

Video Watermarking using Motion Compensated 2D+t+2D Filtering

Deepayan Bhowmik^{*}
Department of Electronic Engineering
The University of Sheffield
Mappin Street, Sheffield, UK, S1 3JD
d.bhowmik@sheffield.ac.uk

Charith Abhayaratne
Department of Electronic Engineering
The University of Sheffield
Mappin Street, Sheffield, UK, S1 3JD
c.abhayaratne@sheffield.ac.uk

ABSTRACT

Frame-by-frame video watermark embedding without considering motion, results in flicker and other motion mismatch artifacts in the watermarked video. Motion compensated temporal filtering (MCTF) provides a better framework for video watermarking by accounting object motion. However, depending on motion and texture characteristics of the video and the choice of spatial-temporal sub band for watermark embedding, MCTF has to be performed either on the spatial domain (t+2D) or in the wavelet domain (2D+t). In this work we propose improved video watermarking schemes by offering a generalized motion compensated 2D+t+2D framework for watermark embedding. An improved MCTF is used by modifying the MCTF update step to follow the motion trajectory in hierarchical temporal decomposition by using direct motion vector fields in the update step and implied motion vectors in the prediction step. The proposed 2D+t+2D framework with the modified MCTF-based watermarking shows better embedding distortion in terms of both mean square error and flicker metric for various combinations of spatial-temporal decompositions, compared to the existing frame-by-frame and t+2D domain video watermarking. The proposed scheme outperforms the conventional t+2D watermarking in terms of robustness performance, particularly for blind watermarking schemes where the motion is estimated from the watermarked video.

Categories and Subject Descriptors

D.2.11 [Software Engineering]: Software Architectures—*Information hiding*; I.4.2 [Image Processing and Computer Vision]: Compression (Coding)—*Video coding*; I.4.9 [Image Processing and Computer Vision]: Applications—*Digital watermarking*

General Terms

Algorithms, Experimentation, Security

^{*}The work of Mr. Bhowmik is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) by an EPSRC-BP Dorothy Hodgkin Postgraduate Award (DHPA).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM&Sec'10, September 9–10, 2010, Roma, Italy.

Copyright 2010 ACM 978-1-4503-0286-9/10/09 ...\$10.00.

Keywords

Motion compensated temporal filtering, video watermarking, 2D+t+2D, discrete wavelet transform.

1. INTRODUCTION

Motion Compensated Temporal Filtering (MCTF) has been successfully used in wavelet based scalable video coding research [5,9]. The idea of MCTF is evolved from 3D subband wavelet decomposition, which is merely an extension of spatial domain transform into temporal domain [15]. 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 trend in transform domain video watermarking.

Often video watermarking schemes are developed by extending image watermarking algorithms. Transform domain, especially, wavelet based image watermarking has been very successful in imperceptibility as well as robustness performance against various image processing attacks. As a successor of the same, several attempts have been made to extend these image watermarking algorithms into video watermarking by using them either on frame-by-frame basis [6–8, 19] or on 3D wavelet decompositions [4, 11, 12].

The frame-by-frame video watermarking considers 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 intended robustness. But frame-by-frame watermarking schemes often perform poorly in terms of robustness against various video processing attacks including temporal desynchronization, video collusion, video compression attacks etc. In order to address these issues, the video temporal dimension is exploited by the spread spectrum domain i.e. DCT and more recently wavelet based 3D decomposition of the host video. In 3D wavelet-based watermarking approaches [4, 11, 12], video is composed into 3D subbands by using separable 3D wavelet transform with shorter mother wavelets, such as Haar. Unfortunately, such naive subband decomposition-based embedding strategies, that do not consider the motion element of the sequence when embedding the watermark, often result in unpleasant flickering visual artifacts. The amount of flickering in watermarked sequences varies according to the texture, color 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 considers motion trajectory during compression coding.

The aim of this paper is to address the consideration of motion

and texture characteristics of the video sequence for extending image watermarking techniques into video. The new proposed approach is evolved from the MCTF based wavelet domain video decomposition concept, as briefed at the beginning of the paper. A few attempts have already been made to investigate the effect of motion in video watermarking attempts on incorporating motion compensation into video watermarking [13, 16, 17]. 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. Here we aim to advance further by investigating along the line of MCTF based wavelet coding to improve the robustness while keeping the imperceptibility or vice versa. Apparent problems of direct use of MCTF and t+2D decompositions in watermarking are three-fold and we offer alternative solutions to address the same.

1) In scalable video coding research it has been evident that video with different texture and motion characteristics leading to its spatial and temporal features perform differently on t+2D domain [5] and its alternative 2D+t domain [1], where MCTF is performed on the 2D wavelet decomposition domain. Further, in 3D subband decomposition for video watermarking, the consideration of motion, thus the use of MCTF, is only required for subbands where the watermarks are embedded. Therefore fixed architectures, such as t+2D or 2D+t, add unnecessary complexity in terms of motion estimation and compensation into the watermarking algorithm.

2) The 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, while keeping majority 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 above, embed the watermark in the low-pass 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 high-pass temporal frames which improves the overall imperceptibility of the watermarked video.

To overcome these shortcomings we propose a flexible 2D+t+2D generalized motion compensated temporal-spatial subband decomposition scheme using a modified motion compensated temporal filtering (MMCTF) scheme for video watermarking. Using the framework we analyze the merits and the demerits of watermark embedding using various combinations of 2D+t+2D structure and propose new video watermarking schemes to improve the imperceptibility and the robustness performance against scalable coded video attacks, such as, Motion JPEG2000, MC-EZBC and H.264-SVC. We also address the issue related to motion estimation from watermarked video without any prior knowledge of original motion information.

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 sub band decomposition framework, and the video watermarking

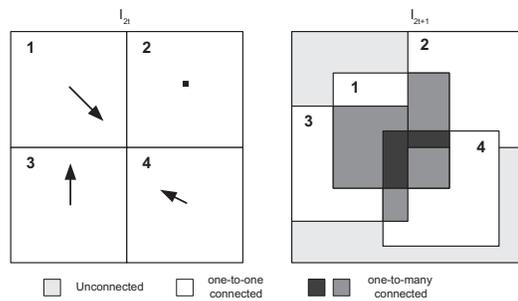


Figure 1: Pixel connectivity in I_{2t} and I_{2t+1} frames.

schemes using the implementation of different subband decomposition schemes are proposed in Section 3. The simulation results of embedding distortion performance of non-blind watermarking is shown in Section 4, followed by conclusions in Section 5.

2. MOTION COMPENSATED 2D+t+2D FILTERING

The generalised spatio-temporal 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 inter-frame pixel connectivity and use the 2D wavelet transform for spatial decomposition. In this section we describe the modified motion compensated temporal filtering (MMCTF) first and then propose the 2D+t+2D general framework based on MMCTF.

2.1 MMCTF

We formulate the MMCTF 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. The I_{2t} frame is partitioned into non-overlapping blocks and for each block, vertical and horizontal displacements are quantified and represented as motion vector fields $\mathcal{V}^{c \rightarrow r}$ and $\mathcal{H}^{c \rightarrow r}$, respectively. In the I_{2t} frame, each block can be one of two types, namely inter and intra blocks, where the motion is only estimated for the former block type only.

Similarly, as far as the I_{2t+1} frame is concerned any pixel can be one of three types, namely, one-to-one connected, one-to-many connected and unconnected (as shown in Figure 1), depending on their connectivity to pixels in the I_{2t} frame following the implied motion vector fields $\mathcal{V}^{c \leftarrow r}$ and $\mathcal{H}^{c \leftarrow r}$, which are simply the directional inverse of the original motion vector fields, $\mathcal{V}^{c \rightarrow r}$ and $\mathcal{H}^{c \rightarrow r}$.

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:

The prediction step:

For one-to-one connected pixels:

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

For one-to-many connected pixels:

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

where J is the total number of connections. For unconnected pixels:

$$I'_{2t+1}[m, n] = I_{2t+1}[m, n]. \quad (3)$$

The above case is similar to the no prediction case as in intra blocks used in conventional MCTF.

The update step:

For inter blocks: Every pixel in an inter block is one-to-one connected with a unique pixel in I_{2t+1} . Then the update step is computed as

$$I'_t[m, n] = I_{2t}[m, n] + \frac{1}{2} I'_{2t+1}[m - \mathcal{H}^{c \leftarrow r}, n - \mathcal{V}^{c \leftarrow r}]. \quad (4)$$

For intra blocks: As there are no motion compensated connections with I_{2t+1} ,

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

Finally these lifting steps are followed by the normalization step.

$$I''_{2t}[m, n] = \sqrt{2} I'_t[m, n], \quad (6)$$

$$I''_{2t+1}[m, n] = \frac{1}{\sqrt{2}} I'_{2t+1}[m, n]. \quad (7)$$

The temporally decomposed frames I''_{2t} and I''_{2t+1} are the first level low and high pass frames and are denoted as L and H temporal subbands. These steps are repeated for all frames in L to obtain LL and LH sub bands and continued to obtain the desired number of temporal decomposition levels. For the inverse transform, the order of operation of steps is reversed and the first operand in lifting steps is changed to subject variable in above equations.

2.2 2D+t+2D framework

As we discussed earlier in Section 1, in a 3D video decomposition scheme, t+2D is achieved by performing temporal decomposition followed by a spatial transform where as 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 analyse both the combinations in order to enhance the video watermarking performance. A common flexible reconfigurable framework, which allows to create such possible combinations, are 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 Discrete Wavelet Transform (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 sequence of spatial LL subbands are temporally decomposed using the MMCTF 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. For example, (032) and (230) parameter combinations

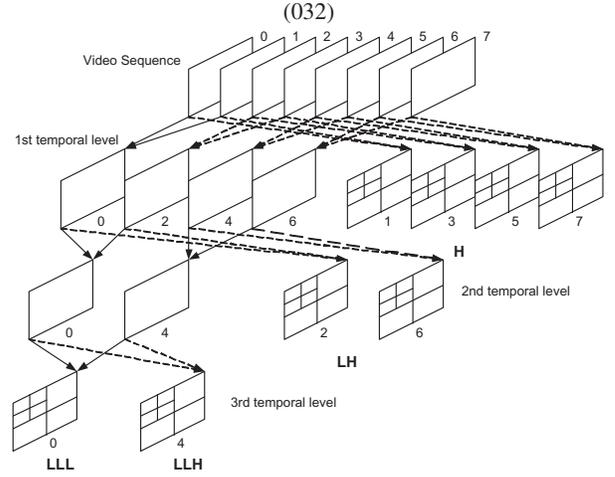


Figure 2: Realization of 3-2 temporal schemes using the 2D+t+2D framework with different parameters: (032).

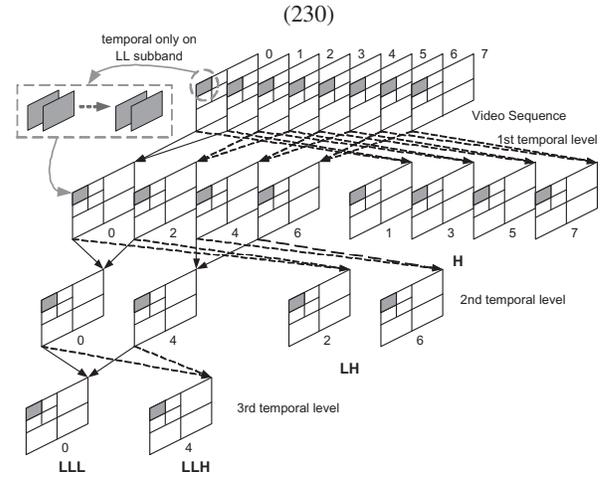


Figure 3: Realization of 3-2 temporal schemes using the 2D+t+2D framework with different parameters: (230).

result in t+2D and 2D+t motion compensated 3D subband decompositions, respectively. The same amount subband decomposition levels can be obtained by also using the parameter combination (131) using the proposed generalized scheme implementation. The combination (002) allows 2D decomposition of all frames for frame by frame watermark embedding. The realization of these examples are shown in Figure 2, Figure 3, Figure 4 and Figure 5. We use the notation (LLL, LLH, LH, H) to denote the temporal subbands after a 3 level decomposition. We have described the the use of this framework in combination with watermarking algorithms, in the next section.

3. VIDEO WATERMARKING IN 2D+t+2D SPATIO-TEMPORAL 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

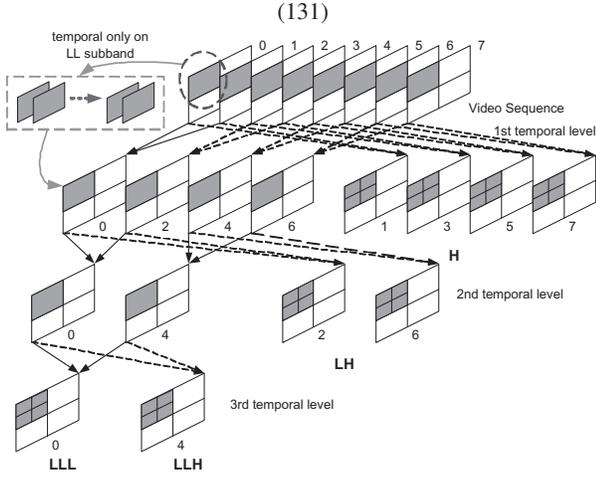


Figure 4: Realization of 3-2 temporal schemes using the 2D+t+2D framework with different parameters: (131).

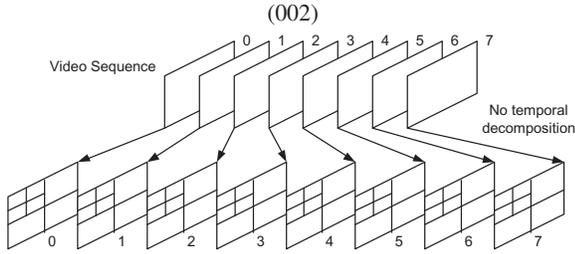


Figure 5: Realization of spatial 2D frame-by-frame scheme using the 2D+t+2D framework with different parameters: (002).

image watermarking algorithms followed by the proposed video watermarking scheme. Then we carried on to analyze various combinations in the proposed video decomposition framework to decide on 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 multi-resolution spatio-frequency representation of the signals, the DWT became the major transform for spread spectrum watermarking. The wavelet domain watermarking algorithms often share a common model. We aim to present a common model suitable for various wavelet based watermarking schemes. Therefore, firstly we have generalized and broadly categorized the wavelet based watermarking schemes into two groups, namely: 1) *Magnitude alteration modification* and 2) *Re-quantization based modification*, and select example cases from each category. These selections also represent 1) *Non-blind* and 2) *Blind* cases of watermarking classification, respectively. A detailed discussion of these categorizations can be found from our previous work [3].

Magnitude Alteration based Watermarking (The Non-blind Case):

For the magnitude alteration category, additive watermarking is used, based on [10]. Magnitude based additive watermarking is

a popular choice for many people, watermarking within the DWT domain, due to its simplicity. Specific values are increased or decreased dependant on the magnitude of the coefficient, by making the modified coefficient a function of the original coefficient. The equation we used for magnitude based additive watermarking can be described as:

$$C'_{s,t}[m, n] = C_{s,t}[m, n] + \alpha C_{s,t}[m, n]W, \quad (8)$$

where $C_{s,t}[m, n]$ is the original decomposed coefficient at s, t spatio-temporal 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. The level adaptive thresholding as described in [10] is also taken into account to avoid watermark embedding in small or nearly zero coefficients to minimize the false detection.

Re-quantization based Watermarking (The Blind Case):

In this category, we used a quantization-based watermarking method as proposed in [18]. Quantization based watermarking relies on modifying various coefficients towards a specific quantization step, δ . The method modifies the median coefficient by using a non-overlapping 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:

$$C'_2 = f(\gamma, C_1, C_3, W), \quad (9)$$

where W is the input watermark sequence, γ is the weighting parameter and $f(\cdot)$ denotes a non-linear transformation and the quantization step δ is defined as:

$$\delta = \gamma \frac{|C_1| + |C_3|}{2}. \quad (10)$$

3.2 Proposed video watermarking scheme

The new video watermarking scheme uses the image watermarking algorithms on spatial-temporal decomposed video. The system block diagrams for watermark embedding, a non-blind extraction and a blind extraction are shown in Figure 6, Figure 7 and Figure 8, respectively.

Embedding:

To embed the watermark, first spatio-temporal decomposition is performed on the host video sequence by applying spatial 2D-DWT followed by temporal MMCTF 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 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 magnitude alteration or a re-quantisation based modification algorithm by selecting various temporal low or high pass frames (i.e. LLL or LLH etc.) and spatial subband within the selected frame. Once embedded, the coefficients follow inverse process of spatio temporal decomposition in order to reconstruct the watermarked video.

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 7 and Figure 8. The watermark coefficients are retrieved by applying 2D+t+2D decomposition on watermarked test video.

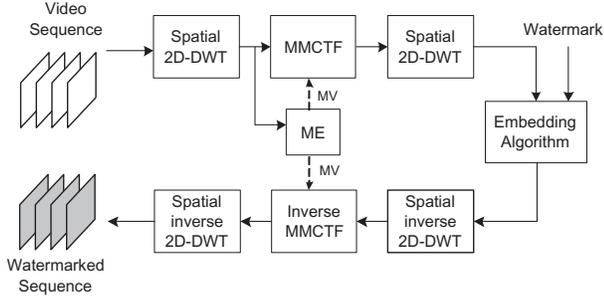


Figure 6: System blocks for watermark embedding scheme in 2D+t+2D spatio-temporal decomposition.

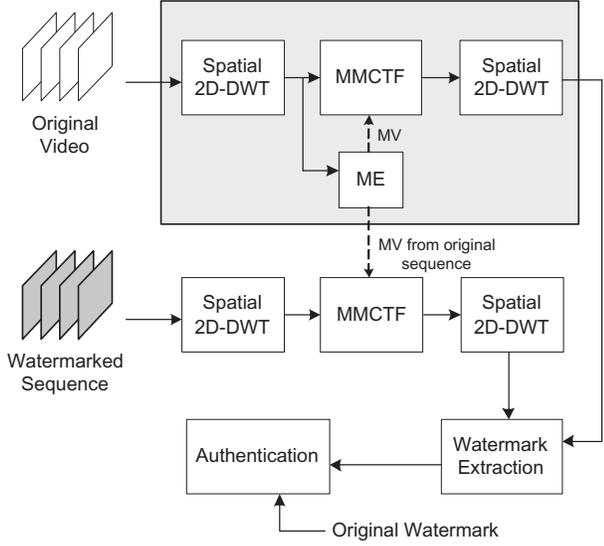


Figure 7: System blocks for non-blind watermark extraction scheme in 2D+t+2D spatio-temporal decomposition.

At this point we need to specifically mention about the motion information retrieval. For a non-blind algorithm the original video sequence is available at the decoder and hence the motion vector is obtained from the original video. After spatio-temporal filtering on test and original video, the coefficients are compared to extract the watermark. On the other hand, 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 spatio-temporal coefficients are obtained for the detection.

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

$$H(W, W') = \frac{1}{L} \sum_{i=0}^{L-1} W_i \oplus W'_i, \quad (11)$$

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.

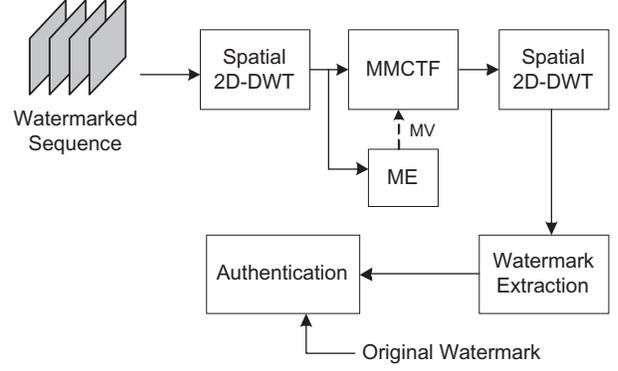


Figure 8: System blocks for blind watermark extraction scheme in 2D+t+2D spatio-temporal decomposition.

3.3 The framework analysis in video watermarking context

Before approaching to the experimental simulations, in this sub section we aim to address the issues related to MMCTF 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.

3.3.1 On improving imperceptibility

In wavelet domain watermarking research, it is well known fact that embedding in high frequency subbands offers better imperceptibility and low frequency embedding provides better robustness. Often wavelet decompositions compact most of the energy in low frequency subbands and leaves lesser 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 shows 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 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 macro block, a fixed size block matching (FSBM) motion estimation with ± 16 search window. The results are shown in Table 1 and the histograms of the coefficients of LLL and LLH are shown in Figure 9. The inner graphs in Figure 9 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 spatio-temporal parameter in the framework to improve the robustness while keeping better imperceptibility.

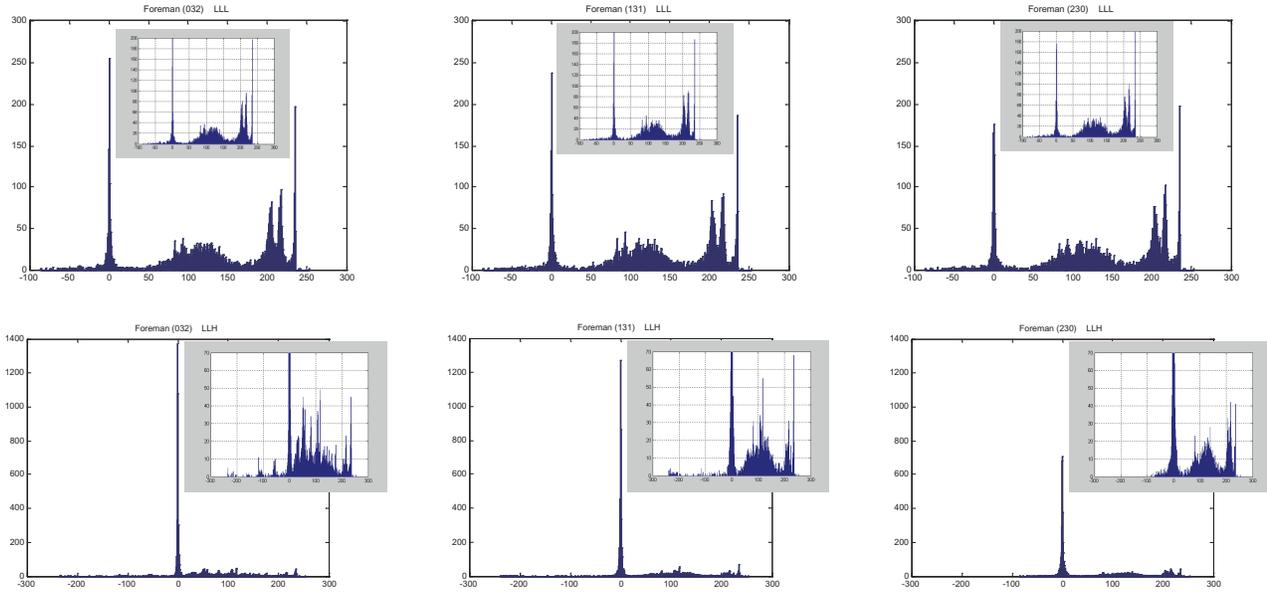


Figure 9: Histogram of coefficients at LL_s for 3^{rd} level temporal low and high frequency frames. Row 1) & 2) represents LLL and LLH temporal frames, respectively and Column 1), 2) & 3) shows 032, 131 and 230 combinations of 2D+t+2D framework.

Table 1: Sum of energy of coefficients at LL_s for first 8 temporal low and high frequency frames of Foreman sequence.

Temporal frames	Sum of Energy		
	032	131	230
LLL	1.49×10^8	1.49×10^8	1.49×10^8
LLH	4.15×10^7	5.17×10^7	6.55×10^7
LH1	3.39×10^7	4.35×10^7	5.41×10^7
LH2	3.14×10^7	3.87×10^7	5.03×10^7
H1	3.03×10^7	3.23×10^7	4.25×10^7
H2	2.68×10^7	3.48×10^7	4.08×10^7
H3	2.76×10^7	3.42×10^7	3.83×10^7
H4	3.42×10^7	3.53×10^7	4.36×10^7

3.3.2 On motion retrieval

In an MCTF based video watermarking scheme motion information contributes at large for temporal decomposition along motion trajectory. The watermarking embedding modification in the temporal domain causes motion mismatch which affects the decoder performance. While original motion information is available for a non-blind watermarking scheme, a motion estimation must be done in the case of a blind video watermarking scheme. In this case, the motion vector is 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, a 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 we investigate the performance based on motion search range (SR) and 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 sub-

band on LLL and LLH temporal frames. Firstly the watermark extraction is done by using the original motion vector and then using various combinations of MB and SR to find the best the motion retrieval parameters. The results are shown in Table 2 using Foreman CIF size video sequence for 032, 131 and 230 spatio-temporal decompositions. The motion is estimated using a fixed size block motion algorithm. Due to the limitations in macro-block size and integer pixel motion search, 16×16 MB search is excluded for 230 decomposition.

The results show that for a MB size more than 8×8 , 2D+t outperform t+2D. In this context the spatio-temporal decompositions can be ranked as $(230) > (131) > (032)$. However in 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.

4. EXPERIMENTAL RESULTS

We used the following experimental setups 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 and the 2D DWT were omitted.

There are two different sets of result obtained for *Foreman* test video sequence to show the embedding distortion and the robustness performance. We have used one non-blind 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 morphological Haar transform was used as the 2D de-

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 Foreman sequence.

(a) LLL						
	Original MV	MV from watermarked video: MB/SR				
		16×16 / ± 32	8×8 ± 32	8×8 / ± 16	4×4 / ± 16	4×4 / ± 8
032	0.03	0.05	0.08	0.08	0.08	0.09
131	0.03	0.04	0.08	0.06	0.13	0.12
230	0.03	-	0.05	0.05	0.15	0.14

(b) LLH						
	Original MV	MV from watermarked video: MB/SR				
		16×16 / ± 32	8×8 ± 32	8×8 / ± 16	4×4 / ± 16	4×4 / ± 8
032	0.17	0.36	0.47	0.44	0.51	0.49
131	0.20	0.25	0.36	0.34	0.48	0.47
230	0.20	-	0.28	0.28	0.38	0.38

compositions following our previous work on optimum wavelet kernel for image watermarking [2].

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 3^{rd} temporal level high pass (LLH) and low pass (LLL) frames to embed the watermark. Other video decomposition parameters are set to: 1) A groups of picture (GOP) size of 8 frames, 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 3.3.2.

For embedding distortion measure we used Mean Square Error (MSE) and also measured the amount of flicker introduced due to watermark embedding by using the flicker metric in the MSU Quality Measurement Tool [14]. The flicker metric compares the flicker content in the watermarked video with respect to the original video. In both metrics the lower values correspond to the better distortion performance. On the other hand the watermarking robustness is represented by Hamming distance as mentioned in Eq. (11) and lower Hamming distance corresponds a better detection performance. Various scalable coded quality compression attacks are considered, such as, Motion JPEG2000 (using *Open JPEG* software code), MC-EZBC scalable video coding (an RWTH Aachen University implementation) and H.264-SVC (scalable extension) using JSVM software (Release 9.15).

The experiments are divided into two sets, one for non-blind watermarking and the other for blind watermarking, described as follows:

Experiment Set 1 (Non-blind watermarking):

A non-blind magnitude alteration based watermarking algorithm is chosen from Section 3.1 for this experimental set. The weighting parameter α is set to 0.1 while using a level adaptive threshold to select the coefficients to embed. The embedding distortion results are shown in Figure 10 and Figure 11 for LLL and LLH frames, respectively with a data capacity of 2096 binary logo based watermark bit. In each of the figures in x-axis, the top row represents the MSE while flicker metrics are shown in the bottom row. The

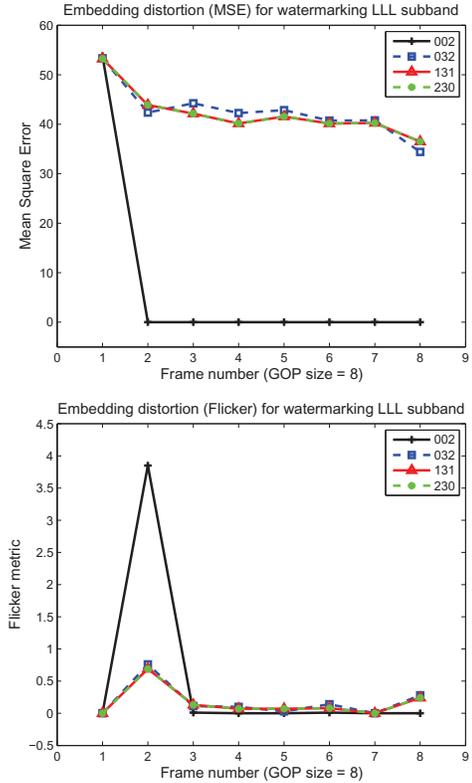


Figure 10: Embedding distortion performance for non-blind watermarking on LLL temporal subband for Foreman sequence. Row 1) & 2) represents the Mean Square Error (MSE) & Flicker metric, respectively.

y-axis of the figures presents various frames within one GOP with the size of 8. It is evident from these results that the proposed framework based methods ((230), (032) and (131)) outperform the frame-by-frame embedding (002) with respect to embedding distortion performance particularly flickering.

The robustness results for this experiment set are shown in Figure 14 where the x-axis represents compression ratio (Motion JPEG) or video bit rate (MC-EZBC and H.264-SVC) and the y-axis shows the corresponding Hamming distances. Rows 1) & 2) show the results for the LLL and LLH frame selections, respectively. Due to the nature of energy distribution, the robustness performances are comparable in the LLL where as any combination of temporal filtering on spatial decomposition (i.e. (131) and (230)) outperforms a conventional t+2D based scheme in the high pass temporal frame LLH . Although a frame-by-frame (002) provides better robustness against video compression, it is wise to avoid such scheme to overcome the flickering issue.

Experiment Set 2 (Blind watermarking):

Here a re-quantization based blind watermarking scheme, described in Section 3.1, is used, while keeping all other common experimental parameters similar to *Experiment Set 1*. The watermark weighting parameter γ is set to 0.1 for LLL and 0.4 for the LLH subband choices. In this case the motion information is obtained from the watermarked test video and the motion parameters are set to the macro block size of 8×8 with a ± 16 search window. The embedding distortion results are shown in Figure 12 and Figure 13 for LLL and

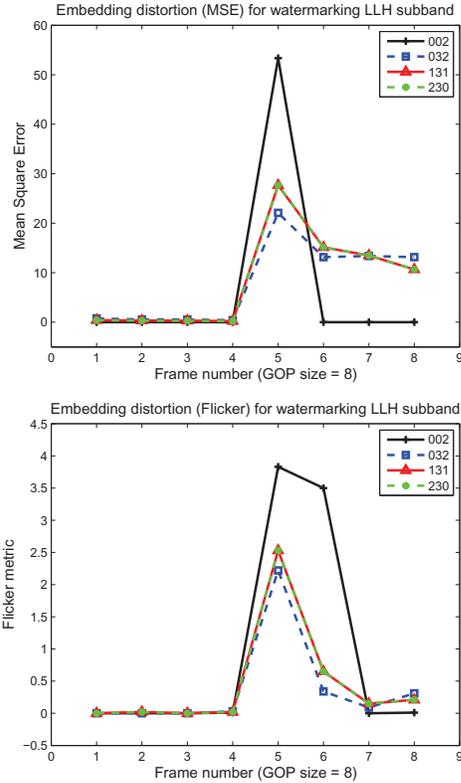


Figure 11: Embedding distortion performance for non-blind watermarking on LLH temporal subband for Foreman sequence. Row 1) & 2) represents the Mean Square Error (MSE) & Flicker metric, respectively.

LLH temporal frames, respectively, with the data capacity of 2112 binary logo based watermark bit. A similar result set, as in the previous experiment setup, is obtained which shows that the MCTF based schemes helps to remove the flicker and the overall performance can be ranked as (230)>(131)>(032)>(002).

The robustness results for the blind watermarking scheme are shown in Figure 15, where the top row shows results for the LLL temporal subband while results for LLH are shown in the bottom row. The columns represent various scalability attacks, Motion JPEG2000, MC-EZBC and H.264-SVC, respectively. In this case, due to the energy distribution in high frequency temporal subbands, the overall robustness performance is poor for watermark embedding in high frequency subbands (LLH).

We have studied in Section 3.3.2 that a 2D+t scheme offers better motion estimation for the blind watermarking algorithm and hence outperforms t+2D in robustness performance. Various spatio-temporal combinations, based on the robustness performance against various attacks, can be ranked as (230)>(131)>(032). With a similar argument regarding the flickering problem, as discussed previously, we exclude the frame-by-frame (002) algorithm while proposing the video watermarking scheme. However for the completeness of this paper we have shown the robustness results for the frame-by-frame (002) watermark embedding.

To conclude the discussion, we suggest that, a choice of 2D+t based watermarking scheme improves the imperceptibility and the robustness performance in a video watermarking scenario for a non-blind as well as a blind watermarking algorithm.

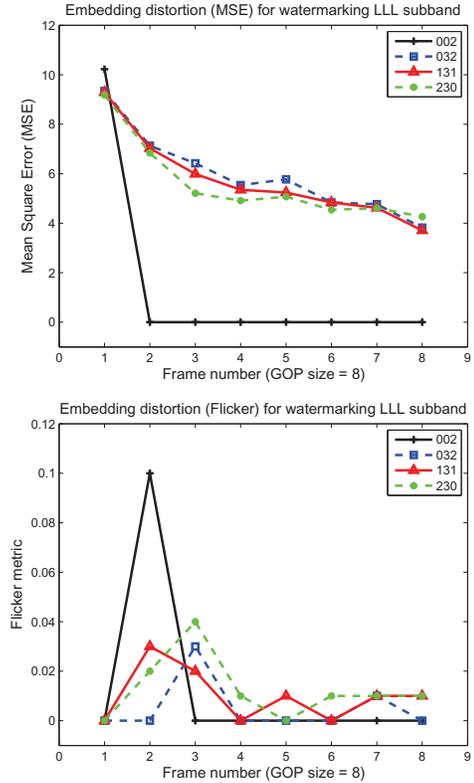


Figure 12: Embedding distortion performance for blind watermarking on LLL temporal subband for Foreman sequence. Row 1) & 2) represents the Mean Square Error (MSE) & Flicker metric, respectively.

5. CONCLUSIONS

In this paper, we have presented a flexible generalized motion compensated temporal-spatial subband decomposition scheme, based on the MMCTF for video watermarking. The MCTF was modified by taking into account the motion trajectory into obtaining an efficient update step. The embedding distortion performance evaluated using both MSE and flicker difference metric shows superior performance for the MMCTF driven 2D+t+2D subband domain watermarking as opposed to frame-by-frame 2D wavelet domain watermarking which does not take motion into account. The proposed subband decomposition also provides low complexity as MCTF is performed only on subbands where the watermark is embedded. The robustness performance against scalable coding based compressions attacks, including Motion JPEG2000, MC-EZBC and H.264-SVC (scalable extension) is also evaluated. The proposed 2D+t based video watermarking scheme within 2D+t+2D filtering framework outperforms conventional t+2D watermarking schemes in a non-blind as well as a blind watermarking scenario.

6. REFERENCES

- [1] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. van der Schaar, J. Cornelis, and P. Schelkens. In-band motion compensated temporal filtering. *Signal Processing: Image Communication*, 19(7):653–673, Aug. 2004.
- [2] D. Bhowmik and C. Abhayaratne. Morphological wavelet

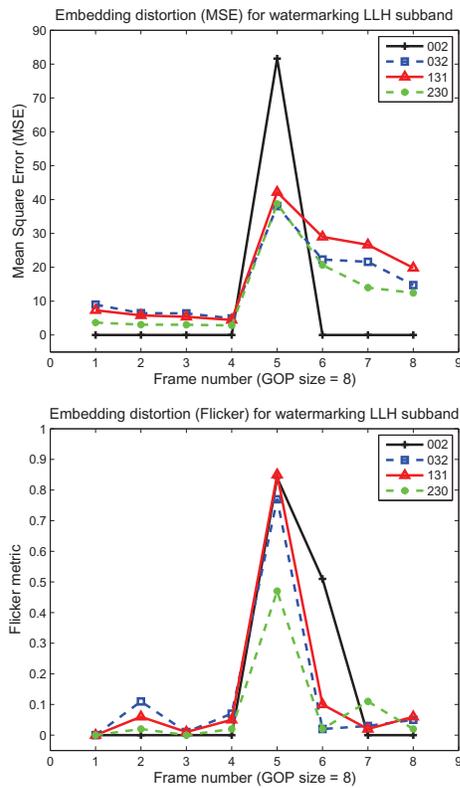


Figure 13: Embedding distortion performance for blind watermarking on LLH temporal subband for Foreman sequence. Row 1) & 2) represents the Mean Square Error (MSE) & Flicker metric, respectively.

domain image watermarking. In *Proc. European Signal Processing Conference (EUSIPCO)*, pages 2539–2543, 2007.

- [3] D. Bhowmik and C. Abhayaratne. A framework for evaluating wavelet based watermarking for scalable coded digital item adaptation attacks. In *Proc. SPIE Wavelet Appl. in Industrial Processing VI*, volume 7248, page 72480M (10 pages), 2009.
- [4] P. Campisi. Video watermarking in the 3D-DWT domain using quantization-based methods. In *Proc. IEEE MMSP*, pages 1–4, 2005.
- [5] S. Choi and J. W. Woods. Motion-compensated 3-D subband coding of video. *IEEE Trans. Image Processing*, 8(2):155–167, Feb. 1999.

- [6] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoan. Secure spread spectrum watermarking for multimedia. *IEEE Trans. Image Processing*, 6(12):1673–1687, Dec. 1997.
- [7] G. Doërr and J.-L. Dugelay. A guide tour of video watermarking. *Signal Processing: Image Communication*, 18(4):263–282, 2003.
- [8] F. Hartung and B. Girod. Watermarking of uncompressed and compressed video. *Signal Processing*, 66(3):283–301, 1998.
- [9] S.-T. Hsiang and J. W. Woods. Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank. *Signal Processing: Image Communication*, 16:705–724, May 2001.
- [10] J. R. Kim and Y. S. Moon. A robust wavelet-based digital watermarking using level-adaptive thresholding. In *Proc. IEEE ICIP*, volume 2, pages 226–230, 1999.
- [11] S.-J. Kim, S.-H. Lee, K.-S. Moon, W.-H. Cho, I.-T. Lim, K.-R. Kwon, and K.-I. Lee. A new digital video watermarking using the dual watermark images and 3D DWT. In *Proc. IEEE Region 10 TENCON*, volume 1, pages 291–294, 2004.
- [12] Y. Li, X. Gao, and J. Hongbing. A 3D wavelet based spatial-temporal approach for video watermarking. In *Proc. IEEE Int'l Conf. on Comput. Intelligence and Multimedia App. (ICCIMA)*, pages 260–265, 2003.
- [13] P. Meerwald and A. Uhl. Blind motion-compensated video watermarking. In *Proc. IEEE ICME*, pages 357–360, 2008.
- [14] V. G. MSU Graphics & Media Lab. MSU quality measurement tool. Available: <http://www.compression.ru/video/> [Accessed: Jan. 15, 2010].
- [15] C. I. Podilchuk, N. S. Jayant, and N. Farvardin. Three-dimensional subband coding of video. *IEEE Trans. Image Processing*, 4(2):125–139, Feb. 1995.
- [16] P. Vinod and P. K. Bora. Motion-compensated inter-frame collusion attack on video watermarking and a countermeasure. *IEE Proceedings on Information Security*, 153(2):61–73, June 2006.
- [17] P. Vinod, G. Doërr, and P. K. Bora. Assessing motion-coherency in video watermarking. In *Proc. ACM Multimedia and Security*, pages 114–119, 2006.
- [18] L. Xie and G. R. Arce. Joint wavelet compression and authentication watermarking. In *Proc. IEEE ICIP*, volume 2, pages 427–431, 1998.
- [19] W. Zhu, Z. Xiong, and Y.-Q. Zhang. Multiresolution watermarking for images and video. *IEEE Trans. Circ. and Syst. for Video Tech*, 9(4):545–550, Jun 1999.

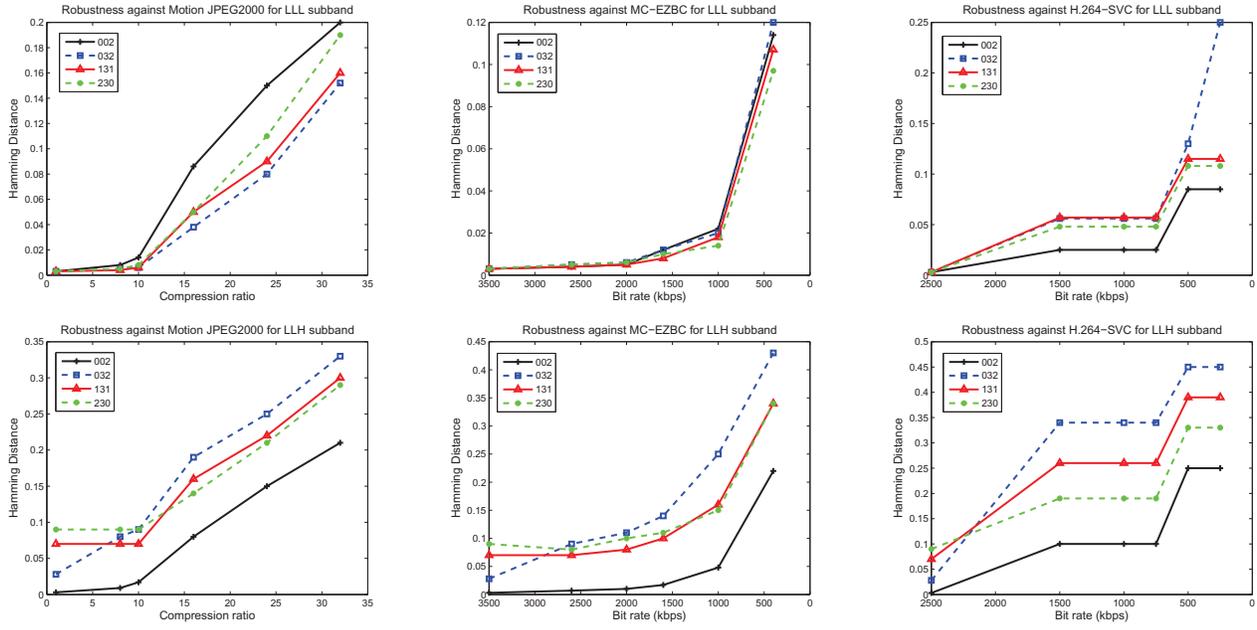


Figure 14: Robustness performance of non-blind watermarking scheme for Foreman sequence. *Column 1, 2) & 3)* show robustness against Motion JPEG2000, MC-EZBC and H.264-SVC, respectively. *Row 1) & 2)* represents the embedding on temporal subbands LLL & LLH, respectively.

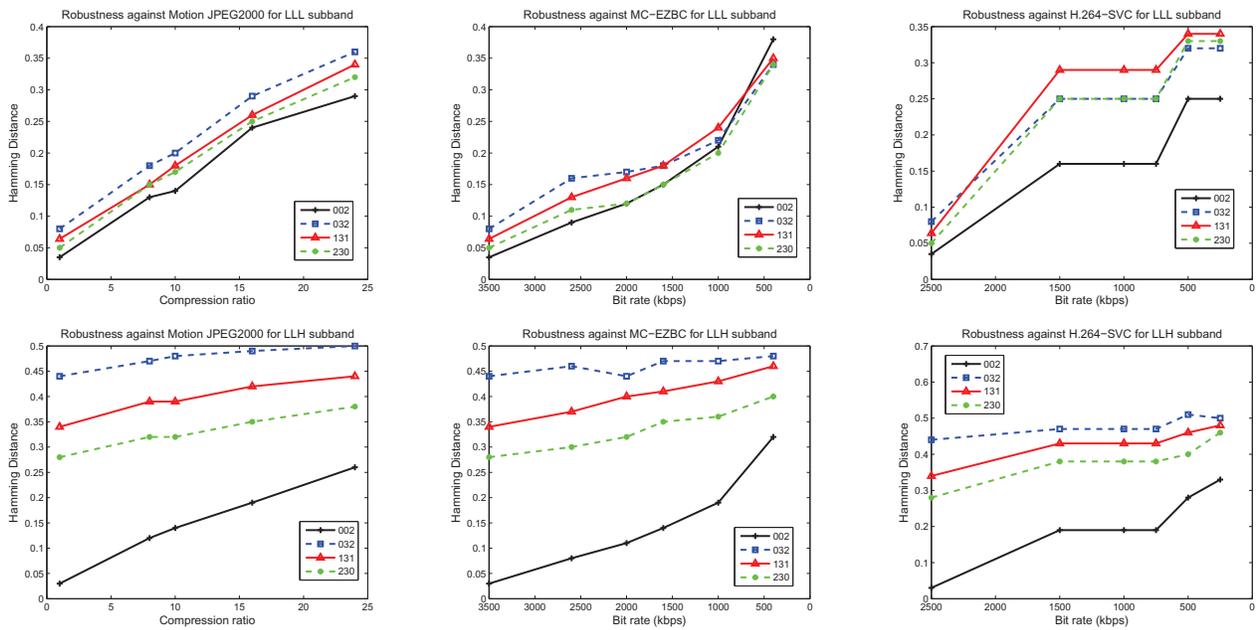


Figure 15: Robustness performance of blind watermarking scheme for Foreman sequence. *Column 1, 2) & 3)* show robustness against Motion JPEG2000, MC-EZBC and H.264-SVC, respectively. *Row 1) & 2)* represents the embedding on temporal subbands LLL & LLH, respectively.