Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

The application of computer multimedia technology in film and television post-production

Pengfei Sun[1,✉], Yue Hu[1], Yan Kang[1]

[1] *Shanghai XingJian College, Shanghai, 200072, China*

ABSTRACT

Computer multimedia technology has brought unprecedented innovation to the film and television production industry. Multimedia technology in film and television post-production mainly focuses on two aspects of image processing and audio processing, this paper selects the skin color enhancement and voice enhancement for further research. Adaptive skin color enhancement method is proposed, IMCRA-OMLSA audio enhancement method is selected, and relevant experiments are designed to compare this paper's method with other classical skin color enhancement and voice enhancement methods respectively, and the effectiveness of this paper's method in skin color enhancement and audio enhancement is examined through the results of subjective and objective evaluation. The accuracy and F1 value of this paper's adaptive skin color detection method are 0.961 and 0.945, respectively, and the performance of skin color detection is good. The adaptive skin color detection method in this paper has the best performance with a comprehensive evaluation score of 6.81. In the objective evaluation of speech enhancement, the PESQ, STOI, WSS, and RMSE values of IMCRA-OMLSA method in this paper are 2.03, 72.36, and 38.06, respectively, which are all optimal results. On subjective evaluation, the MOS value of IMCRA-OMLSA method is 1.88 which is the highest value.IMCRA-OMLSA method has the best performance for speech enhancement.

Keywords: adaptive skin color enhancement, IMCRA-OMLSA, speech enhancement, film and television post-production

1. Introduction

The development of computer network equipment promotes the reform and innovation of China's film and television industry, and a variety of different types of film and television works began to appear in front of people, which enriches people's entertainment life and brings spiritual products to people [1-2]. And with the improvement of people's quality of life and living standards, people began to have higher requirements for film and television works, and people's high requirements for the

✉ Corresponding author.
E-mail address: spf2050@163.com (P. Sun).
Received 20 March 2024; Revised 25 May 2024; Accepted 10 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-449

content and presentation of works to a certain extent promote the optimization and transformation of film and television works [3-5]. Generally speaking, there are three main steps in the production of a qualified film and television work, i.e., pre-preparation, actual shooting and post-production after shooting [6]. In contrast, people usually pay little attention to post-production, i.e., when people watch a movie or television work, most of them pay attention to the actors and the plot, but pay little attention to the post-production related content [7-9]. It is important to know that post-production is an indispensable part of the production of a movie or television work. Multimedia technology as a kind of information technology, its post-production of film and television play a very effective role in the technical support, can effectively enhance the entertainment, viewing and fun of film and television works, and then can enhance the overall quality of film and television works, so that the audience get a better audio-visual experience [10-14]. Therefore, through the application of multimedia technology in film and television post-production research and analysis, aimed at exploring the more flexible application of computer multimedia technology to promote the development and innovation of the film and television industry.

The combination of video and image reconstruction is the main work of film and television post-production, computer multimedia technology has a flexible editing mode of operation, to broaden the editing thinking, the film will be presented in front of people in an optimal way. Literature [15] discusses the role of computer multimedia technology in the late stage of film and television production, the use of its flexible mode of operation to simplify the film and television editing process, improve the efficiency of film and television production, but also through the provision of color rendering and other technical support to enrich the quality of film and television and rendering effects. Literature [16] shows that image splicing technology can enrich the video editing techniques of film and television works, which fully expresses the overall style of film and television works and the director's intention, and at the same time, it is conducive to shaping the image of the characters, rendering the environmental atmosphere, and improving the attractiveness of the video content. Literature [17] points out that the post-production of film and television works need computer multimedia technology to expand the creative space and enhance the performance effect, computer multimedia technology can improve the quality and efficiency of video editing, but also automatically deal with the color rendering of the video resources, in the field of film and television post-production is widely used.

With the help of computer multimedia technology, film and television works can also be produced through special effects will be conceived of the virtual world real to the audience to show out, giving the audience a fantasy world of imagination space. Literature [18] analyzes the role and value of computer digital animation technology in film and television art creation, through the production of special effects greatly enriches the picture texture of film and television works, which significantly improves the audience's audio-visual experience. Literature [19] clarifies that the development of virtual reality technology provides a broader space for film and television creation, the development of computer technology has enriched the film and television special effects technology, but also makes the film and television animation production process more scientific, compared with traditional film and television works, based on virtual reality technology, film and television works of higher popularity. Literature [20] explores the application of computerized 3D real-time drawing technology in film and television post-production, and designs a post-processing system that contains commonly used special effects, which can not only enhance the efficiency of the editor, but also increase the sense of realism and immersion in 3D film and television applications, which is of great practical significance. Literature [21] constructs a deep learning-based detail feature transfer relationship model in multi-frame movies and TSEs images from the perspective of art form change, enhances the detail structure and features of the special effects images by attenuating the detail action effects, and effectively solves the problem of edge blurring and distortion in multi-frame movies and images in the field of art form change.

In this paper, through the application of multimedia technology in film and television post-production, we select two aspects of skin color enhancement and speech enhancement from two fields of image processing and audio processing for in-depth study. Based on the dynamic skin color model, adaptive skin color modeling is carried out for film and television videos, and skin color correction is carried out for characters in film and television works. The IMCRA-OMLSA method is selected for audio enhancement based on the movie and television audio type recognition method and audio enhancement method. The multimedia post-production techniques proposed and selected in this paper are compared with other skin color enhancement and audio enhancement methods, and the efficacy of the adaptive skin color enhancement method and IMCRA-OMLSA audio enhancement method in this paper is determined through subjective and objective evaluations.

2. Application of multimedia technology in film and television post-production

2.1. Multimedia Technology in Film and Television Post-production

Film and television post-production, mainly refers to the film and television team in the end of shooting, the content of film and television works to carry out post-processing, such as in the film and television works to add the corresponding text, special effects, or sound processing, in order to let the film and television works to present a more ideal effect. In terms of the current status of film and television post-production, it mainly relies on graphic software, non-linear editing software, synthesis software, three-dimensional software and other ways to carry out the production.

2.1.1. Image processing. In the post-production of film and television, through the effective application of image digital production technology in multimedia technology, it can realize the optimization of the color of film and television works, and based on this technology, it can also realize the virtualization of film and television works, digital processing, which makes the image picture more vivid and realistic, and enhances the visual expression of film and television works.

In addition, digital color mixing technology is an important link in the post-production of film and television, which can not only enhance the visual effect of the film, but also deepen the emotional expression of the film. The development of digital color mixing technology promotes the innovation of film and television art, so that directors and screenwriters can more richly express their creativity and emotion, so that the film's artistry and viewability have been doubly enhanced.

The first level of color correction is to adjust the color deviation of the whole picture and correct the color balance of the picture, brighten the underexposed phenomenon of the picture, and at the same time darken the overexposed places, and match the color and exposure for the shooting shots of the same scene.

Secondary color grading is to deal with character skin color, sky, water and other details to deal with, reduce the interference of the environment on the characters, to protect the harmony of the picture. Color mixing is mainly for the color design and planning of the screen, and combined with the creative intent, to achieve the secondary light distribution of the screen, and tonal processing, to express the color emotion.

Through digital color mixing technology, film makers can make the picture more deep and hierarchical under the fine control of the light and shadow effect of each scene. In film and television works, the character image is the core of the whole story, and in order to make these characters more vivid, the production team needs to pay attention to several aspects. In the post-production stage, the staff must improve the overall picture effect and the skin color presentation of the characters in different environments with the help of digital color mixing technology and other means to fill in the lack of emotional expression of the characters and the lack of atmosphere. Through clever color mixing, the characters are made more realistic and three-dimensional, guiding the audience to be

able to deeply feel the emotions of the characters, increasing the attractiveness of the work and attracting the attention of more viewers.

2.1.2. Audio processing. Audio post-production is an important part in the production of movie and television works, including steps such as editing, processing and mixing of the recorded material. In the audio editing stage, the original audio material obtained from the shooting site is imported into the audio editing software, and after editing, the useless parts are deleted, only the needed audio clips are retained, and the audio clips are organized and filed according to the scene and content. After that, audio cleaning and restoration, dialog and sound design, audio mixing, stereo and surround sound processing, as well as output and format conversion are carried out, and the processed audio is applied to the final film to bring the audience a high-quality viewing experience.

2.2. Skin Color Enhancement Techniques for Film and Television Characters

Skin tone enhancement is part of image enhancement technology designed to individually color correct or compensate for the sensitive skin areas of the human eye to show a more natural skin tone performance and enhance the display quality of the picture without affecting the overall tonal style of the picture, or other areas.

2.2.1. Video Adaptive Skin Color Modeling. Based on the results of the lens boundary detection, it can be determined whether the current frame belongs to a new model update unit. If no new unit appears, skin color detection continues using the current skin color model parameters [22-23]. If a new unit appears, face detection is started at the current frame and the skin color model is updated when reliable face feature points are obtained. The specific cases can be categorized as follows:

1) No new model update unit is found. Directly use the last stored model parameters for skin color detection.

2) A face is detected at the first frame of the skin color model update unit. Update the skin color model parameters according to the skin color reference points of the face obtained at the current frame, use the set of thresholds for skin color detection, and store the current model parameters into memory.

3) No face was detected at the first frame of the skin color model update unit. Comparison with keyframes of skin color model updating units that have already appeared determines whether similar footage clips have already appeared and a reliable skin color model has been established. In this paper, the length of the unit keyframe queue K is set to 10, i.e., there are $n \in [0,10]$. Denote the penultimate n boundary in K as k':

$$k' = K(10-n) \tag{1}$$

Use equation (1) to calculate the difference between the first frame stored in the penultimate n boundary in K and the current frame i, denoted $S_2'(k',i)$:

$$S_2'(k',i) = \alpha_2 \times S_D(k',i) + (1-\alpha_2) \times S_H(k',i) \tag{2}$$

If $S_2'(k',i)$ is less than T_{2L} it means that the current frame is similar to the stored boundary, read the corresponding skin color threshold in the corresponding skin color model update unit.

4) No keyframe similar to the current frame is found in queue K. It means that the current segment has not appeared, and it is necessary to continue to maintain face detection and appropriately reduce the frequency of face detection.

After the above steps, a number of groups of thresholds from different faces can finally be obtained $T_1, T_2 \ldots T_{\max}$. Let the corresponding upper and lower thresholds of the first group be $T_{1\_Cr\_\max}$, $T_{1\_Cb\_\max}$, $T_{1\_Cr\_\min}$, $T_{1\_Cb\_\min}$, respectively, and the same for the other groups. Thus the total threshold T_{total} can be obtained by equation (3):

$$T_{total} = \begin{cases} T_1 : \{T_{1\_Cr\_\min} \le Cr(i,j) \le T_{1\_Cr\_\max} \\ T_{1\_Cb\_\min} \le Cb(i,j) \le T_{m\_Cb\_\max}\} \\ T_m : \{T_{m\_Cr\_\min} \le Cr(i,j) \le T_{m\_Cr\_\max} \\ T_{m\_Cb\_\min} \le Cb(i,j) \le T_{m\_Cb\_\max}\} \end{cases} \tag{3}$$

Let the binary skin color mask generated by skin color detection be denoted as Mask, and $Mask(i,j)$ be 255 when the $Cb(i,j)$ and $Cr(i,j)$ components of the current pixel satisfy T_{total} at the same time, and 0 otherwise.A large number of unnecessary face detection and skin color model updating are avoided by the above method, which makes the skin color detection faster, and helps to realize real-time processing.

2.2.2. Skin color correction. After obtaining the skin color mask by the adaptive skin color method, the skin color region needs to be corrected.

1) Acquisition of preferred skin color models The test images contain a total of 10 images each from 3 groups of black, yellow, and white races, and each group contains skin color images with different brightness. In this paper, the skin color brightness is divided into 5 levels on average, from dark to bright: very dark, dark, normal, bright, very bright, and each brightness level contains 2 test images.

In the YCbCr color space, nine candidate preference images are obtained by expanding the chromaticity mean value of the skin tone region in each test image by 2 units in all directions, respectively. Firstly, the skin color region is labeled manually, and then the skin color pixels $Cb(i,j)$ are adjusted, and the candidate images are denoted by $l(l = 1,2,\ldots,9)$ to indicate the sequential number from left to right and from top to bottom, and the pixels of the candidate images can be obtained by Eqs. (4) and (5):

$$\Delta Cb_l = 2*[l - 3*((l-1)\backslash 3)] - 4, \Delta Cr_l = 2*((l-1)\backslash 3) - 2 \tag{4}$$

$$Cb_l(i,j) = Cb(i,j) + \Delta Cb_l, Cr_l(i,j) = Cr(i,j) + \Delta Cr_l \tag{5}$$

There are certain patterns in the human eye's preference for skin color. For skin color, the larger the Cb component, the bluer the skin color, the smaller the more yellow, the larger the Cr component, the more red, the smaller the more green.

2) Racial skin color partition. In this paper, the dynamic skin color model is obtained in the skin color detection stage, in order to make the effect of skin color enhancement algorithm more in line with the human eye's preference for different races of skin color, this paper proposes a racial skin color partitioning method.

First, the binary mask generated by the skin color detection is analyzed for the connected domains, and the average chromaticity and luminance values for each connected domain are counted. Then, the obtained mean chromaticity of each region is noted as Cb_m and Cr_m, and the skin color luminance level of the region is determined based on the mean luminance Y_m. Next, using the luminance classes, the probabilities of the three races under the current luminance class condition are obtained and are noted as P_r ($r = 1,2,3$ corresponds to yellow, white, and black races). Then

determine the chromaticity information of the three races under the current luminance condition, noted as Cb_{mr} and Cr_{mr}.

The Euclidean distance between the current region and the three races' skin colors can be calculated through equation (6) and is noted as D_r:

$$D_r = (1 - P_r)\sqrt{(Cb_m - Cb_{mr})^2 + (Cr_m - Cr_{mr})^2} \tag{6}$$

Compare the three Euclidean distance values and use equation (7) to find the shortest distance and record the corresponding race r:

$$D_{r\min} = \min(D_1, D_2, D_3) \tag{7}$$

3) Skin color correction. After obtaining the luminance Y_m and the belonging race r corresponding to each skin color region, it is necessary to offset each skin color region towards the pre-determined preferred skin color. Chromaticity conversion, for a pixel, is the process of converting from the original chromaticity value to the preferred chromaticity value. The corresponding preferred skin color is first obtained based on the luminance Y_m and the race affiliation information r and noted as Cb_p and Cr_p. Then the Cb and Cr chromaticity means of each skin color region are calculated, noted as $\overline{Cb}$, $\overline{Cr}$, and compared to the preferred skin color to obtain the distances from the original skin color mean to the preferred skin color, d_{cb} and d_{Cr}, which represent the chromatic aberration from the preferred skin color:

$$d_{cb} = Cb_p - \overline{Cb}, d_{cr} = Cr_p - \overline{Cb} \tag{8}$$

Since a set of skin color thresholds is used in the skin color detection process to obtain specific skin color regions, when determining the skin color enhancement amplitude, it is necessary to take into account that the colors located near the two sides of the thresholds may be disconnected in the adjustment process. For this reason, the C_{Cb_ref} parameter is introduced to constrain it, with the aim of making the adjustment amplitude of skin color regions close to the thresholds small and the adjustment amplitude of skin colors farther from the thresholds on both sides larger, which is calculated as shown in Eq. (9) and Eq. (10):

$$M_{Cb} = \frac{T_{Cb\_\min} + T_{Cb\_\max}}{2}, M_{Cr} = \frac{T_{Cr\_\min} + T_{Cr\_\max}}{2} \tag{9}$$

$$C_{Cb\_ref_i} = c_0 e^{-\frac{(Cb(i,j) - M_{Cb})^2}{2}}, C_{Cr\_ref_i} = c_0 e^{-\frac{(Cr(i,j) - M_{Cr})^2}{2}} \tag{10}$$

where $Cb(i,j)$, $Cr(i,j)$ represent the Cb and Cr components corresponding to the skin color pixels at the (i,j) locations, respectively, and c_0 represents the overall skin color adjustment amplitude control parameter, which was set to 0.25 based on the experimental subjective effect.After the above analysis, for the pixels within each skin color region, the enhanced components $Cb'(i,j)$, $Cr'(i,j)$ were calculated by Eqs. (11) and (12):

$$Cb'(i,j) = Cb(i,j) + d_{Cb} C_{Cb\_ref_i} \tag{11}$$

$$Cr'(i,j) = Cr(i,j) + d_{Cr} c_{Cr\_ref_i} \tag{12}$$

Adjustment of a connected region can be achieved by the above step, and this step is repeated until all connected regions are processed, so that an image with only the skin color region processed can be obtained, which is recorded as Dst. It also needs to be merged with the original image to obtain the final enhancement effect. In order to avoid the phenomenon of hard transition at the skin tone boundary, the skin tone mask needs to be mean filtered before merging to soften the transition at the

edges, and the original binary image is grayed out to $[0,255]$. Setting the pixel value in the original mask as $Mask(i,j)$ and $\Omega_{(i,j)}$ as the neighborhood of pixel (i,j) containing M points, the pixel value $Mask'(i,j)$ after mean filtering can be obtained by Eq. (13):

$$Mask'(i,j) = \sum_{(x,y)\in\Omega_{(i,j)}} \frac{Mask(x,y)}{M} \tag{13}$$

The original image and the skin color correction result are merged using the Alpha-blending method with $Mask'$ pixel value as the control parameter. The original image before processing is denoted as Src, the processed skin color image as Dst, and the merged result is denoted as Result.The calculation of the fusion of the two is shown in equation (14):

$$\mathrm{Re}\,sult = Src * \alpha + Dst\ * \frac{Mask'}{255} * (1-\alpha) \tag{14}$$

If the skin color model is updated at an intermediate position during the skin color detection phase, it is necessary to use a gradient transition for the next segment of the sequence to avoid flickering. The transition is mainly realized by adjusting the mixing parameter α in Eq. (14). The calculation of α is given in Eq. (15):

$$\alpha = \frac{1}{L}\alpha_0 * m \tag{15}$$

Wherein, L represents the length of the transition interval, α_0 represents a fixed parameter during ordinary frame fusion, m is the number of frames away from the update position, and there are $0 \le m \le L$. This paper sets $L = 50$, $\alpha_0 = 0.8$. By the above method, the processed skin color region can be merged with the original image, and then the subsequent output display can be performed and the next frame processing can be continued.

2.3. Movie and TV Speech Enhancement Technology

2.3.1. Audio Type Recognition Methods. In this paper, the identification of audio types is divided into the identification of background noise types and the identification of target audio types, and the same need to design classifiers to realize the type recognition, about the model of audio type recognition has been a variety of programs can be studied similarly, mainly divided into machine learning-based models and deep learning-based models. Machine learning based models include Gaussian mixture models, Hidden Markov Models, etc. Deep learning based models include Convolutional Neural Networks, Recurrent Neural Networks, Fully Connected Neural Networks, and their various improvement methods.

2.3.2. Audio enhancement methods. Currently, audio enhancement methods can be roughly divided into three categories, namely signal processing-based enhancement methods, statistical model-based enhancement methods, and machine learning-based enhancement methods. Among them, statistical model-based methods are the most commonly used methods for real-time audio enhancement, including the Wiener filtering method and the IMCRA-OMLSA method, etc. The principle is to utilize mathematical and statistical methods to estimate the noise and target audio components corresponding to each frequency point in the noise-containing frequency signal spectrum. In this section, the most typical Wiener filtering method is analyzed and introduced.

Wiener filtering is an optimal linear filtering method that minimizes the mean value of the error between the output audio and the target audio based on the minimum mean square error criterion [24-25]. The specific steps in performing audio enhancement are as follows:

Let the noise-containing frequency signal be $x(n)=s(n)+d(n)$, and the desired signal $\hat{S}$ is obtained after passing through the FIR filter with filter coefficient $h(k)$:

$$\hat{s}(n)=\sum_{k=0}^{M-1}h(k)x(n-k) \tag{16}$$

The minimum mean square error criterion can be utilized to make the acquired desired signal $\hat{s}$ infinitely close to the target signal, as shown in equation (17):

$$E\left\{e(n)^2\right\}=E\left\{(\hat{s}(n)-s(n))^2\right\} \tag{17}$$

by the orthogonality principle:

$$E\left\{x(n-k)e^*(n)\right\}=0,k=0,1,2,\mathrm{L},M-1 \tag{18}$$

The Fourier transform of Eq. (18) yields Eq. (19):

$$H(k)=\frac{P_{xy}(k)}{P_x(k)} \tag{19}$$

where $P_x(k)$ is the power spectral density and $P_{xy}(k)$ is the reciprocal power spectral density of noise and noise-containing frequency signals. Since pure audio and noise are independent of each other, Eq. (20) and Eq. (21) can be obtained:

$$P_{xy}(k)=P_s(k) \tag{20}$$

$$P_x(k)=P_s(k)+P_d(k) \tag{21}$$

Substituting Eq. (20) and Eq. (21) into Eq. (19) yields Eq. (22):

$$H(k)=\frac{P_S(k)}{P_S(k)+P_d(k)} \tag{22}$$

On the frequency domain k, the $\hat{s}(n)$ prediction of the spectrum of the nnd frequency point $\hat{S}(k)$ can be expressed by equation (23):

$$\hat{S}(k)=H(k)\cdot Y(k) \tag{23}$$

In complex real-world scenarios, the method is able to reduce some of the effects of musical noise based on the extracted information. However, similar to spectral subtraction, the method requires prerequisite assumptions such as noise smoothness, and the enhancement effect may be affected in non-smooth noise environments.

2.3.3. IMCRA-OMLSA audio enhancement methods. Cohen et al. first proposed the minimum controlled recursive averaging (MCRA) method in the literature and improved it [26]. The improved MCRA method first determines whether the signal spectrum contains audio components or not, and then performs noise estimation only on the pure noise components, avoiding the influence of audio interference, which is known as the Improved Minimum Controlled Recursive Averaging (IMCRA) method.

According to the way of combining noise and signal in the noise-containing frequency signal, the noise can be classified into two types: additive noise and multiplicative noise, both of which are combined with the signal as shown in Eq. (24) and Eq. (25), respectively:

$$y(t)=x(t)+n(t) \tag{24}$$

$$y(t)=x(t)*n(t) \tag{25}$$

where $y(t)$ is the noise-containing frequency signal at moment t, $x(t)$ is the pure audio signal, and $n(t)$ is the noise signal. In practice this can be done by converting multiplicative noise into additive noise by homomorphic transformation (logarithmic operation), so in the next step the audio signal assuming only additive noise is processed.

A short-time Fourier transform of the preprocessed noise-laden frequency signal is obtained:

$$Y(k,l) = X(k,l) + N(k,l) \tag{26}$$

where k denotes the frequency index and l denotes the frame index.

Assume that $p(k,l)$ is the probability of the presence of the audio signal in the noise-containing signal and $q(k,l)$ is the probability of the absence of the audio signal. The recursive mean of the power spectrum of the estimated noise is calculated according to the IMCRA method:

$$\bar{\lambda}_d(k,l+1) = \tilde{\alpha}_d(k,l)\bar{\lambda}_d(k,l) + \left[1-\tilde{\alpha}_d(k,l)\right]|Y(k,l)|^2 \tag{27}$$

The presence probability $p(k,l)$ of the audio is recursively smoothed in time to obtain $\tilde{\alpha}_d(k,l)$, which is calculated as shown in equation (28):

$$\tilde{\alpha}_d(k,l) = \alpha_d + \left(1-\alpha_d\right)p(k,l) \tag{28}$$

where α_d is the smoothing factor, which is set to a value range of $(0,1)$.

Also in the process of estimating the noise power spectrum, a bias factor is introduced to reduce the effect of distortion:

$$\hat{\lambda}_d(k,l+1) = \beta * \bar{\lambda}_d(k,l+1) \tag{29}$$

where the magnitude of β is determined by the a priori audio non-existence probability. Eq. (30) and Eq. (31) are the definitions of the a priori signal-to-noise ratio and a posteriori signal-to-noise ratio of the signal, respectively:

$$\xi(k,l) = \frac{\lambda_s(k,l)}{\lambda_d(k,l)} \tag{30}$$

$$\gamma(k,l) = \frac{|Y(k,l)|^2}{\lambda_d(k,l)} \tag{31}$$

where, $\lambda_s(k,l)$ is the power spectrum of the pure target audio signal and $\lambda_d(k,l)$ is the power spectrum of the noise signal.

A Gaussian statistical model is utilized in the IMCRA method to estimate the probability of audio presence, which is calculated as follows:

$$p(k,l) = \left\{1 + \frac{q(k,l)}{1-q(k,l)}(1+\xi(k,l))\exp(-v(k,l))\right\}^{-1} \tag{32}$$

where $v(k,l) = \gamma\xi/(1+\xi)$. In order to obtain the probability of absence of audio in the noise-containing frequency signal $q(k,l)$, one minimum search and two smoothing operations are required.

The result of the first power spectrum frequency smoothing is shown in equation (33):

$$S_f(k,l) = \sum_{i-\infty}^{o} b(i)\left|Y(k-i,l)\right|^2 \tag{33}$$

where b is the normalized window function with window lengths $2\omega+1$, $\sum_{i-\omega}^{\omega} b(i) = 1$.

The result of the smoothing operation in the time domain dimension is shown in equation (34):

$$S(k,l)=\alpha_s S(k,l-1)+(1-\alpha_s)S_f(k,l) \tag{34}$$

where α_s is a smoothing parameter in the range of $(0,1)$. The same smoothing parameter α_s is used for subsequent operations when performing the second power spectrum smoothing. It is also necessary to calculate the audio non-existence probability using the a priori signal-to-noise ratio, and Eq. (35) shows the formula for calculating the a priori signal-to-noise ratio in the IMCRA method:

$$\tilde{\xi}(k,l)=\alpha G_{H_1}^2(k,l-1)\gamma(k,l-1)+(1-\alpha)\max\{\gamma(k,l)-1,0\} \tag{35}$$

where α is a parameter that weighs speech distortion against noise suppression and G_{H_1} is a spectral gain function.

An estimate of the probability of audio non-existence can be obtained by the above computational steps, the power spectrum of the noise is further estimated, and then the power spectrum of the pure target audio signal is estimated using the Optimally Modified Logarithmic Spectral Amplitude Method (OMLSA).

The gain function is first computed using the noise power spectrum estimated in the above process, and then the noise-containing frequency signal power spectrum is filtered using this function, and the estimated pure target audio signal is computed using equation (36):

$$\hat{A}(k,l)=G(k,l)Y(k,l) \tag{36}$$

where $\hat{A}(k,l)$ is the estimated pure audio signal, $G(k,l)$ is the estimated gain of the noise-containing frequency signal, and $Y(k,l)$ is the noise-containing frequency signal.

The OMLSA method essentially minimizes the difference between the actual pure audio signal and the estimated pure audio signal, and this difference can be expressed by equation (37):

$$Delta=E\left(\left|\log A(k,l)-\log\hat{A}(k,l)\right|^2\right) \tag{37}$$

where $A(k,l)$ is the actual pure audio signal power spectral amplitude and $\hat{A}(k,l)$ is the estimated pure audio signal power spectral amplitude. After minimizing the difference $Delta$ the gain function $G(k,l)$ can be calculated with the following expression:

$$G(k,l)=G_{H_1}(k,l)^{p(k,l)}G_{\min}^{1-p(k,k)} \tag{38}$$

where G_{H_1} is the spectral gain function. $G_{\min}$ is the gain function when the audio is not present, and the value of $G_{\min}$ is generally set to the minimum signal-to-noise ratio of the noise-containing frequency signal. $p(k,l)$ is the a posteriori probability of the presence of audio.

3. Film and television post-production effects

3.1. Character skin color enhancement effect analysis

3.1.1. Comparison of skin color detection:

1) Image processing comparison. Skin color detection, as a pre-step of skin color enhancement, and its detection performance will have a significant impact on the processing results of skin color enhancement. This section compares the adaptive skin color detection method proposed in Section 2.2 of this paper with the more representative skin color detection methods in the current technology field.

1000 sample images are randomly selected from different human skin color datasets and manually labeled with their corresponding skin color masks. Skin color masks are computed for the sample

images separately using different skin color detection methods. The results of the mask calculation are compared with the manually labeled skin color mask, and objective evaluation indexes such as detection accuracy and F1 score are calculated respectively.

In order to compare the skin color detection performance of different methods with objective criteria from the perspective of quantitative evaluation, the skin color mask is calculated for 1000 sample images using the above methods respectively, and compared with the corresponding standard mask, and the TP, TN, FP and FN of all detection results are counted, and the detection accuracy and F1 scores of each method are calculated throughout the whole experimental process, and the comparison results are shown in Table 1.

As can be seen from the experimental results in Table 1, the three classical skin color model-based skin color detection methods are limited by the model fitting ability, and the detection performance evaluation is not satisfactory. The adaptive skin color detection method in this paper is 0.961 and 0.945 in terms of accuracy and F1 score, respectively, both of which reach more than 0.9, which is significantly better than the adaptive threshold modeling method, adaptive ellipse modeling method, Gaussian mixture modeling method, human eye inspection sampling method and lightweight skin color segmentation network, and achieves a large increase in terms of F1 score when compared with the human eye inspection sampling method.U-Net, DeepLab V3+ and adaptive skin color detection methods show better detection performance than other types of methods. Among them, the adaptive skin color detection method in this paper achieves similar skin color detection performance to U-Net and DeepLab V3+ with a more streamlined model structure.

Table 1. Quantitative comparison of skin color detection performance

Skin color detection	Accuracy	F1
Adaptive threshold modeling method	0.783	0.792
Adaptive elliptical modeling method	0.844	0.808
Gaussian hybrid model method	0.825	0.830
Human eye sampling method	0.923	0.873
Lightweight skin color segmentation network	0.942	0.934
U-Net	0.962	0.945
DeepLab V3+	0.959	0.943
Adaptive skin color detection method	0.961	0.945

2) Comparison of video processing. A number of clips were selected from the video materials chosen for the experiment, including all kinds of topics and scene environments as far as possible, such as sports, movies, advertisements, reality shows, news and interviews, and so on. A certain length of clips were intercepted from each clip, and all the clips were edited and spliced together to form a test video. Different methods are used to detect the skin color of the test video frame by frame, among which, the adaptive skin color detection method in this paper adopts video segmentation and node labeling method for model update guidance. From the starting frame to the 2400th frame of the test video, the skin color detection F1 score of the current detection method in the past 100 frames is calculated every 100 frames. The experimental results are shown in Fig. 1, where ATMM, AEMM, GHMM, HESM, LSCSN, and ASCDM stand for Adaptive Threshold Modeling Method, Adaptive Elliptic Modeling Method, Gaussian Mixture Modeling Method, Human Eye Inspection Sampling Method, Lightweight Skin Tone Segmentation Network, and Adaptive Skin Tone Detection Method proposed in this paper, respectively.

As can be seen from Fig. 1, the adaptive skin color detection method in this paper can maintain a more stable skin color detection performance in different video scenes, and its skin color detection F1 score curve basically maintains above 0.9, which is significantly better than the adaptive threshold modeling method, adaptive ellipsoid modeling method, Gaussian mixture modeling method, human eye inspection sampling method and lightweight skin color segmentation network. Compared with other types of methods, U-Net, DeepLab V3+ and adaptive skin color detection methods show significant performance advantages in different video scenes.

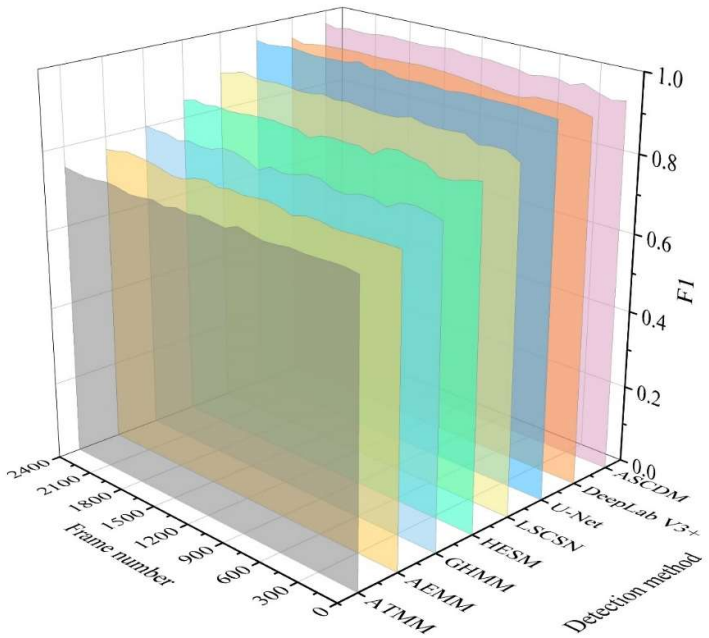
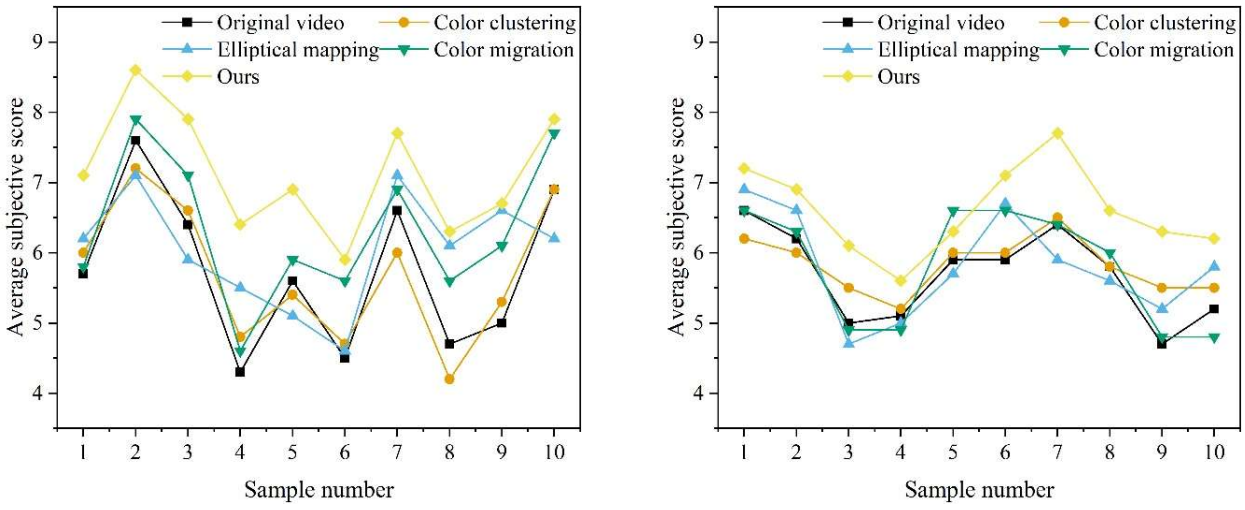


Fig. 1. Comparison of performance of video color test

3.1.2. Skin color enhancement contrast. This section compares the skin color enhancement method based on adaptive skin color modeling in this paper with more representative skin color enhancement methods in the current state of the art. The skin color enhancement comparison experiments are conducted using subjective experimental methods. Ten sample videos with a resolution of 3840×2160 and a frame rate of 60fps were selected from the video clips chosen for the experiment, the video content contained all kinds of scene environments as much as possible, and each video contained three types of racial skin tones: yellow people, white people, and black people. Different skin color enhancement methods were used to enhance the skin color of the eight sample videos respectively, and all video clips including the original video were reordered and played.

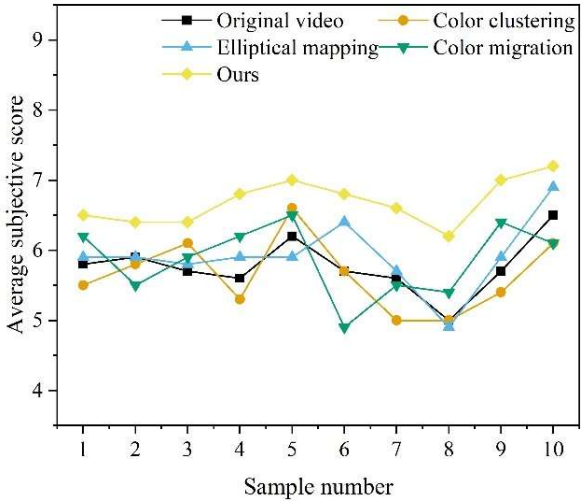
For the subjective experiment, 25 viewers aged between 20 and 50 were invited to serve as testers for the experiment. The subjective experiment utilized a single-stimulus continuous quality evaluation method in which testers subjectively rated the video segments they viewed based on a subjective rating scale without having viewed the original video. Ratings were given for each of the three racial skin color categories in each video segment.

At the end of the experiment, the average subjective scores obtained from all the video clips were calculated separately, and the results are shown in Fig. 2, with (a) to (c) showing the subjective ratings of the experimenters on the color of the yellow, white, and black racial skin tones in each video, respectively. From the experimental results, it can be seen that the chromaticity clustering method and the elliptic mapping method are limited by the less adaptive skin color modeling method, the detection accuracy of human skin color in the video clip is low, the color enhancement effect is not ideal, and the color enhancement for different races of skin color lacks of relevance, and even in the case of individual samples after the skin color enhancement treatment subjective evaluation instead of a decline in the situation. The color migration method obtains higher subjective evaluations in most samples compared with the first two methods. The two skin color enhancement methods proposed in this paper use a more adaptive skin color detection method and color enhancement method, which can achieve effective skin color enhancement for different sample videos, and obtain the highest subjective evaluations compared with the chromaticity clustering method, ellipse mapping method, and color migration method.



(a) The subjective evaluation for yellow race

(b) The subjective evaluation for white race



(c) The subjective evaluation for black race

Fig. 2. Video color enhancement subjective evaluation contrast

According to Fig. 2, the scoring data of each method on 10 samples were averaged respectively, so as to evaluate the skin color enhancement effect of each method on the same type of human race and the comprehensive evaluation effect as shown in Table 2.

From the experimental results in Table 2, it can be seen that the adaptive skin color enhancement methods in this paper obtain a higher subjective evaluation than other methods in all aspects. Among them, the skin color enhancement method based on adaptive skin color modeling in this paper achieves about 18.85% subjective evaluation enhancement in terms of comprehensive evaluation compared with the original video, and obtains the highest subjective evaluation in all aspects.

Table 2. Comparison of video color enhancement comprehensive evaluation

Racial color	Original video	Color clustering	Elliptical mapping	Color migration	Ours
Yellow race	5.73	5.71	6.04	6.32	7.14
White race	5.68	5.82	5.81	5.79	6.60
Black race	5.77	5.65	5.92	5.86	6.69
Comprehensive evaluation	5.73	5.73	5.92	5.99	6.81

3.2. Analysis of movie and TV audio enhancement effects

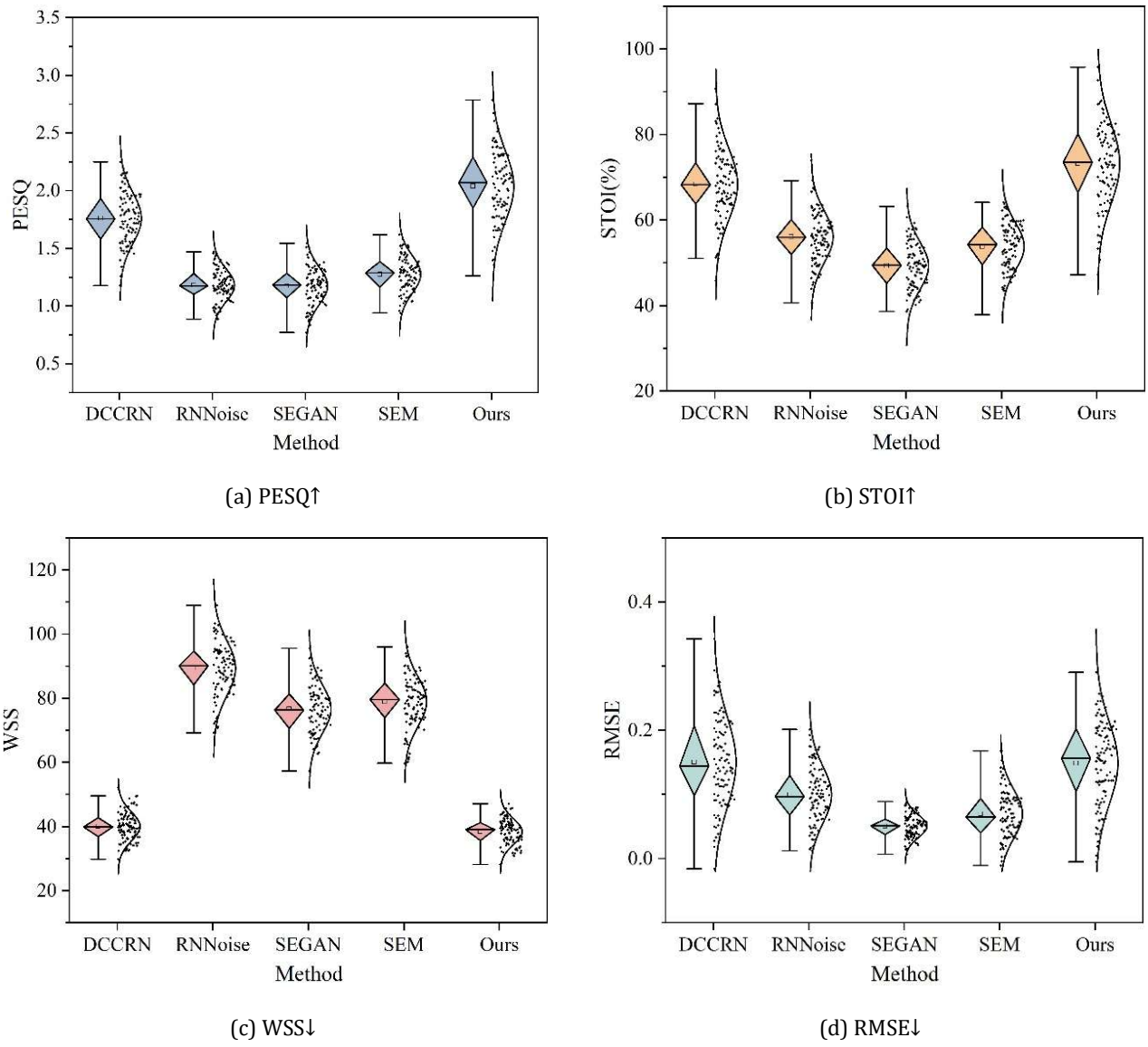
3.2.1. Hyperparameter settings. The effect of different α values on the speech enhancement effect of movie and television productions was investigated and validated with PESQ and STOI metrics on the NTCD-TIMIT validation set. The results are shown in Table 3. It can be seen that the two loss functions have different effects on speech enhancement under different weights. The speech enhancement effect will not always increase with the increase of the weighting of a single loss function, instead, a balance point needs to be found in the experiment to make it better to guide the weighting update to achieve the best speech enhancement result. As can be seen from the table, the optimal results are achieved for both metrics under the NTCD-TIMIT validation set for α = 0.4. This fully demonstrates that, compared to using a single loss function, the loss function used in this paper can provide more information for the weight update of the network, which is conducive to achieving better speech enhancement results.

Table 3. Average value of PESQ and STOI metrics under the NTCD-TIMIT verification dataset

Parameter	PESQ ↑	STOI(%) ↑
α=0	2.19	61.31
α=0.1	2.14	60.27
α=0.2	2.10	62.40
α=0.3	2.10	62.21
α=0.4	2.27	64.41
α=0.5	2.13	61.61
α=0.6	2.10	63.42
α=0.7	2.15	62.44
α=0.8	2.11	62.23
α=0.9	2.17	60.34
α=1	2.16	62.41

3.2.2. Results and Discussion. In order to validate the effectiveness of the IMCRA-OMLSA method, this subsection compares it with the DCCRN, RNNoise, SEGAN and Speech Ehancement (SEM) methods. The average results obtained by each method under each objective evaluation metric on the NTCD-TIMIT test dataset are shown in Figure 3. In this case, the box-and-line plot shows the distribution of the objective metrics results obtained under each method on the NTCD-TIMIT test set. As can be seen in Figure 3, compared to the other methods, the IMCRA-OMLSA method performs slightly worse in terms of RMSE, which is 0.15. However, the RMSE metric measures the error between the model prediction results and the actual observations, and even though it has a certain degree of informativeness for the evaluation of speech quality, it is usually necessary to combine it with other metrics consistent with the auditory perceptions of the human ear to measure speech quality in the actual evaluation. The IMCRA-OMLSA method, on the other hand, obtained the optimal values (2.03, 72.36, 38.06) under the other three objective metrics PESQ, STOI, and WSS. It can also be observed from the boxplot that the IMCRA-OMLSA method has a more even distribution of the

indicator result data on the NTCD-TIMIT test set. It is well illustrated that the speech enhancement results under the IMCRA-OMLSA method obtain higher speech quality and intelligibility as well as better robustness.



(a) PESQ↑ (b) STOI↑

(c) WSS↓ (d) RMSE↓

Fig. 3. Boxplot of the objective evaluation results of each method under Syn-DMU-VDR test set

Under the NTCD-TIMIT dataset, the subjective evaluation results are shown in Table 4 and the MOS violin plot is shown in Fig. 4. It can be seen that the IMCRA-OMLSA method in this paper not only obtains the overall optimal results on the NTCD-TIMIT test set, at the same time, it also obtains the highest MOS value (1.88) on the NTCD-TIMIT real audio test set, and the violin body shape is more uniform, which indicates that the IMCRA-OMLSA method has a more uniform distribution of the MOS data under a single audio , which is more robust to audio enhancement under different signal-to-noise ratios.

In addition, the MOS-N, MOS-C, MOS-D and MOS-L indicators in the NTCD-TIMIT dataset also reflect the effect of noise reduction to a certain extent, which represent four dimensions affecting speech quality: noise degradation perception dimension, coloring dimension, discontinuity dimension and loudness dimension, the values under the four dimensions reflect the degree of difference between before and after the noise-reduced audio, the noise-reduced audio, the The degree of noise degradation is obvious, and the coloring reflects the amount of energy in different

frequency bands, which to some extent reflects the timbre changes. Noise reduction techniques have a greater impact on the discontinuity dimension, which also reflects the advantages and disadvantages of noise reduction techniques to a certain extent. Loudness metrics are physical quantities used to describe the perception of sound intensity. Unlike the decibel (dB) of sound, the loudness metric is based on the characteristics of human hearing and can better reflect the human perception of sound intensity. Higher loudness values also emphasize the strong energy of the speech signal, which reduces audio background noise while the perception of speech loudness by the human ear increases accordingly.

Table 4. The result of subjective evaluation of each method under the NTCD-TIMIT dataset

Method	MOS ↑	NISQA			
		MOS-N	MOS-D	MOS-C	MOS-L
DCCRN	1.68	2.44	2.54	2.31	2.71
RNNoise	1.42	2.23	2.78	2.03	2.52
SEGAN	1.61	3.10	2.53	2.47	5.64
SEM	0.92	2.25	1.86	1.95	2.93
Ours	1.88	2.78	2.99	2.46	2.74

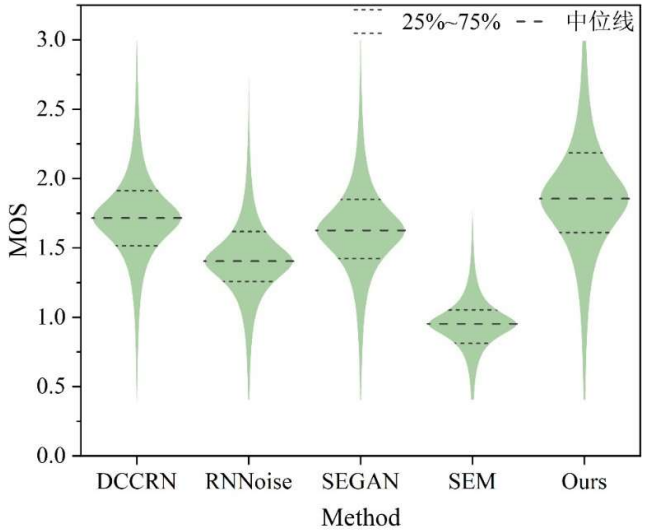


Fig. 4. The violin plot of the MOS metrics results of each method in the NTCD-TIMIT dataset

4. Conclusion

This paper analyzes the multimedia technology in film and television post-production, takes skin color enhancement technology and voice enhancement technology as the research object, and analyzes the application effect of adaptive skin color enhancement method and IMCRA-OMLSA audio enhancement method in this paper.

1) The accuracy and F1 value of U-Net, DeepLab V3+ and the adaptive skin color detection method of this paper are more than 0.9, which is obviously better than the detection performance of other types of methods. The subjective evaluations of this paper's adaptive skin color detection method on the skin color of yellow, white, and black races are 7.14, 6.60, and 6.69, respectively, and the comprehensive evaluation score is 6.81, which is higher than other skin color enhancement methods.

2) When the parameter α is 0.4, both PESQ and STOI metrics are optimized under the NTCD-TIMIT validation set. In this paper, the PESQ, STOI, WSS, and RMSE index scores of IMCRA-OMLSA method in speech enhancement evaluation are 2.03, 72.36, 38.06, and 0.15, respectively, and the

IMCRA-OMLSA method achieves the best results in PESQ, STOI, and WSS. In terms of subjective evaluation, the MOS value of IMCRA-OMLSA method is 1.88, which is the highest value among all speech enhancement methods. IMCRA-OMLSA method is able to achieve the highest speech quality and intelligibility.

Funding

This work was supported by Shanghai Association of Higher Education (SAHE) 2024-2026 Annual planning research topic "Research on the Digital Transformation and Development of Film and Television Majors in Higher Vocational Education under the Background of Artificial Intelligence" (Project Number:2QYB24040).

References

[1] Li, F., & Wang, Z. (2021, July). Application of digital media interactive technology in post-production of film and television animation. In Journal of Physics: Conference Series (Vol. 1966, No. 1, p. 012039). IOP Publishing.

[2] Chen, Z., Zhu, Y., Xu, P., & Liu, R. (2019, November). A brief analysis of the application research of digital media technology--taking digital film and television as an example. In Journal of Physics: Conference Series (Vol. 1345, No. 5, p. 052058). IOP Publishing.

[3] Wang, Y. (2022, January). Film and television special effects production based on modern technology: from the perspective of statistical machine learning. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 833-836). IEEE.

[4] Zhang, W., & Wang, Y. (2020). Analyze the Application of Multimedia Technology in the Special Effects of Film and Television Animation--From the Animated Short Film of" Day after Day". International Journal of Social Science and Education Research, 2(11), 10-15.

[5] Zou, C. (2024). Research on Innovation in Film and Animation Art Design in the Digital Age. Research and Commentary on Humanities and Arts, 2(8).

[6] Peng, M. (2024). The application of digital media technology in the post-production of film and television animation. Media and Communication Research, 5(2), 129-134.

[7] Ying, Z. (2021, April). Interactive film and television animation special effects production techniques in visual design. In Journal of Physics: Conference Series (Vol. 1881, No. 2, p. 022020). IOP Publishing.

[8] Tu, Y. (2024, January). Research on the Application of Digital Media Art in Film and Television Works. In ADDT 2023: Proceedings of the 2nd International Conference on Art Design and Digital Technology, ADDT 2023, September 15–17, 2023, Xi'an, China (p. 13). European Alliance for Innovation.

[9] Xu, X., Yan, H., & Wang, X. (2022). Research on Postproduction of Film and Television Based on Computer Multimedia Technology. Scientific Programming, 2022(1), 4364834.

[10] Zhang, S. (2024). Analysis of the Impact of Multimedia Imaging Technology on Film Media. International Journal of e-Collaboration (IJeC), 20(1), 1-16.

[11] Li, K. (2024). Application of Communication Technology and Neural Network Technology in Film and Television Creativity and Post-Production. International Journal of Communication Networks and Information Security, 16(1), 228-240.

[12] Wang, D., & Lathakumari, K. R. (2023, November). Extraction Algorithm of Film and Television Special Effects Based on Artificial Intelligence Technology. In International Conference on Cognitive based Information Processing and Applications (pp. 83-92). Singapore: Springer Nature Singapore.

[13] Jiang, R., Wang, L., & Tsai, S. B. (2022). An empirical study on digital media technology in film and television animation design. Mathematical Problems in Engineering, 2022(1), 5905117.

[14] Feng, T., Wang, C., & Chen, C. (2022). Creative Effect of Film and Television Advertising Based on Digital Media Interactive Technology. Mobile Information Systems, 2022(1), 6231760.

[15] Lin, Y. (2021, February). On the application of computer combined with multimedia technology in post-production of film and television. In Journal of Physics: Conference Series (Vol. 1744, No. 4, p. 042073). IOP Publishing.

[16] Wang, Z. (2021, May). Analysis on the Application of Video Editing Skills Based on Image Mosaic in Film and Television Works. In 2021 2nd International Conference on Computers, Information Processing and Advanced Education (pp. 1446-1449).

[17] Wan, J. W. (2024). The Application Strategy of Intelligent Computer Multimedia Technology in Film and Television Post Production. Journal of Computers, 35(2), 183-197.

[18] Zeng, R. (2021, May). Research on the application of computer digital animation technology in film and television. In Journal of Physics: Conference Series (Vol. 1915, No. 3, p. 032047). IOP Publishing.

[19] Wang, S., Xu, Q., & Liu, Y. (2021, February). Research on the creation of film and TV works based on virtual reality technology. In Journal of Physics: Conference Series (Vol. 1744, No. 3, p. 032015). IOP Publishing.

[20] Du, N., & Yu, C. (2020, October). Research on special effects of film and television movies based on computer virtual production VR technology. In Proceedings of the 2020 International Conference on Computers, Information Processing and Advanced Education (pp. 115-120).

[21] Qi, D. (2024). Visual Communication Method of Multi-Frame Film and Television Special Effects Images Based on Deep Learning. International Journal of High Speed Electronics and Systems, 2540036.

[22] Zhang, Kun,Wang, Yedong,Li, Wenyuan,Li, Changlu & Lei, Zhichun. (2021). Real-time adaptive skin detection using skin color model updating unit in videos. Journal of Real-Time Image Processing(2), 1-13.

[23] Yen Chih Huang,Huang Pin Yuan & Yang Po Kai. (2020). An Intelligent Model for Facial Skin Colour Detection. International Journal of Optics1-8.

[24] Alireza Mahmoudi Nahavandi. (2018). Noise segmentation for improving performance of Wiener filter method in spectral reflectance estimation. Color Research & Application(3),341-348.

[25] Yanna Ma & Akinori Nishihara. (2014). A modified Wiener filtering method combined with wavelet thresholding multitaper spectrum for speech enhancement. EURASIP J. Audio, Speech and Music Processing(1),1-11.

[26] Mathivanan M. & Pandian S.Chenthur. (2012). Efficient Speech Enhancement Approach based on Minima Controlled Recursive Averaging through Modified Map Criterion using Hidden Markov Model. International Journal of Computer Applications(14),27-34.