Automatic Stereoscopic Video Object-Based Watermarking Using Qualified Significant Wavelet Trees†

Klimis S. Ntalianis[1], Paraskevi D. Tzouveli[1] and Athanasios S. Drigas[2]

[1]National Technical University of Athens, Electrical and Computer Engineering Department, 15773, Athens, Greece
[2]Net Media Lab, NCSR Demokritos, Athens, Greece

ABSTRACT

In this paper a fully automatic scheme for embedding visually recognizable watermark patterns to video objects is proposed. The architecture consists of 3 main modules. During the first module unsupervised video object extraction is performed, by analyzing stereoscopic pairs of frames. In the second module each video object is decomposed into three levels with ten subbands, using the Shape Adaptive Discrete Wavelet Transform (SA-DWT) and three pairs of subbands are formed (HL_3, HL_2), (LH_3, LH_2) and (HH_3, HH_2). Next Qualified Significant Wavelet Trees (QSWTs) are estimated for the specific pair of subbands with the highest energy content. QSWTs are derived from the Embedded Zerotree Wavelet (EZW) algorithm and they are high-energy paths of wavelet coefficients. Finally during the third module, visually recognizable watermark patterns are redundantly embedded to the coefficients of the highest energy QSWTs and the inverse SA-DWT is applied to provide the watermarked video object. Performance of the proposed video object watermarking system is tested under various signal distortions such as JPEG lossy compression, sharpening, blurring and adding different types of noise. Furthermore the case of transmission losses for the watermarked video objects is also investigated. Experimental results on real life video objects indicate the efficiency and robustness of the proposed scheme.

Keywords: video object (VO), Shape Adaptive Discrete Wavelet Transform, visually recognizable watermark pattern, Qualified Significant Wavelet Tree.

1. INTRODUCTION

During the last decade the significant improvement of PCs' computational power and the rapid growth of low-cost portable devices, allowed for easy manipulation, replication and distribution of digital media. However the large amounts of visual information have led to an emerging need for copyright protection of intellectual property. To confront this problem digital watermarking has been proposed as a means to identify the owner of the digital data and detect illegal distribution paths. The watermarking process encodes hidden copyright information into the digital media, by modifying the original data either in the spatial or in the frequency domain.

On the other hand, MPEG-4 has introduced the concept of Video Objects (VOs), which may correspond to semantic entities. Such an object-based representation is very useful for a variety of applications including retrieving and indexing of visual information, efficient image/video coding and image/video editing. These semantic entities make the produced content far more reusable and flexible, leading to a migration from a frame-based to an object-based consideration of digital media [1]. According to the aforementioned concepts, semantic object-based watermarking schemes can allow for hierarchical protection of media content and provide several new functionalities compared to frame-based approaches.

Till now most of the digital watermarking algorithms are frame-oriented. Early techniques embed the watermark in the least significant bits (LSBs) of image pixel [2]. However, this technique and some other proposed improvements [3], [4], except of having relatively low-bit capacity, they are also not resistant enough to lossy image compression, cropping and other image processing attacks. On the contrary, frequency-domain-based techniques are more robust to attacks. In particular Cox et. al. [5] embed a set of i.i.d. sequences, following a Gaussian distribution, into the perceptually most significant frequency components of an image. In [6], visually recognizable patterns are embedded, by selectively modifying the middle frequencies of the image obtained using the DCT transform. Other approaches such as [7], [8], [9] use the Discrete Wavelet Transform (DWT) to hide data in the frequency domain. In most of the aforementioned techniques the watermark is a random sequence of bits and can be detected only by employing a detection theory scheme. Furthermore, all the aforementioned approaches are frame-based and thus semantically meaningful video objects composing a frame may not be sufficiently protected.

On the other hand, limited work has been done in literature towards object-based watermarking. Starting from early works, in [10] a digital watermarking scheme of objects is proposed, based on the 2-D/3-D shape adaptive discrete wavelet transform with Arbitrary Regions Of Support (AROS). The watermark in this scheme is an i.i.d. Gaussian distributed vector variable, added to all high-pass bands of an object in the wavelet domain. In [11], the embedding scheme exploits the shape of video objects and the watermark is a random sequence transformed to fit the scale and orientation of them. However, both approaches use no segmentation algorithm, assuming that VOs are pre-segmented, i.e., they are a priori available. In [12], a motion oriented segmentation algorithm is used to detect VOs and the watermark is a pseudorandom sequence, embedded to the DCT coefficients in an 8×8 block resolution. Nevertheless in this approach, the detected objects are motion regions and therefore this scheme cannot be straightforwardly applied to

† This work is an extension of our previous work: K. S. Ntalianis, N. D. Doulamis, A. D. Doulamis and S. D. Kollias, "Automatic Stereoscopic Video Object-Based Watermarking Using Qualified Significant Wavelet Trees," in Proc. of the *IEEE Intern. Conf. on Consumer Electronics* (ICCE'02), L.A, USA, June 2002.

images where no motion information is available. In the work of [13] a cocktail watermarking technique is proposed where again the watermark is an i.i.d. Gaussian distribution. The proposed system incorporates low-level texture segmentation for object detection, which faces difficulties in correctly separating semantically meaningful entities. Furthermore, in all the aforementioned approaches, the watermark is a random variable sequence instead of a visually recognizable pattern.

Additionally, considering the most recent works, in [14] image segmentation is performed and the extracted salient spatial features are used as reference for compensating usual geometric attacks. During segmentation the largest segments are selected for watermark embedding. In [15] a binary watermark image is embedded into multi-scale feature point-based local characteristic regions in the transform domain. For synchronization purposes characteristic regions are first detected, using SIFT and image normalization. However the detected regions in both schemes may not correspond to semantically meaningful video objects. In [16] watermark synchronization is accomplished by invariant local feature regions, centered at scale-space feature points. Affine covariant regions (ACRs), local circular regions (LCRs), Tchebichef moments, and local Tchebichef moments (LTMs) are investigated to embed and detect the watermark. However all these kinds of regions usually do not efficiently detect semantics. In [17] scale and affine invariant regions are extracted on a de-noised image and non-overlapping invariant regions are selected. The watermark is embedded surrounding the selected invariant regions. In [18] a watermarking scheme is proposed to enhance the capability of resisting VQ attacks, by partitioning the image into interrelated regions with irregular shapes. The authentication watermark is generated by a feedback-based chaotic system and inserted into the two LSBs of the region pixels, while recovery is achieved by a reference sharing mechanism on the encoded DCT coefficients. However interrelated regions also may not correspond to semantically meaningful video objects, while LSB insertion is not robust to signal processing manipulations. In [19] a fragile, blind, high payload capacity, Region of Interest Medical image watermarking (MIW) technique is proposed for grayscale medical images. Aim of this work is to maintain Electronic Patient Report (EPR)/DICOM data privacy and medical image integrity. Another relevant work is presented in [20], where a lossless watermarking technique for Ultrasound images is proposed, based on difference expansion (DE). Its main aim is to hide patient's data and protect the ROI by a tamper detection method. However in both schemes ROIs are detected manually, while watermarks are not designed to survive under signal processing manipulations. The work in [21] focuses on surveillance video, in order to protect its integrity against later manipulation. The system is based on an array construction method using seed sequences, which allows a simple hardware-generated watermark to be inserted into a surveillance camera video stream in realtime within the camera itself. The watermark changes in every frame, and it is possible to infer lost frames. By ceding to the camera the task ROIs detection can be accomplished, however ROIs are not watermarked in order to maintain the local integrity of the video frame. Finally in [22] an Angle Quantization Index Modulation-based watermarking scheme is proposed that considers the statistical behaviour of the region, where a message bit is to be embedded before settling on the size of the quantization step. However in this case regions are just 8×8 pixel blocks without any explicit meaning, which are distinguished based on texture estimation (homogeneous and highly-textured regions).

In this paper, a fully automatic video object-based watermarking scheme is designed and implemented for stereoscopic video sequences. Two of the main contributions of the proposed approach are: (a) The scheme is fully automatic, using an efficient unsupervised video object segmentation scheme, which exploits depth information and (b) visually recognizable patterns such as binary, grayscale or color images are embedded to each video object, in contrast to existing object-based approaches. Thus selection of experimental thresholds during watermark detection is avoided, as the retrieved watermark is recognizable. In particular in the proposed approach initially video objects are unsupervisedly extracted by incorporating the method proposed in [23]. Then, each unsupervisedly extracted video object is decomposed into three levels by applying a shape adaptive discrete wavelet transform (SA-DWT) [24], providing ten subbands. Afterwards, three pairs of subbands are formed, (HL_3, HL_2), (LH_3, LH_2) and (HH_3, HH_2) and the pair with the highest energy content is selected. For this pair Qualified Significant Wavelet Trees (QSWTs) are detected [7] in order to select the coefficients where the watermark should be casted. QSWTs, which are based on the definition of the Embedded Zerotree Wavelet (EZW) algorithm [25], are high-energy paths of coefficients within the selected pair of subbands and enable adaptive casting of watermark energy in different resolutions, achieving watermark robustness. Then, the watermark pattern is redundantly embedded to both subbands of the selected pair, using a non-linear insertion procedure that adapts the watermark to the energy of each wavelet coefficient. Finally, the inverse SA-DWT is applied to produce the watermarked video object. Differences between the original and watermarked video objects are imperceptible to human eyes, while watermarked video objects are robust under different combinations of image processing attacks and transmission losses. Experimental results exhibit the efficiency of the proposed automatic video object-watermarking scheme, in real world sequences.

2. COMPARISON TO OUR PREVIOUS WORK

In this paper we significantly extend and enhance the results and overall presentation of our previous work (see footnote of first page). In particular:

(a) Regarding Section 3 (Unsupervised Video Object Segmentation) an analytical description of the tube-embodied GVF field method is provided. Additionally, compared to our previous work, the extraction technique is applied to several stereoscopic frames (two of which are presented) and not only to one test image.

(b) Regarding Section 4 (Shape Adaptive Discrete Wavelet Transform and Qualified Significant Wavelet Trees), a figure is created (Figure 2) to illustrate the decomposition process and more details are provided, explaining the detection of significant wavelet coefficients. Also another important contribution of our current work is the introduction of "In-Nodes".

(c) Section 5 (Video Object Watermarking: Embedding and Extraction Methods) contains two new figures (Figure 4 and Figure 5) that make much clearer the processes of watermark embedding and extraction. Furthermore the typical QSWTs detection algorithm [7] is modified to take into consideration "In-Nodes", which essentially accelerate the detection process.

(d) Section 6 (Experimental Results) is significantly extended by performing several new experiments. More specifically in our initial work we have tested watermark robustness under salt & pepper noise, Gaussian noise, blur, JPEG lossy compression and sharpening. In this paper many new mixed attacks are included, which present significant interest to the research community: combination of sharpening and bluring, combination of sharpening and bluring under JPEG compression and combination of different JPEG compression ratios under various Bit Error Rates, corresponding to typical mobile radio channels. All these new results provide a more integrated aspect of the presented methodology, leading to clearer conclusions of the advantages of the proposed watermarking scheme.

(e) Finally in terms of bibliography (Section 8), this paper follows the terminology of the updated version of the ISO/IEC 14496-2:2004 standard that focuses on "coding of audio-visual objects". Furthermore regarding state-of-art completeness, this paper contains fourteen new references that extensively cover the topic of video object watermarking.

3. UNSUPERVISED VIDEO OBJECT SEGMENTATION

The first module of the proposed watermarking scheme includes an efficient unsupervised video object segmentation process. In this paper, the fast and accurate video object extraction method proposed in [23] is incorporated, which exploits depth information. In particular, for each stereoscopic pair of frames initially the disparity field is computed followed by an occlusion detection and compensation algorithm [26]. This procedure leads to the estimation of an occlusion compensated depth map. Afterwards, a segmentation algorithm is applied to the depth map, providing a depth segments map. Depth information is an important feature for content description, since usually video objects are composed of regions located on the same depth plane [27]. However, object boundaries (contours) cannot be identified with high accuracy by a depth segmentation algorithm, mainly due to erroneous estimation of the disparity field, even after it has been improved by the occlusion detection and compensation algorithm. For this reason, contours of depth segments are refined using a Gradient Vector Flow (GVF) scheme [28]. In particular, inside each depth segment, a GVF field is computed, which is incorporated for adjusting the depth segment's contour. In our case, the GVF field is estimated only within a tube region, leading to a significant reduction of the computational complexity. The "out"-boundary of the tube coincides to the depth segment's contour, while the "in"-boundary is constructed by shrinking the "out"-boundary using an edge map constraint [23]. Finally an active contour is unsupervisedly initialized onto the "out"-boundary of the tube and is guided to the video object's boundary, driven by the tube-embodied GVF field. In Figure 1 the tube generation for a frame of the "Eye to Eye" sequence and a frame of the standard sequence "Claude" is presented, while in Figures 1(c,f) the detected video objects are shown.

4. SHAPE ADAPTIVE DISCRETE WAVELET TRANSFORM AND QUALIFIED SIGNIFICANT WAVELET TREES

The shape-adaptive discrete wavelet transform (SA-DWT) was proposed in [24], for efficiently coding arbitrarily shaped visual objects. The SA-DWT transforms the samples in an arbitrarily shaped region into the same number of coefficients in the subband domain, while preserving the spatial correlation, locality and self-similarity across subbands. Furthermore for a rectangular region, the SA-DWT becomes identical to the conventional wavelet transform. In the framework of video object watermarking, where regions of arbitrary shape are considered, the SA-DWT should be adopted as it is contour-sensitive, providing exact values of the wavelet coefficients at the border of each video object. On the contrary, the conventional DWT provides wavelet coefficients of usually higher (than the real) values in the borders of video objects, since the area around video objects (background area) is also considered. Thus more reliable QSWTs are detected when using the SA-DWT.

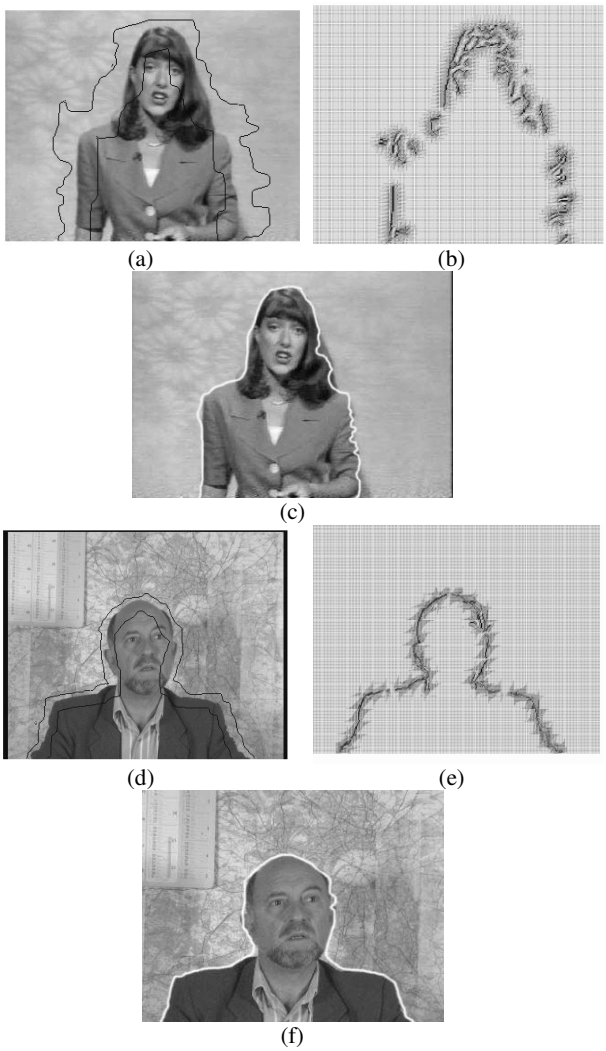


Figure 1: Video object segmentation using the proposed tube-embodied Gradient Vector Flow (GVF) method [23]. (a,d) The "out" and "in" boundaries of the tubes around the video objects for a frame of the "Eye to Eye" sequence and the "Claude" sequence. (b,e) The gradient vector flow fields inside the tubes. (c,f) The detected video objects.

By applying the SA-DWT once to an area of arbitrary shape, four parts of high, middle, and low frequencies, i.e. LL_1,

HL_1, LH_1, HH_1, are produced. Band LL_1 (HH_1) includes low (high) frequency components both in horizontal and vertical direction, while the HL_1 (LH_1), includes high (low) frequencies in horizontal direction and low (high) frequencies in vertical direction. Subband LL_1 can be further decomposed in a similar way into four different subbands, denoted as LL_2, HL_2, LH_2, HH_2 respectively. This process can be repeated several times, depending on the specific application. An example of video object decomposition into three levels with ten subbands using the SA-DWT is depicted in Figure 2. In this figure, a parent-child relationship is defined between wavelet coefficients at different scales, corresponding to the same location. For example, the subbands LH_3, LH_2, LH_1 follow a parent-child relationship. The coefficient at the highest level is called the parent, and all coefficients corresponding to the same spatial location at the lower levels of similar orientation are called children. For a given parent, the set of all coefficients at all finer scales of similar orientation corresponding to the same location
are called descendants. In Figure 2 arrows point from parent pixels/subbands to the respective children pixels/subbands. The wavelet coefficients can be distinguished into two types; the "In-Node" coefficients which belong to the video object area and the "Out-Node" coefficients which do not belong to the video object.

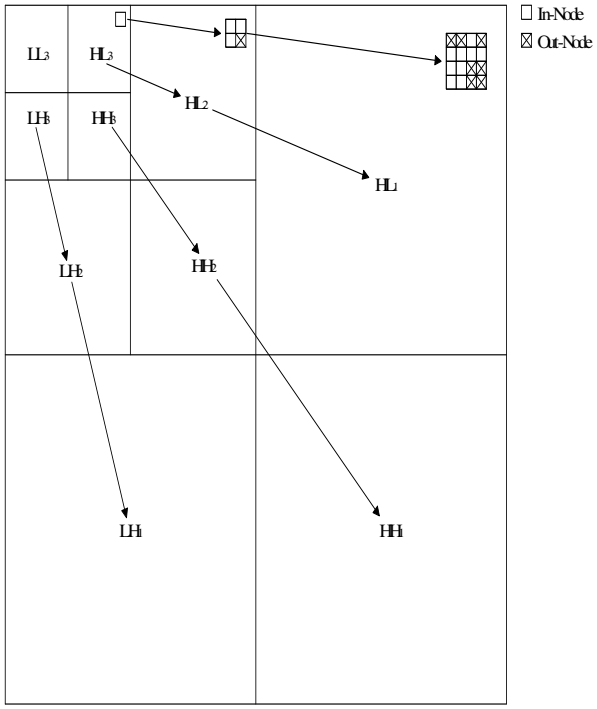


Figure 2: SA-DWT decomposition of a video object. Arrows point from parent subbands to respective children subbands. "In-Node" coefficients are pixels that belong to the video object. "Out-Node" coefficients do not belong to the video object.

In the proposed video object watermarking scheme, coefficients with local information in the subbands are chosen as target coefficients for casting the watermark. Coefficients selection is based on Qualified Significant Wavelet Trees (QSWTs) derived from the Embedded Zerotree Wavelet algorithm (EZW) and the necessary definitions are given below.

Definition 1: An "In-Node" wavelet coefficient $x_n(i,j) \in D$ is a parent of $x_{n-1}(p,q)$, where D is a subband labeled HL_n, LH_n, HH_n, $p=2i-1|2i$, $q=2j-1|2j$, $n>1$, $i>1$ and $j>1$. Symbol | corresponds to the OR-operator. The $x_{n-k}(p,q)$ are called descendants of $x_n(i,j)$, for $1 \le k < n$.

Definition 2: If an "In-Node" wavelet coefficient $x_n(i,j)$ and all its descendants $x_{n-k}(p,q)$ for $1 \le k < n$ satisfy $|x_n(i,j)|<T$, $|x_{n-k}(p,q)|<T$ for a given threshold T, $\forall$ $p=2i-1|2i$, $q=2j-1|2j$, then the tree $x_n \rightarrow x_{n-1} \ldots \rightarrow x_{n-k}$ is called wavelet zerotree [25].

Definition 3: If an "In-Node" wavelet coefficient $x_n(i,j)$ satisfies $|x_n(i,j)|>T$, for a given threshold T, then $x_n(i,j)$ is called a significant coefficient [25].

Definition 4: If an "In-Node" wavelet coefficient $x_n(i,j) \in D$, where D is one of the subbands labeled HL_n, LH_n, HH_n, satisfies $|x_n(i,j)|>T_1$ and its "In-Node" children $x_{n-1}(p,q)$ satisfy $|x_{n-1}(p,q)|>T_2$, for given thresholds T_1 and T_2, $\forall$ $p=2i-1|2i$, $q=2j-1|2j$, then the "In-Node" parent $x_n(i,j)$ and its "In-Node" children $x_{n-1}(p,q)$ are called a Qualified Significant Wavelet Tree (QSWT).

5. VIDEO OBJECT WATERMARKING: EMBEDDING AND EXTRACTION METHODS

After unsupervised video object extraction is accomplished, each video object is decomposed into three levels with ten subbands, using the SA-DWT. In the following, the watermark image, which is a visually recognizable pattern, is redundantly embedded to the host video object, by modifying the QSWT coefficients of one of its subband pairs. In particular, for each video object three pairs of subbands are examined (since the decomposition is performed into three levels) for possible watermark casting; pair P_1 consisting of subbands (HL_3, HL_2), pair P_2 of subbands (LH_3, LH_2) and finally pair P_3 of (HH_3, HH_2). The pair of the highest energy content compared to the other two pairs is selected as the most appropriate for watermark casting. Let us denote as E_{P_k} the energy of P_k, k=1,2,3, which is defined as the sum of the squares of "In-Node" wavelet coefficients of the respective pair of subbands P_k.

$$E_{p_k} = \sum_i \sum_j [x_3(i,j)]^2 + \sum_p \sum_q [x_2(p,q)]^2 \quad k=1,2,3 \tag{1}$$

where $x_3(i,j)$ is an "In-Node" wavelet coefficient of the respective subband, $x_3(i,j) \in R_k$, k=1,2,3, with $R_1=HL_3$, $R_2=LH_3$ and $R_3=HH_3$. Similarly, $x_2(p,q) \in S_k$, k=1,2,3, with $S_1=HL_2$, $S_2=LH_2$ and $S_3=HH_2$.

Then the most appropriate pair for watermark casting is selected as the one that maximizes the energy,

$$\hat{k} = \arg\max_{k=1,2,3,} E_{p_k} \tag{2}$$

5.1 The Watermark Embedding Method

After selecting the pair of subbands with the highest energy content, QSWTs are detected for the selected pair $P_{\hat{k}}$ and the visually recognizable watermark is cast by modifying the values of the detected QSWTs. In order to estimate the QSWTs, we need to determine the two threshold parameters T_1, T_2. In our case, the average values over all "In-Node" wavelet coeffi-

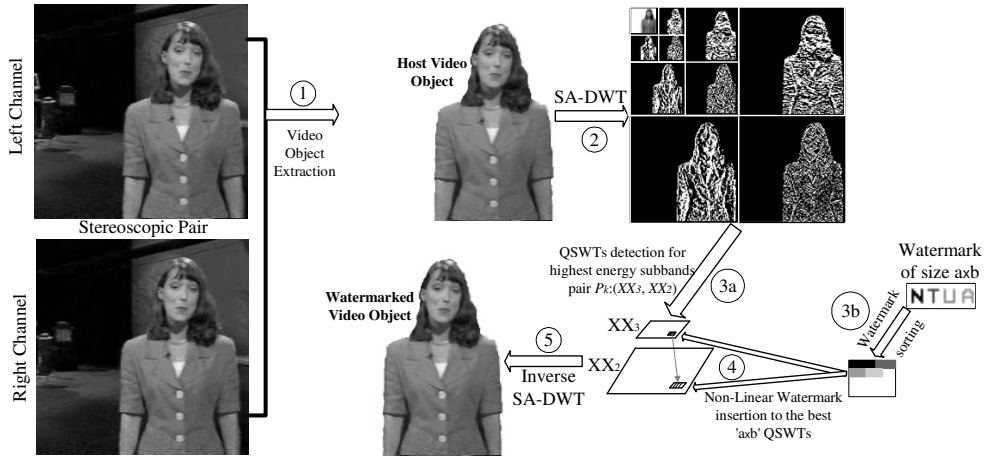


Figure 4: Watermark embedding method.

cients of the respective subbands of the selected pair are used as threshold values,

$$T_1 = \frac{1}{N_1}\sum_i\sum_j \left|x_3(i,j)\right|, \text{ "In-Node" } x_3(i,j) \in R_{\hat{k}} \quad (3a)$$

$$T_2 = \frac{1}{N_2}\sum_p\sum_q \left|x_2(p,q)\right|, \text{ "In-Node" } x_2(i,j) \in S_{\hat{k}} \quad (3b)$$

where N_1 (N_2) is the number of "In-Node" wavelet coefficients of $R_{\hat{k}}$ ($S_{\hat{k}}$).

QSWTs are detected using these threshold values. To better clarify the proposed algorithm, a piece of pseudo-code is given in Figure 3.

```
t=0
  QSWT[t]=∅
    For i=1 to N     /* N x M is the size of subband LH3. */
          For j=1 to M
             If x3(i,j)∈LH3 is "In-Node" AND x3(i,j)≥T1
                If {
                    x2(2*i-1, 2*j-1) ∈LH2 is "In-Node"
                    AND x2(2*i-1, 2*j-1)≥T2
                    AND x2(2*i-1, 2*j) ∈LH2  is "In-Node"
                    AND x2(2*i-1, 2*j) ≥T2
                    AND x2(2*i, 2*j-1) ∈LH2  is "In-Node"
                    AND x2(2*i, 2*j-1) ≥T2
                     AND x2(2*i, 2*j) ∈LH2 is "In-Node"
                     AND x2(2*i, 2*j) ≥T2
                   }
                    QSWT[t]=x3(i, j) + x2(2*i-1, 2*j-1) +
                    +x2(2*i-1, 2*j)+ x2(2*i, 2*j-1)+ x2(2*i, 2*j)
                    t=t+1
                End If
             End If
          End For j
     End For i
```

Figure 3: Pseudo-code for estimating qualified significant wavelet trees (QSWTs).

For simplicity purposes, we assume that pair P_2 has been selected and that all four children $x_2(p,q)$ of parent $x_3(i,j)$ are "In-Node" wavelet coefficients. Other cases, where only some of the children are "In-Node" coefficients, can be addressed in a similar way.

Assuming that the watermark pattern is of size axb, the axb largest values of array QSWT[t] (see Figure 3) are selected to cast the watermark. Let us assume that the $x_3(m,l)$ wavelet coefficient is the nth significant value of array QSWT[t], with $n \le a \cdot b$. Then, the value of $x_3(m,l)$ is modified as

$$x_3'(m,l) = x_3(m,l)\cdot(1 + c_3 \cdot w(m',l')) \quad (4)$$

where $w(m',l')$ is the nth greatest gray-scale value of the digital watermark, c_3 is a scaling constant that balances the robustness of watermark casting and $x_3'(m,l)$ is the modified wavelet coefficient. As is observed, the nth significant value of array QSWT[t] is modified by the nth greatest value of the watermark image. Small values of $x_3(m,l)$ are modified by small values of the watermark image to avoid image artifacts, while when $x_3(m,l)$ is large the watermark energy is increased for robustness.

The child coefficient of $x_3(m,l)$ is modified in a similar way. In particular, among all children of $x_3(m,l)$ the child with the maximum wavelet coefficient is selected and used for watermark casting.

$$x_2'(r,s) = x_2(r,s)\cdot(1 + c_2 \cdot w(m',l')) \quad (5)$$

where $x_2(r,s)$ is the child of $x_3(m,l)$ with the maximum wavelet coefficient value:

$$x_2(r,s) = \max\{\, x_2(2m-1,2l-1),\, x_2(2m,2l-1),\, x_2(2m-1,2l),\, x_2(2m,2l)\} \quad (6)$$

In the previous equation and without loss of generality, we have assumed that all children of $x_3(m,l)$ are "In-Node" wavelet coefficients.

Finally the inverse SA-DWT is applied to the modified and

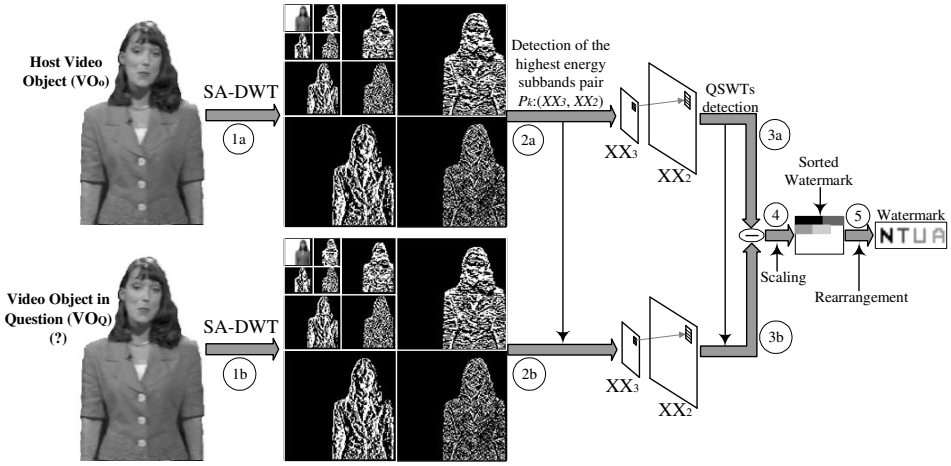


Figure 5: Watermark extraction method.

unchanged subbands to form the watermarked video object. A graphical representation of the watermark embedding method can be seen in Figure 4. In this figure arrows show the flow direction of the process from step 1 (video object extraction) to step 5 (inverse SA-DWT).

5.2 The Watermark Extraction Method

The watermark extraction method uses the original host video object and the scaling constants (c_2 and c_3) to extract the watermark pattern (if present) from the video object under question. Towards this direction the following steps are performed:

Step 1: Initially the original video object V_o and the video object under question V_q are decomposed into three levels with ten subbands using the SA-DWT,

$$V_{o,w} = \text{SA-DWT}(V_o) \tag{7a}$$

$$V_{q,w} = \text{SA-DWT}(V_q) \tag{7b}$$

where $V_{o,w}$ and $V_{q,w}$ correspond to the shape adaptive wavelet transforms of video objects V_o and V_q respectively.

Step 2: The highest energy pair of subbands is detected for the video object $V_{o,w}$ using equation (2). Let us assume that x_3^o is one of the $a \cdot b$ most significant wavelet coefficients of object $V_{o,w}$ in the selected subband pair of the third decomposition level. The respective wavelet coefficient of object $V_{q,w}$ is denoted as x_3^q. Then, the watermark is extracted by solving equation (4) with respect to the gray-scale values of the watermark

$$\hat{w}_3 = (x_3^q - x_3^o)/(x_3^o \cdot c_3) \tag{8}$$

where $\hat{w}_3$ refers to respective estimated gray-scale value of the watermarked pattern as is obtained from the third resolution level. This value may differ from the original watermark values, since several image processing attacks can be performed on the watermarked image.

Similarly, we can estimate the same watermark value from the second decomposition level.

$$\hat{w}_2 = (x_2^q - x_2^o)/(x_2^o \cdot c_2) \tag{9}$$

where x_2^q, x_2^o are the respective wavelet coefficients at second decomposition level of $V_{o,w}$ and $V_{q,w}$.

Step 3: The estimated gray-scale values $\hat{w}_3$ and $\hat{w}_2$ are first averaged and then rearranged to form the watermark pattern. Rearrangement is performed since the values of the watermark pattern have been sorted before watermark casting as described in Section IV. A graphical representation of the watermark extraction method can be seen in Figure 5. In this Figure arrows show the flow direction of the process from step 1 (video objects SA-DWT) to step 5 (watermark rearrangement).

6. EXPERIMENTAL RESULTS

The effectiveness and robustness of the proposed video object watermarking system has been extensively tested under various image processing attacks, using real life stereoscopic video sequences. In Figures 6 and 7 two frames of the "Eye2Eye" sequence are presented, which are used for evaluation purposes. This sequence is a stereoscopic television program of about 25 minutes total duration (12,739 frames at 10 frames/sec) and was produced in the framework of the ACTS MIRAGE project in collaboration with AEA Technology and ITC. Left channels of the stereoscopic pairs are depicted in Figures 6(a) and 7(a), while the unsupervisedly extracted foreground video objects are depicted in Figures 6(b) and 7(b), using the method reported in Section II. In the performed ex-

periments, a grayscale image of size 6×20 pixels containing the characters "NTUA" is used as watermark pattern for the host video object of Figure 6(b) and a binary image of size 8×22 with characters "IVML" for the host video object of Figure 7(b). These two watermark patterns are shown in Figures 6(c) and 7(c).

(a)

(b) (c) (d)

Figure 6: (a) Original left channel of a stereoscopic pair (b) Unsupervisedly extracted original foreground video object (c) Watermark pattern and (d) Watermarked video object.

Then according to the sizes of the watermark images, the top 120 values and the top 176 values of QSWTs are selected for embedding the watermarks in the first and second case respectively. Furthermore for simplicity in our experiments c_2 and c_3 are constants in all frequency bands and equal to c_2=0.1 and c_3=0.15 respectively. The watermarked video objects are depicted in Figures 6(d) and 7(d).

(a)

(b) (c) (d)

Figure 6: (a) Original left channel of a stereoscopic pair (b) Unsupervisedly extracted original foreground video object (c) Watermark pattern and (d) Watermarked video object.

As it can be observed in both cases the embedded watermarks are imperceptible. Additionally Table I contains the extraction results from Figures 6(d) and 7(d) (without any attacks) using the proposed method. In the same table the PSNR values of the video objects after embedding the watermark patterns are also provided. In the performed experiments PSNR is computed by:

$$PSNR = 10\log_{10}\frac{255^2}{\frac{1}{a(V_o)}\sum_i\sum_j\left[V_o(i,j)-V_q(i,j)\right]^2} \quad (10)$$

where $a(\cdot)$ is a function that returns the number of pixels of an arbitrarily shaped region and $V_o(i,j)$, $V_q(i,j)$ are the pixel values of objects V_o and V_q respectively.

TABLE I
WATERMARK EXTRACTION FROM VIDEO OBJECTS OF FIGURES 6(d) AND 7(d) WITHOUT ATTACK

	1st case	2nd case
Embedded watermark		
PSNR (dB)	44.3	45.7
Extracted watermark		

In the following, the robustness of the proposed system under various attacks such as JPEG lossy compression, gaussian noise, blurring, sharpening and lossy transmission is investigated. Furthermore, an objective criterion is used to evaluate how close is the extracted watermark image to the original one. In our case, the correlation coefficient is selected as appropriate similarity measure. Let us denote as $\mathbf{w}$ the vector containing the gray-scale values of the original watermark and as $\mathbf{w}'$ the vector containing the values of the estimated watermark. Then, the standard correlation coefficient is defined as:

$$\rho = \frac{\sum(\mathbf{w}-\overline{w})(\mathbf{w}'-\overline{w'})}{\sqrt{\sum(\mathbf{w}-\overline{w})^2}\sqrt{\sum(\mathbf{w}'-\overline{w'})^2}} \quad (11)$$

where $\overline{w}$ is the mean value of $\mathbf{w}$ and $\overline{w'}$ the mean value of $\mathbf{w}'$.

Correlation can be used as a complimentary criterion to the subjective interpretation of extracted visually recognizable images and it is useful for automatic detection of watermarked video objects.

TABLE II
WATERMARK EXTRACTION FROM VIDEO OBJECT OF FIGURE 6(d) IN CASE OF JPEG COMPRESSION

Compression Ratio	23.4	28.1	35.4
PSNR	34.4	32.1	29.7
Extracted Watermark			
ρ	0.874	0.81	0.773

TABLE III
WATERMARK EXTRACTION FROM VIDEO OBJECT OF FIGURE 7(d) IN CASE OF JPEG COMPRESSION

Compression Ratio	21.9	25.2	31.4
PSNR	37.1	35.7	32.6
Extracted Watermark			
ρ	0.993	0.967	0.931

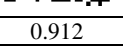
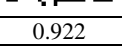
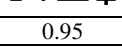
TABLE IV
WATERMARK EXTRACTION FROM VIDEO OBJECT OF FIGURE 6(d) IN CASE OF GAUSSIAN NOISE AND IMAGE PROCESSING ATTACKS

Image Operation	**Gausian Noise**	**Sharpen**	**Blur**
PSNR	29.2	30.2	23.5
Extracted Watermark			
ρ	0.886	0.821	0.93

6.1 Robustness against JPEG Lossy Compression

Table II shows the watermark extraction results from JPEG-compressed versions of the watermarked video object of Figure 6(d), with compression ratios of 23.4, 28.1, and 35.4. Similar results in case of the host video object of Figure 7(d) are presented in Table III, where the compression ratios in this case are 21.9, 25.2 and 31.4. As it can be observed the extracted watermark image is still in viewable even under the highest compression ratios. The difference in quality between the extracted binary and grayscale watermark images is also evident. This is however expected and justified since during binary watermark detection only two levels should be distinguished in contrast to the grayscale case where 256 levels exist. Furthermore high values of both correlation criteria and in all cases are in total agreement to the extracted visually recognizable patterns.

TABLE V
WATERMARK EXTRACTION FROM VIDEO OBJECT OF FIGURE 7(d) IN CASE OF GAUSSIAN NOISE AND IMAGE PROCESSING ATTACKS

Image Operation	**Gausian Noise**	**Sharpen**	**Blur**
PSNR	30.4	32.1	24.8
Extracted Watermark			
ρ	0.912	0.922	0.95

6.2 Robustness against Noise and Image Processing Attacks

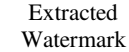
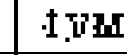
Robustness of the proposed scheme against gaussian noise and image processing attacks such as sharpening and blurring is investigated in this subsection. In particular during transmission, noise may be added to watermarked video objects, which can be modeled in some cases as gaussian noise. On the other hand, sharpening operations are usually performed to enhance the quality of original video objects, while smoothing operations, which blur video objects, are used to decrease artifacts, created by transmission channels of poor quality.

TABLE VI
WATERMARK EXTRACTION FROM VIDEO OBJECT OF FIGURE 6(d) IN CASE OF MIXED IMAGE PROCESSING ATTACKS AND JPEG COMPRESSION

Image Operations	**Sharpen + Blur**	**Sharpen + Blur under JPEG compression ratio =15.7**
PSNR	30.7	27.4
Extracted Watermark		
ρ	0.741	0.653

In Table IV watermark extraction results are presented for video object of Figure 6(d). Similar results for the second case [video object of Figure 7(d)] are depicted in Table V. In both Tables and for all cases the extracted watermark patterns are highly correlated to the original watermarks and are clearly recognizable.

TABLE VII
WATERMARK EXTRACTION FROM VIDEO OBJECT OF FIGURE 7(d) IN CASE OF MIXED IMAGE PROCESSING ATTACKS AND JPEG COMPRESSION

Image Operations	**Sharpen + Blur**	**Sharpen + Blur under JPEG compression ratio =16.3**
PSNR	33.2	29.6
Extracted Watermark		
ρ	0.892	0.847

6.3 Robustness against Combinations of Image Processing Operations and JPEG Compression

A very interesting and common category of attacks combines mixed image processing operations together with JPEG compression. Mixed image processing operations can enhance the overall quality of video objects, while JPEG compression decreases the data size of the final video objects. In our experiments sharpening and blurring operations are performed to the watermarked video objects of Figures 6(d) and 7(d) and then JPEG compression is applied. Tables VI and VII show the watermark extraction results for the two video objects. The video object of Figure 6(d) is enhanced and afterwards compressed with ratio 15.7 providing a PSNR value of 27.4 dB. Similar image processing operations are performed to the video object of Figure 7(d), where now the compression ratio is 16.3 providing PSNR equal to 29.6 dB. Again, in all cases the extracted watermark patterns are highly correlated to the original watermarks, while the contained characters in each pattern are in most cases easily recognizable.

TABLE VIII
WATERMARK EXTRACTION RESULTS FROM VIDEO OBJECT OF FIGURE 6(d), UNDER COMBINATIONS OF JPEG COMPRESSION AND DIFFERENT BER_S

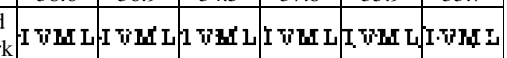
JPEG Compression Ratio	2.6	2.6	2.6	5.1	5.1	5.1
BER	$3x10^{-4}$	$1x10^{-3}$	$3x10^{-3}$	$3x10^{-4}$	$1x10^{-3}$	$3x10^{-3}$
PSNR	37.1	35.9	32.9	36.8	35.2	31.8
Extracted Watermark						
ρ	0.914	0.897	0.853	0.902	0.874	0.835

6.4 Robustness against JPEG Compression and Lossy Transmission

In this subsection the case of JPEG compression and lossy transmission is investigated. Such an attack is common since images may be compressed before transmission, while in unstable QoS networks (e.g. mobile) transmission losses are usual. In our experiments and for each JPEG-compressed watermarked VO, lossy transmission simulations were performed for different Bit Error Rates (BERs). Results are presented for 3 different BERs of $3x10^{-4}$, $1x10^{-3}$ and $3x10^{-3}$, considering that

typical average BERs for cellular mobile radio channels are between 10^{-4} and 10^{-3} [30]. Results of the retrieved watermark patterns for the first and second video objects are given in Tables VIII and IX respectively. As it can be observed from these tables, the proposed system is also robust to this type of attack. Correlation values are high for the extracted watermark patterns, while even under heavy transmission losses retrieved patterns are still recognizable.

TABLE IX
WATERMARK EXTRACTION RESULTS FROM VIDEO OBJECT OF FIGURE 7(d), UNDER COMBINATIONS OF JPEG COMPRESSION AND DIFFERENT BER_S

JPEG Compression Ratio	2.9	2.9	2.9	5.5	5.5	5.5
BER	$3x10^{-4}$	$1x10^{-3}$	$3x10^{-3}$	$3x10^{-4}$	$1x10^{-3}$	$3x10^{-3}$
PSNR	38.6	36.9	34.5	37.8	35.9	33.7
Extracted Watermark						
ρ	0.997	0.983	0.961	0.989	0.976	0.954

7. CONCLUSION

In this paper, a wavelet-based watermarking system is proposed, which embeds visually recognizable watermark patterns such as binary, grayscale or color images, to the most significant wavelet coefficients (QSWTs) of host video objects. Since watermark patterns are recognizable, selection of experimental thresholds during watermark detection can be avoided, in contrast to existing object-based approaches where i.i.d. distributions are embedded.

The system consists of three main modules: (a) unsupervised video object extraction, (b) shape adaptive wavelet decomposition (SA-DWT) and QSWTs detection and (c) watermark embedding. Video objects are automatically extracted using depth information, tube-embodied Gradient Vector Flow fields and active contours. Each video object is then decomposed into three levels with ten subbands using the SA-DWT transform and for the highest energy pair of subbands, QSWTs are estimated. Finally, a watermark pattern is embedded to the best QSWTs of each video object using a non-linear insertion procedure that adapts the watermark pattern to the energy of each specific wavelet coefficient.

Experimental results show that hidden watermarks are perceptually invisible, statistically undetectable and thus difficult to extract without knowledge of the embedding method. Furthermore the watermarks are resistant against several types of plain and mixed image processing attacks. Watermarked video objects are also tested under compression and lossy transmission simulations, providing also very promising results. Additionally, a correlation measure has been adopted for automatic detection of watermark patterns so as to avoid use of text recognition algorithms.

In future research, oblivious watermark retrieval methods should also be investigated. Additionally schemes for directive spreading of watermark information should be implemented to cover all different regions of a video object (e.g. face and body of a human VO). Finally cases of rotation, scaling and cropping attacks, combined with image processing operations, should be analytically investigated.

8. REFERENCES

[1] ISO/IEC 14496-2:2004 *Information technology - Coding of audio-visual objects* - Part 2: *Visual*, 2004.

[2] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in Proceedings *of the IEEE Int. Conf. Image Processing*, vol.2, pp. 86-90, 1994.

[3] N. Nikolaidis, and I. Pitas, "Copyright protection of images using robust digital signatures," in Proceedings *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol.4, pp. 2168-2171, May 1996.

[4] R. Wolfgang and E. Delp, "A watermark for digital image," in Proceedings *Int. Conf. Image Processing*, vol.3, pp. 211-214, 1996.

[5] J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol.6, pp. 1673-1687, Dec.1997.

[6] C.-T. Hsu and J.-L. Wu, "DCT-based watermarking for video," *IEEE Trans. Consumer Electronics*, vol.44, pp. 206-216, Feb. 1998.

[7] M.-S. Hsieh, D.-C. Tseng, and Y.-H. Huang, "Hiding Digital Watermarks Using Multiresolution Wavelet Transform," *IEEE Trans. Industrial Electronics*, vol. 48, no. 5, pp. 875-882, Oct. 2001.

[8] W. Zhu, Z. Xiong, and Y.-Q. Zhang, "Multiresolution watermarking for images and video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no.4, pp. 545-550, June 1999.

[9] H. Daren, L. Jiufen, H. Jiwu, and L. Hongmei, "A DWT-based image watermarking algorithm," in Proceedings *IEEE Int. Conf. Multimedia and Expo*, Tokyo, Japan, 22-25 August, 2001.

[10] X. Wu, W. Zhu, Z. Xiong, and Y.-Q. Zhang, "Object-based multiresolution watermarking of images and video," in Proceedings *IEEE Int. Sym. Circuits and Systems*, Geneva, Switzerland, May 28-31, 2000.

[11] P. Bas, and B. Macq, "A new video-object watermarking scheme robust to object manipulation," in Proceedings *IEEE Int. Conf. Image Processing,* Vol. 2, pp. 526 –529, Oct. 2001.

[12] M. D. Swanson, B. Zhu, B. Chau, and A. H. Tewfik, "Object-based transparent video watermarking," in Proceedings *IEEE Workshop on Multimedia Signal Processing*, New Jersey, USA, June 23-25, 1997.

[13] C.-S. Lu, and H.-Y. M. Liao, "Oblivious cocktail watermarking by sparse code shrinkage: a regional- and global-based scheme", in Proceedings *IEEE Int. Conf. Image Processing,* Vancouver, Canada, vol. III, pp. 13-16, Sept. 10-13, 2000.

[14] A. Nikolaidis and I. Pitas, "Region-Based Image Watermarking," *IEEE Trans. Image Processing*, Vol. 10, No. 11, Nov. 2001.

[15] L. Lia,, J. Qiana,, and J.-S. Panb, "Characteristic region based watermark embedding with RST invariance and high capacity," *Elsevier International Journal of Electronics and Communications*, Vol. 65, No. 5, May 2011.

[16] C. Deng, X. Gao, X. Li and D. Tao, "Robust Image Watermarking Based on Feature Regions," *Studies in Computational Intelligence*, Springer, Vol. 346, 2011.

[17] X. Wang and Z. Guo, "A robust content-based watermarking scheme," *IEEE International Workshop on Multimedia Signal Processing*, 2009.

[18] L. Yang, R. Ni and Y. Zhao, "Segmentation-based Image Authentication and Recovery Scheme Using Reference Sharing Mechanism," *American Journal of Engineering and Technology Research*, Vol. 11, No.6, 2011.

[19] M.K. Kundu and S. Das, "Lossless ROI Medical ImageWatermarking Technique with Enhanced Security and High Payload Embedding," *International Conference on Pattern Recognition*, Turkey, 2010.

[20] O. M. Al-Qershi and B. E. Khoo, "ROI–based Tamper Detection and Recovery for Medical Images Using Reversible Watermarking Technique," *IEEE International Conference on Information Theory and Information Security*, Beijing, China, 2010.

[21] R.-V. Schyndel, "A Hardware-based Surveillance Video Camera Watermark ," *International Conference on Digital Image Computing: Techniques and Applications*, Sydney, Australia, 2010.

[22] L. Coria, P. Nasiopoulos and R. Ward, "A Region-Specific QIM-Based Watermarking Scheme for Digital Images," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Bilbao, Spain, May 2009.

[23] K. S. Ntalianis, N. D. Doulamis, A. D. Doulamis, and S. D. Kollias, "Tube-Embodied Gradient Vector Flow Fields for Unsupervised Video Object Plane (VOP) Segmentation," in Proceedings *IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, October 2001.

[24] S. Li, and W. Li, "Shape-Adaptive Discrete Wavelet Transforms for Arbitrarily Shaped Visual Object Coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no.5, pp. 725-743, August 2000.

[25] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol.41, pp. 3445-3462, Dec. 1993.

[26] A. D. Doulamis, N. D. Doulamis, K. S. Ntalianis and S. D. Kollias, "Efficient Unsupervised Content-Based Segmentation in Stereoscopic Video Sequence," in *Intern. Journal on Artificial Intelligence Tools*, vol. 8, no.6, 2000.

[27] L. Garrido, F. Marques, M. Pardas, P. Salembier and V. Vilaplana, "A Hierarchical Technique for Image Sequence Analysis," in *Proc. of Workshop on Image Analysis for Multim. Interactive Services* (*WIAMIS*), pp. 13-20, Louvain-la-Neuve, Belgium, June 1997.

[28] C. Xu, and J. L. Prince, "Snakes, Shapes, and Gradient Vector Flow," *IEEE Trans. Image Processing*, Vol. 7, No. 3, pp. 359-369, March 1998.

[29] C.-T. Hsu and J.-L. Wu, "Multiresolution watermarking for digital images," *IEEE Trans. Consumer Electronics*, vol. 45, pp. 1097–101, Aug. 1998.

[30] V. Weerackody, C. Podilchuk, and A. Estrella, "Transmission of JPEG-Coded Images over Wireless Channels," *Bell Labs Technical Journal*, Autumn 1996.

Received: November 2006. Accepted: September 2012.