9-2012

Scalable content authentication in H.264/SVC videos using perceptual hashing based on Dempster-Shafer theory

Dengpan YE

Zhuo Wei
Singapore Management University, zhuowei@smu.edu.sg

Xuhua DING
Singapore Management University, xhding@smu.edu.sg

Robert H. DENG
Singapore Management University, robertdeng@smu.edu.sg

Scalable Content Authentication in H.264/SVC Videos Using Perceptual Hashing based on Dempster-Shafer theory

Ye Dengpan[1, 2, 3], Wei Zhuo[2], Ding Xuhua[2], Robert H. Deng[2]

1 School of Computer Science,
Wuhan University, Wuhan, Hubei, 430072, China
2 School of Information System School,
Singapore Management University, 188065, Singapore
3 The Key Laboratory of Aerospace Information Security and Trust Computing,
Ministry of Education, Wuhan, Hubei, 430072, China
E-mail:yedp2001@163.com

Received 30 November 2011; accepted 19 June 2012

Abstract

The content authenticity of the multimedia delivery is important issue with rapid development and widely used of multimedia technology. Till now many authentication solutions had been proposed, such as cryptology and watermarking based methods. However, in latest heterogeneous network the video stream transmission has b een coded in scalable way such as H.264/SVC, there is still no good authentication solution. In this paper, we firstly summarized related works and p roposed a scalable content authentication scheme using a ratio of different energy (RDE) based perceptual hashing in Q/S dimension, which is used Dempster-Shafer theory and combined with the latest scalable video coding (H.264/SVC) construction. The idea of "sign once and verify in scalable way" can be realized. Comparing with previous methods, the proposed scheme based on perceptual has hing outperforms previous works in uncertainty (robustness) and efficiencies in the H.264/SVC video streams. At last, the experiment results verified the performance of our scheme.

Keywords: Dempster-Shafer theory, scalability, video authentication, uncertainty

1. Introduction

Multimedia data is the most important information that human can recognize and understand. With the development of multimedia services and network communication technology nowadays, we ca n conveniently acquire various multimedia information such as video, audio, and others media, even at any time and any place. Due to adopting some new scalable video coding and transmission mechanisms (H.264/SVC) people can e ven consume same video contents in broadcast way via various terminates such as home TV, PC or mobile TV without transcoding [1, 2] (see Figure 1). However the security of delivery of these video contents has become a cruci al problem including solutions to the confidentiality, originality, integrity and so on. Traditional security method such as encry ption and data hashing, such as AES and SHA algorithms etc, are usually adopted to protect data contents. And in later years watermarking methods are developing especially in copyright applications. These two ways are m ain research directions in multimedia security till now. Although it may be achi eved good performance in traditional ways, it seem s not adaptive to the latest heterogeneous multimedia networks in which scalable video coding has been adopted.

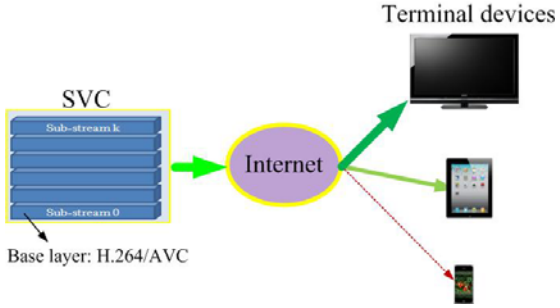

Fig.1. H.264/SVC scalable video distribution.

In this paper we will focus the multimedia content authentication issue. Multimedia authentication techniques have been used for protecting the originality and integrality of multimedia contents and can tell us whether or not the contents received are tampered or authentic [3]. The related works can be divided into two groups including digital signature and watermarking. The former utilizes encryption algorithm to extract hash codes from image or other multimedia data as signature, and the hash code is saved in header file or other extra space transmitting with the image. When authentication is needed, signature produced in the same way will be compared with signature saved before. If they match, then received multimedia data is authentic. The latter utilize semi fragile watermarking to verify the originality of received contents, and hot to design the watermarking algorithm is the key point in authentication. As we know, cryptology based hash or signature will be quite different when introducing even a bit modification at the input. This is not obviously practice in real multimedia transmission or processing applications. And pure watermarking based authentication scheme seems more vulnerable to attacks such as various watermarking attacks or delicate authentication attacks due to watermarks are inherent visual redundancy information in multimedia data.

Perceptual hashing (P-Hashing) is a latest technique for the authentication of multimedia content [12]. It works by computing hash values from features of multimedia data. Compared to conventional hashing and signature, the idea perceptual hashing has the advantage that they will not change when multimedia data undergoes common processing while detecting malicious modifications. This kind of perceptual hashing seems a good option for multimedia content authentication, especially in the public transmission channel. For video authentication, some video hashing had been proposed and can achieve trade-off between robustness and discrimination [19-25]. However, as we mentioned above, there is still no good authentication solution to the scalable video coding (H.264/SVC) streams content in latest heterogeneous networks. In the scalable video coding streams broadcast system, the server coding once and receiver can decoding in many ways. So how to secure the authenticity of SVC contents delivery and make it "Sign once and verify it in scalable way" is the key point of our works. Till now there are some works have been done including cryptology hashing, watermarking method. However, perceptual hashing based scalable authentication method for SVC videos seems a new solution someone seldom proposed.

The main contribution of this paper is that we proposed a new scalable perceptual video hashing can maintain invariance output result when the scalable visual content had not changed and can discriminate malicious modifications. Our scheme is adapted to the three dimensions' structure of H.264/SVC including spatial layers and quality layers as well as temporal layers. The features extracted from quality layers and spatial layers belong to same frame contents were constructed and further produced on invariance perceptual hash bits which are used for authentication in different receivers. Due to our hash bits generated from visual content features, different layers of same content can be authenticated using only one hash. If some layer or packet discarded, it will not affect authentication result. Regarding tampers in temporal axis such as dropping, reordering, the time stamp and NAL packets ID can tell the original order of frames and GOPs and discriminate malicious temporal tampers from normal fame rate changes.

The rest of this paper is organized as follows. In next section we summarized some latest related works about scalable video authentication. Section 3 described our proposed quality or spatial layer based perceptual hashing (Q/S based P-Hashing) scheme and how to achieve the invariance features from different spatial layers and quality layers. In the last section some experiments result in H.264/SVC streams was given as well as conclusions and future works.

2. Related Works

In [3], different approaches of multimedia authentication include conventional cryptography, fragile and semi-fragile watermarking and digital signatures that are based on the image content were summarized. Although the classification and performance comparison analysis for different authentication methods are given by the author in details, the authentication for scalable multimedia distribution was not mentioned. In the early related works for scalable image coding such as JPEG 2000, C Peng et al proposed a flexible and scalable authentication scheme based on the Merkle hash tree and digital signature [8, 19]. And it allows users to verify the authenticity and integrity of different sub-images extracted from a single

compressed codestream protected with a single digital signature.

To the best of our knowledge, the earliest paper about video scalable authentication was by Sun QB et al [5]. They considered three common MPEG video transcoding manipulations and combined error correction coding (ECC) and watermarking to design a content authentication scheme for scalable video streaming. Yan WQ et al proposed a scalable video signature that can authenticate video contents at three hierarchical levels (key-frame, shot and video) [7], but it is implemented in the pixel domain of video frames and not suitable for practice scalable video coding mode, as well as the authentication precision is a little coarse. According to MPEG-2 video authentication, Ye et al used multi-features extracted from coefficients of macro-blocks, time order information and motion information to produce a signature for MPEG authentication. MPEG-4 is the latest media stream processing standard which possesses the important "compress once, decompress many ways" property, Wu YD et al presented three scalable authentication schemes for MPEG-4 streams in multicast and lossy networks [9] and share the novel property of "sign once, verify many ways". In heterogeneous wireless networks, Gabriele et al proposed a loss tolerant video streaming authentication scheme mainly based on the video feature extraction idea [26]. In the experiments, feature difference indicator (FDI) and two attacks are adopted to verify the sensitive of the content feature extraction algorithm. This works can be referred in our feature hashing based scalable authentication scheme. However, all these above works are not implemented in the H.264/SVC we mainly considered in this paper and which is more complex. So it is questionable whether or not they are suitable for scalable authentication for H.264/SVC. Subsequently, we summarized related works exactly in H.264/SVC streams.

In [16, 17], Su-Wan Park et al embedded reversible watermarks into the intra prediction mode (IPM) of H.264/SVC bit streams for authentication and encrypted the IPMs of 4×4 luma block, the sign bits of texture, and the sign bits of MV difference values in the intra frames and the inter frames. If the watermark is correctly detected then the cipher content is decrypted. Although the visual quality of videos is not degraded too much, the watermarking scheme has a little bit-overhead. And it is questionable whether or not watermarking only IPMs can protect the whole content of SVC video streams. In [10, 11], Mokhtarian et al proposed an authentication scheme that accounts for the full scalability of video streams, and enables verification of all possible sub-streams that can be extracted from the original stream, in which the adaptation for spatial, quality and temporal dimension are all considered. In the scheme, the hashing generated from MGS (medium-grained scalable) packets to layers, then to CGS (coarse-grained scalable) and spatial layers, video frames to GOP according to temporal relation constructed a complex hash trees to authentication video data in any layers and this hash trees can also resist packets lossy as well as adapt to different layers discarding case. Although this kind of cryptology hash based scheme achieved security, adaptive and computation performance to some extent, there still some problems can not be easily resolved.

- Due to hash bits generated from every MGS packets and different layers should be attached in respective location, there is too much communication overload even some improvement works was made such as generating hash bits from group truncation unit not individual one.
- Due to cryptology hash can not resist even a bit change in video data, every bit modification of MGS blocks can cause the final frame or GOP can not pass the authentication while the video content is still original.

According to these problems, we combined perceptual video hashing and watermarking to design an authentication scheme adaptive to the three dimensions (S, Q and T) scalability of H.264/SVC video streams.

3. Proposed Scheme

Our scheme belongs to content based authentication. So how to design perceptual hashing that can both meet the requirement of scalability and sensitive to content changes. In addition, the system need some robustness to noise such as some bits change without affect video contents and this maybe often occur in real applications. Our scheme partly was inspired by the works in [5, 6, and 13] and based on the DCT robust features. Before illustrating our scheme in detail, the structure of H.264/SVC streams should be introduced. When we say a frame in SVC, it is constituted by different spatial, quality layers denoted by DID, QID. With the change of frame rate, the temporal scalability is necessary. The

frame in higher temporal layer can only be predicted from the one in lower temporal layers and the TID is the ID of temporal layers. In content authentication, ideal perceptual hash extracted from same content should be the same to each other, no matter what the resolution or quality is. So it is nature idea to extract same features from different spatial (resolution) or CGS (quality) layers included in one frame and can produce several same hashing bits sequence. It is so called Q/S based Hashing in our scheme. Notice that H.264/SVC allows up to 8 spatial and 16 quality layers for one frame, we select the frame of base layer to extract main features and select enhancement frame to extract enhancement features for authentication. This will reduce the hash computation while not missing the significant content need to authenticate. Regarding the temporal tampering, with the TID we can know the correct temporal orders. If attackers forger a TID and disorder or replace the frames, the Q/S based hashing can do the authentication works. In [5, 6] three transcoding include re-quantization, frame resizing and dropping had been considered. In H.264/SVC we firstly considered key frames in two cases, one is changed with different resolution (spatial change) and the other is same resolution with different quality (quality change). The latter case is further divided into different CGS layer and MGS layer.

a) The invariance feature between frames with same resolution in different CGS layers

In H.264/SVC, 4x4 blocks are adopted and transformed into the matrix with16 DCT coefficients. Each coefficient is quantized by a same quantization step. In key frame intra 4x4 blocks (denoted by Io) should be intra coded by different intra prediction modes, the intra prediction pixel block's difference is denoted by Po and residual blocks are transformed into DCT coefficients we denoted by r in Figure 2. The r then quantized by Q1 and denoted the quantization noise is $\Delta_1 \in [-Q_1/2, Q_1/2]$, so

$$b = r - \Delta_1 \tag{1}$$

And denoted quantization noise introduced by Q2 is $\Delta_2 \in [-Q_2/2, Q_2/2]$, so

$$a + b = r - \Delta_2 \tag{2}$$

And when decoding in different quality layers at receiver, [Po+IDCT(a+b)]should be with the same visual feature as the [Po+IDCT(b)] except that different quantization distortion (in Figure 3).

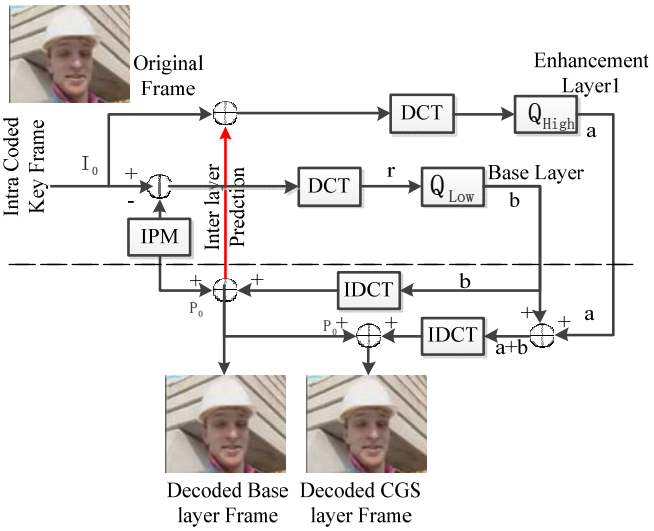

Fig.2. Quality scalability coding in H.264/SVC.

Due to many AC coefficients are zero especial those high frequency ones after quantization, the features extracted from the group of DCT coefficients are more robust than single DCT coefficient and preferable selected units for producing perceptual hash, And it can be predicted that with the increasing of the number of blocks, the invariance relation of DCT frequency energy in the group pair can be maintained due to the quantization distortion can be regarded as the noise with mean value is zero, which can be obtained by the Dempster-Shafer Evidence Theory (D-S) theory. D-S theory considers one or more evidences to enhance the information and decrease the uncertainty. In this paper, we can ensure the robust of features and enhance image quality and energy. Firstly, D-S theory considers the frame of discernment, in this paper, the frame of discernment is composed of group DCT coefficients, and we can combine these together to get better features. Dempster-Shafer theory has five related concepts: basic probability assignment (BPA), combine rules, belief, plausibility, and belief interval [27].

- *BPA*

A basic probability assignment (BPA) represents the degree of supporting for a piece of evidence. Its definition is as following:

$$m(\phi) = 0 \text{ and } \sum_{A \in 2^{\Theta}} m(A) = 1$$

According to the definition, a BPA for {A, B} represents the degree of supporting for {A, B}, not {A} or {B}, or the sum of BPA of {A} and {B}.The sum of all BPAs will equal 1.0. In this paper, we consider the rate of visual feature in the total layer as the input of BPA.

- *Combine Functions*

Multiple evidences can be co mbined using D-S combination rule shown in following equation (1) and (2), which is also called th e orthogonal sum of evidences or mass function, shown as following:

$$m=m_1 \oplus m_2 \oplus \ldots \oplus m_n \quad (1)$$

$$m(A)=\begin{cases} 0 & (X \cap Y=\phi) \\ \dfrac{\sum_{\cap A_i=A} \prod_{i=1}^{n} m_i(A_i)}{1-\sum_{\cap A_i=\phi} \prod_{i=1}^{n} m_i(A_i)} & (X \cap Y \neq \phi) \end{cases} \quad (2)$$

Through the combine functions, we can get the comprehensive features from the group of DCT coefficient and make the features more robust.

- *Belief Function*

Belief represents the total support for a h ypothesis, and will be drawn fro m the BPAs for all subsets o f that hypothesis. The b elief function is defined as equation (3).

$$Bel(A)=\sum_{B \subseteq A} m(B) \quad (3)$$

Through the belief function, we can get the feature extracted from the single DCT coefficient, which is shown as probability.

- *Plausibility Function*

In contrast to belief, plausibility represents the degree to which a h ypothesis cannot be disbelieved or false. Unlike the case in Bayesian probability theory, disbelief is not the complement of belief, but represents the degree of support for all hypotheses that do not intersect

with that hypothesis. For each $A \subseteq 2^{\Theta} h$, plausibility

function is defined as:

$$Pls(A)=\sum_{B \cap A \neq \phi} m(B) \quad (4)$$

Through the plausibility function, we can get the degree of feature that different quantization distortion.

- *Belief Interval*

The above measures provide DS theory with an explicit measurement of ignorance about a hy pothesis and its complement sets. Belief in terval is d efined as an interval [Bel(A), Pls(A)][28]. This can also b e interpreted as the imprecision on the 'true probability' of A.

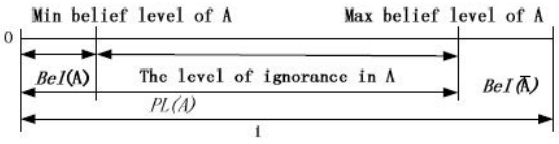

Figure 2. The relation between Bel and Pls

The final belief, plausibility, and belief interval for each of the hypotheses can the n be calculated based on the basic probability assignment using the above equations. Ignorance for the whole set can also be derived. In most cases, after a dding new evidence, the ignorance is reduced [29]. Based on DS theory combination rule, more than two hypotheses can be integrated.

When hypotheses are c ombined by m ultiple pieces of evidence, the DS theory can be used to fuse these evidences. The final result represents the synthetic effects of all evidences [30].

b) The invariance feature between frames with sam e resolution with different MGS quality

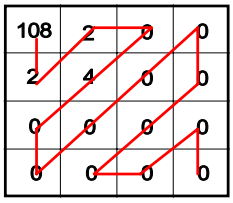

Fig.3. DCT coefficients discarding in zig-zag order with MGS layer.

We divided n blocks into 2 gr oups with n/2 blocks respectively. These n blocks can be generated one feature bit according to th e different energy (DE) between two groups (basic idea can be referred in [18]). It can be expected that the relation of different energy of two groups are invariance as long as the proper n value is given, even if some higher frequency coefficients in certain MGS packet will be d iscarded when quality switch, due to the DC c oefficients are m ost important and usually largest coefficients in 4x4 intra blocks.

c) The relation of residual DCT energy in same key frames with different resolutions

The energy of different block groups reflects the content of video frame such as edges, textures and flat characteristic. It can be expected that the relation of DE will be invariant to the scaling of video frames, which is the same as in the different quality layers. However, in the H.264/SVC bit streams, the residual DCT coefficients in highe r spatial layer are based on the frame difference predicted from the fram e up-sampled from base l ayer, and the quantization factors is even different in respective spatial layer. So the statistica l relation between the residual DCT en ergy and the intra

residual DCT energy in different spatial layers is complex and difficult to describe directly from exist compressed data. In this paper we only consider different quality based invariance relation feature and different resolution based one will be studied in future work.

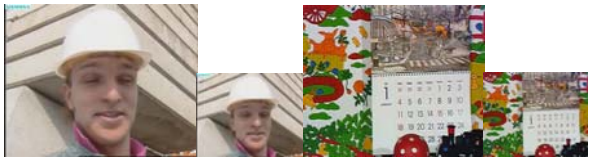

Fig.4. Key frames of "Foreman" and "Mobile" with different resolution (CIF and QCIF).

So we generate different perceptual hash according to different spatial layers and attach the hash bits within one frame. Each hashing bit sequence can be sensitive to tampers in frame with respective resolutions and can further tell the location of tampered blocks group. If I= {I1, I2 ... IM ...} are the input N video key frames and then bi are the 4x4 pixel blocks within one key frame. Denote the energy of one block Ei as follow, λ_j is the weighting factor of the j-th DCT coefficient.

$$E_i = \sum_{0}^{15} \lambda_j C_j^2, j \in [0,15] \tag{4}$$

Select n blocks according to the method in [18], the total energy of sub-group A or B with n/2 blocks and the energy of whole group A+B then can be denoted as follow.

$$E_{AorB} = \sum_{1}^{n/2} E_i, i \in [1, n/2] \tag{5}$$

$$E_{A+B} = \sum_{1}^{n} E_i, i \in [1, n] \tag{6}$$

Different from feature bit generating method, the ratio of group DCT energy difference (RDE) is utilized here. So the final robust feature bit can be extracted as (l denote group index and τ_l denote threshold for group l):

$$f_l = \begin{cases} 0, if \left|E_A - E_B\right| / E_{A+B} \le \tau_l \\ 1, if \left|E_A - E_B\right| / E_{A+B} > \tau_l \end{cases} \tag{7}$$

And final perceptual hash bit sequence was produced by encryption algorithm E (.), r denotes certain resolution and K= {k1, k2... km...} denote the secret key sequence assigned for different key frame M to generate secure perceptual hash Hr.

$$H_r = E_K(F_r), F_r = \{f_1^r, f_2^r, ...f_l^r ...\}, r \in \{1,2,...\} \tag{8}$$

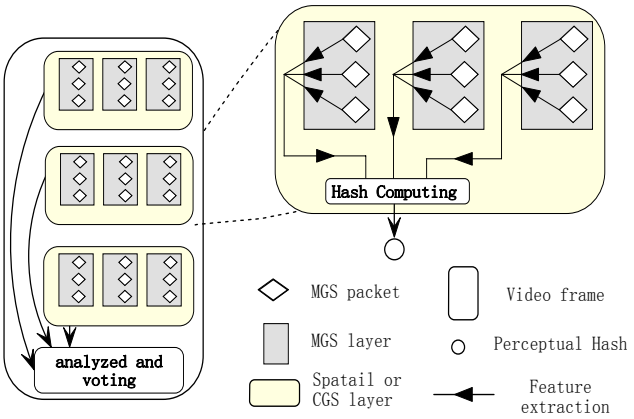

Fig.5. Authenticating a video frame from spatial and quality layer.

When authenticating MGS frame, total of MGS layer should be together to construct one complete block coefficients sequence, include every coefficients such as first 3, latter 3 and at last 10 ones in zigzag order (in figure 3). The video frame based authentication scheme can be described as figure 5, in which one hash bits sequence denoted the content of one resolution frame. All the results based on the every hash bits sequence are finally analyzed and voted to outcome the decision.

4. Simulation and Analysis

We simulate three types of test video, namely "Foreman", "Mobile" (Figure 4), and encoded them using H.264/SVC reference software, called JSVM. Each encoded stream consists of four temporal layers (GoP size 8) and two spatial layers providing CIF and QCIF resolutions. The first frame (frame 0) is the key I frame of GOP which was utilized to extract perceptual hash.

Table 1. Experiment parameters for authentication scheme

Test Video	Foreman (QCIF)	Foreman (CIF)	Mobile (QCIF)
Resolution	176x144	352x288	176x144
DCT feature	Ratio of DE	Ratio of DE	Ratio of DE
Block size	4x4	4x4	4x4
Group size	16x16	32x32	16x16
Weighting factor λ_j	{0.8, 1.1, 1.1, 1.1, 1,1,1,1,1,1,1,1,1,1,1,1}	{0.8, 1.1, 1.1, 1.1, 1,1,1,1,1,1,1,1,1,1,1,1}	{1,0.7,0.7,0.7,1,1,1,1,1,1,1,1,1,1,1,1}
Threshold τ_l	0.1	0.1	0.09
Key I frame index	0	0	0
P-Hashing bit length	99	99	99
Authentication unit	16x16 pixel block	32x32 pixel block	4x4 pixel block
Feature matching when quality changes	All bits	Except 2 bit	All bits
Tamper type	Delete and replacement	-	-
Number of P-hashing per key I frame	number of Spatial layers	-	number of Spatial layers

A) The invariance of RDE based Q-hashing in deferent quality layers

From the feature extraction experiment, in 4 quality frames reconstructed from 4 layers we can get (base layer, CGS, MGS1, and MGS2) 4 feature RDE value sequence. In figure 6, the threshold τ_l are all 0.1, and the weighting factor $\lambda_j, j \in [0,15]$ is {0.8, 1.1, 1.1, 1.1, 1… 1}. It can be seen that according to every group index four feature points are all located in one side of threshold line. This is the best result and so all hash bits sequence generated in formula (7) are all matched. The results verified our RDE feature based P-hashing is quality scalability and adaptive to SVC coding. We further encoded and decoded test video "Mobile" in figure 6, set the weighting factor λ_j={1, 0.7, 0.7, 0.7, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}, the τ_l are all 0.09 and get the same best results (all hashing bits are m atched). However, when applied these into the cif of "foreman", there are two groups unmatched.

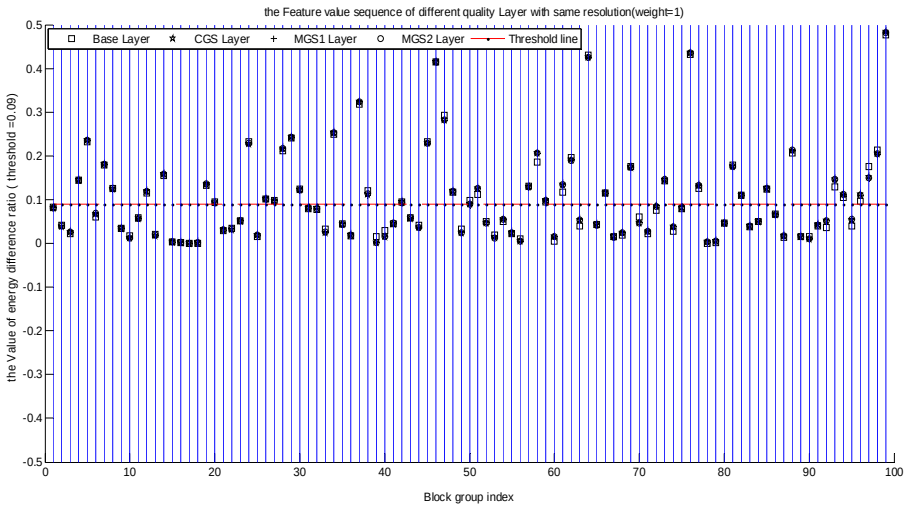

Figure 6. The invariance of RDE between different quality frames (Base, CGS, MGS1, and MGS2)

1

Then we generated corresponding feature hashing bits from several key frame sequences of "foreman" including frame 1, 9, 17, 25, 33, 41 and 49, without tampering. And comparing the feature hashing bits extracted from base layer w ith different quality layer (CGS, MGS1, and MGS2) as well as spatial layer (CIF), the mean value of the verification ratio can be shown as Table 2. The verification ratio means how many video authentication units can be verified without any malicious modifications. The experiment results maintain above high ratio to some extent which verified the efficiency of the proposed scheme.

Table 2. Mean verification ratio of Key frame sequences (1, 9, 17, 25, 33, 41 and 49)

Mean verification ratio	Foreman	Mobile
Base layer	100%	100%
CGS	97.4%	97.11%
MGS1	97.98%	98.56%
MGS2	95.53%	95.67%
Spatial Layer (CIF)	95.09%	94.81%

B) The sensitive of RDE to content tampers

At first, we verified the efficient of our scheme through two classical ta mpers including del eting and replacement attacks in two blocks contents in base layer frame and deleting attacks in respective quality frame (CGS, MGS 1 and M GS 2). The attacks and authentication results are sh own in figure 7 and 8. In figure 7 (a) and (b), the two top and left blocks were directly deleted or dedicate replaced with sophisticated contents, and both modified locations can be rem arked in the figure 7 (c) and (d) in our proposed scheme.

Secondly, according to the related two forging attacks (superimposing attack and picture in picture attack) proposed in [26], we tested the sensitive of our proposed RDE based perceptual hashing to these attacks methods which can prov ide different attacks intensity by adjusting the factor. The authenticated video is Foreman, attack video is Akiyo. In table 3, the RDE feature based hashing bit error rate (FER) in different quality and spatial layers under different attacks was given. For better describing the performance of R DE hashing to Pic-in Pic attacks, here in the b) of table 3 we used the globe FER and local FER to indicate the feature hashing bits changes under attacks with different intensity. It is obvious that the size a of Pic-in-Pic attacks can be used to convert eac h other between Globe FER and L ocal FER. It can be seen that increasing the mf and a, increase the Globe FER values with no exceptions. However, the local FER decreased when the size of Pic-in-Pic attacks increase d. This is because the modification area becam e smaller so that the rate became smaller with the same mismatch feature bits. Our one authentication unit is 16x16 block size. Comparing with the normal verification ratio, the superimposing attacks can be di scriminated from the FER values due to the lowest FER is 13.97% which was higher than the normal ratio (100%-94.81%=5.19%). Nevertheless, in Pic-in-Pic attacks, the situation was more complex when the attack a rea decreased to 1/16 and below. Here the globe and local FER should be both

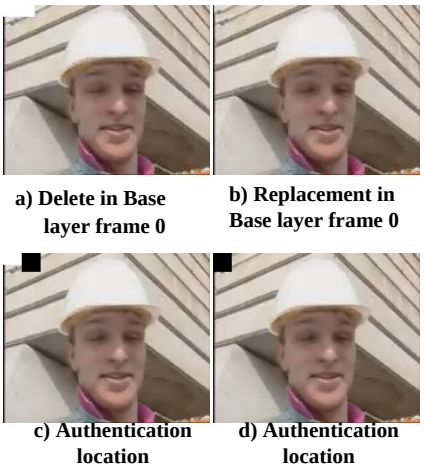

Figure 7. Tampers (delete and replacement) in base layer frame 0 and authentication

Figure 8. Tampers (delete) in respective quality frame 0

and authentication results

considered. So the local RDE hashing bit further extracted to judge whether some area are noise or intended modifications. And the local FER were basically high under the local attacks such as Pic-in-Pic attacks (above 50%) or deleting or replacement and etc (in Figure 7 and Figure 8).

Globe FER=mismatch feature bits/Total number of authentication units (9)

Local FER= mismatch feature bits/Total number of authentication units/a (10)

Figure 9. Foreman-Akiyo Superimposing attack using a merge factor (from left to right 0.2, 0.5, 0.8)

Figure 10. Foreman-Akiyo Picture in picture attack using a factor (from left to right 1/4, 1/16, 1/64)

Figure 11. Mobile- Akiyo Superimposing attack using a merge factor (from left to right 0.2, 0.5, 0.8)

Figure 12. Mobile- Akiyo Picture in picture attack using a factor (from left to right 1/4, 1/16, 1/64)

Table 3. Mean error bit rate of RDE based hashing of Key frame sequences (1, 9, 17, 25, 33, 41 and 49) under different attacks

a) Superimposing attacks

Superimposing Attacks intensity	Foreman/Akiyo	Mobile/Akiyo
	Global FER	Global FER
mf=0.2	13.97%	17.58%
mf=0.5	39.95%	43.89%
mf=0.8	44.34%	48.24%

b) Picture in picture attacks

Pic-in-Pic Attacks intensity	Foreman/Akiyo		Mobile/Akiyo	
	Global FER	Local FER	Global FER	Local FER
a=1/4	12.30%	49.18%	12.94%	42.70%
a=1/16	4.47%	59.00%	4.57%	50.27%
a=1/64	1.30%	85.71%	1.88%	93.06%

C) Related measure matrix compared with the SvcAuth scheme in [10] using JSVM platform.

Table 4. Comparison measure matrix

Performance	Q P-Hashing Scheme	SvcAuth
Computation Cost	Real time (synchronization to encoding and decoding)	higher
Robustness Against Loss	Can be Robust to MGS layer discarding	good
Communication Overhead	99bits per Spatial layer in one I frame	Very high

The robustness and sensitive of the perceptual hashing are the two key points in designing authentication scheme. These performances are base d on the feature selection and ex traction. In this p aper we used ratio of different energy of group blocks to extract the robust feature and computed one hashing sequence for each quality frame. The e xperimental results show that this energy ratio based hashing has good perceptual characteristics and ro bust to fram e quality changes in H.264/SVC while sensitive to content tampering such as deleting and replacement. At same time, the P-has hing can further s how the tampers location. In prev ious method, even a bit modification can c hange the final signature and the authentication result. This is the major difference of content based authentication from cryptology schemes. Due to generating one hashing for different quality layers, the computation cost and overhead caused by hash bits data was g reatly reduced (in Table 2). However, we should notice that the discrimination of R DE based P-Hashing between robustness and sen sitive is still n ot so smart when all

tampers case are considered. In future works we will consider the dynamic adaptive threshold and revise the weighting factor and expect to get better results, especially in different spatial layers.

5. CONCLUSION

In this paper we proposed a scalable robust perceptual hashing (P-Hashing) scheme which can be used in the scalable video stream authentication. Our feature based P-Hashing was designed adaptively to quality scalability and resolution scalability simultaneously, so that some invariance can be achieved between different quality layers and spatial layers in the latest H.264 /SVC coding structure which was verified in the experiments. It can be seen in the simulations that the malicious modification and coding noises can be discriminated using this scalable P-hashing. Comparing with previous works in scalable video authentication [10, 11], the proposed scheme is more efficient in the computation cost and robustness and communication overload. In the experiments we used the ratio of different energy as the features to construct P-Hashing. Here some improve feature extraction and optimal algorithms need to be applied in this scheme in future works as written in experiment analysis. How to achieve better tradition between robustness and sensitivity is the key point of the scalable video authentication and the feature extraction is the most important works should be done firstly in the future works. Consequently a set of DCT or others suitable features extracted from H.264/SVC coding data will be applied in the next step compared experiment and be analyzed.

6. ACKNOWLEDGMENTS

This work was funded by the A*STAR SERC Grant No.1021010027 of Singapore, and Natural Science Foundation of China (60903197, 61272453).

7. REFERENCES

1. M. Wien, H. Schwarz, and T. Oelbaum, “Performance analysis of SVC,” IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9, pp.1194–1203, Sep. 2007.
2. H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
3. J. Apostolopoulos, “Architectural Principles for Secure Streaming &. Secure Adaptation in the Developing Scalable Video Coding (SVC). Standard”, IEEE ICIP, 8-11 October 2006, Atlanta, GA, USA, 729-732.
4. Adil Haouzia and Rita Noumeir, “Methods for image authentication-a survey”, Multimedia Tools and Applications, 2008, 39(1):1-46.
5. Q. Sun, D. He, Z. Zhang, and Q. Tian, “A robust and secure approach to scalable video authentication”, in Proc. Int. Conf. Multimedia Expo (ICME03), Baltimore, MD, Jul. 2003, pp. II209–II212.
6. Qibin Sun, Dajun He, Qi Tian, “A Secure and Robust Authentication Scheme for Video Transcoding”, IEEE Transactions on Circuits and Systems for Video Technology, 16(10):1232 – 1244, 2006.
7. Yan Weiqi et al, “A scalable signature scheme for video authentication”, Multimedia Tools Appl, 34(1): 107-135, 2007.
8. R. H. Deng, D . Ma, W. Shao and Y. Wu, “Scalable trusted online dissemination of JPEG2000 images”, ACM Multimedia Systems Journal, Vol. 11, pp. 60-67, November 2005.
9. Y. Wu and R. H. Deng, “Scalable authentication of MPEG-4 streams”, IEEE Transactions on Multimedia, 8(1):152-161, Feb. 2006.
10. K. Mokhtarian and M. Hefeeda, "Authentication of Scalable Video Streams with Low Communication Overhead“, IEEE Transactions on Multimedia, 12(7), pages 730—742, November 2010.
11. M. Hefeeda and K. Mokhtarian, "Authentication schemes for multimedia streams: quantitative analysis and comparison“, ACM Transactions on Multimedia Computing, Communications and Applications, 6(1), article 6, pages 6:1--6:24, February 2010.
12. Zhu bin, “Encryption and Authentication for Scalable Multimedia Current State of the Art and challenges”, SPIE 5601, 2004.
13. Ye DP et al, “multi-feature based authentication scheme for MPEG videos”, Chinese Journal on communications, 29(2):59-65, 2008.
14. Lina Wang, Xiaqiu Jiang, Shiguo Lian, Donghui Hu and Dengpan Ye, “Image authentication based on perceptual hash using Gabor filters”, Soft Computing, 15 (3):493-504, 2011.
15. Peter Meerwald and Andreas Uhl, “ Robust Watermarking of H.264 SVC- Encoded Video quality and resolution scalability”, IWDW 2010, pp. 15 9–169, 2011.
16. Su-Wan Park and Sang-Uk Shin, “Authentication and Copyright Protection Scheme for H.264 AVC and SVC”, JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 27, 129-142, 2011.
17. Su-Wan Park and Sang-Uk Shin, “Combined Scheme of Encryption and Watermarking in H.264/Scalable Video Coding (SVC)”, New Directions in Intelligent Interactive Multimedia, SCI 142, 351–361, 2008.
18. Yuewei Dai; Stefan Thiemert; Martin Steinebach, “Feature-based watermarking scheme for MPEG-I/II video authentication”, SPIE 5306, pp.325-335.
19. C.-Y. Lin and S.-F. Chang, “Issues and Solutions for Authenticating MPEG Video”, SPIE Security and Watermarking of Multimedia Contents, EI '99, San Jose, CA, vol. 3657, 1999, pp. 5465.
20. Cheng Peng, Robert Deng, Yongdong Wu. Weizhong Shao, “A Flexible and Scalable Authentication Scheme for JPEG2000 codestream,” ACM Multimedia. pp.433-441, Nov. 2003.
21. S. C. Cheung and A. Zakhor, “Efficient video similarity measurement with video signature”, IEEE Trans.

Circuits Syst. Video Technol., vol. 13, no. 1, pp . 5974, Jan. 2003.
22. K. Kashino, T. Kurozumi, and H. Murase, "A quick search method for audio and video signal b ased on histogram pruning", IEEE Trans. Multimedia, vol. 5, no. 3, pp. 348357, Sep. 2003.
23. C. D. Roover, C. D. Vleeschouwer, F. Lefebvre, and B. Macq, "Robust video hashing based on radial projection of key frames", IEEE Trans. Signal Process., vol. 53, no. 10, pp. 40204037, Oct. 2005.
24. B. Coskun, B. Sankur, and N. Memon, "Spatio-temporal transform based video hashing", IEEE Tr ans. Multimedia, vol. 8, no. 6, pp.11901208, Dec. 2006.
25. Peter Meerwald, Andreas Uhl, "Robust watermarking of H.264/SVC-encoded video: quality and resolution scalability", IWDW '10, pp. 159-169, Seoul, Korea, Lecture Notes in Computer Science, 6526 , Springer, October 1 - 3, 2010.
26. Gabriele Oligeri, Stefano Chessa, Gaetano Giunta, "Loss tolerant video streaming authentication in heterogeneous wireless networks", Computer Communications, 2011, 34(11): 1307-1315.
27. M. Beynon, D.Cosker, D. Mar shall. An exp ert system for multi-criteria decision making using De mpster Shafer theory. Expert Systems with Applications 2001 (20), pp.357-367.
28. RA. Dillard. The Dempster-Shafer theory applied to tactical data fusion in an inference system. MIT/ONR, 1982, pp.170-174.
29. M. Beynon, B. Curry, P. Morgan. The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modeling [J]. Omega, 2000 (28), pp.37-50.
30. X. Fan, M. Zuo . Fault dia gnosis of machines based on D－S evidence theory. Part 1: D－S evidence theory and its improvement [J], Pattern Recognition Letters, 2006 (27), pp.366－376.