

## Segmentation and Recognition Algorithm Research of Moving Targets in an Image

Zhao Lixin<sup>1</sup>, Wang Jianping<sup>1</sup>, Zhang Yong<sup>2</sup> Xie Mengqing<sup>3</sup>

<sup>1</sup>School of Electric Engineering and Automation, Hefei University of Technology

<sup>2</sup>Office of Discipline Inspection, Hefei University of Technology, Hefei, China Hefei, China

<sup>3</sup>School of Foreign Studies, Hefei University of Technology, Hefei, China

hfutzlx@163.com, wjphfut@126.com, xgbzhy@hfut.edu.cn, xiaoxieuser@163.com

**Keywords:** motion estimation and compensation; improved HSI frame difference method; segmentation of moving targets; targets classification.

**Abstract.** For image segmentation and recognition problems of moving target in the image, a humanoid intelligently cognitive method has been discovered. The thesis adopts an improved three-step search algorithm to improve the search speed of gray projection algorithm, through improved HSI difference model, adaptive segmentation algorithm and image morphology to automatically extracting moving target area under the condition of non-background image; the thesis introduces some features analysis and definition of moving objects to calculate and match eigenvalues of the moving target area, classifies and recognizes the moving target area. Experimental results show that the method in this thesis can effectively extract the moving targets and quickly achieve classification and reorganization of moving objects.

### Introduction

Moving target detection and identification is the key technology of machine vision research and application and intelligent video surveillance, which has been one of the research focuses and attracted great attention. Current detection methods are: optical flow method[1], background subtraction method [2], three image difference method [3], etc.. Optical flow method is able to detect the object of independence movement, and can be used under conditions that the camera and the target are in relative motion. but optical flow method is computational complexity and time-consuming; as for the background subtraction, moving target has a satisfactory extracting result of when the background image is fixed, but it is greatly impacted by light and shadow; conventional three image difference method is able to meet the moving target detection under more complex scene, but when there is a relative motion of the background, conventional three image difference will make the moving target and background aliasing together, so it will be difficult to accurately detect the moving target.

Aiming at segmentation and recognition problems of moving objects in the image, the thesis adopts an improved three-step search algorithm[4] (ITSS) which has small computation and high quality of motion compensation. After detecting background motion vector, the same background of three adjacent frame images can be stabilized in the same place on the same image through inter-frame motion vector compensation, which can take advantage of the improved HSI frame difference method and adaptive segmentation algorithm to extract the moving target. The thesis analyzes features of three moving objects, human beings, vehicles and animals, defines the difference method and provides computation and match method of eigenvalues in the moving target area, in order to classify and recognize the moving target area.

## Motion Estimation and Compensation in the Image

### A. Gray Projection Algorithm for Non-specific Background Conditions

#### 1) Row and Column Gray Projection

Row and column gray projection is to map the two-dimensional gray information  $V_k(i, j)$  of an image into independent one-dimensional projection sequence in rows X, Y direction. To eliminate the correlated errors caused by different light, projection vector normalization is needed: first, obtaining the average E gray value of the image, in the calculation of a separate sequence of one-dimensional projection values of each pixel are subtracted gradation mean E. Calculated as follows:

$$V_k(i) = \sum_j V_k(i, j) - E \quad (1)$$

$$V_k(j) = \sum_i V_k(i, j) - E \quad (2)$$

Formula (1) and (2) is the normalized row, column gray projection formula. Whereby,  $V_k(i)$  and  $V_k(j)$  are the sums of gray value of the i row and the j column in the k frame image respectively, and  $V_k(i, j)$  is the pixel gray value at (i, j) position of the k frame image.

#### 2) Displacement Correlation Detection

Doing cross-correlation calculation between reference frame R image and the row, column gray projection curves, the row and column displacement vector value of current frame image relative to reference frame image can be determined according to the valley value of two correlation curves. Correlation calculation formulas are as follows:

$$C(w) = \sum_{j=1}^L [G_k(j+w-1) - G_r(n+j)]^2,$$

$$1 \leq w \leq 2n+1 \quad (3)$$

$$\delta_c = n+1 - W_{\min} \quad (4)$$

In formula (3),  $G_k(j)$ ,  $G_r(j)$  is the j gray projection value of the k frame and reference frame r respectively. L is the length of the column and n is the search width of displacement vector relative to the reference frame on one side. Suppose  $W_{\min}$  is the W value when  $C(w)$  is at the minimum, the displacement vector in the vertical direction of the k frame image relative to the reference frame image can be obtained through formula (4). The same process is adapted to obtain the displacement vector in the horizontal direction. Then current image can be moved a corresponding distance in the opposite direction corresponding to the displacement vector of the pixel size, or the previous frame to move along the displacement vector corresponding size, in order to successfully compensate for the background motion image sequence.

### B. Improved Three-step Search Algorithm Detecting Motion Vector

For the existing algorithm, TSS algorithm adopts a loop iteration search strategy to search MBD (Minimum Block Difference) points, i.e., the minimum error point. The algorithm takes half the width of the search window as its initial search step, and each cycle will determine the new MBD point and the search step is halved until the step length is equal to 1 so that the final MBD point is obtained. This thesis uses an improved three-step search algorithm (ITSS) to design the size of diamond search pattern and adopt the adaptive search strategy. Operating steps are as follows:

Step 1, searching 17 points, including nine points in the 9x9 meter and eight points on the big diamond pattern around the center. If MBD point is the center of the search window, the algorithm ends; if MBD point is on the big diamond pattern, go to step 2, otherwise go to step 3.

Step 2, taking the MBD point in Step 1 as the center, reuse small diamond search pattern to search until the MBD point is in the center of search window.

Step 3, halving the search step size and identifying new MBD point until the step size is equal to 1, and the algorithm ends.

Figure1 describes how to use ITSS algorithm to search motion vectors A (-2, -2). First go to step

1 and search 17 points. After obtaining the MBD point, change the big diamond pattern into a small diamond pattern and then re-search four points around the small diamond pattern, shown as circles in the figure. After getting MBD, reusing the small diamond search pattern, shown as triangles in Figure 1, the motion vector A (-2, -2) can be obtained.

C. Compensation of Motion Vectors for Non-specific Background Conditions

For three consecutive frame images with background change for non-specific background conditions, using the improved three-step search algorithm (ITSS) in the gray projection algorithm to estimate the displacement vector of the k frame relative to the k-1 frame background, the displacement vector of the k+1 frame relative to the k frame background. According to displacement vector, compensating the k-1 to k frame, the k + 1 frame to the k frame, so the same background of three adjacent frame images will be stabilized at the same position on the same image. After compensation of three adjacent frame images, background difference is reduced, which image segmentation algorithm can be followed under static background.

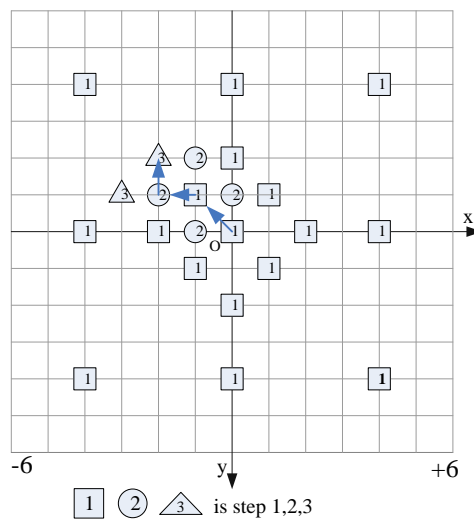


Figure 1. Improved Three-step Search Method Sample

MOVING OBJECT SEGMENTATION IN IMAGE

D. Target Area Divided by Improved HSI Frame Difference Method

1) Improved HSI Frame Difference Method

Hue close to the color perception of human eyes, saturation and brightness properties (H, S, I) model is more suitable for image recognition processing. Therefore, the thesis introduces the improved HSI frame difference model [5].

a) HSI Difference Model of Motion Target Image

Suppose three consecutive frame image sequences collected in the same position by machine vision system are  $f_{k-1}(x,y)$ ,  $f_k(x,y)$ ,  $f_{k+1}(x,y)$ , sensitive coefficients to hue of human eye's imitation is lower than brightness. By sensitive coefficient[6] adjustment of image hue H, saturation S and brightness I attribute, the sensitivity coefficient, and each frame image of the obtained sequence is converted to improved HSI image  $f_i(x,y)$  which is more capable to highlight the moving target, as the following equation:

$$f_i(x,y) = \{W_H H_i(x,y), W_S S_i(x,y), W_I I_i(x,y)\}$$

$$= \{H'_i(x,y), S'_i(x,y), I'_i(x,y)\}, \quad i=k-1, k, k+1 \quad (5)$$

Where,  $W_H$ ,  $W_S$ ,  $W_I$  are the setting hue, saturation and brightness sensitivity coefficient respectively. Through formula (5), difference the k frame image with the k-1 frame image, the k + 1 frame image with the k frame image.

Definition 1, frame difference image computation model is as follows:

$$f_{d1}(x, y) = \{ |H'_k(x, y) - H'_{k-1}(x, y)|, |S'_k(x, y) - S'_{k-1}(x, y)|, |I'_k(x, y) - I'_{k-1}(x, y)| \} \quad (6)$$

$$f_{d2}(x, y) = \{ |H'_{k+1}(x, y) - H'_k(x, y)|, |S'_{k+1}(x, y) - S'_k(x, y)|, |I'_{k+1}(x, y) - I'_k(x, y)| \} \quad (7)$$

Where,  $f_{d1}(x, y)$ ,  $f_{d2}(x, y)$  are results caused by three consecutive frames image sequence. Images after difference contain more noise spots and the opening and closing algorithm in the mathematical morphology can be used as noise filter to remove the interference of noises.

#### b) Adaptive Segmentation of Moving Target Area

There are often dash areas in the target, which is difficult to distinguish. Because shadow pixels and moving target pixels have some same visual characteristics, conventional background subtraction operation can not distinguish moving targets and shadows. Therefore, shadows are also seen as moving targets in conventional methods, which will affect the accuracy of detection and bring tracking and recognition errors to subsequent detection. In RGB space, the consistency of differences between humans' perception and computation, but HSI color space has better color perception consistency, proper for computation of image's color similarity. Hence, in order to take advantage of hue changes to determine the moving target area, two frame difference images are processed as follows:

Definition 2 second frame difference image computation is:

$$Y(x, y) = f_{d2}(x, y) \odot f_{d1}(x, y) \\ = \{ Y_H(x, y) \ Y_S(x, y) \ Y_I(x, y) \} \quad (8)$$

Where,  $Y(x, y)$  is the result of phase computation. Phase computation is the minimum operation on saturation H, brightness S and properties I of the two-frame difference image  $f_{d1}(x, y)$   $f_{d2}(x, y)$ . The hue attribute H of HSI color space image can more accurately reflect color types and has a low sensitivity to changes of external lighting conditions, with relatively stable and narrow range of values. But when brightness I is small, the value of H tends to uncertainty. In order to determine the obtained area resulted by phase computation between results  $f_{d1}(x, y)$   $f_{d2}(x, y)$  of two-frame difference image, the ratio of inter-class and internal spur is taken as the objective function [7] with best division. In order to better determine the adaptive threshold segmentation, the thesis uses the improved HSI frame difference model of double attributes, hue H and brightness I when image computation.

Definition 3 Improved HSI second frame difference model is:

$$K(x, y) = Y_H(x, y) + Y_I(x, y) \quad (9)$$

Where,  $K(x, y)$  is the computational result of frame difference image, and frame difference model is the summation computation of hue H value and brightness I value. Areas that HSI color space image pixel values are bigger than T form  $A1(x, y)$  area. Suppose its variance is  $\delta A1(T)$ , less than or equal to T-formed  $A2(x, y)$  area, suppose its variance is  $\delta A2(T)$ , take advantage of inter-class and internal spur method to obtain the best segmentation value of the moving area  $Ym(x, y)$  of moving targets and invariable background area  $Yb(x, y)$ . Conducting increase in connected domain for corresponding image  $fR(x, y)$  of each connected domain, more intact region  $fk(x, y)$  of moving targets can be obtained.

#### E. Processing and Extracting Moving Target

For the moving target area obtained through the improved HSI secondary frame difference method, using closing operation, filling operation, median filtering of mathematical morphology to process moving target area  $fk(x, y)$ . So the moving object can be extracted. Note each edge of a closed area for a moving target  $Ni$ ,  $i=1, 2, \dots$ .

## Characteristic Extraction and Classification of Moving Target Area

### F. Characteristics Extraction of Moving Target

When the scene contains multiple moving targets, human beings, vehicles and animals on the scene need feature recognition in order to analyze and recognize characteristics of every moving target.

#### 1) Volume Characteristic Extraction of Moving Target

According to the moving target knowledge in the intended recognition (eg: the car occupies larger area than the person of the same position in the image, the person occupies larger area than small animals and other priori knowledge), a threshold  $t_0$  can be set based on the size which the recognition target occupies in the image by the position for a specific scene image region, using the area of moving targets to differentiate. Due to the area of object imaging related to distance, the moving target area needs being equivalent to the same plane as the camera distance. Therefore the depth information of moving target is needed as a reference, and the computation are as follows:

a) Horizontal, vertical axis of center of gravity in the moving target area is calculated as follows:

$$m_{xi} = \frac{\sum_{(x,y) \in F_i} \sum x F_i(x,y)}{\sum_{(x,y) \in F_i} \sum F_i(x,y)} \quad (10)$$

$$m_{yi} = \frac{\sum_{(x,y) \in F_i} \sum y F_i(x,y)}{\sum_{(x,y) \in F_i} \sum F_i(x,y)} \quad (11)$$

Where,  $m_{xi}$ ,  $m_{yi}$  are the  $i$ -th coordinate of gravity moving target  $N_i$ ,  $N_i$  is moving target region of binary image  $F_k(x,y)$ .

#### b) Computation and Track of the Deep Coordinate Value of Center of Gravity in the Moving Target Area

Under the condition of non-specifically known information, depth information computation of the object is a difficult task. The thesis adopts a ranging method which is approximate to pinhole imaging principle using single camera, more accurately calculating the depth distance  $m_z$  in the case of a known camera height.

In practice, the robot camera is placed right in front, which height has been determined to be  $h$ , the focal length has been determined to be  $f$ . The center of gravity of the moving target itself has a certain distance from the ground, so moving target region of the moving object may be directly below the edge of the center of gravity coordinate points in the image represents the point P, specifically computations are as follows:

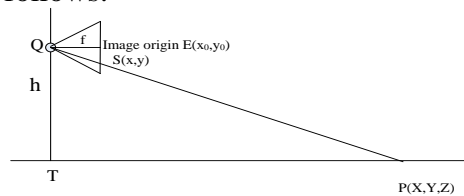


Figure 2. Ranging Diagram of Pinhole Imaging Principle

With figure 2, Q is a video camera, T is the camera projection on the ground, the center point coordinate of the imaging surface is  $E(x_0, y_0)$ , the three-dimensional coordinates of the edge points which is right below the center of gravity of the moving target is  $P(X_P, Y_P, Z_P)$ . The projection coordinate of this point in the image is  $S(X_S, Y_S)$ , so the triangle  $PQT$  is similar to triangle  $QES$  and the horizontal distance  $m_z$  between the object point on the ground and the camera can be obtained through computation.

$$m_Z = f \times h / (Y_S - Y_T) \quad dy \quad (12)$$

Where,  $dy$  is the physical size of the vertical pixel camera imaging,  $h$  is the height between the camera and the ground.

Under pinhole imaging model, the ratio between the imaging area of the moving target and the surface area of the object is equal to the ratio between the focal length  $f$  and the object distance  $mZ$  squared. Through derivation the camera distance is  $mZ$ , so the imaging area  $M$  when the imaging area  $M_1$  of the moving target:

$$M = \frac{m_Z^2}{L^2} M_1 \quad (13)$$

Through (13), the imaging area of the moving target can be normalized as the imaging area obtained when distance camera is at  $L$  position, regarded as a volume feature of the target.

### 2) Shape Characteristics Extraction of Moving Targets

Based on prior knowledge, the maximum ratio between length and width of the car in the image should be more than 1, the maximum ratio between length and width of the person in the image should be less than 1, and the maximum ratio between length and width of the animal in the image should be greater than 1. Suppose  $W_R^i, H_R^i$  is the found maximum width and height of the  $i$ -th moving target, the shape characteristic of the moving target which is also the maximum ratio between length and width movement is:

$$A_i = W_R^i / H_R^i \quad (14)$$

### 3) Head and Shoulder Characteristics Extraction of Moving Targets

Taking 1/4 from the top position of the moving target to get image of head and shoulder parts of the moving target, shape characteristics are extracted for target recognition. There are large differences in shapes of person's head and shoulder areas compared to the top area of other moving targets, therefore it is an effective recognized characteristics. The thesis adopts four affine invariant moments to calculate the obtained upper 1/4 shape characteristics of the target area, detailed calculation seen the reference [9]. Suppose  $E_i$  is the calculated sum of the absolute value of the difference between four affine invariant moments and head and shoulder part experience invariant moments,  $\delta_1$  is a certain tolerance limit.

## G. Classification and Match of Moving Targets

### 1) Classification of Moving Targets

Through the above algorithm, moving targets in the scene can be classified and the target required to be recognized can be found based on the actual situation.

For human beings,  $\{A_i > 1\} \cap \{E_i \in (0, \delta_1)\}$ ;

For cars,  $\{A_i < 1\} \cap \{M_i > t_0\} \cap \{E_i \notin (0, \delta_1)\}$ ;

For animals, others.

### 2) Match of Moving Targets

When the scene contains several moving targets, using the following rules to match for the current detected motion target  $N_i$  ( $i=1, 2, \dots$ ).

Matching rule sets of moving targets:

Rule 1 If  $A_i > 1$  and  $E_i \in (0, \delta_1)$ , Then the moving target is human being.

Rule 2 If  $A_i < 1$  and  $M_i > t_0$  and  $E_i \notin (0, \delta_1)$ , Then the moving target is car.

Rule 3 If  $A_i < 1$  and  $M_i < t_0$  and  $E_i \notin (0, \delta_1)$ , Then the moving target is animal.

Through matching between the detected moving targets and the expected recognition targets respectively, the recognized moving target area can be determined.

## Experimental Results and Analysis

### H. Experimental Conditions and Parameters

A lot of experiments have been done to verify the method proposed in the thesis of moving targets detection in the outdoor scenes. Experimental conditions are as follows: machine vision system camera resolution is  $320 * 240$ ; height  $h$  is 653mm, the focal length  $f$  of 15mm, 25 frames per second, the horizontal and vertical physical size of the camera imaging pixel can be respectively obtained by camera calibration,  $dy = 0.12\text{mm}$ . Sensitive factors under HIS space take  $WH = 0.1$ ,  $WS = 0.1$ ,  $WI = 0.80$ . Suppose moving targets are all normalized to the position which is 5000mm away from camera to get the imaging area, the imaging area threshold  $t_0$  set at this position is 400 pixels, the established tolerance values  $\delta_1$  of the absolute value of the difference between affine invariant moments and capititation shoulder part invariant moments is 100.

### I. Experimental Results

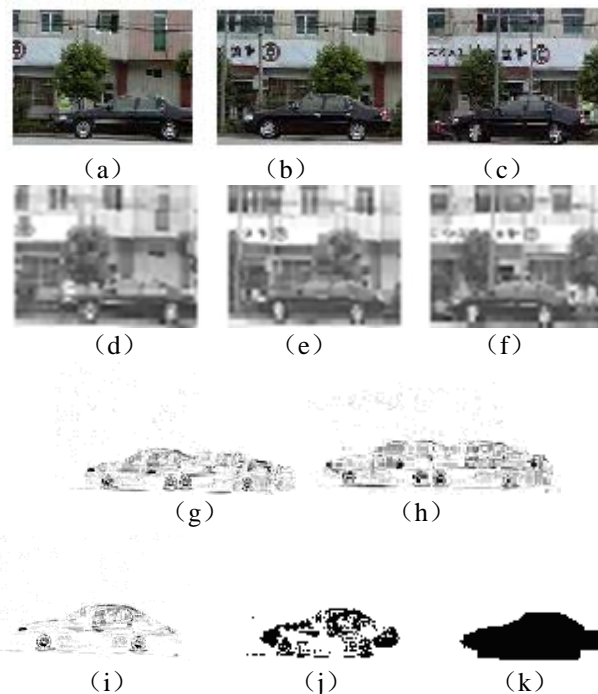


Figure 3. Experiments images

Experimental Steps:

- ① Three frames original images seized by the camera at  $t$  moment are image (a), (b), (c).
- ② Images under new HIS space transferred by equation (5) are image (d) to image (f).
- ③ Respectively differencing image (e) and image (d), and image (f) and image (e), results are image (g) and image (h).
- ④ Doing phase operation for image (g) and image (h) by equation (8), the result is image (i).
- ⑤ Differentiating adaptive binarization threshold of image (i), the binarized image (j) can be obtained.
- ⑥ Using mathematical morphology to process moving target area, image (k) can be obtained.
- ⑦ Through the method in the IV section, moving targets can be recognized: N1 is the car.

## Conclusion

The thesis presents the moving target detection and extraction algorithms based on non-specific

background conditions, applicable to static background and presence of relative motion in the background conditions in order to detect and extract the single object and multi-objects.

(1) The thesis adopts an improved three-step search gray projection algorithm, which improves search speed of motion vector and reduces computational complexity, maintains high search precision of motion vector meanwhile.

(2) Detecting moving targets through the improved HSI three-frame difference method overcomes the noise influence caused by the relative motion of light and background, achieving the real-time detection to moving target areas.

(3) Introducing some features analysis and definition of moving objects, defining the eigenvalues computation of the moving target area which can proceed feature match of the moving object, extract and recognize the required moving target area.

The method in this thesis for problems such as target disappearance of moving objects caused occlusions or high-speed of moving object needs further research.

## References

- [1] WANGXiao wei , NINGGu. A Modified Object Tracking Algorithm Based on Optical-Flow[J]. Editorial Board of Geomatics and Information Science of Wuhan University. 2003,03:351-353, 362.
- [2] DAI Ke xue, LI Guo hui, TU Dan, YUAN Jian. Prospects and Current Studies on Background Subtraction Techniques for Moving Objects Detection from Surveillance Video[J]. Journal of Image and Graphics, 2006 , 11( 7) : 919-917.
- [3] Wang Jing, BaoWen xing . A VIDEO SUPPORTEDMOVING OBJECT DETECTION TECHNIQUE BASED ON DIFFERENCE ALGORITHM[J]. Computer Applications and Software, 2009 , 26( 12) : 26-12.
- [4] XIE Lang, YANG Yan, CAO Yang, et al. An Improved Three-Step Search Algorithm for Block Motion Estimation[J]. Wuhan University Journal(Natural Science Edition), 2005,51(5):625-628.
- [5] WANG Jian-ping,LIU Wei,WANG Jin-ling, A Moving Object Detection and Recognition Method in Video Sequences[J]. Computing Technology and Automation, 2007, 26 (3) :78-80.
- [6] WANG Jian-ping,QIAN Bo,JIANG Tao. Research on the segmentation method of vehicle license plate based on space transform analysis[J]. Journal of Hefei University of Technology(Natural Science), 2004,27 (3) :251-255.
- [7] WANG Jian-ping, QIN Feng. On the adaptive binarization filtering algorithm for gray character images and its application[J]. Journal of Hefei University of Technology(Natural Science), 2004,27 (5) :509-512.
- [8] LI Ling juan, JIA Zhen tang, HE Gui ming. A Robust Video Segmentation Algorithm[J]. Journal of Image and Graphics, 2002, 7 (11) : 1128-1133.
- [9] NI Fuchuan, HE Guiming, LONG Lei. Detection of Human's Head and Shoulder Based on Moment Invariants[J]. Computer Engineering, 2005, 31(10):174-176.