www.cad-journal.net
Computer-Aided Design AND APPLICATIONS

Reconstruction of Film and TV Scenes Based on Computer-Aided Design and Machine Vision

Qingyang Li[1] and Kai Wang[2]

[1]College of Arts, Cheongju University, Cheongju 28503, South Korea,
lqy6137@126.com
[2]School of Literature and Journalism, Yantai University, Yantai 264005, China,
404953kai@ytu.edu.cn

Corresponding author: Kai Wang, 404953kai@ytu.edu.cn

Abstract. The reconstruction of film and TV scenes is an important part of the film and TV production process, which has a decisive impact on the visual effect of the film and the audience's viewing experience. The modeling method of automatically obtaining the 3D geometric structure of natural scenes using 3D reconstruction technology can break away from the tedious manual interaction mode of traditional 3D modeling, making the 3D modeling process simpler and more convenient. This study attempts to apply computer-aided design (CAD) and machine vision technology to the reconstruction of film and TV scenes, aiming to reduce the complexity of the model while ensuring its accuracy, thereby improving the overall efficiency of film and TV scene reconstruction. The study also introduced an assessment function based on wavelet transform (WT) to evaluate the quality of film and TV scene reconstruction. Compared with the WT model, the improved algorithm proposed in this article significantly improves image processing efficiency and reduces processing time. In addition, by introducing lighting and texture information, the reconstructed model has a higher sense of realism, providing the audience with an immersive viewing experience, thereby improving the quality of the viewing experience. The research results have played a crucial role in various stages of film and TV scene reconstruction, bringing higher value and broader creative space to film and TV production.

Keywords: Computer-Aided Design; Machine Vision; Film and TV Scenes; 3D Reconstruction
DOI: https://doi.org/10.14733/cadaps.2024.S15.290-307

1 INTRODUCTION

Machine vision is the computer realization of human vision. Because the human visual system is intelligent and powerful, the information it obtains accounts for all the information obtained by human beings. Moreover, human vision not only obtains information through optical perception but

also processes and reconstructs information, that is to say, it not only contains low-level responses to light-induced stimuli but also converts these stimuli into high-level meaningful content. Image recognition is an important field of deep learning applications. By utilizing deep learning techniques, various elements in images, such as characters, scenery, text, etc., can be efficiently recognized and analyzed. In the field of film and television production, image recognition can help producers, directors, and photographers better understand the content in the shot, thereby better controlling the creative process of film and television works. The application of deep learning in image recognition mainly involves convolutional neural networks (CNNs). CNN is a neural network specifically used for processing image data, which can automatically extract image features, classify, and recognize them. In the field of film and television production, CNN can be used to analyze and identify elements such as scenes, characters, props, etc. in images, helping producers, directors, and photographers make better decisions. Film and television scene reconstruction is the process of using deep learning technology to digitize and transform film and television works. Through deep learning technology, scenes, characters, props, etc. in film and television works can be identified and analyzed, and digital reconstruction and transformation can be carried out. This technology can help producers, directors, and photographers achieve more precise creation and post-production [1]. The design framework based on triangles and networks plays an important role in the reconstruction of film and television scenes. Erdolu [2] explored the principles, applications, and development prospects of this design framework. A triangle-based design framework divides images into triangular meshes to better represent various shapes and structures in the image. The principle of this design framework is relatively simple but very effective. In the reconstruction of film and television scenes, a triangle-based design framework can help producers, directors, and photographers better understand the content in the shot, thereby better controlling the creative process of film and television works. Divide the image into triangular meshes. This segmentation can be achieved manually or through automated algorithms. In the reconstruction of film and television scenes, this segmentation can better identify various elements in the scene, such as characters, props, backgrounds, etc. Build a 3D model based on the segmented triangular mesh. This model can accurately represent various shapes and structures in images. In the reconstruction of film and television scenes, this model can be used for digital reconstruction and transformation of various elements in the scene.

Film and television scene reconstruction is the process of using deep learning technology to digitize and transform film and television works. Through deep learning technology, scenes, characters, props, etc. in film and television works can be identified and analyzed, and digital reconstruction and transformation can be carried out. This technology can help producers, directors, and photographers achieve more precise creation and post-production. The application of deep learning in film and television scene reconstruction mainly involves technologies such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). GAN is a generation model that can generate new works similar to training data through training data. In the field of film and television production, GAN can be used for digital reconstruction and transformation of scenes and characters in film and television works. VAE is an autoencoder that can encode input data into implicit vectors and generate new data from these vectors. In the field of film and television production, VAE can be used for digital reconstruction and transformation of scenes, characters, props, and other elements in film and television works. In the future, deep learning technology may be combined with technologies such as virtual reality (VR) and augmented reality (AR) to provide viewers with a more immersive viewing experience. In addition, deep learning technology can also be used for digital preservation and inheritance of cultural heritage, leaving behind richer historical treasures for humanity [3]. The reconstruction of film and television scenes based on CAD and machine vision refers to the 3D model generated by observing the scene. It depends on the observed geometric shape, the light source and its characteristics of the scene, and the characteristics of the sensor. The modeling method of automatically obtaining the three-dimensional geometric structure of natural scenes using 3D reconstruction technology can eliminate the tedious manual interaction mode of traditional 3D modeling, making the 3D modeling process simpler and more convenient. During the projection process, sensors integrate the spatial relationships, physical characteristics, and surface

reflection characteristics of the 3D scene into the grayscale values of the 3D image. This transformation is irreversible and not unique [4].

Traditional machine learning techniques and CNN have broad application prospects in film and television image analysis. Traditional machine learning technology has advantages such as simplicity, ease of use, and strong interpretability, but it also has disadvantages such as weak model generalization ability. CNN has powerful feature learning and abstraction capabilities, which can automatically learn richer feature information from raw image data. However, it also has drawbacks, such as high model complexity and poor interpretability. Iqbal et al. [5] introduced the applications, advantages, and disadvantages of these two methods in film and television image analysis. Object detection is another important task in medical image analysis, which can help directors quickly and accurately locate the changing regions in the image. Traditional machine learning techniques can detect unknown film and television images by learning from known ones. The reconstruction of film and television scenes is an important part of the film and television production process, which has a decisive impact on the visual effect of the movie and the viewing experience of the audience. Traditional film and television scene reconstruction mainly relies on manual design and construction. Although this method can achieve high scene recovery, it is inefficient, costly, and often difficult to reconstruct complex scenes. With the development of technology, the recognition and interpretation of human emotions have become increasingly important. Especially in fields such as psychology, behavioral science, and human-computer interaction, accurately identifying and understanding human emotions has important application value. Kamble and Sengupta [6] analyzed video scene reconstruction technology based on CAD and machine vision, bringing a new solution for video production. CAD technology can provide accurate numerical models for scene design, and machine vision can extract the geometric structure of 3D scenes from 3D images. The combination of the two can automatically and efficiently complete the reconstruction of complex scenes, greatly improving the efficiency of film and television production.

Machine learning-based evaluation functions can be used to evaluate the quality of film and television scene reconstruction. This evaluation function can analyze and evaluate the reconstructed scene to measure its similarity, restoration degree, and other evaluation indicators with the original scene. One possible approach is to use deep learning models, such as convolutional neural networks (CNN) or generative adversarial networks (GAN), to evaluate the reconstructed scene. This model can accept reconstructed scenes as input and output a score or rating to represent the quality of the scene or its similarity to the original scene. For example, CNN can be used to classify or regress images of reconstructed scenes to evaluate their quality. If the input image is more similar to the original scene image, the evaluation score of the CNN output will be higher. GAN can also be used to generate new scene images and compare them with the original scene images to evaluate their similarity and quality [7]. Krner et al. [8] believe that CAD models are crucial in the reconstruction of film and television scenes. However, creating an accurate CAD model requires a lot of manual involvement, which undoubtedly increases time and resource costs. In addition, some complex shapes and structures may be difficult to describe using CAD models, which requires designers to possess superb skills and experience accurately. CAD models not only require precision but may also involve a large amount of data and details, which may lead to the model being too complex, thereby affecting computational efficiency and real-time performance. Therefore, it is necessary to minimize the complexity of the model while ensuring its accuracy in order to improve computational efficiency. For machine vision, handling complex lighting and shadows is a major challenge. The real lighting and shadow effects are often influenced by many factors, such as the direction, color, and intensity of the light source, as well as the shape and material of the object. To accurately simulate these effects, profound knowledge of image processing and machine learning is required. With the rapid development of technology, 3D reconstruction technology has been widely applied in many fields. Among them, the incremental point cloud compression and remote 3D reconstruction technology based on depth cameras have become a research hotspot. Li et al. [9] introduced the principle, implementation methods, and application scenarios of this technology. Incremental point cloud compression based on depth cameras refers to the use of depth cameras to obtain point cloud data of a scene and compress it through specific algorithms to reduce the size and storage space of the data.

During the compression process, a portion of the data is usually lost, but key geometric shapes and features are retained to ensure the quality of subsequent 3D reconstruction. The principle of incremental point cloud compression is to segment the point clouds in the scene into multiple sub-regions. Then, each subregion is locally compressed and encoded to achieve a compact representation of the data. In addition, some geometric constraints and prior knowledge can be utilized to optimize the data representation further to reduce reconstruction errors and data size. Traditional film and TV scene reconstruction relies on manual design and construction, which is time-consuming and laborious. Based on CAD and machine vision technology, 3D scene information can be extracted from 2D images, and complex scene reconstruction can be completed automatically. The research on the reconstruction technology of film and TV scenes based on CAD and machine vision is helpful in promoting technical progress in the field of film and TV production. By realizing automatic and efficient scene reconstruction, the efficiency of film production can be significantly improved, the cost can be reduced, and the visual effect of the film can be enriched. Through high-precision scene reconstruction, the scene of the film can be more truly restored, making the visual effect of the film more shocking and improving the audience's viewing experience. Aiming at the problems of CAD and machine vision in film and TV scene reconstruction, this article will study how to improve the accuracy and efficiency of CAD models, improve the performance of machine vision under complex lighting and shadow conditions, and optimize the data fusion algorithm between them. In this article, the improved technology is used for empirical research, and the performance and effect of the technology in practical application are verified by selecting representative film and TV scenes. Studying and optimizing the technology of CAD and machine vision is expected to improve the performance and effect of these two technologies in film and TV scene reconstruction and provide new ideas and methods for solving problems in film and TV scene reconstruction. The innovations of this study include:

(1) This study attempts to apply the combination of CAD and machine vision technology to the reconstruction of film and TV scenes. By using CAD technology to provide accurate numerical models, machine vision extracts 3D scene information from 2D images, and the two complement each other, enabling automated and efficient reconstruction of complex scenes.

(2) Traditional CAD models, while pursuing accuracy, often lead to overly complex models, increasing computational and processing difficulties. This study proposes an optimization algorithm aimed at reducing the complexity of the model while ensuring its accuracy, thereby improving the overall efficiency of film and TV scene reconstruction.

(3) When dealing with complex lighting and shadow conditions, machine vision often finds it difficult to extract the geometric structure of the scene accurately. This study improved the performance of machine vision in handling complex lighting and shadow conditions by improving image processing algorithms and introducing advanced deep-learning techniques.

(4) This article introduces a WT-based assessment function to evaluate the quality of film and TV scene reconstruction. This function utilizes the multi-scale analysis characteristics of WT and can comprehensively consider the quality of images in different frequency subbands, providing a comprehensive and objective assessment index for film and TV scene reconstruction.

The article first introduces the significance of the application of CAD and machine vision in film and TV scene reconstruction. Then, an overview of relevant theories was provided, and based on this, the WT and 3D reconstruction of film and TV scene images in this article were proposed; Subsequently, the performance of the proposed film and TV scene reconstruction method in terms of accuracy, efficiency, and realism was verified through experiments; Finally, the research results were summarized and research prospects were proposed.

2 OVERVIEW OF RELEVANT THEORIES

In practical applications, it is necessary to consider the use of efficient algorithms and optimized hardware resources. In addition, due to the diversity and complexity of film and television images, multiple classifiers may be needed to process different types of scenes in order to achieve better

classification results. Lin et al. [10] proposed an effective solution for classifying enhanced film and television image scenes using a multi-layer fractional order machine learning classifier. This classifier can handle various complex and ever-changing scenes and can adapt well to different enhancement techniques and image characteristics. Use a multi-layer fractional order machine learning classifier to classify preprocessed images. This classifier usually consists of multiple convolutional layers, pooling layers, and fully connected layers, which can effectively extract local and global features of the image. At each level, fractional convolution can be used to extract more complex features, such as edges, textures, etc. When training a classifier, it is necessary to prepare a large amount of training data, including various types of film and television images and corresponding labels. You can use existing movie scene classification datasets or build your own datasets. In the reconstruction of film and television scenes, IFS-based image enhancement technology can be used to improve the contrast, clarity, color saturation, etc., of images in order to better display the details and features in the scene. This technology can also be combined with other computer vision and image processing methods, such as edge detection, feature extraction, deep learning, etc., further to improve the quality and accuracy of scene reconstruction. Lin [11] preprocesses the input image, such as noise removal, color balance, histogram equalization, etc., in order to provide a good initial image for subsequent IFS transformations. Using the preprocessed image as the initial value, iteratively apply the local transformations in the defined IFS until the preset stop condition is reached or the maximum number of iterations is reached. Take the final iterated image as the output result and perform necessary post-processing, such as sharpening, color correction, etc., to improve the quality and visual effect of the enhanced image. In the reconstruction of film and television scenes, IFS-based image enhancement technology can be applied to various elements of the scene. By enhancing the details and features of these elements, their form and texture can be better displayed, providing a more realistic and vivid scene experience. In order to protect and repair these precious film films, we can use imaging spectroscopy technology and machine learning algorithms for digital restoration. Imaging spectroscopy technology is an advanced technology that can simultaneously obtain object color and spectral information. Liu et al. [12] used this technique to obtain color and spectral information for each pixel in a movie film. This provides us with a detailed and comprehensive understanding of the film surface, enabling us to identify damaged parts and parts that require repair. The main advantage of imaging spectroscopy technology in digital restoration is its ability to provide accurate color and spectral information. This makes the repair process more precise and avoids repair failures caused by inaccurate colors or spectra. At the same time, this technology also enables us better to understand the material and structure of the film, thus developing more effective repair strategies. When training the model, we need to prepare a set of labeled training data, which includes defective film images and corresponding repaired images. Then, we can repeatedly iterate through the model learning process to identify defects from the original image and generate a repaired image. Once the model training is completed, we can use it to repair new movie film images automatically.

Partial beam adjustment is a method of beam adjustment that defines the adjustment function on the beam and utilizes the characteristics of the beam to solve the adjustment problem. In 3D reconstruction, partial beam adjustment can represent the point cloud data in the reconstructed scene as the intersection points of a series of beams, and the characteristics of these beams can be used to constrain and optimize the point cloud data. In the reconstruction of dynamic scenes, it is necessary to match and register the scene in real-time due to the changes in the position and pose of objects in the scene over time. The method of partial beam adjustment can effectively solve this problem. By representing the feature points in each frame of the image as beams and using the intersection points of the beams to solve for matching points, fast and accurate dynamic scene reconstruction can be achieved. Luo et al. [13] introduced the partial bundle adjustment method in precise 3D reconstruction and explored its application, advantages, and disadvantages in 3D reconstruction. The partial beam adjustment method, as an effective beam adjustment method, can achieve high precision, high adaptability, and high scalability, resulting in 3D reconstruction. On the basis of initial reconstruction, we can use the method of partial bundle adjustment to make more precise adjustments and optimizations. By representing the point cloud data in the reconstructed

scene as the intersection points of a series of light beams and utilizing the characteristics of these beams to constrain and optimize the point cloud data, reconstruction errors and model incompleteness can be effectively reduced [14]. By using advanced algorithms and big data technology, Mo et al. [15] simulated and optimized the design and assembly process of products, thereby improving product quality and performance while reducing costs and improving efficiency. It explores how to obtain design intent through product information modeling to achieve more effective computer-aided intelligent assembly modeling. Product information modeling is a method of describing the shape, size, structure, material, performance, and other attributes of a product by establishing a mathematical model. It can help us better understand the design and functionality of products, and provide necessary data support for computer-aided intelligent assembly modeling. Through computer-aided intelligent assembly modeling, we can simulate the assembly process of products, predict potential problems, optimize assembly line design, and even predict the product's lifecycle and maintenance requirements. This not only helps us improve product quality and performance but also reduces product costs and development time.

With the rapid development of the film and television industry and the increasing demand for emotional expression from audiences, the application of emotion recognition technology in film and television production is becoming increasingly widespread. Su et al. [16] introduced how to use machine learning and deep learning techniques to perform multidimensional emotion recognition on film and television scene images and combine electronic imaging technology to achieve more realistic and rich emotional expression. The basic principle of emotion recognition is to extract features and recognize patterns from input data, such as text, speech, and images, in order to determine the category of emotions expressed. The emotional recognition of film and television scene images mainly involves the extraction and analysis of visual information, such as character expressions, colors, textures, etc., in the images. Emotional recognition of scene images is of great significance in film and television production. The emotional analysis of scene images can help directors better grasp the plot direction and character traits. At the same time, it also helps actors better understand the character's emotions, thereby more vividly portraying the character. The computer-aided digital archiving technology for architectural spatial perception experience has broad application value in multiple fields. For example, it can be used as a material library for architectural education, allowing students to experience and learn various types of architectural spaces intuitively. It can also be used for evaluating building quality by analyzing digital models of building spaces to evaluate their design rationality and performance. In addition, this technology can also be used for energy conservation management in intelligent buildings by simulating and analyzing digital models of building spaces to develop more reasonable energy conservation plans. With the continuous development of technology, we can foresee the application of more advanced computer-aided technology in the digital archiving field of architectural spatial perception experience in the future. Tai and Sung [17] utilize artificial intelligence (AI) technology for automated analysis and evaluation. Utilize cloud computing technology to achieve efficient distributed storage and management, and utilize 5G and IoT technology to achieve real-time collection and feedback of human activity information. The development of these technologies will further promote the progress and expansion of the application scope of computer-aided digital archiving technology for building spatial perception experiences. Wang et al. [18] explored a solution for achieving all-weather, naturally silent speech recognition through machine learning-assisted tattoo electronic devices. Tattood electronic devices are electronic devices that can be implanted into the human skin. They utilize the bioelectricity of the human body to power them and have the advantages of small size, portability, and convenience in use. With the development of technology, the functions and performance of tattoo electronic devices have been greatly improved, which can achieve various functions, such as biological signal detection, health monitoring, information transmission, etc. In addition, tattoo-style electronic devices can also transmit and interact with other devices through wireless communication technology, which provides the possibility for our speech recognition solution. In terms of equipment, tattoo electronic devices use highly sensitive sound sensors and bioelectric energy collectors. When a person is pronouncing, the sound sensor will collect the speech signal and convert it into a digital signal for processing. At the same time, the bioelectric energy collector will collect the bioelectric energy of the human body and

provide power to the equipment. In this way, users do not need additional power supply or specialized recording equipment to collect voice signals. In terms of algorithms, we have adopted deep learning and neural network techniques. A scheme for achieving all-weather, naturally silent speech recognition through machine learning-assisted tattoo electronic devices has been proposed. This scheme combines the advantages of tattoo-style electronic devices and machine learning algorithms and can achieve all-weather, naturally silent speech recognition. The experimental results show that our scheme can achieve high accuracy speech recognition and has strong adaptability and robustness.

Zhao and Cole's [19] use of text mining and machine learning techniques to reconstruct dispersion relationships from a large amount of literature data and predict refractive index has important practical significance. Specifically, text mining technology can extract information related to dispersion and refractive index from a large number of literature, and machine learning algorithms can be used to establish models between this information and material properties. A common method for reconstructing dispersion relationships is to use the Support Vector Machine (SVM) algorithm. An effective method for predicting refractive index is to use neural network models. A neural network is a model that simulates the connectivity of neurons in the human brain, which can predict unknown inputs by learning a large amount of data. By using known refractive index data as a training set, neural networks can learn the relationship between refractive index and other properties of substances and use it to predict the refractive index of unknown substances. In practical applications, we need first to collect a large amount of literature data and use text-mining techniques to extract information related to dispersion and refractive index. Then, we use SVM or neural network models to establish a model between this information and material properties. Finally, we can use this model to predict the dispersion relationship and refractive index of unknown substances. With the rapid development of film and television technology, computer intelligent design has become an important tool in many shooting fields. In the design of film and television scene reconstruction, computer intelligent design also plays a crucial role. Modular design is a method of dividing complex systems into a series of independently exploitable modules, which has broad application prospects in the intelligent design of computer reconstruction for film and television scenes. Modular design has advantages such as improving design efficiency, reducing development costs, and enhancing maintainability [20].

3 WT AND 3D RECONSTRUCTION OF FILM AND TV SCENE IMAGES

3.1 Assessment Function Based on WT

3D scene reconstruction is the process of restoring or generating 3D shapes from 2D images. This process involves a series of complex machine vision and graphics technologies. In film and TV production, 3D scene reconstruction plays a crucial role in creating high-realism visual effects. It not only enhances the visual impact of movies but also provides a richer visual experience. From a technical perspective, 3D scene reconstruction can be divided into two types: active and passive. Active 3D reconstruction technology obtains 3D information about the scene through the use of special hardware devices such as laser scanners, structured light projectors, etc. Passive 3D reconstruction relies on algorithms to analyze and infer ordinary 2D images, thereby obtaining 3D structures. In film and TV production, passive 3D reconstruction technology based on machine vision has received widespread attention and application due to its low equipment requirements and strong adaptability. CAD is a technology that utilizes computer technology to assist designers in various design activities. CAD systems can receive instructions and information input by designers and, after computer processing, generate the results and information required by designers.

CAD technology is widely used in scene design and pre-visualization in film and TV scene reconstruction. By using CAD software, film and TV production personnel can easily create high-precision 3D models and make detailed adjustments and optimizations to the models. In addition, CAD models can also be easily imported into other film and TV production software for subsequent animation, special effects, and other production. Traditional CAD modeling methods

often have low efficiency and require a lot of time and effort when dealing with complex scenes. Therefore, how to combine CAD technology and machine vision technology to achieve efficient and automated film and TV scene reconstruction is an important direction of current research.

Machine vision is the use of computers to simulate human visual functions, extract information from images of objective objects, process and understand it, and ultimately use it for practical detection, measurement, and control. In film and TV scene reconstruction, machine vision can derive 3D information of the scene by analyzing the captured 3D images. For example, through stereo vision matching algorithms, the 3D geometric structure of the scene can be obtained from two images taken from different angles. In addition, machine vision can also be used for tasks such as texture mapping and lighting analysis of scenes, further improving the realism of the scene. It should be noted that although machine vision has great potential in film and TV scene reconstruction, its application also faces some challenges, such as lighting changes, occlusion issues, computational complexity, etc. Therefore, how to overcome these problems and improve the performance and stability of machine vision in film and TV scene reconstruction is also an important content of this study. In the process of 3D reconstruction of movie and television scenes, WT is an important preprocessing step. WT has good time-frequency localization characteristics and can analyze the characteristics of images on multiple scales, so it has been widely used in image processing, image compression, and other fields. In the image WT, 3D discrete WT is the most commonly used method. It regards the image as a 3D matrix and obtains the multi-scale decomposition of the image by performing one-dimensional WT on the rows and columns of the image, respectively. This decomposition can decompose the image into different frequency subbands so that most of the energy of the image is concentrated in a few subbands, which provides convenience for subsequent 3D reconstruction. WT-based assessment function plays a key role in the 3D reconstruction of movie scenes. It can evaluate the image quality after WT and guide the optimization of the 3D reconstruction algorithm.

WT of image is the basis of wavelet applied to image processing and image compression, which is based on 3D discrete WT. The image can be regarded as a 3D matrix, and the image WT can be regarded as a univariate wavelet to decompose the rows and columns of the image, respectively. After doing multi-layer WT on the image, some sub-images with different resolutions will be obtained, as shown in Figure 1.

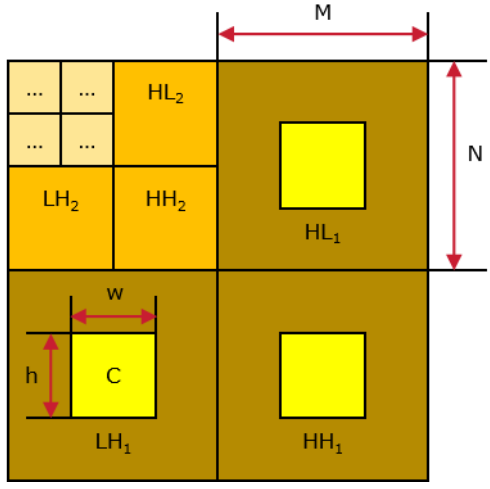


Figure 1: Schematic diagram of image wavelet decomposition.

The clearer the image, the larger the high-frequency coefficient after wavelet decomposition; The more blurry the image (the greater the degree of defocus), the smaller the high-frequency coefficient

after wavelet decomposition, while the low-frequency coefficient increases. Based on this idea, various WT-based clarity assessment functions have emerged.

The square assessment function of the sum of wavelet coefficients (Harr wavelet basis):

$$F_{haar\_2} = \sum_{i=1}^{M}\sum_{j=1}^{N}\left[\left|W_{HL2}\ i,j\right| + \left|W_{LH2}\ i,j\right|\right]^2 \tag{1}$$

Wavelet transform (db6 wavelet basis) M_{WT}^2 operator assessment function:

$$F_{db\_6} = \frac{1}{wh}\left[\begin{array}{l}\sum_{i=1}^{M}\sum_{j=1}^{N} W_{LH1}\ i,j\ - \mu_{LH1}\ ^2 \\ +\sum_{i=x_0}^{w}\sum_{j=y_0}^{h} W_{HL1}\ i,j\ - \mu_{HL1}\ ^2 \\ +\sum_{i=x_0}^{w}\sum_{j=y_0}^{h} W_{HH1}\ i,j\ - \mu_{HH1}\ ^2\end{array}\right] \tag{2}$$

Where μ_{LH1}, μ_{HL1} and μ_{HH1} are the mean values of wavelet coefficients in the corresponding local assessment window.

Assessment function of high-frequency and low-frequency comprehensive change characteristics (db6 wavelet base);

$$F_{db6\_3} = \frac{M_H^2}{M_L^2} \tag{3}$$

$$M_H^2 = \sum_{n=1}^{3}\left[\sum_{i,j\in S_{LHn}} W_{LHn}^2\ i,j\ + \sum_{i,j\in S_{HLn}} W_{HLn}^2\ i,j\ + \sum_{i,j\in S_{HHn}} W_{HHn}^2\ i,j\right] \tag{4}$$

$$M_L^2 = \sum_{i,j\in S_{LL3}} W_{LL3}^2\ i,j \tag{5}$$

Among them, S_{LHn}, S_{HLn} and S_{HHn} are the decomposition windows of high-frequency band after N-layer decomposition of db6 wavelet respectively, and the corresponding high-frequency wavelet coefficients are W_{LH_n}, W_{HL_n} respectively. W_{LL3} is the low-frequency coefficient of layer 3 wavelet transform.

The choice of wavelet basis directly affects the effectiveness of the assessment function. In the field of image processing, commonly used wavelet bases include haar, db, biro, and sym. In the same experimental environment, different wavelet bases were used to evaluate the film and TV images in Figure 5, and the results are shown in Figure 2. The results indicate that the db2 and sym2 wavelet bases perform well in terms of resolution and sensitivity. In addition, under the same vanishing moment order conditions, the DB wavelet has the minimum support length and computational complexity, making it the preferred wavelet basis for image decomposition.

Generally speaking, when choosing a wavelet for image processing, we always want to choose a wavelet with a high vanishing moment order, because a high vanishing moment usually means better frequency resolution and finer image feature extraction ability. The higher the vanishing moment order, the greater the computational complexity, which may increase the computational complexity and time cost. Therefore, in practical application, it is necessary to make a trade-off between the order of vanishing moment and the amount of operation. Figure 3 shows the results of the assessment and comparison by taking 3 ~ 6 vanishing moments of the db wavelet. When the vanishing moment order is 6, the DB wavelet has the best assessment accuracy and sensitivity. This

shows that when selecting wavelet bases, we should consider the order of vanishing moment and the amount of operation to choose the wavelet bases that are most suitable for the current task.

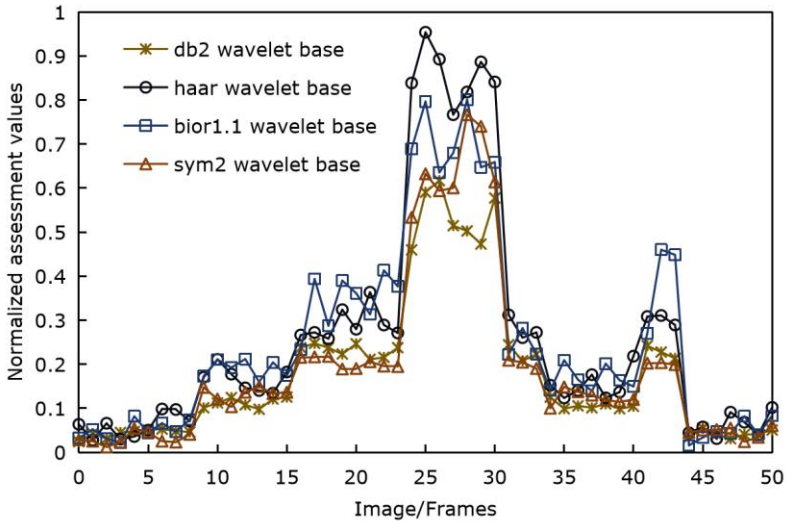


Figure 2: Comparison of assessment performance of db, haar, coif, and sym wavelet bases.

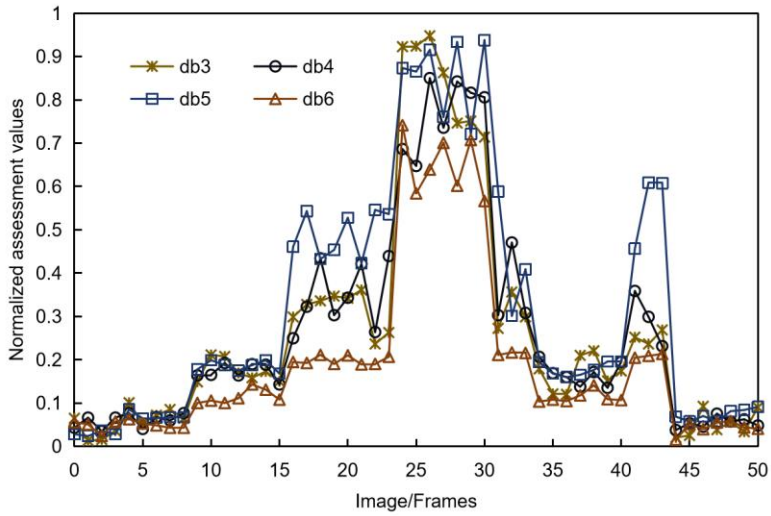


Figure 3: Comparison of assessment performance of db wavelet with different vanishing moment orders.

As the number of decomposition layers increases, the wavelet coefficients in the decomposed subgraph gradually move towards the low-frequency part. In this case, the assessment window focuses more on the approximate parts of the image when selecting, and may overlook the details that reflect the clarity of the image. Figure 4 shows the comparison of assessment performance analysis after using db6 wavelet for 1-layer, 2-layer, and 3-layer decomposition. From the graph, it can be clearly seen that the assessment performance of layer 1 decomposition is significantly better than that of layer two and layer three decomposition, with good resolution and unimodal performance. In addition, as the number of wavelet decomposition layers increases, the

computational load will also gradually increase. Therefore, when considering the balance between computational complexity and assessment performance, a 1-layer db6 wavelet decomposition can be chosen as the image processing method.

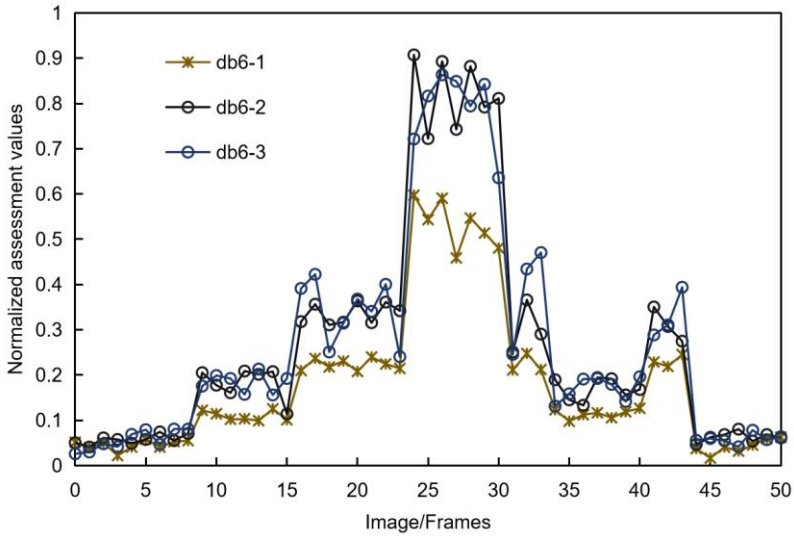


Figure 4: Comparison of the assessment performance of DB6 wavelet transform in layers 1, 2, and 3.

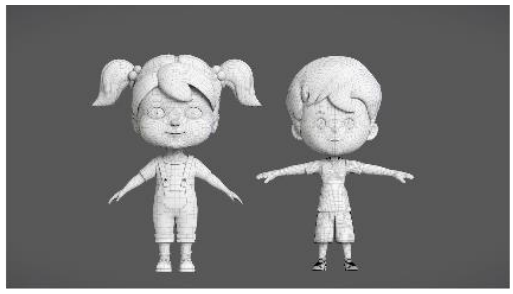
3.2 3D Reconstruction of Film and TV Scenes

This study aims to explore the possibility of integrating CAD and machine vision technology in the field of film and TV scene reconstruction. By combining the precise numerical modeling capabilities of CAD with the technology of machine vision to extract 3D scene information from 2D images, we expect to be able to automate and efficiently reconstruct complex scenes. As shown in Figure 5, the input of this study is a video or image sequence captured from multiple perspectives of a scene and its corresponding depth map for each frame. The research goal is to reconstruct the precise 3D geometric structure of the scene by deeply fusing these depth map sequences. Firstly, sample and backproject each depth map frame to obtain the corresponding 3D point set for each frame. In this process, some imprecise 3D points are eliminated through the statistical analysis of depth error in reprojection to improve the accuracy of reconstruction. Then, the multi-frame 3D points are collected and combined into the overall 3D point cloud of the scene. This process will eliminate redundant 3D points, ensuring the clarity and accuracy of the point cloud data. Finally, a complete geometric model of the scene is extracted using the fused 3D point cloud. This model will accurately and comprehensively reflect the 3D structure of the original scene.

When a 3D object in the real world is projected onto a 3D plane, it is necessary to establish a 3D rectangular coordinate system on the projection plane, which is the image coordinate system. The image coordinate system can be subdivided into ideal image coordinate system x_u, y_u , real image coordinate system x_d, y_d , and pixel coordinate system x_p, y_p . The relationship between the above coordinate systems is shown in Figure 6.

When projecting 3D objects from the real world onto a 3D plane, in order to accurately describe and locate these projections, it is necessary to establish a 3D Cartesian coordinate system on the projection plane, which is called the image coordinate system. The image coordinate system plays a crucial role in the field of machine vision and image processing, providing a way for research to describe and manipulate images on a 3D plane. The ideal image coordinate system is a theoretical coordinate system that assumes that the image is not affected by any distortion or noise. In this

coordinate system, there is a simple linear relationship between the points of the image and those in 3D space, which greatly simplifies the complexity of image processing and analysis. The real image coordinate system considers various factors in reality, such as lens distortion, lighting conditions, etc., which can lead to a certain degree of image distortion. In the real image coordinate system, it is necessary to preprocess and correct the image to reduce or eliminate these distortions, so that the image processing results are closer to the real situation. The pixel coordinate system describes image coordinates in units of pixels. In this coordinate system, each pixel has a fixed coordinate value, which enables precise positioning and manipulation of each pixel in the image in research.



(a) Initial model for deep feature fusion

(b) Reconstructed 3D character model

(c) Local features of 3D model

Figure 5: 3D reconstruction results of a movie scene example.

There is a rigid transformation relationship between the world coordinate system and the camera coordinate system:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + l \tag{6}$$

The basic formula of lens imaging in physics is defined as:

$$\frac{1}{f}=\frac{1}{m}+\frac{1}{n} \tag{7}$$

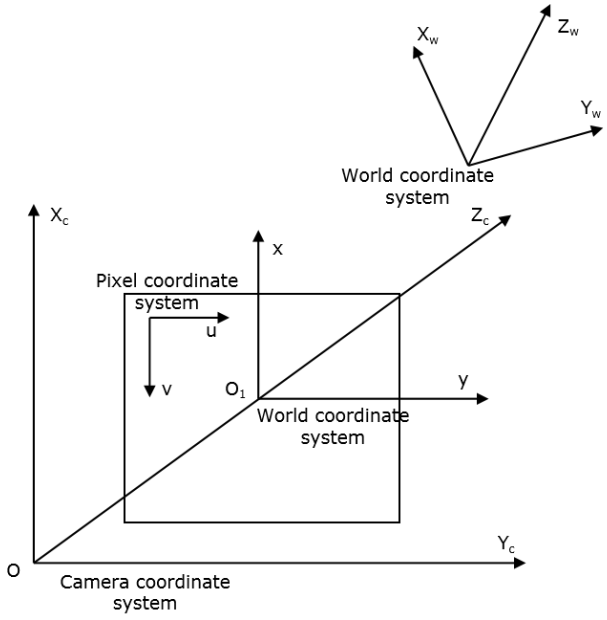


Figure 6: Several coordinate systems.

Where f is the focal length of the lens, m is the image distance, and is the object distance.

In perspective projection, because $n>>f$, it is approximately considered that $m=f$, that is, the image distance is equal to the focal length. According to the perspective projection of the pinhole imaging model, the central projection equation expressed by non-homogeneous coordinates is:

$$\begin{cases} x_u = f\dfrac{x_c}{z_c} \\ y_u = f\dfrac{y_c}{z_c} \end{cases} \tag{8}$$

Homogeneity means:

$$z_c\begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix}=\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{9}$$

x_u and y_u represent the 3D central projection coordinates in the ideal image coordinate system. The conversion formula from world coordinates to ideal coordinates can be written as:

$$\lambda\begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix}=K\begin{bmatrix} R & -Rt \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{10}$$

Where R is a rotation matrix, t is a translation vector, and $K = diag[f, f, 1]$ is the internal parameter matrix of the camera in an ideal situation.

4 EXPERIMENTAL RESULTS

4.1 Simulation Performance Testing

The experimental goal is to verify the performance of the proposed film and TV scene reconstruction method in terms of accuracy, efficiency, and realism. In order to comprehensively evaluate the performance of the method proposed in this article, the experiment used film and TV images of different scales and complexities as the experimental dataset. These datasets contain images from various scenes and lighting conditions to verify the adaptability of the method. Perform necessary preprocessing operations on the original image, such as denoising and image enhancement, to improve image quality and accuracy of subsequent processing. Using CAD technology for 3D reconstruction based on extracted feature points. By matching feature points, the 3D geometric structure of the scene is restored, and the corresponding CAD model is generated. Optimize the generated CAD model to increase its realism. Finally, the model is displayed in a virtual environment through rendering techniques. Firstly, different scale film and TV images were analyzed for operational efficiency using different node count methods, as shown in Figure 7.

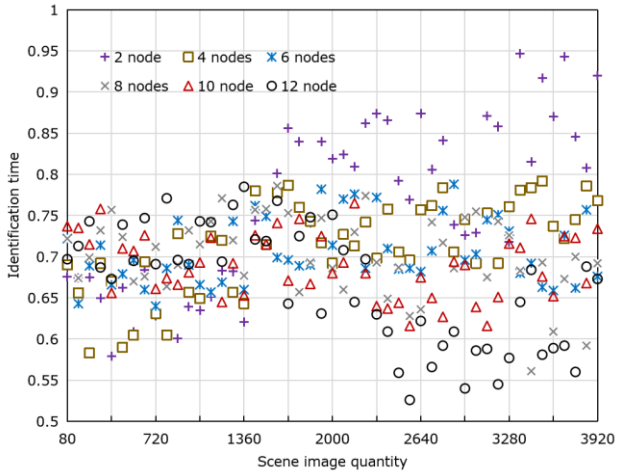


Figure 7: Time consumption of film and TV image recognition.

In cases where the scale of film and TV images is relatively small, the time required for image recognition increases with the number of nodes in the image. When the image size is small, adding nodes means increasing computational complexity, and due to the limited information content of the image itself, this increase does not bring significant performance improvement. As the image size gradually increases, the recognition efficiency of multiple nodes shows a significant upward trend. This is because, for large-scale images, the time and resources required for single-node processing greatly increase, while multi-node parallel processing can effectively share this computational burden, thereby achieving a reduction in processing time. In the process of reconstructing film and TV scenes, the scale of images that need to be processed is often large, so how to use multiple nodes to improve processing efficiency is a very important issue. The results indicate that for large-scale film and TV images, using multi-node processing can significantly improve processing efficiency. Figure 8 clearly shows the time consumption comparison between the proposed method and the traditional WT model in film and TV image processing.

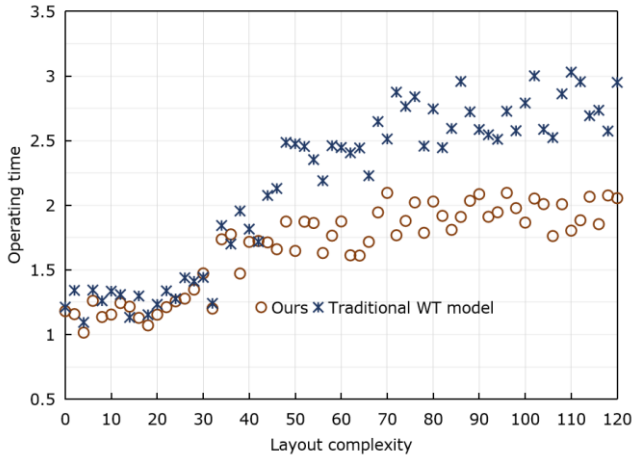


Figure 8: Time-consuming film and TV image processing.

The digital processing of film and TV images using traditional WT models takes significantly longer. This is because traditional WT may have redundant calculations in the processing process, and the optimization degree of the algorithm may not be as good as the improved algorithm in this article. In contrast, the improved algorithm presented in this article exhibits higher image processing efficiency and significantly reduces the required processing time.

4.2 Example Generated Based on A Single Scene Model

Figure 9 shows an original scene image captured by a camera. Figure 10 represents the virtual environment display effect after the 3D reconstruction of the original scene image based on technology. From the original scene image to the reconstructed model, a clear 3D structure restoration can be seen. This is thanks to the 3D coordinates of the feature points obtained through calculation. These coordinates restore the spatial structure information in the scene, making the flat image 3D, and providing a rich 3D environment for subsequent film and TV production.

Figure 9: Image of the original scene.

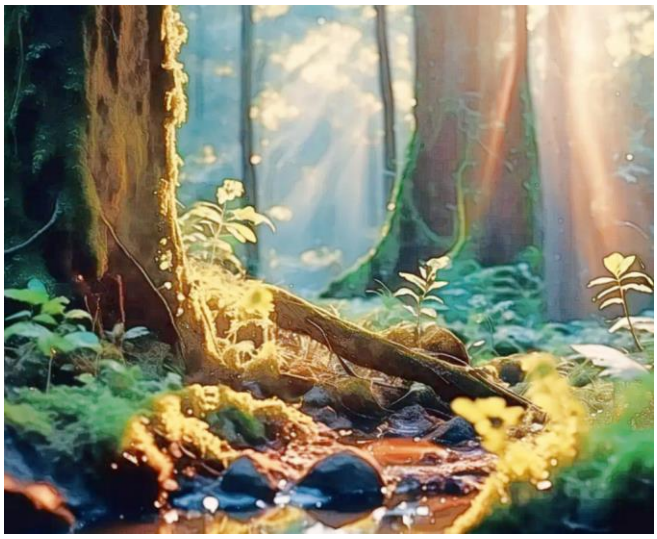


Figure 10: Model generated in a virtual environment.

In order to make the reconstructed model more realistic, this study further added lighting and texture information. The addition of lighting makes the model present a realistic light and shadow effect, and different parts exhibit different light and dark changes due to differences in lighting angle and intensity, which makes the model look more layered. The addition of texture information injects life into the model. Whether it is a large surface structure or a small decoration, fine texture maps are obtained, which makes the model visually comparable to real scenes. Extracting 3D information from 2D images and utilizing this information to construct realistic 3D scenes not only demonstrates the power of technology but also brings higher artistic value to film and TV production. In future film and TV production, using this technology, we can more freely create various complex scenes, bringing viewers a richer and more immersive visual experience.

The assessment function based on WT provides an effective quality assessment mechanism for film and TV scene reconstruction. This assessment function can quantify the quality of reconstructed images, guide algorithm optimization, and ensure that the final 3D scene reconstruction results have high accuracy and realism. By comparing the processing time of the traditional WT model with the method proposed in this article, it can be seen that the improved algorithm in this article is more efficient in film and TV image processing. This improvement in efficiency means that a significant amount of time and cost can be saved in the reconstruction process of film and TV scenes, thereby improving the overall production efficiency and better meeting the needs of film and TV production. By comparing the original scene images with the reconstructed model, the actual effect of this method in film and TV scene reconstruction can be clearly seen. The 3D coordinate extraction of feature points restores the structural information in the scene, making the flat image 3D and rich in hierarchy. The increased lighting and texture information further enhances the realism of the model, making the audience feel as if they are in the real world. The research results have played a crucial role in various stages of film and TV scene reconstruction, bringing higher value and broader creative space to film and TV production.

5 CONCLUSIONS

The emergence of film and TV scene reconstruction technology based on CAD and machine vision has brought new solutions to film and TV production. CAD technology can provide accurate numerical models for scene design, while machine vision can extract the geometric structure of 3D scenes from

3D images. The combination of the two can automate and efficiently complete the reconstruction of complex scenes. Traditional CAD models, while pursuing accuracy, often lead to overly complex models, increasing computational and processing difficulties. This study provides accurate numerical models through CAD technology, and machine vision extracts 3D scene information from 2D images. The two complement each other, enabling automated and efficient reconstruction of complex scenes. Compared to the traditional WT model, the improved algorithm in this article greatly improves the efficiency of image processing and reduces processing time. Moreover, by increasing lighting and texture information, the reconstructed model presents a higher sense of realism, immersing the audience and improving the viewing experience.

The integration of CAD and machine vision technology has brought more efficient and realistic scene reconstruction methods to film and TV production, injecting new vitality into the further growth of the film and TV industry. This article has demonstrated the effectiveness of CAD and machine vision technology in film and TV scene reconstruction. In the future, further exploration can be made of their combination with other technologies, such as deep learning and virtual reality, to improve the accuracy and interactive experience of reconstruction.

Qingyang Li, https://orcid.org0009-0004-3477-8184
Kai Wang, https://orcid.org/0000-0002-0906-7872

REFERENCES

[1] Buyukdemircioglu, M.; Kocaman, S.: Reconstruction and efficient visualization of heterogeneous 3d city models, remote sensing, 12(13), 2020, 2128. https://doi.org/10.3390/rs12132128
[2] Erdolu, E.: Lines, triangles, and nets: A framework for designing input technologies and interaction techniques for computer-aided design, International Journal of Architectural Computing, 17(4), 2019, 357-381. https://doi.org/10.1177/1478077119887360
[3] Fu, K.; Shi, W.; Ke, J.; Guo, K.: Image restoration and quantitative metallographic tissue based on machine vision, Journal of Electronic Imaging, 31(5), 2022, 051412-051412. https://doi.org/10.1117/1.JEI.31.5.051412
[4] Houssein, E.-H.; Hammad, A.; Ali, A.-A.: Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review, Neural Computing and Applications, 34(15), 2022, 12527-12557. https://doi.org/10.1007/s00521-022-07292-4
[5] Iqbal, S.-N.; Qureshi, A.; Li, J.; Mahmood, T.: On the analyses of medical images using traditional machine learning techniques and convolutional neural networks, Archives of Computational Methods in Engineering, 30(5), 2023, 3173-3233. https://doi.org/10.1007/s11831-023-09899-9
[6] Kamble, K.-S.; Sengupta, J.: Ensemble machine learning-based affective computing for emotion recognition using dual-decomposed EEG signals, IEEE Sensors Journal, 22(3), 2021, 2496-2507. https://doi.org/10.1109/JSEN.2021.3135953
[7] Kim, H.-J.; Chong, M.; Rhee, T.-G.; Khim, Y.-G.; Jung, M.-H.; Kim, Y.-M.; Chang, Y.-J.: Machine-learning-assisted analysis of transition metal dichalcogenide thin-film growth, Nano Convergence, 10(1), 2023, 10. https://doi.org/10.1186/s40580-023-00359-5
[8] Krner, A.; Born, L.; Bucklin, O.: Integrative design and fabrication methodology for bio-inspired folding mechanisms for architectural applications, Computer-Aided Design, 133(80), 2020, 102988. https://doi.org/10.1016/j.cad.2020.102988
[9] Li, Y.; Gao, J.; Wang, X.: Depth camera based remote three-dimensional reconstruction using incremental point cloud compression, Computers & Electrical Engineering, 2022, 99, 2022, 107767. https://doi.org/10.1016/j.compeleceng.2022.107767
[10] Lin, C.-H.; Wu, J.-X.; Li, C.-M.; Chen, P.-Y.; Pai, N.-S.; Kuo, Y.-C.: Enhancement of chest X-ray images to improve screening accuracy rate using iterated function system and multilayer fractional-order machine learning classifier, IEEE Photonics Journal, 12(4), 1-18. https://doi.org/10.1109/JPHOT.2020.3013193

[11] Lin, C.-J.: Topological vision: applying an algorithmic framework for developing topological algorithm of architectural concept design, Computer-Aided Design and Applications, 16(3), 2019, 583-592. https://doi.org/10.14733/cadaps.2019.583-592
[12] Liu, L.; Catelli, E.; Katsaggelos, A.; Sciutto, G.; Mazzeo, R.; Milanic, M.; Walton, M.: Digital restoration of colour cinematic films using imaging spectroscopy and machine learning, Scientific reports, 12(1), 2022, 21982. https://doi.org/10.1038/s41598-022-25248-5
[13] Luo, K.; Pan, H.; Zhang, Y.: Partial bundle adjustment for accurate three-dimensional reconstruction, IET Computer Vision, 13(7), 2019, 666-675. https://doi.org/10.1049/iet-cvi.2018.5564
[14] Ma, K.; Mao, Z.; He, D.: Design a network architectural teaching system by auto CAD, Computer-Aided Design and Applications, 17(S2), 2020, 1-10. https://doi.org/10.14733/cadaps.2020.S2.1-10
[15] Mo, S.; Xu, Z.; Tang, W.: Product information modeling for capturing design intent for computer-aided intelligent assembly modeling, Journal of Northwestern Polytechnical University, 40(4), 2022, 892-900. https://doi.org/10.1051/jnwpu/20224040892
[16] Su, Z.; Liu, B.; Zhou, X.; Ren, H.: Multidimensional sentiment recognition of film and television scene images, Journal of Electronic Imaging, 30(6), 2021, 063014-063014. https://doi.org/10.1117/1.JEI.30.6.063014
[17] Tai, N.-C.; Sung, L.-W.: Digital archiving of perceptual experiences of an architectural space with computer-aided methods, Computer-Aided Design and Applications, 17(3), 2019, 585-597. https://doi.org/10.14733/cadaps.2020.585-597
[18] Wang, Y.; Tang, T.; Xu, Y.; Bai, Y.; Yin, L.; Li, G.; Huang, Y.: All-weather, natural silent speech recognition via machine-learning-assisted tattoo-like electronics, npj Flexible Electronics, 5(1), 2021, 20. https://doi.org/10.1038/s41528-021-00119-7
[19] Zhao, J.; Cole, J.-M.: Reconstructing chromatic-dispersion relations and predicting refractive indices using text mining and machine learning, Journal of Chemical Information and Modeling, 62(11), 2022, 2670-2684. https://doi.org/10.1021/acs.jcim.2c00253
[20] Zhao, S.; Xu, C.; Jiang, H.; Bu, C.: The strategy and method of intelligent computer-aided design for modular gun family, Acta Armamentarii, 42(4), 2021, 723-733. https://doi.org/10.3969/j.issn.1000-1093.2021.04.006