A VSS Algorithm based on Multiple Features for Object Tracking

Bin Xu
Department of Mechanical-Electrical Engineering, North China Institute of Science and Technology, Sanhe, 065201 China
Email: 200602006@163.com

Xiaoju Shen
Department of Management, North China Institute of Science and Technology, Sanhe, 065201 China
Email: 49609856@qq.com

Feiji Ding
Department of Mechanical-Electrical Engineering, North China Institute of Science and Technology, Sanhe, 065201 China
Email: simon.space@163.com

***Abstract*—A variable search space (VSS) approach according to the color feature combined with point feature for object tracking is presented. Mean shift is a well-established and fundamental algorithm that works on the basis of color probability distributions, and is robust to given color targets. As it solely depends upon back projected probabilities, it may miss the targets because of illumination and noise. To overcome the flaw, we proposes VSS algorithm based on the color and robust feature of the detected object. The proposed algorithm can solve the problem that the color of the detected object is similar to the background, and achieve better real-time tracking due to change the search window's size. Experimental work demonstrates that the presented method is robust and computationally effective.**

***Index Terms*—Meanshift, Scale Invariant Feature Transform, target tracking, variable search space**

I. Introduction

Visual object tracking in complicated environments is an important topic within the field of computer vision and is widely used in various applications such as vehicle navigation[1], intelligent surveillance, traffic monitoring, and robot vision[2]. The target tracking indicates tracing the process of targets when they shift over a set of pictures. An excellent tracking approach should be able to work well in many real circumstances, such as background clutters, object occlusions, and different illuminations [3].

On the basis of its feature values, the methods of target tracking can be divided as color-based method, model-based method, and boundary-based method. As the simple implementation to facilitate fast calculation, the color-based method is widely exploited in object tracking. Classical color-based algorithms are mean-shift[4] and Camshift[5], which is used for the color probability distribution of object tracking. However, a drawback of this algorithm is vulnerable to different illumination changed and background with similar colors. Thus, two or more methods are combined to achieve more robust tracking effects[6].

The proposed algorithm in this paper combines color feature with point feature, which is an effective enhancement for meanshift using SIFT algorithm. The key to the method is that extracts the color feature from the detected object, and reduces the search space of extracting the feature points[7]. This method may change the global search into the local search so as to decrease the processing time, and settle the similar color background problem and enhance the property of target tracking[8]. The major results of this paper are as follows: this approach grounded on color and point features is presented for object tracking to obtain accurate tracking, and the proposed algorithm improves the classical algorithms of meanshift and SIFT in tracking and performs effectively in complex real environments.

II. Finding The Region of Interent via Camshift Algorithm Based on Color Feature

The Camshift algorithm is derived from the original Mean Shift algorithm based on color for tracking. The meanshift algorithm, which repeatedly moves a mass center to the average of data points in its neighborhood, is a statistical approach. It is effective and robust to use for the conditions such as clustering, tracking, and probability density estimations. However, Mean Shift algorithm based on static probability distributions, is not updated unless the object obviously changes in shape, color or size, While Camshift employs continuously adaptive probability distributions.

Camshift basically climbs the gradient of a background probability distribution calculated from regulated color histogram to seek the neighbourhood peak in a search region. The mean location of the object is found by computing 0th, 1st and 2nd order picture moments:

a) Compute the 0th moment

doi:10.4304/jsw.8.12.3029-3034

$$M_{00} = \sum_x \sum_y P(x, y) \tag{1}$$

b) Find the first moment for x and y

$$M_{10} = \sum_x \sum_y xP(x, y) \tag{2}$$

$$M_{01} = \sum_x \sum_y yP(x, y) \tag{3}$$

c) Find the second moment for x and y

$$M_{20} = \sum_x \sum_y x^2 I(x, y) \tag{4}$$

$$M_{02} = \sum_x \sum_y y^2 I(x, y) \tag{5}$$

d) Compute the centroid of search window

$$x_c = \frac{M_{10}}{M_{00}}; y_c = \frac{M_{01}}{M_{00}} \tag{6}$$

Where $P(x, y) = h(I(x, y))$ is the probability distribution at position x, y. The search window $I(x, y)$ is calculated from the histogram h of I. The mean position of the target can be computed with (x_c, y_c).

While its aspect ratio:

$$ratio = \frac{M_{20}}{x_c^2} \bigg/ \frac{M_{02}}{y_c^2} \tag{7}$$

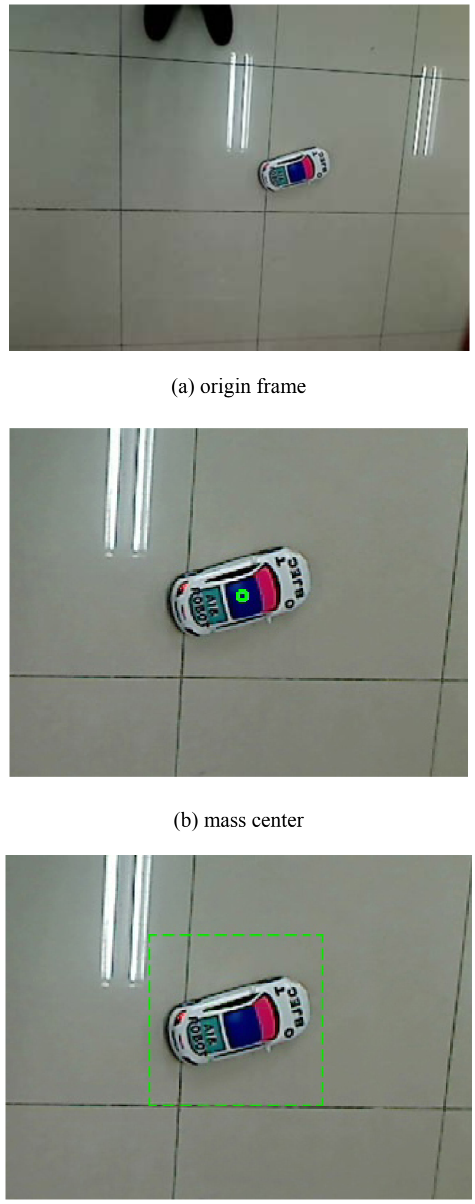
is used to update the search window with

$$width = 2M_{00} \cdot ratio,\ heigth = 2M_{00} \cdot ratio \tag{8}$$

The Mean Shift algorithm is carried out by iteratively computing new values of (x_c, y_c), which is the mass center of the search window. The algorithm can be stopped in the case where 0th is zero, it shows no difference in the window of zero intensity. The position and dimensions in the search window are rescaled until convergence. Fig. 1 shows the processing of the generation of the ROI region using the Camshift algorithm in AI car. Table 1 shows the Camshift algorithm steps.

TABLE I
CAMSHIFT ALGORITHM IS SUMMARIZED AS FOLLOWS
(INTEL CORPORATION,2001)

Step 1.Set the region of interest (ROI) of the probability distribution image to the entire image.
Step 2.Select an initial location of the Mean Shift search window. The selected location is the target distribution to be tracked.
Step 3.Calculate a color probability distribution of the region centred at the Mean Shift search window.
Step 4.Iterate Mean Shift algorithm to find the centroid of the probability image. Store the zeroth moment (distribution area) and centroid location.
Step 5.For the following frame, center the search window at the mean location found in Step 4 and set the window size to a function of the zeroth moment. Go to Step 3.



(a) origin frame

(b) mass center

(c) search window

Figure 1. The ROI in AI car

III. THE EXTRACTION OF FEATURE POINT

Scale Invariant Feature Transform was presented by Lowe, for extracting highly feature point from images[9]. The major steps of SIFT are detecting scale space extrema, localizing keypoints, assigning orientation and calculating descriptor[10]. In this paper, we make use of detected feature points to implement reliable matching of the same target between different pictures and generate a window for the tracking.

The first stage is seeking throughout all scales and locations in image. It is invariant to scale and orientation that potential interest points extracted in a difference -of-Gaussian image are[11]. The function of scale space in an picture is shown as follows:

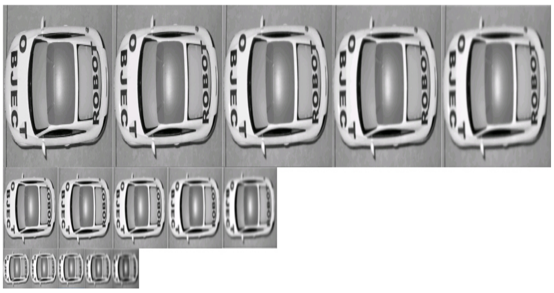
$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{9}$$

Here I(x,y) represents the input picture, G(x,y,σ) denotes a variable scale Gaussian, and * represents the convolution symbol. Two successive Gaussians are separated by a constant multiplicative factor k, and the DOG image is defined as such function:

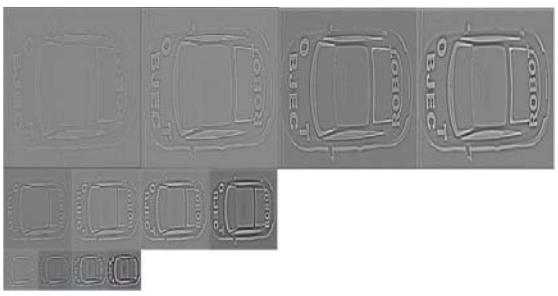
$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) * I(x,y,\sigma) \end{aligned} \tag{10}$$

Fig. 2 shows Gaussian and its DOG pyramids of the object in AI car. Fig. 2a is a set of scale space pyramid images, which is built by the input image convolved with Gaussian repeatedly. The DOG images shown in Fig. 2b are made of adjacent subtracted Gaussian images. Then, the Gaussian image is down-sampled by the factor of 2, and the process is done repeatedly.

(a) object



(b) Gaussian pyramid image

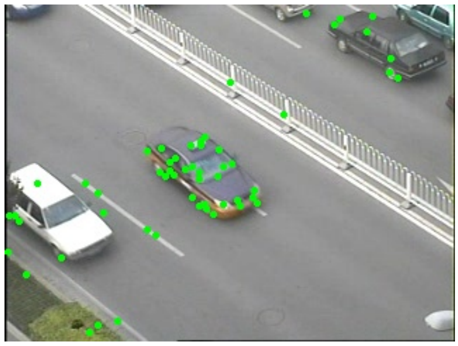


(c) DOG pyramid image

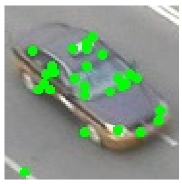
Figure 2. The sampled Gaussian and DOG images of the object

A sample point must be the local maxima or minima of DOG, which is in contrast with its 26 neighbors from the current ,above and below scales. At each candidate point, it is fit to look for the location and scale. In order to improve the stability and robustness of selected point, lots of keypoints will be removed among the candidate points, in case that the structure of points has low contrast, or the value of each point is estimated below a threshold. Fig. 3 shows the sift points extracted by the SIFT algorithm in AI car. There are 63 feature points in the whole frame shown in Fig. 3a, and 29 feature points in the ROI. Obviously, it is to decrease the search space of the frame that will reduce processing time of extraction, and avoid to extract the points unrelated to the object.

(a) The original frame



(b) 63 points in the whole frame



(c) 29 points in the ROI

Figure 3. The detection of feature points via SIFT in traffic video (No.279 frame)

IV. The VSS Algorithm Based on Multiple Features

The proposed VSS algorithm involves three steps. The first step is seeking the ROI of target by the meanshift approach based on color feature. The second is the extraction of SIFT points within the detected ROI. The last step is the recognition and tracing of the moving target. At last, the presented tracking method looks for rapidly the most accurate target by decreasing the search space. During Camshift tracking, a histogram of color is easily be computed, and the size of the search window can be updated. Due to noise and illumination, it is sensitive to some resembled colored background regions and objects that disturb the process of tracking. Thus, to offset the drawback, the VSS algorithm will generates a variable search space with candidate points for the accurate tracing of moving targets. A flowchart of the VSS algorithm is shown in Fig. 4.

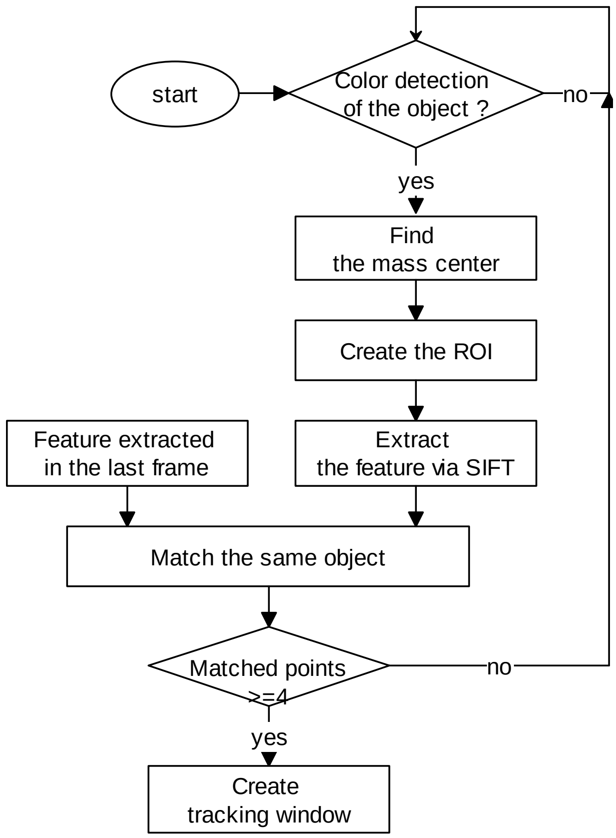


Figure 4. The flowchart for the VSS algorithm

The ROI of the object is set as the finding of the maximization via meanshift algorithm in each frame. The size of ROI is a variable region according to the size of the object. The next step involves that Gaussian pyramid pictures are made in the region, DOG pyramid pictures are made to Gaussian pyramid pictures. In the DOG pyramid pictures, maxima and minima candidate key points are found. Using the SIFT, feature points are extracted within the ROI only though screening. It should be used to implement stable matching of the same target between the ROI and original object image. We then take four feature points among correct matched points which have maximum or minimum axis value, and generate a rectangle window with them. The region of the rectangle represents the target that we are tracking.

The VSS algorithm is as follow:

Step 1 set a variable region as ROI using meanshift

Step 2 extract the robust feature with SIFT

Step 3 match feature points to make sure the identified object in the ROI is the same in the original object image

Step 4 make a rectangle with four correct-matched points in ROI, which have maximum or minimum axis value

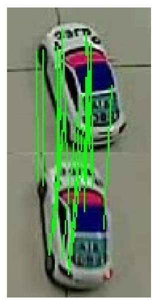
Step 5 stop tracing unless the number of correct matched points is under four

if not, repeat from steps 1 to 4

Fig. 5 shows the process of the generation of the VSS. In Fig. 5, 5b is the ROI that is found by color feature and represented by the green rectangle. To compensate for the weakness that it is sensitive to illumination and similar color background, we extract robust matching points within the ROI. This method reduce the calculation time greatly because of changing the search space from the whole frame to the ROI. It set four features into min_x, min_y, max_x and max_y, which generate a red rectangle in the ROI. Therefore, the tracking window then changes from the ROI to the region in red rectangle. Producing a variable tracking window in each picture can enhance the accuracy of the target tracking property.

(a) No.228 frame in AI car

(b) The ROI via the Camshift algorithm



(c) Matching the same object (d) Creating a tracing window

Figure 5. The processing of No.228 frame in AI car using the VSS algorithm

Fig. 6 The images of single target tracking result using the proposed algorithm in AI car and traffic video.

V. EXPERIMENTAL RESULTS

The presented approach is executed in Visual C++ and OpenCV. Table II denotes the detailed information for each video sequence. TWO video sequences are used in the experiments, AI car and traffic videos. In order to recognize the object handily, we create the original object image that is the standard image of the ideal object region.

In the experiments, the results shows the comparison of mean processing time between the proposed algorithm and meanshift. In case of AI car, it takes 0.052 s to process one frame, and spends 0.088 s on processing one frame. Due to extracting feature points, it is slower than the algorithm based on color. However, it is enough time for real-time tracking.

TABLE II
THE INFORMATION OF THE EXPERIMENTED VIDEO

Name	Total frame	Frame rate (frame/s)	size	No. of object
AI car	501	15	640*480	Single
Traffic	3269	25	320*240	multiple

Table 3 indicates the difference of the accuracy between the algorithms. To estimate the tracing accuracy, the function is shown as follows:

$$\text{accuracy}(\%) = \frac{\text{total frames} - \text{miss frames}}{\text{total frames}} \times 100 \quad (11)$$

Generally the accuracy of the presented algorithm is higher than other used methods. The average accuracy of the proposed algorithm is 95.63% in traffic video. And, the proposed method can increase the tracking accuracy at about 3.34%.

TABLE III
THE RESULT OF THE ACCURACY FOR THE ALGORITHMS

Name		Correct frame	Result (%)
AI Car	Camshift	488	97.41
	The proposed Algorithm	491	98
Traffic	Camshift	3017	92.29
	The proposed Algorithm	3126	95.63

Fig. 6 indicates the resulting pictures of one target tracing via different methods in AI car and traffic video. As the Fig. 6 shows, a red rectangle is the proposed algorithm, and its size denotes the window of tracking. The proposed algorithm offer more accurate tracking information. The size of the red rectangle is up to the size of the detected object. The object's size change a bigger one, while the red rectangle enlarge its size to the big, and vice versa. It is clear that the window created by the proposed algorithm adapted to the size of the target properly.

VI. CONCLUSION

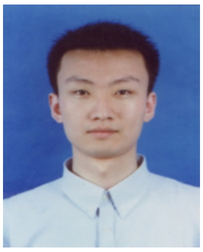
In this paper, we present VSS algorithm based on multiple feature to achieve the exact tracing of moving targets. So as to reduce the processing time to identify the object, it make use of the color feature of the tracked object to change the search space from a whole frame to the region of ROI only. The point feature is used to match the same object between the ROI and original object image to prohibit missing the object. Furthermore, the four matched feature points which have maximum or minimum axis value, generates a rectangle window to obtain an accurate tracking. Through the experiments, the proposed algorithm shows more effective and robust in tracking applications.

ACKNOWLEDGMENT

This research was supported by the Fundamental Research Funds for the Central Universities 3142013066.

REFERENCES

[1] S. Avidan, "Support vector tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.26, no. 10, pp. 1064-1072, Aug. 2004.
[2] X. Zhang, H. Liu, and X. Li, "Target Tracking for Mobile Robot Platforms via Object Matching and Background Anti-matching," *Journal of Robotics and Autonomous Systems*, vol. 58, no. 11, pp. 1197-1206, 2010.
[3] J-S Hu, C-W Juan, J-J Wang, "A spatial-color mean-shift object tracking algorithm with scale and orientation estimation," *Pattern Recogn Lett*. 29(16), 2165–2173(2008). doi: 10.1016/j.patrec. 2008. 08. 007
[4] Y Cheng, "Mean shift, mode seeking, and clustering." *IEEE Trans Pattern Anal Mach Intell*. 17(8), 790–799(1995). doi: 10.1109/34.400568
[5] G.R. Bradski. Computer vision face tracking for use in a perceptual user interface,1998.
[6] Y Akazawa, YOkada, K Niijima, "Robust tracking algorithm based on color and edge distribution for real-time video based motion capture systems," *in IAPR workshop on Machine Vision Applications*, pp. 60–63(2002)
[7] D.G. Lowe, "Distinctive image features from scale-invariant keypoints." *IJCV*. 60(2), 91–110(2004)
[8] Xu Bin, Yu Naigong, "Study on real-time feature extraction based on color label and scale invariant feature transform." *Computer Measurement & Control*. 2011. 19(2):409-411
[9] Yang Liu, Ma Zhao, Qinzhang Wu, "A new hybrid grid multiple model estimation based on STF." *Journal of Computers*, Vol 7, No 7 (2012), 1607-1614
[10] Zhibin Xue, Jianchao Zeng, Caili Feng, Zhen Liu, "Swarm target tracking collective behavior control with formation coverage search sgents & globally asymptotically stable analysis of stochastic swarm." *Journal of Computers*, Vol 6, No 8 (2011), 1772-1780
[11] Xiaoming Liu, Lian Pan, "The study of target tracking based on ARM embedded platform." *Journal of Computers*, Vol 7, No 8 (2012), 2015-2023



Bin Xu was born in 1980, and received his M.S. degree in Pattern Recognition and Intelligent System from Beijing University of Technology, China, in 2011. He is a lecturer in Department of Mechanical Electrical Engineering at North China Institute of Science and Technology, Sanhe, China. His major field of study include computer vision.

His research interests include image processing and artificial intelligence robot.

Xiaoju Shen was born in 1980. She is a lecturer at Department of Management, North China Institute of Science and Technology, Sanhe, China. Her research interests include economics and computer applications.

Feiji Ding was born in 1989, and received his bachelors degree from North China Institute of Science and Technology, Sanhe, China. His major field of study include mechanical design, manufacturing and automation.

His research interests include industrial robot and visual simulation.