Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science (EECSS 2015)
Barcelona, Spain, July 13 - 14, 2015
Paper No. 330

A Spatial Feature Descriptor for Object Tracking

Zhu Li, Qinghua Sheng, Wenyu Cai
Hangzhou Dianzi University
Xiasha Higher Education Zone, Hangzhou, China
First.lz1126@hdu.edu.cn; Second.Sheng7@hdu.edu.cn; Third.caiwy@hdu.edu.cn

Abstract- Object tracking for image sequences captured with a moving camera is very important for several applications such as Robot Vision, SLAM, ITS, and video surveillance systems. However, it is difficult to realize accurate tracking using only local feature descriptor such as HSV histograms, edge histograms, HOG histograms, and SIFT features, because it is affected by several phenomena such as illumination change, viewpoint change, size change, and noise. In order to realize robust tracking, global spatial distribution should be considered. In this paper, we propose a method that considersboth local feature descriptor and global spatial distribution. Experiments show that the proposed method realizes robust tracking.

Keywords: Object Tracking Feature Spatial GA Assignment

1 Introduction

Object tracking for image sequences captured with a moving camera plays an important role for motion picture analysis. Many feature descriptors such as HSV[1], HOG[2], and SIFT[3] have been proposed for applications of tracking object. Mean-Shift[4] and Particle Filter[5] are very popular in current research. In general, Mean-Shift calculates the similarities between templates and objects by distance measure on HSV histograms. In the algorithm of Particle Filter, the weight of each particle describes its likelihood and the weights of all particles represents the estimate of the posterior. The weight of each particle can be computed using many methods, the most common one of them is template matching which normally calculates the similarities by using HSV histograms. However, object tracking in such a situation poses several challenges such as illumination change, viewpoint change, size change, and noise. These phenomena lead to inaccurate object tracking because they considerably alter the visual features of the object in the current frame from those of the template. In order to resolve such issues, several studies have been conducted with the aim of determining a feature descriptor that is robust to changes in the environment. The SIFT feature is robust to many conditions and can obtain the relation between each part of an object and the corresponding part of the template. However, the descriptor is not always stable when the local features of the object in the current frame change because the feature information of certain parts of an object is completely different from that of the parts of a template. For tracking and matching a moving object, we believe a stable feature descriptor must adopt not only the local information of one points but also the global spatial distribution of a feature points set in the object as well.

In this paper, we propose a new method to describe the feature of an object. This method employs both local feature descriptor and global spatial distribution. This method formulates the object tracking problem as an optimization problems. The energy function minimizes the total changes of local and global constraints. However, this matching problem is a quadratic assignment problem that is NP-hard. Presently, there is no effective method to solve this problem within a short processing time. Therefore, we adopt genetic algorithm[6] to search the approximate solution.

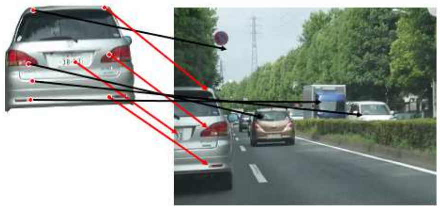


Fig. 1: Black lines represent inaccurate matching caused by vanishment of the object.

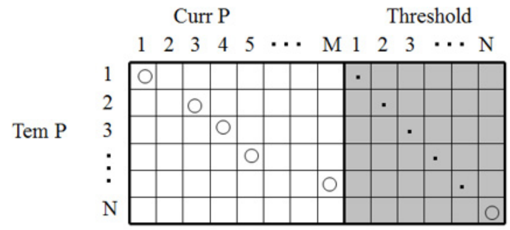


Fig. 2: Expansion of basic cost matrix considering vanishment.

Some related studies have reported the use of the spatial constraint. Graph structure based methods[7], which perform graph matching between 2 sets of feature points, are only applicable to the situations wherein the structure of the graph of objects changes slightly. These methods are not applicable if the graph structure of the target object in the scenes cannot be created. This problem is very common because stable feature points cannot always be extracted from the target object in scenes. On the other hand, our method, which does not require the graph structures of the moving object exhibits higher robustness. Another related study is shape tracking. For example, the method in [8] and [9]use level sets to track contour of moving objects. The method[10] calculates similarity by the combination of four features which are HSV, vertical edge, horizontal edge, and diagonal edge. Different from these methods, our method that focus on the global spatial distribution of a feature points set of an object.

2 Linear Assignment Tracker

2.1 Creating the Basic Cost Matrix

In this section, we describe our tracker based on linear assignment. First, we extract SIFT feature points from the current frame and the template. If we assume that the numbers of feature points on the current frame and the template are M and N, respectively. Then we build an $M \times N$ array that consists of feature similarities (actually difference measure of distance between two orientation histograms) between the current frame's points(Curr P) and the template's points(Temp P). Next, one-to-one matching between Curr P and Temp P should be performed in such a way that the total distance is minimized.

2.2 Expansion of Basic Cost Matrix Considering Vanishment

However, it is impossible to archive the one-to-one matching when the object is occluded or vanish in the scene. Under these situations, the presence of similar points in the current frame results in inaccurate matching, as shown in the Fig. 1. In order to solve this problem, we expand the cost matrix. We add a Vanish array that consists of a predetermined threshold value. The final cost matrix is shown in Fig. 2. In the Vanish array, only the diagonal elements can be selected. If a Temp P is matched with this part of the cost matrix, it is regarded as a vanishing point or an occluded point. The final size of the cost matrix becomes $N \times (M+N)$. The arithmetic expression of the matching problem is written down as follows:

Minimize

$$C = \sum_{i=1}^{M} \sum_{j=1}^{N+M} p_{ij} c_{ij},$$

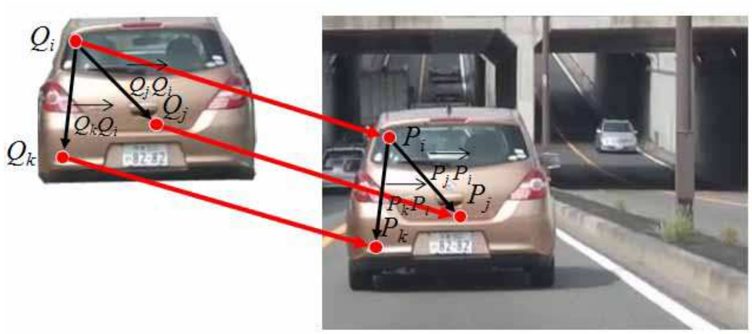


Fig. 3: The global spacial distribution of a feature points set in an object is expressed by the relative positions and calculated by vectors.

subject to

$$\sum_{j=1}^{N+M} p_{ij} = 1 \quad i = \{1,2,\ldots,N\},$$
$$\sum_{i=1}^{N} p_{ij} \leq 1 \quad j = \{1,2,\ldots,N+M\},$$
$$p_{ij} = \{0,1\},$$

$$c_{ij} = \begin{cases} dist\{CurrP_i, TempP_j\} \\ \quad i = \{1,\ldots,M\}, \\ \quad j = \{1,\ldots,N\}, \\ threshold\ for\ vanishing \\ \quad i = \{1,\ldots,M\}, \\ \quad j = \{K \times N+1,\ldots,N+M\}. \end{cases}$$

$dist\{CurrP_i, TempP_j\}$**:** Distance between feature point i of the current frame and feature point j of the template.

$threshold\ for\ vanishing$**:** A predetermined threshold value. If this value is matched, this block is regarded as vanishing.

This is a type of linear assignment problem and can be solved by the Hungarian method[11].

3 Descriptor of Global Spacial Distribution

As described in the introduction, the information of global spacial distribution is very important for matching between feature points sets. The global spacial distribution can be expressed by the relative positions of feature points. Therefore, we adopted vector to express the relative positions. As shown in Fig. 3, template's points Q_i, Q_j and Q_k are matched with the current frame's points P_i, P_j and P_k, respectively. The matching considering the spacial distribution should minimize two values that represent the variations of relative positions. The first value is the angular variation that is calculated as follows:

$$A_{ijk} = \| \arccos \frac{\overrightarrow{P_kP_i} \bullet \overrightarrow{P_jP_i}}{| P_kP_i || P_jP_i |} - \arccos \frac{\overrightarrow{Q_kQ_i} \bullet \overrightarrow{Q_jQ_i}}{| Q_kQ_i || Q_jQ_i |} \| . \tag{1}$$

Although the A_{ijk} is minimized, there is no guarantee that the relative positions of points are same. Therefore, we add one more constraint that represents the variations of endpoints calcualted as follows:

$$R_{ijk} = \| \left(\overrightarrow{P_iP_j} - \overrightarrow{P_iP_k}\right) - \left(\overrightarrow{Q_iQ_j} - \overrightarrow{Q_iQ_k}\right) \| . \tag{2}$$

Then the constraint considering the global spacial distibution is calculated by the following equation:

$$S = V_d + \eta V_a. \tag{3}$$

Here, η is a waighting factor. The V_a and V_d are calculated as follows:

$$V_a = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} A_{ijk}, \tag{4}$$

$$V_d = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} R_{ijk}. \tag{5}$$

Considering both the color feature and global spacial distribution simultaneously, the optimization problem can be expressed by the following equation:

$$E = w * C + S, \tag{6}$$

where C is the total cost of color features in Eq. 1, S is the total structure variation in Eq. 3, and w is a weighting factor.

4 Solution of Optimization Problem by Adopting Genetic Algorithm

The optimization of E in Eq. 6 becomes a quadratic assignment problem that is NP-hard. As there is no effective method for solving this problem within a short processing time, we adopt genetic algorithm (GA)[6] to search for the approximate solution in this paper.

4.1 Genetic Algorithm for Template Matching

In this subsection, we explain the outline of the GA. First, we create a square cost matrix with dummy values. The real size of the cost matrix is $(N+M) \times (N+M)$. For convenience in explanation, we assume that $(N+M) = 8$ and provide an example of an 8×8 cost matrix, as shown in Fig. 4(a). Two rows are filled with dummy values. Next, a candidate solution is encoded to create individuals, as shown in Fig. 4(b). The next step involves generation of new individuals. As shown in Fig. 5, we employ order crossover (OX)[12] for crossover processing and exchange mutation (EM) for mutation processing. The fitness of an individual

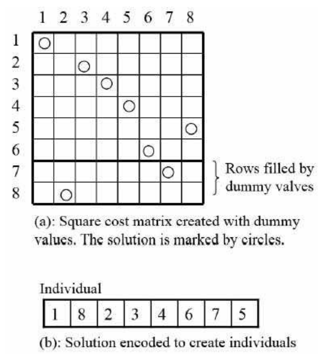


Fig. 4: Example of encoding candidate solution. The part of the cost matrix with dummy values is not employed in the calculation of fitness.

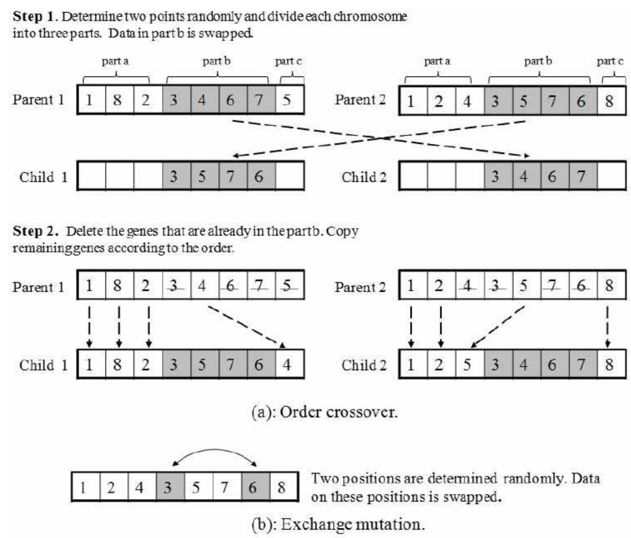


Fig. 5: Order crossover and exchange mutation.

Table 1: Parameters of Experiments

Image size	320×240	w in Eq. 6	30
Population size of GA	500	Crossover rate of GA	0.8
Mutation rate of GA	0.5	Number of generations of GA	200
PC for experiments	Core 2 Duo 3.00GHz PC with 2GB RAM, WindowsXP		

is calculated according to Eq. 6. Here, the part of the cost matrix with dummy values is not employed in the calculation of fittness. Tournament selection is then employed to select individuals with the best fitness to form a new population. The generation process is repeated until a fixed number of generations has been reached.

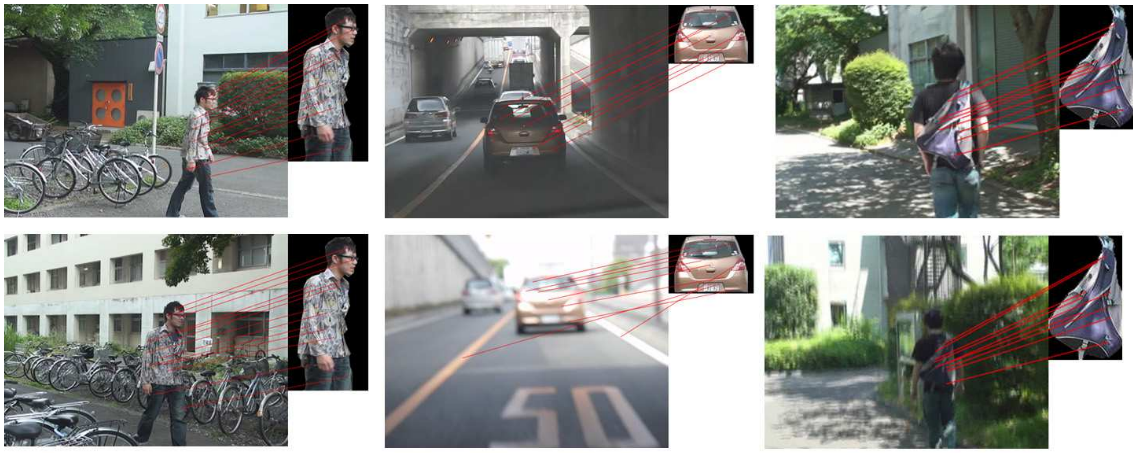
5 Experimental Results

5.1 Parameters and Environment of Experiments

In our experiments, we employ two groups of data. One is our data set containing outdoor scenes and the other is the benchmark data set from [13]. We normalize the Bhattacharyya distance in the range from 0 to 1000. All parameters are listed in Table 1. The code of SIFT is obtained from Rob Hess's website[14]. The processing time is approximately $0.5 \sim 2s$ per frame. We aim to improve this processing time using parallel processing and a hardware accelerator in our future works.

5.2 Experimental Results of Outdoor Data

We present the result of our data sets of real outdoor scenes including the tracking of a knapsack, a car and a person. There are many challenging issues in these image sequences. In Fig. 6(a), the target object in the frame objects vary greatly. In Fig. 6(b) and (c), illumination conditions change abruptly. There are many



(a): Person (b): Car (c): bag

Fig. 6: Results of ourdoor data sets.

similar bjects in scenes of Fig. 6(b). Moreover, fig. 6(b) and (c) are the tracking results of image sequences captured by a hand-held camera. The images appeared blurred because of the movement of the camera. As can be observed from these results, the proposed method realizes robust tracking and the corresponding relations between objects and templates are obtained accurately. In Fig. 6(b), there are some mis-matching. The resason is feature points can not be extracted on the target object in the frame.

5.3 Experimental Results of Bench Data

Another group of experimental results is obtained using benchmark data downloaded from Daivd Ross's website[13]. The proposed method is tested by using his data and comparing proposed method with his method which is presented in [15]. The proposed method is tested by using his data and comparing proposed method with his method which is presented in [15]. The results are shown in fig. 7. The proposed method performs well even for the frames where David's method fails.

5.4 Comparison with Recent Tracking Methods

Finally, the proposed method is compared with some representative methods reported in recent years. Figure 8 shows some frames of the results. Figure 8(a), (b) ,(c) and (d) are results compared with TSS method[16], MIL method[17], P-N method[18], and MK method[19]. These result are obtained from Baiyang Liu's homepage[20]. Compared with results obtained by these methods, the proposed method provide tracking and extraction results with higher precision.

6 Conclusion

In this paper, we proposed a tracking method that considers both local feature and global spacial distribution. This method is robust to many problems. In our future research, we will focus on improving the processing time through parallel processing and using a hardware accelerator. Moreover, we aim to introduce the spatio-temporal constraint to further enhance the robustness of the proposed method.

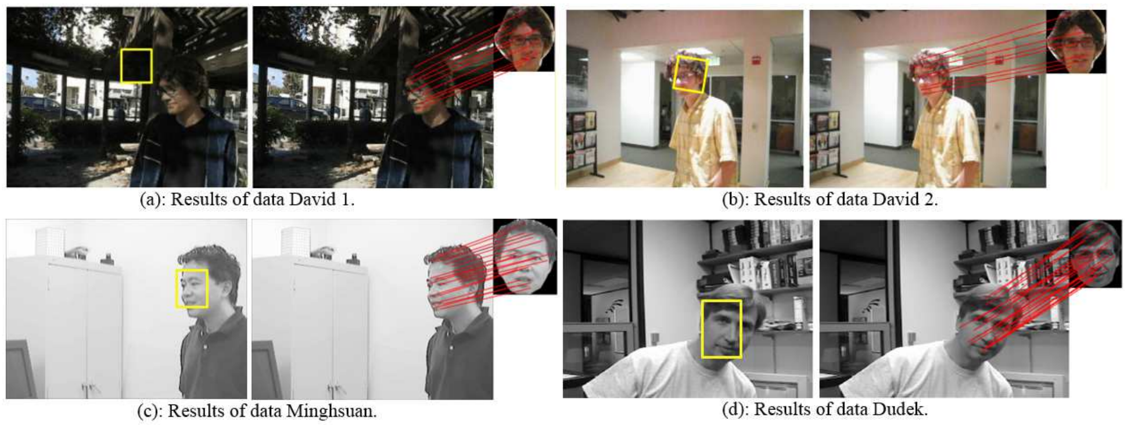


Fig. 7: Comparison with David's method.The proposed method provides tracking and extraction results with higher precision and performs well even for the frames where David's method fails.

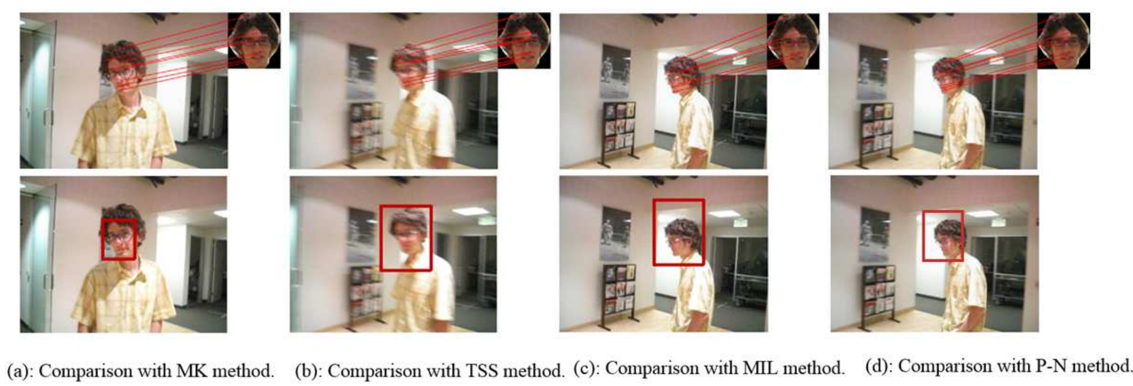


Fig. 8: Comparison with TSS method[16], MIL method[17], P-N method[18], and MK method[19].

Acknowledgments

This paper is supported by the National Nature Science Foundation of China (No. 61302134).

References

K. Nummiaro, E. Koller-meier, and L.V. Gool, "A color-based particle filter," Image and Vision Computing, pp.53–60, 2002.

N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE International Conference on Computer Vision and Pattern Recognition, pp.886–893, 2005.

D. Lowe, "Object recognition from local scale-invariant features," Proceedings of the International Conference on Computer Vision 2, pp.1150–1157, 1999.

D. Comaniciu, P. Meer, and S. Member, "Mean shift: A robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, pp.603–619, 2002.

P. Pe'rez, C. Hue, J. Vermaak, and M. Gangnet, “Color-based probabilistic tracking,” European Conference on Computer Vision, pp.661–675, 2002.

D.E. Goldberg, “Genetic algorithms in search, optimization, and machine learning,” Addison-Wesley, pp.661–675, 1989.

T. Caetano, J. McAuley, L. Cheng, Q. Le, and A. Smola, “Learning graph matching,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.31, pp.1048–1058, 2009.

A. Yilmaz, X. Li, and M. Shah, “Contour based object tracking with occlusion handling in video acquired using mobile cameras,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.26, no.11, pp.1531–1536, 2004.

C. Bibby and I. Reid, “Robust real-time visual tracking using pixel-wise posteriors,” Proceedings of European Conference on Computer Vision, 2008.

K. She, G. Bebis, H. Gu, and R. Miller, “Vehicle tracking using on-line fusion of color and shape features,” IEEE Int. Conf. on Intelligent Transportation Systems, pp.731–736, 2004.

“Munkres assignment algorithm,”
http://csclab.murraystate.edu/bob.pilgrim/445/munkres.html.

L. Davis, “Applying adaptive algorithms to epistatic domains,” Proceedings of the International Joint Conference on Artificial Intelligence, 1985.

“Data sets from david ross,”
http://www.cs.toronto.edu/dross/ivt/.

“Code of sift keypoint detector,”
http://web.engr.oregonstate.edu/ hess/index.html.

D.A. Ross, J. Lim, R.S. Lin, and M.H. Yang, “Incremental learning for robust visual tracking,” International Journal of Computer Vision, vol.77, pp.125–141, 2008.

B. Babenko, M. hsuan Yang, and S. Belongie, “Visual tracking with online multiple instance learning,” IEEE International Conference on Computer Vision and Pattern Recognition, 2009.

B. Liu, L. Yang, J. Huang, P. Meer, L. Gong, and C. Kulikowski, “Robust and fast collaborative tracking with two stage sparse optimization,” The 11th European Conference on Computer Vision, 2010.

Z. Kalal, J. Matas, and K. Mikolajczyk, “P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints,” Conference on Computer Vision and Pattern Recognition, 2010.

B. Liu, J. Huang, L. Yang, and C. Kulikowski, “Robust tracking using local sparse appearance model and k-selection,” IEEE International Conference on Computer Vision and Pattern Recognition, 2011.

“Baiyang liu's homepage,”
http://paul.rutgers.edu/ baiyang/spt/.