

## Research Article

# Visual Sequence Algorithm for Moving Object Tracking and Detection in Images

Renzheng Xue <sup>1</sup>, Ming Liu <sup>1</sup> and Xiaokun Yu <sup>2</sup>

<sup>1</sup>School of Computer and Control Engineering, Qiqihar University, Qiqihar, Heilongjiang 161006, China

<sup>2</sup>Department of Computer Science, Heilongjiang Communications Polytechnic, Qiqihar, Heilongjiang, China

Correspondence should be addressed to Renzheng Xue; 201812210601036@zcmu.edu.cn

Received 15 October 2021; Revised 21 November 2021; Accepted 29 November 2021; Published 27 December 2021

Academic Editor: Yuvaraja Teekaraman

Copyright © 2021 Renzheng Xue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Objective.** The effects of different algorithms on detecting and tracking moving objects in images based on computer vision technology are studied, and the best algorithm scheme is confirmed. **Methods.** An automatic moving target detection and tracking algorithm based on the improved frame difference method and mean-shift was proposed to test whether the improved algorithm has improved the detection and tracking effect of moving targets. The algorithm improves the traditional three-frame difference method and introduces a single Gaussian background model to participate in target detection. The improved frame difference method is used to detect the target, and the position window and center of the target are determined. Combined with the mean-shift algorithm, it is determined whether the template needs to be updated according to whether it exceeds the set threshold so that the algorithm can automatically track the moving target. **Results.** The position and size of the search window change as the target location and size change. The Bhattacharyya similarity measure  $\rho(y)$  exceeds the threshold  $r$ , and the target detection algorithm is successfully restarted. **Conclusion.** The algorithm for automatic detection and tracking of moving objects based on the improved frame difference method and mean-shift is fast and has high accuracy.

## 1. Introduction

With the rapid development of the economy and the accelerating process of urbanization, a large number of people flock to cities, which makes the traffic safety and social security problems gradually show a complicated growth trend. Social organizations are increasingly dependent on video surveillance. However, traditional video surveillance only stores the image information collected by cameras. Normally, image information is not analyzed by traditional video surveillance, and a lot of subsequent analysis work is left to the staff. Workers working in this environment for a long time will be fatigued and inefficient. Large monitoring errors will occur, and sometimes the staff will even ignore important information, resulting in irreparable losses. In addition, video monitoring is carried out manually, which has the risk of revealing the privacy of others. When there is a safety problem, we can only watch the surveillance video, which delays the timely handling of the incident.

With the rapid development of computer hardware and software technology, the computer is running faster and faster, which makes the algorithms widely used in the fields of digital image processing, computer vision, artificial intelligence, and so on. In intelligent video surveillance, by adopting computer vision and artificial intelligence algorithms, the computer can be like a human, which can understand the content in the image and analyze and judge the content of the monitoring screen. And when an abnormal situation occurs, it can be judged in advance and reported to the staff quickly, thus helping the staff to deal with the abnormal situation in time, improving the efficiency of event handling and reducing the workload of the staff. Computer vision plays an important role in real life, and it is widely used in many fields.

Campbell proposed the optical flow method, which is characterized by an instantaneous rate when the target moves pixels in the image area. In addition, the optical flow method studies the changes of gray value in the time

domain, which can be roughly divided into three categories: matching, frequency domain, and gradient. The matching-based approach is to define velocity as displacement. The method based on the frequency domain makes use of the frequency and phase information of different output filters. Although this method has high accuracy, it requires a large number of calculations [1]. In the study carried out by Defaria et al., the gradient-based method uses the gray time difference of an image sequence to calculate the speed of each pixel point. The algorithm is simple and effective, but the parameters have a great influence on the results of practical application, and there are also some problems. The advantages of the optical flow method are target detection and strong adaptability in any scene, but the disadvantages are that noise, occlusion, lighting, and other factors will affect the results, and the calculation is large, requiring good hardware support [2]. Liddo et al. proposed the mixed Gaussian model method for background modeling in the VSAM system and applied it in more complex scenes, such as shaking water waves, sudden changes in light intensity, storms, and other harsh weather conditions. As the updating speed of the dynamic background model is variable and fast, foreground objects may appear in the background; otherwise, some background objects in the background will be removed [3].

In all these application fields, the detection and tracking of moving objects is an essential link. It is an important cornerstone of subsequent target recognition and image understanding, and its results will directly affect the subsequent processing. Due to the increasing ability of computer processors to process data, many algorithms in computer vision have been realized and are widely used. Experts from all over the world have carried out a lot of research on them and put them into practical products. As an important research direction of computer vision, in the fields of video surveillance, intelligent temperature monitoring systems, and early warning systems, it has been continuously improved and applied by researchers and engineers, so it has gained more room for development. The rapidity and accuracy of moving target detection directly affect the subsequent target tracking and recognition.

Based on this, this paper puts forward the computer vision technology and different algorithms on the image moving target detection and tracking effect, which confirms the optimal algorithm scheme. Based on the improved frame difference method and mean-shift automatic detection and tracking algorithm of moving targets, check whether the improved algorithm has improved the detection and tracking effect of moving targets.

## 2. Target Tracking Algorithm

- (1) Tracking method based on contour: according to the energy of the contour curve, the contour curve will gradually approach the actual contour edge of the moving target under the joint action of external force and binding force. Only when the energy of the contour curve of the object edge reaches the minimum, the real contour of the moving target can be

obtained. It is difficult to get the precise initialization contour of the moving object; that is, it is very difficult to initialize the boundary contour. Because it is very difficult to update the contour of the moving target in the tracking process, the effect of this method is not good when the tracking target has an occlusion problem [4].

- (2) Model-based tracking method: in this method, an appropriate geometric model is first established, and then model matching is carried out to realize the tracking process, and the model should be updated with timely type parameters. There are three models of moving objects: line model, 2D contour model, and 3D stereo model. The line graph model is a simple model to describe real objects; 2D projection models are very intuitive; the 3D model is relatively complex, but the model makes full use of the essential characteristics of the target object, making the tracking process more stable, especially in the complex environment. The advantage is more obvious. Compared with other tracking methods, it has higher reliability. The shortcoming lies in the large amount of calculation, which is time-consuming and difficult to meet the requirements of real-time [5].
- (3) Region-based tracking method: the main idea of the region tracking method is to take the common features of the connected region of the moving target as the tracking basis, such as the target texture information, color information, and so on. The selected feature information is called the reference template. The reference template determined by the first frame of the video sequence is defined as the target template, and the correlation between the target template and the candidate template in the subsequent video sequence is calculated to determine the location of the target, so as to complete the target tracking process. In practical applications, especially in complex environments, multiple features are often selected for tracking. Because the feature information based on this method is not exactly the same, the tracking accuracy and stability are better when the target object has no occlusion problem, but when the moving target has shadow and occlusion problems, the tracking processing is still difficult [6].

### 2.1. Detection and Extraction of Moving Objects by the Improved Frame Difference Method

*2.1.1. Single Gaussian Model to Extract Static Background.* When the background environment is simple and the camera is fixed, the brightness value of the background pixels of each frame captured by the camera is near a relatively stable value, which accords with Gaussian distribution [7]. In a period of time, each pixel obeys a Gaussian distribution with mean value  $\mu$  and standard variance  $\sigma$ , and the Gaussian distribution of each pixel is independent. Therefore, a Gaussian model can be established for each pixel as follows:

$$P(i_{x,y}) = \frac{1}{\sqrt{2\pi\delta_{x,y}^2}} e^{-\left(\frac{i_{x,y} - \mu_{x,y}}{\delta_{x,y}}\right)^2 / 2}, \quad (1)$$

where  $i_{x,y}$  represents the value of a certain pixel;  $(x, y)$  indicates the coordinates of the pixel in the frame [8];  $\mu_{x,y}$  represents the mean value of the Gaussian model of this pixel;  $\delta_{x,y}$  is the variance corresponding to the pixel; and  $P(i_{x,y})$  indicates the probability that the pixel value of this point is  $i_{x,y}$ .

When no moving object enters, the value  $P(i_{x,y})$  of each pixel changes within a stable level, so it can be determined that these pixels are all background pixels [9]. When a moving target enters, the changes of pixels in the area where the target is located do not conform to Gaussian distribution, so these points can be judged as foreground pixels. The mean and standard variance of the set of pixels  $(x, y)$  in  $T$  frame are  $\mu_i$  and  $\sigma_i$ .

$$\mu_i = \frac{1}{T} \sum_{i=0}^{T-1} g_i, \quad (2)$$

$$\sigma_i = \sqrt{\frac{1}{T} \sum_{i=0}^{T-1} (g_i - \mu_i)^2},$$

where  $g_i$  is the gray level of any pixel  $(x, y)$  in the  $i$ th frame. If  $G(x, y)$  is the gray value of any point in the video sequence,  $k$  is an empirical value between 2 and 3. When  $|g_{i+1}(x, y) - \mu_i(x, y)| \geq k\sigma_i(x, y)$ , it is considered as the front scenic spot [10]. When  $|g_{i+1}(x, y) - \mu_i(x, y)| < k\sigma_i(x, y)$ , this point is considered as the background point. The extracted background frame is marked as  $\text{Dif}_{p_M}$ .

**2.1.2. Detection of Moving Targets.** In this paper, the improved five-frame difference algorithm is adopted, and two frames separated by one in five consecutive frames are used for frame difference. The obtained frame difference is ‘‘AND’’ with the background frame, and then the three groups of results are ‘‘OR,’’ so that the moving target can be detected more completely [11]. Let the five consecutive frames in the video image sequence be as follows:  $p_1, p_2, p_3, p_4$ , and  $p_5$ , in which  $p_3$  is an intermediate frame. The result obtained by using two frames separated by one frame as the frame difference is as follows:

$$\begin{aligned} \text{Dif}_{31} &= |p_3 - p_1|, \\ \text{Dif}_{42} &= |p_4 - p_2|, \\ \text{Dif}_{53} &= |p_5 - p_3|. \end{aligned} \quad (3)$$

Then, the three-frame difference results  $\text{dif}_{p_{31}}, \text{dif}_{p_{42}}$ , and  $\text{dif}_{p_{53}}$  are, respectively, operated with the frame difference  $\text{Dif}_{p_M}$  between the intermediate frame and the background frame, which is as follows:

$$\begin{aligned} D_1 &= \text{Dif}_{p_{31}} \cap \text{Dif}_{p_M}, \\ D_2 &= \text{Dif}_{p_{42}} \cap \text{Dif}_{p_M}, \\ D_3 &= \text{Dif}_{p_{53}} \cap \text{Dif}_{p_M}. \end{aligned} \quad (4)$$

At last,  $d_1, d_2$ , and  $d_3$  are OR-operated to get  $d$ , which is the detected moving target. There are some small holes in the

moving target detected by the abovementioned improved method, which can be eliminated by the expansion and corrosion method of binary image in image morphology [12].

**2.1.3. Extracting Moving Objects.** When a moving target is detected, the moving area of the target can be determined. In order to extract the moving target, it is necessary to remove a large number of shadows contained in the moving target area, which seriously interferes with the edge detection of the moving target [13]. Shadow detection methods are based on shadow features and geometric models. Among them, the detection method based on shadow feature is to process the image by detecting the geometric feature, color, and brightness of shadow [14]. Because the existence of shadows only affects the brightness of pixels, it does not have much influence on the color of pixels. In this paper, a shadow detection method based on HSV color space and edge detection are used to extract moving objects. First of all, the edge detection method is used to get the edge of the target moving area. Then, the shadow of the moving area of the target is detected by the HSV method, and the edge of the shadow part is obtained by edge detection. Finally, accurate target edge information can be obtained by subtracting the shadow edge from the edge of the target moving area.

There are two main approaches to visual tracking: one is the bottom-up approach, and the other is the top-down approach. This paper adopts a bottom-up approach, combined with the theory of visual computing, which can be divided into three stages: layer vision, middle layer vision, and high-level vision. Low level vision to middle level vision is an image feature description, while middle level vision to high level vision is a 2.5-dimensional description, and high level vision is a 3 dimensional description.

## 2.2. Mean-Shift Target Tracking

**2.2.1. Overview of Mean-Shift Algorithm.** The mean-shift algorithm is a semi-automatic tracking method, which selects the moving target by manually determining the search window in the initial tracking frame [15]. The histogram distribution of the search window weighted by the kernel function is calculated, and the histogram distribution of the corresponding window of the current frame is calculated by the same method. Based on the principle of maximum similarity between the two distributions, the search window moves to the real position of the target in the direction of maximum density increase. For a finite set  $a$  in  $n$ -dimensional Euclidean space  $x$ , the mean-shift form at  $x \in X$  is as follows:

$$M_h(x) = \frac{\sum_{i=1}^n G((x_i - x)/h) \omega(x_i) (x_i - x)}{\sum_{i=1}^n G((x_i - x)/h) \omega(x_i)} - x. \quad (5)$$

Given an initial point  $x$ , the kernel function  $G(x_i)$ , and the weight function  $\omega x_i$ , the allowable error is  $\xi$ . The first term on the left side of the formula is defined as  $m_h(x)$ .

$$m_h(x) = \frac{\sum_{i=1}^n G((x_i - x)/h)\omega(x_i)(x_i - x)}{\sum_{i=1}^n G((x_i - x)/h)\omega(x_i)}. \quad (6)$$

At first, calculate  $m_h(x)$ , and iterate the calculation result of  $m_h(x)$  with  $x$ ; if  $\|m_h(x) - x\| < \xi$ , stop the iterative process.

*2.2.2. The Improved Frame Difference Method Is Combined with the Mean-Shift Algorithm.* In order to solve the problem that mean-shift must enter the search window manually, the target edge extracted by the improved frame difference method is used to update the mean-shift search window [16]. Assuming that the bandwidth  $h$ ,  $x_1, x_2, \dots, x_n$  of the initial kernel window are the sampling points included in the kernel window and  $x_0$  is the center of the target, the histogram distribution of the target is as follows:

$$\hat{q}_u = C \sum_{i=1}^n k\left(\left\|\frac{(x_i - x_0)}{h}\right\|\right) \delta[b(x_i) - u], \quad u = 1, 2, \dots, m, \quad (7)$$

where  $k$  is the kernel function;  $M$  is the number of eigenvalues in the feature space;  $\delta$  is the Kronecker function;  $B(x_i)$  is the characteristic value corresponding to pixel  $x_i$ ;  $C$  is the normalized coefficient; and  $H$  is the bandwidth of the kernel function. Therefore, object tracking can be simplified as finding the optimal  $y$ , so that  $p_u(y)$  is most similar to  $\hat{q}_u$ . The similarity between  $p_u(y)$  and  $\hat{q}_u$  is measured by the Bhattacharyya coefficient, which is as follows:

$$\rho(y) = \sum_{u=1}^m \sqrt{\hat{p}_u(y)\hat{q}_u}. \quad (8)$$

The traditional mean-shift method does not update the template when calculating the distribution of target histogram  $q_u$ , but always uses the template established by the initial frame, which leads to the deviation between the histogram distribution of the target template and  $q_u$ , which easily leads to “following and losing” of the target and cannot adapt to the size change of the target. In this paper, the Bhattacharyya similarity measure coefficient  $\rho(y)$  is introduced to judge whether the template needs to be updated. When tracking the target, if the variation of  $\rho(y)$  exceeds a certain threshold value  $r$ , it means that the template deviation is too large. At this time, the improved frame difference method is rerun to extract the moving target, update the size and position of the tracking window, and update the histogram distribution  $q_u$  of the target [17].

Image segmentation is an important method in computer moving target tracking. Aiming at the vehicle tracking problem, Robert edge detection operator is used to detect the edge of the target vehicle, which greatly improves the segmentation accuracy. In addition, we propose a vehicle target segmentation strategy with the maximum interclass variance, and the experimental results show that the algorithm has a good segmentation effect.

### 3. Results

In this paper, a video taken indoors is used as the video source to verify the effectiveness of the algorithm. The algorithm is run on MATLAB to verify the effectiveness of the improved frame difference method by observing whether a large number of holes appear in the result of target detection. The tracking effect can be judged by observing whether the size and center position of the search box change in real time with the change of the target [18]. This is to verify that the Bhattacharyya similarity measure coefficient  $\rho(y)$  has exceeded the threshold  $r$  and to determine whether it has successfully restarted the target detection algorithm and updated the search window. If the target detection result is clear and no void appears, and the size and position of the tracking window of the mean-shift algorithm can be updated in real time according to the change of the target object, the validity of the algorithm can be verified. The experiment steps are as follows:

- (1) Read the video sequence.
- (2) Extract static background and five consecutive images  $p_1, p_2, p_3, p_4$ , and  $p_5$ .
- (3) Make the frame difference between the intermediate frame and the extracted static background to obtain  $\text{Dif}_{p_M}$ .
- (4) Use the 5-frame difference algorithm mentioned above, the moving object  $D$  is detected.
- (5) The HSV transform is carried out on the detected target area, and combined with edge detection information, the shadow is eliminated, and the target frame and center point are extracted.
- (6) Take the target frame and center point obtained in step (5) as the template of the mean-shift algorithm.
- (7) Track the target frame by frame for subsequent frames, and judge whether the similarity measurement coefficient  $\rho(y)$  of Bhattacharyya exceeds the threshold. If it does not exceed the threshold, continue tracking until the end of the video frame. If it exceeds the given threshold, proceed to step (2) from the current frame, and redetect and obtain the target window. The experimental results of target detection are shown in Figures 1(a)–1(d).

As can be seen from Figure 1, the image processed by the traditional three-frame difference method has obvious holes and a blurred target contour. The image contour obtained by the improved algorithm is clear, which shows that the algorithm has high extraction accuracy and ideal effect. The experimental results of target tracking are shown in Figures 2(a)–2(d). Figures 2(a) to Figure 2(d) are tracking effect diagrams of different frames. As shown in Figure 2, with the change of the target position and size, the position and size of the search window also change. It shows that the variation of the Bhattacharyya similarity measure coefficient  $\rho(y)$  exceeds the threshold  $R$ , and the target detection

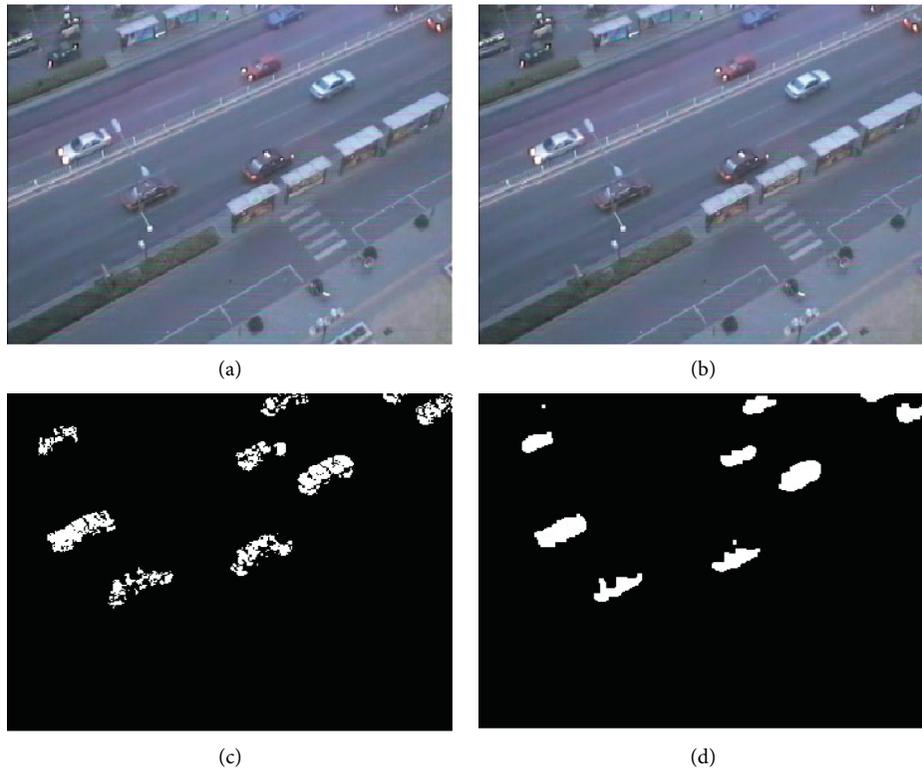


FIGURE 1: Test results of target detection. (a) 20th frame of original video, (b) 20th frame of original video, (c) traditional three-frame difference, and (d) improved results.

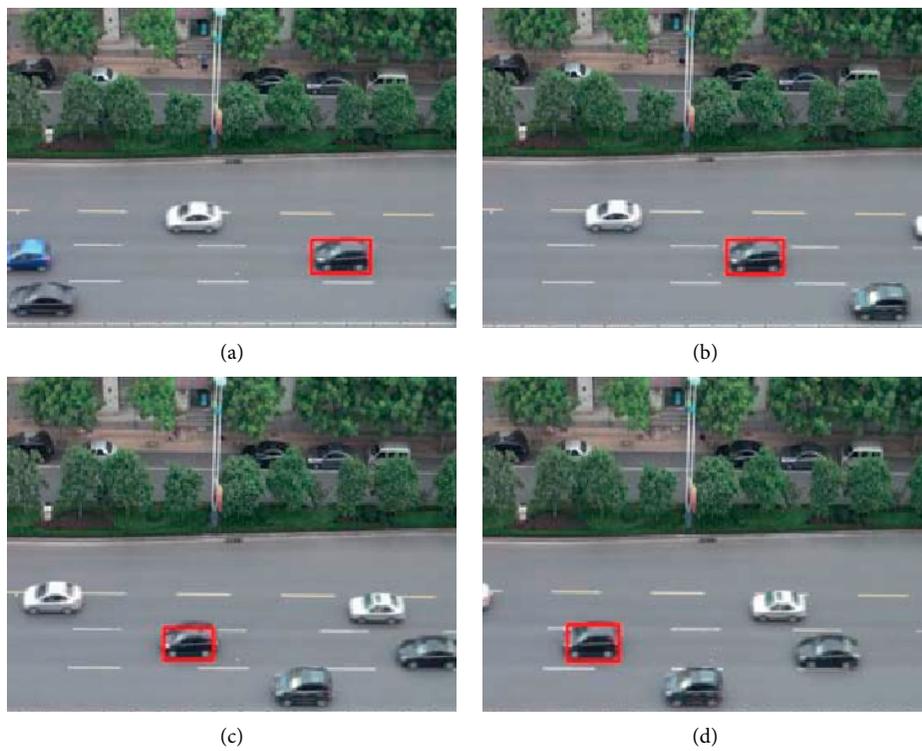


FIGURE 2: Target tracking results.

algorithm is successfully restarted. In the whole tracking process, there is no “follow-up” phenomenon. Besides, no one is involved in the whole process, which realizes the automatic tracking of the target.

The key of video surveillance is the detection and tracking of moving objects. Broadly speaking, this technology combines the methods of image recognition and target tracking to detect moving objects in an image sequence on the basis of analyzing static and dynamic images. In a static image, the distribution of information is constant in time and variable in space (it changes with space). And the continuous static image sequence reflects the dynamic image changes in time. Therefore, the distribution characteristics of the spatial location information density of dynamic images also change with time. Moving target detection is realized by using the time correlation between continuous frames of a static image and a dynamic image. Firstly, the moving information of the target in the dynamic image is detected, and then the following target recognition and tracking process is completed. The so-called moving target detection is to extract and separate the motion information of interest to researchers from the original video stream and separate the information features of moving objects from the background. This is regarded as the first step in the application of image analysis, intelligent monitoring, and other technologies. This paper only focuses on target tracking in visible light, and there is no indepth study on other cases. For example, targets at night will be affected by light changes, which may involve the processing of targets in infrared video images; night conditions require the system to have higher robustness, in addition to the influence of climate conditions.

#### 4. Conclusion

Based on computer vision technology, in the process of detecting and tracking the moving object in the image, the improved frame difference method and mean-shift algorithm combined with the moving object detection and tracking algorithm can effectively overcome the shortcomings of the traditional frame difference method and mean-shift algorithm under the condition of static background, the position and size of the search window change as the target location and size change. The Bhattacharyya similarity measure  $\rho(y)$  exceeds the threshold  $r$ , and the target detection algorithm is successfully restarted. The method adopted by the program is simple, and the calculation speed is fast. It can meet the requirements of real-time detection and tracking.

#### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### Acknowledgments

This study was supported by Heilongjiang Province Education Department Scientific Research Project (Project no. 135409224).

#### References

- [1] M. D. Campbell, “Understanding and quantifying bias in visual fisheries surveys using advanced technology,” *Fisheries*, vol. 44, no. 5, pp. 212–215, 2019.
- [2] M. Defaria, D. Oliveiradantas, M. Ruizalves, B. Demedeiros, and L. Docarmo, “Visual acuity screening, photoscreening and dispensing of glasses with ready to clip technology,” *Revista Brasileira de Oftalmologia*, vol. 79, no. 5, pp. 309–314, 2020.
- [3] A. D. Liddo, N. P. Souto, and B. Plüss, “Let’s replay the political debate: hypervideo technology for visual sense-making of televised election debates,” *International Journal of Human-Computer Studies*, vol. 145, no. 1, Article ID 102537, 2020.
- [4] U. Karamchandani, S. Erridge, K. Evans-Harvey, A. Darzi, J. Hoare, and M. Sodergren, “1190 eye-tracking technology differentiates visual gaze patterns between trainee endoscopists according to a validated objective skills assessment scale,” *Gastrointestinal Endoscopy*, vol. 91, no. 6, pp. AB113–AB114, 2020.
- [5] Y. B. Blokhinov, V. A. Gorbachev, A. D. Nikitin, and S. V. Skryabin, “Technology for the visual inspection of aircraft surfaces using programmable unmanned aerial vehicles,” *Journal of Computer and Systems Sciences International*, vol. 58, no. 6, pp. 960–968, 2019.
- [6] Y.-h. Pan, “On visual knowledge,” *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 8, pp. 1021–1025, 2019.
- [7] Y. Han, A. Wu, L. Zhu, and Y. Yang, “Visual commonsense reasoning with directional visual connections,” *Frontiers of Information Technology & Electronic Engineering*, vol. 22, no. 5, pp. 625–637, 2021.
- [8] J.-z. Xia, Y.-h. Zhang, H. Ye et al., “Supoolvisor: a visual analytics system for mining pool surveillance,” *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 4, pp. 507–523, 2020.
- [9] J. Lu, Y. Peng, G.-J. Qi, and J. Yu, “Guest editorial introduction to the special section on representation learning for visual content understanding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 9, pp. 2797–2800, 2020.
- [10] T. Gorichanaz, “Understanding and information in the work of visual artists,” *Journal of the Association for Information Science and Technology*, vol. 71, no. 6, pp. 685–695, 2020.
- [11] M. Chegini, J. Bernard, J. Cui et al., “Interactive visual labelling versus active learning: an experimental comparison,” *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 4, pp. 524–535, 2020.
- [12] S.-w. Li, Q.-b. Jiang, Q.-j. Zhao, L. Lu, and Z.-l. Feng, “Asymmetric discriminative correlation filters for visual tracking,” *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 10, pp. 1467–1484, 2020.
- [13] G. E. Lancioni, M. F. O’Reilly, J. Sigafos et al., “Smartphone-based technology to support functional occupation and mobility in people with intellectual disability and visual impairment,” *Advances in Neurodevelopmental Disorders*, vol. 3, no. 3, pp. 334–342, 2019.

- [14] Y. Xie, J. Xiao, K. Huang, J. Thiyagalingam, and Y. Zhao, "Correlation filter selection for visual tracking using reinforcement learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 192–204, 2020.
- [15] Y. Xie, J. Xiao, K. Huang, J. Thiyagalingam, and Y. Zhao, "Correlation filter selection for visual tracking using reinforcement learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 192–204, 2020.
- [16] H.-W. Chae, J.-H. Choi, and J.-B. Song, "Robust and autonomous stereo visual-inertial navigation for non-holonomic mobile robots," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9613–9623, 2020.
- [17] P. Spagnolo, H. Aghajan, G. Bebis et al., "Guest editorial introduction to the special issue on large-scale visual sensor networks: architectures and applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 4, pp. 1249–1252, 2021.
- [18] L. Deng, M. Yang, Z. Liang, Y. He, and C. Wang, "Fusing geometrical and visual information via superpoints for the semantic segmentation of 3d road scenes," *Tsinghua Science and Technology*, vol. 25, no. 4, pp. 498–507, 2020.