Vision Based People Tracking System

Boukerch Haroun, Luo Qing Sheng, Li Hua Shi, Boukraa Sebti

Abstract**—**In this paper we present the design and the implementation of a target tracking system where the target is set to be a moving person in a video sequence. The system can be applied easily as a vision system for mobile robot. The system is composed of two major parts the first is the detection of the person in the video frame using the SVM learning machine based on the "HOG" descriptors. The second part is the tracking of a moving person it's done by using a combination of the Kalman filter and a modified version of the Camshift tracking algorithm by adding the target motion feature to the color feature, the experimental results had shown that the new algorithm had overcame the traditional Camshift algorithm in robustness and in case of occlusion.

Keywords**—**Camshift Algorithm, Computer Vision, Kalman Filter, Object tracking.

Open Science Index, Computer and Information Engineering Vol:13, No:11, 2019 waset.org/Publication/10010893

I. Introduction

TARGET identification and target tracking are important research topics in the computer vision field. This importance is due to the numerous research fields which include these two topics such as vehicle navigation, video surveillance, industrial automation and security, mobile robot and human interaction.

The first step of any tracking system is selecting the target which is in our case a moving person in the scene. The researches in [1]-[5] indicates that sliding window classifiers are presently the predominant method being used in object detection, or more specifically, human detection, due to their good performance. The sliding window approach is that for each image is scanned from the top left to the bottom right with rectangular sliding windows in different scales. For each sliding window, certain features such as edges, image patches, and wavelet coefficients are extracted and fed to a classifier, which has been already trained by labeled training data. The classifier will classify the sliding windows, which bound a person, as positive samples, and the others as negative samples. Currently, the Support Vector Machine (SVM) and variants of boosted decision trees are two leading classifiers for their good performance and efficiency [6]. In this paper we used a sliding window approach for the detection, based on the Histograms of Oriented Gradients (HOG) to detect the person [1], Using an SVM classifier.

The second step is the tacking, in recent research the Objects could be represented by a single point [7] or a set of points [8], by bounding Boxes (rectangle, ellipse) [9], silhouettes and contours for tracking complex non-rigid objects [10].

Boukerch Haroun is with College of Electro Mechanics, Beijing institute of technology, Beijing, China (e-mail: harounbest@hotmail.com).

Pr. Luo Qing Sheng and Boukraa Sebti are with College of Electro Mechanics, Beijing institute of technology, Beijing, China.

Li Hua Shi is with school of Mechanical Engineering, Beijing institute of technology, Beijing, China.

Many algorithms of object tracking have been proposed and developed such as the Mean shift which is an efficient pattern matching algorithm with no parameter estimation, and can be combined with other algorithms. It uses the kernel function histogram model of the target object. Meanshift is not sensitive for part of block rotation and deformation, but the track is easy to lapse when blocked a large area on the target [11], [12]. Continuously Adaptive Mean Shift tracking (CAMSHIFT) is the solution of the Meanshift problem where It combines the basic Mean-shift algorithm with an adaptive region-sizing step but the disadvantage of Camshift is affected by the probability distribution of the color image which means it's not very useful where the object's color is not different with Background's or a total occlusion happens [13]-[16]. Kalman filter is an efficient recursive filter, and provides an efficient method to calculate the state estimation process. It can estimate of the unknown past and current state of the signal, even to estimate the future state and can deal with occlusions problem [17]-[19].

Recently many researches were made to outcome the limits of the previous algorithm by generating a new hybrid algorithm which is a combination of two or more algorithms. In [12] and [20] the authors combine the Camshift and the Kalman algorithms by using the moving target information to improve the matching accuracy of the candidate target and object target to handle the occlusion problem. In [21] the authors had Extend the CAMSHIFT algorithm with the scene depth information; from the disparity image they were able to increase the tracking quality with acceptable losses in the execution time. Other work use the combination Meanshift, Camshift, Kalman filter [22]. Some researches modified the traditional Camshift algorithms to get better results [23]. By The use of multi-dominant color object localization and track the dominant color object separately [24] by adding feature to the Camshift beside the color feature such as the motion information [25].

In this paper we propose a vision system to detect and track a moving person in a scene, the detection part were done by using the sliding window approach based on the HOG feature, for the tracking of the target we had proposed a hybrid algorithm which combine the kalman filter with a modified version of the Camshift tracking algorithm by adding the target motion feature to the color feature, the experimental results shows that the new algorithm has overcame the traditional Camshift. The implementation of the system was done using OpenCv [26].

II. People Detection

In this section, the system architecture of our human detection system is presented (Fig. 1). This human detection system is using the Sliding window detection systems which scan the image at all relevant positions and scales to detect a person, consequently there are two major components [27]. The feature component encodes the visual appearance of the person; in this study we had used the Histogram of Oriented Gradients [1] as the main feature descriptor set to locate humans in the video scenes. Whereas the classifier determines for each sliding window independently whether it contains the person or not, a linear SVM classifier were used [1].

INRIA Person Dataset [28] was used which contain 2416 64*128 positive images (contain a person) and 347 64*128 negative image (don't contain a person) Fig. 2 shows some samples.

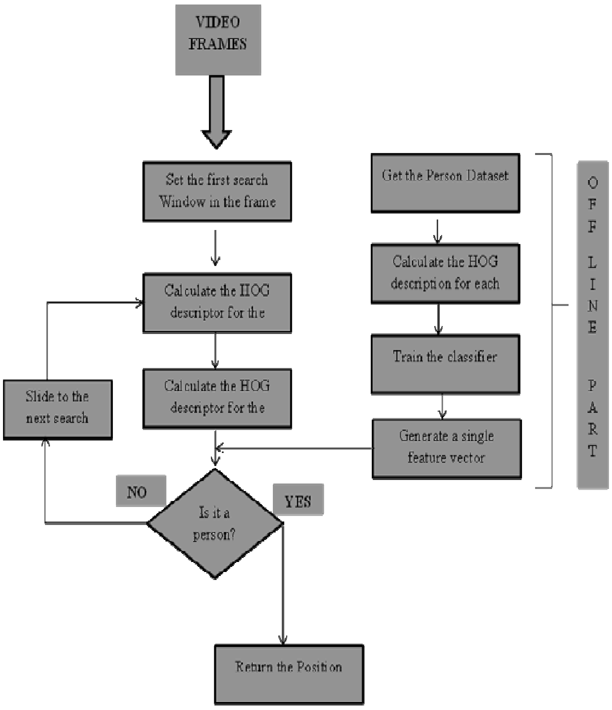


Fig. 1 Person detection system

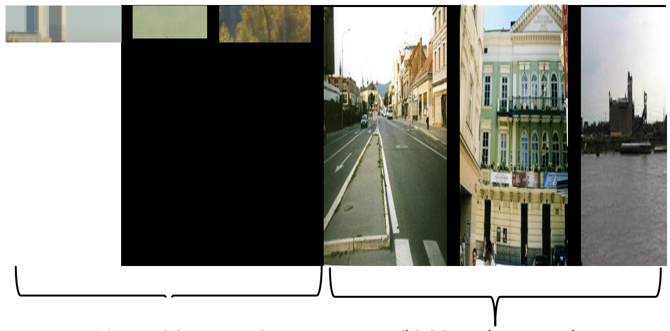
The algorithm can be split to two parts the first one is an offline the training the second one is online

A. Offline Treatment

1. For each image calculates the HOG features and keeps track of their classes (positive, negative)
2. Train the SVM classifier using machine learning algorithm [29]
3. Generate from the calculated support vectors and SVM model a single detecting feature vector.
4. Set the sliding window parameter

B. Online

1. Read the input video frame by frame
2. For each window in the current frame calculate the hog descriptor
3. Compare windows features with the descriptor vector decide where if it's contain a person or not if yes return the position and draw a boundary box around the person.
4. If no person found in the current frame pass to the next frame.



(a) Positive sample (b) Negative sample

Fig. 2 Sample of INIRIA person data set

III. People Tracking

A. Camshift Algorithm:

The CAMSHIFT algorithm can be summarized with these steps [30]:

1. Choose the initial region of interest, which contains the object we want to track.
2. Make a color histogram of that region as the object model.
3. Make a probability distribution of the frame using the color histogram. As a remark, in the implementation, they use the histogram back projection method.
4. Based on the probability distribution image, find the center mass of the search window using mean-shift method.
5. Center the search window to the point taken from step 4 and iterate step 4 until convergence.
6. Process the next frame with the search window position from the step 5.

In order CAMSHIFT can track colored object, it needs a probability distribution image. They use the HSV color system and using only hue component to make the object's color 1D histogram [30] this histogram is stored to convert next frames into corresponding probability of the object. The probability distribution image itself is made by back projecting the 1D hue histogram to the hue image of the frame. The result called backproject image. CAMSHIFT is then used to track the object based on this backproject image. In terms of statistics, the values stored in backProjection represent the probability that a pixel in the Image belongs to the object area.

The mean location of the probability image inside search window is computed using

$$M_{00} = \sum_x \sum_y I(x,y) \quad (1)$$

$$M_{10} = \sum_x \sum_y XI(x,y) \quad (2)$$

$$M_{01} = \sum_x \sum_y yI(x,y) \quad (3)$$

$$(x_c\,, y_c) = \left({}^{M_{10}}/_{M_{00}}, {}^{M_{01}}/_{M_{00}} \right) \quad (4)$$

where M00 in (1) is the zeroth moment, M10 in (2) and M01 in (3) are the first moments, these moments could be used to compute the next center position of the tracking window xc and yc as shown in (4).

$$M_{20} = \sum_x \sum_y x^2 \times I(x,y) \quad (5)$$

$$M_{02} = \sum_x \sum_y y^2 \times I(x, y) \quad (6)$$
$$M_{11} = \sum_x \sum_y x \times y \times I\,(x, y) \quad (7)$$

The orientation of the major axis and the scale of the distribution are determined by finding an equivalent rectangle that has the same moments as those measured from the 2D probability distribution image. The orientation is defined by (8), using the first and second moments in (5)-(7).

$$\theta = \frac{\tan^{-1}\left(\frac{2\times\left(\frac{M_{11}}{M_{00}} - x_c \times y_c\right)}{\left(\frac{M_{20}}{M_{00}} - x_c^2\right) - \left(\frac{M_{02}}{M_{00}} - y_c^2\right)}\right)}{2} \quad (8)$$

The first two eigenvalues (the length and width of the probability distribution) are calculated in closed form as follows in (9)-(11):

$$a = \frac{M_{20}}{M_{00}} - x_c^2 \quad (9)$$
$$b = 2 \times \left(\frac{M_{11}}{M_{00}} - x_c \times y_c\right) \quad (10)$$
$$c = \frac{M_{02}}{M_{00}} - y_c^2 \quad (11)$$

The length and width of the distribution around the centroid is then given by (12) and (13):

$$l = \sqrt{\frac{(a+c) + \sqrt{b^2 + (a-c)^2}}{2}} \quad (12)$$
$$w = \sqrt{\frac{(a+c) - \sqrt{b^2 + (a-c)^2}}{2}} \quad (13)$$

B. Lucas-Kanade Algorithm

The aim of the Lucas-Kanade Algorithm is, for a given set of points in a video frame find those same points in the next frame. Or for given point [u_x,v_y] in frame F; find the point $[u_x + \delta x\,, u_y + \delta y]^t$ in image F1 that minimizes ε as shown in (14) [25]:

$$\varepsilon(\delta_x, \delta_y) = \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} (F_1(x, y) - F_2(x + \delta_x, y \; + \delta_y)) \quad (14)$$

C. Kalman Filter

The Kalman filter, also known as linear quadratic estimation (LQE), is an algorithm that uses a series of measurements observed over time. The input parameters of the filter are, respectively, the position of the object in the image at time k, the size of the object and the width and length of the search window of the object which varies due to the mobility of the object during the sequence. These parameters represent the state vector and measurement vector of the Kalman filter. In general a Kalman filter is a process that requires the following steps:

- The estimate, which updates the position of the object.
- The measure in our case we used the combination of Camshift and lucase-canade.
- The prediction, which calculates the position of the object in the next frame.

Because of the algorithm's recursive nature, it can run in real time using only the present input measurements and the previously calculated state; no additional past information is required.

D. Overview of the Proposed Algorithm

For the people tracking system we first had used the Camsihft algorithm but the tests has shown some insufficiency of robustness and occlusion handling problems.

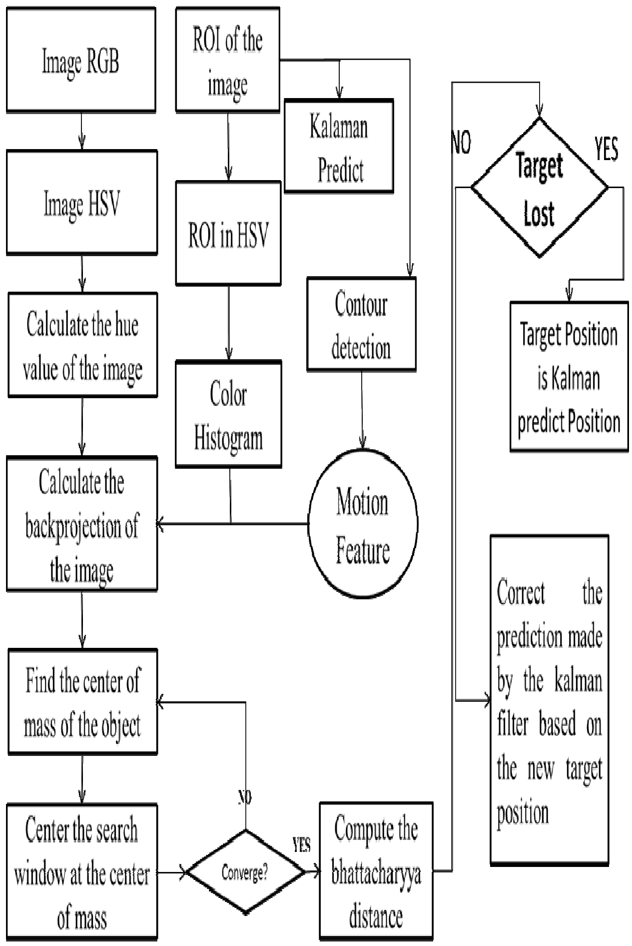


Fig. 3 Flow chart of the tracking system

To improve the robustness we had added the motion features to the color features already used by the Camshift By combining the Lucas-Canade with it

To handle the occlusion problems we had added the Kalman filter to the Camshift, so the combination of the three previous algorithms had generated a new algorithm as follows:

1. predict the position of the target in the current frame using the kalman predict block
2. calculate the hue value of the current frame and the color histogram of the region of interest (ROI)
3. calculate the backprojection of the current image
4. find the contour in the ROI, threshold the contours based on their area
5. run lucas-canade algorithm, to find the position of the contours in the current frame based on the position of the contours in the previous frame
6. update the backprojection image by giving more weightage to the pixels found by the lucas-canade algorithm

7. run the Camshift algorithm using the new backprojection image
8. calculate the Bhattacharyya distance between the previous target position and the new target position to find if there is an occlusion or loss of the target

a. no occlusion detected: correct the prediction made by the kalman filter based on the new target position
b. Occlusion detected: the new position of the target is based on the prediction block of the kalman filter.

The proposed algorithm is shown in Fig. 3

IV. EXPERIMENTAL RESULTS

The system was implanted on a PC with Intel celron E3400 2.60Ghz processor and 2G main memory, built-in video capture card .In the IDE Vs2012, by using the OpenCv library. The experimental result as follows. Fig. 4 shows identification of the target in several video sequences we had used where we had always positive detection of the person and none negative detection; Fig. 5 shows the result of the proposed tracking algorithm. Through comparing and analyzing the experimental results between the camshift algorithm and the algorithm we proposed, we can get a conclusion that both methods can complete tracking. But the proposed algorithm shows strength in the occlusion cases where the camshift will loss the target, the tests shows that using the modified backprojection image had increased the robustness against the background similarity with object.

V.CONCLUSION

In this paper, a vision based people tracking system is proposed. Where several algorithm where combined to develop the whole system. A sliding window using HOG descriptor approach was used to detect the person in the frame. For tracking, a combination of Kalman filter and a modified version of Cmashift algorithm were used to track the person in the next frames. The experimental result demonstrated the efficiency and reliability of this algorithm proposed. As further work we will try to add other features to the camshift to add it's robustness against background similarity with the object.

Fig. 4 Person detection in the image

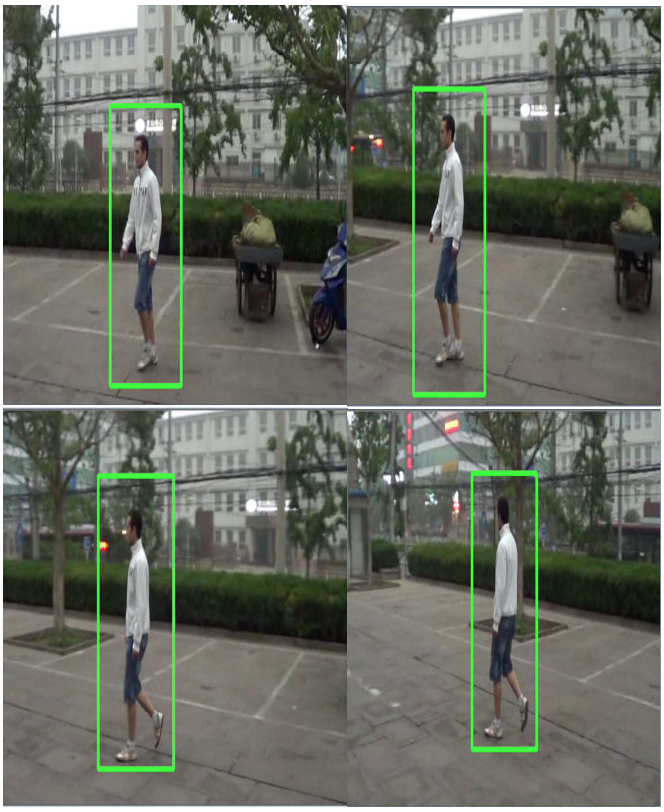


Fig. 5 Person tracking

REFERENCES

[1] Dalal, N. and Triggs, B. Histograms of oriented gradients for human detection. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR-05), 2005 886–893.
[2] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In CVPR, 2008.
[3] C. H. Lampert, M. B. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In CVPR, 2008
[4] O. Tuzel, F. Porikli, and P. Meer. Human detection via classification on Riemannian manifolds. In CVPR, pages 1–8, 2007
[5] S. Munder and D. Gavrila. An experimental study on pedestrian classification. IEEE Trans. Pattern Anal. Mach. Intell., 28(11):1863–1868, Nov. 2006
[6] Wang, X., Han, T. X., & Yan, S. An HOG-LBP human detector with partial occlusion handling. In Computer Vision, 2009 IEEE 12th International Conference on (pp. 32-39). September 2009.
[7] Veenman, C. Reinders, M., and Backer, E. 2001. Resolving motion correspondence for densely moving points. IEEE Trans. Patt. Analy. Mach. Intell. 23, 1, 54–72
[8] Serby, D., Koller-Meier, S., and Gool, L. V. 2004. Probabilistic object tracking using multiple features. In IEEE International Conference of Pattern Recognition (ICPR). 184–187
[9] Comaniciu, D., Ramesh, V., Andmeer, P. 2003. Kernel-based object tracking. IEEE Trans. Patt. Analy. Mach Intell. 25, 564–575
[10] Yilmaz, A., Li, X., and Shah, M. 2004. Contour based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Trans. Patt. Analy. Mach. Intell. 26, 11, 1531–1536.
[11] Cheng, Yizong (August 1995). "Mean Shift, Mode Seeking, and Clustering". IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE) 17 (8): 790–799
[12] Huang, S., & Hong, J. (2011, April). Moving object tracking system based on camshift and Kalman filter. In Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference on (pp. 1423-1426). IEEE
[13] G. Bradski, and T. Ogiuchi, and M. Higashikubo, "Visual Tracking

Algorithm using Pixel-Pair Feature". International Conference on Pattern Recognition, no. 4, pp. 1808–1811. 2010)
[14] Y. Ruiguo, and Z. Xinrong, "The Design and Implementation of Face Tracking in Real Time Multimedia Recording System". IEEE Transaction, no. 3, pp. 1–3. 2009
[15] E. David, and B. Erich, and K. Daniel, and S. Anselm, "Fast and Robust Camshift Tracking". IEEE Transaction, no. 8, pp. 1–8. 2010
[16] Boubou, S., Kouno, A., & Suzuki, E. (2011, December). Implementing Camshift on a Mobile Robot for Person Tracking and Pursuit. In Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on (pp. 682-688). IEEE
[17] X. Sun and H. Yao, and S. Zhang, "A Refined Particle Filter Method for Contour Tracking". Visual Communications and Image Processing, no. 8, pp. 1–8. 2010
[18] Welch, G., & Bishop, G. (1995). An introduction to the Kalman filter
[19] Salmond, D. (2001, October). Target tracking: introduction and Kalman tracking filters. In Target Tracking: Algorithms and Applications (Ref. No. 2001/174), IEE (pp. 1-1). IET
[20] Wang, X., & Li, X. (2010, December). The study of MovingTarget tracking based on Kalman-CamShift in the video. In Information Science and Engineering (ICISE), 2010 2nd International Conference on (pp. 1-4). IEEE
[21] Kovacevic, J., Juric-Kavelj, S., & Petrovic, I. (2011, May). An improved CamShift algorithm using stereo vision for object tracking. In MIPRO, 2011 Proceedings of the 34th International Convention (pp. 707-710). IEEE
[22] Salhi, A., & Jammoussi, A. Y. (2012). Object tracking system using Camshift, Meanshift and Kalman filter. World Academy of Science, Engineering and Technology
[23] Li, J., Zhang, J., Zhou, Z., Guo, W., Wang, B., & Zhao, Q. (2011, October). Object tracking using improved Camshift with SURF method. In Open-Source Software for Scientific Computation (OSSC), 2011 International Workshop on (pp. 136-141). IEEE
[24] Hidayatullah, P., & Konik, H. (2011, July). CAMSHIFT improvement on multi-hue and multi-object tracking. In Electrical Engineering and Informatics (ICEEI), 2011 International Conference on (pp. 1-6). IEEE
[25] Fahad Fazal Elahi Guraya, Pierre-Yves Bayle and Faouzi Alaya Cheikh, People Tracking via a Modified Camshift Algorithm, DCABES;2009
[26] Gary Bradski, Adrian Kaehler, "Learning OpenCV", O'Reilly, 2008
[27] Schiele, B., Andriluka, M., Majer, N., Roth, S., & Wojek, C. (2009, May). Visual people detection: Different models, comparison and discussion. In Proceedings of the IEEE ICRA 2009 workshop on people detection and tracking (Vol. 12)
[28] http://pascal.inrialpes.fr/data/human/
[29] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 27.
[30] Bradski, G. R. "Computer Vision Face Tracking for Use in a Perceptual User Interface". Intel Technology Journal, 2(2), 13-27, 1998