A Robust Object Shape Prediction Algorithm in the Presence of White Gaussian Noise

M. Khansari

Digital Media Lab, Computer Eng. Department, Sharif University of Technology

Khansari@mehr.sharif.edu

H. R. Rabiee

Digital Media Lab, Computer Eng. Department, Sharif University of Technology

rabiee@sharif.edu

M. Asadi

Digital Media Lab, Computer Eng. Department, Sharif University of Technology

m_asadi@ce.sharif.edu

M. Ghanbari

Digital Media Lab, Computer Eng. Department, Sharif University of Technology

Department of Electronics System Engineering, University of Essex

ghan@essex.ac.uk

Abstract

This paper presents a shape prediction algorithm in a noisy video sequence based on pixel representation in the undecimated wavelet domain. In our algorithm for tracking of user-defined shapes in a noisy sequence of images, the amplitude of coefficients in the best basis tree expansion of the undecimated wavelet packet transform are used as feature vectors (FVs). FVs robustness against noise has been achieved through inherent denoising and edge component separation in the best basis selection algorithm. The algorithm uses these FVs to track the pixels of small square blocks located at the vicinity of the object boundary. Searching for the best-matched block has been performed using conventional block matching algorithm in the wavelet domain. Our experimental results show that the algorithm is robust to noise in case of object's shape translation, rotation and/or scaling and can be used to track both rigid and non-rigid shapes in image sequences.

Keywords—Object tracking; Undecimated Wavelet Packet; White Gaussian Noise; Shape extraction

1. Introduction

One of the general problems in the applied image and video processing is how to alleviate the effect of noise. Noise may be defined as any unwanted signal that interferes with the communication, measurement or processing of an information-bearing signal. 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 shape 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 success of a noise processing method depends on its ability to characterize and model the noise process, and to use the noise characteristics to discriminate the signal from the noise.

The wavelet transform which is both linear and shift-invariant is the Undecimated Wavelet Transform (UWT). It also gives a denser approximation to the continuous wavelet transform than the approximation provided by the orthonormal discrete wavelet transform.

UWT is one of the variants of the wavelet transform. UWT is simply the conventional filter bank implementation without down sampling, so that the low-low (LL), low-high (LH), high-low (HL), and high-high (HH) sub-bands remain at full size. Therefore, all the properties of the original signal are redundantly available in the transform domain. These over-complete libraries of the waveforms which span the signal space redundantly provide an adaptive selection of the basis for representation of the original signal. Some algorithms for adaptive selection of the best basis have been proposed in the literature. They remove the redundancy in the signal expansion and have a noise reduction property [1, 2].

The UWT can also be used to generate an Undecimated Wavelet Packet Transform (UWPT) tree. Contrary to the general wavelet transform where only the LL subbands are expanded in the lower levels, in the UWPT all the subbands have a full decomposition up to a predefined level. Therefore, the UWPT has a shift-invariant property [3].

In this paper, we have developed a new noise robust algorithm for tracking the user-defined shapes in noisy the image and video sequences by using the features generated in the UWPT for small blocks approximating the shape.

After surveying the literature of the undecimated wavelet transform in section 2, we present our new algorithm in section 3. Section 4 illustrates the experimental results and in section 5 conclusions and the future work are presented.

2. Undecimated Wavelet Transform

Many applications of the undecimated wavelet transform in image and video processing specially for noise/speckle reduction or removal have been presented in the literature [3]. Argenti et al. [4] proposed a method to denoise dependent additive distortion using undecimated wavelets. They considered a noise model to cover all types of noise processes. Noise reduction was performed by adaptive re-scaling of the generated coefficients of undecimated wavelet decomposition in Wiener-like filtering. Undecimated decomposition which is shift-invariant helps to prevent the ringing impairment compared to critically-subsampled wavelet denoising.

Lang et al. [5] presented a nonlinear noise reduction algorithm, which uses an Undecimated, shift-invariant, nonorthogonal wavelet transform. They used thresholding in the wavelet transform domain by following the Coifman [1] best basis selection algorithm. Their approach is a repeated application of the original Donoho and Johnstone method for different shifts. Their algorithm significantly reduced noise compared to the original wavelet based approach for a large class of signals.

Argenti and Alparone [6] proposed a speckle removal in SAR images using the undecimated wavelet transform. Their approached to the problem was based on minimum mean-squared error (MMSE) of the filtered undecimated wavelet coefficients by means of an adaptive rescaling of the detail coefficients, whose amplitudes are divided by the variance ratio of the noisy coefficients to the noise-free ones. All the above quantities are analytically calculated from the speckled image. The variance and autocorrelation of the fading variable and the wavelet filters only, without resorting to any model to describe the underlying backscatter. The absence of decimation in the wavelet decomposition avoids typical impairments often introduced by subsampled wavelet-based denoising

Carre et al. [7] proposed two denoising methods of the EHG signal by undecimated wavelets. Different source of noise create interfering signals with overlapping spectra in the EHG signal. The first proposed method uses the algorithm “a trou” with nonsymmetrical filters that result in a rapid and satisfactory denoising. The second algorithm exploits orthogonal wavelets and the result of the thresholding corresponds to the average of all circulant shifts denoised by a decimated wavelet transform.

Strickland and Hahn [8] developed a 2-stage wavelet transform for detecting and segmenting microcalcifications in mammograms. The first stage is based on an undecimated wavelet transform. Detection occurs at the HH and the combined LH+HL subbands. The second stage is designed to overcome the limitations of the simplistic Gaussian assumption and provides an accurate segmentation of the calcification boundaries. Detected pixel sites in HH and LH+HL are dilated then weighted before performing the inverse wavelet transform.

3. The Proposed Algorithm

In our proposed algorithm, object shape prediction in a noisy environment is performed by tracking some selected feature points located near the object’s boundary at a reference frame in the wavelet domain. A selected feature point is “near” a boundary if the distance between this point and the boundary is less than a predefined threshold T. To achieve the best prediction of the object’s shape, the number of these points (M) can be computed based on the threshold and size of the frame. Let S_t be the set of M feature points which approximate the object's shape at frame t

$$S_t = [s_t^1, s_t^2, \ldots, s_t^M], \quad s_t^i = [x_t^i, \; y_t^i] \qquad (1)$$

Where $[x_t^i, \; y_t^i]$ is the coordinates of point s_t^i at frame t.

At the next frame, t+1, each s_t^i undergoes a transformation and is represented by s_{t+1}^i. If S_{t+1} is known, the object's shape at frame t+1 can be reconstructed. Our approach to track s_t^i exploits the shift invariant property of the Undecimated Wavelet Packet Transform (UWPT) to extract an invariant feature vector for a pixel. UWPT also alleviates the problem of subband aliasing associated with the decimated transform such as (DWT, FT, etc). The method is based on our previous works [9, 10, 11].

For the sake of robustness of pixel tracking, for each s_t^i we define a square Q_t^i centered at s_t^i. The pixels within the Q_t^i are used to find the correct location of s_t^i in frame t+1 (s_{t+1}^i).

The algorithm tracks square Q_t^i in the next frame and finds the best matched Q_{t+1}^i and hence s_{t+1}^i. Having found S_{t+1}, the new object's shape in frame t+1 can simply be reconstructed.

The problem that needs to be specified is that how the algorithm generates the feature vector for each pixel in Q_t^i using the wavelet packet tree and tracks the generated feature vectors in frame t+1.

The wavelet packet tree is generated by the Undecimated Wavelet Packet Transform (UWPT). UWPT has two properties, which make it suitable for generating invariant and noise robust features corresponding to each pixel.

1- It has the shift invariant property. Due to this property, when pixels of Q_t^i move from frame t to new positions in frame t+1 (translation), there would be little changes in the value of the wavelet coefficients. Therefore, feature vectors that are 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. This feature simplifies the feature extraction procedure.

In general, biorthogonal wavelet bases which are particularly useful for object detection [10] could be used to generate the UWPT tree. The procedure for generating a feature vector for each pixel in frame t can be summarized as follows:

Stage 1:

1- Generate UWPT for frame t.

2- Perform entropy-based algorithms for the best basis selection [1] and prune the wavelet packet tree. The output of this step is an array of node numbers of the UWPT tree, which specify the best basis.

Having considered the second property of the UWPT, feature vector (FV) for each pixel in frame t (therefore in Q_t^i) can be simply created by selecting the corresponding wavelet coefficients in the best basis nodes of step 2. Therefore, the number of elements in FV is the same as the number of best basis nodes.

Step 2 is only performed for the reference frame and then the determined nodes' numbers of the best basis in this step are used to prune the UWPT tree of the successive frames. Therefore, the process of creating FVs for the pixels in the successive frames is simplified.

The procedure to match Q_t^i in frame t to Q_{t+1}^i in frame t+1 is as follows (stage 2):

1- Assume a search window in frame t+1 centered at pixel $[x_t^i, y_t^i]$.

2- By performing the procedure in stage 1, we have the FV for pixels in both Q_t^i and the search window

3- Choose a search block with the same size as Q_t^i to sweep the search window

4- Find the best match of Q_t^i in the search window by finding the minimum sum of the Euclidean distances between the search block and Q_t^i pixels' FVs (e.g. full search algorithm in the search window).

4. Experiments and results

In our experimental studies to simulate the noise and shape prediction for the noisy frames by the proposed algorithm, the following procedure has been followed:

1- The user generates the object's shape in the noise-free reference frame by defining the points that approximate the object's shape in the reference frame. Then a function automatically generates the selected feature points as the input for the algorithm. The number of points depends on the size of the object and the block size. The points are selected in a way that their centered blocks without overlapping cover the object's entire boundary.

2- In each successive frame, a White Gaussian Noise (WGN) with a defined energy (standard deviation) is added to the frame.
3- Based on the proposed algorithm, new object's shape will be predicted in the noisy frame.

Performance of the generated feature vectors for tracking of objects contaminated by the quantization noise has also been evaluated by conducting experiments on the compressed video [10]. Different parameters of the algorithm such as block size, search area, and the number of UWPT decomposition level have been reported before [11].

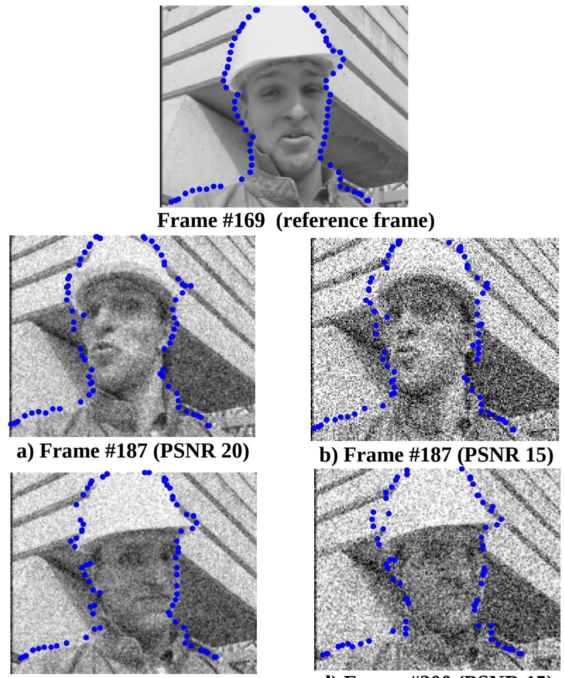
The algorithm has been implemented with 5x5 pixel blocks in a search window of ± 7 pixels, and 4 levels of UWPT decomposition. The selected wavelet family to generate the UWPT is bior2.2 [12]. In BiorNr,Nd, Nr and Nd represent the number of vanishing moments for the synthesis and analysis wavelets. The presence of spikes in bior2.2 makes it suitable for object detection applications [13].

The simulation has been conducted on frames 170 to 200 of the foreman, carphone and frames 1 to 100 of the bream sequences for different PSNR values (10, 15, 20, 25, and 30 dB). Selected range of the frames for the simulation includes shape deformation and different types of object's shape movement such as translation, rotation, and scaling. The results have been compared subjectively

Figure 1 shows the noise-free reference frame (frame 169) with its manually generated object's shape along with the results of object's shape prediction for noisy frames 187 and 200 for two different PSNR values of 15 and 20 dB.

As shown in figure 1, the frame 187 has an abrupt change including translation, rotation and deformation for some pixels at the borders. As the generated features are robust to noise, the algorithm can find new border pixels and predict the new shape both for PSNR=15 and 20 dB well.

The object's shape in frame 200 is similar to the reference frame except it has a small translation and rotation. At PSNR=15 dB some pixels at the left side of the foreman's hat are scattered, but one can use an interpolation algorithm to draw the object's border.



Frame #169 (reference frame)

a) Frame #187 (PSNR 20) **b) Frame #187 (PSNR 15)**

c) Frame #200 (PSNR 20) **d) Frame #200 (PSNR 15)**

Figure 1. Tracking foreman's shape in the "foreman" sequence, where the object's shape has translation and rotation: a & c) for PSNR= 20 dB b & d) for PSNR=15 dB

The algorithm has been applied on a moving object with a considerable motion at each frame in a noisy environment. Coast-guard is a test video sequence with such behavior. To cope with the high speed movement of the boat at some frames, we have used a wider search area. For instance, at frame 72, the search windows size is ±30 pixels. Having selected the pixels' position at the first frame, after passing 72 frames, the algorithm is still able to find the correct position of the pixels even in such a large search area (Fig. 2. Col. 2).

a

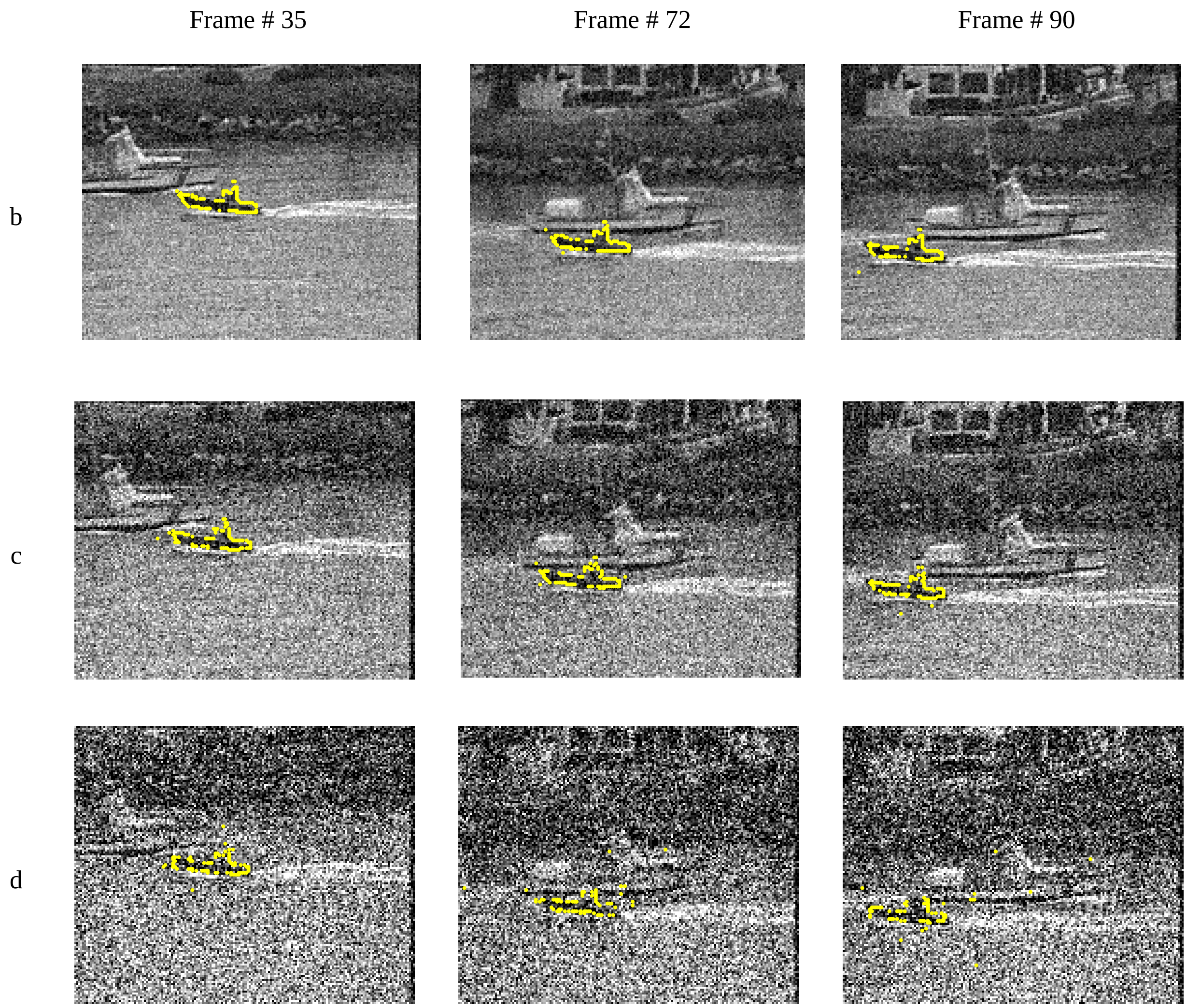
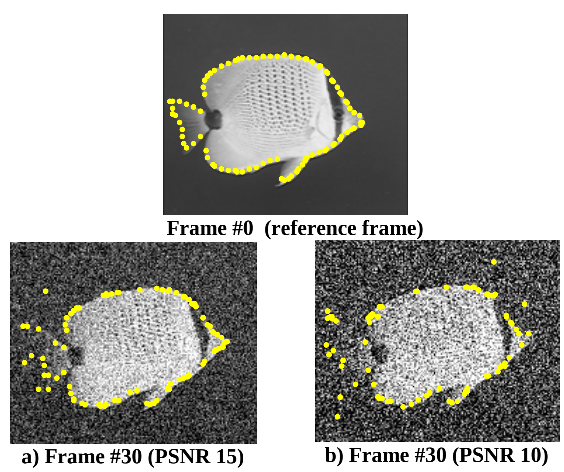


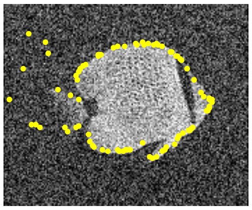
Figure 2. Tracking coast-guard in noisy enviroment a) frame # 0 (reference frame) b) for PSNR = 20 c) for PSNR = 15 d) for PSNR = 10

The effect of noise at the different PSNR values (10, 15, 20) on this sequence has been evaluated. The proposed method can find the matched pixels well and re-construct the object's shape (Fig. 2) in the PSNR as low as 10. The corresponding video sequences which indicates the tracking results at different PSNR values are available to download

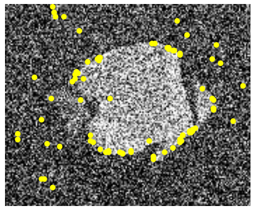
(http://mehr.sharif.edu/~khansari).

The Bream sequence contains a fish swimming and changing direction. Figure 3, shows the snapshots for the bream sequence. Frame 0 is the noise-free reference frame in which the object’s shape is generated manually.





c) Frame #100 (PSNR 15)



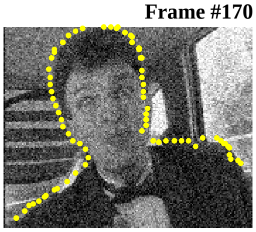
d) Frame #100 (PSNR 10)

Figure 3. Tracking bream's shape in the "bream" sequence, where there is translation , object zoom in (scaling), and a small rotation: a & c) for PSNR=15 dB b & d) for PSNR=10 dB

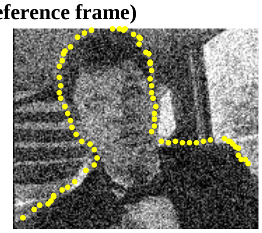
As shown in figure 3, frame 30 has translation and small rotation. In frame 100, object's shape undergoes scaling (zooming in) and translation. In spite of scattering some points near the bream's tail, the algorithm can follow the actual object's border well. In case of Bream's tail, the pixels at the tail do not have a clear edge, and their luminance value is close to the background. Therefore, the resulting tracked pixels are not valid and cause corruption at the reconstructed object's boundary. The algorithm performs well for PSNRs as low as 10 dB.

Figure 4, shows the simulation results for the carphone video sequence. Frame 170 is the noise-free reference frame in which the object's shape is generated manually.

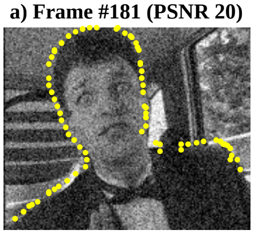
Frame #170 (reference frame)



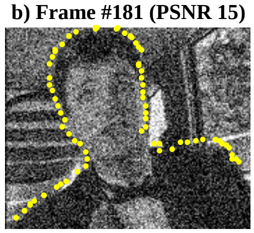
a) Frame #181 (PSNR 20)



b) Frame #181 (PSNR 15)



c) Frame #189 (PSNR 20)



d) Frame #189 (PSNR 15)

Figure 4. Tracking carphone's shape in the "carphone" sequence, where there is object zoom in (scaling): a & c) for PSNR=20 dB b & d) for PSNR=15 dB

As shown in figure 4, frame 181 has object's shape scaling (zooming in) and frame 189 has both scaling and translation. In both cases the results of shape prediction for PSNRs 15 and 20 dB are visually acceptable and follow the actual object's shape.

5. Conclusions

A new object shape tracking method in noisy video frames based on pixel features in the wavelet domain has been proposed. Shape prediction has been achieved by tracking small squares centered at selected pixels near the object's boundaries. For this purpose, a special feature vector corresponding to each pixel in the square has been generated. These features are extracted from the Undecimated Wavelet Packet Transform (UWPT) of the squares. UWPT tree pruning results in an inherent property of noise removal in the wavelet domain. Considering the shift invariant property of the UWPT, a search of reference block by employing generated feature vectors of pixels can simply be carried out in the search window of the next frame

The simple search method along with the aforementioned properties of the UWPT can successfully track objects even at the presence of noise and shape translation, rotation and zooming. Suitable features for selected pixels have been formed due to the inherent denoising and edge component separations in the best basis selection algorithm. Mahalanobis distance measure could provide a better separation among the FVs, but it more computational complex compared to the Euclidean distance measure. Simulation results show that the method is robust against WGN even with the PSNR as low as 10 dB and considerable object's motion.

Acknowledgment

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

References

[1] 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.

[2] 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.

[3] H. Guo, "Theory and Applications of the Shift-Invariant, Time-Varying and Undecimated Wavelet Transforms," Master Thesis, Rice University, 1995.

[4] F. Argenti, G. Torricelli, L. Alparone "Signal-Dependent Noise Removal in the Undecimated Wavelet Domain," IEEE ICASSP, pp. 3293-3296, Orlando, Florida, USA, May 2002.

[5] M. Lang, H. Guo, J.E. Odegard, C.S Burrus, R.O Wells, Jr. MeVis, "Noise Reduction Using an Undecimated Discrete Wavelet Transform," IEEE Signal Processing Letters, pp. 10-12, Jan 1996.

[6] F. Argenti, L. Alparone, "Speckle removal from SAR images in the undecimated wavelet domain," IEEE Transactions on Geoscience and Remote Sensing, Vol. 40, No. 11, pp. 2363- 2374, Nov 2002.

[7] P. Carre, H. Leman, C. Fernandez, C. Marque, "Denoising of the uterine EHG by an undecimated wavelet transform," IEEE Transactions on Biomedical Engineering, Vol. 45, No. 9, pp. 1104-1113, Sep 1998.

[8] R.N. Strickland, H. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms," IEEE Transactions on Medical Imaging, Vol. 15, No. 2, pp. 218-229, Apr 1996.

[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]M. Khansari, H. R. Rabiee, M. Asadi, M. Ghanbari, M. Nosrati, M. Amiri, "A Quantization Noise Robust Object's Shape Prediction Algorithm," in European Signal Processing Conference (EUSIPCO), Antalya, Turkey, September 4-8, 2005.

[11]M. Khansari, H. R. Rabiee, M. Asadi, M. Ghanbari, M. Nosrati, M. Amiri, "An Adaptive Semi-Automatic Video Object Extraction Algorithm Based on Joint Transform and Spatial Domains Features," International Workshop on Content-Based Multimedia Indexing (CBMI), Riga, Latvia, June 21-23, 2005.

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

[13]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.