CROWDED SCENCE OBJECT TRACKING IN PRESENCE OF GAUSSIAN WHITE NOISE USING UNDECIMATED WAVELET FEATURES

ABSTRACT

In this paper, we propose a new noise robust algorithm for object tracking in the crowded video scenes. The algorithm exploits the properties of undecimated wavelet packet transform (UWPT) coefficients and texture analysis to track arbitrary objects. The coefficients of the UWPT of a user-specified region at the reference frame construct a Feature Vector (FV) for every pixel in that region. Optimal search for the best match of the region in successive frames is then performed by using the generated FVs inside an adaptive search window. Adaptation of the search window is achieved by inter-frame texture analysis to find the direction and speed of the object motion. Noise robustness has been achieved through inherent noise suppression in the generation process.

Experimental results show a good performance for object tracking in contaminated crowded scenes with Gaussian White Noise even in presence of partial occlusion.

1. INTRODUCTION

With the emergence of interactive multimedia systems, tracked objects in an object tracking process in video sequences, can be used for many applications such as video surveillance, visual navigation and monitoring, content-based indexing and retrieval, object-based coding, traffic monitoring, sports analysis for enhanced TV broadcasting and video post-production.

The Introduction of noise to an object tracking system, specially tracking in crowds, makes the tracking a sophisticated problem. The noise effect on a received image and video sequence passing through a channel has direct impact on the visual representation of the received signal. In many applications, this effect is more important. For instance, object tracking may have some characteristics such as shape rotation and scaling, changing the color, non-uniform object movement, and changing in the background that make the tracking in the presence of noise more complicated

The color histogram is typically used to model the targets to combat partial occlusion, non-rigidity, and to some extent noise [1, 2]. As color histogram only describes the global color distribution and ignores spatial or layout of the colors, tracked objects are easily confused by a background with similar colors. Moreover, it can not deal easily with illumination changes and lacks exploiting any spatial information. Therefore, feature description based on color histogram for target tracking particularly in the crowded scenes, where similar small objects exist will most likely fail.

More recently, mean-shift tracking algorithms that use color histogram has been successfully applied in object tracking and proved to be robust to partial occlusions [1, 3-5].; to the best of our knowledge these algorithms has not been reported neither for object tracking in crowded scenes nor for noise robustness.

Color histogram has also been integrated in probabilistic frameworks such as Bayesian and particle filters [2, 7, 9]. Comparative evaluation of different tracking algorithms shows that this family of techniques is more robust to noise than the other well known techniques [10].

In addition, most of the previous work has been evaluated on simple scenarios: either a walking person who has been occluded by another person in the reverse direction [2, 9, 10] or a football player with a known color outfit who has been occluded by the opposite player [8] and not for more complex scenes such as dense crowds of very close and similar objects.

The general drawback of these techniques is that similar objects (e.g. heads in a crowd) hardly change the color histogram and hence impair their reliability. Other challenging issues of the aforementioned methods are robustness against noise and stability of the selected features in presence of various object transformations [10].

The main contribution of this paper is evaluation and analysis of the adaptation of a feature vector generation and block matching algorithm in the Undecimated Wavelet Packet Transform (UWPT) domain [11] for tracking objects in crowded scenes [13] in presence of Gaussian White Noise (WGN). In contrast to the conventional methods for object tracking which using spatial domain features with the limited functionality in noisy sequences, it introduces a new transform domain tracking algorithm that can manage crowded scene object tracking in a noisy environment.

After presenting the algorithm in section 2, performance of the proposed algorithm under various conditions is evaluated in section 3. Finally, section 4 provides the concluding remarks and future works.

2. THE PROPOSED ALGOTRITHM

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), temporal tracking and update the search window comprise the main elements of the algorithm. These are briefly presented in the following subsection.

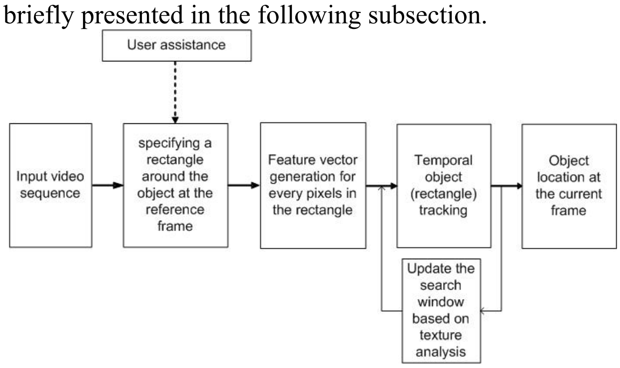


Figure 1. A general block diagram of the proposed algorithm

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 [15, 17].

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. As the approximation subband provides an average of the signal based on the number of levels at the UWPT tree, we prune the tree to have the most coefficients from the approximation subbands. This type of basis selection gives more weight to the approximations which are useful for the intended application. For example, in Fig.2, we may let $x = (W_4^A, W_4^D)$. For our application, this type of basis selection is more reasonable; because the comparison in the temporal tracking part of the algorithm is carried out between two regions that are represented by similar approximation and detail sub-bands.
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.

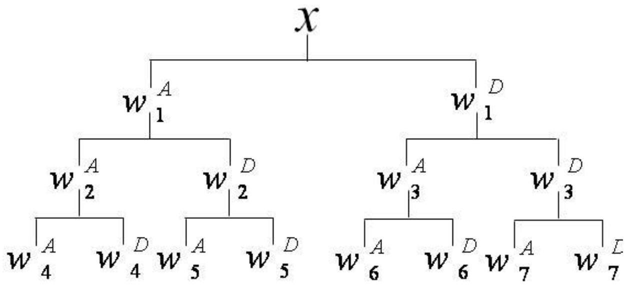


Figure 2. Undecimated Wavelet Packet Transform tree for one dimensional signal *x*

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 [12] 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 error into the following frames and causes loss of tracking, in particular for noisy sequences.

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 [14]. 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 [14]. The direction and speed of the motion is estimated based on the temporal difference histogram.

3. EXPERIMENTAL RESULTS

Results of the proposed tracking method have been compared with the well-known color histogram based tracking algorithms with two different matching distance measures; i.e. Chi-square and Bhattacharyya. Therefore, figures are presented in three parts: a) Color Histogram based tracking with Chi-square matching distance measure (briefly CHC), b) Color Histogram based tracking with Bhattacharya matching distance measure (briefly CHB), and c) our proposed algorithm (briefly WBMA[1]).

Moreover, we have defined a method for objective evaluation of the tracking techniques; the Euclidian distance of the center of the gravity of the tracked and actual object. Here at the start of tracking, a bounding rectangle centered at the center of the gravity of the desired object is drawn. In the following frames, the bounding rectangle represents the tracked object and its distance with the center of the gravity of the actual object is measured (Fig. 4).

We have used biorthogonal wavelet bases, which are particularly useful for object detection to generate the UWPT tree [18]. In fact, the presence of spikes in the biorthogonal wavelet bases makes them suitable for target tracking applications. In all experiments of this section, 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 of the synthesis and analysis filters [16]. In the color histogram based algorithm implementation, the number of color bins has been set to 32.

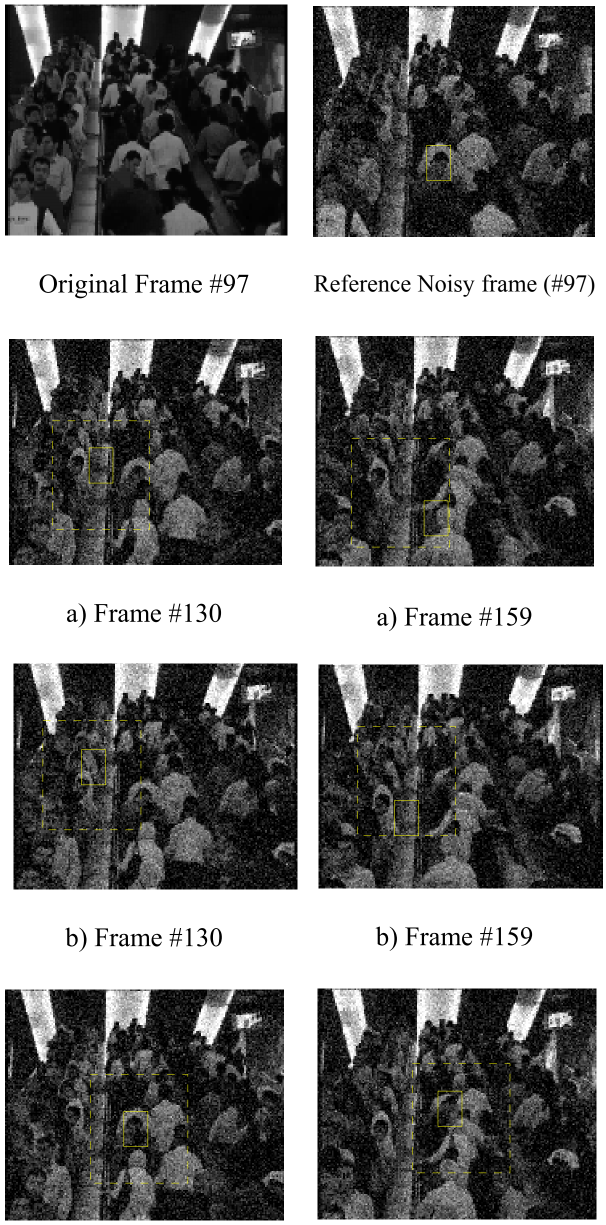
To evaluate the algorithms in a real environment setting, we have applied them to different real video clips of Tehran Metro Stations in cooperation with the Tehran Metro authorities. In all the shown snap shots, solid rectangles correspond to the rectangles around the objects and the rectangles with dashed lines represent the search window. Note the difficulty in tracking of heads in a crowded scene, as there are several near distance similar objects.

For each tracking result, the corresponding complete video clips has been generated that is available through Internet[2] for more detailed subjective comparison

Fig. 3 shows the result of tracking of a person moving up the stairs and away from the camera in the metro station in presence of additive white Gaussian noise (PSNR 20 dB). This type of noise is very common with low light video, especially in undergrounds. Again, the reference frame which is contaminated with noise is frame #97. To highlight the resilience of the proposed algorithm against noise, we have used a big enough bounding box (23x16) and therefore a large search area (62x69). Empirical parameters to find the direction and speed of the motion for updating the search window was set to d=1 and k=3 for Fig. 3.

The target object is stepping up the stairs with a constant speed and its movement exhibited a small amount of zooming out, some degree of rotation of the head, and partial occlusion (e.g. frame #159).

As shown in Fig. 4, CHB losses the tracks from the first frame after the reference frame. It backs to the target randomly at frame #112 (for only one frame) and losses the track completely in all successive frames.



Original Frame #97 | Reference Noisy frame (#97)

a) Frame #130 | a) Frame #159

b) Frame #130 | b) Frame #159

c) Frame #130 | c) Frame #159

Figure 3. 1) Tracking a man stepping up the stairs, in presence of partial occlusions, zooming out and additive Gaussian White Noise (PSNR = 20 dB); a) CHC, b) CHB, and c) WBMA.

.

CHC is showing an unreliable tracking. Until frame #114, it flips flop between an inaccurate box near the target and some other far from the original target. From frame #115 the resulted tracked object by CHC is getting far from the expected target.

As shown in Fig. 4, both of the color histogram based algorithms can not handle partial occlusion and zooming out in the sequence. The presence of noise has degraded the performance of theses algorithms tremendously; without having any ill effect on our method.

[1] Wavelet-based Block Matching Algorithm
[2] http://mehr.sharif.edu/~khansari

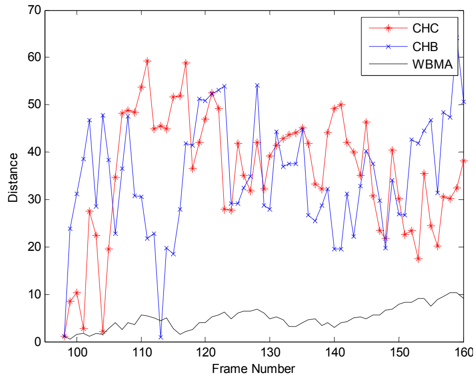


Figure 4. Distance between the center of tracked bounding box and the expected center at each frame of the Fig. 3 sequence.

4. CONCLUSIONS AND FUTURE WORKS

A new adaptive object tracking algorithm for noisy 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 in the presence of GWN in a reasonable number of successive frames. The experimental results confirmed the efficiency of our algorithm in tracking the object in noisy crowded scenes.

A memory-based FV updating mechanism combined with Kalman filters for search area prediction can improve the performance of our algorithm in presence of abrupt zooming in/out or object scaling.

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.

REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-Based Object Tracking,” IEEETrans. Pattern Anal. Mach. Intell., vol. 25, no. 5, pp. 564–577, May 2003.

[2] P. Perez, C. Hue, J. Vermaak, and M. Gangnet,“ Color-based probabilistic tracking,” In Proc. European Conf. on Computer Vision, Copenhagen, Denmark, vol. I, pp. 661–675, 2002.

[3] D. Comaniciu and P. Meer, “Mean shift: A Robust Approach Toward Feature Space Analysis”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 5, pp. 603–619, May 2002.

[4] Z. Zivkovic, B. Krose, “An EM-like algorithm for color-histogram based object tracking,” Proceedings of the International Conference on Computer Vision and Pattern Recognition, 2004.

[5] Dong Xu, Yimin Wang, and Jinwen An, “Applying a New Spatial Color Histogram in Mean-Shift Based Tracking Algorithm,” Image and Vision Computing New Zealand, 2005.

[6] D. Comaniciu, V. Ramesh, and P. Meer, “Real-Time Tracking of Non-Rigid Objects Using Mean Shift,” Proc. IEEE Conf. on CVPR, pp. 142–149, 2000.

[7] A.D. Jepson, D.J. Fleet, T.F. El-Maraghi, “Robust Online Appearance Models for Visual Tracking,” Proceedings of the International Conference on Computer Vision and Pattern Recognition, vol. I, pp. 415-422, 2001.

[8] A. Jacquot, P. Sturm O. Ruch,” Adaptive Tracking of Non-Rigid Objects Based on Color Histograms and Automatic Parameter Selection ,” IEEE Workshop on Motion and Video Computing (WACV/MOTION'05), vol. 2, pp. 103-109, 2005.

[9] B. Han, C. Yang, R. Duraiswami and L. Davis, “Bayesian Filtering and Integral Image for Visual Tracking. Invited to special session of Real-Time Object Tracking,” Algorithms and Evaluation in Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Montreux, Switzerland, 2005.

[10] B. Deutsch, Ch. Gr¨aßl, F. Bajramovic, J. Denzler, “A Comparative Evaluation of Template and Histogram Based 2D Tracking Algorithms,” Pattern Recognition, 27th DAGM Symposium, Springer, pp. 269–276, 2005.

[11] 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, pp. 961-4, 2003.

[12] M. Khansari, H. R. Rabiee, M. Asadi, M. Ghanbari, M. Nosrati, M. Amiri, “A Shape Tracking Algorithm Based on Generated Pixel Features by Undecimated Wavelet Packet,” CSICC, 10th Annual Computer Society of Iran Computer Conference, Tehran, Iran, 2005.

[13] M. Khansari, H. R. Rabiee, M. Asadi, P. Khadem Hamedani, M. Ghanbari, “Adaptive Search Window for Object Tracking in the Crowds using Undecimated Wavelet Packet Features,” WAC, World Automation Congress, Budapest, Hungary, July 24-26, 2006.

[14] V. E. Seferidis and M. Ghanbari, “Adaptive Motion Estimation Based on Texture Analysis,” IEEE Transactions on Communications, vol. 42, no. 2-4, pp. 1277-1287, 1994.

[15] M. Vetterli, J. Kovacevic, Wavelets and Subband Coding, Prentice Hall PTR,1st, edition, 1995.

[16] Daubechies, Ten lectures on wavelets, CBMS, SIAM, 61, pp. 271-280, 1994.

[17] H. Guo, “Theory and applications of the shift invariant, time-varying, and undecimated wavelet transform,” Master's Thesis, Rice University, Houston, TX, 1995.

[18] 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.