DCT-BASED WATERMARKING FOR VIDEO

Chiou-Ting Hsu[1] and Ja-Ling Wu[1,2]
[1]Communication and Multimedia Lab.
Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan
[2]Department of Information Engineering
National Chi-Nan University
Puli, Nantou, 545, Taiwan
E-mail: candy@cmlab.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 to hide secret information into the signals so as to discourage unauthorized copying or attest the origin of the media. In the proposed method, we take advantage of prediction types of MPEG bitstreams to embed watermarks into both intraframe and non-intraframe 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 cropping 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**

  The watermark should be imperceptible so as not to affect the viewing experience of the image or the audio quality of the original signal.

- **Undeletable**

  The watermark must be difficult or even impossible to remove by a hacker, at least without obviously degrading the original signal.

Contributed Paper
Manuscript received January 12, 1998

- **Statistically Undetectable**

  A 'pirate' should not be able to detect the watermark by comparing several watermarked signals belonging to the same author.

- **Robust to Lossy Data Compression**

  The watermark should be survive the lossy compression techniques like JPEG and MPEG which are commonly used for transmission and storage.

- **Robust to Signal Manipulation and Processing Operations**

  The watermark should still be retrievable even if common signal processing operation are applied, such as signal enhancement, geometric image operations, noise, filtering, etc.

- **Unambiguous**

  Retrieval of the watermark should be unambiguously identify the owner, and the accuracy of identification should degrade gracefully in the face of attack.

Several watermarking techniques for text, image and audio data have been developed.

The watermarking techniques for text are applied to either an image representation of the document or to a document format file. The watermarking can be achieved by altering the text formatting, or by altering certain characteristics of textual elements (e.g., characters). Three methods have been proposed for applying watermarking to text images: line-shift coding, word-shift coding and feature coding, which can be used either separately or jointly [4]. Of course, the watermark can be defeated by retyping all the text in the document.

There are several different methods [5-8] to enable image watermarking in the spatial domain. One disadvantage of spatial domain watermarks—besides the trade-off between invisibility and robustness—is that a common picture-cropping operation can eliminate the watermark [9]. The simplest watermarking method in spatial domain is to just flip the least significant bit (LSB) of chosen pixels in an image. A uniform distributed pseudo random number generator is used to map these bits onto a path of randomly selected pixel locations within the limit of the image. At each location the LSB of the pixel value is forced to match the value of the corresponding watermark bit [5]. However, this works well only if the image will not be subject to any modification, such as color modification done by an image editor or lossy compression.

Image watermarking can be applied in the transform domain [10-12]. These methods are similar to spatial domain watermarking in that the values of selected frequencies can be altered. Because high frequencies will be lost by compression or scaling, the watermark signal is applied to lower frequencies, or better yet, applied adaptively to frequencies containing important elements of the original picture. Upon inverse transformation, watermarks applied to frequency domain will be dispersed over the entire spatial image, so these methods are not as susceptible to defeat by cropping as the spatial technique. However, the trade-off between invisibility and robustness is greater here.

In [13], a digital watermarking approach for audio signals is proposed. In this method, the watermark is generated by filtering a PN-sequence with a filter that approximates the frequency masking characteristics of the Human Auditory System (HAS). It is then weighted in the time domain to account for temporal masking.

In this paper, we present a DCT-based 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 [11], applying the same watermarking method into the intraframe and embedding the watermark into the non-intraframe by extending the residual mask into temporal domain (i.e. 3-D residual pattern) could achieve a robust video watermarking which could survive the attack of 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) [14] 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 bidirectional predicted/interpolated frame (B-frame), which are described as follows [15]:

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 8×8 blocks of pixels, and the 2-D DCT is then applied independently to each block. This operation results in an 8×8 block of DCT coefficients in which most of the energy in the original block is typically concentrated on 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 location information 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 bidirectional 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. Figure 1 shows an example of a GOP structure.

3. WATERMARKING APPROACHES

The DCT-based watermarking method [11] 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 spatial neighboring relationship, is extended to 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 therefore, the quantization table applied in lossy compression always reflects the human visual system which is less sensitive to quantization noise at higher frequencies. 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 8×8 blocks of pixels, and the 2-D DCT is applied independently to each block. Then, pick up the coefficients of middle-frequency range from the DCT coefficients, an example of defining the middle-frequency coefficients is shown in Figure 2. A 2-D sub-block mask, as shown in Fig. 3, is used in order to compute the residual pattern from the chosen middle-frequency coefficients. For example, if $a=b=c=0$, $d=-1, x=1$, then the polarity is a binary pattern (0 or 1) which represents the coefficients at the position of the current block is larger (polarity=1) or less (polarity=0) than the coefficient at the corresponding position of the previous block.

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 applied. 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. However, in this image-dependent permutation, both variances of image blocks and watermark blocks are sorted and mapped accordingly so that the invisibility of the watermarked image will be improved.

After the binary residual patterns of the transformed intraframe are obtained, for each marked pixel of the permuted watermark, modify the DCT coefficients according to 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. Fig. 4 shows the embedding steps of intraframe watermarking.

The extraction of watermark requires the original frame, the watermarked frame and also the digital watermark. First of all, both the original frame and the watermarked frame 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. Fig. 5 shows the extracting steps of the intraframe watermarking.

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 reconstruct the images. In order to embed the watermark into such kind of motion compensated images, the residual patterns of neighboring blocks are extended to 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. Figure 6 shows the residual mask used for P frame. For example, if $a=b=\cdots=h=0$, $x=-1$, $y=1$, then the residual values are the polarity between frame-to-frame differences (with zero-motion vectors). Reverse the polarity of residual pattern by modifying the coefficients of the current frame to embed the watermark.

For bidirectional predicted/interpolated B frame, the residual mask is designed between current B frame and its past and future reference frames. Figure 7 shows the residual mask used for B frame. For example, if $a=b=\cdots=h=0$, $a'=b'=\cdots=h'=0$, $x=x'=(-1/2)$, $y=1$, then the residual values are the polarity between current block and the interpolation of past and future blocks (with zero-motion vectors). The polarity of residual pattern is also 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 considering the effect of the MPEG quantizer, which is applied to each DCT coefficient and is responsible for the lossy nature. Two default quantization tables were defined in MPEG standard, one is for intraframe coding and another is for non-intraframe coding. MPEG allows the quantization operation to achieve a higher level of adaptation, and this is a key factor in achieving good picture quality in MPEG. 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. Figure 8 shows the default quantization matrix used in MPEG-1 standard.

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: one luminance Y and two color differences Cb and Cr. The picture size for Y is 352 × 288, while the Cb and Cr are subsampled to 176 × 144. Only the luminance components are considered during our video watermarking. The binary watermark is set to be 128 × 128.

The GOP structure is assumed as shown in Fig. 1. Each frame is watermarked by their associate watermarks, as shown in Fig. 9. Figure 10 shows one watermarked intraframe and its 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, 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 value.

Cropping 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. Figure 11 shows the cropped version of the watermarked intraframe, and the missing portions are filled with zero values to extract the watermark. Figure 12 shows the relationship between the NC value and the cropping ratio, and the effect of cropping operation to NC is almost linear.

Figure 13 shows that when the watermarked sequences are under compression / decompression attack with different compression ratio from 9.25~29.0, the extracted watermarks for I and P frames are still survive with acceptable quality, while the extracted watermark for B frame is degraded and harder to discriminate subjectively.

Figure 14 shows the relationship between the NC values with the cropping ratio for compressed watermarked sequences. Although the quality of extracted watermark is degraded more for B-frames, the effects of cropping ratio for intraframe and non-intraframe are still linear.

Figure 15 shows the effect of compression for watermarked sequences, when the compression ratio increases, the picture quality is decreased accordingly.

Figure 16 shows that as the compression ratio increases, the NC values decreased accordingly. In most cases, the extracted watermarks for I and P frames are with much higher NC values than that of the B-frames.

5. Conclusion

In this paper, a DCT-based watermark embedding algorithm in considering the MPEG structure 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 embedding 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 Commum.*, vol. 13, no. 8, Oct. 1995.

[5] S. Walton, "Image Authetication 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 International Conference On Image Processing*, pp. 219-222, Sep. 1996.

[9] H. Berghel and L. O'Gorman, "Protecting ownership rights through digital watermarking," *Computer Magazine*, pp. 101-103, July 1996.

[10] E. Koch and J. Zhao, "Towards Robust and Hidden Image Copyright Labeling," *Proc. IEEE Nonlinear Signal and Image Processing*, pp. 452-455, June 1995.

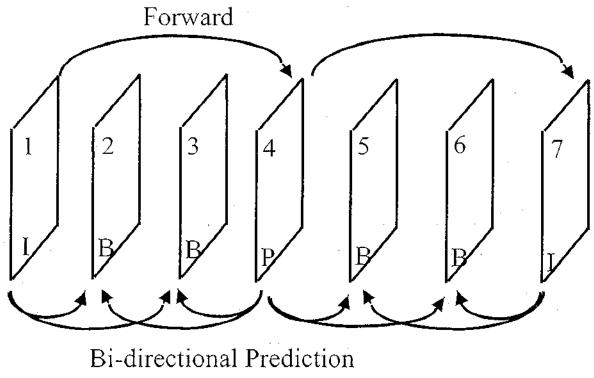
[11] C. T. Hsu and J. L. Wu, "Hidden Signatures in Images," Proceeding of *International Conference On Image Processing*, pp. 223-226, Sep. 1996.

[12] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *Tech. Rep. 95-10*, NEC Research Institute, 1995.

[13] L. Boney, A. H. Tewfik and K.N. Hamdy, "Digital Watermarks for Audio Signals," *IEEE Proceedings of Multimedia'96*, pp. 473-480, 1996.

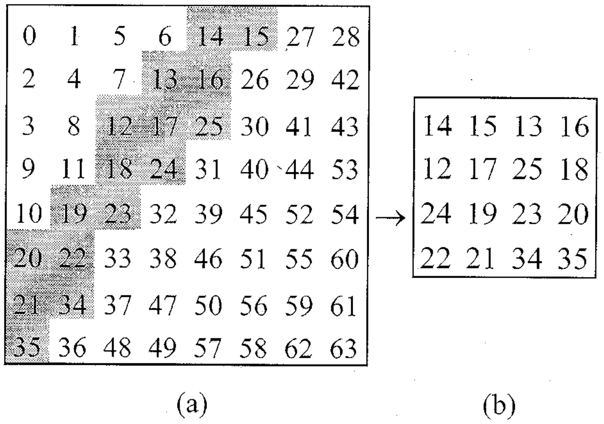
[14] ISO/IEC 11172-2, "Coding of Moving Pictures and Associated Audio - for Digital Storage Media at up to About 1.5 Mbits/s - Part 2 Video."

[15] 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.



Transmission order
1 4 2 3 7 5 6
I P B B I B B

Figure 1: An example of group of pictures, note that B-frame 5,6 and I-frame 7 are part of the next GOP.



0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

→

14	15	13	16
12	17	25	18
24	19	23	20
22	21	34	35

(a) (b)

Figure 2: An example of defining the middle-frequency coefficients, in which the coefficients are picked up in zigzag-scan order and then reordered into a block of 4×4. (a) shows the zigzag ordering of DCT coefficients and the middle frequency coefficients are shown in the shadow area, and (b) shows how the picked up coefficients are reordered into the 4×4 block.

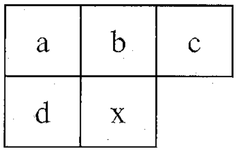


Figure 3: The residual mask, where each square represents a transformed block, and position x stands for the current processing block.

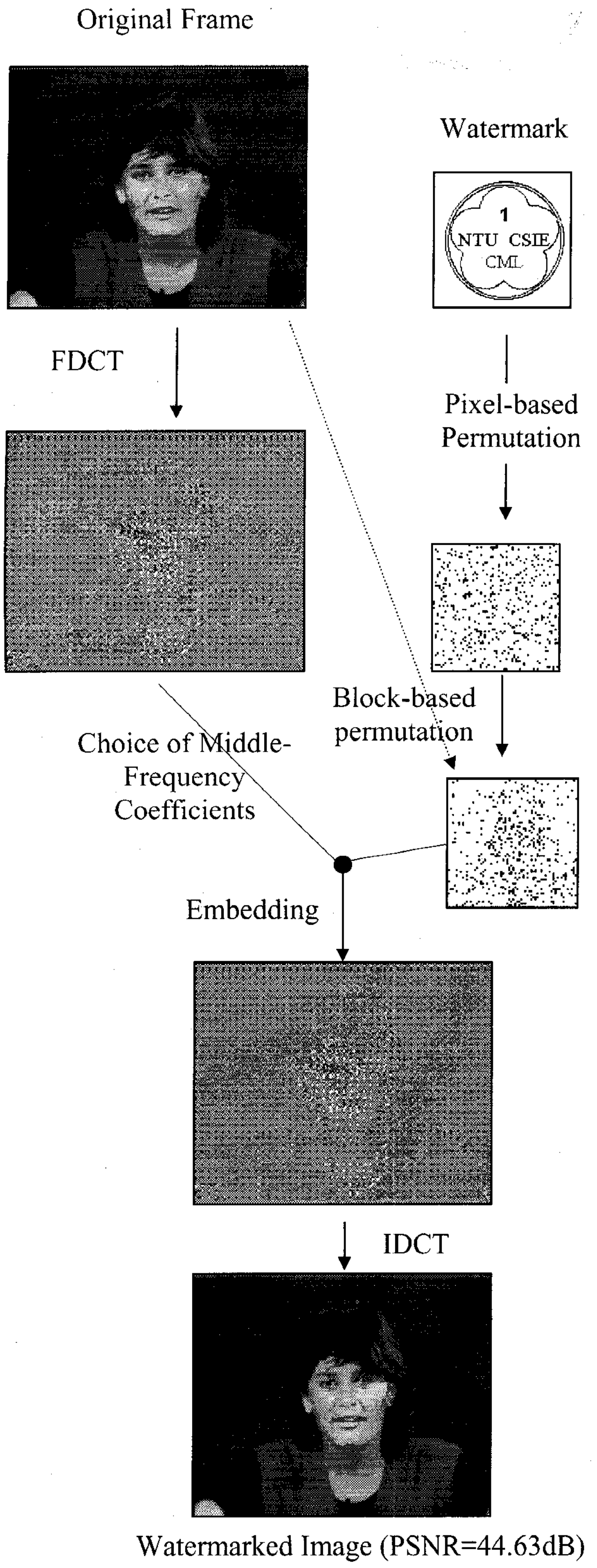


Figure 4: Intraframe watermarking embedding steps.

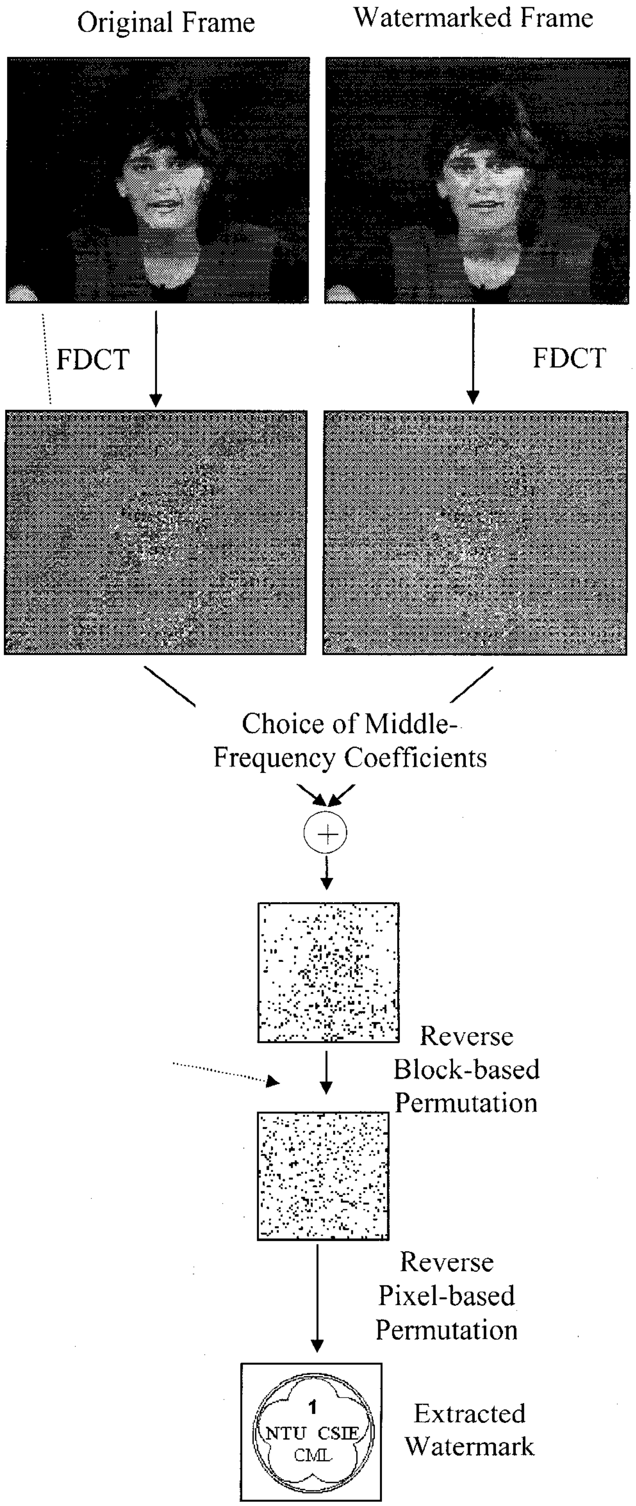


Figure 5: Intraframe watermarking extracting steps.

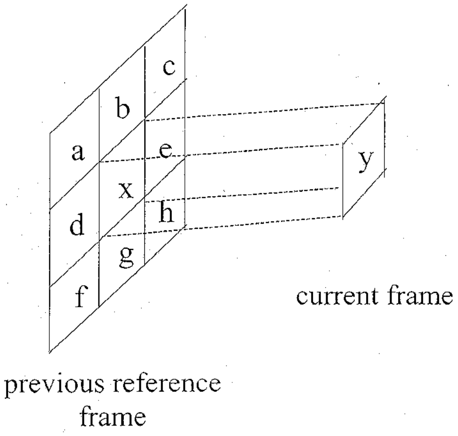


Figure 6: The residual mask for P frame.

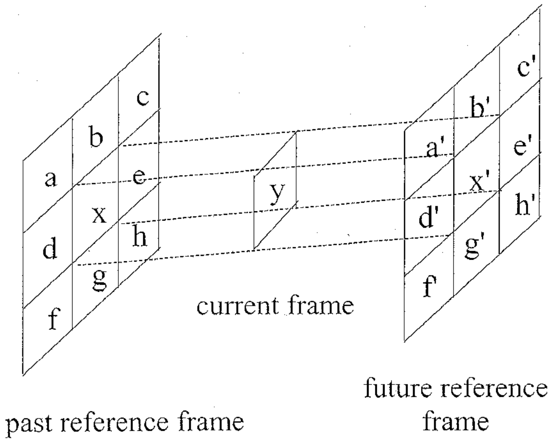


Figure 7: The residual mask for B frame.

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	37	40
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

(a)

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

(b)

Figure 8: Default quantization table of MPEG, (a) intra-frame (b) inter-frame.

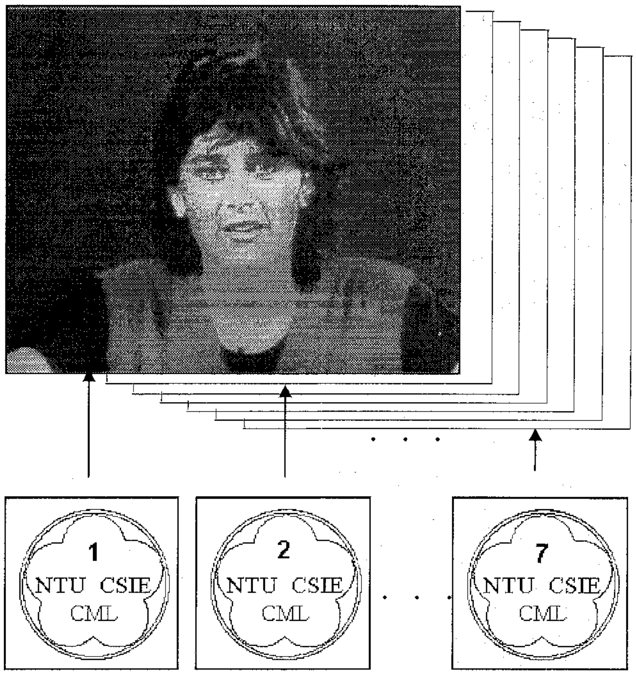
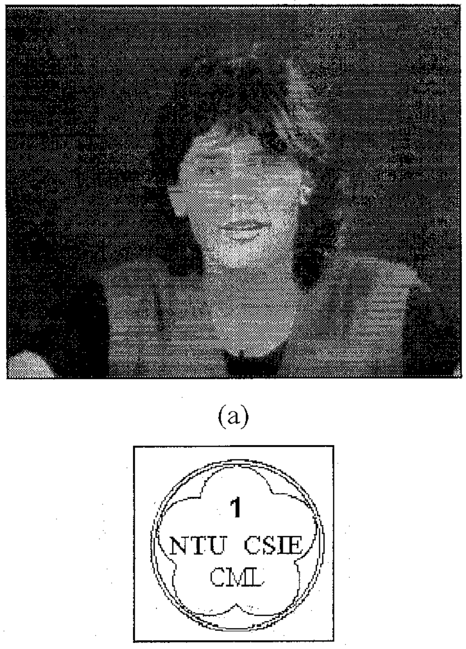


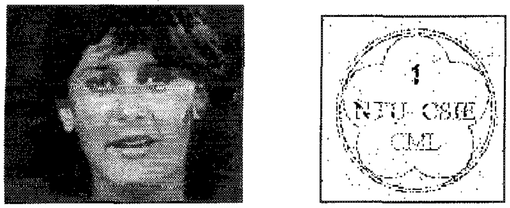
Figure 9: Original video sequences and their associate watermarks.



(a)

(b)

Figure 10: (a) Watermarked picture (of frame 1) with PSNR=44.63dB and (b) is the extracted watermark (NC=1).



(a) (b)

Figure 11: Cropped version of Fig. 10 (a), and (b) is the extracted watermark after cropping operation.

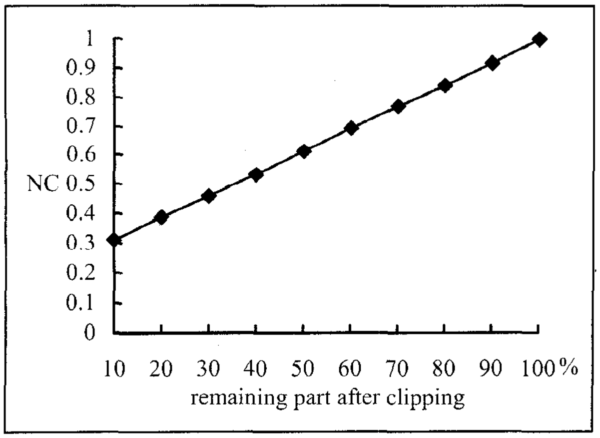
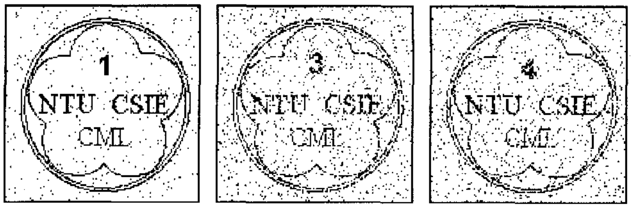
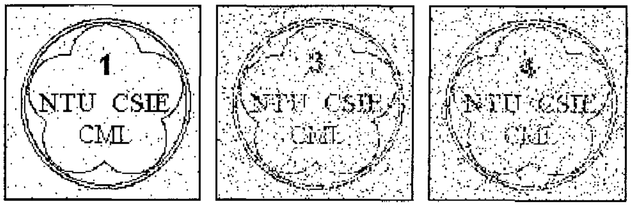


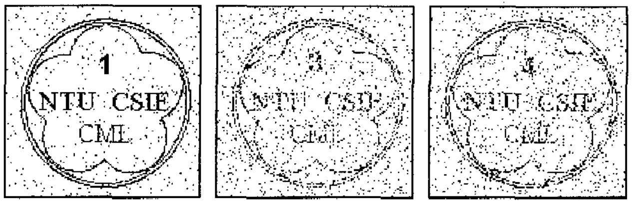
Figure 12: The relationship between the clipping ratio and NC values (Note that, the NC values after clipping are almost the same for Intra- and non-intraframes).



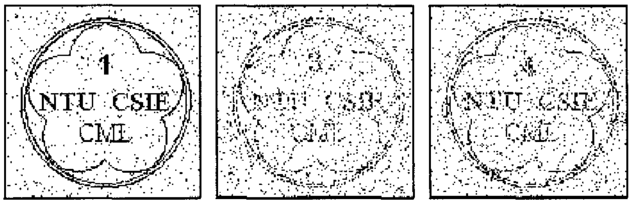
(a) I (frame 1) B (frame 3) P (frame 4)



(b) I (frame 1) B (frame 3) P (frame 4)



(c) I (frame 1) B (frame 3) P (frame 4)



(d) I (frame 1) B (frame 3) P (frame 4)

Figure 13: Extracted watermark for compressed watermarked sequences (a) with compression ratio=9.25, (b) with compression ratio 18.4, (c) with compression ratio 24.34, and (d) with compression ratio 29.0.

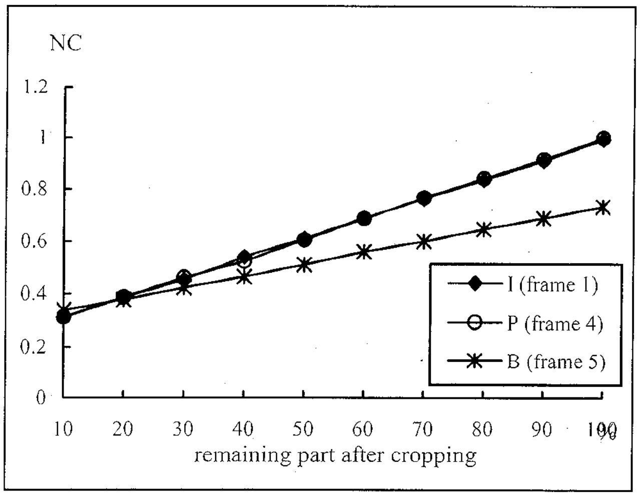


Figure 14: The relationship between the cropping ratio and NC values for compressed sequences with compression ratio=9.25.

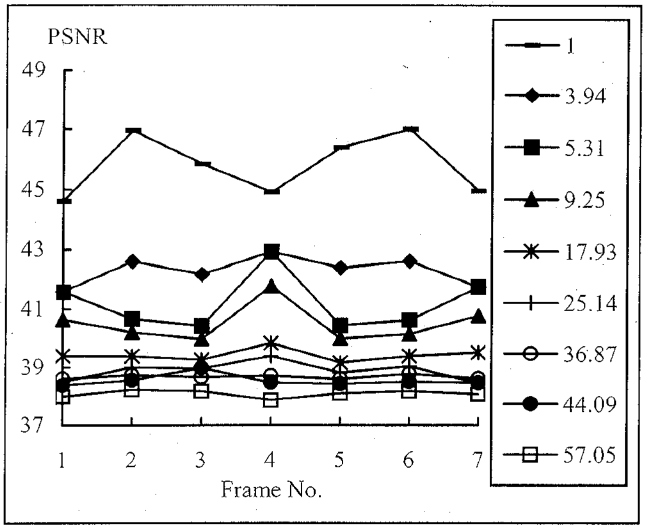


Figure 15: The relationship between compression ratio and objective picture quality. As the compression ratio increases from 1, 3.94 to 57.05, the PSNR is decreased accordingly. Note that, the PSNR of watermarked non-intraframe (without compression, i.e. compression ratio=1) is higher than that of intraframe.

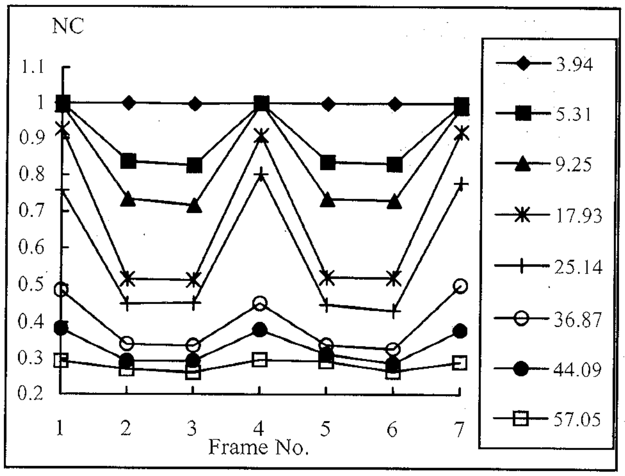


Figure 16: The relationship between compression ratio and quantitative measurement NC values. As the compression ratio increases from 3.94 to 57.05, the NC values are decreased accordingly. In most cases, the extracted watermarks for I and P frames are with much higher NC values than that of B-frames.