PERCEPTUAL HASH FOR COMPRESSED MEDICAL VIDEO

Randheer Bagi, Tanima Dutta and Hari Prabhat Gupta

Dept. of CSE, IIT (BHU) Varanasi, India

ABSTRACT

The perceptual hash dependent on the content of the multimedia object is widely used to protect from copy attack, a watermark estimation attack, that causes protocol ambiguity in a watermarking system. Frame averaging is mainly used for watermark estimation. Motion coherency is essential to resist temporal frame averaging. Compressed domain techniques have less complexity as full decoding and re-encoding is not required. As far we know, no perceptual hash based on motion coherency especially in compressed domain is yet explored. Our main contributions are: 1) motion coherent macroblocks are detected in compressed domain; 2) a perceptual hash is extracted from robust compressed domain features, like, luminance intra prediction modes, zero or nonzero value DC coefficients and chrominance modes of motion coherent macroblocks; 3) finally, the robustness of motion coherent perceptual hash is experimentally verified against different attacks. A feature based comparative study and time complexity of our method are also discussed.

Index Terms— Compressed domain, perceptual hash, motion coherent, copy attack, temporal frame averaging

1. INTRODUCTION

Copyright protection is essential in view of the growing popularity of video sharing websites. It deals with not only whether a copy occurs in a query video but also where the copy is located and where the copy is originated from. The technique of embedding digital signature into video streams for copyright protection, content authentication, and ownership integrity is called digital video watermarking. Video signals are often stored and transmitted in a compressed format. In many applications, complete decoding of video sequences is not feasible. Consequently, compressed video watermarking [1] have gained more attention as full decoding and re-encoding of the video signal is not required for watermark embedding or detection. Different video compression standards like H.264 [2] have emerged recently. The goal of each standard is to provide more compressed data along with better visual quality. H.264 is the most efficient and latest compression standard utilized in a wide range of applications.

Video watermarking technique has to be robust; subsequent processing of watermarked data should not impair the detection of embedded information. The copy attack [3] causes protocol ambiguity in a watermarking system. Such attacks can successfully remove the hidden watermark without sacrificing media quality. Copy attack [3] has been developed to create the false positive problem; i.e., a situation in which one can successfully detect a watermark from unwa-termarked data [4]. Watermark estimation is used to realize a copy attack, which is usually accomplished by means of a denoising procedure. Estimation of watermark is mostly performed using frame averaging [5]. Motion coherency is a desirable property to resist temporal frame averaging [5].

For authentication purpose, perceptual watermark should be robust against unmalicious distortions, but sensitive to ma-licious manipulations [6]. Lu and Hsu [7, 8] have embedded perceptual watermark to protect from watermark estimation attacks like copy attack. The perceptual watermark is gen-erated from the difference between AC coefficients in 8 8 blocks [7, 8]. As per the literature [7, 8], embedding a ro-bust perceptual watermark is one of the efficient ways to resist from copy attack.

The quantized statistics of wavelet coefficients are used as descriptors to generate image hashing and provide secu-rity using random tiling transform [9]. The hashing scheme in [10] has incorporated pseudo random projection into the fourier-mellin transform to achieve robustness to geometric operations, but suffers from some classical attacks like ad-ditive noise. Khelifi and Jiang [11] have introduced a vir-tual watermark detector that tuned the false alarm probabil-ity while detecting the pseudo random sequence. Hashes are generated by thresholding pseudo random sequences. Monga and Mihcak [3] proposed an image hashing method based on non-negative matrix factorization using local feature points in spatial domain to make a trade off between geometric invari-ance and robustness against classical attacks. Lv and Wang [4] have designed an image hashing method using local fea-ture points based on shift invariant feature transform and har-ris operator (SIFT-H) and embedded in shape and contexts based descriptors.

In [5], motion incoherent components are detected using motion compensated temporal frame averaging. Motion coherent blocks are quite stable to synchronization error and robust to common attacks [5]. The perceptual watermark can be generated from those blocks which are motion coherent so that such blocks remain robust to manipulations. Moreover, forced tempering on motion coherent blocks will cause sig-nificant degradation in visual quantity.

From the aforementioned literature on perceptual watermark, it is clear that most of the watermarks are generated in spatial domain features, where complete decoding and re-encoding of the compressed video is required. A compressed domain based perceptual watermark is therefore yet to explore. Moreover, if the portion of the video from where the perceptual watermark is generated is not motion coherent, then that portion can be easily altered. As far we know, no perceptual watermark is generated based on motion coherency in compressed domain.

The goal of this paper is to generate a motion coherent perceptual hash for low bit compressed videos, like H.264 that can withstand watermark estimation attacks, such as, copy attack. The complete procedure is performed in compressed domain to avoid further decoding and re-encoding. Frame averaging is widely used for watermark estimation. Motion coherency is an important property to avoid temporal frame averaging. The portions of video from where the hash is generated are motion coherent so that such portions remain robust against different distortions. Motion coherency is determined within a short video neighborhood, where frame correlation is very high when no scene change is detected. Furthermore, hash is generated from structural information of macroblocks in I-frames, which are very crucial for a video. Any manipula-tion in such features of I-frames will cause significant degra-dation in visual quality. The robustness of the extracted hash is verified against different attacks.

The rest of the paper is organized as follows. In Section 2, necessary preliminaries of H.264/AVC are presented. The proposed method is described in Section 3. Simulation results are given in Section 4. Finally, Section 5 concludes the paper.

2. H.264/AVC PRELIMINARIES

H.264/AVC [2] compressed videos have three types of frames, i.e., Intra frame, Predictive frame, and Bi-predictive frame denoted by I-frame P-frame, and B-frame, respec-tively. In coding each of the aforementioned frame types, first the color transform changes the color space from RGB (i.e., red, green, and blue) to YCbCr, where Y is the luminance component and Cb and Cr are the chrominance components. Each color space component is divided into non-overlapping blocks. Block sizes for the luminance component in I-frames are 4 4 and 16 16. H.264 encoder employs nine differ-ent intra prediction modes for 4 4 blocks and four intra prediction modes for 16 16 blocks to encode I-frames.

In the proposed method, only luminance intra prediction modes of 4 4 blocks are used. This is due to the fact that in I-frames, 16 16 blocks are chosen in smooth regions, while 4 4 blocks are selected in detailed areas [1]. The nine in-tra prediction modes denoted by prediction modes for 4 4 blocks are categorized into four groups as: DC mode (2), hor-izontal modes (1, 6, 8), vertical modes (0, 3, 4, 5, 7), and diagonal modes (3, 4) as similar modes may be converted to each other after embedding or re-encoding. DC mode, horizontal mode, vertical mode, and diagonal mode are denoted by f00, 01, 10, 11g, respectively. The prediction modes and their corresponding directions are depicted in Fig. 1.

Chrominance components of I-frames are divided into non-overlapping 8 8 blocks and chrominance blocks have four intra prediction modes denoted by chroma mode, shown in Fig. 2. The chroma modes, i.e., vertical mode, horizontal mode, DC mode, and plane mode are denoted f00, 01, 10, 11g, respectively. The coefficients in a 4 4 block in zigzag sequence order are shown in Fig. 3(c), where C(0) is DC coefficient and the rest are AC coefficients. In each 4 4 block, a one bit binary parameter (DC parameter) is assigned the value f0,1g based on zero and nonzero values of DC coef-ficient, is depicted in Fig. 3(b). Due to synchronization error, magnitude of DC coefficient may change, but zero coefficient does not become nonzero or vise versa. If due to some attack, zero DC coefficient becomes nonzero or vise versa, then the visual quality of that video will degrade significantly.

The aforesaid compressed domain features, i.e, luminance intra prediction modes, chrominance modes, and DC coefficients of a macroblock are robust features as per the literatures [1] and [12]. Therefore, if a hash is extracted from such features, then the hash will also be robust. The robustness of the hash is experimentally verified in Section 4.

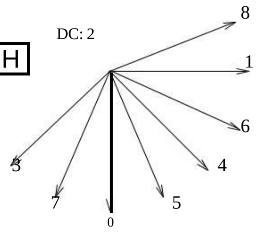
3. PROPOSED METHOD

The motivation of this paper is to generate a perceptual hash based on motion coherent blocks in H.264/AVC compressed videos. Within a short video neighborhood (SVN), video frames are visually coherent when no scene change is detected. Motion coherency can be detected easily within a SVN. As mentioned in Section 1, motion coherency is important to protect from watermark estimation attacks [5]. A block denotes a 4 4 block in the rest of this paper. The detection procedure of motion coherent blocks and generation of perceptual hash is performed in compressed domain, so computationally less expensive. In this paper, a motion coherent perceptual (MCP) hash is generated from structural information of blocks in I-frames of compressed videos.

Video frames are partitioned into non-overlapping blocks of size 4 4 within a SVN. Motion vector, motion threshold, prediction mode, chroma mode, and DC parameter of current I-frame f for a block B(f; i; j) are denoted by MV (f; i; j), M_{th}, P M(f; i; j), CM(f; i; j), and DC(f; i; j) respectively. Current I-frame and its consecutive previous and next I-frame are denoted by f, f k, and f +k, respectively. k is the length of the group of pictures (GOP). Mostly, H.264/AVC has only one I-frame in each GOP.

The complete process of hash generation is performed in

M	A	B	C	D	E	F	G	H
I	a	b	c	d				
J	e	f	g	h				
K	i	j	k	l				
L	m	n	o	p				



Mode	Name
0	Vertical
1	Horizontal
2	DC
3	Diagonal Down-Left
4	Diagonal Down-Right
5	Vertical-Left
6	Horizontal-Down
7	Vertical-Right
8	Horizontal-Up

Fig. 1. Luminance intra prediction modes of 4 4 blocks [2].

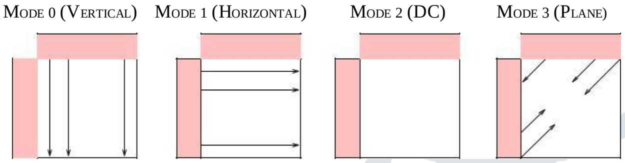
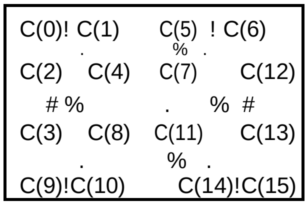


Fig. 2. Chrominance prediction mode of 8 8 blocks [2].



DC Parameter	
DC coefficient	Value
= 0	0
> 0	1
< 0	1

(a) (b)

Fig. 3. Part (a) of the figure illustrates the zigzag scan order of coefficients in 4 4 blocks [2] and part (b) of the figure depicts the value of DC parameter.

two steps; 1) estimation of motion coherent blocks in I-frames is done in Section 3.1 and 2) generation of perceptual hash is performed in Section 3.2.

3.1. Estimation of Motion Coherent Blocks in I-frames

The I-frames are intra coded, so value of motion vector is zero. The pseudo motion vector for I-frames are estimated from P-frames. Some authors have suggested different ways to find pseudo motion vector for I-frames. To detect reliable moving blocks, estimation of motion coherent blocks in I-frames directly from compressed video are processed in the following three steps:

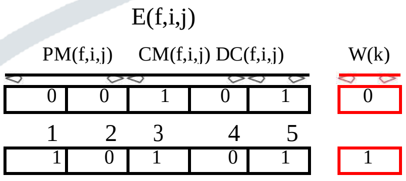
Step 1: Divide P-frames into non-overlapping blocks. Calculate motion vectors for all blocks in P-frames. Normalize each motion vectors to ensure that it points directly to the location in the immediate previous frame [13]. Normaliza-tion simplifies the referencing relationship since H.264/AVC supports multiple reference frames. Motion vectors of intra coded blocks are zero in P-frames. Estimate motion vector for intra-coded blocks from neighboring blocks. Smooth all normalized motion vectors by using a 3 3 median filter [14] and finally forms a complete motion vector field.

Step 2: Assign pseudo motion vector MV (f; i; j) to each block in I-frames by interpolating motion vectors at the same location from the nearest P-frames.

Step 3: Apply accumulation operator [13] to all motion vectors to enhance coherent motion and suppress noisy motion. Accumulation may increase nonzero motion vectors so remove those accumulated motion vectors which have a magnitude of zero before accumulation. Discontinuity in motion magnitude and direction is checked using spatial and temporal confidence measure [15].

3.2. Generation of Motion Coherent Perceptual (MCP) Hash

After the estimation of pseudo motion vectors MV (f; i; j) for I-frames, motion coherent blocks are tracked in compressed domain within a SVN. A block B(f; i; j) having absolute value of motion vector greater than motion threshold (MV$_{th}$) in current frame f will have its motion coherent block B(f k; i^0 ; j^0) in previous frame f k and B(f + k; i^{00} ; j^{00}) in next frame f + k. All blocks having absolute value of motion vec-tor between zero and M$_{th}$ are assumed blocks with a noisy motion vector so they are not considered as blocks in motion.



Parameters	Values
P M(f; i; j)	00
	01
	10
	11
CM(f; i; j)	00
	01
	10
	11
DC(f; i; j)	0
	1

(a) (b)

Fig. 4. Majority-win process: In part (a) of the figure, hash bit is 0 when the number of 0's is greater than the number of 1's in E(f; i; j) otherwise hash bit is 1 and part (b) of the figure shows all possible values of P M(f; i; j), CM(f; i; j), and DC(f; i; j).

In majority-win process, a five bit binary array E(f; i; j) for each block B(f; i; j) is calculated. First two bit is assigned for prediction mode, next two bit is allotted for

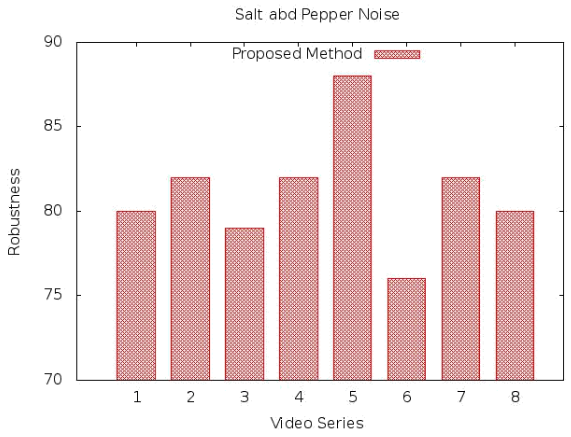


Fig. 5. The robustness of the proposed method against salt and pepper noise

chroma mode, and last bit gives the value of DC parameter. Once, E(f; i; j) is estimated, the kth hash bit W (k) is determined. If the number of 1 is greater than the number of 0 in E(f; i; j), then the hash bit is 1. Similarly, if the number of 0 is greater than the number of 1 in E(f; i; j), then the hash bit is 0. In Fig. 4(a), the majority-win process is shown, where the red box is hash bit W (k) and black box is E(f; i; j). The hash bit (W (k)) is 0 when P M(f; i; j) = 00, CM(f; i; j) = 10, and DC(f; i; j) = 1 i.e. three 0 and two 1 so number of 0 ¿ number of 1, hence W (k) = 1. Simi-larly, W (k) is 1 when P M(f; i; j) = 10, CM(f; i; j) = 10, and DC(f; i; j) = 1. All possible values of P M(f; i; j), CM(f; i; j), and DC(f; i; j) are shown in Fig. 4(b).

4. EXPERIMENTAL RESULTS

The proposed method is implemented using H.264/AVC [2] reference software JM 17.2. The video sequence is in Quarter Common Intermediate Format (QCIF) where frame Resolu-tion is 176 144. Frame Rate is 30 frames per second. The intra period is 7, i.e., fIBPBPBPg, Quantization Parameter (QP) is 28 for both I-frame and P-frame, and search range for motion estimation is [-32, 32]. Three reference frames are used. The fast full search in JM is employed and all intra and inter prediction sizes are enabled. B-frames are not used as reference frames. 8 video sequences of 140 frames are taken. The dataset [16, 17] for medical videos are used.

In our experimentation, motion threshold (M_{th}) is taken as 1. In Table 1, a feature based comparative result of the proposed MCP hash with recent existing schemes like [7, 8, 11, 4] are shown, where key points are selected blocks based on which the perceptual hash is generated.

The robustness of the extracted hash against salt and pep-per noise based on perceptual similarity defined in [8] Defini-tion 4 are depicted in Figure 5.

In the state of the art literature, no motion coherent perceptual hash extracted only from compressed domain features is present. Therefore, the proposed MCP hash is not com-pared with any literature. Furthermore, the proposed hash is not compared with any spatial domain hash as the comparison will be unfair with respect to time complexity. The time com-plexity of any spatial domain hash is much higher for com-pressed videos as complete decoding and re-encoding of the compressed video is required.

5. CONCLUSION

In this paper, a motion coherent perceptual hash for low bit compressed video like H.264/AVC is proposed. Motion co-herent blocks are estimated for I-frames using compressed domain features. A hash is generated from structural infor-mation of coherent motion macroblocks in I-frame in com-pressed domain.

In future, robustness of the proposed motion coherent per-ceptual hash from different attacks can be further analyzed. The proposed motion coherent perceptual hash can be used as watermark to resist watermark estimation attack like copy attack and can be compared state of the art copy attack resist-ing watermarking schemes.

Acknowledgements

This work is supported by Science and Engineering Research Board (SERB) file number ECR/2017/002419, project enti-tled as A Robust Medical Image Forensics System for Smart Healthcare, and scheme Early Career Research Award.

6. REFERENCES

[1] A. Mansouri, A. Aznaveh, F. Torkamani, and F. Ku-rugollu, “A Low Complexity Video Watermarking in H.264 Compressed Domain,” IEEE Transaction on In-formation Forensics and Security, vol. 5, no. 4, pp. 649– 657, 2010.

[2] I. Richardson, The H.264 Advanced Video Compression Standard, Wiley Publication, 2010.

[3] V. Monga and B. Evans, “Perceptual Image Hashing Via Feature Points: Performance Evaluation and Tradeoffs,” IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3452 –3465, 2006.

[4] Xudong Lv and Z. Wang, “Perceptual Image Hashing Based on Shape Contexts and Local Feature Points,” IEEE Transactions on Information Forensics and Security, vol. 7, no. 3, pp. 1081–1093, 2012.

[5] V. Pankajakshan, G. Doerr,¨ and P. Bora, “Detection of motion-incoherent components in video streams,” IEEE

Table 1. A Feature Based Comparative Study

	Methods			
Features	[7, 8]	[11]	[4]	MCP hash
Domain for key points estimation	Spatial	Spatial	Spatial	Compressed
Feature for key points estimation	Mesh	High Pass Filtered	SIFT-H	Motion Coherency
Hashing	Image	Image	Image	Video
Domain for hash generation	Compressed	Spatial	Spatial	Compressed
Features for hash generation	Difference between AC coefficients	Mean of coefficients	Shape and context based descriptors	Prediction Mode, Chroma Mode, and DC parameter

Transaction on Information Forensics and Security, vol. 4, no. 1, pp. 49–58, 2009.

[6] M. Tagliasacchi, G. Valenzise, and S. Tubaro, “Hash-Based Identification of Sparse Image Tampering,” IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2491–2504, 2009.

[7] C. Lu, S. Sun, C. Hsu, and P. Chang, “Media hash-dependent image watermarking resilient against both geometric attacks and estimation attacks based on false positive-oriented detection,” IEEE Transactions on Mul-timedia, vol. 8, no. 4, pp. 668–685, 2006.

[8] C. Lu and C. Hsu, “Near-Optimal Watermark Estimation and Its Countermeasure: Antidisclosure Watermark for Multiple Watermark Embedding,” 2007.

[9] R. Venkatesan, S. Koon, M. Jakubowski, and P. Moulin, “Robust image hashing,” in Proceedings of ICIP, 2000, vol. 3, pp. 664–666.

[10] A. Swaminathan, Y. Mao, and M. Wu, “Robust and se-cure image hashing,” IEEE Transactions on Information Forensics and Security, vol. 1, no. 2, pp. 215–230, 2006.

[11] F. Khelifi and J. Jiang, “Perceptual Image Hashing Based on Virtual Watermark Detection,” IEEE Transac-tions on Image Processing, vol. 19, no. 4, pp. 981–994, 2010.

[12] M. Noorkami, Secure and Robust Compressed-Domain Video Watermarking for H.264, Ph.D. Thesis. Georgia Institute of Technology, 2007.

[13] Z. Liu, Y. Lu, and Z. Zhang, “Real-time spatiotemporal segmentation of video objects in the H.264 compressed domain,” Journal of Visual Comunication and Image Representation, vol. 18, no. 3, pp. 275–290, 2007.

[14] P. Dong, Y. Xia, L. Zhuo, and D. Feng, “Real-time mov-ing object segmentation and tracking for H.264/AVC surveillance videos,” in Proceedings of ICIP, 2011, pp. 2309–2312.

[15] R. Wang, H. Zhang, and Y. Zhang, “A confidence measure based moving object extraction system built for compressed domain,” in Proceedings of ISCAS, 2000, vol. 5, pp. 21–24.

[16] “The Kvasir Dataset,” http://datasets.simula.no/kvasir/, 2018.

[17] “Hamlyn Centre Laparoscopic / Endoscopic Video Datasets,” http://hamlyn.doc.ic.ac.uk/vision/, 2018.