

# DIGITAL WATERMARKING FOR VIDEO

Chiou-Ting Hsu<sup>1</sup> and Ja-Ling Wu<sup>1,2</sup>

<sup>1</sup>Communication and Multimedia Lab.,  
Department of Computer Science and Information Engineering,  
National Taiwan University, Taipei, Taiwan

<sup>2</sup>Department of Information Engineering,  
National Chi-Nan University, Puli, Nantou, 545, Taiwan  
E-mail: f0506010@csie.ntu.edu.tw

**Abstract:** In this paper, an MPEG-based technique for embedding digital “watermarks” into digital video is proposed. Watermarking technique has been proposed as a method by hiding secret information into the signals so as to discourage unauthorized copying or attest the origin of the media. In our proposed method, we take advantage of prediction types of MPEG standard to watermark both intraframe and non-intraframe blocks with different residual masks. The experimental results show that the proposed watermarking technique results almost invisible difference between the watermarked frames and the original frames, and also robust to clipping operations and MPEG compression.

## 1. INTRODUCTION

As the computers are more and more integrated via the network, the distribution of digital media is becoming faster, easier, and requiring less effort to make exact copies. One of the major impediment is the lack of effective intellectual property protection of digital media to discourage unauthorized copying and distribution.

In analog world, a painting is signed by the artist to attest the copyright, an identity card is stamped by the steel seal to avoid forgery, and the paper money are identified by the embossed portrait. Such kind of hand-written signatures, seals or watermarks have been used from ancient times as a way to identify the source, creator of a document or a picture. For example, a priceless painting of the eleventh-century in National Palace Museum named “Travelers on a Mountain Path” had not been identified as the genuine work of Fan Kuan until Fan’s signature is found between the woods behind a group of travelers of the painting [1].

However, in the digital world, digital technology for manipulating images has make it difficult to distinguish the visual truth [2]. Besides, the characteristics of digitization bring significant changes in copyright issues, such as: ease of replication, ease of transmission and plasticity of digital media, which creates an urgent need to intellectual property protection on the digitally recorded information [3].

In face of the challenges derived from the characteristics of digital media and internet revolution, digital watermarking has been proposed as a way to claim the ownership of the source and the owner himself. The watermark must be embedded into the media so that the watermarked signal is undetectable by the user. Moreover, the watermark should be:

- Imperceptible
- Undeletable
- Statistically Undetectable
- Robust to Lossy Data Compression
- Robust to Signal Manipulation and Processing Operations
- Unambiguous

Several watermarking techniques for text, image and audio data have been developed [4-12].

In this paper, we present a watermarking technique for video sequences. In most of the video coding standards (such as H.261, MPEG-1 and MPEG-2), the hybrid motion compensation/discrete cosine transform (MC/DCT) coding method is commonly used. Based on the DCT-based image watermarking technique[10], applying the same watermarking method into the intraframe and inserting of the watermark in non-intraframe by extending the residual mask into temporal domain (i.e. 3-D residual pattern) could achieve a video watermarking which could survive under MPEG compression.

This paper is organized as follows. A brief background about MPEG is described in Section 2. The video watermarking techniques for both intraframe and non-intraframe are described in Section 3. Section 4 shows the experimental results. Finally, Section 5 concludes the paper.

## 2. MPEG CODING STANDARD

The MPEG-1 (Moving Picture Expert Group) [13] video compression standard is in a hybrid DCT/DPCM block-based scheme. Three types of pictures are considered: Intraframe (I-frame), forward-predicted frame (P-frame), and bi-directional predicted/interpolated frame (B-frame), which are described as follows [14]:

Since video is a sequence of still images, it is

possible to compress a video signal using techniques similar to JPEG. Such methods of compression are called *intraframe* coding techniques, where each frame of video is individually and independently compressed. Intraframe coding exploits the spatial redundancy that exists between adjacent pixels of a frame. A frame is first divided into  $8 \times 8$  blocks of pixels, and the 2-D DCT is then applied independently on each block. This operation results in an  $8 \times 8$  block of DCT coefficients in which most of the energy in the original block is typically concentrated in a few low-frequency coefficients. A quantizer is applied to each DCT coefficient that sets many of them to zero. This quantization is responsible for lossy nature of the compression algorithm. Compression is achieved by storing only the coefficients that survive the quantization operation and by entropy coding their locations and amplitudes.

Temporal redundancy results from a high degree of correlation between adjacent frames. In computing the frame-to-frame difference, a block-based motion compensation approach is adopted for P-frame. A block of pixels (target block) in the frame to be encoded is matched with a set of blocks of the same size in the previous frame (reference frame). The block in the reference frame that best matches the target block is used as the prediction for the latter. This best-matching block is associated with a motion vector that described the displacement between it and the target block. The motion vector is also encoded along with the prediction error. The prediction error is encoded using the DCT-based intraframe encoding technique.

In bi-directional prediction, some of the video frames are encoded using two reference frames, one in the past and one in the future. A block in those frames can be predicted by another block from the past reference frame (forward prediction), or from the future reference frame (backward prediction), or by average of the two blocks (interpolation). Frames that are bidirectionally predicted are never themselves used as reference frames.

A video sequence is divided into a series of GOPs (Group of Pictures), where each GOP contains an I-frame followed by an arrangement of P-frames and B-frames.

### 3. WATERMARKING APPROACHES

The DCT-based watermarking method [10] has been developed for image watermarking which could survive several kinds of image processings and lossy compression. In order to extend the watermarking technique into video sequences, the concept of temporal prediction exploited in MPEG is considered. For intraframe, the same techniques of image watermarking is applied, while for non-intraframe, the residual mask, which is used in image watermarking to obtain the spatially neighboring relationship, is extended into temporal domain according to the type

of prediction coding.

In the proposed method, the GOP structure exploited in the MPEG compressing coding is assumed to be known in advance. Based on the GOP structure, the watermarking technique for intraframe and non-intraframe are described in detail as below.

#### 3.1 Watermarking for intraframe

I-frame is encoded without reference to any other images. In considering the JPEG-like coding technique, a DCT-based watermarking method is developed to provide a invisible watermark and also survive the lossy compression.

The human eyes are more sensitive to noise in lower frequency range than its higher frequency counterpart, while the energy of most natural images are concentrated on the lower frequency range and then the quantization table applied in lossy compression reflects the human visual system which is less sensitive to quantization noise at higher frequencies. Therefore, in order to invisibly embed the watermark and to survive the lossy data compression, a reasonable trade-off is to embed the watermark into the middle-frequency range of the image. Besides, to prevent expert to extract the hidden information directly from the transformed domain, the watermarks are embedded by modifying the relationship of the neighboring blocks of middle-frequency coefficients of the original image instead of embedding by an additive operation.

The original image is divided into  $8 \times 8$  blocks of pixels, and the 2-D DCT is applied independently on each block. Then, pick up the coefficients of middle-frequency range from the DCT coefficients. A 2-D sub-block mask is used in order to compute the residual pattern from the chosen middle-frequency coefficients.

Let the digital watermark be a binary image. A fast 2-D pseudo random number traversing method is used to permute the watermark so as to disperse its spatial relationship. In addition to this pixel-based permutation, a block-based permutation according to the variances of both the image and the watermark is also used. Although the watermark is embedded into the middle frequency coefficients, for those blocks with little variances (i.e. the blocks contain the low frequency contents), the modification of DCT coefficients will introduce quite visible artifacts. In this image-dependent permutation, both variances of image blocks and watermark blocks are sorted and mapped accordingly to improve the invisibility.

After the residual pattern is obtained, for each marked pixel of the permuted watermark, modify the DCT coefficients according the residual mask, so that the corresponding polarity of residual value is reversed. Finally, inverse DCT of the associated result to obtain the watermarked image.

The extraction of watermark requires the original image, watermarked image and also the digital

watermark. First of all, both the original image and the watermarked images are DCT transformed. Then, make use of the chosen middle-frequency coefficients and the residual mask to obtain the residual values. Perform the exclusive-or (XOR) operation on these two residual patterns to obtain a permuted binary signal. Reverse both the block- and the pixel-based permutations to get the extracted watermark.

### 3.2 Watermarking for non-intraframe

P-frame is encoded relative to intraframe or other P-frame. B-frame is derived from two other frames, one before and one after. These non-intraframes are derived from other reference frames by motion compensation, which use the estimated motion vectors to construct the images. In order to insert the watermark into such kind of motion compensated images, the residual patterns of neighboring blocks are extended into temporal domain and other parts of image watermarking techniques could be applied directly into non-intraframes.

For forward predicted P frame, the residual mask is designed between the P-frame and its reference I or P frame. That is, the watermarks are embedded by modifying the temporal relationship between the current P frame and its reference frame.

For bi-directional predicted/interpolated B frame, the residual mask is designed between current B frame and its past and future reference frames. The polarity of residual pattern is reversed to embed the watermark.

Note that, in order to survive the MPEG compression, the middle-frequency DCT coefficients must be modified to reverse the polarity of residual pattern in consideration of the MPEG quantizer, which is applied to each DCT coefficient and is responsible for the lossy nature. Two default quantization tables are defined in MPEG, one is for intraframe and another is for non-intraframe. MPEG allows the quantization operation to achieve a higher level of adaptation, a key factor in achieving good picture quality. Therefore, if the DCT coefficients are modified according to larger quantization factor, the watermark would survive larger compression ratio, and on the other hand, could be with poor invisibility.

## 4. EXPERIMENTAL RESULTS

The "Miss America" video sequences are used as our test sequences and CIF (Common Intermediate Format) format is applied. CIF pictures are composed of three components: luminance Y and color differences Cb and Cr. The picture size for Y are 352 × 288, and the Cb and Cr are subsampled to 176 × 144. Only the luminance components are used during our video watermarking. The binary watermark is set to be 128 × 128.

The GOP structure is assumed as I B B P B B I. Each frame is watermarked by their associate watermarks. Fig. 1 shows the watermarked intraframe and the extracted watermark.

After extracting the watermark, the user can compare the results with the referenced watermark subjectively. However, since the subjective measurement will be affected by the factors of image size, expertise of the observers, the experimental conditions, and etc., a quantitative measurement is required to provide objective judgment of the extracting fidelity. Therefore, a similarity measurement of the extracted and the referenced watermarks can be defined as:

$$\text{Normalized correlation: } NC = \frac{\sum_i \sum_j W(i, j) \hat{W}(i, j)}{\sum_i \sum_j [W(i, j)]^2}$$

which is the cross-correlation normalized by the reference watermark energy to give unity as the peak correlation.

Clipping operation is usually used to edit the special effect on video sequences. Since there are two permutation methods used to disperse the watermark, it is very hard to detect or remove the watermark by just cutting off some part of the frame. Fig. 2 shows the clipped version of the watermarked intraframe, and the missing portions are filled with zero values to extract the watermark.

Figure 3 shows that when the watermarked sequences are under compression / decompression with compression ratio 9.25, the extracted watermarks for I and P frame are still survive with acceptable quality, while the extracted watermark for B frame is degraded and harder to discriminate subjectively.

Figure 4 shows the relationship between the NC values with the clipping ratio for compressed watermarked sequences. Although the quality of extracted watermark is degraded more for B-frames, the effects of clipping ratio for intraframe and non-intraframe are still linear.

Figure 5 shows that as the compression ratio increases, the NC values are decreased according. In most cases, the extracted watermarks for I and P frames are with much higher NC values than that of B-frames.

## 5. CONCLUSION

In this paper, an MPEG-based watermark embedding algorithm is proposed. In order to provide the proposed watermarking algorithm the ability of being: undeletable, perceptually invisible, statistically undetectable, robustness to lossy compression, and survival of video manipulation and processing. The inserting and extracting methods based on the DCT-based intra-mode and temporal predictive non-intra mode have been included in the proposed approach. Experimental results are presented to claim the invisibility and robustness of the proposed video watermarking process.

## References

- [1] National Palace Museum, <http://www.npm.gov.tw>.

- [2] W. J. Mitchell, "When is Seeing Believing?," *Scientific Amer.*, pp. 68-73, Feb. 1994.
- [3] P. Samuelson, "Digital Media and the Law," *Communications of ACM*, vol. 34, no. 10, pp. 23-28, Oct. 1991.
- [4] J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O'Gorman, "Electronic Marking and Identification Techniques to Discourage Document Copying," *IEEE J. Selected Areas Comm.*, vol. 13, no. 8, Oct. 1995.
- [5] S. Walton, "Image Authentication for a Slippery New Age," *Dr. Dobb's Journal*, pp. 18-26, April 1995.
- [6] W. Bender, D. Gruhl and N. Morimoto, "Techniques for Data Hiding," *Proc. of SPIE*, vol. 2420, Feb. 1995.
- [7] O. Bruyndonckx, J. J. Quisquater and B. Macq, "Spatial Method for Copyright Labeling of Digital Images," *Proc. IEEE Nonlinear Signal and Image Processing*, pp. 456-459, June 1995.
- [8] R. B. Wolfgang and E. J. Delp, "A Watermark for Digital Images," *Proceeding of ICIP*, pp. 219-222, Sep. 1996.
- [9] E. Koch and J. Zhao, "Towards Robust and Hidden Image Copyright Labeling," *Proc. IEEE Nonlinear Signal and Image Processing*, pp. 452-455, June 1995.
- [10] C. T. Hsu and J. L. Wu, "Hidden Signatures in Images," *Proceeding of ICIP*, pp. 223-226, Sep. 1996.
- [11] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *Tech. Rep. 95-10*, NEC Research Institute, 1995.
- [12] L. Boney, A. H. Tewfik and K.N. Hamdy, "Digital Watermarks for Audio Signals," *IEEE Proceedings of Multimedia'96*, pp. 473-480, 1996.
- [13] ISO/IEC 11172-2, "Coding of Moving Pictures and Associated Audio - for Digital Storage Media at up to About 1.5 Mbits/s - Part 2 Video."
- [14] R. Aravind, G. L. Cash, D. L. Duttweiler, H. M. Hang, B. G. Haskell and A. Puri, "Image and Video Coding Standards," *AT&T Technical Journal*, pp. 67-89, Jan/Feb. 1993.

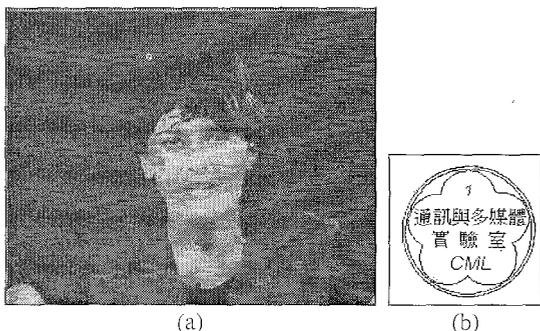


Fig. 1: (a) Watermarked frame with PSNR=44.63dB and (b) the extracted watermark (NC=1).

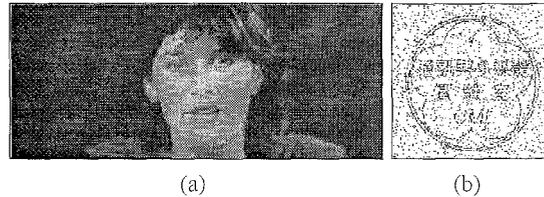


Fig. 2: (a) Clipped version of Fig. 1(a). (b) The extracted watermark after clipping operation (NC=0.6115).

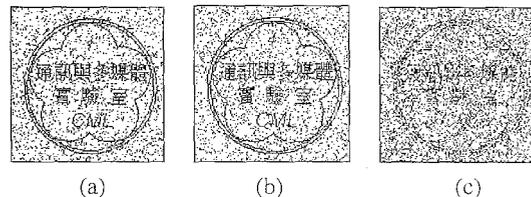


Fig. 3: Extracted watermark for compressed watermarked sequences with compression ratio=9.25, (a) I (frame 1), (b) P (frame 4) and (c) B (frame 5).

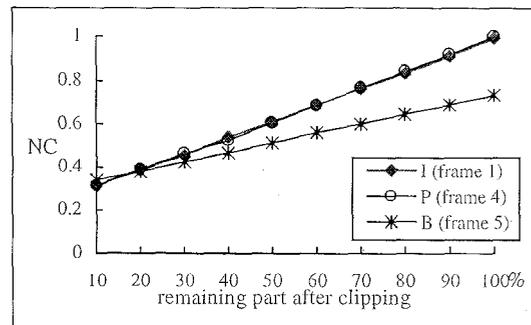


Fig. 4: The relationship between the clipping ratio and NC values for compressed sequences with compression ratio=9.25.

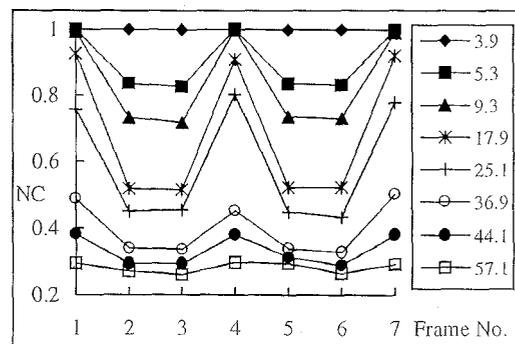


Fig. 5: The relationship between compression ratio and quantitative measurement NC values. As the compression ratio increases from 3.9 to 57.1, the NC values are decreased accordingly. In most cases, the extracted watermarks for I and P frames are with much higher NC values than that of B-frames.