



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/75659/>

Article:

Bhowmik, Deepayan and Abhayaratne, Charith (2012) 2D+t Wavelet Domain VideoWatermarking. *Advances in Multimedia*, 2012. 973418. ISSN: 1687-5680

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Research Article

2D+t Wavelet Domain Video Watermarking

Deepayan Bhowmik and Charith Abhayaratne

Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 3JD, UK

Correspondence should be addressed to Charith Abhayaratne, c.abhayaratne@sheffield.ac.uk

Received 29 November 2011; Revised 20 January 2012; Accepted 21 January 2012

Academic Editor: Chong Wah Ngo

Copyright © 2012 D. Bhowmik and C. Abhayaratne. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel watermarking framework for scalable coded video that improves the robustness against quality scalable compression is presented in this paper. Unlike the conventional spatial-domain ($t + 2D$) water-marking scheme where the motion compensated temporal filtering (MCTF) is performed on the spatial frame-wise video data to decompose the video, the proposed framework applies the MCTF in the wavelet domain ($2D + t$) to generate the coefficients to embed the watermark. Robustness performances against scalable content adaptation, such as Motion JPEG 2000, MC-EZBC, or H.264-SVC, are reviewed for various combinations of motion compensated $2D + t + 2D$ using the proposed framework. The MCTF is improved by modifying the update step to follow the motion trajectory in the hierarchical temporal decomposition by using direct motion vector fields in the update step and implied motion vectors in the prediction step. The results show smaller embedding distortion in terms of both peak signal to noise ratio and flickering metrics compared to frame-by-frame video watermarking while the robustness against scalable compression is improved by using $2D + t$ over the conventional $t + 2D$ domain video watermarking, particularly for blind watermarking schemes where the motion is estimated from the watermarked video.

1. Introduction

Several attempts have been made to extend the image watermarking algorithms into video watermarking by using them either on frame-by-frame basis or on 3D decomposed video. The initial attempts on video watermarking were made by frame-by-frame embedding [1–4], due to its simplicity in implementation using image watermarking algorithms. Such watermarking algorithms consider embedding on selected frames located at fixed intervals to make them robust against frame dropping-based temporal adaptations of video. In this case, each frame is treated separately as an individual image; hence, any image-watermarking algorithm can be adopted to achieve the intended robustness. But frame-by-frame watermarking schemes often perform poorly in terms of flickering artefacts and robustness against various video processing attacks including temporal desynchronization, video collusion, video compression attacks, and so forth. In order to address some of these issues, the video temporal dimension is exploited using different transforms, such as discrete Fourier transform (DFT), discrete cosine transform (DCT),

or discrete wavelet transform (DWT). These algorithms decompose the video by performing spatial 2D transform on individual frames followed by 1D transform in the temporal domain. Various transforms are proposed in 3D decomposed watermarking schemes, such as 3D DFT domain [5], 3D DCT domain [6], and more popularly multiresolution 3D DWT domain watermarking [7, 8]. A multilevel 3D DWT is performed by recursively applying the above-mentioned procedure on low-frequency spatiotemporal subband. Various watermarking methods similar to image watermarking are then applied to suitable subbands to balance the imperceptibility and robustness. 3D decomposition-based methods overcome the issues like temporal desynchronization, video format conversion, and video collusion. However, such naive subband decomposition-based embedding strategies without considering motion element of the sequence during watermark embedding often result in unpleasant flickering visual artefacts. The amount of flickering in watermarked sequences varies according to the texture, colour, and motion characteristics of the video content as well as the watermark strength and the choice of frequency subband used for

watermark embedding. At the same time, these schemes are also fragile to video compression attacks, which consider motion trajectory during compression coding.

In order to address such issues stated above, we have extended image watermarking techniques into video considering the motion and texture characteristics of the video sequence using wavelet-based motion compensated $2D + t + 2D$ filtering. The proposed approach is evolved from the motion compensated temporal filtering- (MCTF-) based wavelet domain video decomposition concept. MCTF has been successfully used in wavelet-based scalable video coding research [9, 10]. The idea of MCTF was originated from 3D subband wavelet decomposition, which is merely an extension of spatial domain transform into temporal domain [11]. But 3D wavelet decomposition alone does not decouple motion information and it is addressed by using temporal filtering along the motion trajectories. This MCTF-based video decomposition technique motivates a new avenue in transform domain video watermarking. Few attempts have already been made to investigate the effect of motion in video watermarking by incorporating motion compensation into video watermarking algorithms [12–14]. In these investigations, the sequence is first temporally decomposed into Haar wavelet subbands using MCTF and then spatially decomposed using the 2D DCT transform resulting in the decomposition scheme widely known as $t + 2D$.

In this paper, we aim to advance further by investigating along the line of MCTF-based wavelet coding to propose a robust video watermarking scheme against scalable content adaptation, such as Motion JPEG 2000, MC-EZBC, or H.264-SVC, while keeping the imperceptibility. Apparent problems of direct use of MCTF and $t + 2D$ decompositions in watermarking are three fold.

- (1) In scalable video coding research, it is evident that video with different texture and motion characteristics leading to its spatial and temporal features perform differently on $t + 2D$ domain [9] and its alternative $2D + t$ domain [15], where MCTF is performed on 2D wavelet decomposition domain. Further, in 3D subband decomposition for video watermarking, the use of MCTF is only required for subbands where the watermarks are embedded. Hence, the motion estimation and compensation on full spatial dimension ($t + 2D$ case) add unnecessary complexity to the watermarking algorithm.
- (2) Conventional MCTF is focused on achieving higher compression and thus gives more attention on the prediction-lifting step in MCTF. However, for watermarking, it is necessary to follow the motion trajectory of content into low-frequency temporal subband frames, in order to avoid motion mismatch in the update step of MCTF when these frames are modified due to watermark embedding.
- (3) $t + 2D$ structure offers better energy compaction in the low-frequency temporal subband, keeping most of the coefficient values to very small or nearly zero in high-frequency temporal subbands. This is very useful during compression but leaves very little room for

watermark embedding in high-frequency temporal subbands. Therefore, for a robust algorithm, most of the MCTF domain watermarking schemes, mentioned before, embed the watermark in the lowpass temporal frames. On the other hand, $2D + t$ provides more energy in high-frequency subbands, which enables the possibility to embed and recover the watermark robustly using highpass temporal frames which improves the overall imperceptibility of the watermarked video.

To overcome these shortcomings, we propose MCTF-based 3D wavelet decomposition scheme for video sequences and offer a flexible $2D + t + 2D$ generalized motion compensated temporal-spatial subband decomposition scheme using a modified MCTF scheme for video watermarking. Using the framework, we study and analyze the merits and the demerits of watermark embedding using various combinations of $2D + t + 2D$ structure and propose new $2D + t$ video watermarking algorithms to improve the robustness performance against quality scalable video compression.

The rest of the paper is organized as follows. In Section 2, the modified MCTF scheme is presented along with the new $2D + t + 2D$ subband decomposition framework. The video watermarking algorithms using the implementation of different subband decomposition schemes are proposed in Section 3. The analysis of the framework is described in Section 4. The experimental results are shown and discussed in Section 5 followed by the conclusions in Section 6.

2. Motion Compensated Spatiotemporal Filtering

The generalized spatiotemporal decomposition scheme consists of two modules: (1) MCTF and (2) 2D *spatial frequency decomposition*. To capture the motion information accurately, we have modified the commonly used lifting-based MCTF by tracking interframe pixel connectivity and use the 2D wavelet transform for spatial decomposition. In this section, first we describe the MCTF with implied motion estimation and then propose the $2D + t + 2D$ general framework.

2.1. MCTF with Implied Motion Estimation. We formulate the MCTF scheme giving more focus into the motion trajectory-based update step as follows. Let I_t be the video sequence, where t is the time index in display order. We consider two consecutive frames I_{2t} and I_{2t+1} , as the current frame (c) and the reference frame (r), respectively, following the video coding terminology. In traditional motion estimation for lifting-based MCTF [9], I_{2t+1} frame usually partitioned into nonoverlapping blocks and for each block, motion is estimated from I_{2t} frame using a block matching algorithm. In this case only, two types of pixel connectivity are considered, (1) pixels are connected or (2) unconnected. In the case of several pixels are connected to the same pixel in the reference frame, only one of them is categorized as a connected pixel. The temporal frames are derived using

the subband analysis pair by replacing the I_{2t} as the low-frequency temporal frame (L) and I_{2t+1} as the high-frequency temporal frame (H).

Connected pixels:

$$\begin{aligned} L[m - \mathcal{H}^{c-r}, n - \mathcal{V}^{c-r}] \\ = \frac{1}{\sqrt{2}} I_{2t+1}[m, n] + \frac{1}{\sqrt{2}} I_{2t}[m - \mathcal{H}^{c-r}, n - \mathcal{V}^{c-r}], \end{aligned} \quad (1)$$

$H[m, n]$

$$= \frac{1}{\sqrt{2}} I_{2t+1}[m, n] - \frac{1}{\sqrt{2}} I_{2t}[m - \mathcal{H}^{c-r}, n - \mathcal{V}^{c-r}],$$

where \mathcal{V}^{c-r} and \mathcal{H}^{c-r} represent the motion vector fields: vertical and horizontal displacements of the nonoverlapping blocks, respectively.

Unconnected pixels:

$$L[m, n] = \frac{2 I_{2t}[m, n]}{\sqrt{2}}. \quad (2)$$

For the unconnected pixels in I_{2t+1} , the scaled displaced frame differences are substituted to form the temporal high subband.

As stated in the introduction, such a traditional scheme gives more attention on the prediction-lifting step in MCTF to reduce the prediction error in high-frequency subband. This is useful in a compression scenario. However, in the case of watermarking, we account the object motion within low-frequency temporal frames to avoid motion mismatch in update step when these frames are modified due to watermark embedding. To address this, we have used MCTF with implied motion estimation, which allows opportunity to embed the watermark in any chosen low- or high-frequency temporal frame. At the same time, as opposed to the traditional scheme, we consider the relative contributions of one-to-many connected pixels and this is important to capture the motion information accurately during MCTF operation.

In the proposed scheme, the I_{2t} frame is partitioned into nonoverlapping blocks and for each block, vertical and horizontal displacements are quantified and represented as motion vector fields \mathcal{V}^{c-r} and \mathcal{H}^{c-r} , respectively. Figure 1 shows an example how the four nonoverlapping blocks in the current frame (I_{2t}) are moved in different direction in the next frame (I_{2t+1}). In the I_{2t} frame, each block can be one of two types, namely, inter- and intrablocks, where the motion is only estimated for the former block type. Similarly, in the I_{2t+1} frame, any pixel can be one of three types, namely, one-to-one connected (point A), one-to-many connected (point B and C), and unconnected (point D) (as shown in Figure 1), depending on its connectivity to pixels in the I_{2t} frame. The connectivity follows the implied motion vector fields \mathcal{V}^{c-r} and \mathcal{H}^{c-r} , which are simply the directional inverse of the original motion vector fields, \mathcal{V}^{c-r} and \mathcal{H}^{c-r} .

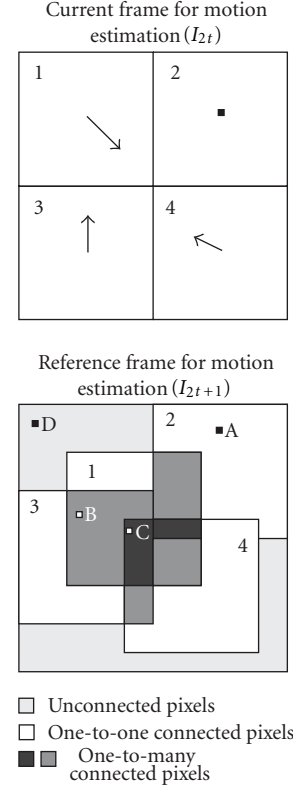


FIGURE 1: Pixel connectivity in I_{2t} and I_{2t+1} frames.

Considering these block and pixel classifications, the lifting steps for pixels at positions $[m, n]$ in frames I_{2t} and I_{2t+1} (i.e., $I_{2t}[m, n]$ and $I_{2t+1}[m, n]$) performing the temporal motion compensated Haar wavelet transform are defined as follows.

Forward Transform

The Prediction Step. For one-to-one connected pixels,

$$I'_{2t+1}[m, n] = I_{2t+1}[m, n] - I_{2t}[m + \mathcal{H}^{c-r}, n + \mathcal{V}^{c-r}]. \quad (3)$$

For one-to-many connected pixels,

$$I'_{2t+1}[m, n] = I_{2t+1}[m, n] - \frac{1}{J} \sum_{i=0}^{J-1} I_{2t}[m + \mathcal{H}_i^{c-r}, n + \mathcal{V}_i^{c-r}], \quad (4)$$

where J is the total number of connections.

For unconnected pixels,

$$I'_{2t+1}[m, n] = I_{2t+1}[m, n]. \quad (5)$$

The last case is similar to the no prediction case as in intrablocks used in conventional MCTF.

The Update Step. For interblocks, every pixel in an interblock is one-to-one connected with a unique pixel in I_{2t+1} . Then, the update step is computed as

$$I_{2t}[m, n] = I_{2t}[m, n] + \frac{1}{2} I'_{2t+1}[m - \mathcal{H}^{c-r}, n - \mathcal{V}^{c-r}]. \quad (6)$$

For intrablocks, as there are no motion compensated connections with I_{2t+1} ,

$$I'_{2t}[m, n] = I_{2t}[m, n]. \quad (7)$$

Finally, these lifting steps are followed by the normalization step:

$$\begin{aligned} I''_{2t}[m, n] &= \sqrt{2} I'_{2t}[m, n], \\ I''_{2t+1}[m, n] &= \frac{1}{\sqrt{2}} I'_{2t+1}[m, n]. \end{aligned} \quad (8)$$

The temporally decomposed frames I''_{2t} and I''_{2t+1} are the first level low- and highpass frames and are denoted as L and H temporal subbands. These steps are repeated for all frames in L to obtain LL and LH subbands and continued to obtain the desired number of temporal decomposition levels.

Inverse Transform

For the inverse transform, the order of operation of steps is reversed as stated follows.

First, the decomposed coefficients are passed through an unnormalization step followed by the inverse lifting steps:

$$\begin{aligned} I'_{2t}[m, n] &= \frac{1}{\sqrt{2}} I''_{2t}[m, n], \\ I'_{2t+1}[m, n] &= \sqrt{2} I''_{2t+1}[m, n]. \end{aligned} \quad (9)$$

The inverse update step:

$$I_{2t}[m, n] = \begin{cases} I'_{2t}[m, n] - \frac{1}{2} I'_{2t+1}[m - \mathcal{H}^{c-r}, n - \mathcal{V}^{c-r}] & \text{for interblocks,} \\ I'_{2t}[m, n] & \text{for intrablocks.} \end{cases} \quad (10)$$

The inverse prediction step:

$$I_{2t+1}[m, n] = \begin{cases} I'_{2t+1}[m, n] + I_{2t}[m + \mathcal{H}^{c-r}, n + \mathcal{V}^{c-r}] & \text{for one-to-one connected pixels,} \\ I'_{2t+1}[m, n] + \frac{1}{J} \sum_{i=0}^{J-1} I_{2t}[m + \mathcal{H}_i^{c-r}, n + \mathcal{V}_i^{c-r}] & \text{for one-to-many connected pixels,} \\ I'_{2t+1}[m, n] & \text{for unconnected pixels.} \end{cases} \quad (11)$$

2.2. 2D + t + 2D Framework. In a 3D video decomposition scheme, t + 2D is achieved by performing temporal decomposition followed by a spatial transform whereas in case of 2D + t, the temporal filtering is done after the spatial 2D transform. Due to its own merit and demerit, it is required to analyze both the combinations in order to enhance the video

watermarking performance. A common flexible reconfigurable framework, which allows creating such possible combinations, is particularly useful for applications like video watermarking. Here, we propose the 2D + t + 2D framework by combining the modified motion compensated temporal filtering with spatial 2D wavelet transformation.

Let $(s_1 t s_2)$ be the number of decomposition levels used in the 2D + t + 2D subband decomposition to obtain a 3D subband decomposition with motion compensated t temporal levels and s spatial levels, where $s = s_1 + s_2$. In such a scheme, first the 2D DWT is applied for an s_1 level decomposition. As a result, a new sequence is formed by the low-frequency spatial LL subband of all frames. Then, the sequences of spatial LL subbands are temporally decomposed using the MCTF with implied motion estimation into t temporal levels. Finally, each of the temporal transformed spatial LL subbands are further spatially decomposed into s_2 wavelet levels.

For a t-s motion compensated temporal subband decomposition, the values of s_1 and s_2 are determined by considering the context of the choice of temporal-spatial subbands used for watermark embedding. From now onwards, in this paper, we will use exact values of s_1, t, s_2 to represent various combinations of spatiotemporal decomposition, that is, $s_1 t s_2$. For example, $s_1 = 0, t = 3, s_2 = 2$ (032), and $s_1 = 2, t = 3, s_2 = 0$ (230) parameter combinations result in t + 2D and 2D + t motion compensated 3D subband decompositions, respectively. The same amount of subband decomposition levels can be obtained by also using the parameter combination $s_1 = 1, t = 3, s_2 = 1$ (131) using the proposed generalized scheme implementation. The combination $s_1 = 0, t = 0, s_2 = 2$ (002) allows 2D decomposition of all frames for frame-by-frame watermark embedding. The realizations of these examples are shown in Figure 2. We use the notation (LLL, LLH, LH, H) to denote the temporal subbands after a 3 level decomposition. We have described the use of this framework in combination with watermarking algorithms, in the next section.

3. Video Watermarking in 2D + t + 2D Spatiotemporal Decomposition

We propose a new video watermarking scheme by extending the wavelet-based image watermarking algorithms into 2D + t + 2D framework. In this section, we briefly revisit the wavelet-based image watermarking algorithms followed by the proposed video watermarking scheme. Then, we carry on to analyze various combinations in the proposed video decomposition framework to decide unique video embedding parameters, such as (1) choice of temporal subband selection and (2) motion estimation parameters, to retrieve the motion information from watermarked video.

3.1. Wavelet-Based Watermarking. Due to its ability for efficient multiresolution spatiofrequency representation of the signals, the DWT became the major transform for spread-spectrum image watermarking [16–22]. A broad classification of such wavelet-based watermarking algorithms can be found in [23]. In this paper, we have chosen commonly

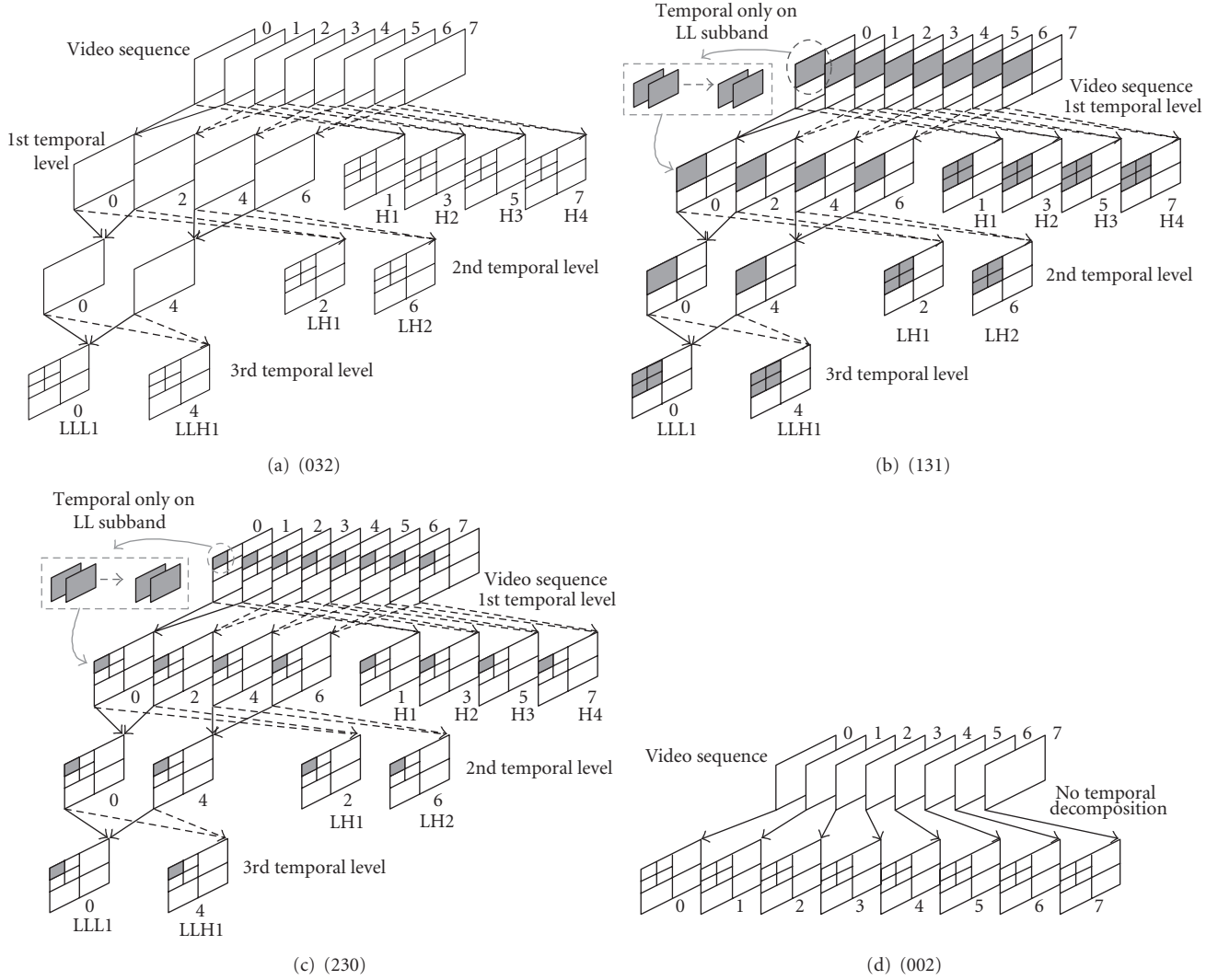


FIGURE 2: Realization of 3-2 temporal schemes using the 2D + t + 2D framework with different parameters: (a) (032), (b) (131), (c) (230), and (d) (002).

used example algorithms to represent nonblind and blind watermarking algorithmic classes.

3.1.1. The Nonblind Case. A magnitude alteration-based additive watermarking is chosen as a nonblind case. In such an algorithm, coefficient values are increased or decreased depending on the magnitude of the coefficient, by making the modified coefficient a function of the original coefficient:

$$C'_{s,t}[m, n] = C_{s,t}[m, n] + \alpha C_{s,t}[m, n] W, \quad (12)$$

where $C_{s,t}[m, n]$ is the original decomposed coefficient at s, t spatiotemporal subband, α is the watermark weighting factor, W is the watermark value to be embedded, and $C'_{s,t}[m, n]$ is the corresponding modified coefficient.

3.1.2. The Blind Case. In this category, we used an example blind watermarking algorithm as proposed in [20, 24], by

modifying various coefficients towards a specific quantization step, δ . The method modifies the median coefficient by using a nonoverlapping 3×1 running window, passed through the entire selected subband of the wavelet decomposed image. At each sliding position, a rank-order sorting is performed on the coefficients C_1, C_2 , and C_3 to obtain an ordered list $C_1 < C_2 < C_3$. The median value C_2 is modified to obtain C'_2 as follows:

$$C'_2 = f(\gamma, C_1, C_3, \delta, W), \quad (13)$$

where W is the input watermark sequence, γ is the weighting parameter, $f()$ denotes a nonlinear transformation, and $\delta = (\gamma(|C_1| + |C_3|)/2)$ is the quantization step.

3.2. The Proposed Video Watermarking Scheme. The new video watermarking scheme uses the above algorithms on spatial-temporal decomposed video. The system block diagrams for watermark embedding, a nonblind extraction

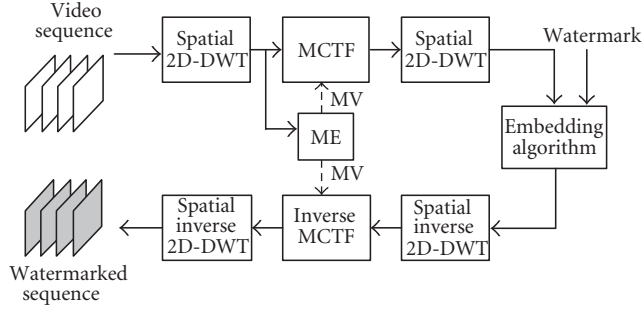


FIGURE 3: System blocks for watermark embedding scheme in $2D + t + 2D$ spatiotemporal decomposition.

process, and a blind extraction process are shown in Figures 3, 4(a), and 4(b), respectively.

3.2.1. Embedding. To embed the watermark, first spatiotemporal decomposition is performed on the host video sequence by applying spatial 2D-DWT followed by temporal MCTF for a $2D + t$ (230) or temporal decomposition followed by spatial transform for a $t + 2D$ (032). In both the cases, the motion estimation (ME) is performed to create the motion vector (MV) either on the spatial domain ($t + 2D$) or on the approximation subband in the frequency domain ($2D + t$) as described in Section 2.2. Other combinations, such as 131 and 002, are achieved in a similar fashion. After obtaining the decomposed coefficients, the watermark is embedded either using nonblind (12) or a blind watermarking algorithm (13) by selecting various temporal low- or highpass frames (i.e., LLL or LLH, etc.) and spatial subband within the selected frame. Once embedded, the coefficients follow inverse process of spatiotemporal decomposition in order to reconstruct the watermarked video.

3.2.2. Extraction and Authentication. The extraction procedure follows a similar decomposition scheme as in embedding and the system diagram for the same is shown in Figure 4. The watermark coefficients are retrieved by applying $2D + t + 2D$ decomposition on watermarked test video. For a nonblind algorithm, the original video sequence is available at the decoder and hence the motion vector is obtained from the original video. After spatiotemporal filtering on test and original video, the coefficients are compared to extract the watermark. In case of a blind watermarking scheme, the motion estimation is performed on the test video itself without any prior knowledge of original motion information. The temporal filtering is then done by using the new motion vector and consequently the spatiotemporal coefficients are obtained for the detection.

The authentication is then done by measuring the Hamming distance (H) between the original and the extracted watermark:

$$H(W, W') = \frac{1}{L} \sum_{i=0}^{L-1} W_i \oplus W'_i, \quad (14)$$

where W and W' are the original and the extracted watermarks, respectively. L is the length of the sequence and \oplus represents the XOR operation between the respective bits.

4. The Framework Analysis in Video Watermarking Context

Before approaching to the experimental results, in this section, we aim to address the issues related to MCTF-based video watermarking of the proposed framework. Firstly, to improve the imperceptibility, an investigation is made about the energy distribution of the host video in different temporal subbands, which is useful to select the temporally decomposed frames during embedding. Then, an insight is given to motion retrieval for a blind watermarking scheme, where no prior motion information is available during watermark extraction and this is crucial for the robustness performance.

4.1. On Improving Imperceptibility. In wavelet domain watermarking research, it is a well-known fact that embedding in high-frequency subbands offers better imperceptibility and low-frequency embedding provides higher robustness. Often wavelet decompositions compact most of the energy in low-frequency subbands and leave less energy in high-frequencies and due to this reason, high-frequency watermarking schemes are less robust to compression. Therefore, increase in energy distribution in high-frequency subbands can offer a better watermarking algorithm.

In analyzing our framework, the research findings show that different $2D + t + 2D$ combinations can vary the energy distribution in high-frequency temporal subbands and this is independent of video content. To show an example, we used *Foreman* sequence and decomposed using 032, 131, and 230 combinations in the framework and calculate the sum of energy for first two GOP each with 8 temporal frequency frames, namely, LLL, LLH, LH1, LH2, H1, H2, H3, and H4. In all cases, we calculate the energy for the low-frequency (LL_s) subband of spatial decomposition. Other input parameters are set to 8×8 macroblock, a fixed-size block matching motion estimation with ± 16 search window. The results for percentage of energy (of a GOP) in each temporally decomposed frame are shown in Figure 5 and the histograms of the coefficients for 032, 131, and 230 of LLL and LLH are shown in Figure 6. The inner graph in Figure 6 represents the zoomed version of the local variations by clipping the y -axis to show the coefficient distribution more effectively. From the results, we can rank the energy distribution in high-frequency temporal subbands as: (230) > (131) > (032). This analysis guides us to select optimum spatiotemporal parameter in the framework to improve the robustness while keeping better imperceptibility. We have performed the experimental simulation on 8 test videos: (*Foreman*, *Crew*, *News*, *Stefan*, *Mobile*, *City*, *Football*, and *Flower garden*) and all of them follows a similar trend.

4.2. On Motion Retrieval. In an MCTF-based video watermarking scheme, motion information contributes at large for

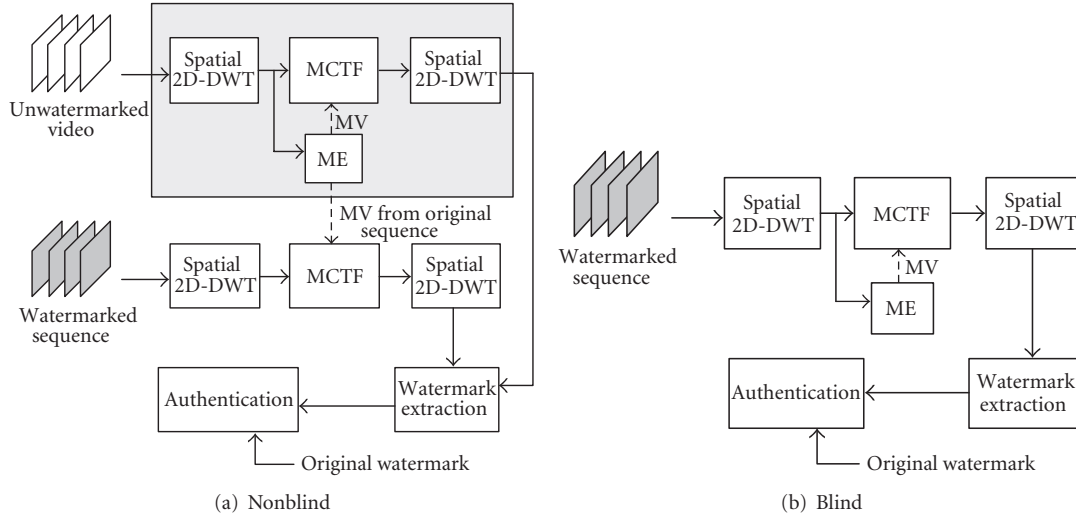


FIGURE 4: System blocks for watermark extraction scheme in 2D + t + 2D spatiotemporal decomposition.

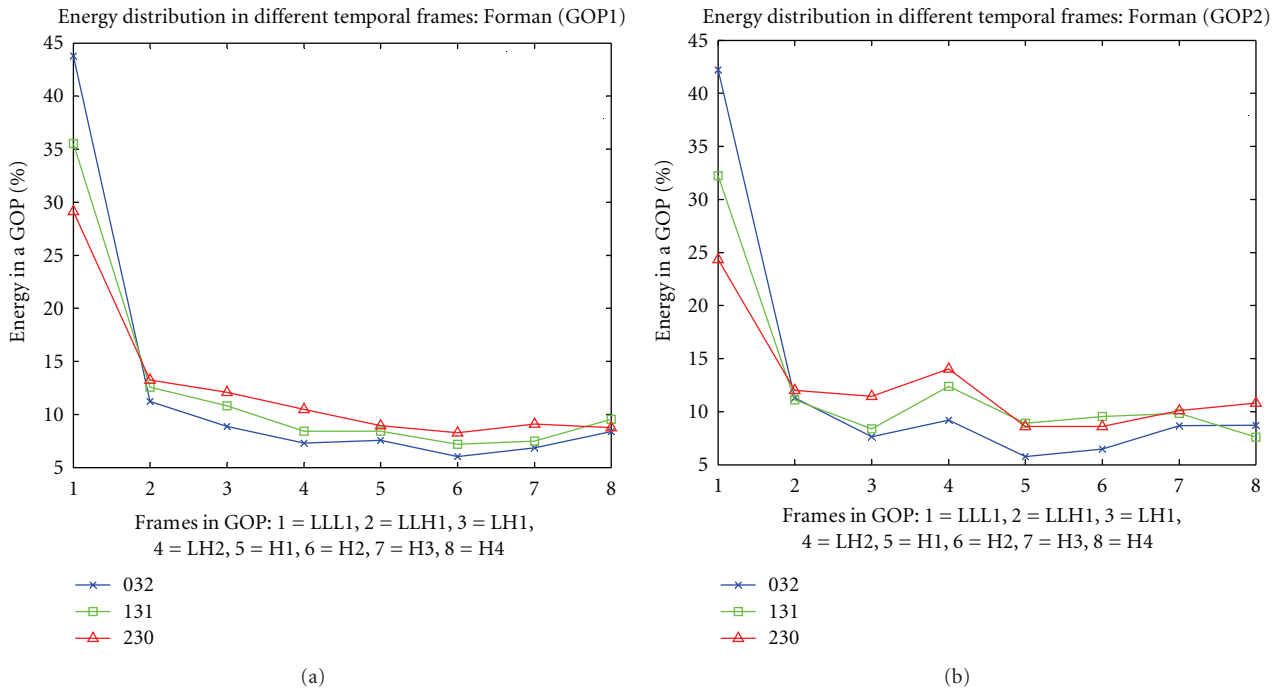


FIGURE 5: Percentage of energy (of a GOP) in each temporally decomposed frame. Energy calculation considers the energy of the coefficients at LL_s for first two GOP each with 8 temporal low and high frequency frames of *Foreman* sequence. (a) GOP1 and (b) GOP2.

temporal decomposition along motion trajectory. The watermark embedding modification in the temporal domain causes motion mismatch, which affects the decoder performance. While original motion information is available for a nonblind watermarking scheme, motion estimation must be done in the case of a blind video watermarking scheme. In this case, the motion vectors are expected to be retrieved from the watermarked video without any prior knowledge of the original motion vector (MV). Our study shows that, in such a case, more accurate motion estimation is possible by choosing the right 2D + t + 2D combination along with an

optimum choice of macro block (MB) size. At the same time, we investigate the performance, based on motion search range (SR). Experimental performance shows that effectively SR has lesser contribution towards motion retrieval. The experiment set is organized by studying the watermarking detection performance by measuring Hamming distance of a blind watermark embedding at LL_s spatial subband on LLL and LLH temporal frames. The watermark extraction is done by using various combinations of MB and SR to find the best motion retrieval parameters. The results are shown in Tables 1 and 2 using average of the first 64 frames from *Foreman* and

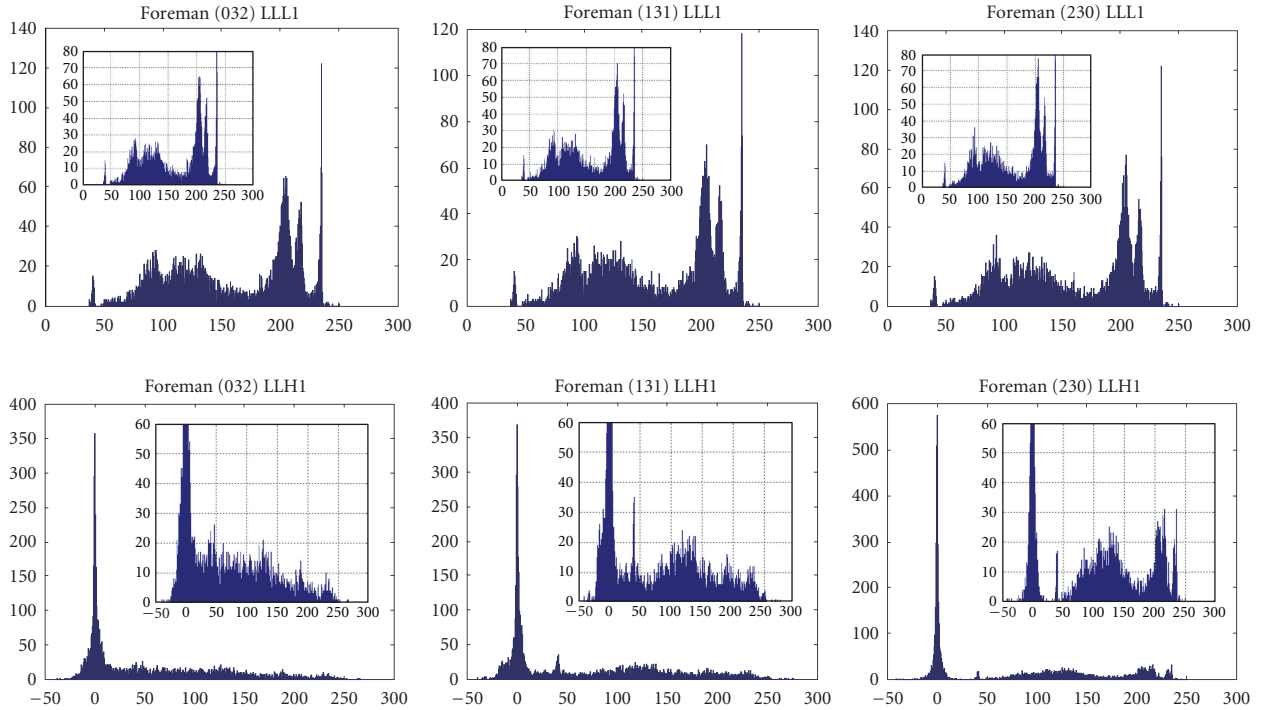


FIGURE 6: Histogram of coefficients at LL_s for 3rd level temporal low and high-frequency frames (GOP 1) for *Foreman* sequence. Row (1) and (2) represents LLL and LLH temporal frames, respectively, and Column (1), (2) and (3) shows 032, 131, and 230 combinations of $2D + t + 2D$ framework.

TABLE 1: Hamming distance for blind watermarking by estimating motion from watermarked video using different macro block size (MB) and search range (SR). Embedding at LL_s on frame: (a) LLL and (b) LLH on *Foreman* sequence (average of first 64 frames).

		(a) LLL						
		MV from watermarked video: MB/SR						
		32×32	16×16	16×16	8×8	8×8	4×4	4×4
		$/\pm 64$	$/\pm 64$	$/\pm 32$	$/\pm 32$	$/\pm 16$	$/\pm 16$	$/\pm 8$
032		0.02	0.03	0.02	0.03	0.03	0.04	0.04
131		—	0.02	0.03	0.03	0.03	0.08	0.07
230		—	—	—	0.03	0.03	0.08	0.07
		(b) LLH						
		MV from watermarked video: MB/SR						
		32×32	16×16	16×16	8×8	8×8	4×4	4×4
		$/\pm 64$	$/\pm 64$	$/\pm 32$	$/\pm 32$	$/\pm 16$	$/\pm 16$	$/\pm 8$
032		0.15	0.29	0.29	0.40	0.39	0.49	0.49
131		—	0.22	0.21	0.29	0.28	0.44	0.44
230		—	—	—	0.23	0.22	0.30	0.30

Crew CIF size video sequences, respectively, for 032, 131, and 230 spatiotemporal decompositions. Due to the limitations in macroblock size and integer pixel motion search, 32×32 MB search is excluded for 131 decomposition and 32×32 , 16×16 MB searches are excluded for 230 decomposition. It is noted that, in the video compression schemes, 16×16 is the most commonly used MB while in this paper we have used

various other MB sizes to investigate the effect on watermark retrieval.

The results show that for an MB size more than 8×8 , $2D + t$ outperform $t + 2D$. In this context, the spatiotemporal decompositions can be ranked as $(230) > (131) > (032)$. In the case of 131 or 230, the motion is estimated in hierarchically downsampled low-frequency subband. Therefore,

TABLE 2: Hamming distance for blind watermarking by estimating motion from watermarked video using different macro block size (MB) and search range (SR). Embedding at LL_s on frame: (a) LLL and (b) LLH on *Crew* sequence (average of first 64 frames).

(a) LLL							
MV from watermarked video: MB/SR							
	32×32	16×16	16×16	8×8	8×8	4×4	4×4
	$/\pm 64$	$/\pm 64$	$/\pm 32$	$/\pm 32$	$/\pm 16$	$/\pm 16$	$/\pm 8$
032	0.03	0.06	0.05	0.09	0.09	0.09	0.09
131	—	0.03	0.03	0.07	0.07	0.14	0.13
230	—	—	—	0.03	0.03	0.15	0.12

(b) LLH							
MV from watermarked video: MB/SR							
	32×32	16×16	16×16	8×8	8×8	4×4	4×4
	$/\pm 64$	$/\pm 64$	$/\pm 32$	$/\pm 32$	$/\pm 16$	$/\pm 16$	$/\pm 8$
032	0.17	0.24	0.23	0.36	0.36	0.48	0.47
131	—	0.16	0.16	0.23	0.23	0.41	0.38
230	—	—	—	0.17	0.17	0.28	0.27

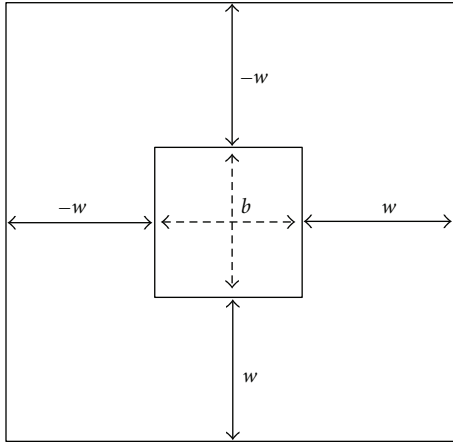


FIGURE 7: Exhaustive search complexity for a motion block.

number of motion vector reduces accordingly for a given macroblock size. This offers two-fold advantages.

(1) *Complexity.* The search range during the motion estimation is either half or quarter size of the full-resolution motion estimation. As a result, the searching time and computation complexity reduces significantly as follows: Let us assume motion is estimated for MB of $b \times b$ with SR $w \times w$ as shown in Figure 7. The complexity, \mathcal{O} , is calculated based on the number of search operations as given in (15):

$$\mathcal{O} = T(2w + 1)^2, \tag{15}$$

where $T = MN$ is total number of pixels. As motion is estimated only on the downsampled low-frequency component, we can rewrite (15) as

$$\mathcal{O} = \frac{MN}{s_1 s_1} (2w + 1)^2, \tag{16}$$

where s_1 is the 1st spatial decomposition in the proposed scheme. Now, SR $w \times w$ is a constant considering any given column in Tables 1 and 2 and hence it is evident that the complexity is inversely proportional to s_1^2 :

$$\begin{aligned} \mathcal{O} &\propto \frac{1}{s_1^2} (2w + 1)^2, \\ &\propto \frac{1}{s_1^2}. \end{aligned} \tag{17}$$

Therefore, the complexity of various spatiotemporal decomposition can be ranked as $(230) < (131) < (032)$, that is, complexity of proposed $2D + t$ scheme is much lesser when compared to traditional $t + 2D$.

(2) *MV Error Reduction.* At the same time, for blind motion estimation, less number of motion vector needs to be estimated at the decoder resulting in more accurate motion estimation and higher robustness. It is evident from Tables 1 and 2 that if the same number of motion vectors are considered, that is, 32×32 MB for 032, 16×16 MB for 131, and 8×8 MB for 230, the robustness performance is comparable for all three combinations. However, in LLL subband of $2D + t$, for a smaller MB, such as 4×4 , more motion mismatch is observed as motion estimation is done in a spatially decomposed region. Now, using the analysis, above, we have designed experiments to verify our proposed video watermarking schemes for improved imperceptibility as well as robustness against scalable video compressions.

5. Experimental Results and Discussion

We used the following experimental setup for the simulation of watermark embedding using the proposed generalized $2D + t + 2D$ motion compensated temporal-spatial subband scheme. In order to make the watermarking strength constant across subbands, the normalization steps in the MCTF

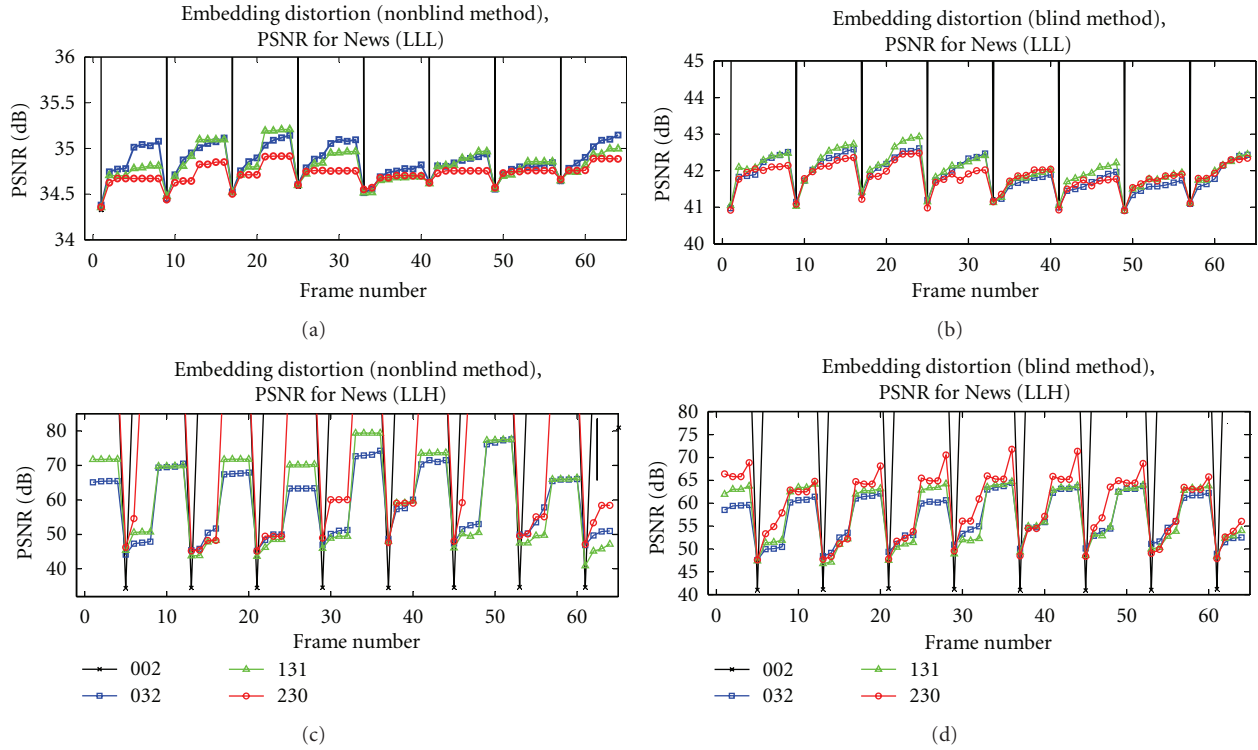


FIGURE 8: PSNR for nonblind and blind watermarking on LLL and LLH temporal subbands for *News* sequence. (a) LLL (nonblind), (b) LLL (blind), (c) LLH (nonblind), and (d) LLH (blind).

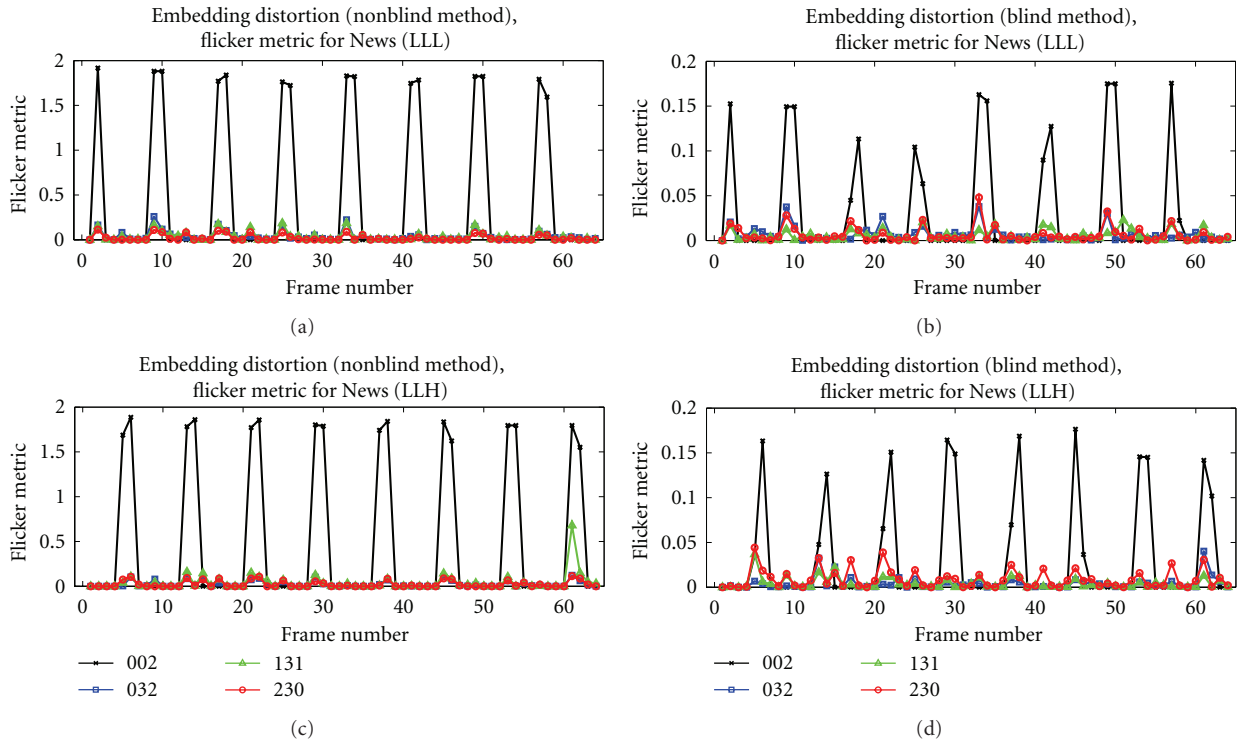


FIGURE 9: Flicker metric for nonblind and blind watermarking on LLL and LLH temporal subbands for *News* sequence. (a) LLL (nonblind), (b) LLL (blind), (c) LLH (nonblind), and (d) LLH (blind).

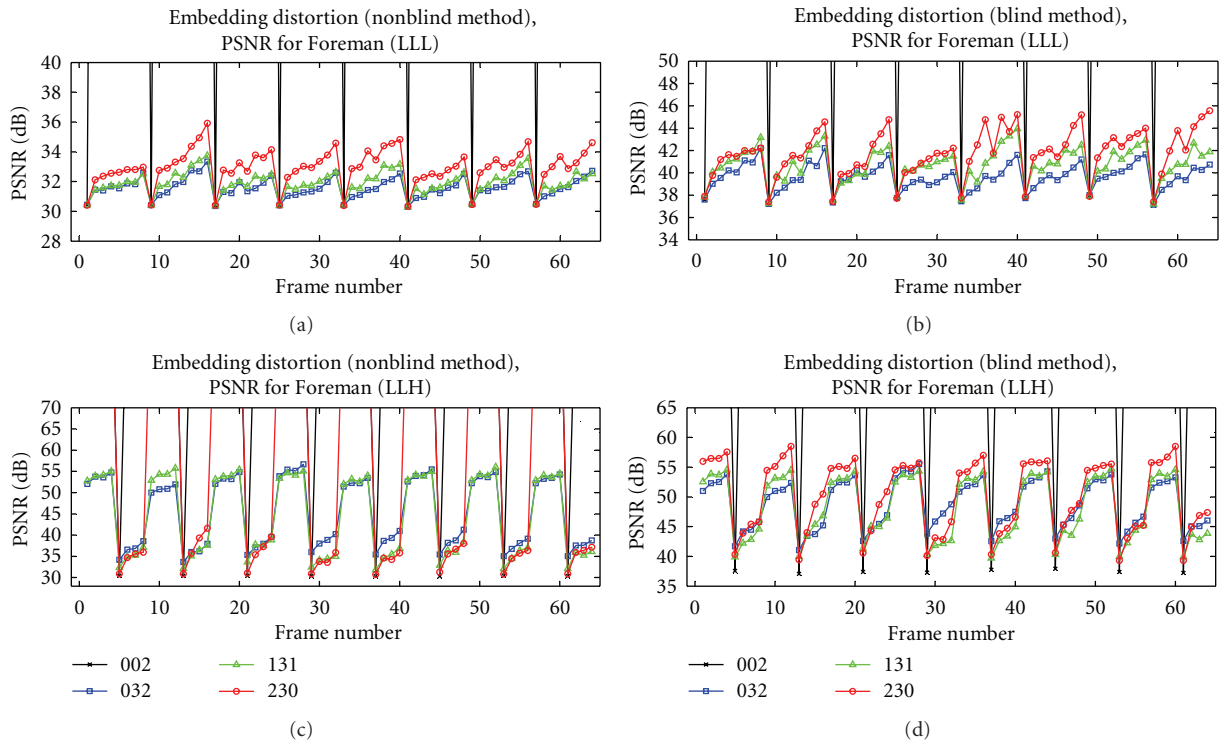


FIGURE 10: PSNR for nonblind and blind watermarking on LLL and LLH temporal subbands for *Foreman* sequence. (a) LLL (nonblind), (b) LLL (blind), (c) LLH (nonblind), and (d) LLH (blind).

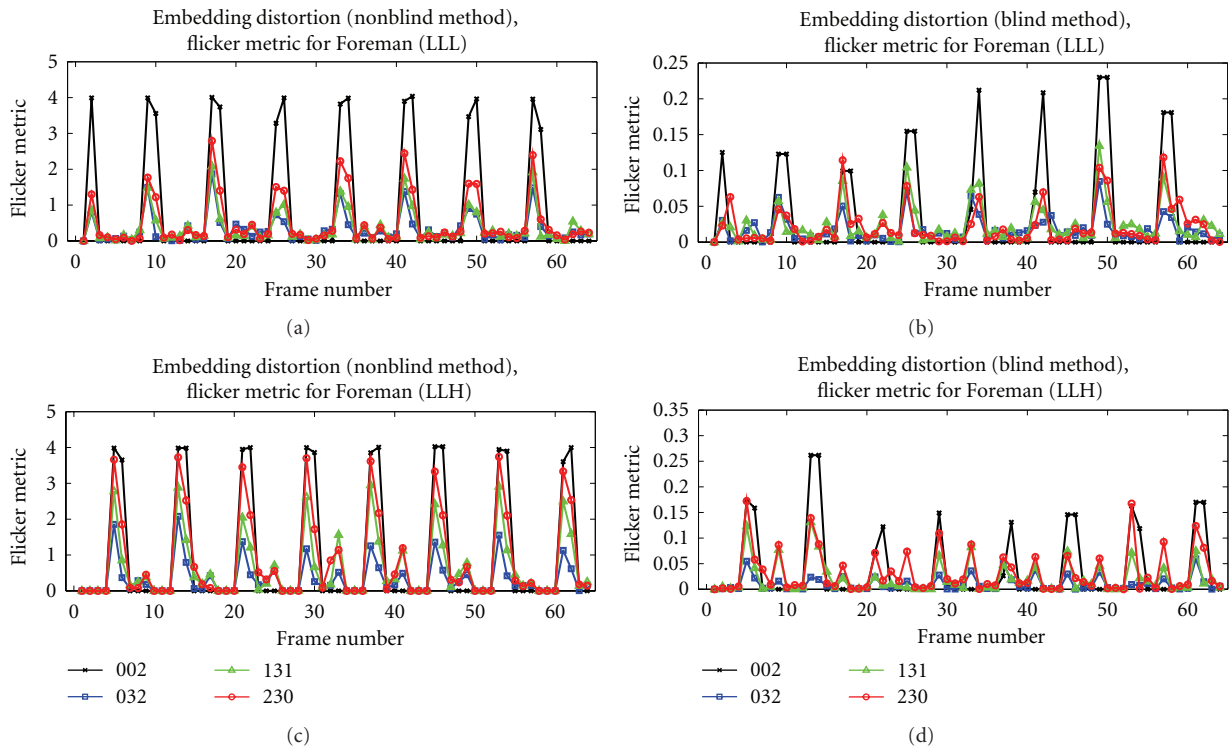


FIGURE 11: Flicker metric for nonblind and blind watermarking on LLL and LLH temporal subbands for *Foreman* sequence. (a) LLL (nonblind), (b) LLL (blind), (c) LLH (nonblind), and (d) LLH (blind).

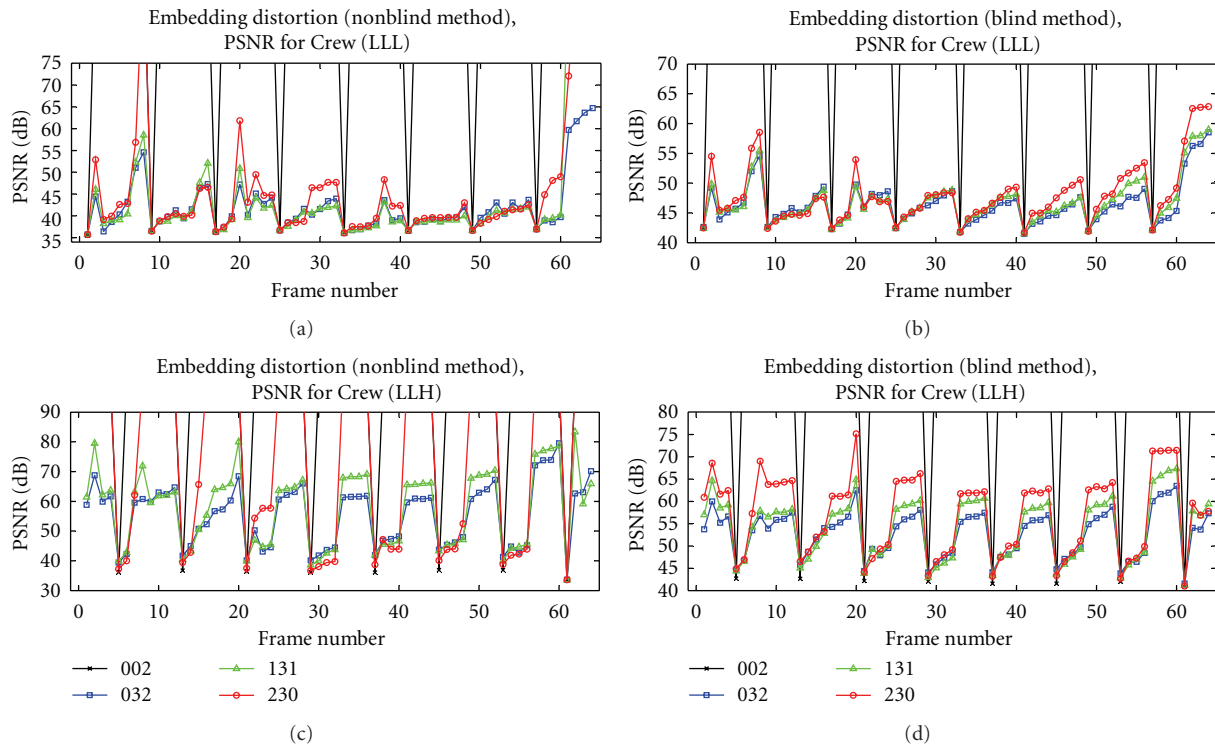


FIGURE 12: PSNR for nonblind and blind watermarking on LLL and LLH temporal subbands for *Crew* sequence. (a) LLL (nonblind), (b) LLL (blind), (c) LLH (nonblind), and (d) LLH (blind).

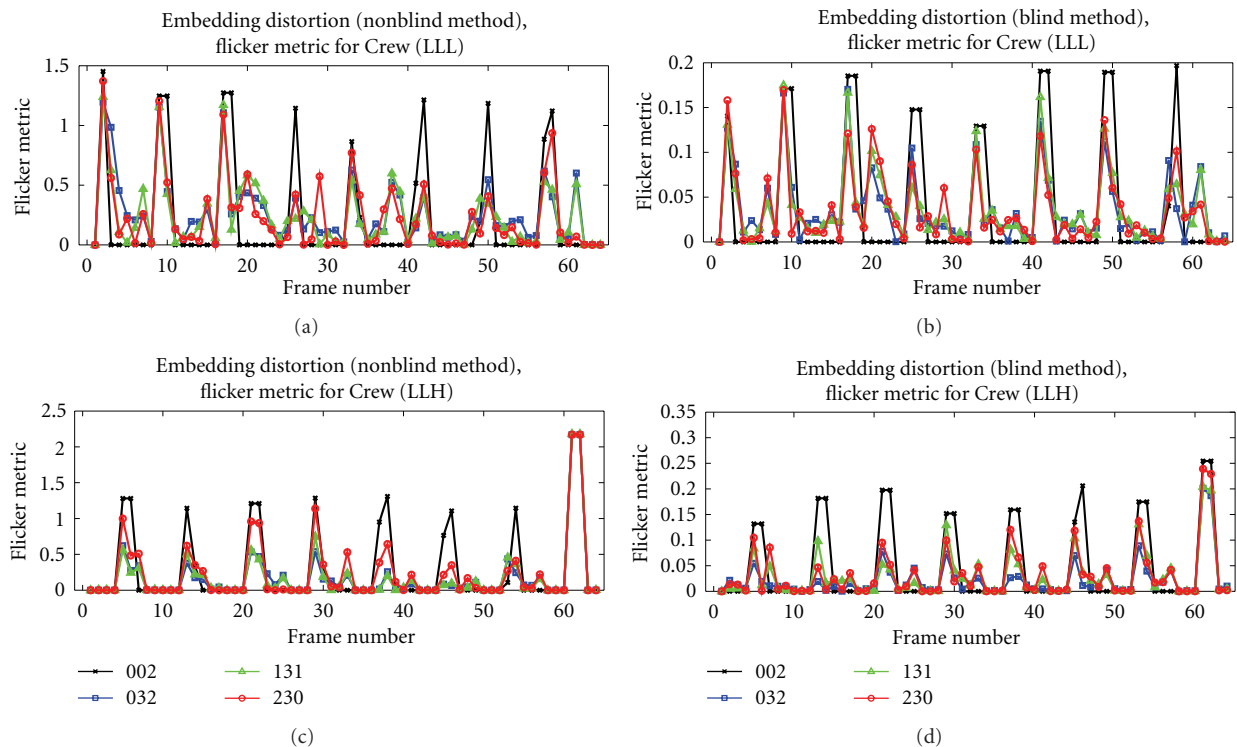


FIGURE 13: Flicker metric for nonblind and blind watermarking on LLL and LLH temporal subbands for *Crew* sequence. (a) LLL (nonblind), (b) LLL (blind), (c) LLH (nonblind), and (d) LLH (blind).

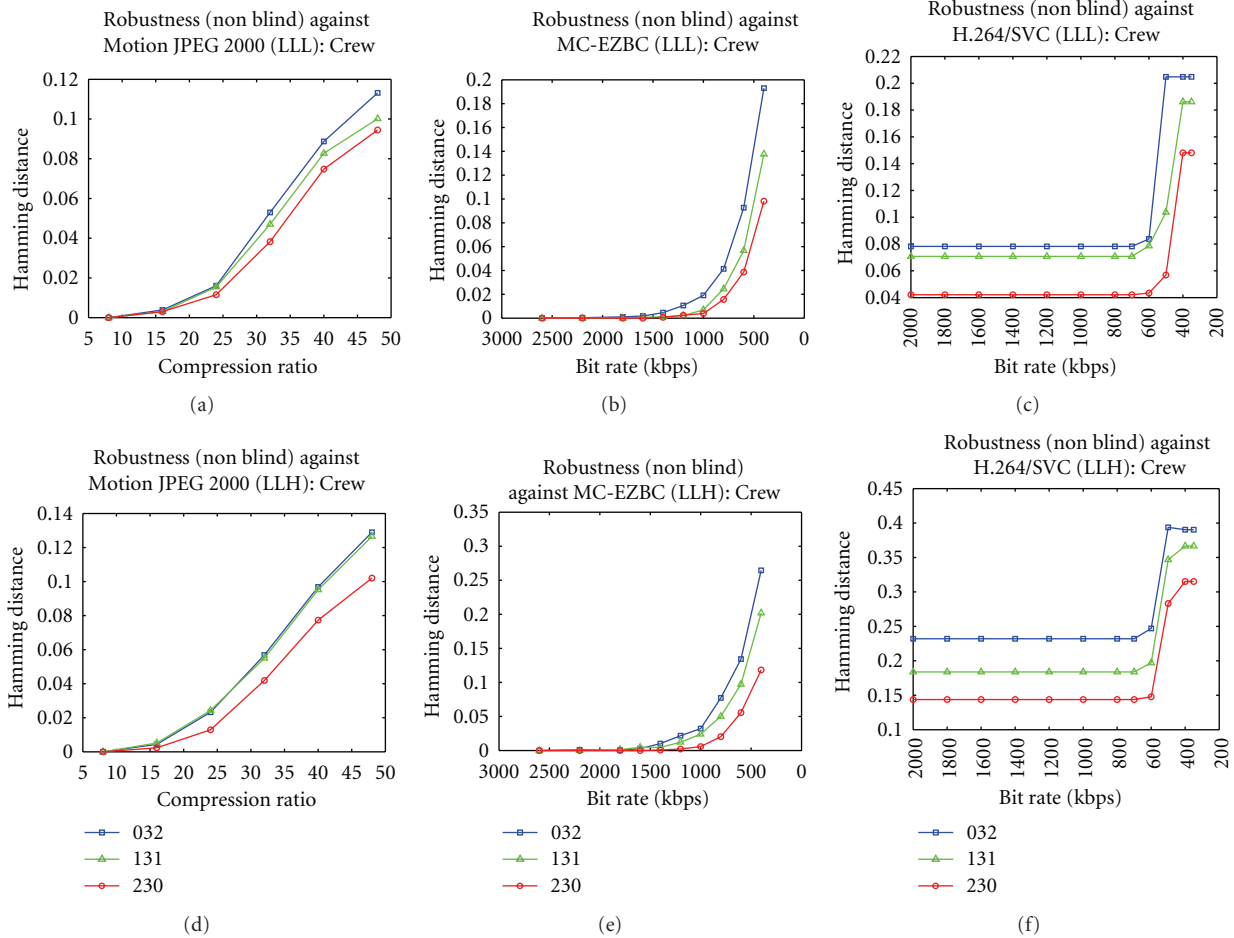


FIGURE 14: Robustness performance of nonblind watermarking scheme for Crew sequence. Columns (1), (2) and (3) show robustness against Motion JPEG 2000, MC-EZBC, and H.264-SVC, respectively. Rows (1) and (2) represent the embedding on temporal subbands LLL & LLH, respectively.

and the 2D DWT were omitted. There are two different sets of results obtained to show the embedding distortion and the robustness performance using luma component of 8 test video sequences (4 : 2 : 0 YUV): (*Foreman, Crew, News, Stefan, Mobile, City, Football, and Flower garden*). However, within the scope of this paper, three test sequences are chosen to show the results according to their object motion activity, that is, high-motion activity (*Crew*), medium-motion activity (*Foreman*), and low-motion activity (*News*). We have used one nonblind and one blind watermarking scheme as example cases, described in Section 3.1. For the simulations shown in this work, the four combinations (032), (230), (131), and (002) were used. In each case, the watermark embedding is performed on the low-frequency subband (LL_s) of 2D spatial decompositions due to its improved robustness performance against compression attacks in image watermarking. In these simulations, the 9/7 biorthogonal wavelet transform was used as the 2D decompositions.

Based on the analysis in the previous section, here we explored the possibility of watermark embedding in high-frequency temporal subband and investigate the robustness performance against compression attacks, as high-frequency

subband can offer improved imperceptibility. In the experiment sets, we chose third temporal level highpass (LLH) and lowpass (LLL) frames to embed the watermark. Other video decomposition parameters are set to (1) 64 frames with GOP size of 8, (2) 8×8 macro block size, and (3) a search window of ± 16 . The choice of macro block size and search window are decided by referring the motion retrieval analysis in Section 4.2.

For embedding distortion measure, we used peak signal to noise ratio (PSNR) and also measured the amount of flicker introduced due to watermark embedding. Fan et al. [25] defined a quality metric to measure flicker in intracoded video sequences. In our experiments, we have measured flicker in a similar way by calculating the difference between average brightness values of previous and current frames and used the flicker metric in the MSU quality measurement tool [26]. The flicker metric here compares the flicker content in the watermarked video with respect to the original video. In these metrics, higher PSNR represents lower embedding distortion and for flicker, the lower values correspond to the better distortion performance. On the other hand, the watermarking robustness is represented by Hamming

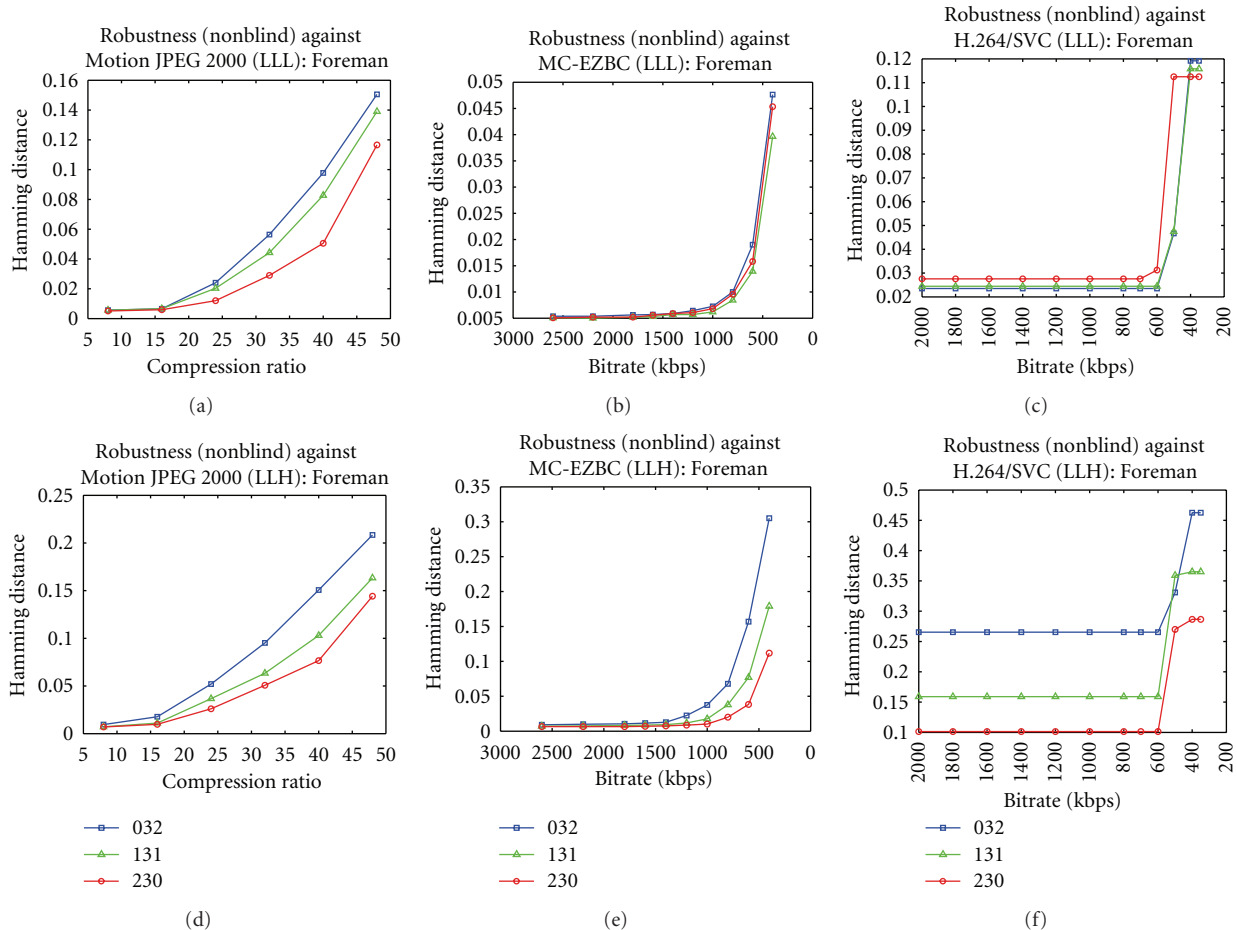


FIGURE 15: Robustness performance of nonblind watermarking scheme for *Foreman* sequence. Columns (1), (2) and (3) show robustness against Motion JPEG 2000, MC-EZBC, and H.264-SVC, respectively. Rows (1) and (2) represent the embedding on temporal subbands LLL & LLH, respectively.

distance as mentioned in (14) and lower Hamming distance corresponds to better detection performance. Various scalable coded quality compression attacks are considered, such as Motion JPEG 2000, MC-EZBC scalable video coding, and H.264/AVC scalable extension (H.264-SVC). In these experiments, low-frequency spatial LL subband are selected within LLL and LLH temporal subbands. Therefore, the scheme is robust against respective spatial and temporal scalability. For example, the algorithm is robust against spatial scalability up to quarter resolution and temporal scaling up to LH and H frames. The results show the mean value of Hamming distance for average of first 64 frames of test video set.

The experiments are divided into two sets, one for embedding distortion analysis and the other for robustness evaluation. In all the experimental setup, we considered two watermarking algorithms, one each from nonblind (Section 3.1.1) and blind (Section 3.1.2) category. The weighting parameters α and γ are set to 0.1. In case of nonblind algorithm, the level adaptive thresholding as described in [22] is taken into account to avoid watermark embedding in small or nearly zero coefficients to minimize the false detection. The watermarking payload is set to 2000

bits and 2112 bits using a binary logo for all combinations and every sequences for nonblind and blind watermarking methods, respectively.

5.1. Embedding Distortion Analysis. The embedding distortion results in terms of PSNR are shown in Figures 8, 10, and 12 for *News*, *Foreman*, and *Crew* sequences, respectively, for nonblind and blind watermarking methods. In each of the figures, x -axis shows the frame number and y -axis represents the PSNR. The flickering results are shown in Figures 9, 11, and 13 for *News*, *Foreman*, and *Crew* sequences, respectively. In these figures, the y -axis represents the flicker metric as discussed in the previous section.

From the results for LLL subband, it is evident that although the PSNR performances are comparable, proposed MCTF-based methods ((032), (131), and (230)) outperform the frame-by-frame embedding (002) in addressing the flickering problem. In all four combinations, the sum of energy in LLL subband are similar and resulting in comparable PSNR. However, in the proposed methods, the error (i.e., PSNR) is propagated along the GOP due to hierarchical temporal decomposition along the motion trajectory and the error

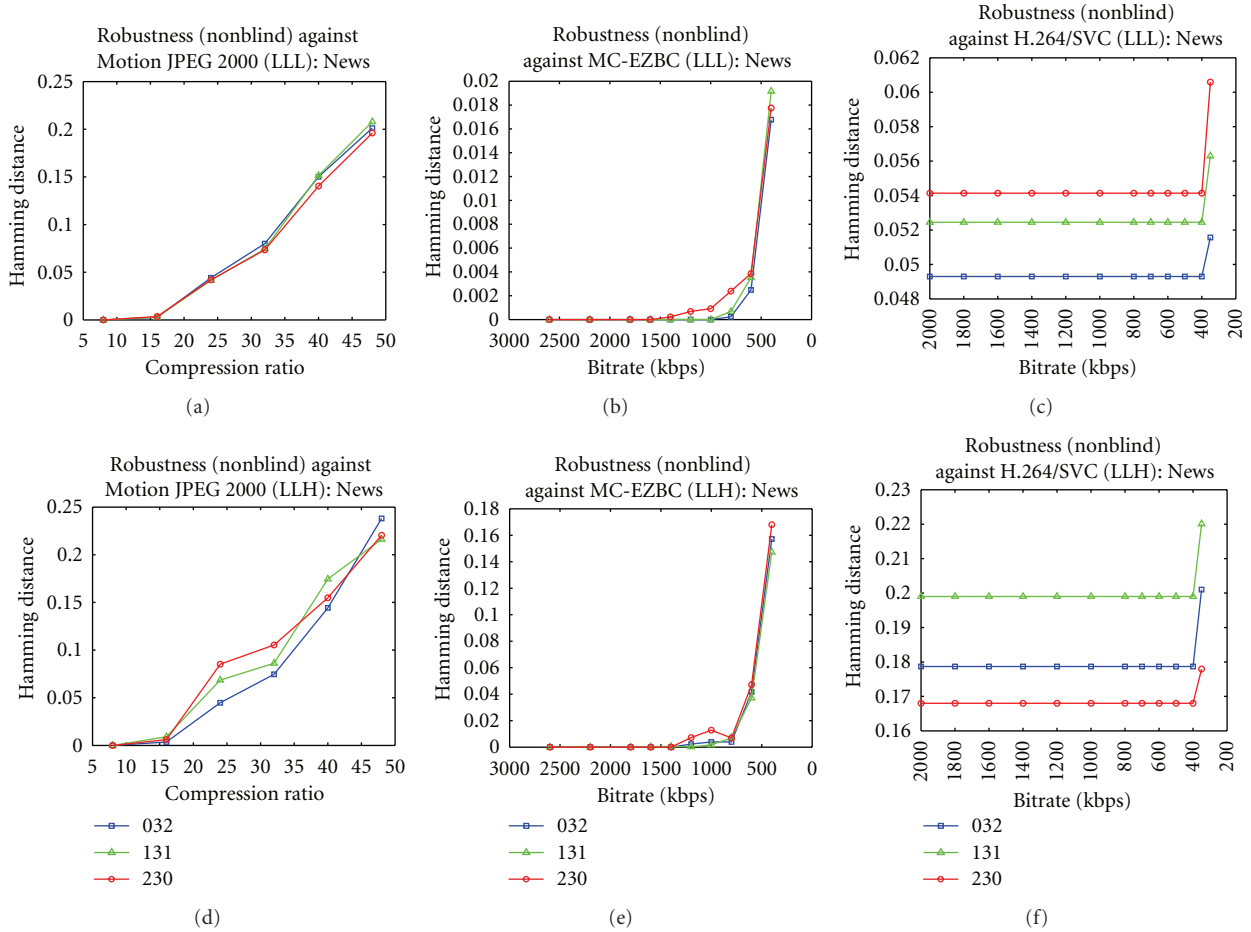


FIGURE 16: Robustness performance of non-blind watermarking scheme for *News* sequence. *Columns* (1), (2) and (3) show robustness against Motion JPEG 2000, MC-EZBC, and H.264-SVC, respectively. *Rows* (1) and (2) represent the embedding on temporal subbands LLL and LLH, respectively.

propagation along the motion trajectory addressed the issues related to flickering artifacts. On the other hand, for LLH subband, due to temporal filtering, the sum of energy is lesser and the four combinations can be ranked as $032 < 131 < 230 < 002$. Hence, the PSNR and flickering performance for this temporal subband can be ranked as $032 > 131 > 230 > 002$. Therefore, while choosing a temporally filtered high-frequency subband, such as LLH, LH, or *H*, the proposed MCTF approach also outperforms the frame-by-frame embedding in terms of PSNR while addressing the flickering issues. It is evident that flickering due to frame-by-frame embedding is increasingly prominent in the sequences with lower motion (e.g., *News* > *Foreman* > *Crew*) and is successfully addressed by the proposed MCTF-based watermarking approach.

5.2. Robustness Performance Evaluation. The robustness results for the nonblind watermarking method are shown in Figures 14, 15, and 16 for *Crew*, *Foreman*, and *News* sequences, respectively. The *x*-axis represents the compression ratio (Motion JPEG 2000) or bitrates (MC-EZBC and H.264-SVC) and *y*-axis shows the corresponding Hamming

distances. *Columns* (1) and (2) show the results for the LLL and LLH frame selections, respectively. The robustness performances shows that 2D + t, for example, any combination of temporal filtering on spatial decomposition (i.e., (131) and (230)) outperforms a conventional t + 2D scheme. The experimental robustness results for blind watermarking method are shown in Figures 17, 18, and 19 for *Crew*, *Foreman*, and *News* sequences, respectively. *Column 1* shows results for the LLL temporal subband while results for LLH are shown in *Column 2*. The rows represent various scalability attacks, Motion JPEG 2000, MC-EZBC, and H.264-SVC, respectively. In this case, the motion information is obtained from the watermarked test video. Similar to the nonblind watermarking 2D + t again outperforms a conventional t + 2D scheme such as in [14]. We now analyze the obtained results by grouping it by selection of temporal subband, by embedding method, and by compression scheme.

5.2.1. Selection of Temporal Subband. The low-frequency temporal subband (LLL) offers higher robustness in comparison to high-frequency LLH subband. This is due to more

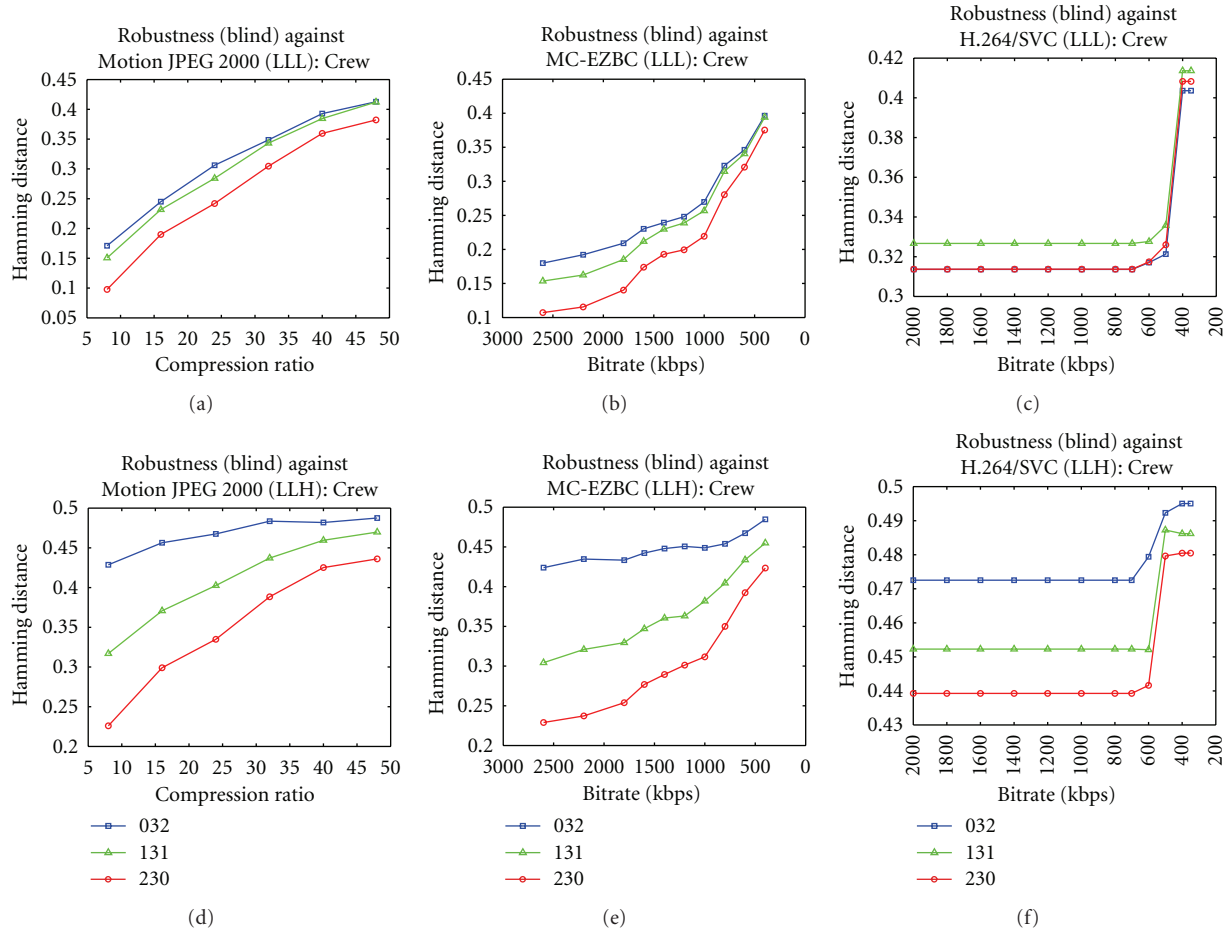


FIGURE 17: Robustness performance of blind watermarking scheme for *Crew* sequence. Columns (1), (2) and (3) show robustness against Motion JPEG 2000, MC-EZBC, and H.264-SVC, respectively. Rows (1) and (2) represent the embedding on temporal subbands LLL and LLH, respectively.

energy concentration in LLL subband after temporal filtering. Within the temporal subbands, in LLL subband, various spatiotemporal combinations perform equally as the energy levels are nearly equal for 032, 131, and 230. However, 230 performs slightly better due to lesser motion-related error in spatially scaled subband. On the other hand, for LLH subband, we can rank the robustness performance as $230 > 131 > 032$ as a result of the energy distribution ranking of these combinations in Section 4.1.

5.2.2. Embedding Method. For a nonblind case, the watermark extraction is performed using the original host video and hence the original motion vector is available at the extractor which makes this scheme more robust to various scalable content adaptation. On the other hand, as explained before, the blind watermarking scheme neither have any reference to original video sequence nor any reference motion vector. The motion vector is estimated from the watermarked test video itself which results in comparatively poor robustness. The effect of motion related error is more visible in LLH subband as the motion compensated temporal highpass frame is highly sensitive to motion estimation accuracy and

so the robustness performance. As discussed in Section 4.2 in case of a $2D + t$ (i.e., 230), the error due to motion vector is lesser compared to $t + 2D$ scheme and hence offers better robustness ($230 > 131 > 032$).

5.2.3. Compression Scheme. We have evaluated our proposed algorithm against various scalable video compression schemes, that is, Motion JPEG 2000, MC-EZBC, and H.264-SVC. First two video compression schemes are based on wavelet technology whereas more recent H.264-SVC uses layered scalability using base layer coding of H.264/AVC.

In Motion JPEG 2000 scheme, the coding is performed by applying 2D wavelet transform on each frame separately without considering any temporal correlation between frames. In the proposed watermarking scheme, the use of 2D wavelet transform offers better association with Motion JPEG 2000 scheme and hence provides better robustness for $2D + t$ combination for LLL and LLH. Also in the case of LLH subband, a better energy concentration offers higher robustness to Motion JPEG 2000 attacks. The robustness performance against Motion JPEG 2000 can be ranked as $230 > 131 > 032$.

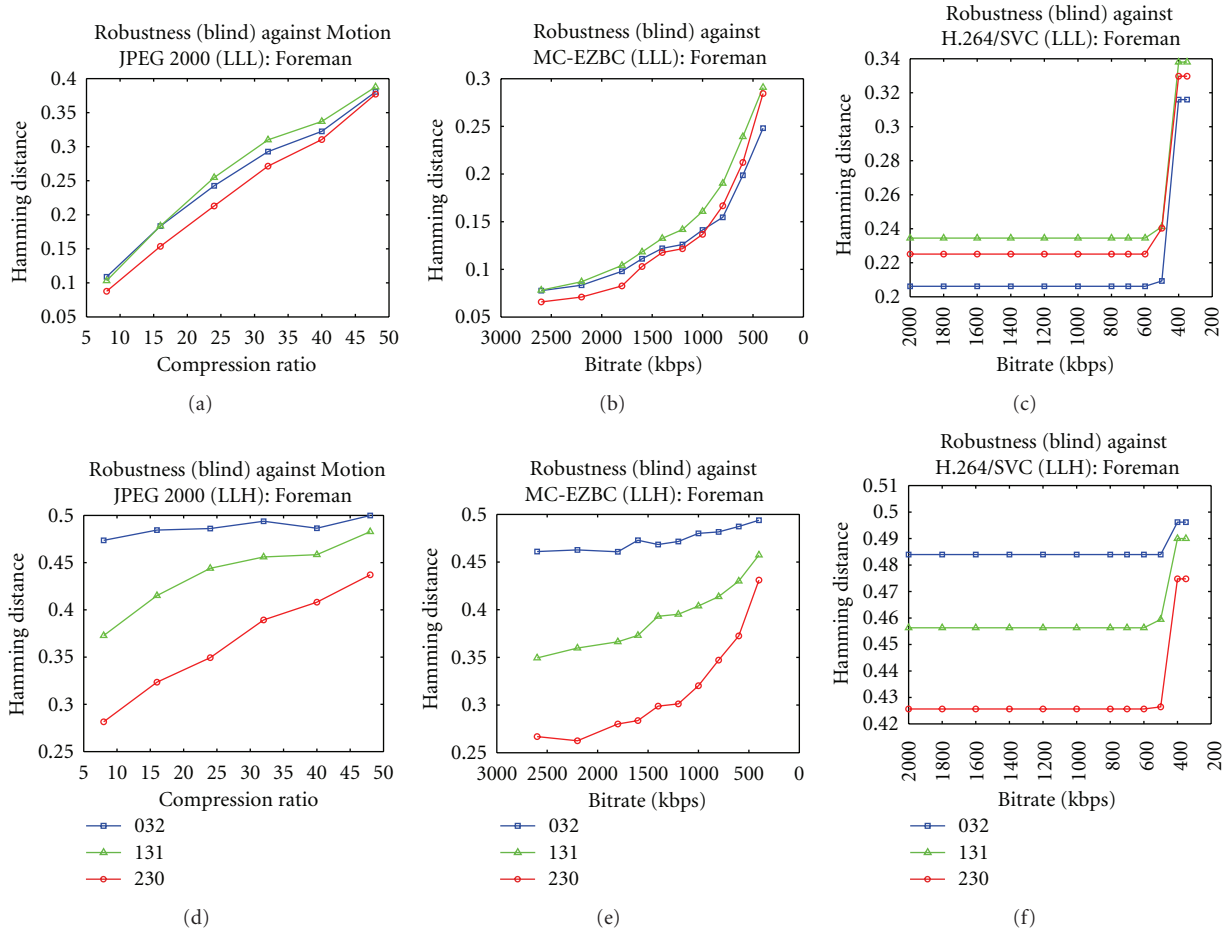


FIGURE 18: Robustness performance of blind watermarking scheme for *Foreman* sequence. *Columns* (1), (2) and (3) show robustness against Motion JPEG 2000, MC-EZBC, and H.264-SVC, respectively. *Rows* (1) and (2) represent the embedding on temporal subbands LLL and LLH, respectively.

MC-EZBC video coder uses motion compensated 1D wavelet transform in temporal filtering and 2D wavelet transform in spatial decomposition. In compression point of view, MC-EZBC usually encodes the video sequences in t+2D combination due to better energy compaction in low-frequency temporal frames. But in watermarking perspective, higher energy in high-frequency subband can offer higher robustness. The argument is justified from the robustness results where results for LLL subbands are comparable, but a distinctive improvement is observed in LLH subband and based on the results the robustness ranking for MC-EZBC can be done as 230 > 131 > 032.

Finally, we have evaluated the robustness of the proposed scheme against H.264-SVC, which uses inter- /intra motion compensated prediction followed by an integer transform with similar properties of DCT transform. Although the proposed watermarking and H.264-SVC video coding scheme do not share any common technology or transform, the robustness evaluation of the proposed method, against H.264-SVC, has been carried for the completeness of the paper for different scalable video compression schemes. The results provide acceptable robustness. However, for a blind watermarking scheme in LLH subband, proposed schemes

performs poorly due to blind motion estimation. Similar to previous robustness results, based on energy distribution and motion retrieval argument, here we can rank the spatiotemporal combinations as 230 > 131 > 032. In a specific example case, H.264-SVC usually gives preference to intraprediction to the sequences with low global or local motion, as in *News* sequence and hence exception in robustness performance to H.264-SVC is noticed for the proposed scheme.

It is evident that, due to close association between the proposed scheme and MC-EZBC, robustness of the proposed scheme offers best performance against MC-EZBC-based content adaptation. To conclude this discussion, we suggest that a choice of 2D + t watermarking scheme improves the imperceptibility and the robustness performance in a video watermarking scenario for a nonblind as well as a blind watermarking algorithm.

6. Conclusions

In this paper, we have presented a new motion compensated temporal-spatial subband decomposition scheme, based on

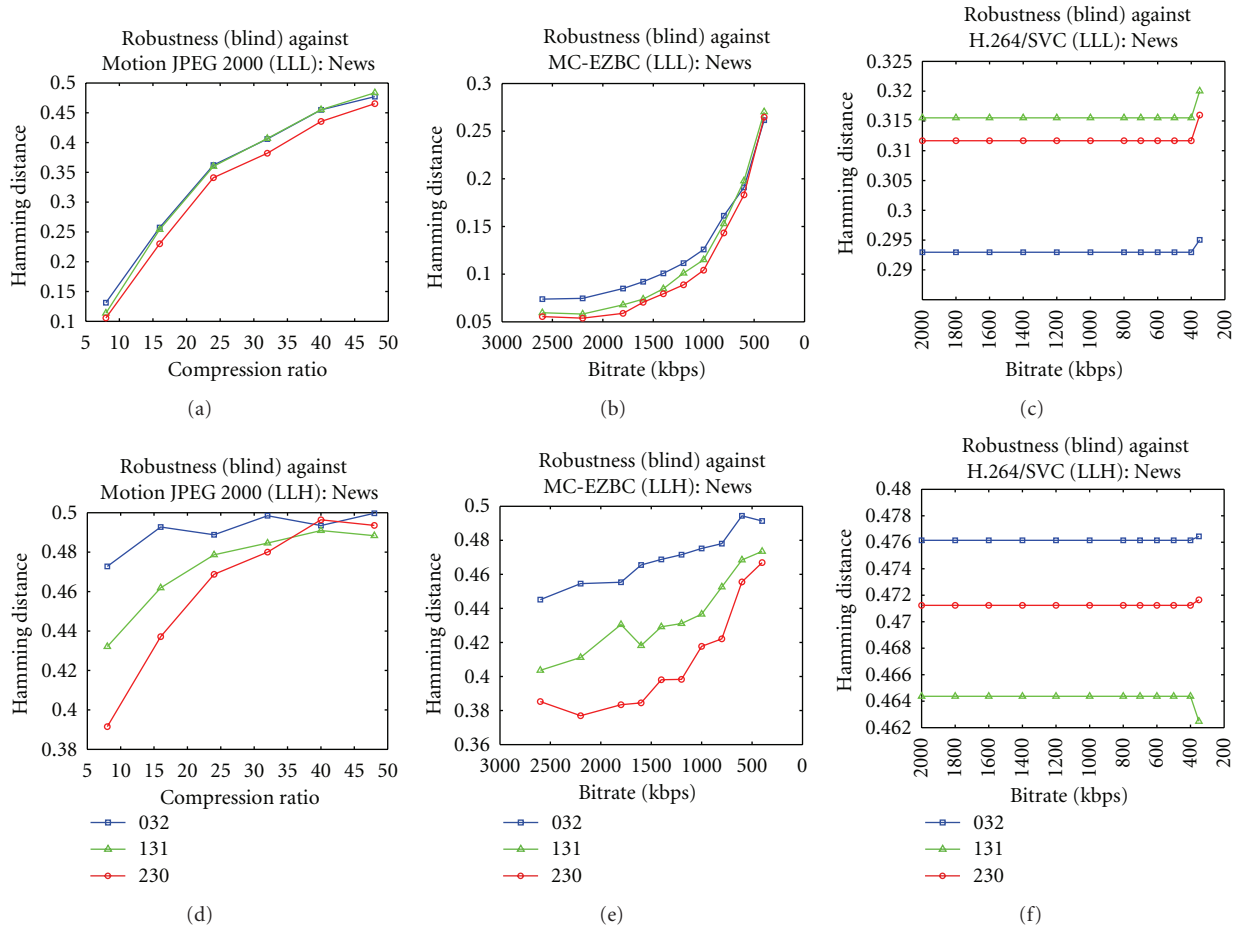


FIGURE 19: Robustness performance of blind watermarking scheme for *News* sequence. *Columns* (1), (2) and (3) show robustness against Motion JPEG 2000, MC-EZBC, and H.264-SVC, respectively. *Rows* (1) and (2) represent the embedding on temporal subbands LLL and LLH, respectively.

the MCTF with implied motion estimation for video watermarking. The MCTF was modified by taking into account the motion trajectory in obtaining an efficient update step. The proposed 2D + t domain watermarking offers improved robustness against scalable content adaptation compared to state-of-the-art conventional t + 2D video watermarking scheme in a nonblind as well as a blind watermarking scenario. The robustness performance is evaluated against scalable coding-based quality compressions attacks, including Motion JPEG 2000, MC-EZBC, and H.264-SVC (scalable extension). The proposed subband decomposition also provides low complexity as MCTF is performed only on subbands where the watermark is embedded.

Acknowledgment

This work is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) by an EPSRC-BP Dorothy Hodgkin Postgraduate Award (DHPA).

References

- [1] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Processing*, vol. 66, no. 3, pp. 283–301, 1998.
- [2] H. Inoue, A. Miyazaki, T. Araki, and T. Katsura, "Digital watermark method using the wavelet transform for video data," in *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems (ISCAS '99)*, vol. 4, pp. V-247–V-250, June 1999.
- [3] G. Doërr and J. L. Dugelay, "A guide tour of video watermarking," *Signal Processing*, vol. 18, no. 4, pp. 263–282, 2003.
- [4] M. P. Mitrea, T. B. Zaharia, F. J. Preteux, and A. Vlad, "Video watermarking based on spread spectrum and wavelet decomposition," in *Wavelet Applications in Industrial Processing II*, vol. 5607 of *Proceedings of the SPIE*, pp. 156–164, 2004.
- [5] F. Deguillaume, G. Csurka, J. J. O'Ruanidh, and T. Pun, "Robust 3D DFT video watermarking," in *Security and Watermarking of Multimedia Contents*, vol. 3657 of *Proceedings of the SPIE*, pp. 113–124, 1999.
- [6] J. H. Lim, D. J. Kim, H. T. Kim, and C. S. Won, "Digital video watermarking using 3D-DCT and intracubic correlation," in

- Security and Watermarking of Multimedia Contents III*, vol. 4314 of *Proceedings of the SPIE*, pp. 64–72, 2001.
- [7] S. J. Kim, S. H. Lee, K. S. Moon et al., “A new digital video watermarking using the dual watermark images and 3D DWT,” in *Proceedings of the IEEE Region Conference (TENCON '04)*, vol. 1, pp. 291–294, 2004.
- [8] P. Campisi and A. Neri, “Video watermarking in the 3D-DWT domain using perceptual masking,” in *IEEE International Conference on Image Processing (ICIP '05)*, pp. 997–1000, September 2005.
- [9] S. J. Choi and J. W. Woods, “Motion-compensated 3-D subband coding of video,” *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 155–167, 1999.
- [10] S. T. Hsiang and J. W. Woods, “Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank,” *Signal Processing*, vol. 16, no. 8, pp. 705–724, 2001.
- [11] C. I. Podilchuk, N. S. Jayant, and N. Farvardin, “Three-dimensional subband coding of video,” *IEEE Transactions on Image Processing*, vol. 4, no. 2, pp. 125–139, 1995.
- [12] P. Vinod and P. K. Bora, “Motion-compensated inter-frame collusion attack on video watermarking and a countermeasure,” *IEE Proceedings on Information Security*, vol. 153, no. 2, pp. 61–73, 2006.
- [13] P. Vinod, G. Doërr, and P. K. Bora, “Assessing motion-coherency in video watermarking,” in *Proceedings of the Multimedia and Security Workshop*, pp. 114–119, September 2006.
- [14] P. Meerwald and A. Uhl, “Blind motion-compensated video watermarking,” in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '08)*, pp. 357–360, June 2008.
- [15] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. Van Der Schaar, J. Cornelis, and P. Schelkens, “In-band motion compensated temporal filtering,” *Signal Processing*, vol. 19, no. 7, pp. 653–673, 2004.
- [16] F. Huo and X. Gao, “AWavelet based image watermarking scheme,” in *Proceedings of the IEEE International Conference on Image Processing*, pp. 2573–2576, Atlanta, Ga, USA, 2006.
- [17] C. Jin and J. Peng, “A robust wavelet-based blind digital watermarking algorithm,” *Information Technology Journal*, vol. 5, no. 2, pp. 358–363, 2006.
- [18] M. A. Suhail, M. S. Obaidat, S. S. Ipson, and B. Sadoun, “A comparative study of digital watermarking in JPEG and JPEG 2000 environments,” *Information Sciences*, vol. 151, pp. 93–105, 2003.
- [19] M. Barni, F. Bartolini, and A. Piva, “Improved wavelet-based watermarking through pixel-wise masking,” *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 783–791, 2001.
- [20] L. Xie and G. R. Arce, “Joint wavelet compression and authentication watermarking,” in *Proceedings of the International Conference on Image Processing (ICIP '98)*, vol. 2, pp. 427–431, October 1998.
- [21] D. Kundur and D. Hatzinakos, “Digital watermarking using multiresolution wavelet decomposition,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '98)*, vol. 5, pp. 2969–2972, May 1998.
- [22] J. R. Kim and Y. S. Moon, “Robust wavelet-based digital watermarking using level-adaptive thresholding,” in *International Conference on Image Processing (ICIP '99)*, pp. 226–230, October 1999.
- [23] D. Bhowmik and C. Abhayaratne, “A framework for evaluating wavelet based watermarking for scalable coded digital item adaptation attacks,” in *Wavelet Applications in Industrial Processing VI*, vol. 7248 of *Proceedings of the SPIE*, San Jose, Calif, USA, January 2009.
- [24] P. Meerwald, “Quantization watermarking in the JPEG2000 coding pipeline,” in *Proceedings of the 5th International Working Conference on Communication and Multimedia Security*, pp. 69–79, 2001.
- [25] X. Fan, W. Gao, Y. Lu, and D. Zhao, “Flicking reduction in all intra frame coding,” Tech. Rep. JVT-E070, 2002.
- [26] MSU Graphics & Media Lab VG, MSU Quality Measurement Tool, <http://www.compression.ru/video/>.