ROBUST CONTENT-BASED VIDEO WATERMARKING EXPLOITING MOTION ENTROPY MASKING EFFECT

Amir Houmansadr, Hamed Pirsiavash

Department of electrical engineering, Sharif University of Technology, Azadi Ave., Tehran, Iran

Shahrokh Ghaemmaghami

Electronics Research Center, Sharif University of Technology, Azadi Ave., Tehran, Iran

Keywords: Digital watermarking, Video sequence, Entropy masking.

Abstract: A major class of image and video watermarking algorithms, i.e. content-based watermarking, is based on the concept of Human Visual System (HVS) in order to adapt more efficiently to the local characteristics of the host signal. In this paper, a content-based video watermarking scheme is developed and the concept of entropy masking effect is employed to significantly improve the use of the HVS model. Entropy masking effect states that the human eye's sensitivity decreases in high entropy regions, i.e. regions with spatial or temporal complexity. The spatial entropy masking effect has been exploited in a number of previous works in order to enhance the robustness of image-adaptive watermarks. In the current research, we use the temporal entropy masking as well to achieve a higher performance in video watermarking. Experimental results show that more robust watermarked video sequences are produced considering temporal entropy masking effect, while the watermarks are still subjectively imperceptible. Robustness enhancement is a function of temporal and spatial complexity of the host video sequences.

1 INTRODUCTION

During the past two decades, digital industries have shown an explosive growth. Digital media has captured homes and offices by a storm through the internet and high quality digital disks such as CD's and DVD's. Unfortunately, copyright protected digital media can be easily copied and distributed without permission of the content owner. This causes a huge damage to the relevant industries especially multimedia producing companies. As a result, enforcement of the intellectual property rights has become a critical issue in recent years and a vast amount of fund has been allocated to research in this area.

At the end of the 20th century, digital watermarking was introduced as a complementary solution for the aim of protecting digital owners' rights. A digital watermark is a visible or an invisible mark, which is embedded into the digital media such as audio, image, and video to identify the customer and/or providing additional information about the media such as its producer, the customer's permissions, etc. Regarding the applications, watermarking schemes are classified into two major classes: source-based and destination-based schemes (Podilchuk, 1998). Source-based watermarking schemes in which a watermark identifying the owner is inserted into different distributions of a digital media aim the ownership identification/authentication. This is while destination-based schemes intend to trace lawbreaking customers by inserting a distinct watermark identifying each customer in the digital production. In this category, it is assumed that the original media is available during the detection process, while in the first category the watermark extraction should be performed just using the watermarked media.

The embedded mark should have some significant features to promise its functionality (Pan, 2004). The watermarked media must be perceptually similar to the original host signal, i.e. the mark existence in the media should be inaudible or invisible (imperceptibility requirement). In addition,

the watermark should survive in the digital media after performing various attacks. The watermarking attacks consist of intentional attacks made to the marked media to remove or change the watermark and various signal processing schemes such as compression, resampling, halftoning, cropping, etc. This feature of the watermarking schemes is addressed as robustness requirement. In the case of destination-based watermarks, another characteristic should also be considered defining how widely it can be distributed. This is called watermark's capacity, which states the ability to detect different watermarks with a low probability of error, as the number of distinct watermarks increases. There is always a tradeoff between these three requirements.

Different models of the Human Visual System (HVS) have been proposed to be used in the image and video compression applications. Due to the fact that both compression algorithms and watermarking schemes aim to find redundant data of the digital media (to get removed in the former and to be used for inserting watermark in the latter), visual models employed for the compression applications can also be used in the watermarking schemes.

To be imperceptible, an image or video watermark should consider the characteristics of the HVS. Depending on how HVS models are used, watermarking schemes can be classified into two major categories: image-independent and image-adaptive (or generally content-based) watermarking schemes (Wolfgang, 1993). Algorithms belonging to the first class are based on the Modulation Transfer Function (MTF) of the human eye only, but do not mention any particular characteristic of the particular image or video frames. On the other hand, image-adaptive watermarking schemes depend not only on the frequency response of the human eye, but also on the properties of the image itself. Consequently, image-adaptive watermarking schemes can maximize the watermark robustness, while satisfying the transparency requirement. In other words, a content-based watermark is perceptually adapted to local characteristics of the host image or video. The main theme of this paper is to develop a content-based watermarking scheme for video host signals.

Various applications have been proposed for video watermarking (Doerr, 2003). Video watermarking has been suggested for preventing illegal copying in the future DVD devices (Bloom, 1999). Also video watermarking can be used for the aim of automatically checking whether a specified program, e.g. commercial advertisement, has been broadcasted by some channels in a specific period of time (Depovere, 1999). Inserting a unique watermark ID for each customer transmitted through Pay-Per-View and Video-On-Demand services enables it to trace back traitor users (Lin, 2001). But its first targeted application is copyright protection, in which the owner inserts its registered watermark into the digital property to prove its paternity in the case of finding an illegally copied version (Qiao, 1998).

In this paper, a content-based method for watermarking of video streams is developed regarding motion entropy of the host video. In Section 2, some discussion on the content-based watermarking schemes is given and the entropy masking effect is introduced. Section 3 extends image watermarking methods to the video contents. The proposed algorithm using motion entropy is described in Section 4. Section 5 presents implementation results and a comparison with the traditional methods. Finally, Section 6 concludes the paper.

2 CONTENT-BASED WATERMARKING SCHEMES

In this section, some image-adaptive watermarking methods based on DCT transform are described. The IA-DCT method proposed by *Podilchuk et al.* is a good paradigm of image-adaptive watermarking schemes (Podilchuk, 1998). The method is the dual of the image-independent scheme proposed by *Cox et al.* in (Cox, 1995). In the IA-DCT method, *Watson*'s visual model (Watson, 1993) is used to insert watermark in the DCT coefficients of the host image. First, the image is divided into 8*8 non-overlapping blocks of pixels and then the watermark signal, which is a Gaussian zero-mean random process with variance equal to one, is inserted in the DCT transform of the blocks, as:

$$X'_{b,u,v} = \begin{cases} X_{b,u,v} + jnd_{b,u,v} * w_{b,u,v} & if \quad X_{b,u,v} \geq jnd_{b,u,v} \\ X_{b,u,v} & otherwise \end{cases} \quad (1)$$

where $X_{b,u,v}$ is the (u,v)-th coefficient of DCT transform over *b*'th block, *jnd* is the corresponding Just Noticeable Difference (JND) which is evaluated using Watson's visual model, $w_{b,u,v}$ is the corresponding watermark bit, and $X'_{b,u,v}$ is the resultant watermarked DCT coefficient.

In (Watson, 1997), Watson introduced a new masking effect called entropy masking, which is due to unfamiliarity of the observer. In fact, entropy masking is due to weakness of human brain in processing simultaneous complex phenomena. By

performing a number of experiments, Watson showed that this masking phenomenon is distinct from previously known masking effects such as contrast masking. Contrast masking refers to the early visual system effective gain reduction, whereas entropy masking takes place in the human's brain. According to this phenomenon, as the number of image details increases, the HVS decreases its sensitivity to the image details. Likewise, HVS decreases its sensitivity to regions in a video stream where a lot of motion complexity exists.

Entropy masking, as defined and introduced by Watson in (Watson, 1997), has rarely been exploited in the watermarking schemes. In some earlier works, entropy masking has been confused by the contrast masking in an awkward manner. This is due to the fact that spatial complexity results in both higher contrast masking and higher entropy masking. However, these two kinds of masking stem from absolutely different sources. In such works, entropy masking has been treated as the contrast masking; so the extreme capability of entropy masking has not been exploited in these works and not a high improvement has been achieved compared to traditional schemes. Exploiting entropy masking in applying visual models to watermarking schemes can lead to drastic improvements in the watermarking system's functionality, i.e. its robustness. In our previous work, the spatial entropy masking has been exploited to enhance the watermark power in IA-DCT scheme; hence the watermark resilience against common attacks is improved. It has been shown that the achieved enhancement is clearly higher than what is gained in similar works. The Shannon's entropy of pixels intensity histogram within each non-overlapping block is defined as its complexity and is used for elevating JND coefficients in high spatial complexity regions.

3 EXTENSION OF IMAGE WATERMARKING SCHEMES TO VIDEO

In this section, previously discussed image watermarking algorithms will be extended to video. MPEG standard is the most popular video compression method; hence, the watermarking algorithm should be robust against this attack.

In MPEG-1 compression method, some frames are coded as intraframes and the others are predicted from them. In order to reduce the prediction error

propagation, for each group of frames (typically 12 consequent frames), one intraframe should be included. In addition, because of large prediction error in shot change frames, these frames are usually considered as intraframes. According to MPEG-1 standard, intraframes should be coded similar to JPEG standard algorithm for still images. The other frames are predicted from their neighborhood frames using block-wise motion vectors and the difference frame is coded in the DCT transform similar to JPEG algorithm using a different quantization matrix. During this process, most details of predicted frames are omitted, so the watermark inserted in these frames cannot be detected efficiently.

In (Wolfgang, 1993), the IA-DCT technique is simply extended to video streams. The watermark detection from predicted frames is very difficult and inaccurate; hence the watermark should be inserted in the intraframes. Because the compression procedure is independent of watermarking system, intraframe positions are not known; so we prefer to embed the same watermark in the frames belonging to each group of typically 12 frames. Consequently, in the detection procedure, the detection correlation is computed for all frames. This correlation value will be significantly higher for intraframes.

In this research, we extend the IA-DCT for video streams considering both spatial and temporal entropy masking effects. Results of these implementations are presented in Section 5.

4 THE PROPOSED ALGORITHM

In this section, our proposed method for video watermarking which uses motion information is discussed. In content-based image watermarking methods, human eye sensitivity to the spatial characteristics is considered, and for instance in (Houmansadr, 2005) which is discussed briefly in section 2, the watermark energy is increased in spatially complex regions of the image, considering entropy masking effect, in order to achieve better results. In video watermarking, temporal sensitivity should be considered as another important factor. As mentioned earlier, entropy masking is due to weakness of human brain in processing simultaneous complex phenomena. In the case of an image, these complex phenomena can be high entropy information corresponding to complex regions of the image. Likely, brain's weakness in processing high entropy regions of a video signal, i.e. moving parts, leads to an increase in the detection thresholds of the noticeable differences (JND), which can be exploited to strengthen the
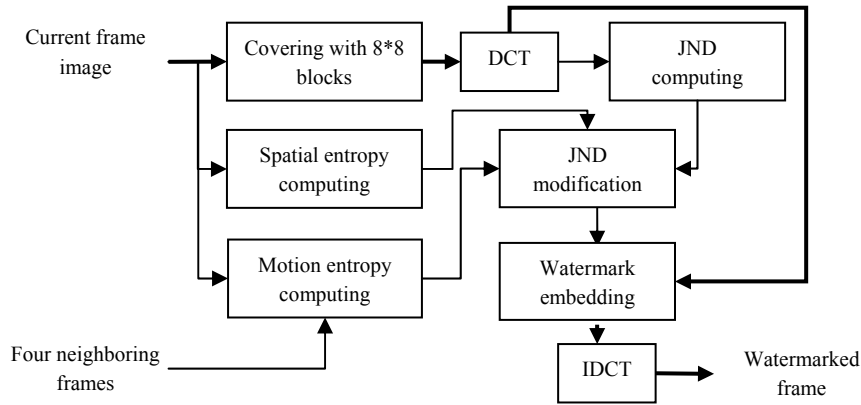
Figure 1: Block diagram of the proposed watermarking method.

embedded watermark. This fact, motion entropy masking effect, has also been noticed in the literature without any adequate reasoning. *Wang et al.* in (Wang, 2002) state that the HVS has lower sensitivity to moving regions of scenes.

To achieve this aim, we need a criterion to measure the temporal complexity of video frames. A primary approach is to extend the spatial entropy blocks in temporal domain and compute the Shannon's entropy for pixels in some cubes instead of blocks. Yet, according to our experiments, this method is inefficient because motion cannot considerably change the pixel values histogram in the cubes.

The overall block diagram of the proposed method is presented in Figure 1. First, each frame is divided into 8*8 non-overlapping blocks of pixels. The resultant frames are transformed to the DCT domain and the JND coefficients are evaluated using the Watson's visual model. Then spatial complexity and temporal complexity are evaluated and used to modify the JND values in respect to the entropy masking effect. The modified JND coefficients are as follows:

$$jnd^{*}=jnd.Spatial\_mask(1+ \alpha.Motion\_mask) \quad (2)$$

where *jnd* is the unmodified JND coefficient

derived from the Watson's visual model, *Spatial_mask* is the spatial entropy masking, *Motion_mask* is the temporal complexity parameter, and α is a parameter which should be adjusted tentatively. Increasing this parameter leads to visible artifacts in the moving areas of the video sequence. This parameter is set to 0.001 in our implementations. The modified JNDs are used in (1) to insert the watermark in the DCT domain of the video frames and an IDCT transform produces the final watermarked video frames.

In (Houmansadr, 2005), we have computed the spatial entropy mask using pixel values histogram of each block. Because few pixels, 8*8=64, are considered in the entropy calculation, the result is not so confident. Hence, the intensity values are quantized to 32 levels. In this research, i.e. video watermarking, the quantized histogram is smoothed using a Gaussian kernel to suppress the sparseness effect perfectly. In fact, having a small number of input data makes the histogram impulsive and leads to inaccurate Shannon's entropy value. The linear function that elevates the spatial entropy mask from the spatial complexity parameter is optimized to new modifications in this research as:

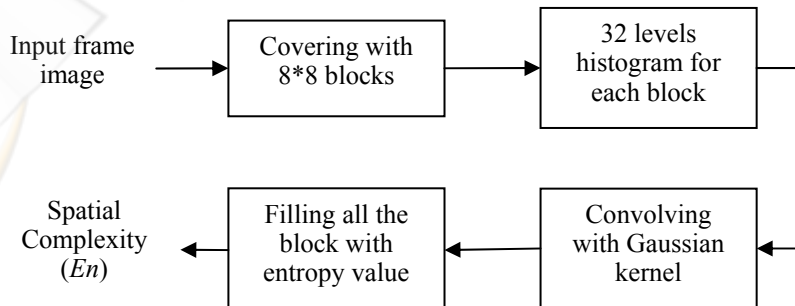$$Spatial\_mask=(0.7*En+0.4) \quad (3)$$

Figure 2: Block diagram of the spatial complexity computation.

Table 1: Results of the subjective experiments (MOS of 5) on the video sequences watermarked with three different algorithms.

Sequence name	IA-DCT	Spatial entropy mask	Entropy and motion mask
Mom and daughter	4.3	4.0	4.1
Coastguard	4.5	4.4	4.4
Foreman	4.3	4.4	4.4
Claire (fast)	4.4	4.4	4.6
Average	4.37	4.3	4.37

where *En* is the spatial complexity parameter evaluated by the Shannon's entropy as discussed earlier. The spatial complexity computing block diagram is shown in Figure 2.

The block diagram of the motion mask production process is shown in Figure 3. First, the frame to be watermarked and its four neighboring frames are considered, and the variance of the intensity values for each pixel is computed for these five frames. Then, the mean of these variance values over 8*8 non-overlapping blocks are computed. This simply computed value for each block represents the temporal complexity of that particular block because it is higher for blocks having larger motion. This is the *Motion_mask* parameter used in (2) with which α parameter has been adjusted in a trade off between robustness and imperceptibility.

5 IMPLEMENTATION RESULTS AND COMPARISONS

The proposed algorithm is implemented in *Matlab* and tested on 200 frames of some standard sequences. A randomly generated watermark is embedded in the frames of the video using IA-DCT, spatial entropy based, and the proposed method prior to the MPEG-1 coding. A set of subjective experiments has been performed to consider the transparency of the watermarked video sequences. As table 1 illustrates, output sequences are subjectively similar to the original ones, and the watermark is perfectly invisible. The correlation values between recovered watermark and the original are computed for all frames of the sequences. Mean and maximum values of the correlation for different sequences are normalized to the IA-DCT method considered as the reference algorithm and are presented in Table 2. In all cases, the maximum value corresponds to one of the intraframes.

As shown in Table 2, correlation of the proposed method in all cases significantly improves over the two previous methods. The increased correlation makes the algorithm more robust and allows the system designer to decrease the watermark energy and makes it more invisible. Claire sequence in our test is a little faster than the ordinary ones and has a large flat background which has very low spatial complexity. So, the spatial entropy masking cannot increase it so much, but using the motion mask, the performance is increased considerably because of high watermark embedding capacity in fast moving regions.
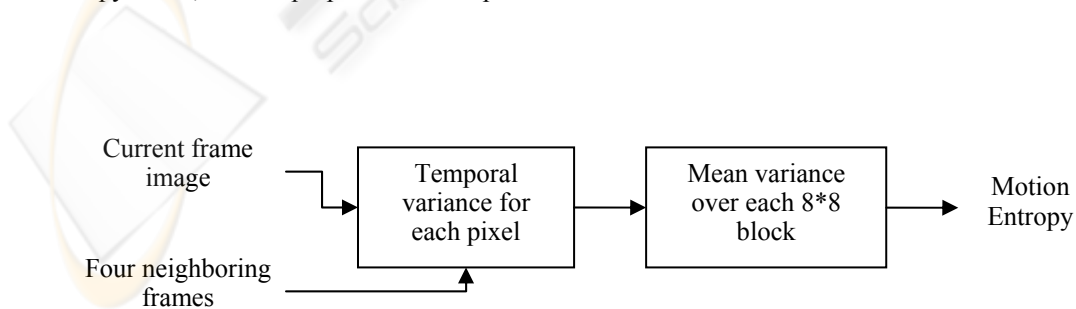

Figure 3: Motion mask computing block diagram.

Table 2: Normalized correlation of original and detected watermarks on MPEG-1 coded watermarked video for some sample sequences.

Sequence name	Mean correlation		Maximum correlation	
	Spatial entropy mask	Entropy and motion mask	Spatial entropy mask	Entropy and motion mask
Coastguard	2.1137	2.7333	1.8176	2.3077
Mom and daughter	1.2309	2.2448	1.1410	2.1239
Foreman	1.4471	2.4762	1.1202	2.6176
Claire (fast)	1.1000	3.0600	0.9817	2.4292
Average	1.4729	2.6286	1.2651	2.3696

Table 3: Mean normalized correlation results for 30 different random croppings.

Sequence name	Spatial entropy mask	Entropy and motion mask
Mom & daughter	1.5211	2.7978
Coastguard	1.5266	2.1205
Foreman	1.1889	1.9922
Claire (fast)	1.0784	4.0784
Average	1.3748	2.5795

The effect of different watermarking attacks is investigated on the proposed power-improved watermarking schemes. The watermarking algorithm should be robust against cropping attack. In these implementations, 30 different randomly selected regions are cropped (keeping one-sixteenth of the original) and the watermark is detected. Table 3 shows the mean of the normalized correlation results. From the table, considering motion entropy in addition to the spatial entropy masking results in more robustness to the cropping, especially in video sequences with high spatial and temporal complexity.

Frame rotating with different degrees is also investigated on the watermarked video sequences. Tables 4 and 5 compare the correlation coefficient of three watermarking schemes for 5 and 10 degrees of rotation, respectively. Again, the correlation coefficients show considerable elevation in respect to the IA-DCT method, particularly for high entropy video sequences.

Also, we passed the watermarked streams through Additive White Gaussian Noise (AWGN) channels with different PSNR values and compared the resulted correlation values. Tables 6 to 8 illustrate these correlations for different PSNR values. Again, it is observed that considering motion mask and spatial entropy effects in the watermark insertion stage leads to more robust watermarked sequences, and the robustness enhancement is a function of the spatial and temporal complexity of the video stream.

Table 4: Mean normalized correlation results after rotation by 5 degrees.

Sequence name	Spatial entropy mask	Entropy and motion mask
Mom & daughter	1.4012	2.6821
Coastguard	1.4116	2.0523
Foreman	1.0962	1.8862
Claire (fast)	1.0565	3.9562
Average	1.2413	2.6442

Table 5: Mean normalized correlation results after rotation by 10 degrees.

Sequence name	Spatial entropy mask	Entropy and motion mask
Mom & daughter	1.4035	2.6855
Coastguard	1.3865	2.0985
Foreman	1.1168	1.8654
Claire (fast)	1.0432	3.8562
Average	1.2375	2.6264

Table 6: Mean normalized correlation results for AWGN channel with PSNR=42.

Sequence name	Spatial entropy mask	Entropy and motion mask
Mom & daughter	1.3561	2.4235
Coastguard	1.3265	1.8956
Foreman	1.0921	1.5786
Claire (fast)	1.0562	3.5890
Average	1.2077	2.3716

Table 7: Mean normalized correlation results for AWGN channel with PSNR=36.

Sequence name	Spatial entropy mask	Entropy and motion mask
Mom & daughter	1.4312	2.5684
Coastguard	1.4290	1.9453
Foreman	1.1036	1.6521
Claire (fast)	1.0486	3.9231
Average	1.2531	2.5222

Table 8: Mean normalized correlation results for AWGN channel with PSNR=30.

Sequence name	Spatial entropy mask	Entropy and motion mask
Mom & daughter	1.4301	2.5622
Coastguard	1.4358	1.9602
Foreman	1.1213	1.7562
Claire (fast)	1.0584	3.9650
Average	1.2614	2.5609

The proposed algorithm has a little more computational complexity over previous methods; hence it could be used in real time applications. By using motion vectors in the motion mask computing process, the algorithm could perform better while is more time consuming. By increasing the watermark energy, it might be visible. Therefore, it might be observed because of watermark stationary, particularly in moving areas. In these cases, watermark could be moved in agreement with the motion vectors. In order to have a fast algorithm, these approaches are not implemented in this research.

Three consecutive frames of "Mom and daughter" sequence, motion mask, spatial entropy mask, the final watermark in DCT domain and spatial domain, and the watermarked image for the central frame are shown in Figure 4. It is obvious that motion mask has high values for mother's hand and lips which move fast.
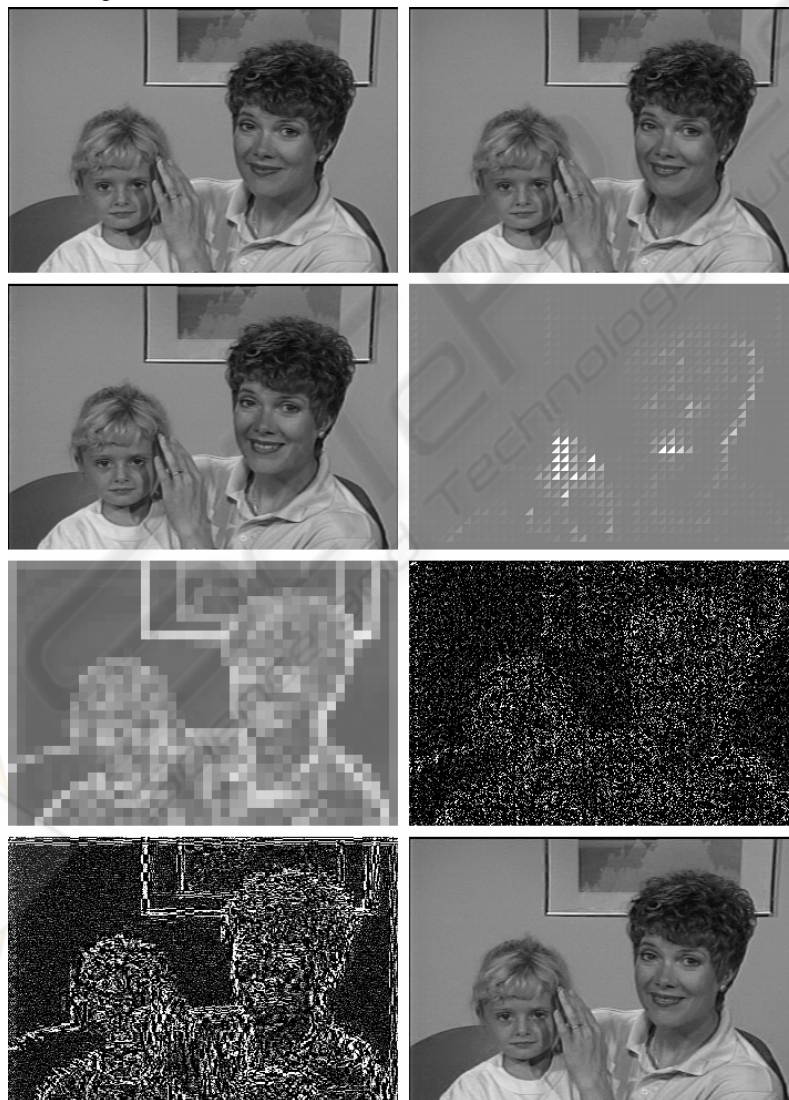

Figure 4: Outputs for Mom and daughter sequence. From top to bottom and left to right: (a) frame 10, (b) frame 12, (c) frame 14, (d) motion mask for (b), (e) spatial entropy mask for (b), (f) embedded watermark in DCT domain, (g) embedded watermark in spatial domain, (h) watermarked output for (b), where Sub-figures (d-g) are exaggerated.

6 CONCLUSIONS

In this paper, a content-based video watermarking algorithm is proposed which uses temporal complexity, i.e. motion, in addition to spatial entropy to better adapt the watermark to the video contents. In other words, because of invisibility of high spatial frequencies in moving fields, the watermark energy can be increased in these regions. In this paper a simple method for motion mask calculation is proposed. This algorithm is implemented in Matlab and tested for some standard sequences. The experiments show considerable improvement in watermark resilience against different watermarking attacks in comparison with the traditional content-based algorithms.

REFERENCES

Bloom, J., Cox, I., Kalker, T., Linnartz, J.-P., Miller, M., and Traw, C., 1999. Copy protection for DVD video, *Proceedings of the IEEE, 87 (7)*. pp. 1267–1276.

Cox, I. J., Kilian, J., Leighton, T., and Shamoon, T., 1995. Secure spread spectrum watermarking for multimedia. *NEC Research Institute Tech.*, Rep. 95-10.

Depovere, G., Kalker, T., Haitsma, J., Maes, M., De Strycker, L., Termont, P., Vandewege, J., Langell, A., Alm, C., Normann, P., O'Reilly, G., Howes, B., Vaanholt, H., Hintzen, R., Donnely, P., Hudson, A., 1999. The VIVA project: digital watermarking for broadcast monitoring. In *Proceedings of the IEEE International Conference on Image Processing, Vol. 2*, pp. 202–205.

Doerr, G., Dugelay, JL., 2003. A guide tour of video watermarking, *Signal Processing: Image Communication, Elsevier Science, vol. 18*. pp. 263-282.

Houmansadr, A., Ghaemmaghami, S., 2005. Robustness enhancement of content-based watermarks using entropy masking effect. *Lecture Notes in Computer Science, vol. 3710*, Springer-Verlag, pp. 444-458.

Lin, E., Podilchuk, C., Kalker, T., Delp, E., 2001. Streaming video and rate scalable compression: what are the challenges for watermarking? In *Proceedings of SPIE 4314, security and Watermarking of Multimedia Content III*, pp. 116–127.

Pan, J. S., Huang, H. C., Jain, L. C., 2004. *Intelligent watermarking techniques*, World Scientific Publishing Co. Pte. Ltd., Singapore.

Podilchuk, C. I., Zeng, W., 1998. Image-adaptive watermarking using visual models. *IEEE Journal on selected areas in communications, vol. 16, no. 4*, pp. 525-539.

Qiao, L., Nahrstedt, K., 1998. Watermarking methods for MPEG encoded video: toward resolving rightful ownership. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, pp. 276–285.

Wang, Y., Ostermann, J., and Zhang, Y., 2002. *Video processing and communications*, Prentice Hall.

Watson, A. B., 1993. DCT quantization matrices visually optimized for individual images. In *Proceedings of SPIE Conf. Human Vision, Visual Processing, and Digital Display IV, vol. 1913*, pp. 202–216.

Watson, A. B., Borthwick, R., and Taylor, M.,1997. Image quality and entropy masking. In *Proc. SPIE Conf., vol. 3016*.

Wolfgang, R. B., Podilchuk, C. I., and Delp, E. J., 1993. Perceptual watermarks for digital images and video. In *Proc. of the IEEE, Vol. 87, No. 7*, pp. 1108-1126.