A QUANTIZATION NOISE ROBUST OBJECT'S SHAPE PREDICTION ALGORITHM

M. Khansari[1], H. R. Rabiee[1], M. Asadi[1], , M. Nosrati[1], M. Amiri[1], M. Ghanbari[1,2]

[1]Digital Media Lab, Advanced Information and Communication Technology Center, Sharif University of Technology
Azadi Ave., Tehran, Iran
Phone: + (9821)6165169, fax: + (9821)6047662, Emails: Khansari@mehr.sharif.edu,
rabiee@sharif.edu, asadi@dml.aictc.com, mohadese_nosrati@yahoo.com, amiri@mehr.sharif.edu,
Web: http://www.aictc.com/dml
[2]Department of Electrical Engineering, University of Essex, England
Email: ghan@essex.ac.uk
Web: http://privatewww.essex.ac.uk/~ghan/

ABSTRACT

This paper introduces a quantization noise robust algorithm for object's shape prediction in a video sequence. The algorithm is based on pixel representation in the undecimated wavelet domain for tracking of the user-defined shapes contaminated by the compression noise of video sequences. In the proposed algorithm, the amplitude of coefficients in the best basis tree expansion of the undecimated wavelet packet transform is used as feature vectors (FVs). FVs robustness against quantization 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 neighborhood of the object boundary. Searching for the best matched block has been performed through the use of conventional block matching algorithm in the wavelet domain [9].

Our experimental results show that the algorithm is robust against the quantization noise of rigid/non-rigid object's shape translation, rotation and/or scaling.

1. INTRODUCTION

One of the general problems, which are raised 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. There are two types of noises which are very important in image and video systems, namely; quantization (compression) noise and channel noise.

Quantization noise is inherent in the amplitude quantization process in image and video encoders. This process is mainly used to control the trade off between compression and quality.

This type of noise has direct impact on the decoded image and video visual quality. 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 which make the tracking of noisy objects more complicated.

In this paper, we have developed a new noise robust algorithm for tracking the user-defined shapes in noisy image and video sequences by using the features generated in the UWPT domain for selected pixels 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

Undecimated Discrete Wavelet Transform (UDWT) is one of the variants of the wavelet transform. UDWT 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 an inherent noise reduction property [1, 2].

The UDWT can also be used to generate an Undecimated Wavelet Packet Transform (UWPT) tree. Contrary to the Dyadic 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, time-varying property [3].

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 a Wiener-like filtering. Undecimated decomposition which is time-invariant helps to prevent the ringing impairments 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 the 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 a noisy coefficient to the noise-free one. All the 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 sources of noise create interfering signals with overlapping spectra in the EHG signal. The first proposed method uses the algorithm “a trou” with nonsymmetrical filters which 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 the 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 frame 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*

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

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 invariance property of the Undecimated Wavelet Transform (UWT) to extract an invariant feature vector for a pixel [3].

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 time 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 which 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 are all of the same size, which are also 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

indices of the UWPT tree which specify the best basis.

3. 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 index 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

Experiments have been conducted to simulate the effect of quantization noise on the proposed shape prediction algorithm. To produce a noisy frame a standard video/image codec has used. By changing the parameters of the codec, one can find the appropriate parameters to re-generate the frame with a pre-defined quantization noise. This results in frames with specific bit rate/PSNR. The following procedure has been followed to verify the algorithm.

1- Encoding the frames of an image / video sequence with the specific quantization noise (bit rate).

2- Generating user-defined object's shape in the noisy reference frame by defining the points which approximate the shape in the reference frame. Then, a function automatically generates the selected feature points as the input for the algorithm.

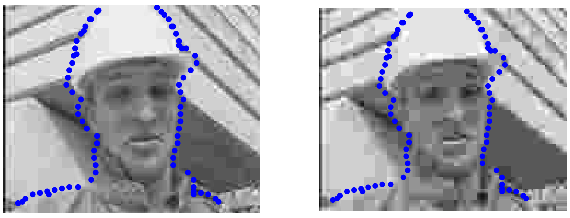
3- Predicting the user-defined shape by applying the proposed method in the successive frames.

The performance of block representation in the wavelet domain [8] and tracking the generated feature vectors for pixels in the noisy frames have been evaluated by conducting experiments with different PSNRs and bit rates.

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 was biort2.2. 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 [9]. The performance evaluation of the object shape tracking has been done subjectively.

The simulation has been conducted on frames 170 to 200 of the foreman sequence for two different PSNR values (31 and 33 dB). This range of the frames includes different object's shape deformation such as translation, rotation, and a small scaling.

Figure 1 shows the noisy reference frame (frame 169) with two different PSNR values (31 and 33 dB) in which the object's shape is generated manually.



a1) Frame #169 (0.45 bpp, PSNR 32.86 dB) a2) Frame #169 (0.32 bpp, PSNR 30.87)

Fig. 1. Noisy reference frame (frame 169 of foreman sequence) for two PSNR values.

Figure 2 illustrates the results of object's shape prediction for selected noisy frames (173, 178, 188, and 190). In these frames the blocking artifacts smooth out the shape's border and complicate shape tracking for spatial domain features such as color, texture and shape. On the contrary, the proposed algorithm features in the wavelet domain are robust to a greater extent.

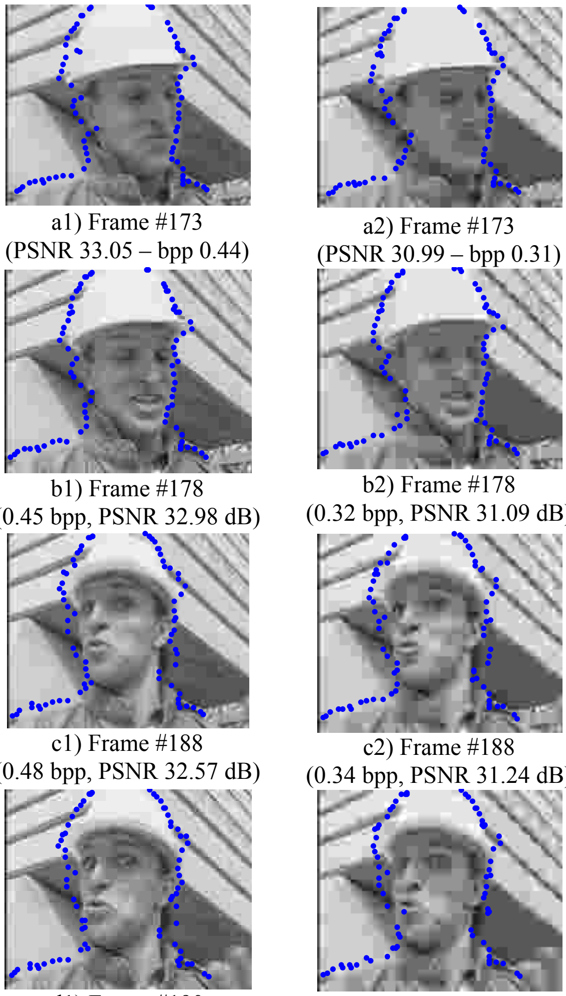
The border points in frame 173 (Fig. 2. a1 and a2) have translation and rotation. As the generated features are quantization noise robust, the algorithm can find the new border pixels and predict the new shape for both PSNR values.

The object's shape in frames 178 is similar to the reference frame except it has a little translation and rotation. In this case shape deformation is overcome by the proposed method and results in a satisfactory object's shape.

In frames 188 and 190 (Fig. 2.) the object's shape has many changes in comparison to the reference frame. Both frames have scaling and translation to some extent. Frame 190 has a small scaling (zooming in) with regards to frame 188.

In this case, some pixels at the right side of the Foreman's neck are distorted, but one can use an interpolation algorithm to reconstruct the object's border.

In both cases the results of shape prediction are visually acceptable and follow the actual object's shape.



a1) Frame #173 (PSNR 33.05 – bpp 0.44)
a2) Frame #173 (PSNR 30.99 – bpp 0.31)
b1) Frame #178 (0.45 bpp, PSNR 32.98 dB)
b2) Frame #178 (0.32 bpp, PSNR 31.09 dB)
c1) Frame #188 (0.48 bpp, PSNR 32.57 dB)
c2) Frame #188 (0.34 bpp, PSNR 31.24 dB)
d1) Frame #190 (0.47 bpp, PSNR 32.78 dB)
d2) Frame #190 (0.33 bpp, PSNR 31.13 dB)

Fig. 1. Tracking foreman's shape in the "foreman" sequence, where the object's shape has translation, rotation, and scaling.

5. CONCLUSIONS

A new object shape tracking method for video frames contaminated by the quantization noise 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 frame. Pruning of the UWPT tree results in an inherent property of noise removal in the wavelet domain. Considering the time 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 simulation results show the robustness of the algorithm against the blocking artifacts which have been caused by quantization noise especially in low bit rates.

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 discriminant measure could provide a better separation among FVs, but it is more computational complex compared to the Euclidean distance measure. The proposed method can be integrated into a semi-automatic video object plane (VOP) generation that can be used in the MPEG4 encoders.

ACKNOWLEDGMENTS

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

6. 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] 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.