Application and Simulation of Machine Vision and Computer Aided Design in Film and TV Art Creation

Xiang Li[1] and Lifen Jiang[2]

[1]School of Journalism and Communication, Hunan Mass Media Vocational and Technical College, Changsha 400100, China, xiang_l@cmu.ac.th
[2]School of Film and Television Arts, Hunan Mass Media Vocational and Technical College, Changsha 400100, China, jiang_lif@cmu.ac.th

Corresponding author: Lifen Jiang, jiang_lif@cmu.ac.th

Abstract. Traditional character animation production methods often rely on the experience of animators and manual adjustments, which is not only inefficient, but also difficult to realistically reproduce the performance details of actors. The expression capture algorithm based on machine vision (MV) has brought a revolutionary technological means to film & TV art creation. This article proposes an expression capture algorithm based on MV, aiming to use MV to automatically and accurately capture actors' expressions, and apply it to computer-aided design (CAD) of character animation to improve the fidelity of character performance. This algorithm is based on advanced computer vision technology and deep learning (DL) models, and achieves high-precision capture of actor expressions by detecting, tracking, and analyzing facial feature points. This algorithm exhibits excellent performance in processing low brightness facial images, not only in brightness adjustment, but also in presenting richer image details. Meanwhile, compared to the wavelet algorithm, the algorithm proposed in this article has higher classification accuracy, lower identification error, and higher operational efficiency. For film & TV art creation, this means higher production efficiency and less post production adjustment time.

Keywords: Film and TV Art Creation; Machine Vision; Computer Aided Design; Expression Capture
DOI: https://doi.org/10.14733/cadaps.2024.S14.77-91

1 INTRODUCTION

As an important part of film & TV works, the fidelity of character animation is directly related to the audience's viewing experience. Almeida et al. [1] explored how to use facial expression recognition-based techniques to recognize and classify emotions in movies. Facial expression recognition mainly relies on computer vision and deep learning technology. By training deep learning models to recognize key features of facial expressions, the model can learn facial expressions in different emotional states, such as happiness, sadness, anger, etc. In movie emotion recognition, facial expression recognition can help us capture subtle emotional changes, thereby more accurately understanding the emotional direction of the movie. Film emotion recognition technology based on facial expression recognition can be applied in many fields, such as film review, film and television production, audience behavior analysis, and so on. In movie reviews, this technology can help critics better understand and analyze the emotional expressions in movies. In film and television production, this technology can help directors better control the emotional direction of the film and improve its artistic effect. In audience behavior analysis, this technology can help analyze the audience's emotional response to the movie, thereby guiding the production and marketing of the movie. Traditional methods of character animation often rely on the experience and manual adjustment of animators, which is not only inefficient, but also difficult to truly restore the performance details of actors. With the increasing demand for low latency applications such as virtual reality (VR), augmented reality (AR), and real-time interaction, higher requirements have been put forward for efficient representation and rendering of 3D models. As the main manifestation of these applications, the adaptive representation of dynamic 3D meshes is of great significance for improving rendering efficiency and optimizing user experience. Arvanitis et al. [2] explored the principles, methods, and applications of this adaptive representation. Adaptive representation is a method of dynamically adjusting representation accuracy based on model complexity and scene characteristics. In dynamic 3D meshes, adaptive representation optimizes the number of vertices, patches, and level of detail of the mesh to achieve more efficient rendering and smoother interactive experience. The core principle is to dynamically simplify the grid based on model features and scene information, while maintaining visual accuracy. In real-time simulation, adaptive representation can achieve real-time rendering of large-scale models, providing more accurate and intuitive data visualization for real-time decision-making.

Therefore, how to accurately and efficiently capture actors' expressions and apply them to character animation has become an urgent problem in the field of film & TV production. In today's digital age, machine vision technology has become an important pillar in the field of artificial intelligence. It is widely used in various scenarios, such as autonomous driving, robot navigation, product quality inspection, etc. In order to use machine vision technology more effectively, Carter and Carter [3] analyzed a method that can accurately describe and understand images. This is where critical image synthesis (CIS) methods come into play. Critical image synthesis is a new method of image understanding that helps machines understand and describe the content of images by generating a series of synthesized images based on image content. This method combines computer vision, machine learning, and image processing technologies to extract key information from raw images and convert it into machine readable descriptions. Machine vision and critical image synthesis are technically complementary. Machine vision can help us extract useful information from images, while critical image synthesis can transform this information into machine readable descriptions. This combination can enable machines to more accurately understand image content and make more accurate decisions in various application scenarios. As an important branch of artificial intelligence (AI), MV aims to simulate and realize human visual perception and understanding through computer vision algorithms and models. In today's digital age, movies have become an indispensable part of people's daily lives. The movie rating system is a key tool to help audiences find movies that they may be interested in. However, most existing rating systems are based on human expert review or user voting, which often has subjectivity and one-sidedness. To address this issue, García et al. [4] proposed a movie metaphor rating optimizer using machine learning and evolutionary algorithms. Using natural language processing techniques to extract keywords and phrases from the text description of movies. At the same time, we also extract visual features of movies through image analysis technology. Train preprocessed data using various machine learning algorithms (such as random forests, neural networks, etc.) to learn the mapping relationship between movie features and ratings. In order to further improve the accuracy and fairness of ratings, we use evolutionary algorithms (such as genetic algorithms, particle swarm optimization algorithms, etc.) to optimize the parameters of machine learning models. In the field of film & TV production, MV is expected to change the traditional animation production method and realize more efficient and accurate expression capture and application.

The computer-aided art design and production technology based on video streaming mainly processes and analyzes video data through computers. Then, based on this information, carry out artistic design and production. The main process of this technology includes four parts: video stream acquisition, feature extraction, artistic rendering, and interactive design. Video stream acquisition mainly involves obtaining raw video data through cameras or input devices. Feature extraction is the extraction of useful features from videos, such as color, shape, motion, etc. Artistic rendering is the process of converting videos into artwork based on extracted features; Interactive design refers to the interactive editing and modification of artistic works by users. In the field of film and television production, this technology can help directors better control the color, rhythm, and emotional elements of a film, and improve the artistic effect of the film. In the field of game design, this technology can be used to create more realistic game scenes and characters, improving the playability of games. In the field of virtual reality, this technology can be used to create more realistic virtual scenes and characters, improving the immersion of virtual reality. In the field of education and entertainment, this technology can be used to create more vivid and interesting educational games and interactive animations, enhancing learners' interest and motivation in learning [5]. With the rapid development of the digital era, image subtitles have become one of the important ways for people to obtain information. However, for those with visual impairment or reading disabilities, obtaining information from images is a difficult task. To address this issue, Jaiswal [6] developed an automated image subtitle generation system using cognitive Internet of Things (IoT) and machine learning (ML) methods. Cognitive Internet of Things is an intelligent network that can learn and adjust itself, which can understand and predict the behavior and state of objects through data analysis. In image subtitle generation, cognitive Internet of Things can use deep learning algorithms to identify objects and scenes in images, thereby determining the theme and content of the image. Cognitive IoT can also use technologies such as optical character recognition (OCR) to recognize text information in images and convert it into editable text formats. Cognitive Internet of Things can generate more accurate and natural subtitles by comprehensively analyzing image and text information, understanding the context and meaning of images.

Jin and Yang [7] analyzed the application of computer-aided design software in environmental art design teaching. Computer assisted design software can assist environmental art designers in accurately constructing three-dimensional models and performing precise composition and layout. This allows designers to repeatedly modify and optimize in a virtual environment, improving the efficiency and accuracy of the design. Through computer-aided design software, environmental art designers can easily select and apply various materials and textures. This allows designers to consider the texture, color, lighting effects, etc. of materials during the design process, improving the visual effect and realism of the design. Computer aided design software is usually equipped with powerful rendering and special effects functions, which can simulate real natural lighting effects, shadow effects, dynamic effects, etc. This allows designers to create more visually impactful and realistic environmental art works. The application of computer-aided design software can combine the theory and practice of environmental art design, improving students' understanding and application ability of theoretical knowledge. Through practical operation, students can better grasp the skills and methods of environmental art design. The combination of 3D reality technology and CAD has had a profound impact on animation design. Firstly, this combination improves the efficiency and accuracy of design, allowing designers to quickly implement design concepts. Secondly, this combination provides designers with more creative space and possibilities, allowing them to create more visually impactful and artistic animation shapes. Finally, this combination also makes the animation design more realistic and natural, improving the audience's viewing experience. Jing and Song [8] explored the application and impact of this combination in animation design. By utilizing 3D reality technology, designers can more accurately construct the body proportions, muscle structures, and clothing design of characters, making their images more realistic and three-dimensional. At the same time, CAD is used to adjust and optimize details such as texture, color, materials, etc., making character design more refined. 3D reality technology can simulate real natural environments, buildings, and scenes, making animation scenes more vivid and three-dimensional. CAD can assist designers in precise layout and design, enhancing the visual effect and artistic sense of the scene.

In this article, an expression capture algorithm based on MV is proposed, which aims to capture the expression of actors automatically and accurately by using MV and apply it to character animation, thus improving the fidelity of character performance. Based on advanced computer vision technology and DL model, the algorithm can capture the actor's expression with high precision by detecting, tracking and analyzing facial feature points. Specifically, firstly, the image processing technology is used to preprocess the input actor's face image and extract the key feature points. Then, through the designed DL model, learn and model the mapping relationship between the actor's expression and feature points. Finally, the trained model is applied to real-time character animation to realize expression migration and mapping. In this way, the actor's expression can be quickly and accurately captured in the role animation, making the performance of the role more vivid and realistic. Through this technology, I hope to provide a powerful and efficient tool for film & TV producers, simplify the animation production process, reduce the cost, and enhance the realism and texture of animation works. Compared with the traditional research of film & TV art creation, this article has the following innovations:

(1) Compared with the traditional methods of manual animation or sensor-based expression capture, this method can capture actors' expressions automatically and in real time by using MV.

(2) In this study, DL model is used to learn and model the complex mapping relationship between actors' expressions and facial feature points. This method can capture and reproduce richer and more subtle expression details and improve the fidelity of character animation.

(3) The algorithm can handle the task of expression capture under different actors, different expressions and different lighting conditions, showing high flexibility and adaptability. This innovative method improves the efficiency of animation production and reduces the production cost.

This article first introduces the importance of expression capture in film and TV art creation, and points out the limitations of traditional methods. Then, an expression capture algorithm based on technology (MV) is proposed, and the principle and flow of the algorithm are described. Then, by comparing the results with wavelet algorithm, the advantages of this algorithm in expression classification accuracy, identification error and response time are demonstrated. Finally, the full text is summarized, and the significance of this algorithm for film and TV art creation is pointed out.

2 THEORETICAL BASIS

In today's data-driven era, the combination of geographic information science (GIS) and artificial intelligence (AI) has become increasingly important. GeoAI, also known as geographic artificial intelligence, is an emerging field that applies AI technology to processing and analyzing large amounts of geographic data. Through technologies such as deep learning and machine learning, GeoAI has made significant progress in image recognition, spatial analysis, and predictive modeling. Li and Hsu [9] discussed the latest applications and developments of GeoAI in geography. Large scale image analysis is an important application field of GeoAI. By utilizing artificial intelligence technology, we can efficiently and accurately analyze satellite images, remote sensing images, and GIS data. In addition, generative adversarial networks (GANs) have also demonstrated potential in image generation and super-resolution, which is of great significance for the analysis and utilization of remote sensing images. With the rapid development of technology, the application of computer-aided design (CAD) software in the field of contemporary art is becoming increasingly widespread. This technology, with its unique advantages, provides artists with infinite creative space and possibilities. However, how to effectively combine computer-aided design with contemporary art teaching, with creativity as the center, and cultivate students' artistic design and innovation abilities is still a problem worth exploring. Liu and Yang [10] explored the computer-assisted teaching model of contemporary art centered on creativity and its application. Students use computer-aided design software to conduct specific design practices based on the set theme and creative thinking. Teachers should provide necessary guidance and feedback to students during this process. After completing the design, organize students to evaluate and reflect. The evaluation not only focuses on the final work, but also considers students' creative thinking and innovation process. Reflection helps students summarize their experiences and further improve their design abilities. Provide students with the opportunity to be exposed to real design problems, cultivate their ability to solve practical problems, and develop innovative thinking.

In the era of big data, how to effectively process and analyze a large amount of information is an important challenge. Film reviews are an example of this, as they contain various emotions and opinions about movies, which are of great value for movie recommendation systems, box office predictions, and audience feedback. However, traditional text analysis methods cannot handle this unstructured information. To this end, we propose a machine learning method based on facial image recognition and emotion detection to extract emotional information from movie reviews. Mazhar et al. [11] conducted tests on multiple datasets and the results showed that this method can effectively extract emotional information from movie reviews with high accuracy. In addition, we also found that the results of facial image recognition and emotion detection have a significant impact on the final classification results. This means that in order to improve accuracy, we need to maintain high accuracy in these two steps. With the continuous development of computer vision technology, image labeling technology has also been widely applied. Image labeling technology is a method that combines image classification, recognition, and search techniques to achieve rapid image retrieval and management by labeling and classifying images. Sager et al. [12] conducted an investigation and analysis of image labeling technology in computer vision applications. Image labeling technology is a method that combines image feature extraction and classification techniques to achieve rapid image retrieval and management by labeling and classifying images. On e-commerce platforms, rapid retrieval and management of products can be achieved by labeling and classifying product images. In an intelligent monitoring system, rapid retrieval and management of targets can be achieved by labeling and classifying the targets in the surveillance video. In the field of cultural heritage protection, rapid retrieval and management of cultural relics can be achieved by labeling and classifying cultural relic images. Distinguishing and recognizing voluntary and involuntary facial expressions is of great significance for interpersonal communication, psychological research, and machine learning and artificial intelligence applications. By deeply understanding this difference, we can more accurately interpret others' emotional states and intentions, thus enabling more effective communication and response. Meanwhile, with the continuous development of machine learning and artificial intelligence technology, we have reason to believe that more innovative applications in the future will rely on this ability to distinguish and recognize. Swaminathan and Arock [13] explored how to distinguish and recognize voluntary and involuntary facial expressions, and the importance of this distinction in interpersonal communication and psychological research. There are differences in muscle movement patterns between voluntary and involuntary facial expressions. For example, when people intentionally laugh, the muscle movement pattern is different from the movement pattern when they are truly happy. This difference can be identified through machine learning and artificial intelligence algorithms. Micro expressions are brief, involuntary facial expressions that typically last for no more than 500 milliseconds. They are often used as indicators of emotional state in psychological research.

The symmetry of animation teaching organizational forms and social networks in multimedia assisted art teaching has brought new perspectives and opportunities to art education. Xie [14] explored the principles, methods, and applications of this symmetry. The organizational form of animation teaching, with its unique flexibility, enables teaching content to be presented to students in a more attractive way. Through animation, teachers can more intuitively demonstrate art skills and creative processes, while also stimulating students' interest and participation. Social networks provide an interactive platform for art teaching. Here, students can share their works, discuss the creative process, and learn and gain inspiration from others' works. This interactivity not only enhances students' learning experience, but also promotes their social development. The symmetry of the organizational form and social network of animation teaching lies in their multimedia-based approach, which enhances the interactivity and experience of art teaching. Animation can provide vivid visual effects, while social networks provide real-time communication and feedback. The combination of the two makes art teaching more diverse, diverse, and effective. Machine learning has also emerged in the field of art, providing artists with new creative tools. However, although machine generated artworks have made technological breakthroughs, people's acceptance of them varies depending on cultural background. Therefore, Xu et al. [15] aimed to explore people's reactions to machine generated artworks in different cultural backgrounds. Through analyzing the data, we have identified some interesting trends. Firstly, the user's response to machine generated artworks is generally positive. However, when considering cultural factors, we found some differences. In some cultures, users are more inclined to accept machine generated artworks as innovative and interesting. In other cultures, users are more inclined to believe that these artworks lack human emotions and creativity. These results pose new challenges to our understanding. On the one hand, it indicates that people's acceptance of machine generated artworks is to some extent malleable and can be changed through education and publicity. On the other hand, this also reminds us that when promoting machine generated artworks, we need to consider that users from different cultural backgrounds may have different reactions.

With the intensification of global market competition, understanding and meeting customer needs has become more important than ever for enterprises. The product ecosystem, as a comprehensive perspective that integrates product design, manufacturing, recycling, and reuse, has attracted widespread attention in the industry. In this process, how to effectively conduct customer demand analysis in order to better design and launch products that meet market demand is a challenge that every enterprise needs to face. Zhou et al. [16] used an e-commerce company as an example to analyze customers' purchase history and feedback comments using machine learning algorithms. Through clustering algorithms, they identified specific customer groups, such as young mothers, technology enthusiasts, health food enthusiasts, etc., and each group had significantly different needs for the product. By using decision tree algorithms, they identified some key factors that affect customers' purchasing decisions, such as price, brand, product functionality, etc. Finally, based on these analysis results, they made adjustments to product design and marketing strategies, thereby increasing sales. In the field of biomedicine, accurate molecular diagnosis has a decisive impact on disease prevention, early detection, and treatment. Machine learning enhanced surface enhanced spectroscopy (ML-SES) is a new technology that optimizes spectral data analysis and signal enhancement through machine learning algorithms. Zhou et al. [17] utilized machine learning algorithms to learn a large amount of spectral data to establish more accurate prediction models and achieve high-precision detection of biomolecules. ML-SES has been widely used in various molecular diagnostic scenarios, including detection of disease markers, drug response monitoring, and gene mutation identification. For example, by using ML-SES, researchers can detect cancer markers with high sensitivity and specificity, thereby achieving early diagnosis of cancer. In addition, ML-SES is also used to monitor patients' drug response to optimize treatment plans. By applying machine learning algorithms, we can extract more useful information from spectral data, thereby achieving more accurate analysis and recognition of biological molecules.

3 EXPRESSION CAPTURE ALGORITHM BASED ON MV

Film and TV artistic creation is a process of telling stories, expressing emotions and shaping roles through various means such as images and audio. In the creation of film and TV art, character animation is a vital link, which can give the role vitality and let the audience feel the plot of the story and the emotion of the role more deeply. Therefore, how to accurately and vividly present the expression of the role has become an important topic in the creation of film and TV art. The introduction of MV provides a new solution and realization method for this topic. MV is an important branch of AI and computer vision, and its goal is to make computers perceive and understand images and videos like humans. MV plays a core role in capturing the expression of movie characters. It can extract the key feature points of the face from the original image data by means of image processing, feature detection and target detection, and accurately identify and analyze the facial expression. These technical means provide reliable basis and support for expression capture. DL is a branch of machine learning, which learns to extract useful feature representations from raw data by constructing deep neural network models, and further completes tasks such as classification and regression. In the film and TV character expression capture, DL can be used for the task of expression identification. By training the deep neural network model, we can learn the complex mapping relationship between expressions and facial features, and realize the accurate classification and identification of facial expressions. The introduction of DL greatly improves the accuracy of expression capture.

CAD is a method that uses computer technology and graphics principles to assist design activities. Through CAD software, the facial structure of the character can be finely modeled, and combined with MV, the captured expression can be migrated to the character model to realize the generation of character animation. Therefore, CAD provides a visual and operable tool for capturing movie and television character expressions, and improves the quality of character animation. With the rapid growth of film and TV technology and the increasing improvement of the audience's quality of film and TV works, the expression capture and expression in role animation has become the core link of film and TV art creation. Traditional expression capture methods, such as image processing-based technology and sensor-based technology, meet the requirements of expression capture to some extent, but their performance is limited when dealing with complex situations, low-quality images or scenes with high real-time requirements. In this article, an expression capture algorithm based on MV is proposed, which combines DL technology to achieve accurate and efficient expression capture and improve the fidelity of film & TV character animation. Firstly, the MV is used to preprocess the input actor's facial image and extract the key feature points. Then, with the help of DL technology, a neural network model is constructed to learn and model the mapping relationship between actor expressions and feature points. By training and optimizing the model, the actor's expression can be accurately captured and classified. Finally, the captured expressions are applied to the character model to generate realistic character animation. The operation flow of expression capture in this article is shown in Figure 1.

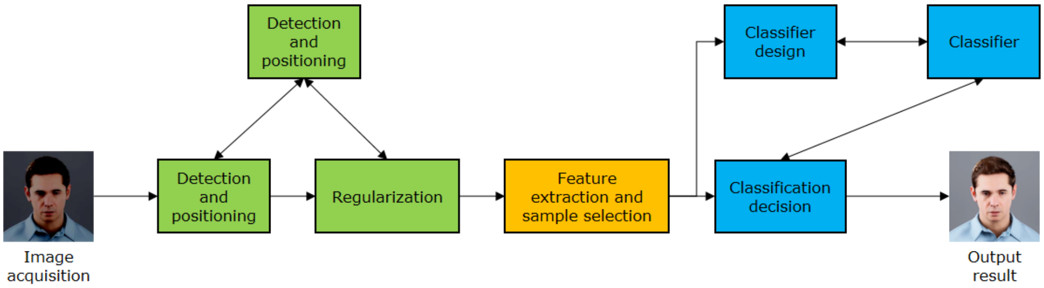


Figure 1: Expression capture operation process.

The input actor's face image is preprocessed, including denoising, illumination normalization, face alignment and so on, in order to improve the stability of subsequent feature detection. Using computer vision technology, the key feature points of face, such as the contour and shape information of eyes, eyebrows and mouth, are extracted from the preprocessed image. Based on DL technology, a neural network model of expression classification is constructed. The model receives facial features as input and outputs the probability distribution of expression categories.

The input of the model is a preprocessed face image. The preprocessing steps usually include face detection, face alignment and normalization to ensure the consistency and standardization of the input image. By stacking multiple convolution layers, the model can abstractly extract the low-level to high-level features of the image layer by layer. After the convolution layer, the activation function is usually used for nonlinear transformation to increase the expressive ability of the model. Pool layer is used to reduce the dimension of features and introduce some spatial invariance. After several rounds of convolution and pooling operations, the model uses the fully connected layer to integrate and classify the extracted features. The fully connected layer transforms a two-dimensional feature map into a one-dimensional feature vector, and maps features with labels through connection weights. For the task of expression classification, the output layer usually uses the softmax function to map features to the probability distribution of each expression category. The DL model of expression feature detection is shown in Figure 2.

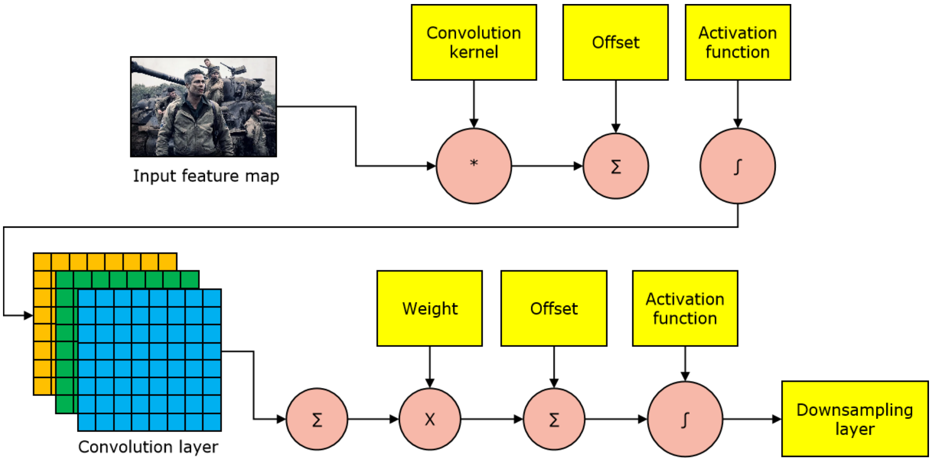


Figure 2: DL model of expression feature detection.

On the basis of facial expression classification, further utilizing DL technology, a regression model is trained to convert the classified facial expression categories into continuous facial expression parameters. These parameters can describe the intensity and details of expressions, providing richer information for the subsequent generation of character animations. Apply the captured expression parameters to the character model to generate character animations with corresponding expressions. This can be achieved by adjusting the shape, texture, and animation curves of the character model. The final generated character animation will have realistic expressions corresponding to the actor's performance. During the training stage, the model optimizes the parameters of the convolutional kernel using a backpropagation algorithm to minimize the loss between the predicted results and the actual labels. By iteratively training a large number of training samples, the model can learn effective expression of facial expression features and achieve accurate classification of facial expression images. Data augmentation technology increases the diversity of data by randomly transforming the original image. The transfer learning strategy utilizes pre trained models on large-scale datasets to achieve adaptation to the tasks in this article through fine-tuning, accelerating model convergence and improving performance.

Assume that the input size of l layer is $C^l \times H^l \times W^l$ tensor, C^l is the quantity of input channels of l layer. Then, corresponding to the convolution layer with C^{l+1} hidden neurons, the output of the corresponding position is as follows:

$$y_{d,i^l,j^l} = \sigma\left(\sum_{c^l=0}^{C^l}\sum_{i=0}^{h^l}\sum_{j=0}^{w^l} p_{d,c^l,i,j} \times x_{c^l,i^l+i,j^l+j} + b_d\right) \tag{1}$$

Where d is that numb of neurons in the l lay; i^l and j^l represent location information. There are the following constraints:

$$0 \le i^l \le H^l - h^l + 1 \tag{2}$$

$$0 \le j^l \le W^l - w^l + 1 \tag{3}$$

Where p is the convolution kernel parameter; b is the bias parameter in convolution; $\sigma \cdot$ is the activation function.

When an actor performs, the expression capture algorithm will capture real-time changes in the actor's expression and convert them into expression parameters. These parameters describe the dynamic changes of various facial feature points, such as eye opening and closing, eyebrow lifting, etc. CAD software receives these parameters and drives the corresponding feature points of the character model to undergo animation deformation, thereby transferring the actor's expression to the character model in real-time. After facial expressions are transferred, CAD software typically provides animation editing tools that allow animators to further adjust and optimize character animations.

4 APPLICATION OF EXPRESSION CAPTURE ALGORITHM AND CAD IN FILM AND TV ART CREATION

Through the MV based expression capture algorithm proposed in this article, film and TV producers can accurately capture actors' expressions and apply them to character animation. This enables character animation to present more vivid and realistic facial expressions, improving the fidelity of character performance. During the viewing process, the audience can more deeply feel the emotions and inner world of the characters, enhancing the viewing experience. The traditional process of character animation production usually relies on the experience of animators and manual adjustments, which is not only inefficient, but also difficult to restore the performance details of actors. The application of expression capture algorithms can automatically and accurately capture actors' expressions and map them to character models, greatly improving the efficiency of animation production. Animators can focus more on creativity and storytelling, enhancing the artistic value of film and TV works.

By using specific transfer algorithms, such as neural network-based style transfer algorithms, the extracted style features are applied to content images. This process requires introducing the color and texture of the style image as much as possible while preserving the basic structure of the content image. By iteratively optimizing the process, continuously adjusting the generated images to maintain consistency in content and approach the target style as closely as possible. In this process, a loss function is usually used to measure the gap between the generated image and the target style and content image, and optimization algorithms are used to minimize this loss. After optimization, the resulting image may require some post-processing steps, such as color correction, contrast adjustment, etc., to obtain the final style transfer result. The image style migration process generated by film & TV animation is shown in Figure 3.

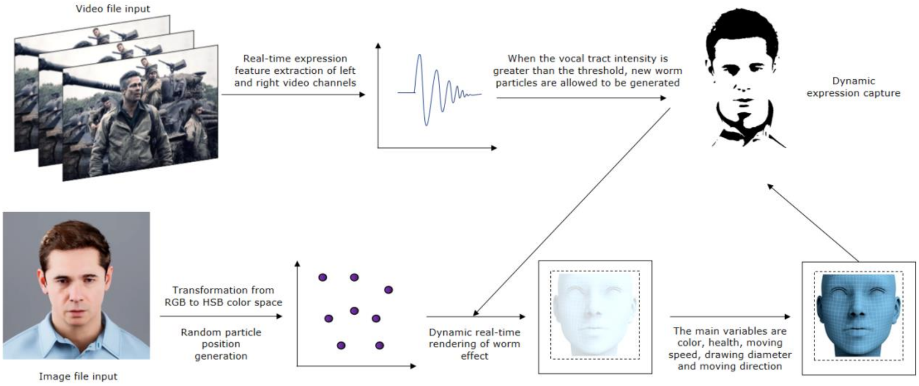


Figure 3: Operation process of image style migration generated by film and TV animation.

The expression capture algorithm can record and analyze the performance details of actors, and provide a rich expression library and action library for character animation. Film and TV producers can flexibly choose and combine different expressions and actions according to the needs of story and character, and create more diversified and personalized role performance. This has brought a broader creative space and unlimited possibilities for film and TV art creation. Using CAD, film & TV producers can carry out fine role design. Through professional 3D modeling software, highly realistic character models can be created, and the facial features and expression details of actors can be accurately restored. Suppose there are l samples randomly and independently extracted from the unknown probability distribution function to form a training sample set:

$$x_i, y_i \ , i = 1,2,3,\ldots,l \quad x_i \in R^d \tag{4}$$

Among them, $y_i \in \ +1,-1$ is the category identification of two types of samples. The goal of learning is to construct the following optimization problems:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \tag{5}$$

$$s.t. \quad y_i \ w^T x_i + b \ \geq 1 - \xi_i \tag{6}$$

$$\xi_i \geq 0, i = 1,2,3,\ldots,l \tag{7}$$

Mapping the low-frequency component image $X_i \middle| i = 1,\ldots,M$ of each known face to the subspace spanned by the feature face to obtain the m-dimensional vector Y_k:

$$\begin{cases} Y_k = W^T \ X_k - \mu \\ k = 1,\ldots\ldots,N_c \end{cases} \tag{8}$$

Where N_c is the known number of people, $W = \left[\omega_1, \omega_2, \ldots, \omega_d\right]$ is the face projection space, and μ is the average vector of all training samples X_i:

$$\mu = \frac{1}{M} \sum_{i=1}^{N} X_i \tag{9}$$

The threshold value θ_c of the distance between faces is defined as:

$$\begin{cases} \theta_c = \frac{1}{2} \max_{j,k} \left\| Y_j - Y_k \right\| \\ j,k = 1,\ldots,N_c \end{cases} \tag{10}$$

Mapping the low-frequency component image X to be identified to a feature space to obtain a vector Y:

$$Y = W^T \ X - \mu \tag{11}$$

The distance between Y and each face set is defined as:

$$Y_k = W^T \ X_k - \mu \tag{12}$$

In order to distinguish human face from non-human face, it is also needed to calculate the distance ε between the low-frequency component image of the original image X and the image X_f reconstructed from the feature space:

$$\varepsilon = \left\| X - X_f \right\|^2 \tag{13}$$

Among them:

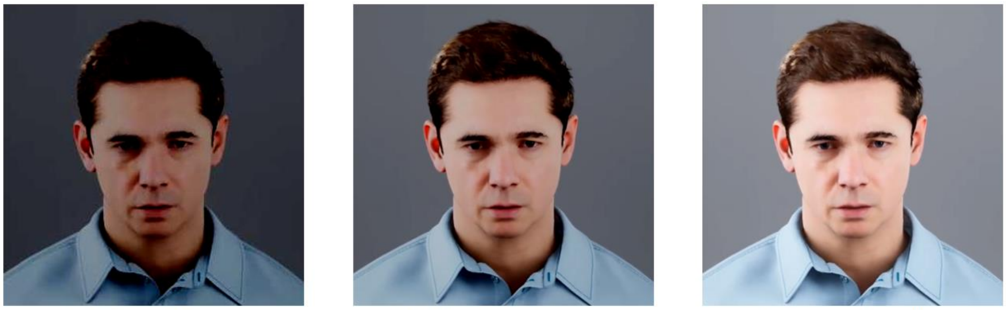
$$X_f = UY + \mu \tag{14}$$

In addition to character design, CAD can also be applied to the design of scenes and props. Film and TV producers can use CAD software to build exquisite scenes and props models, and combine film and TV special effects technology to create fantastic and shocking visual effects. Through the 3D model generated by CAD software, film & TV producers can carry out preliminary work such as camera movement and lighting adjustment in the virtual environment, and realize the preview and trial screening of film and TV works.

5 RESULTS AND ANALYSIS

The experiment aims to verify the advantages of the expression capture algorithm based on MV proposed in this article in film and TV art creation. Compared with the traditional wavelet algorithm, the performance of this algorithm in expression classification accuracy, identification error and response time is evaluated. The experiment uses an open facial expression data set, which contains facial image samples with different brightness, angle and expression changes. The face image is preprocessed, including face detection, alignment, normalization and other operations to ensure the consistency of data.

The algorithm in this article has higher processing effect when dealing with low brightness face images, and can better preserve and enhance the details and brightness of the images. This is of great significance to the expression capture of movie characters, because in the actual shooting stage, due to the influence of light, environment and other factors, the brightness of face images may be reduced, thus affecting the accuracy of expression capture. The image processed by the algorithm in this article not only has improved brightness, but also has clearer details, and the overall image quality has been significantly improved. This provides higher quality image resources for subsequent film and TV production, which is conducive to improving the visual effect and impression experience of film and TV works. Figure 4 shows the original face image and the processing results.



(a) Original image (b) Wavelet algorithm (c) Paper algorithm

Figure 4: Original face image and processing results.

This advantage may come from the outstanding performance of the algorithm in feature detection and robustness, which directly affects the value of the algorithm in practical application. The expression classification accuracy of different algorithms is shown in Figure 5.

In this article, the algorithm adopts a deeper neural network structure and more advanced feature detection technology, so as to extract key features related to expressions from face images more effectively. These features are very important for expression classification, so more effective feature detection directly improves the classification accuracy. Although wavelet transform can extract the frequency domain features of images, it may not be able to extract high-level and abstract features like DL for the complex task of expression classification. Wavelet algorithm is more dependent on image quality, such as brightness and contrast. When the image quality declines, such as low brightness, the classification accuracy of wavelet algorithm may be greatly affected.

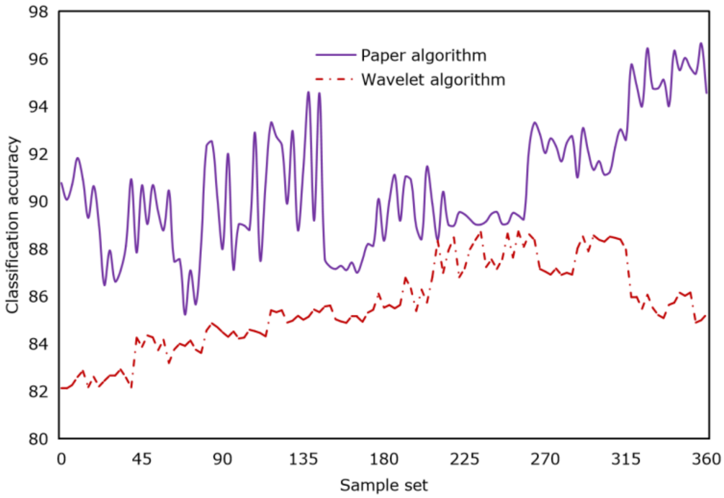


Figure 5: Expression classification accuracy of different algorithms.

Figure 6 shows the comparison of the expression identification error between this algorithm and wavelet algorithm. Figure 7 shows the comparison of response time of different algorithms, and it can be seen that the response time of this algorithm is obviously lower than that of wavelet algorithm.

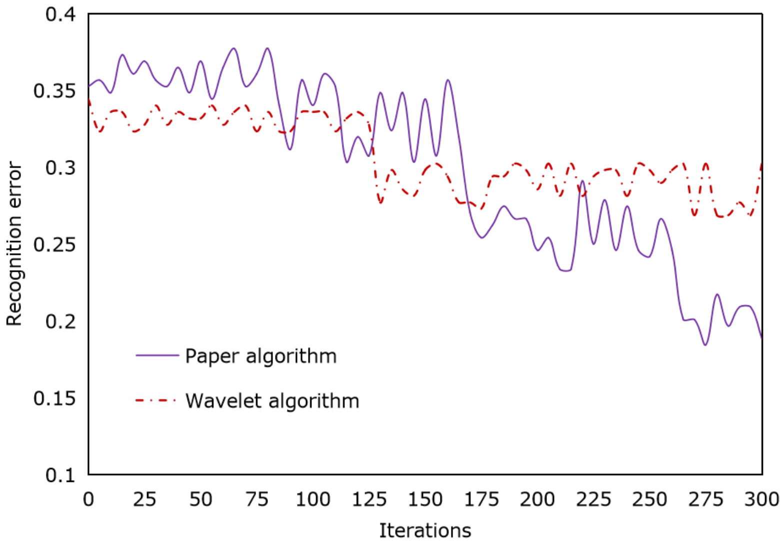


Figure 6: Identification errors of different algorithms.

The algorithm in this article may have stronger ability in feature detection, and can capture the key features related to expressions more accurately, thus reducing the identification error. In contrast, wavelet algorithm may be relatively simple or not precise in feature detection, resulting in relatively high identification error. The algorithm in this article makes more effective use of computing resources, for example, by parallel computing or optimizing the operation of computing matrix, which speeds up the processing process.

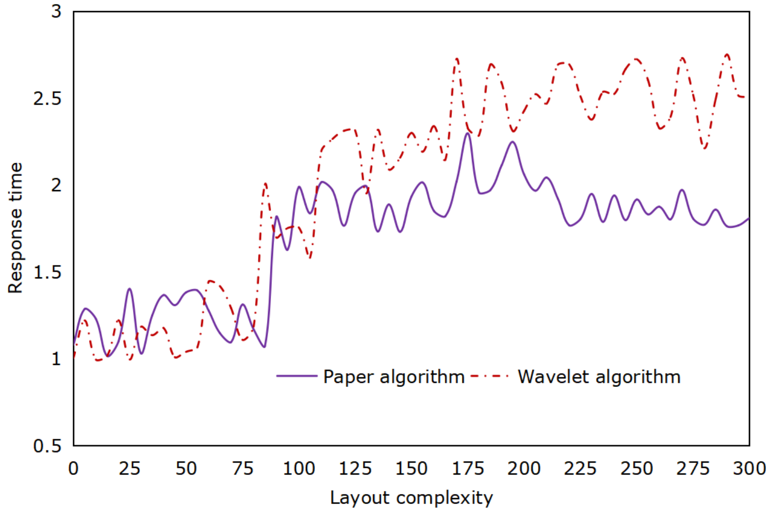


Figure 7: Response time of different algorithms.

This optimization can significantly reduce the running time of the algorithm, thus reducing the response time. The reduction of identification error indicates that the accuracy of expression identification has been improved, and the reduction of response time means that the processing efficiency of the algorithm has also been optimized. These advantages make the algorithm in this article more competitive and practical in practical application.

In film & TV art creation, the expression and emotional transmission of the role are the core of the soul of the work. Through the summary of the above results, we can see that the algorithm in this article has shown remarkable advantages in expression capture, classification and identification. Accurate expression capture can ensure that the emotional expression in character animation is more real and delicate, and enhance the emotional resonance of the audience. The fast response time and high-precision classification of the algorithm can significantly improve the efficiency of film & TV production and reduce the time and cost of post-modification and adjustment. Because the algorithm has a good processing effect on low-brightness images and expressions in complex situations, it provides a wider material selection and creative space for film & TV creation.

6 CONCLUSIONS

Character animation, as an important component of film & TV works, its fidelity directly affects the audience's viewing experience. In the field of film & TV production, MV is expected to change traditional animation production methods and achieve more efficient and accurate expression capture and application. This article proposes an expression capture algorithm based on MV, aiming to use MV to automatically and accurately capture actors' expressions, and apply it to character animation to improve the fidelity of character performance. From the perspective of facial expression capture, the algorithm presented in this article exhibits excellent performance in processing low brightness facial images. It not only performs well in brightness adjustment, but also presents richer image details. In terms of accuracy in facial expression classification and identification, it has been proven that our algorithm has higher classification accuracy and lower identification error compared to wavelet algorithm. From a practical application perspective, the algorithm proposed in this article also significantly reduces response time, which means higher production efficiency and less post adjustment time for film & TV art creation.

Overall, this algorithm brings higher authenticity and accuracy to film & TV art creation, providing strong technical support for film & TV production personnel. In future research and applications, this technology based on MV and DL is expected to play a greater role in fields such as film & TV, animation, and games, helping art creation reach new heights.

7 ACKNOWLEDGEMENT

This work was supported by Hunan Federation of Social Sciences Fund Project "Research on the Path of 'Quality Control at the Source' for New Mainstream Films and Televisions in Splendid Xiaoxiang", project number XSP22YBC209.

Xiang Li, https://orcid.org/0009-0000-7073-3646
Lifen Jiang, https://orcid.org/0009-0001-3444-5126

REFERENCES

[1] Almeida, J.; Vilaça, L.; Teixeira, I.-N.; Viana, P.: Emotion identification in movies through facial expression recognition, Applied Sciences, 11(15), 2021, 6827. https://doi.org/10.3390/app11156827

[2] Arvanitis, G.; Lalos, A.-S.; Moustakas, K.: Adaptive representation of dynamic 3D meshes for low-latency applications, Computer Aided Geometric Design, 73(8), 2019, 70-85. https://doi.org/10.1016/j.cagd.2019.07.005

[3] Carter, R.; Carter, R.-A.: Machine visions: Mapping depictions of machine vision through critical image synthesis, Open Library of Humanities, 9(2), 2023, 1-33. https://doi.org/10.16995/olh.10077

[4] García, O.-R.-H.; García, S.-P.; Merelo, G.-J.-J.: StarTroper, a film trope rating optimizer using machine learning and evolutionary algorithms, Expert Systems, 37(6), 2020, e12525. https://doi.org/10.1111/exsy.12525

[5] Guo, S.; Li, X.: Computer aided art design and production based on video stream, Computer-Aided Design and Applications, 2020, 18(S3), 2020, 70-81. https://doi.org/10.14733/cadaps.2021.S3.70-81

[6] Jaiswal, T.: Image captioning through cognitive IOT and machine-learning approaches, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(9), 2021, 333-351. https://doi.org/10.17762/turcomat.v12i9.3077

[7] Jin, H.; Yang, J.: Using computer-aided design software in teaching environmental art design, Computer-Aided Design and Applications, 19(S1), 2021, 173-183. https://doi.org/10.14733/cadaps.2022.S1.173-183

[8] Jing, Y.; Song, Y.: Application of 3D reality technology combined with CAD in animation modeling design, Computer-Aided Design and Applications, 18(S3), 2020, 164-175. https://doi.org/10.14733/cadaps.2021.S3.164-175

[9] Li, W.; Hsu, C.-Y.: GeoAI for large-scale image analysis and machine vision: Recent progress of artificial intelligence in geography, ISPRS International Journal of Geo-Information, 11(7), 2022, 385. https://doi.org/10.3390/ijgi11070385

[10] Liu, F.; Yang, K.: Exploration on the teaching mode of contemporary art computer aided design centered on creativity, Computer-Aided Design and Applications, 19(S1), 2021, 105-116. https://doi.org/10.14733/cadaps.2022.S1.105-116

[11] Mazhar, T.; Malik, M.-A.; Nadeem, M.-A.; Mohsan, S.-A.-H.; Haq, I.; Karim, F.-K.; Mostafa, S.-M.: Movie reviews classification through facial image recognition and emotion detection using machine learning methods, Symmetry, 14(12), 2022, 2607. https://doi.org/10.3390/sym14122607

[12] Sager, C.; Janiesch, C.; Zschech, P.: A survey of image labelling for computer vision applications, Journal of Business Analytics, 4(2), 2021, 91-110. https://doi.org/10.1080/2573234X.2021.1908861

[13] Swaminathan, A.; Arock, M.: VIVAFER - voluntary and involuntary actions-based facial expression and recognition, International Journal of Business Innovation and Research, 26(4), 2021, 467-487. https://doi.org/10.1504/IJBIR.2021.119815
[14] Xie, Z.: Symmetry for multimedia-aided art teaching based on the form of animation teaching organization and social network, Symmetry, 12(4), 2020, 671. https://doi.org/10.3390/sym12040671
[15] Xu, K.; Liu, F.; Mou, Y.; Wu, Y.; Zeng, J.; Schäfer, M.-S.: Using machine learning to learn machines: A cross-cultural study of users' responses to machine-generated artworks, Journal of Broadcasting & Electronic Media, 64(4), 2020, 566-591. https://doi.org/10.1080/08838151.2020.1835136
[16] Zhou, F.; Ayoub, J.; Xu, Q.; Jessie, Y.-X.: A machine learning approach to customer needs analysis for product ecosystems, Journal of Mechanical Design, 142(1), 2020, 011101. https://doi.org/10.1115/1.4044435
[17] Zhou, H.; Xu, L.; Ren, Z.; Zhu, J.; Lee, C.: Machine learning-augmented surface-enhanced spectroscopy toward next-generation molecular diagnostics, Nanoscale Advances, 5(3), 2023, 538-570. https://doi.org/10.1039/d2na00608a