FEATURE DIFFERENCE ANALYSIS IN VIDEO AUTHENTICATION SYSTEM

Dajun He[1], Zhiyong Huang[2], Ruihua Ma[1] and Qibin Sun[1]

[1]Institute for Infocomm Research (I2R)
21 Heng Mui Keng Terrace, Singapore, 119613
Email: {djhe, ruihua, qibin}@i2r.a-star.edu.sg

[2]School of Computing
National University of Singapore, Singapore, 117543
Email: huangzy@comp.nus.edu.sg

ABSTRACT

In most video authentication systems, the difference between features of the original and received videos is often used to decide the authenticity. In this paper, by employing mutual information to represent the similarity between the original and received video frames, we theoretically analyze the relationship between the feature difference and the video distortion. The relationship we derived is applied to estimate the maximum allowable feature difference in a video authentication system and to show how the feature difference varies with quantization step. Experimental results demonstrate that the derived relationship is reasonable and helpful for designing a robust video authentication system.

1. INTRODUCTION

In most video applications, original video may undergo various processing before reaching final users. So a distortion between the original video and the received video may exist. This may be either incidental distortion, which is introduced by normal video processing such as compression, resolution conversion and geometric transformation, or intentional distortion, which is introduced by malicious attack. A robust video authentication system should, hence, tolerate the incidental distortion while being capable of detecting the intentional distortion. Many researchers have designed robust image/video authentication algorithms to meet above requirements based on watermarking strategy, which is sometimes termed "self-embedding" authentication system [1] as shown in Figure 1. In this system, a robust and important feature of the image/video is extracted and embedded into the image/video at the sending site; the detector retrieves this original feature from the watermark and compares it with the feature extracted from the received image/video to determine the authenticity of the image/video. If the difference exceeds a threshold, the received image/video will be claimed as an un-authentic image/video. This threshold, which should be determined before an authentication system is designed, refers to the maximum feature difference between the original video and the video that has undergone various normal video processing.

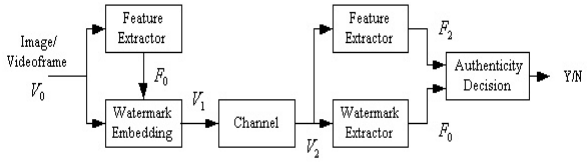

Figure 1 Block-diagram of a "self embedding" authentication system

Dittmann [2] and Queluz [3] used the edge of image as the feature to generate a digital signature in their authentication systems. Although they claimed that this feature is robust to high quality compression and scaling, a threshold is still used to improve the robustness. Nevertheless, the authors did not mention how this threshold is obtained. In our previous works [4, 5], a semi-fragile watermark instead of digital signature is used in a video authentication system. The Error Correction Coding (ECC) scheme and cryptographic hash scheme were employed to improve the system's robustness and security. However, a threshold was also needed to decide the authenticity of video objects. And this threshold was determined in an empirical way. It requires a lot of computation; and the effectiveness of the threshold itself also poses a problem.

In this paper, we will adopt a theoretical approach to derive the threshold based on analyzing the relationship between the feature difference and the video distortion. Moreover, using the derived relationship, we will also show how the feature difference varies with the quantization step in video compression. The paper is organized as follows: the relationship between the feature difference and the video distortion is analyzed in Section 2. Experimental results are presented in Section 3. We conclude the paper in Section 4.

2. FEATURE DIFFERENCE ANALYSIS

2.1 Mutual information and feature difference

Mutual information is a basic concept in information theory. It measures the amount of the information that one random variable contains about another random variable [6]. The definition is as follows:

$$I(X;Y) = \sum_{x\in\widetilde{X}}\sum_{y\in\widetilde{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \quad (1)$$

where X is a discrete random variable with set $\widetilde{X}$ and probability density function $p(x)$; Y is a discrete random variable with set $\widetilde{Y}$ and probability density function $p(y)$; and $p(x,y)$ is the joint probability density function of X and Y. To compute mutual information, the probability density functions must be computed or estimated. It is worth noting that in real applications, the estimation is crucial but not always trivial.

Mutual information has been widely used to measure the similarity between two images, especially in image registration. For example, two images are considered geometrically aligned if the mutual information of image intensity values is maximized [7].

As mentioned in the Introduction Section, in video authentication, we are concerned about the feature difference between the original video frame and the distorted video frame to determine whether the distorted one is still authentic. When the feature is well selected to represent the video frame, we should be able to express the feature difference between two video frames in terms of mutual information of these two video frames in the following linear form:

$$d(F_0,F_2) = L_f - \alpha * I(V_0;V_2) \quad (2)$$

where V_0 and F_0 represent the original video frame and its feature respectively; V_2 and F_2 represent the distorted video frame and its feature respectively; and $I(V_0;V_2)$ is the mutual information between the original video frame and the distorted video frame. L_f is the size or length of the selected feature.

Intuitively, the relationship between the feature difference and the video distortion should meet following two requirements:

a. The difference should be zero if the original video frame and its distorted video frame are identical.
b. The difference should be maximized if the original video frame and the distorted video frame are independent.

Thus, we have

$$\alpha = \frac{L_f}{I(V_0;V_0)} = \frac{L_f}{H(V_0)} \quad (3)$$

where $H(V_0)$ represents the entropy of the original video frame [6].

Equation (3) can also be acquired from the angle of information theory because the entropy of an image can be used to represent the complexity of this image. Given a video frame, more complexity it is, the larger its entropy is. On the other hand, the size of feature extracted from a video frame increases with the complexity of this video frame. Thus, a relationship between entropy of a video frame and the size of feature, which is similar to equation (3), must exist.

Substituting equation (3) into (2), equation (2) can be rewritten as

$$d(F_0,F_2) = L_f - L_f * \frac{I(V_0;V_2)}{H(V_0)} \quad (4)$$

This is the relationship between the feature difference and the video distortion. From this equation, we can know that the feature difference between two video frames only depend on the mutual information of these two video frames because $H(V_0)$ will be fixed once a video frame is given. Therefore, the computation of the feature difference becomes that of the mutual information. Again, mutual information computation is a process of the distribution estimation for video frames.

Now let's take a look back at L_f. In [4], features are all converted into Quasi-Gray binary code, called feature code, to ensure that one-bit change in feature code only represents one unit modification on the feature of video content so that the difference between two features can be measured by just calculating the Hamming distance. So, in this paper, we will assume that all features are converted into Quasi-Gray binary code. Thus, the two terms "feature difference" and "feature distance" are interchangeable in this paper.

In the next subsection, two important applications of the derived relationship will be introduced: one is to estimate the maximum difference between feature of the original video frame and that of the processed video frame if the video processing is acceptable; the other is to show how the feature difference varies with the quantization step in video compression.

2.2 Two applications

2.2.1 Maximum allowable feature difference

As shown in Figure 1, if the video only undergoes normal video processing, the distortion introduced by watermark embedding and video processing must be imperceptible. In other words, this distortion should be limited. Let D represents the maximum allowable distortion between

V_0 and V_2, the maximum allowable feature difference can be calculated as follows:

$$\max(d(F_0;F_2)) = \max_{E(d(V_0;V_2))\le D}(L_f - \alpha * I(V_0;V_2)) \quad (5)$$

$$= L_f - \alpha * \min_{E(d(V_0;V_2))\le D}(I(V_0;V_2)) \quad (6)$$

$$= L_f - \alpha * R(D) \quad (7)$$

The second term of equation (7) is the definition of Rate Distortion function. Since $R(D)$ is a non-increasing convex function of D [6], the maximum feature difference could be calculated if the maximum allowable distortion D is known. For a given video frame, we consider its Just Noticeable Difference (JND) as the maximum acceptable distortion in the video authentication system. Thus, the maximum feature difference can be obtained on the JND.

2.2.2 Feature difference and video compression

In video compression, video quality degradation mainly comes from quantization. So we will look for the relationship between the feature difference and the quantization step since the video compression is considered as a normal processing in video authentication. Let C_i be the original DCT coefficient and q_i the quantization step. Then, the reconstructed DCT coefficient ($C_i^{'}$) is

$$C_i^{'} = C_i + \Delta(q_i) \quad (8)$$

where $\Delta(q_i)$ is considered as an additive uniformly distributed noise [8]; and the probability density function of $\Delta(q_i)$ is given by

$$p(\Delta(q_i)) = \begin{cases} 1/q_i & if \quad |\Delta(q_i)| < q_i/2 \\ 0 & Others \end{cases} \quad (9)$$

Using the same notation in Section 2.1, the difference between features of the original video frame and the reconstructed video frame can be calculated as

$$d(F_0;F_2) = L_f - a * I(V_0;V_2)$$
$$= \alpha * (H(V_0) - I(V_0;V_2)) \quad (10)$$

According to properties of entropy and mutual information,

$$H(V_0) - I(V_0;V_2) = H(V_0/V_2) \quad (11)$$

$$= H((V_0 - V_2)/V_2) \quad (12)$$

Following that conditioning reduces the entropy, we further get

$$H(V_0) - I(V_0;V_2) \le H(V_0 - V_2) \quad (13)$$

$$= H(\sum_i (C_i - C_i^{'})) \quad (14)$$

Using the theorem, termed as independence bound on entropy, in information theory, we have

$$H(\sum_i (C_i - C_i^{'})) \le \sum_i H(C_i - C_i^{'}) \quad (15)$$

$$= \sum_i H(\Delta(q_i)) \quad (16)$$

$$= \sum_i \log(q_i) \quad (17)$$

Equation (17) is the entropy of variable $\Delta(q_i)$ with a uniform distribution. Therefore, the upper bound for feature difference between the original video frame and the reconstructed video frame can be finally expressed as

$$d_{\max}(F_0;F_2) = \alpha * \sum_i \log(q_i) \quad (18)$$

If all the quantization steps are identical to be q, equation (18) can be further written as

$$d_{\max}(F_0;F_2) = \beta * \log(q) \quad (19)$$

This relationship clearly indicates how compression affects the feature difference between the original video frame and its compressed version.

3. EXPERIMENATAL RESULTS

In this section, we use the feature selected in [4] for evaluation. It is a 44 bits binary data. Please refer to [4] for more detail on feature selection. During evaluation, video "Akiyo" is used as the testing video.

Firstly, we compute the maximum allowable feature difference based on the JND given by Watson [9]. During computation, the DCT coefficients are classified into 64 independent channels by placing the coefficients in the same position in the DCT blocks into the same channel. These 64 channels are scanned in Zig-Zag order before the first 30 channels are selected for computation. This is in line with the fact that features selected in video authentication system usually only represent the low and middle frequency information due to the requirement of robustness. We also assume that the channels are Gaussian channels except that the DC channel is assumed to be a Laplacian channel. The upper bound of the maximum allowable feature difference is shown in Figure 2. From this figure, we can see that the maximum allowable feature differences are quite stable within the whole video sequence. Similar results have also been obtained in evaluating other testing videos. This indicates that the maximum allowable feature difference is a value almost independent of video content. Thereafter, the maximum allowable feature difference could be calculated before a robust video authentication system is designed. Note that, however, the calculated value "6" is different from the value "3" that we obtained in experiments [4].

This is due to two factors: one is that the value in Figure 2 is an upper bound; the other is that the selected feature only partially reflects the information in video frames and is not as sensitive as expected. That is, we need to understand that there is always a trade-off between robustness and sensitivity for feature selection in video authentication.

Secondly, we evaluate the feature difference between the original video frame and the reconstructed video frame. The relationship between the feature difference and quantization step is shown in dashed line in Figure 3. For comparison, we also test the theoretical relationship based on equation (19), shown in solid line in Figure 3. The experimental results are not very close to the analytical results derived in Section 2 but agree in terms of tendency.

4. CONCLUSIONS

In this paper, we have derived an analytical relationship between the feature difference and the video distortion based on Mutual Information for video authentication. To evaluate its validity, we applied it to estimate the maximum allowable feature difference, which is an important parameter in designing a robust video authentication system. In addition, we also showed how feature difference varies with the quantization step in video compression. Experimental results have confirmed the validity of our analytical results and the usefulness of the derived relationship in the design of a robust video authentication system.

The same approach can also be applied to relate the feature difference to the geometrical manipulations such as rotation and scaling. In the future, we will investigate how to select/combine different features according to the theoretical results in the design of a robust video authentication system.

11. REFERENCES

[1] Martinian, E.; Wornell, G.W.; and Chen, B., "Authentication with Distortion Criteria", Submitted to *IEEE Trans. Inform. Theory*.

[2] Ditmann, J.; Steinmetz, A.; and Steinmetz, R., "Content-based digital signature for motion pictures authentication and content-fragile watermarking", *Multimedia Computing and Systems*, 1999. IEEE International Conference on, Volume: 2, 1999, Page(s): 209 -213 vol.2.

[3] Queluz, M.P., "Towards robust, content based techniques for image authentication", *Multimedia Signal Processing*, 1998, IEEE Second Workshop on, 1998, Page(s): 297 –302.

[4] Dajun He; Qibin Sun; and Qi Tian, "A Semi-fragile Object Based Video Authentication System", *Circuits and Systems*, 2003, ISCAS '03, Proceedings of the 2003 International Symposium on, Volume: 3, May 25-28 2003 Page(s): 814 -817.

[5] Qibin Sun; Shih-Fu Change; Maeno, K; and Suto,M , " A New Semi-fragile Image Authentication Framework Combining ECC and PKI Infrastructures", IEEE International Symposium on *Circuits and Systems*, 2002, ISCAS 2002, Volume: 2, Page(s): 440-443.

[6] T. M. Cover, J.A. Thomas, *Elements of Information Theory*, New York: John Wiley & Sons, 1991.

[7] Maes, F.; Collignon, A.; Vandermeulen, D.; Marchal, G.; and Suetens, P., **"**Multi-modality image registration by maximization of mutual information", *Mathematical Methods in Biomedical Image Analysis*, 1996., Proceedings of the Workshop on , 21-22 June 1996, Page(s): 14 -22.

[8] Kundur, D., "Implications for high capacity data hiding in the presence of lossy compression"**,** *Information Technology: Coding and Computing*, 2000, Proceedings. International Conference on, 27-29 March 2000, Page(s): 16 -21

[9] Watson, A.B., "Visually optimal DCT quantization matrices for individual images", Data Compression Conference, 1993. DCC '93, 30 March-2 April 1993, Page(s): 178 -187.

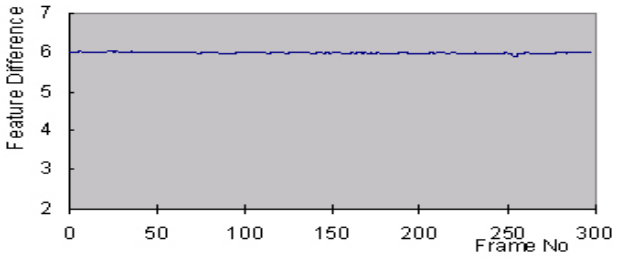

Figure 2 Maximum allowable feature differences for video "Akiyo". The horizontal axis represents frame number. From the figure, we can find that all differences are around 6.

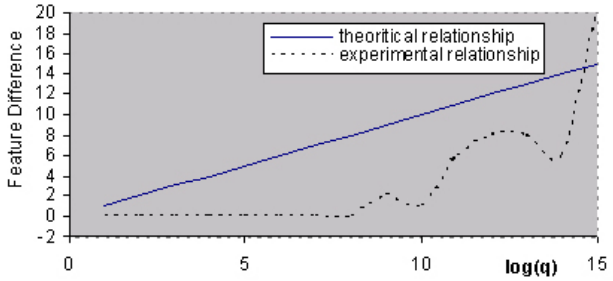

Figure 3 The relationship between feature difference and quantization step. The solid line represents the relationship based on theory; the dashed line is the experimental result. Two results agree in terms of tendency.