3D VIDEO WATERMARKING USING DT-DWT TO RESIST SYNTHESIS VIEW ATTACK

Shuvendu Rana, Arijit Sur

Multimedia Lab, Department of Computer Science and Engineering
IIT Guwahati, Assam, India 781039
Emails: shuvendu, arijit@iitg.ernet.in

ABSTRACT

In this paper, a 3D video watermarking scheme is proposed for depth image based rendering (DIBR) based multi view video plus depth (MVD) encoding technique. To make the scheme invariant to view synthesis process in DIBR technique, watermark is inserted in a center view which is rendered from left and right views of a 3D video frame. A low pass center view, obtained from the motion compensated temporal filtering over all the frames of a GOP, is used for embedding to reduce the temporal flickering artifacts. To make the scheme invariant to the DIBR process, 2D DT-DWT block coefficients of low-pass center view are used for embedding by exploiting its shift invariance and directional property. A comprehensive set of experiments have been carried out to justify the robustness of the proposed scheme over existing schemes with respect to compression of the 3D-HEVC video codec and synthesis view attack.

Index Terms— depth-image-based rendering (DIBR); 3D high-efficient-video-coding (3D-HEVC); dual-tree discrete-wavelet-transform (DT-DWT); watermarking; synthesis view

1. INTRODUCTION

Huge advancement in Internet technology and wide availability of the cheaper display devices made the 3D video more attractive due to its immersive experience. Due to the very nature of the digital media, it is easy to copy and redistribute the media content. This makes the video piracy and content tampering as a serious threat over efficient and secure video communication specially over Internet. Watermarking is being regarded as an efficient DRM (digital right management) tool for video transmission. Although, quite a few works have been reported on the video watermarking, it is observed that relatively less attention has been paid on 3D video watermarking until recently. In 3D viewing, the immersive binocular vision [1] is generated by incorporating the virtual depth plane which can be obtained by capturing image or video by more than one cameras (primarily with one left and one right). In this kind of 3D virtualization, a visual discomfort, called vertical and horizontal parallax [2] may be occurred in time of viewing the 3D video. To reduce this discomfort, synthetically generated intermediate views can be used to produce smoother display. Recently, depth image based rendering (DIBR) technique [3] has been introduced for generating this synthetic view and the corresponding display is popularly known as auto-stereoscopic display where the conventional multi view coding (MVC) technology has been extended as multi view video with depth (MVD) [4]. This MVD based 3D video coding is deployed in recent most video standard popularly known as 3D HEVC.

The 3D depth vision for the human perception is generated by merging two or more than two camera views (corresponding to the left eye and right eye views). Hoffman et al. [1] have shown that the common pixels move horizontally with common left and right views respectively. This makes a common region, called dependent view and an uncommon region, called independent view, for the left and right views. To achieve higher compression efficiency, few intermediate views are synthesized using the DIBR technique. One of the main challenge of 3D video watermarking is to secure not only both left and right views but these synthesized views also. In the existing 3D watermarking schemes, watermark is inserted either in both the views [5, 6] or in the synthesized center view [7]. In DIBR technique, it is observed that only one full view (left or right having both dependent and independent parts) along with independent part of other view(s) are communicated to the receiver side. So the watermark embedded in the other views (where only independent parts are encoded) may not be properly extracted. Moreover, in case of MVD based encoding, the view synthesis process using depth (say view synthesis attack) itself may destroy the watermark. In that case, a compromised receiver can generate a synthesized view from watermarked left and right views. Thus the existing 3D watermarking scheme is not very suitable for handling the DIBR technique used in MVD based encoding.

It has been observed in the literature that few image based 3D watermarking schemes are extended for the video. For example, a depth based image watermarking is proposed by Campisi [8] where authors did not consider the dependent and the independent views, which makes the scheme relatively less robust against DIBR technique. In another scheme, Lin and Wu proposed a multiple object based watermarking scheme [9] for 3D images. But in case of video, these schemes may not survive the 3D-video compression process due to the motion estimation and quantization. Garcia and Dugelay [10] have proposed a object based watermarking scheme for 3D video. This scheme may suffer from spatial view de-synchronization error since the dependent and the independent view in successive frames are different. This may also suffer from temporal de-synchronization as the object motion is not considered. Recently, Kim et al. proposed a 3D image watermarking using DT-CWT [7]. It is observed that the scheme may not fully resilient to the DIBR process and type II collusion attack [11]. Moreover, this image based scheme may suffer the motion compensation process when is applied to the video watermarking. In another work, Lee et al. proposed a perceptual depth based 3D video watermarking scheme [5] where the watermark is embedded with the hidden pixels and high motion Z-axis using DIBR. Since the watermark is embedded by altering the Z-axis only, the rendering process (DIBR) is not hampered due to embedding. But, since the watermark region is not synchronized with the continuous frames in the GOP (group of picture), watermark bits may be destroyed at the time of motion compensation. Also the scheme is not resilient against synthesis view attack and collusion attack. Recently, Rana and Sur proposed an independent region based watermarking scheme [6] using Z-axis where the watermark is embedded on the independent part of the left and right view for a video frame. The main motivation for using Z-axis based scheme is that the independent view is mutually exclusive and the watermark region could not be colluded. But due to the partial presence of the independent view in the synthesized view, the watermarking scheme could not resist the synthesis view attack.

There are few relevant observations be made from the above discussion. Firstly, most of the existing schemes are not resilient to DIBR based encoding and view synthesis process. Secondly, although independent view based embedding is not visually perceptible due to parallax display [2], it can easily be destroyed by modifying independent region without substantial perceptual degradation. Thirdly, with the availability of the depth information at the receiver, it is possible to regenerate the other views from a single view which may be a serious threat over view based watermarking. Finally, since there is a strong similarity between different views of a frame, generation of watermarks for different views should be homogeneous to resist the collusion attack. Taking these facts into consideration, in this paper, a 3D blind video watermarking scheme is proposed which embeds watermark in the synthesized center view to resist the synthesis view attack. The challenge of the proposed scheme is to protect the left and right views which are generated from the center view by a reverse synthesis process and to handle the temporal de-synchronization due to motion in the successive frames. To achieve this, motion compensated temporal filtering is done on center views of the successive video frames and DT-DWT [12] coefficients of the low pass center view are used for watermark embedding. In this scheme, blockwise approximate shift invariant DT-DWT [12] is employed to make the watermark robust against DIBR technique. A comprehensive set of experiments are carried out to justify the applicability of the proposed scheme over the existing literature. The rest of the paper is organized as follows. In Sec. 2, proposed watermarking scheme is presented. The experimental results are presented in the Sec. 3 and finally the paper is concluded in Sec. 4.

2. PROPOSED SCHEME

In the proposed scheme, watermark is embedded in the synthesized center view which is obtained by rendering the left and right views using DIBR. After embedding, the watermarked center view is inverse rendered to get watermarked version of the left and right views which are being communicated to the receiver side. So the primary goal of this work is to embed the watermark in center view in such a way that the watermark propagates to left and right views during inverse rendering process. An efficient block selection strategy has been employed to meet this requirement. Moreover, DT-DWT coefficients of a selected block are used for embedding to make the scheme invariant to the inverse rendering process. The details of block and coefficient selection processes are discussed in the successive subsections. It is observed in Sec. 1 that most of the existing 3D watermarking schemes did not consider the motion along the temporal axis which may produce temporal flickering due to embedding. To reduce this temporal artifacts, the motion compensated temporal filtering has been employed on successive center views in a group of picture (GOP) and a low pass version is chosen for embedding. In the proposed scheme, since the dependent portion of both left and right views are watermarked with homogeneous watermark (generated from same center view), the type II collusion attack can also be resisted.

2.1. Embedding Zone Selection

In the proposed scheme, a GOP of a cover 3D video sequence is taken for watermarking. Firstly, left and right views for each frame of a 3D video are identified and a center view corresponding to each frame has been generated. Then all the center views for the GOP are subjected to motion compensated temporal filtering as discussed in [13] to get the low pass version of such views. In this paper, DCT based motion compensated temporal filtering (MCDCT-TF) is used which was first used by Atta et al. for temporal scalable video coding in [14] where they have used 3/2 temporal DCT to extract 2 low pass frames from a GOP of 3 frames. In the proposed scheme, $k \times 1$ (where k is the number of frames in a GOP) temporal DCT is used to extract 1 low pass frame from GOP of k frames as discussed in [13].

2.1.1. *Block Partitioning*

The low pass center view obtained from the MCDCT-TF is partitioned into non overlapping blocks of size $M \times N$. It is observed that the disparity between left and right views is approximately around the 200 pixels for the natural 3D video frames. It implies that the disparity between center view and any of the left or right view will be around 100 pixels (half of the disparity between left and right). To assure the retaining of the watermarkable coefficients in the reverse rendering process, the block width of the center view for embedding should be more than 100 pixels (the disparity between center view and any of the left or right view). Keeping this fact in mind, the block size of the center view for embedding is taken as 16×256. The blocks of the low pass version of the center view are subjected to 2D dual-tree discrete wavelet transform (DT-DWT) which is discussed in the next subsection.

2.1.2. *2D Dual Tree Discrete Wavelet Transform (DT-DWT)*

As it is observed in the literature that the dual tree complex wavelet transform (DT-CWT) or its real counter part (DT-DWT) are approximately shift-invariant [12] and can be used to handle the disparity occurred due to rendering the left or right views from the center view [7]. It is also observed that the PSNR between of center view with left or right view for H_1, H_2, H_5, H_6 coefficients are higher than that of H_3, H_4 coefficients where H_1, H_2, H_3, H_4, H_5, H_6 are representing the coefficients having orientation angle $15°$, $45°$, $75°$, $-75°$, $-45°$, $-15°$ respectively for 2D-DT-DWT transformation. It is observed in the experimentation that the absolute difference between H_1 and H_2 and that of H_5 and H_6 are generally very less (approximately 5% of each other). Additionally, absolute difference between $|\frac{1}{n}\sum\limits_n (H_1 - H_6)|$ and $|\frac{1}{n}\sum\limits_n (H_2 - H_5)|$ is negligibly small (almost close to zero).

2.2. Watermark Embedding

To embed the binary watermark sequence (W), the 3^{rd} level 2D DT-DWT coefficients (say H_1^3, H_2^3, H_5^3, H_6^3) are used. Let $H_1^3 - H_2^3$ and $H_6^3 - H_5^3$ are taken as ζ_1 and ζ_2. In this embedding process, if watermark bit is "0", then $\sum\limits_n |\zeta_1|$ should be greater than $\sum\limits_n |\zeta_2|$ by atleast τ value and if the watermark bit is "1", then $\sum\limits_n |\zeta_1|$ should be less than $\sum\limits_n |\zeta_2|$ by atleast τ value where τ is a threshold and is defined as $(\sum\limits_n |\zeta_1| + \sum\limits_n |\zeta_2|) * \alpha$. In case, if watermark bit is "0" but the corresponding embedding rule ($\sum\limits_n |\zeta_1| - \sum\limits_n |\zeta_2| \geq \tau$) does not hold, the coefficient H_2^3, H_5^3 are modified to meet the requirement. Similar measure is taken for embedding bit "1". The watermark embedding process is described by the following Eq. 1

$$\left.\begin{array}{l}
if\ W_i == 0 \\
\quad if\ \sum\limits_n |\zeta_1| - \sum\limits_n |\zeta_2| \geq \tau\ then \qquad no\ change; \\
\quad else \\
\quad\quad H_{2_n}^3 = H_{2_n}^3 - (H_{1_n}^3 - H_{2_n}^3) * (\alpha + \kappa) \\
\quad\quad H_{5_n}^3 = H_{5_n}^3 + (H_{6_n}^3 - H_{5_n}^3) * (\alpha + \kappa) \\
else \\
\quad if\ \sum\limits_n |\zeta_2| - \sum\limits_n |\zeta_1| \geq \tau\ then \qquad no\ change; \\
\quad else \\
\quad\quad H_{2_n}^3 = H_{2_n}^3 + (H_{1_n}^3 - H_{2_n}^3) * (\alpha + \kappa) \\
\quad\quad H_{5_n}^3 = H_{5_n}^3 - (H_{6_n}^3 - H_{5_n}^3) * (\alpha + \kappa)
\end{array}\right\} \quad (1)$$

where α is the robustness threshold, and κ is defined as $\frac{\sum\limits_n |\zeta_1| - \sum\limits_n |\zeta_2|}{\sum\limits_n |\zeta_1| + \sum\limits_n |\zeta_2|}$. From experimental observations, the value of α is taken as 0.02.

Algorithm 1: Watermark Embedding $(V_L, V_R, D_L, D_R, \alpha, W)$

Input: V_L: Left eye video, V_R: Right eye video, D_L: Left eye video depth, D_R: Right eye video depth, α: Watermark strength, and W: Watermark bit stream

Output: V_{WL}:Watermarked left eye video, V_{WR}:Watermarked right eye video

begin

1. The center view video is synthesized using the left video V_L, left eye video depth D_L, right video V_R and the right eye video depth D_R.
2. According to the GOP size, the center view video frames are motion compensated to a single frame using the motion vector.
3. A temporal DCT is done over the motion compensated frame and low-pass frame is extracted.
4. Partition the low-pass temporal filtered frame in to non overlapping $M \times N$ blocks as described in Sec. 2.1.1.
5. **for** *each Block* **do**
   (a) Use $3^r d$ level of 2D DT-DWT and select the coefficients of $15°$, $45°$, $75°$, $-75°$, $-45°$, $-15°$ orientation angle (H_1^3, H_2^3, H_5^3, H_6^3).
   (b) Embed the watermark with H_1^3, H_2^3, H_5^3, H_6^3 using the Eq. 1
   (c) Use $3^r d$ level of inverse 2D DT-DWT to generate the temporal low-pass filtered watermarked frame
6. Use inverse temporal DCT and inverse motion compensation to generate the watermarked center view.
7. Using DIBR technique, the left and right views are inverse rendered from the watermarked center view and they are used to replace the dependent part to generate the watermarked left and right video V_{WL}, V_{WR}.
8. **return** (V_{WL}, V_{WR})

The watermarked coefficients are subjected to 3^{rd} level of 2D inverse dual-tree DWT (IDT-DWT) and subsequently to the inverse MCDCT-TF to generate the watermarked center view for the video frame. These watermarked center view is inverse rendered to generate the dependent portion of the left and right view and finally merged to the independent view to get the watermarked views. The overall embedding process is narrated in Algo. 1 and the block diagram is depicted in Fig. 1.

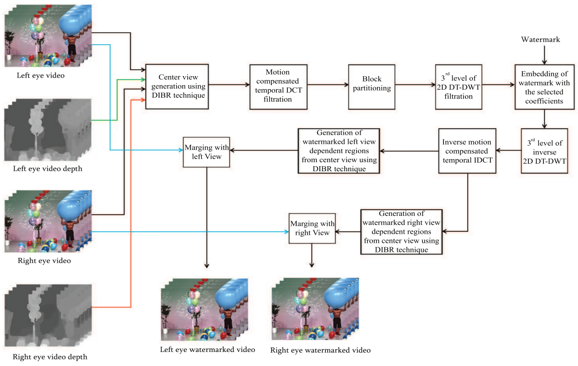

Fig. 1. Watermark embedding model

2.3. Watermark Extraction

As a view invariant watermarking, the watermark can be extracted from any views of a frame (left, right or any other synthesized views) in the proposed scheme. In the watermark extraction scheme the 3^{rd} level 2D DT-DWT is carried out over the motion compensated temporal filtered blocks of the video frames (any view). Let $H_1'^3$, $H_2'^3$, $H_5'^3$, $H_6'^3$ are the filtered high frequency watermarked coefficients of orientation angle $15°$, $45°$, $-45°$, $-15°$ and let $H_1'^3 - H_2'^3$ and $H_6'^3 - H_5'^3$ are taken as ζ_1' and ζ_2'. The watermark is extracted using ζ_1' and ζ_2' as described in Eq. 2.

$$\left.\begin{array}{l} if\ \sum\limits_{n} |\zeta_1'| > \sum\limits_{n} |\zeta_2'|\ then \\ \quad W_i' = 0 \\ else \\ \quad W_i' = 1 \end{array}\right\} \tag{2}$$

where W' is the extracted watermark. The overall watermark extraction process is narrated in Algo. 2 and the block diagram is depicted in Fig. 2.

Algorithm 2: Watermark Extraction (V_W)

Input: V_W: Watermarked video
Output: W':Extracted Watermark
begin

1. According to the GOP size, the video frames of Watermarked video (V_W) are motion compensated to a single frame using the motion vector.
2. A temporal DCT is done over the motion compensated frame and low-pass frame is extracted.
3. Partition the low-pass temporal filtered frame in to non overlapping $M \times N$ blocks as described in Sec. 2.1.1.
4. **for** *each Block* **do**
   (a) Use 3^rd level of 2D DT-DWT and select the coefficients of $15°$, $45°$, $75°$, $-75°$, $-45°$, $-15°$ orientation angle (H_1^3, H_2^3, H_5^3, H_6^3).
   (b) Extract the watermark W' by comparing H_1^3, H_2^3, H_5^3, H_6^3 using the Eq. 2
5. **return** (W')

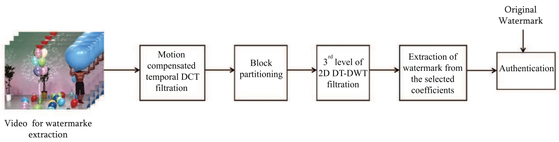

Fig. 2. Watermark extraction model

3. EXPERIMENT RESULTS

For the experimental purpose, the proposed scheme is tested over a set standard 3D video sequences (for example, Balloons, Champagne_tower, Kendo, MicroWorld, Shark etc.) with different camera views. The experiment results of the proposed scheme is compared with the Lee's scheme [5] and Rana's scheme [6] as these two schemes are most recent watermarking scheme for DIBR based 3D video sequence.

3.1. Visual Quality

To evaluate the visual quality of the proposed scheme, Peak Signal-to-Noise Ratio (PNSR), Structural Similarity (SSIM) and Visual Information Fidelity pixel domain (VIFp) are measured. The comparison results for proposed scheme with Lee's scheme [5] and Rana's scheme [6] are tabulated in Table 1 for these visual quality measures. From Table 1, it is observed that the proposed scheme gives almost comparable result for the PSNR, SSIM, VIFp against the Lee's scheme [5]. It shows a bit inferior results against Rana's scheme [6] as only independent part of the left or right view is alter in the Ranas scheme [6].

		Average PSNR	Average SSIM	Average VIFp
Proposed scheme	Balloons	45.12	0.9980	0.9896
	Shark	45.22	0.9984	0.9899
Lee's scheme	Balloons	45.35	0.9978	0.9882
	Shark	44.24	0.9973	0.9861
Rana's scheme	Balloons	49.55	0.9985	0.9912
	Shark	52.14	0.9982	0.9904

Table 1. Average PSNR, SSIM, VIFp comparison result of the proposed scheme with Lee's scheme [5] and Rana's [6]

3.2. Robustness

The robustness of the proposed scheme is evaluated against 3D HEVC compression by measuring the hamming distance between original and extracted watermark. To show the robustness against synthesis view attack, watermark is extracted from intermediate synthesis views generated by left and right view. For experimental purpose, camera views 2 and 4 is taken as left and the right view. Camera views 2 and 4 are used to generate the synthesis camera views 2.5, 3, 3.5 video sequence using DIBR technique. Table 2 tabulates the comparison result of the proposed scheme with Lee's scheme [5] and Rana's scheme [6] for Balloons video and Shark video of camera views 2, 2.5, 3, 3.5, 4.

Video	View number	Proposed scheme	Lee's scheme	Rana's scheme
Balloons video	2	0.1811	0.3846	0.1253
	4	0.1781	0.3696	0.1024
	3	0.0926	0.6612	0.6011
	2.5	0.1304	0.4805	0.5126
	3.5	0.1253	0.4712	0.5211
Shark Video	2	0.1755	0.4236	0.0825
	4	0.1719	0.3983	0.0787
	3	0.0892	0.5824	0.5686
	2.5	0.1223	0.5102	0.5482
	3.5	0.1205	0.4996	0.4562

Table 2. Hamming distance comparison result of the proposed scheme with Lee's scheme [5] and Rana's [6] for Balloons video and Shark video

It is observed from the Table 2 that proposed scheme outperforms other two schemes, Lee's scheme [5] and Rana's scheme [6], for synthesized views (i.e. for 3, 2.5 and 3.5). It gives a bit inferior results than the Rana's scheme [6] when extracting from original embedded views (i.e. 2 and 4 views) but outperforms the Lee's scheme [5]. To test the robustness of the proposed scheme against collusion attack, the left (and the right) view video is rendered to right (and left) view video and colluded with the right (and the left) eye video. The extracted watermark from the colluded video is depicted in Table 3. It is observed that the proposed scheme performs better against collusion attack than Lee's scheme [5] and Rana's scheme [6] because similar watermarks are embedded in the homogeneous (depended) regions for the left and right views.

Video	Proposed scheme	Lee's scheme	Rana's scheme
Balloons	0.1961	0.4216	0.4141
Shark	0.1902	0.4656	0.3988

Table 3. Average hamming distance comparison result of left ans right view of the proposed scheme with Lee's scheme [5] and Rana's scheme [6] after collusion attack using DIBR technique for Balloons video and Shark video

4. CONCLUSION

In this paper, a shift invariant DT-DWT based watermarking scheme has been proposed to resist DIBR based view synthesis process. Motion compensated temporal filtering is used over the synthesized center views to make the scheme robust against 3D-HEVC compression. To improve the robustness, 3^{rd} level 2D DT-DWT has been used for watermark embedding and selected coefficients are perturbed to make the scheme invariant to view synthesis process and collusion attack. The experiment results demonstrate the applicability of the proposed scheme over the existing methods.

REFERENCES

[1] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision*, vol. 8, no. 3, pp. 33, 2008.

[2] Robert Patterson, "Human factors of 3-d displays," *Journal of the Society for Information Display*, vol. 15, no. 11, pp. 861–871, 2007.

[3] C. Fehn and R.S. Pastoor, "Interactive 3-dtv-concepts and key technologies," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 524–538, March 2006.

[4] Ying Chen, Gerhard Tech, Krzysztof Wegner, and Sehoon Yea, *Test Model 8 of 3D-HEVC and MV-HEVC*, Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3, jct3v-h1003 edition, Mar.-Apr. 2014.

[5] Min-Jeong Lee, Ji-won Lee, and Heung-Kyu Lee, "Perceptual watermarking for 3d stereoscopic video using depth information," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Seventh International Conference on*, Oct 2011, pp. 81–84.

[6] Shuvendu Rana and Arijit Sur, "Blind 3d video watermarking based on 3d-hevc encoder using depth," in *Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing*, New York, 2014, ICVGIP '14, pp. 52:1–52:8, ACM.

[7] Hee-Dong Kim, Ji-Won Lee, Tae-Woo Oh, and Heung-Kyu Lee, "Robust dt-cwt watermarking for dibr 3d images," *Broadcasting, IEEE Transactions on*, vol. 58, no. 4, pp. 533–543, Dec 2012.

[8] Patrizio Campisi, "Object-oriented stereo-image digital watermarking," *Journal of Electronic Imaging*, vol. 17, no. 4, pp. 043024–043024–5, 2008.

[9] Yu-Hsun Lin and Ja-Ling Wu, "A digital blind watermarking for depth-image-based rendering 3d images," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 602–611, June 2011.

[10] E. Garcia and J. Dugelay, "Texture-based watermarking of 3d video objects," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 8, pp. 853–866, Aug 2003.

[11] P. Vinod and P.K. Bora, "Motion-compensated inter-frame collusion attack on video watermarking and a countermeasure," *Information Security, IEE Proceedings*, vol. 153, no. 2, pp. 61 – 73, june 2006.

[12] Ryan Anderson, Nick Kingsbury, and Julien Fauqueur, "Determining multiscale image feature angles from complex wavelet phases," in *Image Analysis and Recognition*, Mohamed Kamel and Aurlio Campilho, Eds., vol. 3656 of *Lecture Notes in Computer Science*, pp. 490–498. Springer Berlin Heidelberg, 2005.

[13] Shuvendu Rana, Nilkanta Sahu, and Arijit Sur, "Robust watermarking for resolution and quality scalable video sequence," *Multimedia Tools and Applications*, pp. 1–30, 2014.

[14] R. Atta and M. Ghanbari, "Spatio-temporal scalability-based motion-compensated 3-d subband/dct video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 1, pp. 43 – 55, jan. 2006.