, and then, the similarity of each head and tail frame is compared. The threshold to find out the segments with lens changes is set, all the segments with lens change as far as possible are found out, and the segments are excluded without lens change [5]. Due to the homogeneity of frames within a shot, the difference between frames within a shot follows a normal distribution. Therefore, the threshold is selected with the help of Gaussian probability density function. The influence of flash may only last for a few frames, so the step n can be appropriately increased to calculate the similarity index between frames, so that the frames with continuous influence of flash can be skipped after $2n + 1$ frames.

2.2. Extracting Video Key Frames Based on Digital Media Technology.

In film and television postproduction, the images in the lens are very similar. After dividing different shots by shot boundary detection, the rich and comprehensive frames in each shot are selected as key frames by using digital media technology [6]. Videos taken in the same scene often have a lot of redundant information, and key frames can be used to represent the main content of the video. In film and television postproduction, it is easy to select too many key frames when there is rapid movement of objects or camera jitter [7]. The appearance of a wrong frame will cause the calculation error of the projection transformation matrix and have a great impact on the effect of film and television postproduction. In order to reduce the redundancy of the extracted key frames and maintain the smooth effect of lens edge protection, this paper extracts the key frames based on digital media technology. This can ensure the completeness of video structured information.

The process of extracting video key frames based on digital media technology is shown in Figure 2.

This method integrates the underlying features and semantic information of the target, selects the key frames based on the similarity of the target information, and can effectively extract the frames with key targets as the key frames. Each color can have a fixed analog and digital representation in the color space. HSV color space, which is closer to people's perception of color than RGB, can intuitively express the hue, brightness, and brightness of color, which is convenient for color comparison [8]. HSV color space entropy makes video color information representation, which is more in line with human visual characteristics. Motion vector is used to measure the change degree of the target and content in the video. It is often to divide the video frame into different blocks and compare the similarity between different blocks in the front and back frames. The displacement between similar blocks is called motion vector [9]. The pixels are subtracted from the square window between the reference image and the target image to search for the matching area. If the difference is zero, the two images match completely, and the central pixel of the area with the smallest difference is the matching point. The diamond search method is used to estimate the motion vector of each block, and the motion vector components in the x-axis direction and y-axis direction are obtained. The motion vector features of video frames are calculated as follows:

$$\delta(\alpha_m) = \sum_u (\|x\| + \|y\|), \qquad (2)$$

where δ represents the motion vector of the video frame; u represents the total number of video frames; x and y represent vector components. After calculating the motion vector characteristics of each frame, the interframe motion region is obtained. The current background is obtained by Gaussian mixture background modeling, and then, the region of nonbackground pixels is obtained by using the background model and the subtraction of the current frame. In order to estimate the parallax of a point in the left figure, a reference area around the point is defined first; then, within the search range of the right image, a preset matching strategy is used to find the closest area to it. According to the color space entropy, motion vector, and motion region, the bottom feature vector of each frame image is formed. The video data are synchronized and preprocessed. The first frame is the key frame by default. The video frames are shot clustered according to the underlying features of the video frames. According to the frequency of shot switching, they can be divided into videos with stable video background and videos with changeable background, and then, appropriate feature weight adjustment is carried out during feature fusion [10]. If it is a nonkey frame, the projection matrix calculated from the previous pair of key frames is directly used for interframe projection transformation [11]. Then, the adaptive threshold algorithm is used to filter the primary key frames in the fusion feature curve. The filtering process of the primary key frame is shown in Figure 3.
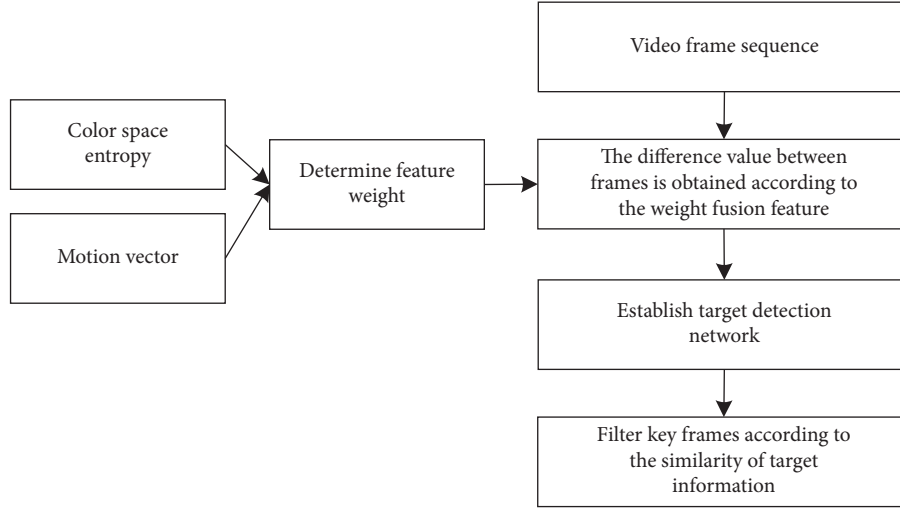
FIGURE 2: Process of extracting video key frames based on digital media technology.
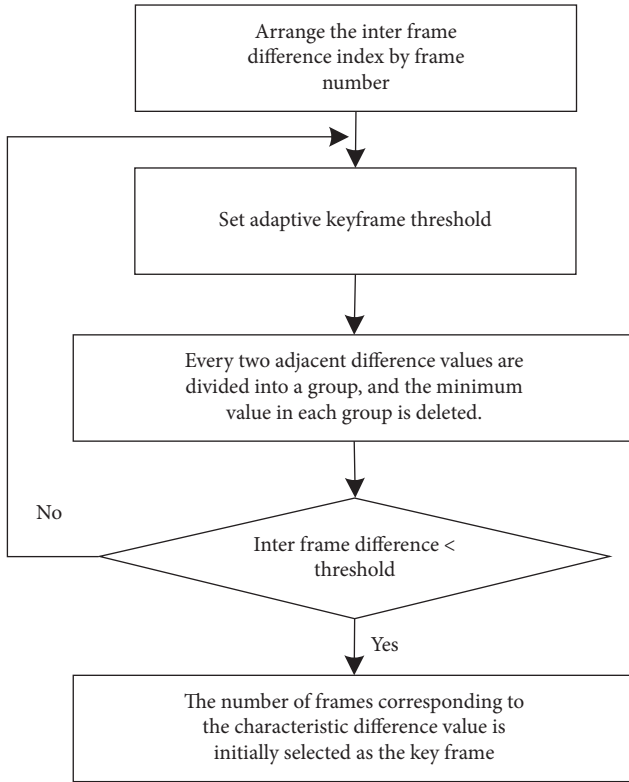

FIGURE 3: Screening process of primary key frames.

of the target category, number, and position in the primary key frame.

2.3. Establishing Film and Television Postproduction Model. Based on the extraction of video key frames, a film and television postproduction model is established. Background information extraction is of great significance for video image stabilization and postediting. This is because using the time feature matching points at the background can more accurately estimate the camera path, so as to improve the effect of video image stabilization. The key point of video splicing to ensure real-time performance is how to deal with the correlation between frames. Finally, when handling the relationship between frames, we should also ensure the splicing quality of video images and ensure that there are no obvious ghosting and suture traces in the spliced video, which can not affect the video appearance [12]. At the same time, the transformation model is calculated by using the spatial matching points at the background, which can realize the seamless stitching of the image background area. Then, a suture line in the background area is found, and the images of the corresponding frames of the video together with the suture line are clipped, which reduces the image ghosting phenomenon caused by the inability to make foreground and background objects at the same time in the presence of parallax. The image registration technology has realized the matching and registration of the pixels in the overlapping area. Then, the image interpolation technology aligns the two images in the same coordinate system. Finally, the stitched image can be obtained by using the fusion algorithm through a certain correspondence. The mathematical representation of this correspondence is the homography transformation matrix [13]. In the process of dynamic video mosaic, firstly, the foreground and background of key frame video image are segmented; secondly, the feature points located in the foreground are removed by using the image pixel label information; finally, the temporal feature matching points and spatial feature matching points are left

The adaptive threshold needs to be set for the first filtering of key frames. The calculation method is as follows:

$$\vartheta = \frac{1}{u-1}\left[\beta\left(\alpha_1,\alpha_2\right)+\beta\left(\alpha_2,\alpha_3\right)+\cdots+\beta\left(\alpha_{m-1},\alpha_m\right)\right], \quad (3)$$

where ϑ represents the adaptive threshold. Finally, the target detection network YOLOv3 is used to obtain the target category, number, and position in the primary key frame, and the final key frame is filtered by comparing the similarity
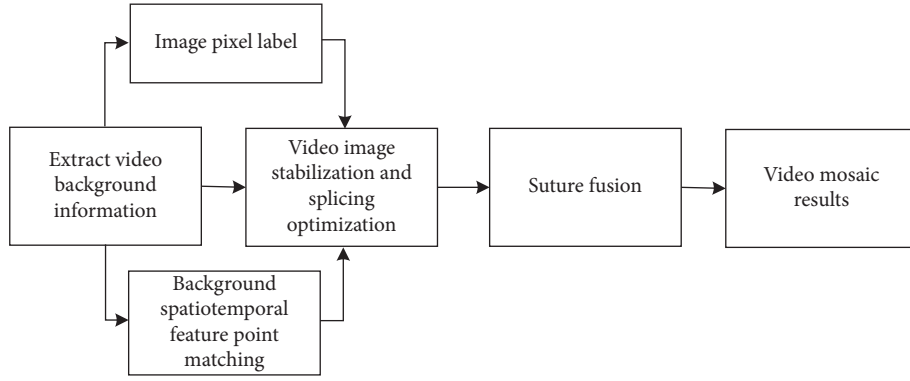
FIGURE 4: Dynamic video splicing model of film and television postproduction.

in the background to complete the dynamic video splicing of postproduction. The dynamic video splicing model is shown in Figure 4.

Video dynamic stitching can be seen as the process of labeling each pixel in the image. Different labels represent the different classification of pixels. The homography matrix not only determines the transformation relationship of points on two images but also directly determines the quality of stitched images. For the same mosaic image, or the image with little change in the position of pixels on the image, the homography matrix is almost the same [14]. From this feature, we can consider starting with the interframe similarity of video pictures and reducing repeated calculation by detecting repeated frame pictures, so as to improve the real-time performance of video splicing. The conditional random field is a probability graph model, which can model the label of pixels as random variables and design the objective function. By minimizing the objective function, the optimal image segmentation result is obtained. The expression of the objective function is as follows:

$$G = \sum_c f(z_1) + \sum_r h(z_1, z_2), \qquad (4)$$

where G represents the objective function; z_1, z_2 indicate the allocation method of two groups of labels; f is a univariate potential function, which represents the probability of assigning labels to pixels; h is the binary potential function, which represents the similarity of a pair of pixels; c, r represent the total number of labels and pixels. Because the full convolution neural network can output the probability that each pixel in the image is divided into a certain category, the univariate potential function can be calculated by the full convolution neural network. The two-frame difference method is used to screen and compare the repetition rate between the current frame and the previous frame. The frames whose repetition rate exceeds the threshold can be considered as the picture does not change, and the previously calculated homography matrix is directly used for splicing. For the frames that do not reach the threshold, it is considered that the picture has changed, and then, the homography matrix is recalculated. This decision method based on two-frame difference method is essentially the embodiment of "and" logical relationship. The more similar the pixels are, the greater the penalty value when they are marked as different labels. The two-frame difference method can capture the time when the jitter is relatively severe according to the similarity of the left and right frame images, and the projection transformation matrix in the subsequent stitching process is updated [15]. By minimizing the objective function of the conditional random field, the most likely label can be assigned to each pixel in the input image. Combined with the location information of feature points and classification label information in the image, the feature matching points in the foreground can be separated from the feature matching points in the background, leaving the temporal feature matching points and spatial feature matching points in the background. The matching points are substituted into the joint optimization framework of video image stabilization and splicing. RANSAC algorithm is used to purify the feature point pair and calculate the projection transformation matrix at this time. If it is a nonkey frame, the projection matrix calculated from the previous pair of key frames is directly used for interframe projection transformation. Finally, the left and right frames are fused to form a complete mosaic image. In the research process, the video image can be divided into multiple grids, each grid corresponds to a local area in the image, and the transformation matrix of multiple groups of grid graphics is used to represent the original motion path of the camera. In this way, the mesh transformation can be smoothed to achieve the purpose of stabilizing the video. So far, the design of film and television postproduction methods based on digital media technology has been completed.

3. Experiment

3.1. Experimental Preparation. In this paper, the experimental hardware used Intel Xeon E5-1650 V4 CPU, 3.6 GHz $*$ 12, 62 GB memory server, and the software used 64 bit Linux operating system, Matlab R2015a simulation. The data set used in the experiment is UCF101. The dataset contains 101 types of actions, and each class contains 25 videos, totaling 2525 videos. The average length of each video is 3s, the frame rate is 30 frames per second, and the size of each video is 320 $*$ 240 pixels. Each type of video is spliced together according to the type to generate 101 long videos, which are used as the videos to be postproduced. The

TABLE 2: Comparison of stable scores of film and television postproduction methods.

Number of tests	Film and television postproduction method based on digital media technology	Film and television postproduction method based on deep learning	Film and television postproduction method based on SIFT	Film and television postproduction method based on clustering algorithm
1	0.562	0.414	0.456	0.368
2	0.526	0.478	0.524	0.361
3	0.581	0.455	0.511	0.377
4	0.572	0.492	0.502	0.358
5	0.648	0.485	0.483	0.385
6	0.626	0.463	0.472	0.399
7	0.554	0.452	0.494	0.416
8	0.613	0.475	0.521	0.402
9	0.642	0.421	0.472	0.390
10	0.621	0.484	0.490	0.382

TABLE 3: Comparison of splicing errors of film and television postproduction methods.

Number of tests	Film and television postproduction method based on digital media technology	Film and television postproduction method based on deep learning	Film and television postproduction method based on SIFT	Film and television postproduction method based on clustering algorithm
1	1.854	2.624	2.443	3.167
2	1.941	2.697	2.451	3.284
3	1.772	2.848	2.629	3.058
4	1.886	2.655	2.546	3.163
5	1.856	2.886	2.684	3.025
6	1.924	2.952	2.456	2.934
7	1.832	2.860	2.632	2.901
8	1.813	2.721	2.751	3.128
9	1.702	2.615	2.528	3.146
10	1.725	2.602	2.312	3.257

videos to be processed were randomly divided into 10 groups for testing. In order to verify the application effect of the film and television postproduction method based on digital media technology, the video processing effect of the proposed method was compared with that of the film and television postproduction method based on deep learning, SIFT, and clustering algorithm.

3.2. Experimental Results and Analysis. In this paper, the stability score and splicing error are used to measure the stability of postproduction video and the effect of video editing. Taking into account the motion trajectory of feature points with a length greater than 1/10 of the total number of video frames, the stability score is calculated.

$$\lambda = \frac{\sum_\tau d_1}{\tau \sum_\tau d_2}, \qquad (5)$$

where λ represents the stability fraction; τ represents the total number of feature tracks; d_1 represents the true length of the feature track; d_2 represents the linear distance from the initial feature point to the end feature point. The higher the stability score of postediting, the more stable the generated video is. The common methods of film and television postproduction include deep learning, clustering algorithm, and SIFT. The comparison results of digital media technology and stability scores are shown in Table 2.

According to the comparison results in Table 2, the stability score of film and television postproduction method based on digital media technology is 0.595, which is 0.133, 0.102, and 0.211 higher than postproduction methods based on deep learning, SIFT, and clustering algorithm, respectively. The postproduction method of digital media technology uses the time feature matching points in the background, which can eliminate the interference of motion foreground and better recover the camera motion, so as to improve the stability of video editing.

The projection error of feature points near the splicing line of film and television postproduction clips is used to measure the splicing quality of the splicing results. The smaller the projection error is, the higher the image splicing quality is. Taking all frames of the video into account, the average splicing error is taken as the splicing error of the spliced video. The lower the average splicing error obtained from later clips, the higher the accuracy of video clips. The splicing error comparison results of various film and television postproduction methods are shown in Table 3.

According to the comparison results in Table 3, the average splicing error of film and television postproduction methods based on digital media technology is 1.831, which is 0.915, 0.712, and 1.275 lower than postproduction methods based on deep learning, SIFT, and clustering algorithm, respectively. The postproduction method of digital media technology can align the background area more accurately, so as to improve the accuracy of video splicing. Therefore,

the method designed in this paper can reduce the phenomenon of video blur and ghosting in film and television postproduction and achieve a more natural video postprocessing effect.

4. Conclusion

Compared with images, video content is richer and can bring people a better visual experience. Multimedia technology and other high-tech means make the film art creation enter the simulation stage of the simulation era, transform the aesthetic elements of the basic aesthetic category, form the film aesthetic characteristics of the digital era, make the digital film aesthetics have unique vision and connotation, and bring the freedom and transcendence of aesthetic experience to the audience. The main conclusions of this study are as follows:

(1) This paper proposes a film and television postproduction method based on digital media technology, which can improve the stability of postproduction video and the effect of video editing

(2) In the future work, we need to further reduce the complexity of the video splicing algorithm and hope to output the spliced video in real time, so that digital media technology can be applied to more fields.

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.

Acknowledgments

References

[1] Z. Wang, "On the development of film and television production and communication in the new media era," *International Journal of Intelligent Information and Management Science*, vol. 10, no. 1, pp. 6–8, 2021.

[2] A. Kadi, "Talk about the process of film and television production," *International Journal of Social Science and Education Research*, vol. 3, no. 9, pp. 112–116, 2021.

[3] H. Nandini, M. Chethan, and B. Rashmi, "An efficient method for video shot transition detection using probability binary weight Approach," *International Journal of Computer Vision and Image Processing*, vol. 11, no. 3, pp. 1–20, 2021.

[4] A. Nedić and S. Lee, "On stochastic subgradient mirror-descent algorithm with weighted averaging," *SIAM Journal on Optimization*, vol. 24, no. 1, pp. 84–107, 2014.

[5] S. Zhou, X. Wu, Y. Qi, S. Luo, and X. Xie, "Video shot boundary detection based on multi-level features collaboration," *Signal, Image and Video Processing*, vol. 15, no. 3, pp. 1–9, 2020.

[6] N. Kumar, Y. Sharma, and S. Parul, "Key feature extraction for video shot boundary detection using CNN," *International Journal of Recent Technology and Engineering*, vol. 8, no. 5, pp. 4763–4769, 2020.

[7] H. Nicolas, "Shadow synthesis for video postproduction," *IEEE Signal Processing Letters*, vol. 12, no. 4, pp. 321–324, 2005.

[8] N. Astawa, K. Darma Putra, M. Sudarma, and R. Hartati, "Enhancement image intensity of HSV color space," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2s11, pp. 2583–2585, 2019.

[9] C. Wei, Z. Li, W. Martin et al., "Motion vector coding and block merging in the versatile video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3848–3861, 2021.

[10] Y. Guo, Q. Min, W. Guan, and J. Li, "Detection of video shot changes based on image matching," *Applied Mechanics and Materials*, vol. 263, no. 263, pp. 2064–2069, 2012.

[11] P. Yu, D. Yu, and A. Popkov, "Entropy-randomized projection," *Automation and Remote Control*, vol. 82, no. 3, pp. 490–505, 2021.

[12] Z. Zhang and W. Xue, "Video image mosaic via multi-module cooperation," *Optoelectronics Letters*, vol. 17, no. 11, pp. 688–692, 2021.

[13] D. Li, L, Chen, W. Bao, J. Sun, B. Ding, and Z. Li, "An improved image registration and fusion algorithm," *Wireless Networks*, vol. 27, no. 27, pp. 1–15, 2020.

[14] C. Du, J. Yuan, J. Dong, L. Li, M. Chen, and T. Li, "Gpu based parallel optimization for real time panoramic video stitching," *Pattern Recognition Letters*, vol. 133, no. 133, pp. 62–69, 2020.

[15] 文. Wen Gang, 苏. Su Dongqi, 隋. Sui Yongxin, 杨. Yang Huaijiang, and 孔. Kong Fanlin, "Calibration of fizeau wavelength-tuned phase-shifting interferometer based on two-frame differential average phase-shifting algorithm," *Chinese Journal of Lasers*, vol. 40, no. 11, p. 1108004, 2013.