# CAD Technology Under the Background of Internet of Things and Its Application in Video Automatic Processing

Shihui Zhang[1], Xiqing Zhao[2], Naidi Liu[3,*] and Huanhuan Gao[4]

[1]College of Information Science and Engineering, Hebei North University, Zhangjiakou 075000, China, zhangshihui@hebeinu.edu.cn
[2]College of Information Science and Engineering, Hebei North University, Zhangjiakou 075000, China, zxqlytqq@163.com
[3]College of Information Science and Engineering, Hebei North University, Zhangjiakou 075000, China, liunaidi@hebeinu.edu.cn
[4]College of Information Science and Engineering, Hebei North University, Zhangjiakou 075000, China, gaohuanhuanghh@outlook.com

Corresponding author: Naidi Liu, liunaidi@hebeinu.edu.cn

**Abstract.** In the past few years, industrial software has entered a relatively mature stage in terms of product technology development and function growth mode, so the functions are more convergent, and the adopted technology can no longer be a direct tool for competition. In fact, in the IoT (Internet of things) solution of CAD (Computer Aided Design), it is not only complete in application, but also practical. At present, CAD is widely used in automatic video processing, which provides a lot of convenience for automatic video processing. Video target segmentation combines the related technologies of target tracking and image segmentation, and can track and segment rigid or flexible targets. In this paper, an asymmetric edge adaptive filtering algorithm based on bilateral filter is improved and implemented. This filtering algorithm can smooth the depth image while retaining the edge information, and can predict the position of the target to achieve the re-acquisition of the target after short occlusion. There is no big change in recall and accuracy between this algorithm and the traditional algorithm, but the time consumption of this algorithm is only 12.5% of that of the traditional algorithm. It can basically meet the requirements of real-time, which proves the effectiveness of this algorithm.

**Keywords:** IoT; CAD; Automatic video processing
**DOI:** https://doi.org/10.14733/cadaps.2023.S2.11-21

## 1 INTRODUCTION

Digital image processing technology is mainly used in two fields, one is to analyze and improve the image, the other is to make the machine automatically understand the image and store, transmit and display it. Internet of things (IOT) and CAD (Computer Aided Design) both consider the needs of information systems from the commercial perspective of manufacturing enterprises. The Internet of things considered by Paterson et al. [1] has evolved from solving business problems related to products to solving problems related to enterprise management. Whether it is the generalization of image processing or the need for video pattern recognition, different moving object segmentation processing is required.

Subtitles contained in the video stream express rich semantics and can play an effective role in the analysis and understanding of the original video stream. For example, subtitles in video news reports generally describe the time, place and person of the reported news. Keye and Gammon [2] introduced the basic idea of multimedia DM (data mining) and gave a prototype of multimedia mining system. Santos et al. [3] presented a complex DM method based on HMM (Hidden Markov model). The main problem to be solved is to identify audio and video at first, and on this basis, a complete set of complex DM models and algorithms are established. Li et al. [4] put forward an algorithm to smooth the image by using median filter to remove noise and impurities, then use edge detection operator to detect the edge, and design an edge filter to remove the non-text edges. Li et al. [5] designed a stroke filter based on the characteristic of uniform stroke width of text, and on this basis, combined with connected domain analysis or SVM (support vector machine) to locate and verify text blocks. Liu [6] combines color and local texture features, clusters them step by step from top to bottom, and locates the text area by combining connected domain analysis. In order to avoid extracting the subtitle region of every frame in the video, Zhang and Zhu [7] also proposed a text event detection algorithm, which can detect the frames with words and reduce unnecessary computation.

There are many difficulties in locating subtitles in video streams. Chen et al. [8] Analyzed the size change of video stream subtitles. In the same scene video image sequence, large and small subtitles will appear at the same time. The fonts of video subtitles are various, and the text of different languages is also different. Wang et al. [9] believes that even in the same language, there are characters with different shapes. In order to make the fused video produce a reasonable occlusion effect, when only the video stream is used as the input information, the video layering technology can be used maximum line of the foreground in the scene, and even if the object stops moving in a certain sequence, it can keep track of it. At the same time, the segmentation effect of subsequent frames depends on the initial segmentation effect of the first frame. The steps of video automatic processing are relatively complicated, and CAD is relatively mature in technology after years of progress and development. At present, it is widely used in video automatic processing, which plays a very important role in video automatic processing.

## 2 RESEARCH METHOD

### 2.1 Multi-Target Tracking and Segmentation of Video

In the process of analyzing different pixel objects, the different differences between images are compared one by one, and the changes of condition parameters of different sequences are analyzed. Assuming that the local illumination remains unchanged, the parameters of the image remain unchanged. Then, according to the target state of the previous frame, Rajab et al. [10] use the motion continuity of adjacent frames to detect the target of the current frame. In the aspect of automatic video processing, only by combining theory with practice can we design high-quality products. At present, the use of computer-aided technology has brought us many advantages, but it also has some disadvantages. In the design, we must make good use of these advantages brought by computer-aided. The method of this paper is divided into three steps; Motion recognition, motion boundary extraction and contour connection. The flow chart is shown in Figure 1.

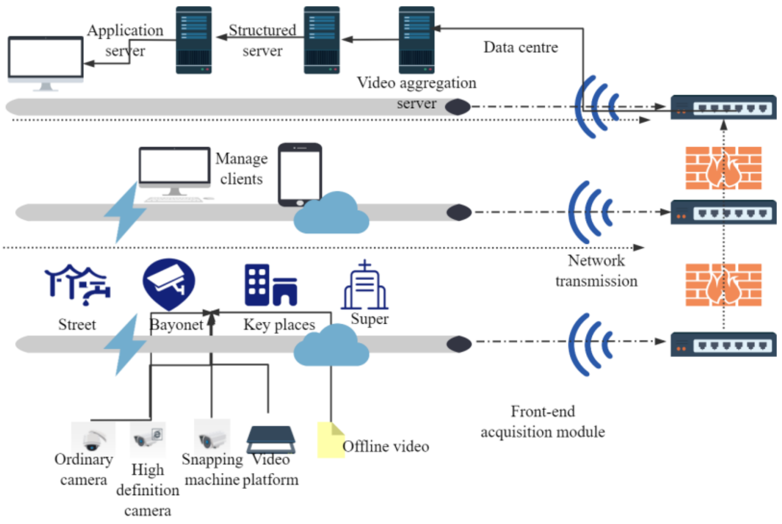


**Figure 1**: Segmentation result graph.

Liu et al. [11] considers the target trajectory as a sequence of positions of tracked features in 2D images or 3D worlds. By collecting the fine image distinguishing features, this paper retains the relevant information that needs to be verified in the next step. In this way, by combining different adjacent pixels, we can finally simulate the super pixels.

The boundary of the object must be continuous, and the piecewise function is used to calculate the cost of the path between the tracking points before and after:

$$D_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

Where $D_{i,j}$ represents the distance between the current tracking point $i$ and its next tracking point $j$.

The video object segmentation method proposed in this chapter is region based. When establishing the data items of energy function, super pixels are used as data points. This can reduce the redundant information of pixels. Anvaripour et al. [12] provided a more convenient image feature calculation method to reduce the amount of calculation for subsequent tasks. It retains the boundary information of the original image, and the segmentation effect is better than the above methods. And the calculation is simple, resources are saved, and the number of finally generated super pixels can be conveniently adjusted. In an image, the basic characteristics of a pixel can be expressed as:

$$f(x,y) = \left[I(x,y), |I_x(x,y)|, |I_y(x,y)|, |I_{xx}(x,y)|, |I_{yy}(x,y)|\right]^T \quad (2)$$

Optical flow method is suitable for accurate analysis, and the motion parameters of the target can be obtained. It is suitable for multi-target motion analysis, which can solve the problems of occlusion overlap that traditional feature-based moving target detection is difficult to solve. Target correlation is a very important problem in multi-target tracking. In multi-target tracking, the detected targets in each frame need one-to-one correspondence, that is, to ensure that:

$$x_k^t = x_k^{t+1} \quad (3)$$

Where $x_k^t$ represents the label of the $k$ th target at time $t$, and the label of each frame should be consistent. The compressed feature similarity is obtained by NB(Naive Bayes) method, and two similarity matrices are obtained. Next, the matching between the target candidate region and the model is carried out by calculating the matching with the maximum similarity.

Here, we first give the specific flow description of multi-target detection algorithm in dynamic background, and then our video experiment. The specific flow description of the algorithm is shown in Figure 2 below:

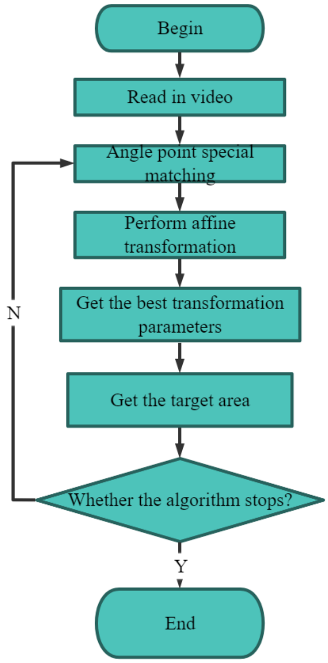


**Figure 2**: Algorithm flow of multi-target detection in dynamic background.

We can see the whole process of multi-target detection algorithm in dynamic background. This algorithm can basically detect multi-targets in dynamic background, and feasibility meets the requirements of the next target tracking. However, this method can only be applied to scenes that are relatively small to the target, and the background should not be too complicated. If there are a lot of real-time changing clutter in the background, or the target is too large, the motion compensation will fail, resulting in the failure of subsequent processing.

## 2.2 Depth-based Virtual Viewpoint Synthesis

Computer-aided demonstration can be used to demonstrate different assembly relationships. If the assembly relationships do not meet the relevant requirements, then the finder can display them, which is convenient for us to modify the wrong parts. Due to the complexity of background images, some background images also show the characteristics of subtitle blocks, and are wrongly judged as subtitle blocks.

The knowledge of graphology shows that subtitles are usually grouped together along the horizontal direction, and color correlation diagram is a way of an image. This feature not only depicts the proportion of pixels in a certain color in the whole image, but also reflects the spatial correlation between different color pairs.

$I$ represents all the pixels of the whole image, and $I_{c(i)}$ represents all the pixels of color $c(i)$. The color diagram can be expressed as:

$$\gamma_{i,j}^{(k)} = \underset{P_1 \in I_{c(i)}, P_2 \in I}{P_r} \left[ |p_1 - p_2| = k \right] \tag{4}$$

The output image of CAD is displayed on a full screen, so the resolution of the output image is the resolution of the display. After the user finishes drawing the pattern structure, the polygon area of each pixel in the output image will be calculated and recorded. Firstly, the objects in 2D plane scene are projected into the real 3D scene through the distance relationship of the depth of objects in the scene, and then the real 3D scene and its relative position relationship with the virtual viewpoint are projected again. Calculate the depth information of each pixel in the image in the actual scene:

$$Z = \frac{Z_{\min} - Z_{\max}}{255} d + Z_{\max} \tag{5}$$

Formula $d$ is the gray value of image depth information, and its value is mapped to the real scene by linear transformation of gray value, ranging from 0 to 255. The farthest scene in the image is represented by the minimum depth value $Z_{\min}$, and the nearest scene in the image is represented by the maximum depth value $Z_{\max}$. $Z$ represents the depth value of the current pixel in the real scene. In two-dimensional space, it is defined as the following formula:

$$G(u,v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(u^2+v^2)/(2\sigma^2)} \tag{6}$$

Where is the filter radius $r^2 = u^2 + v^2$, Different normal distribution results are related to the standard variance state of the surface contour in two-dimensional space. By distinguishing and transforming different non-zero pixels, this paper studies the original matrix graphics. The main block diagram is shown in Figure 3:

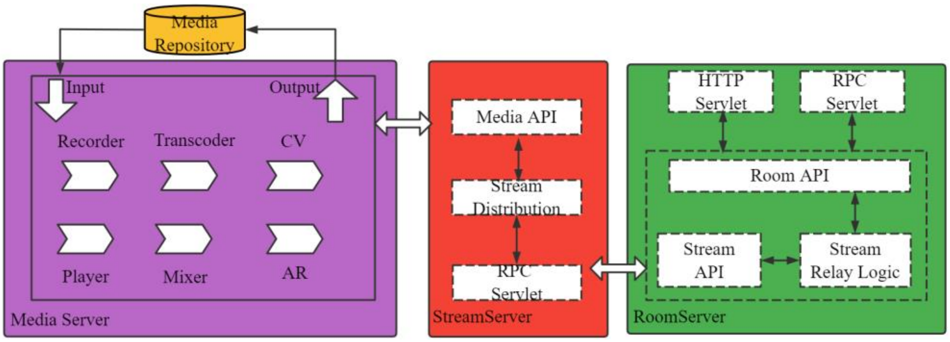


**Figure 3**: 2D to 3D block diagram.

Two video images and depth images to be processed are needed, one tilt image is the current frame, the depth image is its corresponding depth image, and the other is a related post, providing the motion vector of the current frame. Before repairing the edge information of holes, find out the holes that should be repaired first and repair them, and then take the last repaired image as the initial image of this repair every time, and so on.

Video sequence images have the feature of redundancy in time, and moving target tracking based on correlation method just uses this feature of video to find the corresponding relationship between pixels in two adjacent frames of images. The most critical step in correlation tracking is the correlation operation. Only by choosing the appropriate correlation coefficient criterion according to the characteristics of background and target can the target be tracked accurately. The correlation coefficient with the corresponding area in the template image is determined by Formula (7):

$$R(x,y)=\sum_{u=0}^{U}\sum_{v=0}^{V}\left|M(u,v)-S(x+u,y+v)\right| \tag{7}$$

Among them, $0\le x\le X-U+1, 0\le y\le Y-V+1$.

The gray histogram of an image is the basis of various spatial domain processing technologies, and histogram equalization can effectively expand the dynamic range of an image, so that the contrast of the image is enhanced and the details are more vivid. For digital images, the normalized transformation form is as follows:

$$s_k=T(r_k)=\sum_{i=0}^{k}P_r(r_i)=\sum_{i=0}^{k}\frac{n_i}{n}\Big|_{k=0,1,\cdots,L-1} \tag{8}$$

In the formula, $n$ is the sum of the number of pixels in the image; $n_i$ is the number of pixels whose gray level is $r_i$; $P_r(r_i)$ is the probability density of the gray level pixels of $r_i$ accounting for the gray level of the whole image; $s_k$ is the gray level. The gray value of the degree level $r_k$ after the equalization transformation.

Feature-based tracking mainly includes feature extraction and feature matching, and the key lies in feature detection, expression and similarity measurement. The advantage of this method is that even if the target is partially occluded, as long as some features can be seen, the tracking task can be completed. The tracking method based on feature matching is shown in Figure 4:

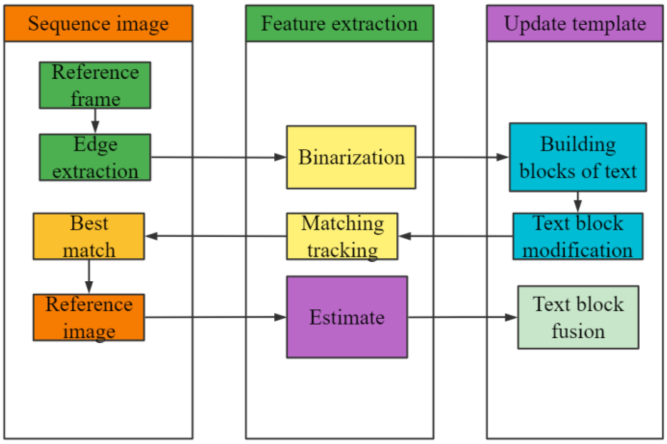


**Figure 4**: Schematic diagram of signature tracking.

Rough detection of text blocks is carried out on edge images, and the detection is based on the arrangement characteristics of text, that is, a text block is composed of multiple characters, and the characters are arranged horizontally or vertically. It is necessary to refine the roof belt in the gradient amplitude image, that is, only the points with the largest amplitude change in the local range are kept, and the points with gradual amplitude change are deleted. This process is called non-maximum suppression, which will generate refined edges. When the number of pixel connections is 1, it can be deleted. This can ensure that the communication characteristics are not eroded.

## 3 RESULT ANALYSIS

We can use an observation sequence for training to adjust the model parameters to obtain a model that can be used to identify other input data. Maximum likelihood estimation training is a general and easy-to-understand technique, but this iterative parameter maximum likelihood estimation can only converge on the local optimum, and in fact, there is no effective method to optimize the model so that the probability obtained under this parameter set can be globally maximized.

In this study, a media library composed of 400 video (with audio) clips is established, of which 300 are used for training models, and the sample data used for training are used to train the corresponding models according to their respective categories. In order to compare with the traditional methods, we separate the video and audio in the samples to train the traditional HMM accordingly. The final test results are shown in Table 1.

| *SNR* | *Video clip HMM* | *Audio frequency HMM* | *Audio and video HMM* |
|---|---|---|---|
| 10 | 88.3% | 98.6% | 98.6% |
| 20 | 86.2% | 95.2% | 97.9% |
| 30 | 87.4% | 97.8% | 98.1% |
| 40 | 87.9% | 98.2% | 96.6% |

**Table 1**: Test result.

It can be seen from Table 1 that the audio-video HMM obtains a higher recognition rate by combining the recognition results of the two input streams, and the recognition rate shows an approximately linear trend due to the change of SNR (signal-to-noise ratio). The use of computer-aided design can also improve the quality of design. There are many advanced technologies in CAD, which makes the design more perfect, thus greatly improving the quality of design.

In the experiment, the starting and ending time of batch processing is calculated, and the average value obtained by removing the processed frames is removed. Rough detection can correctly detect most text blocks in the video, but there are many false alarms in the detection results, and the accuracy rate is not high. Comparing this algorithm with ref [10], the results are shown in Figure 5 and Figure 6. From the comparison results, it can be seen that the two algorithms have no big changes in recall and accuracy, but the time consumption of this algorithm is only 12.5% of that of ref [10].

The ref [10] algorithm expands the edge moderately on each scale edge image in order to overcome the edge breakage, but this aggravates the adhesion, so the candidate text areas given in the rough detection are relatively large, which increases the computation of subsequent processing. For example, when searching connected domain and initially constructing text blocks, the search scope can be greatly reduced after sorting; The existing features are used to screen candidates step by step, and the amount of computation in subsequent processing is reduced.

Sparse optical flow graph can be obtained by sampling the edges and inflection points of objects in the video image with non-uniform hook. The edges and inflection points are obtained from the motion information between the front and back frames. Triangle-based sparse-to-dense conversion method has low computational complexity and less memory requirements. However, when the sampled image points can't accurately approach the grid of the image, this method will easily lead to delineating the wrong object edges. In addition, dense features are easily affected by the error that sparse points are located at the edge to be interpolated.

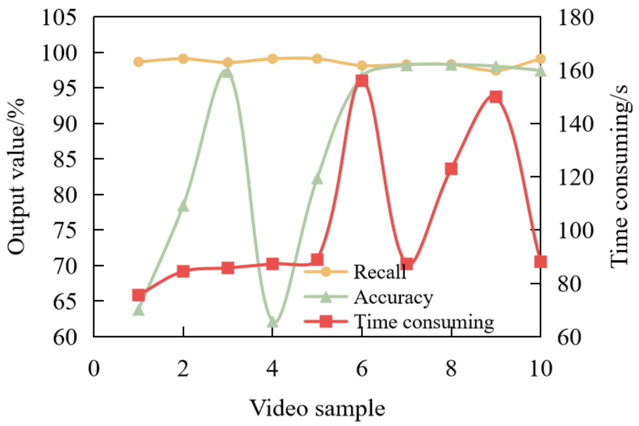


**Figure 5**: The performance of each link of text location algorithm in this paper.

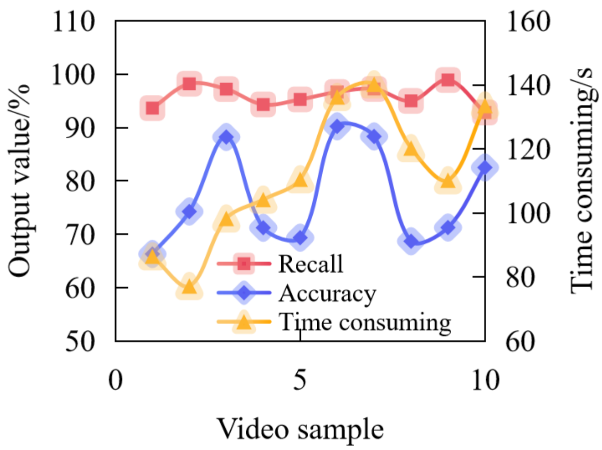


**Figure 6**: Performance of each link in text location of ref [10] algorithm.

At present, many depth map generation algorithms are based on a single depth cue dimension to obtain depth maps. These methods can achieve good results for a specific type of video image information, but there are so many types of scenes in 2D video images that it is difficult to find an algorithm that can generally adapt to all types of scene information. Finally, it is convenient to convert motion information into depth information, with simple conversion principle and low computational complexity. In order to enhance the depth perception of moving objects, moving objects are used as foreground and other information as background.

Perform scene processing on the ideal original image and the original image respectively, so that it has too dark, too bright and dim scene effects. Then, it is enhanced by various algorithms, and following Figure 7 and Figure 8.

By comparing the parameters of the dark scene images, it can be found that the adaptive enhancement algorithm to other algorithms in all performance indexes, and has excellent performance in image contrast, detail expression and so on. Only the adaptive enhancement algorithm based on dynamic scene estimation proposed in this paper has the best data performance in all scenes and the strongest adaptability. Most of the other algorithms have good results in dealing with a specific scene, but they are unsatisfactory in other situations.

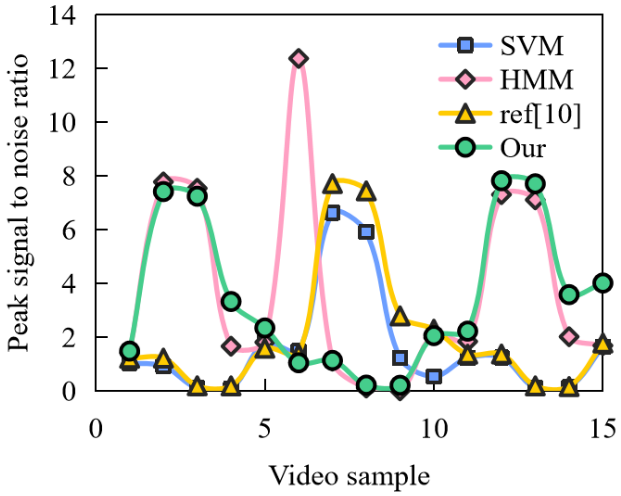


**Figure 7**: Comparison of the processing effects of enhancement algorithms in the scene of passing the scene.

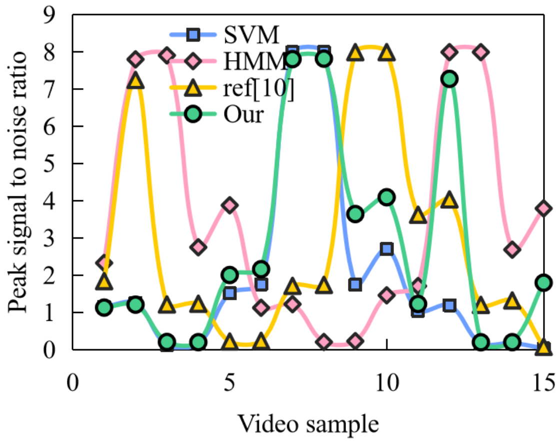


**Figure 8**: Compare the processing efficiency of enhancement algorithms in over-bright scenes.

For image enhancement algorithms, only data comparison can't fully explain the processing effect of the algorithm, and the visual perception of human eyes is often a more important evaluation method. Therefore, the design of video multi-target tracking processing software is very important. Through the software, we can not only obtain the overall tracking result of target tracking, but also display a single target separately, and make further statistics on the time of each link of the algorithm. It takes a long time to repair a single image; The time complexity of layered hole repair algorithm is low, but the foreground edge has certain blurring effect. Depth adaptive cavity repair algorithm can effectively eliminate the fuzzy effect of layered cavity repair.

The enhanced text block is binarized and connected domain calibrated, features are extracted, connected domain is screened, interference is removed, and a clean binary image with definite polarity is generated. When judging the polarity, if the judgment is wrong, the correct recognition result can't be obtained, all the previous processes will be wasted. When screening the connected domain, if the interference connected domain error in the background is retained. However, they are far apart in RGB space, so they are classified into different categories during clustering. It is very difficult to extract the features of characters from the clustering results. Through the model analysis of video under the Internet. The model graph of spatial surface is transformed.This phenomenon is called "color transition". This problem is discussed in literature, but there is no mature solution.

## 4 CONCLUSION

At present, CAD develops rapidly with the changes of society, so designers must update themselves in time according to the constant changes of design techniques and design concepts, so that they can master new techniques and concepts. And internet plus IoT are increasingly discussed by the industry. However, as terminal users and manufacturing enterprises, they still need to start from their own actual business, because both IoT and Internet are tools. Based on the traditional decision theory algorithm, this paper determines that the pixel to be processed and the median pixel are potential noise points. The reliability of target association is increased, and at the same time, the space-time constraints and optical flow tracking are applied, which can match the target very accurately, and it is robust to the occlusion and deformation of the target. From the comparison results, it can be seen that there is no big change in the recall rate and accuracy rate of the two algorithms, but the time consumption of this algorithm is only 12.5% of that of the traditional algorithm. The subjective processing effect and objective data index of the algorithm are better than those of similar algorithms.

## 5 ACKNOWLEDGEMENT

This work was supported by Fundamental Research Funds for the Universities of Hebei Province Education Department: Research on classroom behavior recognition under the goal of improving teaching effect (No. JYT2020046); Hebei higher education teaching reform and research practice project: Construction and practice of PBL teaching project library under the background of curriculum ideology and Politics -- Taking the course of programmable logic devices and VHDL design as an example (No. 2020GJJG601).

*Shihui Zhang*, https://orcid.org/0000-0002-7104-6861
*Xiqing Zhao*, https://orcid.org/0000-0001-9098-7168
*Naidi Liu*, https://orcid.org/0000-0001-7098-5632
*Huanhuan Gao*, https://orcid.org/0000-0001-8970-7325

## REFERENCES

[1] Paterson, L.; May, F.; Andrienko, D.: Computer aided design of stable and efficient oleds, Journal of Applied Physics, 128(16), 2020, 160901. https://doi.org/10.1063/5.0022870

[2] Keye, S.; Gammon, M.-R.: Development of deformed computer-aided design geometries for the sixth drag prediction workshop, Journal of Aircraft, 55(4), 2018, 1401-1405. https://doi.org/10.2514/1.C034428

[3] Santos, I.; Castro, L.; Rodriguez, F.-N.: Artificial neural networks and deep learning in the visual arts: A review, Neural Computing and Applications, 33(1), 2021, 121-157. https://doi.org/10.1007/s00521-020-05565-4

[4] Li, W.; Qi, L.; Guo, Y.; Zhang, Z.; Wang, Z.: Application value of cta in the computer-aided diagnosis of subarachnoid hemorrhage of different origins, Journal of Healthcare Engineering, 2021(1), 2021, 1-8. https://doi.org/10.1155/2021/6638610

[5] Li, H.; Zhang, H.; Zhao, Y.: Design of computer-aided teaching network management system for college physical education, Computer-Aided Design and Applications, 18(S4), 2021, 152-162. https://doi.org/10.14733/cadaps.2021.S4.152-162

[6] Liu, X.: Training strategies for practical ability of college students majoring in computer-aided design, International Journal of Emerging Technologies in Learning (iJET), 15(16), 2020, 134. https://doi.org/10.3991/ijet.v15i16.15935

[7] Zhang, X.; Zhu, X.: Efficient and de-shadowing approach for multiple vehicle tracking in aerial video via image segmentation and local region matching, Journal of Applied Remote Sensing, 14(1), 2020, 1. https://doi.org/10.1117/1.JRS.14.014503

[8] Chen, Y.; Li, Y.; Wang, J.: Remote aircraft target recognition method based on superpixel segmentation and image reconstruction, Mathematical Problems in Engineering, 2020(1), 2020, 1-9. https://doi.org/10.1155/2020/608768

[9] Wang, C.; Pei, J.; Wang, Z.; Huang, Y.; Yang, J.: When deep learning meets multi-task learning in sar atr: simultaneous target recognition and segmentation, Remote Sensing, 12(23), 2020, 3863. https://doi.org/10.3390/rs12233863

[10] Rajab, K.-Z.; Wu, B.; Alizadeh, P.; Alomainy, A.: Multi-target tracking and activity classification with millimeter-wave radar, Applied Physics Letters, 119(3), 2021, 034101. https://doi.org/10.1063/5.0055641

[11] Liu, J.; Gong, S.; Guan, W.; Li, B.; Liu, J.: Tracking and localization based on multi-angle vision for underwater target, Electronics, 2020, 9(11), 1871. https://doi.org/10.3390/electronics9111871

[12] Anvaripour, M.; Saif, M.; Ahmadi, M.: A novel approach to reliable sensor selection and target tracking in sensor networks, IEEE Transactions on Industrial Informatics, 16(1), 2020, 171-182. https://doi.org/10.1109/TII.2019.2916091