

Adaptive Search Window for Object Tracking in the Crowds using Undecimated Wavelet Packet Features

M. Khansari, Digital Media Lab, Department of Computer Engineering, Sharif University of Technology, Iran, khansari@mehr.sharif.edu

H. R. Rabiee, Department of Computer Engineering of Sharif University, and Iran Telecommunication Research Center, Iran, rabiee@sharif.edu

M. Asadi, Digital Media Lab, Department of Computer Engineering, Sharif University of Technology, Iran, m_asadi@ce.sharif.edu

P. Khadem Hamedani, Digital Media Lab, Department of Computer Engineering, Sharif University of Technology, Iran, khadem@dml.aictc.ir

M. Ghanbari, Department of Electronic Systems Engineering, University of Essex, England, ghan@essex.ac.uk

ABSTRACT

In this paper, we propose an adaptive object tracking algorithm in crowded scenes. The amplitudes of Undecimated Wavelet Packet Tree coefficients for some selected pixels at the object border are used to create a Feature Vector (FV) corresponding to that pixel. The algorithm uses these FVs to track the pixels of small square blocks located at the vicinity of the object boundary. The search window is adapted through the use of texture information of the scene by finding the direction and speed of the object motion. Experimental results show a good object tracking performance in crowds that include object translation, rotation, scaling and partial occlusion.

KEYWORDS: Object tracking, Undecimated Wavelet Packet Transform, Motion Direction, Texture Analysis, Crowded Scenes

1. INTRODUCTION

Object tracking is one of the challenging problems in image and video processing applications. The extracted objects in video sequences can be used in many applications such as video surveillance, visual monitoring, content-based indexing and retrieval, traffic monitoring, and video post-production. Various techniques for video object tracking have been proposed in the literature [1-9]. Object tracking in video sequences vary according to user interaction, tracking features, motion-model assumption, temporal object tracking, and update procedures. The temporal object tracking methods can be classified into four groups: region-based, contour/mesh-based, model based, and feature based methods [10].

In the region-based methods information such as motion, color, and texture of the regions are used to track the regions. By using a combination of these regions, one can construct the intended object's shape [11]. Contour-based methods try to track an object by following the pixels on the object boundary by building the contour of the object. These methods make use of the motion

information to project the contour and then adapt it to the detected object in the next frame [6]. In model-based methods the parameterized object model provides the priori information [13].

Most of the works in tracking humans in crowded scenes use model-based approaches within simple regions such as rectangles and ellipses along with estimation techniques to track humans in a crowded scene [14, 15], or simply use some heuristics such as vertical projection of the foreground to detect regions of interest in crowded scenes [16].

In this paper, we have developed adaptations to our previous works [17, 18, 19] for tracking of user-defined rectangles encompassing object of interest in the crowded video scenes. The algorithm uses an inter-frame texture analysis scheme [20], to update the search window location for the successive frames. The key advantage of UWPT is that it is redundant, shift invariant, and it gives a denser approximation to continuous wavelet transform than the approximation provided by the orthonormal discrete wavelet transform [21, 22].

In section 2, the proposed algorithm and search window updating mechanisms is presented. Section 3 illustrates experimental results in the various conditions. Finally, section 4 provides concluding remarks and the future works.

2. THE PROPOSED ALGORITHM

In the proposed algorithm, object tracking is performed by temporal tracking of a rectangle around the object at a reference frame. Figure 1 shows a block diagram of the system where generation of the feature vector (FV) and update the search window comprise the main elements of the algorithm. These are briefly presented in the following subsection.

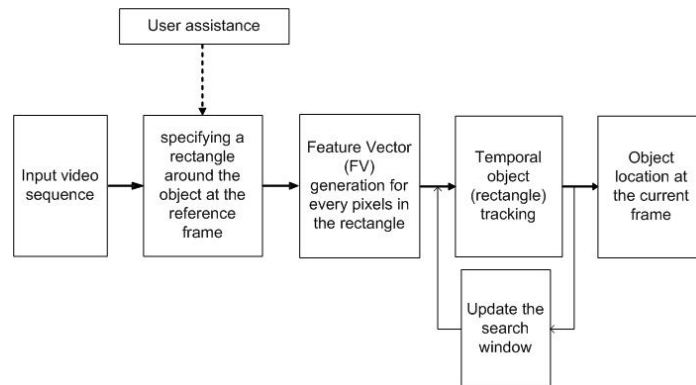


Fig. 1. A general block diagram of the proposed algorithm

2.1 Feature Vector Generation

The Undecimated Wavelet Packet Transform (UWPT) has two properties making it suitable for generating invariant and robust features corresponding to each pixel [23, 24].

1. It has the shift invariant property. Consequently, feature vectors based on the wavelet coefficients in frame t , can be found again in frame $t+1$.
2. All the subbands in the decomposition tree have the same size equal to the size of the input frame (there is no down sampling). This property simplifies the feature extraction process (Fig. 2).

The procedure for generating a FV for each pixel in region r (containing the target object) at frame t can be summarized in the following steps:

1. Generate UWPT for region r (note that the UWPT is constructed with padding zero when needed).
2. Since the approximation provides an average of the signal based on the number of levels at the UWPT tree, the tree is pruned to have most coefficients from the approximation subbands. This type of basis selection gives more weight to the approximations while

considering the details. For our application, this type of basis selection is more reasonable; because the comparison in the object temporal tracking part of the algorithm is carried out between two regions represented by similar approximation and detail sub-bands. The output of this step is an array of node numbers of the UWPT tree specifying the selected basis for the successive frame manipulations.

3. The FV for each pixel in region r can be simply created by selecting the corresponding wavelet coefficients in the selected basis nodes of step 2. Therefore, the number of elements in FV is the same as the number of selected basis nodes.

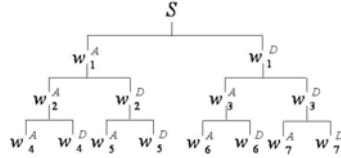


Fig. 2. Undecimated Wavelet Packet Transform tree for one dimensional signal S

2.2 Temporal Tracking

The procedure to search for the best-matched region is similar to the general block-matching algorithm, except it exploits the generated FV of the aforementioned procedure rather than the luminance of the pixels. To find the best match of rectangle r in frame t to r in frame $t+1$ the minimum sum of the Euclidean distances between the search rectangle and FVs of the pixels within region r is calculated (e.g. full search algorithm in the search window).

2.3. Search Window Updating Mechanism

In previous work [19] we have updated the search window (SW) center based on the center of the rectangle around the object at the current frame. This approach is simple but propagates any mismatch into the following frames and causes loss of tracking, in particular for occluded object. The object motion speed and direction is used to update the location of the SW. This idea can be implemented efficiently by using the inter-frame texture analysis technique [20]. The temporal difference histogram of two blocks belonging to current and successive frames, defined as the absolute difference of gray level values between pairs of pixels in the blocks. Several features such as coarseness and directionality can be derived from the temporal difference histogram [20]. The direction and speed of the motion is estimated based on the temporal difference histogram. The main advantages of this algorithm are as follows.

1. It can track both rigid and non-rigid objects without any pre-assumption, training, or object shape model.
2. It can efficiently track the objects in the crowded video sequences
3. It is robust to the different object transformation such as translation and rotation.
4. In case of perspective transformation due to the object scaling in the scene, the algorithm can handle the event.
5. Partial or full occlusions of the object can successfully be overcome in the successive frames.

3. EXPERIMENTAL RESULTS

Since there is no universally agreed method of evaluating the performance of object tracking for crowded scenes, we have analyzed our results, subjectively. The outcomes of tracking in various conditions are shown with their associated SW in this section. Throughout the experiments it was no scene cuts were assumed.

The bi-orthogonal wavelet bases was used to generate the UWPT tree. We have used 3 levels of UWPT tree decomposition with wavelet family of *Bior2.2*. In $BiorN_r.N_d$, N_r and N_d represent

the number of vanishing moments for the synthesis and analysis wavelets [23].

To evaluate the algorithm in a realistic environment, we have applied it to different real-time video clips in cooperation with Tehran Metro authorities. These video clips show moving crowds at different parts of the metro such as getting on/off from the train and up/down the stairs. We have also compared updating the SW center based on the center of the rectangle around the object with updating the SW based on the direction and speed of the object motion. In all the figures, smaller rectangles correspond to the rectangles around the objects and larger rectangles correspond to the SW.

Fig. 3 shows the result of tracking a man, coming down the stairs in a crowded metro station. Empirical parameters to find the direction and speed of the motion for updating the SW was set to $d = 1$ and $k = 3$. The object is moving down with a constant speed with small amount of zooming out, some degree of rotation of the head and some cross movements. As the tracking results show, the object of interest has been successfully tracked by the algorithm that use updated SW based on the direction and speed of the object motion in the presence of complete occlusion and zoom out.

Note that in all figures, prefixes "a" denotes results of the algorithm used to update the SW based on the direction and speed of the object motion and prefix "b" denotes the algorithm used to update the SW center based on the center of the rectangle around the object respectively.

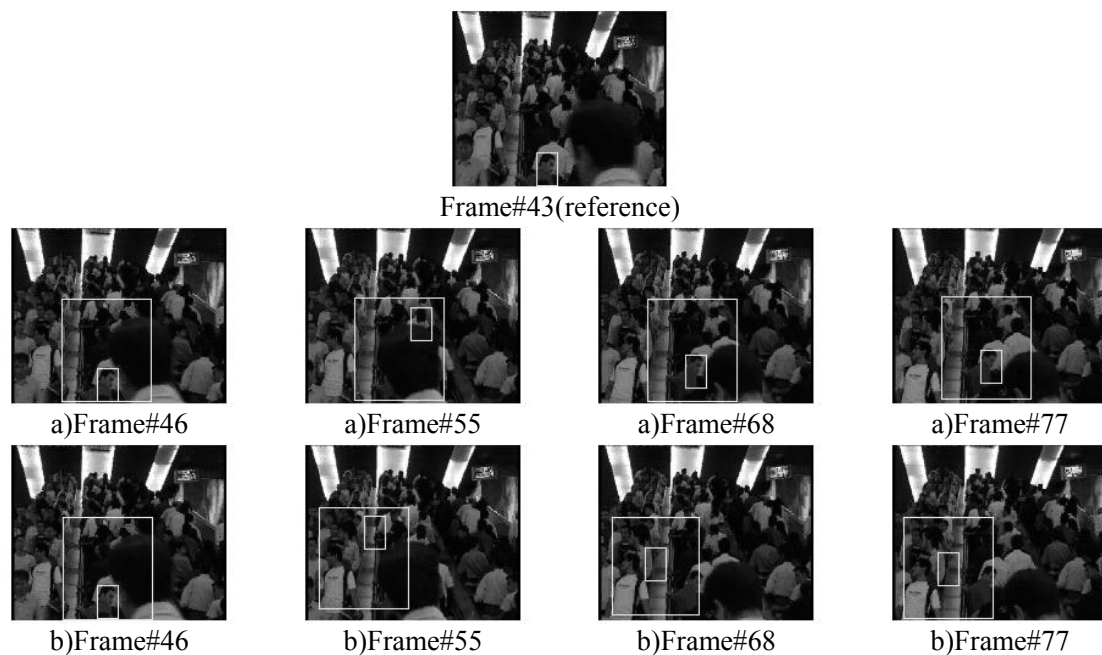


Fig. 3: Tracking a man going up from the stairs a) $k=3$, $d=1$

Fig. 4 shows the result of tracking where the people are getting off the train. The object passes through the crowd where he experiences full occlusions and some zooming effects in a number of frames. As is demonstrated in the Fig. 4, the algorithm with an update on the SW based on the direction and speed of motion can successfully handle complete occlusions even in the presence of zooming effects. This is due to the robustness of FVs and adaptability of the SW.

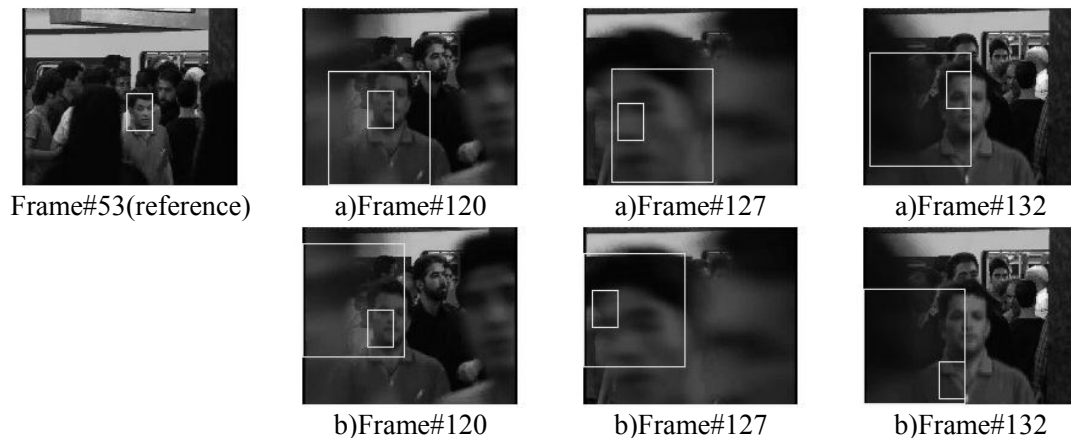


Fig. 4. Tracking a man getting off the train a) $k=4$, $d=1$

4. CONCLUSIONS AND FUTURE WORKS

A new adaptive object tracking algorithm for crowded scenes has been proposed. The algorithm uses pixel features in the wavelet domain with a novel search window updating mechanism based on texture analysis to track the objects in crowded scenes. Based on the properties of the UWPT, existence of individual robust FVs for each pixel, and the adaptive search window, this method can handle complex object transformation including translation, small rotation, scaling and partial or complete occlusions in a reasonable number of successive frames. The experimental results confirmed the efficiency of our algorithm in tracking the object in crowded scenes.

One of the open problems is the integration of algorithm with some spatial domain features such as color and edge for a better update of feature vector and search window. This improvement should handle special cases such as when the tracked object falls outside the search window, in the presence of occlusion, noise, abrupt transformation and zooming.

5. ACKNOWLEDGMENTS

This research has been funded by Iran Telecommunication Research Center and the Advanced Information and Communication Technology Research Center (AICTC) of Sharif University of Technology.

6. REFERENCES

- [1] R. N. Strickland and H. I. Hahn, "Wavelet transform methods for object detection and recovery," *IEEE Transaction on Image Processing*, Vol. 6, No. 5, pp. 724-735, May 1997.
- [2] C. Gu, M. Lee, "Semiautomatic Segmentation and Tracking of Semantic Video Objects," *IEEE Transaction for circuit and systems for video Technology*, Vol. 8, No. 5, pp. 572-584, 1998.
- [3] P. Salembier, F. Marques, "Region-Based Representations of Images and Video: Segmentation Tools for Multimedia Services," *IEEE Transaction for circuit and systems for video Technology*, Vol. 9, No. 8, Dec. 1999.
- [4] J. Guo, C.-C.J Kuo, *Semantic Video Object Segmentation for Content-based Multimedia Applications*, Kluwer Academic Publisher, 2001.
- [5] Cavallaro, "From visual information to knowledge: semantic video object segmentation, tracking and description," PhD thesis, EPFL Switzerland, 2002.
- [6] Ç. E. Erdem, A. M. Tekalp, B. Sankur, "Video Object Tracking With Feedback of Performance Measures," *IEEE Transaction for circuit and systems for video Technology*, Vol. 13, No. 4, pp. 310-324, 2003.

- [7] D. Comaniciu, V. Ramesh, P. Meer: Kernel-Based Object Tracking, *IEEE Trans. Pattern Analysis Machine Intell.*, Vol. 25, No. 5, 564-575, 2003.
- [8] X.S. Zhou, D. Comaniciu, A. Gupta: An Information Fusion Framework for Robust Shape Tracking, *IEEE Trans. PAMI*, Vol. 27, No. 1, 115-129, 2005.
- [9] M. Amiri, H. R. Rabiee, F. Behazin, M. Khansari, "A new wavelet domain block matching algorithm for real-time object tracking," IEEE ICIP, September 14-17, Barcelona, Spain, 2003.
- [10] D. Wang, "Unsupervised video segmentation based watersheds and temporal tracking," *IEEE Trans. CSVT*, Vol. 8., No. 5, pp. 539-546, 1998.
- [11] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Computer Vision*, vol. 1, no. 4, pp. 321--331, 1987.
- [12] B. Günsel, A. M. Tekalp, and P. J.L. van Beek, "Content-based access to video objects: Temporal segmentation, feature extraction and visual summarization," *Signal Processing (special issue)*, vol. 66, no. 2, pp. 261-280, April 1998.
- [13] Y. Chen, Y. Rui, and T. S. Huang, "JPDAF based HMM for real-time contour tracking," in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, 2001, pp. 543-550.
- [14] T. Zhao, R. Nevatia, "Tracking Multiple Humans in Crowded Environment," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04).
- [15] C. Belezni1, B. Fruhstick, H. Bischof, and W. Kropatsch4, "Model-Based Occlusion Handling for Tracking in Crowded Scenes," Preprint, 2005.
- [16] S. Haritaoglu, D. Harwood and L. S. Davis, "Real-Time Surveillance of People and Their Activities," *IEEE Trans. PAMI*, vol.22, no.8, 2000.
- [17] M. Khansari, H. R. Rabiee, M. Asadi, M. Ghanbari, M. Nosrati, M. Amiri, "A Quantization Noise Robust Object's Shape Prediction Algorithm," EUSIPCO, European Signal Processing Conference, Antalya, Turkey, September 4-8, 2005.
- [18] M. Khansari, H. R. Rabiee, M. Asadi, M. Nosrati, M. Amiri, , M. Ghanbari "Object Shape Prediction in Noisy Video Based on Undecimated Wavelet Packet Features," 12th International Multimedia Modelling (MMM) Conference, Beijing, China, Jan. 2-4, 2006.
- [19] M. Khansari, H. R. Rabiee, M. Asadi, M. Ghanbari, M. Nosrati, M. Amiri, "A Semi-Automatic Video Object Extraction Algorithm Based on Joint Transform And Spatial Domain Features," ` , International Workshop on Content-Based Multimedia Indexing, Riga, Latvia, June 21-, 2005.
- [20] V. E. Seferidis and M. Ghanbari, "Adaptive Motion Estimation Based on Texture Analysis," IEEE Trans. on Communications, vol. 42, no. 2/3/4, 1994.
- [21] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," IEEE Transaction on Information Theory, Special Issue on Wavelet Transforms and Multiresolution Signal Analysis, Vol. 38, pp. 713-718, Mar. 1992.
- [22] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate distortion sense," *IEEE Transaction on Image Processing*, Vol. 2, pp. 160-175, Apr. 1993.
- [23] Daubechies, *Ten lectures on wavelets*, CBMS, SIAM, 61, 1994, 271-280.
- [24] Theory and Applications of the Shift-Invariant, Time-varying Undecimated Wavelet Transforms," MS Thesis, Rice University, 1995.