A Robust Tracking Algorithm Based on HOGs Descriptor

Daniel Miramontes-Jaramillo[1], Vitaly Kober[1,2],
and Víctor Hugo Díaz-Ramírez[3]

[1] CICESE, Ensenada, B.C. 22860, México
dmiramon@cicese.edu.mx, vkober@cicese.mx
[2] Department of Mathematics, Chelyabinsk State University, Russian Federation
[3] CITEDI-IPN, Tijuana, B.C. 22510, México
vhdiaz@citedi.mx

Abstract. A novel tracking algorithm based on matching of filtered histograms of oriented gradients (HOGs) computed in circular sliding windows is proposed. The algorithm is robust to geometrical distortions of a target as well as invariant to illumination changes in scene frames. The proposed algorithm is composed by the following steps: first, a fragment of interest is extracted from a current frame around predicted coordinates of the target location; second, the fragment is preprocessed to correct illumination changes; third, a geometric structure consisting of disks to describe the target is constructed; finally, filtered histograms of oriented gradients computed over geometric structures of the fragment and template are matched. The performance of the proposed algorithm is compared with that of similar state-of-the-art techniques for target tracking in terms of objective metrics.

1 Introduction

Increasing available computing power has made real-time tracking feasible. Object tracking systems are used for applications such as video surveillance, motion based recognition, and vehicle navigation. Tracking requires processing large amounts of data. Two approaches can be taken: reduce the amount of information to be processed, and carry out the processing faster. In the former approach, features are usually computed. A feature extractor ideally outputs a small number of features. Matching these features across frames yields the displacement information. When the camera rate is high, preprocessing might be done by subtracting the background of a given frame from the next one, so that only information in the area where movement took place is left in the frame. The tracking quality can be affected by the presence of additive sensor's noise and cluttering background, geometric distortion of a target, occlusion, exiting and re-entering of the target to the observed scene, illumination changes, and real-time requirements. In this paper, we propose a tracking algorithm, which deals with the problems using matching of filtered histograms of oriented gradients computed in circular sliding windows.

E. Bayro-Corrochano and E. Hancock (Eds.): CIARP 2014, LNCS 8827, pp. 54–61, 2014.

Basically, there are three approaches for representation of an object of interest in descriptor based tracking algorithms [1]: by keypoints or features, by silhouette, and by kernel. The first one utilizes a set of features describing the object and further used for matching [2]. The most popular matching algorithms based on keypoints are SIFT [3] and SURF [4]. These algorithms and their variants can be used for designing various tracking algorithms. Recently, a tracking-learning-detection (TLD) algorithm for real-time target tracking was introduced [5]. This method learns past detection errors and self-adjusts to avoid the errors in the future. The second approach uses the object contour for matching in each frame. Such descriptor is flexible to shape changes of the object silhouette. For instance, a tracking method based on the mean-shift algorithm and silhouette descriptor has been recently proposed [6]. Finally, the kernel approach uses area and pixel characteristics of the object to generate statistical and structural invariants [7].

In this paper we suggest an algorithm exploiting the last approach. In general, the proposed algorithm consists of the following stages: preprocessing, matching, and prediction. The preprocessing stage carries out illumination normalization [8,9] over a selected (predicted) fragment of a current frame. The second stage computes descriptors (filtered histograms of oriented gradients [10]) over geometric structures [11] of the fragment and template and performs matching. The last stage predicts the position of the object in the next frame using a kinematic model [12,13]. The fragment of interest in the next frame is formed around the predicted target location. This helps to reduce significantly the processing time. The performance of the proposed algorithm in a test database is compared with that of the SIFT and SURF based tracking algorithms in terms of accuracy and processing time.

2 Proposed Approach

2.1 Preprocessing

First, a geometric structure describing a target and consisting of disks in an image fragment extracted from a frame to be processed. Let us define a set of circular windows$\{W_i,\, i = 1, ..., M\}$ in a target fragment as a set of closed disks:

$$W_i = \left\{ (x, y) \in \mathbb{R}^2 \ : (x - x_i)^2 + (y - y_i)^2 \le r \right\}, \qquad (1)$$

where (x_i, y_i) are the coordinates of the center and r is the radius of the disks. M is the number of circular windows filling inside an object of interest in the fragment. Numerous experiments have shown that the number of circular windows may be chosen from 2 to 4 to yield good matching performance. The disks form a geometric structure that runs across a frame fragment with distances $\{D_{ij}\}$ between the window centers and angles $\{\gamma_i\}$ between every three adjacent centers of the circular windows defined as follows:

$$\left\{ D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2},\, i = 1, ...M;\, j = i + 1, ..., M \right\}, \qquad (2)$$

$$\left\{\gamma_i = \cos^{-1}\left[\frac{D_{i,j+1}^2 + D_{i,j+2}^2 - D_{i,j+3}^2}{2D_{i,j+1}D_{i,j+2}}\right], i = 1, ..., M-2\right\}. \tag{3}$$

Histograms of oriented gradients are good descriptors for matching because they possess a high discriminant capability and robust to small image deformations such as rotation and scaling. The histograms are calculated over the sliding geometric structure.

Next, we describe the suggested preprocessing for illumination correction [9]. Assume that a frame fragment is distorted by a slow-varying illumination function. If a frame fragment $\{f(x,y) : (x,y) \in R_f\}$ is sufficiently small then the signal can be considered uniformly illuminated in the fragment area R_f. In this case the correction can be carried out as follows:

$$\hat{f}(x,y) = a_{x,y} f(x,y) + b_{x,y}, \tag{4}$$

where $a_{x,y}$ and $b_{x,y}$ are coefficients, which can be computed with the least mean square estimation. The mean-squared-error (MSE) between $f(x,y)$ and target $t(x,y)$ inside the region of support R_t of the target can be written as:

$$MSE(\alpha,\beta) = \sum_{(x,y)\in R_t} \left(a_{\alpha,\beta} f(x+\alpha, y+\beta) + b_{\alpha,\beta} - t(x,y)\right)^2, \tag{5}$$

the estimates of $a_{x,y}$ and $b_{x,y}$ are given by

$$a_{\alpha,\beta} = \frac{\frac{1}{N}\sum_{(x,y)\in R_t} t(x,y)\, f(x+\alpha, y+\beta) - \mu_t \mu_f(\alpha,\beta)}{\mu_t^2(\alpha,\beta) - \mu_f^2(\alpha,\beta)}, \tag{6}$$

$$b_{\alpha,\beta} = \mu_t - a_{\alpha,\beta}\mu_f(\alpha,\beta), \tag{7}$$

where N is the number of signal elements in R_t; μ_t and μ_f are the sample mean values of the target and the frame fragment inside the region of support of the target, respectively.

2.2 Matching

At each position of the *ith* circular window on a frame fragment we compute gradients inside the window with the help of the Sobel operator [14]. Next, using the gradient magnitudes $\{Mag_i(x,y) : (x,y) \in W_i\}$ and orientation values $\{\varphi_i(x,y) : (x,y) \in W_i\}$ quantized for Q levels, the histogram of oriented gradients can be computed as follows:

$$HoG_i(\alpha) = \begin{cases} \sum\limits_{(x,y)\in W_i} \delta(\alpha - \varphi_i(x,y)), & Mag_i(x,y) \geq Med \\ 0, & otherwise, \end{cases} \tag{8}$$

where $\alpha = \{0, ..., Q-1\}$ are histogram values (bins), Med is the median value inside of the circular window, and $\delta(z) = \begin{cases} 1, & z=0 \\ 0, & otherwise \end{cases}$ is the Kronecker delta function. The calculation in Eq.(8) requires approximately $\left[\pi r_i^2\right]$ addition operations. In order to reduce computational complexity the calculation of the histograms at the sliding window position k can be performed in a recursive manner as follows:

$$HoG_i^k(\alpha) = HoG_i^{k-1}(\alpha) - \sum_{(x,y)\in OutP_i^{k-1}} \delta\left(\alpha - Out\varphi_i^{k-1}(x,y)\right) + \sum_{(x,y)\in InP_i^k} \delta\left(\alpha - In\varphi_i^k(x,y)\right), \tag{9}$$

where $OutP_i^{k-1}$ is a set of outgoing orientation values whose pixels belong to the half of the perimeter of the sliding window at step $k-1$; and InP_i^k is a set of incoming orientation values whose pixels belong to the half of the perimeter of the sliding window at step k. The computational complexity of this calculation is approximately $[2\pi r_i]$ addition operations. The recursive calculation can be used along columns as well as rows.

We utilize a normalized correlation operation for comparison of the histograms of the target and frame fragments. Let us compute a centered and normalized histogram of oriented gradients of the target as follows:

$$\overline{HoG_i^R}(\alpha) = \frac{HoG_i^R(\alpha) - Mean^R}{\sqrt{Var^R}}, \tag{10}$$

where $Mean^R$ and Var^R are sample mean and variance of the histogram, respectively.

The correlation output for the ith circular window at position k can be computed with the help of the fast Inverse Fourier Transform [14] as follows:

$$C_i^k(\alpha) = IFT\left[\frac{HS_i^k(\omega)\, HR_i^*(\omega)}{\sqrt{Q\sum_{q=0}^{Q-1}\left(HoG_i^k(q)\right)^2 - \left(HS_i^k(0)\right)^2}}\right], \tag{11}$$

where $HS_i^k(\omega)$ is the Fourier Transform of the histogram of oriented gradients inside of the ith circular window over the frame fragment, and $HR_i(\omega)$ is the Fourier Transform of $\overline{HoG_i^R}(\alpha)$; the asterisk denotes complex conjugate. The correlation peak is a measure of similarity of the two histograms, which can be obtained as follows:

$$P_i^k = \max_{\alpha}\left\{C_i^k(\alpha)\right\}. \tag{12}$$

The correlation peaks are in the range of $[-1, 1]$. We assign a correlation peak threshold value Th_Q that yields a trade-off between the probabilities of miss and false alarm errors for Q histogram bin. In order to take final decision about the presence of the reference object at the position k, the distances $\{D_{ij}\}$ between the window centers and angles $\{\gamma_i\}$ between every three adjacent centers of circular windows of the geometric structure are considered. Computation of the centered and normalized histograms for all circular windows over the target image as well as their Fourier Transforms can be done as preprocessing.

2.3 Prediction

After a frame at discrete time τ was processed, we save its state in the form of a vector $[k_\tau, \phi_\tau]$, where the values of the vector are the position $k = (x, y)$ and direction ϕ of the object. The vector can be rewriten as $[x_\tau, y_\tau, \phi_\tau]$. To speed-up the tracking, a prediction stage is implemented. In order to improve the state estimates for the next frame $\tau + 1$, we take into account information from past and current frames and state vectors to predict the next frame vector $[x_{\tau+1}, y_{\tau+1}, \phi_{\tau+1}]$. A state-space motion model [12] can be utilized. The target behavior is described by a coordinated turn model [15] as follows:

$$\begin{aligned}
x_{\tau+1} &= x_\tau + \frac{\sin(\phi_\tau \varDelta)}{\phi_\tau}\hat{x}_\tau - \frac{1-\cos(\phi_\tau \varDelta)}{\phi_\tau}\hat{y}_\tau + A_{x,\tau}\frac{\varDelta^2}{2}, \\
y_{\tau+1} &= y_\tau + \frac{1-\cos(\phi_\tau \varDelta)}{\phi_{tau}}\hat{x}_\tau + \frac{\sin(\phi_\tau \varDelta)}{\phi_\tau}\hat{y}_\tau + A_{y,\tau}\frac{\varDelta^2}{2}, \\
\hat{x}_{\tau+1} &= \cos(\phi_\tau \varDelta)\,\hat{x}_\tau - \sin(\phi_\tau \varDelta)\,\hat{y}_\tau + A_{x,\tau}\varDelta, \\
\hat{y}_{\tau+1} &= \sin(\phi_\tau \varDelta)\,\hat{x}_\tau - \cos(\phi_\tau \varDelta)\,\hat{y}_\tau + A_{y,\tau}\varDelta, \\
\phi_{\tau+1} &= \phi_\tau + A_{\phi,\tau},
\end{aligned} \tag{13}$$

where x_τ and y_τ are the position of the target in frame τ in Cartesian coordinates, $\hat{x}_\tau$ and $\hat{y}_\tau$ are velocity components in x and y directions, ϕ_τ is the target angular rate, $A_{x,\tau}$ and $A_{y,\tau}$ are random variables representing acceleration in x and y directions, and $A_{\phi,\tau}$ is the angular acceleration. Actually, the predicted position of the target does not coincide with the actual target position. So, we take a frame fragment around the predicted coordinates for further precise matching

3 Experimental Results

In this section we present computer simulation results. The experiment is carried out using 10 synthetic sequences, each one consists of 240 frames. Sequences contain different trajectories of a target. The target has the size of 144 × 144 pixels, circular windows have the radius of $r = 28$ pixels. Frames of the video sequences are of size 640 × 480 pixels. The target is taken from the Amsterdam Library of Object Images[16], while the video sequences are generated from scene images.

Each sequence is composed with arbitrary target orientations (in-plane rotation ranging from 0° to 360°) and geometric distortions (out-of-plane rotation ranging from 0° to 35°and scaling by a factor of $[0.8, 1.2]$). The parameters of the proposed algorithm are as follows: $M = 2$, $Q = 64$, and $Th_Q = 0.7$. The algorithm was implemented in a standard PC with an Intel Core i7 processor with 3.2 GHz and 8 GB of RAM using OpenCV with multithreading from OpenMP library.

The performance of the proposed algorithm was compared with that of similar tracking algorithms implemented on the base of the SIFT and SURF matching algorithms using OpenCV. Note that tracking algorithms based on the SIFT and SURF have been recently proposed [17,18], however, their codes are not available.

The performance of the tested algorithms for in-plane/out-of-plane rotations is shown in Fig. 1. It can be seen that the proposed algorithm yields the best in-plane rotation invariance among the tested algorithms and similar performance with the SIFT algorithm for out-of-plane rotation.

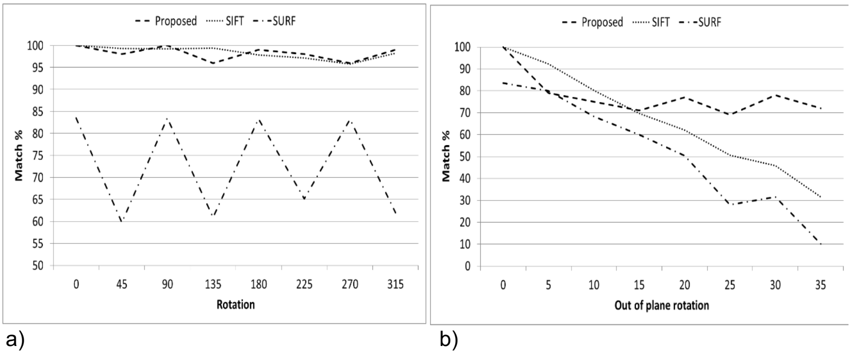

Fig. 1. Matching results in video frames for: a) in-plane rotation, b) out-of-plane rotation

Table 1 shows the performance of the tested algorithms in terms of average miss and false alarm error rate with in-plane and out-of-plane rotations and slight scaling of input scenes by a factor of $[0.8, 1.2]$. It can be seen that the proposed algorithm yields the best performance for in-plane and out-of-plane rotations and similar performance to that of the SIFT based algorithm for scaling.

Figure 2 shows the performance of the tested algorithms in terms of frames per second (FPS) over video sequences. As it is expected, the SIFT based algorithm is slowest, whereas the SURF based algorithm is fastest. The proposed algorithm is able to track objects with speed of 20 FPS using no specialized hardware such as GPUs or FPGAs.

Table 1. Performance of the tested tracking algorithms in terms of average miss and false alarms error rate over 10 video sequences

Algorithm	Rotation in-plane	Rotation out-of-plane	Scale
Proposed	1.25%	22.38%	5.89%
SIFT	1.64%	33.48%	5%
SURF	27.35%	48.54%	25.56%

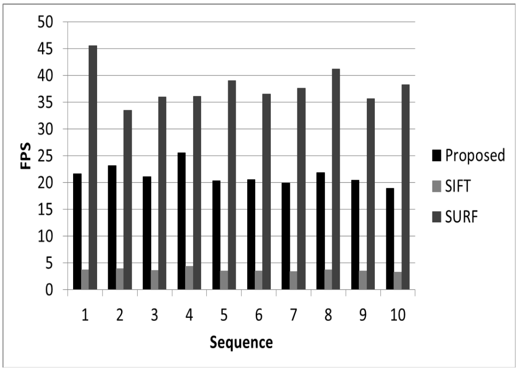

Fig. 2. Speed of tracking for the tested algorithms over 10 image sequences

4 Conclusion

In this paper we proposed a robust tracking algorithm based on HOGs descriptor computed over circular windows. The proposed algorithm employs a prediction stage based on modeling the kinematic behavior of a target in two-dimensional space. Based on predicted states the algorithm extracts from the input frame a small fragment to perform accurate and fast target state estimation by HOGs matching. According to computer simulation results the proposed algorithm showed a superior performance in terms of tracking accuracy and speed of processing comparing with similar tracking techniques based on features matching. It is expected that implementation of the proposed algorithm with GPU devices will help us to achieve 30 FPS rate of processing.

Acknowledgments. This work was supported by the Ministry of Education and Science of Russian Federation (grant 2.1766.2014K).

References

1. Yilmaz, A., Javed, O., Shah, M.: Object Tracking: A Survey. ACM Computer Surveys 38(4), 45 p. (2006)
2. Sethi, I., Jain, R.: Finding trayectories of feature points in a molecular image secuence. IEEE Transactions on Pattern Analysis and Machine Intelligence 9(1), 56–73 (1987)
3. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proc. Int. Conference on Computer Vision, vol. 2, pp. 1150–1157 (1999)
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. Comput. Vis. Image Underst. 110(3), 346–359 (2008)
5. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-Learning-Detection. IEEE Trans. Pattern Anal. 34(7), 1409–1424 (2012)
6. Talu, F., Turkoglu, I., Cebeci, M.: A hybrid tracking method for scaled and oriented objects in crowded scenes. Expert Systems with Applications 38, 13682–13687 (2011)
7. Nejhum, S., Ho, J., Yang, M.H.: Online visual tracking with histograms and articulating blocks. Computer Vision and Image Understanding, 901–914 (2010)
8. Díaz-Ramírez, V.H., Kober, V.: Target recognition under nonuniform illumination conditions. Appl. Opt. 48, 1408–1418 (2009)
9. Martínez-Díaz, S., Kober, V.: Nonlinear synthetic discriminant function filters for illumination-invariant pattern recognition. Opt. Eng. 47(6) (2008)
10. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)
11. Miramontes-Jaramillo, D., Kober, V., Díaz-Ramírez, V.H.: CWMA: Circular Window Matching Algorithm. In: Ruiz-Shulcloper, J., Sanniti di Baja, G. (eds.) CIARP 2013, Part I. LNCS, vol. 8258, pp. 439–446. Springer, Heidelberg (2013)
12. Rong Li, X., Jilkov, V.P.: Survey of maneuvering target tracking. Part I. dynamic models. IEEE Trans. on Aerosp. Electron. Sys. 39(4), 1333–1364 (2003)
13. Díaz-Ramírez, V.H., Picos, K., Kober, V.: Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. Opt. Comm. 323, 32–43 (2014)
14. Pratt, W.K.: Digital Image Processing. John Wiley & Sons (2007)
15. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behavior. IEEE Trans. Syst. Man Cybern. C Appl. Rev. 34(3), 334–352 (2004)
16. Geusebroek, J.M., Burghouts, G.J., Smeulders, A.W.M.: The Amsterdam library of object images. Int. J. Computer Vision 61(1), 103–112 (2005), `http://staff.science.uva.nl/~aloi/`
17. Zhou, H., Yuan, Y., Shi, C.: Object Tracking Using SIFT Features and Mean Shift. Comput. Vis. Image Underst. 113(3), 345–352 (2009)
18. Zhou, D., Hu, D.: A robust object tracking algorithm based on SURF. In: Int. Conf. on Wireless Comm. Sign. Proc., pp. 1–5 (2013)