A Proposed Intelligent Denoising Technique for Spatial Video Denoising for real-time applications

Associated Prof. Dr. AMANY SARHAN
Prof. Dr. MOHAMED T. FAHEEM
Department of Computer and Automatic Control Engineering,
Faculty of Engineering, Tanta University
Eng. RASHA ORBAN MAHMOUD
Nile Institute of Commerce & Computer Technology, Mansoura

Abstract:

With the wide spread of video usage in many fields of our lives, it becomes very important to develop new techniques for video denoising. Spatial video denoising using wavelet transform has been the focus of the current researches as it requires less computation and more suitable for real-time applications. Two specific techniques for spatial video denoising using wavelet transform are considered in this work: 2D Discrete Wavelet Transform (2D DWT) and 2D Dual Tree Complex Wavelet Transform (2D DTCWT). We performed an analytical analysis to investigate the performance of each of these techniques. From this analysis, we found out that each of these techniques has its advantages and disadvantages. The first technique gives less quality at high levels of noise but consumes less time while the second gives high quality video while consuming long. Based on this in this work, we introduce an intelligent denoising system that makes a tradeoff between the quality of the denoised video and the time required for denoising. The system first estimates the noise level in the video frame then accordingly chooses the proper of the two denoising techniques to apply on the frame. The simulation results show that the proposed system is more suitable for real-time applications where the time is critical while giving high quality videos especially at low to moderate levels of noise.

KeyWords: - Video denoising, 2D wavelet, dual-tree complex, wavelet transform.

I. Introduction

The recent advancement in multimedia technology has promoted an enormous amount of research in the area of image and video processing. Included in the many image and video processing applications, such as compression, enhancement, and target recognition, is preprocessing functions for noise removal. Noise removal is one of the most common and important processing steps in many image and video systems. Because of the importance and commonality of preprocessing in most image and video systems, there has been an enormous amount of research dedicated to the subject of noise removal, and many different mathematical tools have been proposed [2].

Noise refers to unwanted stochastic variations as opposed to deterministic distortions such as shading or lack of focus. It can be added to the video signal or multiplied with the video signal. It can also be signal dependent or signal independent [6]. Based on its spectral properties, noise is further classified as white or color noise. Many types of noise effect charge-coupled device (CCD) cameras such as photon shot noise and read out noise. Photon shot noise is due to the random arrival of photons at the sensor, which is governed by Poisson distribution. Other sources of noise include output amplifier noise, camera noise and clock noise, which can be combined in a single equivalent Gaussian noise source called read out noise. Because of the high counting effect of Photon arrivals and according to the central limit theorem, the aggregate noise effect can be well approximated by Gaussian distribution. Consequently, in this paper, an Additive White Gaussian Noise (AWGN) model is assumed. The choice is also motivated by AWGN being the most common noise model for TV broadcasting [6].

Spatial video denoising techniques use both the two dimensional Dual Tree Complex Wavelet Transform (2D DTCWT) and three dimensional Dual Tree Complex Wavelet Transform (3D DTCWT)[11], temporal video denoising techniques use temporal filtering only [15], while spatio-temporal video denoising techniques use combination of spatial and temporal denoising [15].

The need for fast and accurate video noise estimation algorithms rises from the fact that many fundamental video processing algorithms such as compression, segmentation, motion estimation and format conversion adapt their parameters and improve performance when the noise is known. The effectiveness of video processing methods can be significantly reduced in the presence of noise. When information about the noise becomes available, processing can be adapted to the amount of noise to provide stable processing methods [5].

A noise estimation technique calculates the level of white Gaussian noise, which is the most commonly assumed noise type in video processing applications, contained in a corrupted video signal. When noise variance becomes available, video denoising algorithms (e.g., 2D Discrete Wavelet Transform (2D DWT) and 2D Dual Tree Complex Wavelet Transform (2D DTCWT)) can be adapted to the amount of noise for improved performance.

Video noise can be estimated spatially or temporally. A widely used spatial noise estimation method calculates the variance (as a measure of homogeneity) over a set of image blocks and averages the smallest block variance as an estimate of the image noise variance. Spatial variance-based methods tend to overestimate the noise in less noisy images and underestimate it in highly noisy and textured images. Therefore, measures other than the variance were introduced in [5] to determine homogeneous blocks. Temporal noise estimation evaluates noise using motion information [14]. Such approach is very expensive for hardware implementations with estimation accuracy not significantly more precise than spatial methods.

This paper aims to introduce a novel intelligent denoising system for spatial video denoising. Two of the most widely used denoising techniques will be used namely: 2D DWT and 2D DTCWT. We conducted an comprehensive analysis of both techniques to identify the strength and weakness points of each. From this analysis, we found that each technique has its advantages and disadvantages. DWT has the advantage of consuming minor in the denoising process; however, it has the disadvantages of producing less quality video at high levels of noise compared to the DTCWT. DTCWT has the advantage of producing high quality video at high levels of noise while having the disadvantages of consuming minor in the denoising process compared to the DWT, which makes it unsuitable for real-time applications. Therefore, a tradeoff has to be made between the quality of the produced video and the time consumed for denoising between the two techniques to get the benefits and discarding the disadvantages of both.

Based on this analysis, we propose a novel intelligent denoising system that makes use of both spatial denoising techniques. The system ha intelligence in that it has a component in the system, called the noise estimator, that that estimates the noise contained in the frame (as we will work on spatial domain). The system accordingly makes a decision on how to handle this frame based on the noise level presented in it. The intelligent choice of the applied denoising technique will lead to an efficient and flexible system for fast and reliable spatial video denoising.

II. Video Denoising Techniques

Denoising is still one of the most fundamental, widely studied, and largely unsolved problems in video processing. The purpose of denoising (or restoration) is to estimate the original video (or a "better" representative of it) from noisy data. Many methods for video denoising have been suggested, but the wavelet transform has been viewed by many as the preferred technique for noise removal [10]. Rather than a complete transformation into the frequency domain, as in DCT or FFT, the wavelet transform produces coefficient values which represent both time and frequency information. The hybrid spatial-frequency representation of the wavelet coefficients allows for analysis based on both spatial position and spatial frequency content. The hybrid analysis of the wavelet transform is excellent in facilitating video denoising algorithms [13].

A. Video Denoising Techniques Based on Wavelet Transform

In recent years, the multiresolution analysis, more specifically the wavelet transform, has shown considerable success in signal denoising. Wavelet analysis is a powerful and popular tool for the analysis of nonstationary signals. The wavelet transform is a joint function of a time series of interest x (t) and an analyzing function or wavelet $\psi(t)$. This transform isolates signal variability both in time t, and also in "scale" s, by rescaling and shifting the analyzing wavelet. The wavelet itself can be said to play the role of a lens through which a signal is observed. Therefore, it is important to understand how the wavelet transform depends upon the wavelet properties [9, 10]. There are two famous types of video denoising that use wavelet transform namely; 2D Discrete Wavelet Transform (2D DWT) and 2D Dual Tree Complex WT (2D DTCWT).

B. 2-D Discrete Wavelet Transform (2D DWT)

The 2D DWT is a very modern mathematical tool. It is used in compression, denoising and watermarking applications. It is built with separable orthogonal mother wavelets, having a given regularity. The DWT gives a multiscale representation of a signal x (n). The DWT is implemented by iterating the 2-channel analysis filter bank described above. Specifically, the DWT of a signal is obtained by recursively applying the lowpass/highpass frequency decomposition to the lowpass output as illustrated in the diagram, see Fig.1. The diagram illustrates a 3-scale DWT. The DWT of the signal x is the collection of subband signals. The inverse DWT is obtained by iteratively applying the synthesis filter bank [13].

DWT has the following advantages:

- Multi-scale signal processing technique.
- Number of significant output samples is very small and hence the extracted features are well characterized.
- Straightforward computation technique.

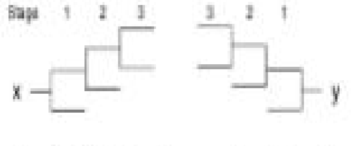

Figure 1. DWT Multi-scale representation of a signal x.

Although the Discrete Wavelet Transform (DWT) in its maximally decimated form (Mallat's dyadic filter tree [4]) has established an impression, its use for other signal analysis and reconstruction tasks has been hampered by two main disadvantages:

- Lack of shift invariance, which means that small shifts in the input signal can cause major variations in the distribution of energy between DWT coefficients at different scales.
- Poor directional selectivity for diagonal features, because the wavelet filters are separable and real.

The 2D DWT is simply the application of the 1D-WT repeatedly to first horizontal data of the image, then the vertical data of the image. The discrete wavelet transform [4] is an algorithm for computing the coefficients $s_{j,k}$ and $d_{j,k}$ in the wavelet expansion of a signal.

$$f(x)=\sum_k s_{j,k}\phi_{j,k}(x)+\sum_k d_{j,k}w_{j,k}(x)+\sum_k d_{j-1,k}w_{j-1,k}(x)+\ldots+\sum_k d_{1,k}w_{1,k}(x) \quad (1)$$

Where j is the number of multiresolution components (or scales), and k ranges from 1 to the number of coefficients in the specified component. ϕ is the scaling function and the w is the wavelet function through dilation and translation as follows

$$\phi_{j,k}(x)=2^{-j/2}\phi(2^{-j}x-k) \quad and \quad w_{j,k}(x)=2^{-j/2}w(2^{-j}x-k) \quad (2)$$

The scaling $\phi(x)$ function is the solution of the dilation equation

$$\phi(x)=\sqrt{2}\sum_k c_k\phi(2x-k) \quad (3)$$

Where the coefficients c_k must satisfy the following conditions [4]:

Unit vector: $\sum_k |c_k|^2 = 1$.

Double-shift: $\sum_k c_k c_{k-2m} = 0$, m =1,2,…, p-1.

Approximation of order p: $\sum_k (-1)^k k^m c_k = 0$, $m = 0, 1, \ldots, p-1$. Where p = (number of coefficients)/2.

While, the wavelet function $w(x)$ can be derived from the corresponding scaling function by taking difference. For the four-coefficient scaling function, the wavelet equation is expressed as:

$$w(n)=\sqrt{2}\sum_k d_k\phi(2n-k) \quad (4)$$

Where $d_k = (-1)PP^{2x-k-1}c_{2p-1-k}$

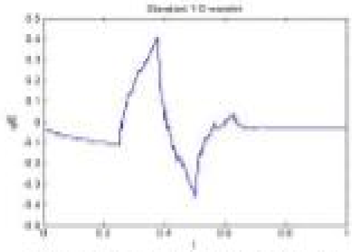

Figure 2. Mother wavelet function (Daubechie's 4).

More precisely, the expansion in (1) for any arbitrary signal $f(x)$ may take the form

$$f(x)=\sum_{k=-\infty}^{\infty} a_k\phi(x-k)+\sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} a_{j,k}w(2^j x-k) \quad (5)$$

Where the coefficients are given by

$$a_k=\int_{-\infty}^{\infty} f(x)\phi(x-k)dx. \quad and \quad a_{j,k}=\int_{-\infty}^{\infty} f(x)w(2^j x-k)dx$$

This wavelet series expansion decomposes $f(x)$ into an infinite summation of wavelets at different scales. For computing the coefficients a_k and $a_{j,k}$ in (5) when $f(x)$ is sampled over some certain interval, the discrete wavelet transform is employed.

To use the wavelet transform for image processing we must implement a 2D version of the analysis and synthesis filter banks. Fig. 3 shows 2-Channel Perfect Reconstruction Filter Bank.

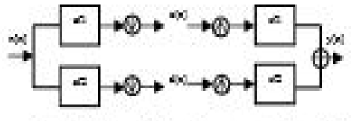

Figure 3. Channel Perfect Reconstruction Filter Bank.

C. 2-D Dual Tree Complex WT (2D DTCWT)

The dual-tree CWT comprises of two parallel wavelet filter bank trees that contain carefully designed filters of different delays that minimize the aliasing effects due to downsampling [4]. The dual-tree CDWT of a signal $x(n)$ is implemented using two critically-sampled DWTs in parallel on the same data, as shown in Fig. 4. The transform is two times expansive because for an N-point signal it gives 2N DWT coefficients. If the filters in the upper and lower DWTs are the same, then no advantage is gained. Therefore, the filters are designed in a specific way such that the subband signals of the upper DWT can be interpreted as the real part of a complex wavelet transform and subband signals of the lower DWT can be interpreted as the imaginary part. When designed in this way the DTCDWT is nearly shift invariant, in contrast to the classic DWT.

Moreover, the dual-tree complex DWT can be used to implement 2D wavelet transforms where each wavelet is oriented, which is especially useful for image processing. (For the 2D DWT, recall that one of the three wavelets does not have a dominant orientation.) The DTCWT outperforms the critically-sampled DWT for applications like image denoising and enhancement.

One of the advantages of the DTCWT is that it can be used to implement 2D wavelet transforms that are more selective with respect to orientation than is the 2D DWT [13]. Let *w2* represent the parent of *w1* (*w2* is the wavelet coefficient at the same spatial position as w1, but at the next coarser scale) [7]. Then:

$$y = w + n$$

Where $w = (w1, w2)$, $y = (y1, y2)$ and $n = (n1, n2)$. The noise values $n1$, $n2$ are zero-mean Gaussian with variance sigma (σ) [11], [12]. Based on the empirical histograms, the following non-Gaussian bivariate equation was used [11].

$$p_w(w) = \frac{3}{2\pi\sigma^2} \cdot \exp\left(-\frac{\sqrt{3}}{\sigma}\sqrt{w_1^2 + w_2^2}\right) \quad (6)$$

With this equation, *w1* and *w2* are uncorrelated, but not independent [12]. The MAP estimator of *w1* yields the following bivariate shrinkage function [1], [8].

$$\hat{w}_1 = \frac{\left(\sqrt{y_1^2 + y_2^2} - \frac{\sqrt{3}\sigma_n^2}{\sigma}\right)_+}{\sqrt{y_1^2 + y_2^2}} \cdot y_1 \quad (7)$$

In general, the DTCWT has the following properties:

- Good directional selectivity in 2-dimentions (also true for higher dimensionality m-D);
- Perfect reconstruction (PR) using short linear-phase filters;
- Limited redundancy, independent of the number of scales, 2^m:1 for m-D;

Efficient order-N computation- only twice the simple DWT for 1-D (2^m times for m-D).

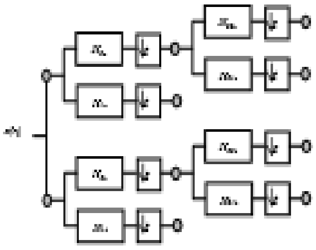

Figure 4. The Dual-Tree complex DWT of a signal x.

III. Noise Level Estimation

The effectiveness of video processing methods can be significantly reduced in the presence of noise. The level of noise can affect the performance of video denoising algorithms. Other video processing algorithms such as compression, segmentation, motion estimation and format conversion adapt their parameters and improve performance when the noise is known. Algorithms for estimating the AWGN variance are either temporal or spatial [5]. There exist few methods for purely temporal noise estimation such as [14]. These methods are challenged by the presence of object or global motion. Motion detection or motion compensation is commonly used as countermeasures. Hence, method in this area such as [14] requires more memory and is, in general, tends to be computationally expensive with estimation accuracy not significantly more precise than spatial methods [5].

For the spatial noise estimation method, the level of noise in a given digital frame is estimated from the noisy frame data. From [2] a median value of the $Y_{hh0}[\,]$ subband is used in the estimation process as follows.

$$\hat{\sigma}_n = \frac{Median\left(\left|Y_{hh0}[\,]\right|\right)}{0.6745} \quad (8)$$

Where $Y_{hh0}[\,]$ are the noisy wavelet coefficients in the high-high band of the 0^{th} scale.

Because the vast majority of useful information in the wavelet domain is confined to few and large coefficients, the median can effectively estimate the level of noise (i.e. the average level of the useless coefficients) without being adversely influenced by useful coefficients [2, 3].

It is worth to denote that the human visual system appears to have sensitivity thresholds. So, all frames with σ value less than 5dB look equally clean to human eye, but once the σ value exceeds 20 dB, the frame simply looks bad. This notation is very essential to us and will affect the decision making stage in the proposed system as will be discussed.

IV. Comparison criteria

In this work, we analyze the usage of two of the famous wavelet transform image denoising techniques in the spatial video denoising namely; 2D DWT and 2D DTCWT. Both techniques were developed originally for image denoising and have been used in spatial video denoising. However, there has not been any analysis of their performance or a comparison between them yet. So, in this work we intend to introduce such a comparison study to facilitate the choice between them in the different applications.

We will work on the spatial domain where we split the video stream into a number of frames (images). Then for each of these frames we apply the 2D DWT or the 2D DTCWT technique for denoising. We believe that even if the DTCWT denoising technique worked well on the image scale denoising, it may have to be studied closely to ensure its effectiveness on video denoising. We will concentrate on the spatial denoising in this work hoping to extend it to both temporal and spatio-temporal domains.

Validation of the performance of both techniques was done by comparing their resultant video criteria and this was