A Survey of Moving Target Detection Methods Based on Machine Vision

Tao Hu
College of Mechanical and Automotive Engineering
Shanghai University of Engineering Science
Shanghai China
507926646@qq.com

Jian Mao
College of Mechanical and Automotive Engineering
Shanghai University of Engineering Science
Shanghai China
(Corresponding author):jmao@sues.edu.cn

ABSTRACT

In recent years, with the development of relevant technologies in the field of machine vision, the processing of visual image information has become the focus of research. Among them, the detection of moving targets is a very important research direction in the field of machine vision, which lays a foundation for the recognition of moving targets and tracking of moving targets. The task of moving target detection is to identify the physical movement of the target in a specific area. In this paper, the relevant image processing techniques used in the process of moving target detection are briefly described, including image preprocessing, image segmentation, feature extraction and so on. Then it describes the algorithms commonly used for moving target detection in recent years, including background difference method, inter-frame method, optical flow method, and compares the advantages and limitations of these methods. In view of the shortcomings of these methods, it summarizes the previous solutions. Finally, the improvement of these algorithms in recent years is pointed out.

CCS CONCEPTS

•Computing methodologies~Artificial intelligence~Computer vision

KEYWORDS

Machine vision, Moving target detection, Image segmentation, Feature extraction, Survey

1 Introduction

The general process of machine vision system is to automatically acquire the desired target image, extract the corresponding features of the target image, analyze and process it, and finally make some decisions based on the results. Moving target detection technology is one of the functions of machine vision system, which means that the desired moving target can be segmented from the surrounding environment for a given video sequence. For any video analysis system, there are three key processes, namely, moving target detection, target tracking and behavior analysis. Moving target detection is the basis of target tracking and behavior analysis, and the detection results will directly affect the subsequent processing, so moving target detection is very important. At present, there is no general moving target detection algorithm that can be applied to all situations, and appropriate algorithms need to be selected for some specific environments. At the same time, each algorithm has its own advantages and disadvantages and application scope. At present, the general motion detection algorithms include background difference method, interframe method and optical flow method. This paper briefly describes some basic processes of image processing, then summarizes these traditional algorithms, and compares the three methods, points out the advantages and limitations of these methods, as well as previous solutions. Finally, the improvement of these motion detection methods in recent years is summarized.

EITCE 2020, November 6–8, 2020, Xiamen, China

ACM ISBN 978-1-4503-8781-1/20/11…$15.00
https://doi.org/10.1145/3443467.3443810

2 Preprocessing of machine vision images

Image preprocessing plays a key role in the process of machine vision image target recognition.Usually, the image acquired by the moving target detection system has some noise, or the image quality is not high due to the condition limitation, so it cannot be used directly in the visual system.Therefore, it is necessary to preprocess the image to improve the image quality.The purpose of image preprocessing is to eliminate useless noise in the image, recover useful information, and enhance the detectability of useful information. It usually includes graying, geometric transformation, and image enhancement[1]. The main methods of image graying include component method, average method, maximum method and weighted average method[1].Image geometric transformation, also known as image space transformation, processes the acquired image through image translation, image mirroring, image transposing, image rotation, image scaling and other geometric transformations[1], which are used to modify the random error of the position of the positive

image acquisition instrument and the systematic error in the acquisition system.Since this transformation relationship may result in the pixel of the output image being mapped to the non-integer coordinates of the input image, interpolation algorithm is needed. The common interpolation algorithms include nearest neighbor interpolation, bilinear interpolation and bicubic interpolation.Image enhancement can be divided into spatial domain method and frequency domain method[1].At present, the airspace method is widely used, which is divided into gray transformation and neighborhood denoising algorithm.Among them, gray transformation is operated on a single pixel of the image, and the gray value of the image pixel is changed by point operation.Neighborhood denoising algorithms are divided into smoothing and sharpening, both of which are operated by image filtering.Image filtering is to eliminate the noise of the image while keeping the image features unchanged, which can be divided into linear filtering and nonlinear filtering[1][3].Because nonlinear filtering can not only protect the image details but also remove the noise, it has become a research hotspot in the current image filtering methods.The most commonly used methods of nonlinear filtering are Kalman filtering and particle filtering. Kalman filtering is simple and easy to implement with good robustness[4], and is widely applied in the field of machine vision tracking.

3 Segmentation of machine vision images

Image segmentation refers to dividing the image into several non-overlapping sub-regions according to the features of color, gray scale, texture and shape, so that these features of the same sub-region show certain similarity, but show obvious difference in different sub-regions. Image segmentation is a key step for automatic recognition of image targets in machine vision. The image quality after segmentation has an important influence on the image analysis and processing. How to quickly and accurately segment the target from the complex image has been the research focus of domestic and foreign scholars[1]. Common methods of image segmentation include region segmentation, threshold segmentation and edge detection[5]. In recent years, the deep learning segmentation method[6] has also been proposed. The basic idea is to obtain the linear decision function by training the multilayer perceptron, and then use the decision function to classify the image pixels to achieve the purpose of segmentation. However, this method also has some disadvantages, that is, it needs a lot of training data and does not have a general neural network structure.

4 Feature extraction

Feature extraction plays an important role in the accuracy and speed of moving object detection in machine vision. Generally speaking, image feature extraction mainly includes four methods: color feature, shape feature, spatial relation feature and texture feature[7]. Among them, the color feature is described by the color feature of the image or image area, which has integrity. This method is suitable for describing the image that is difficult to be automatically segmented. The commonly used color feature extraction methods include color histogram, color set, and color moment, etc. Its disadvantage is that it is insensitive to the change of the direction and size of the image area, and cannot well retrieve the local features of the target in the image. There are two methods to extract and represent shape features, one is regional features, the other is contour features. The disadvantage is that the comprehensive description of the target requires high calculation time and storage. Spatial relation feature extraction can enhance the description and distinguishing ability of image content, but its disadvantage is that it is sensitive to scale change, inversion and rotation of image target. In addition, in order to better retrieve the target in practical application, in addition to the spatial relation features, it should be combined with other features. Image texture feature extraction can effectively describe regional features. Meanwhile, texture features also have rotation invariability and strong anti-noise and anti-interference ability[8-9]. However, its disadvantage is that it may be affected by light and reflection, and the texture deviation may increase with the change of image resolution. Because of its high extraction efficiency, image feature extraction method has been widely used at present, and feature extraction technology has also been paid attention by many researchers.

5 The method of moving target detection

5. 1 Background difference method

Background difference method, also known as background subtraction, is one of the most commonly used detection methods for moving objects in complex scenes. This method is mainly applied to the recognition and segmentation of moving objects in static background. The principle is that the first K frames of the video sequence are statistically averaged to obtain the background model, and then the image to be detected and the background model are made differential to obtain the difference image. Generally speaking, there is a significant difference between the gray scale of the moving target and the background gray scale, which makes the difference image only show a large gray value in the target area, and the moving target area is further obtained through binarization and morphological processing[10]. The general steps of the background difference method are as follows:
The first K frame images are averaged to obtain the background model I_{bk}

$$I_{bk} = \frac{1}{K}\sum_{k=1}^{K} I_k(x,y) \tag{1}$$

The image to be detected and the background model are computed differentially, and the absolute value is taken to obtain the difference image $D_n(x,y)$

$$D_k(x,y) = \left| I_k(x,y) - I_{bk}(x,y) \right| \tag{2}$$

The binary image M_k is obtained by comparing each pixel value in the difference image with the threshold value T

$$M_k = \begin{cases} 1 & D_k(x,y) \geq T \\ 0 & D_k(x,y) < T \end{cases} \quad (3)$$

The final target image is obtained by morphological processing. The implementation process of background difference method is shown in Figure 1.

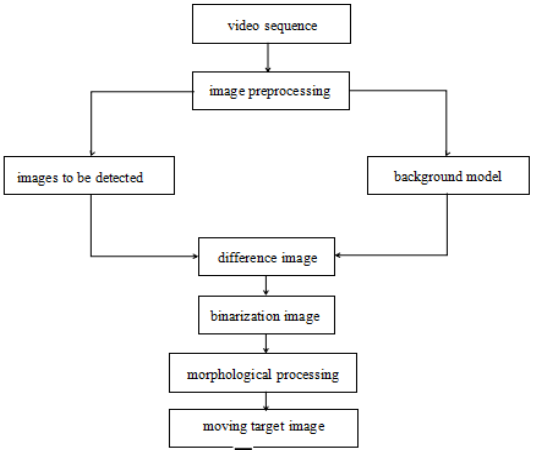

Figure 1: General implementation flow of background difference method

The performance of this approach depends on the construction of the background model. In practical situations, due to the dynamic background, illumination, shadow and other factors, the background image obtained through the statistical average of the previous K frames cannot well reflect the current scene. The improved background difference method[11] differentiates the pixel with its corresponding background pixel and its 8 neighborhood pixel, takes the absolute value, forms a difference sequence, and then assigns the minimum value of the difference sequence to the pixel. The improved background difference method can be used to deal with dynamic background, such as the slight motion of some objects. This method can also be used to eliminate the "ghost" because it takes into account the neighborhood pixels of the pixel.

5. 2 Interframe difference method

The inter-frame method is to obtain the inter-frame difference image by means of the displacement of the moving target between two adjacent frames, where the background of the image is almost constant, and then make a difference between the two adjacent frames, that is, the pixel value of the corresponding position of the two frames is subtracted. By setting an appropriate threshold value, and then comparing it with the given threshold value, if the pixel value on the difference image is greater than the threshold value, it means that the pixel point is the point on the target pixel; If the pixel value on the difference image is less than the threshold, that pixel is considered to be the pixel on the background. The final output image is the binary image that separates the background from the foreground, and then the target region image is obtained through the subsequent processing of the binary image. The general steps are as follows:

The difference operation is performed between the image in frame K and the image in frame K-1

$$D_k(x,y) = \left| I_k(x,y) - I_{k-1}(x,y) \right| \quad (4)$$

Where, $I_k(x,y)$ and $I_{k-1}(x,y)$ is the image of the k frame and the k-1 frame, and $D_k(x,y)$ is the image after difference.

Set a certain threshold value T , when $D_k(x,y)$ less than the threshold value T is judged as background pixel, when $D_k(x,y)$ greater than the threshold value T is judged as target pixel. The target image is obtained by morphological processing.

$$M_k(x,y) = \begin{cases} 1 & D_k(x,y) > T \\ 0 & D_k(x,y) \leq T \end{cases} \quad (5)$$

The general implementation process of frame difference method is shown in Figure 2.

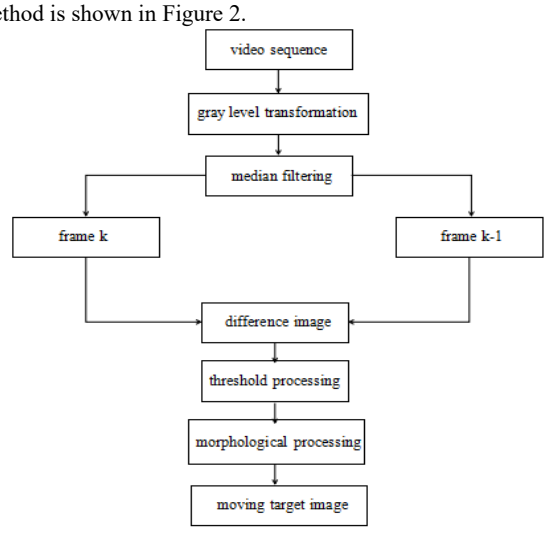

Figure 2: General implementation **flow of interframe method**

Interframe difference method is simple, small computation and strong real time. According to the principle of frame difference method, it is not sensitive to the slow change of light in the scene and has strong self-adaptability to dynamic environment[12]. This method is suitable for the situation where the detection precision is not high and the real-time requirement is high. However, the frame difference method also has some disadvantages. The detection result of the algorithm also depends on the displacement of the moving target of adjacent frames. If the displacement is too

large, one moving target may be mistakenly detected as two targets. When there is a large uniform gray area on the surface of the target, it is easy to produce a "void" phenomenon inside the moving entity.

Aiming at the disadvantages of the above two frame difference method, that is, the target area obtained will be larger than the actual target, which is easy to cause the phenomenon of "double shadow", and the overlapping area cannot be effectively detected, which will cause the phenomenon of "void". Three frame difference method is based on the two frames difference method, by getting three consecutive frames of video sequence images, and then let the adjacent two difference between two consecutive frames, and the results obtained by the binarization processing, finally to be two binary logic "and" of image, two images of the public areas, so as to eliminate the phenomenon of "ghosting", to produce the final motion target area.

5. 3 Optical flow method

Optical flow is the instantaneous velocity of pixels moving on the imaging plane. In an image plane, each pixel in the image set to a velocity vector to form optical flow field, because of the moving target and the image background region light flow, leading to optical flow field changes, but the image in the background region generated by the motion is continuous uniform, so you can according to the change trend of image pixels to implement for the moving target detection. Common optical flow calculation methods include Lucas-Lanada algorithm and Horn-Schunck algorithm. The derivation process of optical flow method is as follows:

The gray value of the pixel (x, y) of the image at time T is I (x, y, t), and the gray value of the pixel at the same position at the next time is I (x+dx, y+dy, T +dt), assuming that the gray value is unchanged

$$I(x+d_x, y+d_y, t+d_t) = I(x, y, t) \tag{6}$$

Since the change of gray level is continuous and uniform, the Taylor series expansion of gray level at the next moment is as follows

$$I(x+d_x, y+d_y, t+d_t) =$$

$$I(x, y, t) + \frac{\partial I}{\partial x} d_x + \frac{\partial I}{\partial_y} d_y + \frac{\partial I}{\partial t} d_t + \varepsilon \tag{7}$$

Where ε is the infinitesimal term of the second order, and further

$$\frac{\partial I}{\partial x}\frac{d_x}{d_t} + \frac{\partial I}{\partial y}\frac{d_y}{d_t} + \frac{\partial I}{\partial t} = 0 \tag{8}$$

So $u = \frac{dx}{dt}, v = \frac{dy}{dt}$, the basic constraint equation of optical flow field is as follows

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{9}$$

The purpose of the optical flow method is to calculate the value of velocity vector U and V through the formula, so as to separate the moving target. Generally, the optical flow value can only be found in the area of the moving target. If the velocity direction and velocity are greater than the given threshold value, it can be identified as the moving target; otherwise, it will be considered as the background.

To sum up, by comparing these methods, the advantage of optical flow method is that it can detect independent moving targets without knowing any information of the scene in advance, and obtain complete information of moving targets, which is suitable for dynamic background. However, the optical flow method has high computational complexity and is sensitive to noise. Background difference method has the advantages of low computational complexity, easy implementation, good real-time performance and accurate acquisition of moving target information. It is suitable for static background. But the background difference method depends on the establishment of the background model and is easily affected by background dynamics, illumination, shadow and other factors. The frame difference method is insensitive to the change of light intensity and has good real-time performance. But for fast moving targets, the higher the speed is, the larger the target range is extracted, and it is easy to turn one target detection into two targets. At the same time, it is easy to produce "empty" phenomenon in the area with uniform gray change in the target area. Aiming at the defects of these traditional moving target detection algorithms, scholars at home and abroad have proposed many improved algorithms, such as background difference method combined with frame difference method, mixed Gaussian model and so on[13].

6 Recent research trends

Duan Suolin et al[14]. improved the traditional three frame difference method. Although the traditional three-frame difference method can effectively suppress the double shadow phenomenon, it is difficult to detect the overlapping part between the targets, and there are still holes. The method differentiates six consecutive frames of images and corrects the dynamic threshold value determined by the maximum variance method to eliminate the void. At the same time, Sobel edge detection operator is used to extract the edge features of the target, and the complete edge of the target is obtained by difference of six frames of images.

Li Yuan et al[15]. proposed an adaptive three-frame difference method based on mean background model. Former k frame is used to establish the method of average background model as three frame difference method in the middle of the frame, through three frame difference, will get the difference of the results to choose the adaptive threshold for binarization processing, the two detected target "and" operation, finally, morphology processing, filtering, etc., to obtain the real position of the target. This method is not susceptible to light changes and other minor changes, and can effectively overcome the cavity and edge loss.

Zhang Yanyan et al[16]. proposed an algorithm combining Lucas-Kanade optical flow method with the image segmentation

method of the maximum inter-class variance. The algorithm firstly grays the image sequence and then computations the optical flow field of two adjacent frames of images by Luck-Kanade. The foreground was extracted by the image segmentation method of the maximum inter-category variance, the image area with discontinuous optical flow was regarded as the foreground area, and the area with continuous optical flow was regarded as the background area. Then the optimal threshold was obtained through the operation of the maximum inter-category variance method, and the target image was obtained by binarization and morphological processing of the image. Compared with the traditional moving target detection algorithm, this algorithm can extract the moving target more completely and the detection effect is better.

Guan Hongyun et al[17]. proposed a background difference improved algorithm based on feature fusion, and made use of spatio-temporal local binary mode and target detection algorithm based on color feature fusion to effectively make up for their respective shortcomings in foreground detection. Considering the confidence and similarity scores of the current pixel points and background model features, the background probability is obtained and foreground segmentation is carried out, and then the currently detected background pixels are used to update the background template, so as to better solve the problem of foreground detection under complex background.

Chen Yuan et al[18]. proposed an improved moving target detection algorithm combining the background difference method and the inter-frame method. The mean background modeling with illumination compensation model was used to reduce the influence of illumination mutation on the effectiveness of target detection. Then, the three-frame difference method based on the maximum fuzzy entropy threshold is used to process the video image to reduce the probability of detecting the incomplete target contour. Finally, the two improved algorithms are combined to obtain a comprehensive foreground object. The improved algorithm can effectively improve the influence of light change and enhance the accuracy of detection.

7 Conclusion

Moving object detection is one of the challenging subjects in the field of machine vision. In this paper, three classical moving target detection algorithms, namely optical flow method, frame difference method and background difference method, as well as the improved three-frame difference method based on the inter-frame method, are firstly summarized and studied, their advantages and disadvantages are analyzed, and some new methods are introduced. Finally, new algorithms in the field of moving target detection in recent years are summarized. These algorithms are applied in specific moving target detection scenes, bringing better detection effect.

REFERENCES

[1] It_job. An overview of preprocessing methods for image processing[EB/OL].https://blog.csdn.net/it_job/article/details.2017-12-18.

[2] Zhang Zhongliang. A Review of image target recognition methods based on machine vision[J]. Technology and Innovation,2016(14):32.

[3] Wang Jin, LI Shaohua, Xie Shouyong.Research on object recognition method based on machine vision[J]. Journal of Southwest Normal University (Natural Science edition),2015,40(6) :130.

[4] Houtekamer P L,Mitchell H L. A sequential ensemble kalman filters for atmo spheric data assimilation[J].Monthly Weather Review,2001,129(1):123.

[5] Hou Hongying, Gao Tian,Li Tao.Overview of image segmentation methods [J]. Computer knowledge and technology,2019,15(5):176.

[6] Wang Qiuping, Zhang Zhixiang, Zhu Xufang. A review of image segmentatio n methods[J].Information Recording Materials,2019,20(7):12.

[7] Wang Zhirui, Yan Cailiang. A review of image feature extraction methods [J]. Journal of jishou university (natural science edition),2011,32(05):43-47.

[8] Ohanian P P,Dubes R C. Performance evaluation for four classes of textural features[J]. Pattern Recognition,1992,25(8):819.

[9] Clausi D A,Yue B. Comparing co－occurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery[J]. IEEE Transactions on Geoscience and Remote Sensing,2004, 42(1) :215.

[10] Song Jiasheng. Research on moving target detection and tracking algorithm in video sequence image[D].Guangzhou: South China University of Technology,2014 .

[11] Xu Guoliang, Zhou Hang, Yuan Liangyou. Using mixed gaussian and topolog ical structure of the human "ghost" suppression algorithm [J/OL]. Journal of int elligent system: 1-9 [2020-07-22]. http://kns.cnki.net/kcms/detail/23.1538.TP.2 0200714.1417.018.html.

[12] Ci Wenyan. Review of methods for detecting moving targets [J]. Information Technology,2016(12):93-96+100.

[13] Zhang Dongmei, Wu Jie, Li Pitin. Overview of motion target detection Algor ithms based on machine vision [J]. Intelligent Computer and Applications,202,1 0(03):192-195+201.

[14] Duan Suolin, Gao Renzhou, Liu Fu, Liu Maomao, Wang Yifan, Pan Lizheng. Prospect detection of fusion improved frame difference and visual background extraction algorithm [J]. Minicomputer system,2019,40(09):1903-1908.

[15] Li yuan, Hou honglu. Adaptive three-frame difference algorithm based on im proved mean modeling[J]. Electronic measurement technology,2019,42(03):21-24.

[16] Zhang yanyan, Lou li, Liang Shuo. Research on moving Target Detection Tec hnology based on Improved Optical Flow Algorithm[J]. Intelligent Computer a nd Application, 2012,8(01):55-58.

[17] Guan Hongyun, Su Zhentao, Wang Chen. Background difference algorithm b ased on feature fusion [J/OL]. Electronic science and technology, 2020 (12) : 1-6 [2020-07-22]. http://kns.cnki.net/kcms/detail/61.1291.TN.20200106.1536.01 6.html.

[18] Chen Yuan, Hu Na, Yu Qiuyue. Moving object Detection using background d ifference method and frame difference method[J]. Modern Computer,2019(34): 50-53.