RESEARCH ARTICLE OPEN ACCESS

Combination of Biometric Information into Video Watermarking

Hai Truong [1], Thanh Tung Nguyen [2]
Faculty of Computer Science and Engineering [1] & [2]
Ho Chi Minh University of Technology
Vietnam

ABSTRACT
Nowadays, when the Internet is widely utilized by almost every individual, sharing information among people is becoming extremely easy. Stormy expansion in the digital techniques and Internet usage has leaded to challenging problems such as illegal re-distribution videos, un-copyright audio or music clips, and manipulating the content of digital media. Watermarking is a techniques of embedding personal data (could be private or public) into the original data to prevent illegitimate exploitation. The embed data could also be extracted again to certify the copyright, or verification. Besides, biometrics has been widely applied in various applications because of its safety and convenience. Biometric information is attached with each individual, which obstructs intruders to counterfeit the embedded data. As a result, combining biometrics into watermarking is not only the solution for the mentioned problems but also delivering the convenience when this approach is applied on mobile phones which are integrated with camera and microphone for efficiently biometrics collection.

Keywords :— Steganography, Video Watermarking, Audio Watermarking, Frequency Domain Watermark, Biometric Security.

I. WATERMARKING OVERVIEW

A. Introduction

Securing the digital data has turned into significant matters due to the quick development of Internet and the increasing number of sharing documents all over the world. Copyright protection of digital media, therefore, requires complex algorithms to overcome the illegal exploitation. The concerns of security, copyright protection and data validation could be achieved by utilizing watermarking. In tradition, a watermark is a text, logo, or image which is impressed onto paper to verify the evidence of authenticity. Digital watermark is an extension of the original one when this technique is applied on digital world. This is the process of embedding personal data, signal into the original media such as image, video, audio and text. This could be handled by manipulating the content of the media. These watermarks should remain intact under transformation and transmission, allowing a user to protect the ownership rights under the digital form. Later in the extraction step, these data could be retrieved to reveal the real identity of the media.

B. Watermarking classification

The watermarking approached could be classified based on different criteria [4,11]:

- According to human perception, the watermarking techniques could be visible [13] or invisible [1]. Visible watermarking embeds the data which could be seen by human's eyes to ensure the copyright information. On the other hand, invisible watermarking embeds secret data which could not or extremely hard to be recognized. The advantage of invisible watermarking is not to effect the media quality.
- According to detection process, the watermark algorithms could be classified into visual and blind categories. Visual watermarking requires the original data for detection whereas blind watermarking does not need the original one for testing.
- According to the domain, watermarking is divided into Spatial Domain Watermarking (SDW) and Transform (Frequency) Domain watermarking (TDW) [12]. The SDW techniques embed and detect the secret data based on spatial pixels' values such as luminance, chrominance, color space or overall video frame. However, some common digital signal processing such as dropping, scaling, rotating or other transforming could damage the watermark data and weaken the robustness of the approach. On the other hand, TDW modifies the coefficients of the video sequence's frames [14]. Some of common transforms are Discrete Cosine Transform, Discrete Fourier Transform and Discrete Wavelet Transform which are all more robust to distortions. Finally, the inverse process is applied to obtained the personal data. The benefit of this approach is that embedded data is dispensed irregularly over the image which impedes the intruder to detect and manipulate the watermark data. A typical process of TDW encoding and decoding is shown in Figure 1. and Figure 2.

C. The watermarking characteristic

Each category of watermarking approach possesses different properties which determines the effectiveness of the system. Different applications require different degrees of these characteristics and how to achieve a good balance among the properties is a challenging mission. In general, a watermark approach should gain the following attributes [8]:

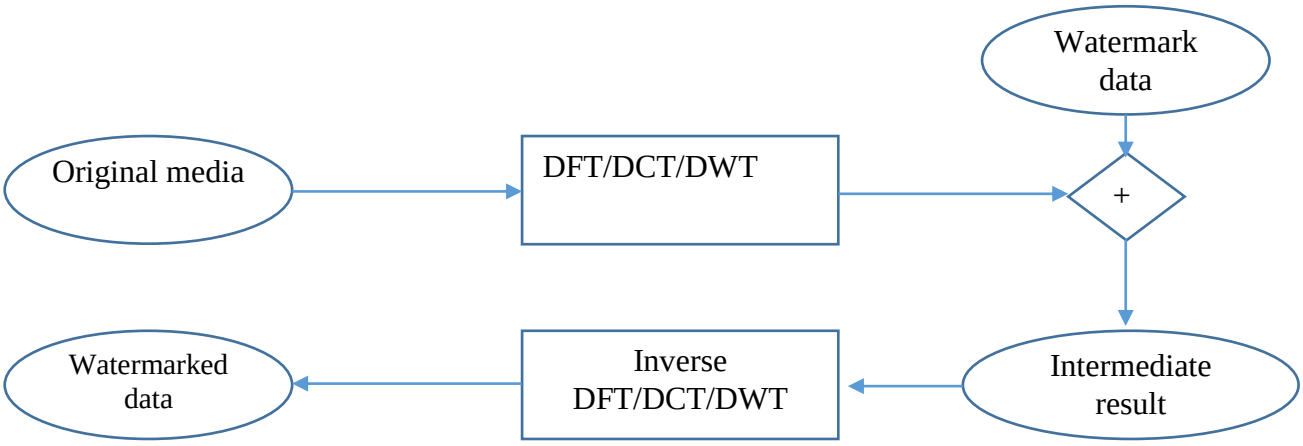

Fig. 1 Encoding in Transform Domain Watermarking

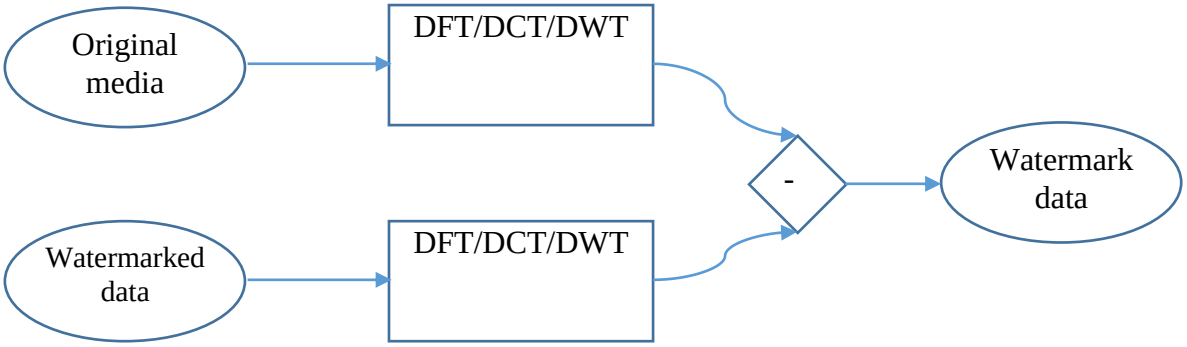

Fig. 2 Encoding in Transform Domain Watermarking

- Robustness: Robustness refers to the ability of extracting the embedded data after the variety of operations or attack. The watermarked data should remain intact and be resistant to various attack and common signal processing such as rotating, scaling and format converting.
- Security: the ability for a malicious user could extract the data without knowing the embedding algorithm, detector and at least one watermarked data. A common way for enhancing the security is to widen the embedded space, or increase the key size split into small pieces of cover image.
- Capacity: is defined as the largest quantity of bits of information that could be embedded in the original data as well as the ability to embed multiple watermarks in a document.
- Fidelity: is determined by the ability that a watermark can not be detected by ears or eyes. The fidelity property assures that only the authorized person can extract the watermark data through a special processing of watermark detector.
- Complexity: describes the effort and time needed for encoding and decoding watermark. In real applications, this parameter should be fairly fast and low computational complexity.

II. BIOMETRIC OVERVIEW

A. Motivation for combining biometric into video watermarking

In recent years, with the development of technologies, biometrics-based information is becoming potential in many fields such as authentication, security and watermarking. Biometrics (from the Greek bio = life, metric = degree) illustrates the ability of authentication by means of physiological features such as face, voice, fingerprint or behavioural features such as gait, typing speed [2]. Instead of using traditional watermark data such as text, logo, or pictures, utilizing biometrics in watermarking system possesses these advantages:

- Watermarking by traditional text or logo is usually visible which can effect on the media quality. Whereas, biometric watermarking is invisible and reduces the influences on the video and music.
- Because text and logo are visible watermark, they are easily recognized and removed. For example, an unauthorized person could simply place a bigger logo at the point of watermarked data to remove it. On the other hand, biometric watermark is invisible, therefore, intruder may not recognize the existence of the secret data for destruction.

- Text and logo could be counterfeited. Biometric features are unique to individuals and could not be forgotten and lost. As a result, faking biometric data consumes a large amount of time and effort of the intruder. An unauthorized person must contain the biometric trait, feature extraction method, and biometric embedding algorithm for destroying the embedded data.

B. Face biometric feature extraction

A feature extraction technique is employed to extract the face feature. Among the class of feature extraction techniques, PCA (Principal Component Analysis) is the most popular ones for face recognition because of its outperformance and precision [3].

In Eigenfaces method [5], the PCA is applied on a set of training images to construct a standardized face ingredient, which is called eigenface. Assume that the set contains M training images I_1, I_2, I_3,..., I_M. The average face of set is defined as:

$$\psi = \frac{1}{M}\sum_{i=1}^{M} I_i \qquad (1)$$

The distance between each I_i image and the average one is:

$$\phi_i = I_i - \psi \qquad (2)$$

Then, the covariance matrix is determined by:

$$\text{Cov} = \frac{1}{M}\sum_{i=1}^{M} \phi_i \phi_i^T = AA^T \qquad (3)$$

where the matrix $A = [\phi_1 \phi_2 \dots \phi_M]$.

M eigenvectors together with M eigenvalues could be achieved by solving the covariance matrix Cov. The eigenvalues indicate the significance of corresponding eigenvector. After that, *R* largest eigenvectors $(R \leq M)$ are selected as the face space which could be denoted as: $S = [s_1 s_2 \dots s_R]_{N^2 xR}$. s_i is the eigenface and these eigenfaces are orthogonal to each other. The image of a user can be transformed to the R-dimensional face space by linear mapping. The result vector is considered as the feature data which are extracted from the input face biometric.

$$\Omega = s^T(I - \psi) = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_R \end{bmatrix}_{Rx1} \qquad (4)$$

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III. THE PROPOSED SYSTEM

A. Overview

An overview of the proposed system is shown in Figure 3. The proposed system is designed for deploying in mobile devices which are already integrated with camera or microphone for easily biometric collection. At the first step, when the user views a video on the smart phone, their biometric is concurrently extracted via the front camera. These images are first preprocessed in order to increase the quality by noise reducing and face region detection. After that, the face image is fetched through the bio-feature extraction for retrieving feature vector. The vector is then driven to the quantization step to retrieve stability degree of face features, and are converted in to bit string for being capable with the input format of LSB (Least Significant Bit) algorithm [9] for audio watermarking and FDW (Frequency Domain Watermarking) for video watermarking. At the same time, the media file is processed to export data packets which contain audio packet and video packet separately. These packets are then extracted to structured data with list of features in order for data to be embedded. At the final step, invisible watermark approach is applied on the audio and video. Utilizing invisible watermark maintains the media quality and resists the data from transformation attacks such as dropping, scaling, or converting data format. The other components are detailed characterized in the following sections.

B. Bio-feature extraction

As mentioned in the previous part, eigenface is selected for retrieving biometric data. In our experiment, a training set of 15 images are selected and 5 face images of users are taken during the time of playing the media. As the result, the output of the Feature Extraction component is 5 features vector F_i = (F_{i1}, F_{i2}...F_{i15}) with F_{ij} are real numbers. The dimension of a feature vector is equal to the number of images in the training set. The reason to collect multiple input images (in our case is 5) is that biometric is an inexact matching, therefore, retrieving several images with different user 's poses facilitates the system to extract the stability degree of the input data and enhance the precision of feature vectors. This requirement is performed by the quantization step.

C. Quantization

After the feature extraction step, it is difficult to expect 5 output feature vectors Fi to be exactly similar. This could be caused by the variation of many parameters when capturing the image such as the environmental brightness, face position or direction. However, these vectors certain is similar with some degree of accuracy because they are all features of one person. The first purpose of Quantization step is to generate a quantization vector vector Q = $(q_1, q_2, \dots, q_n)$ which extracts the stability degree of the set of input vectors by (5).

$$q_i = \left\langle Max\ q \mid \lfloor 10^q * F_{ji} \rfloor = \lfloor 10^q * F_{li} \rfloor, \forall j, l \in [1,k], j \neq l \right\rangle \qquad (5)$$

An example with k = 2, n = 4, F_1 = (0.756, 0.24, 0.598768, 0.3) and F_2 = (0.758, 0.2, 0.598762, 0.3) then the quantization vector Q will be $Q = (q_1, q_2, q_3, q_4)$ = (2, 1,5,1).

The second purpose of the Quantization step is to transform F_i from continuous domain into discrete domain (integer

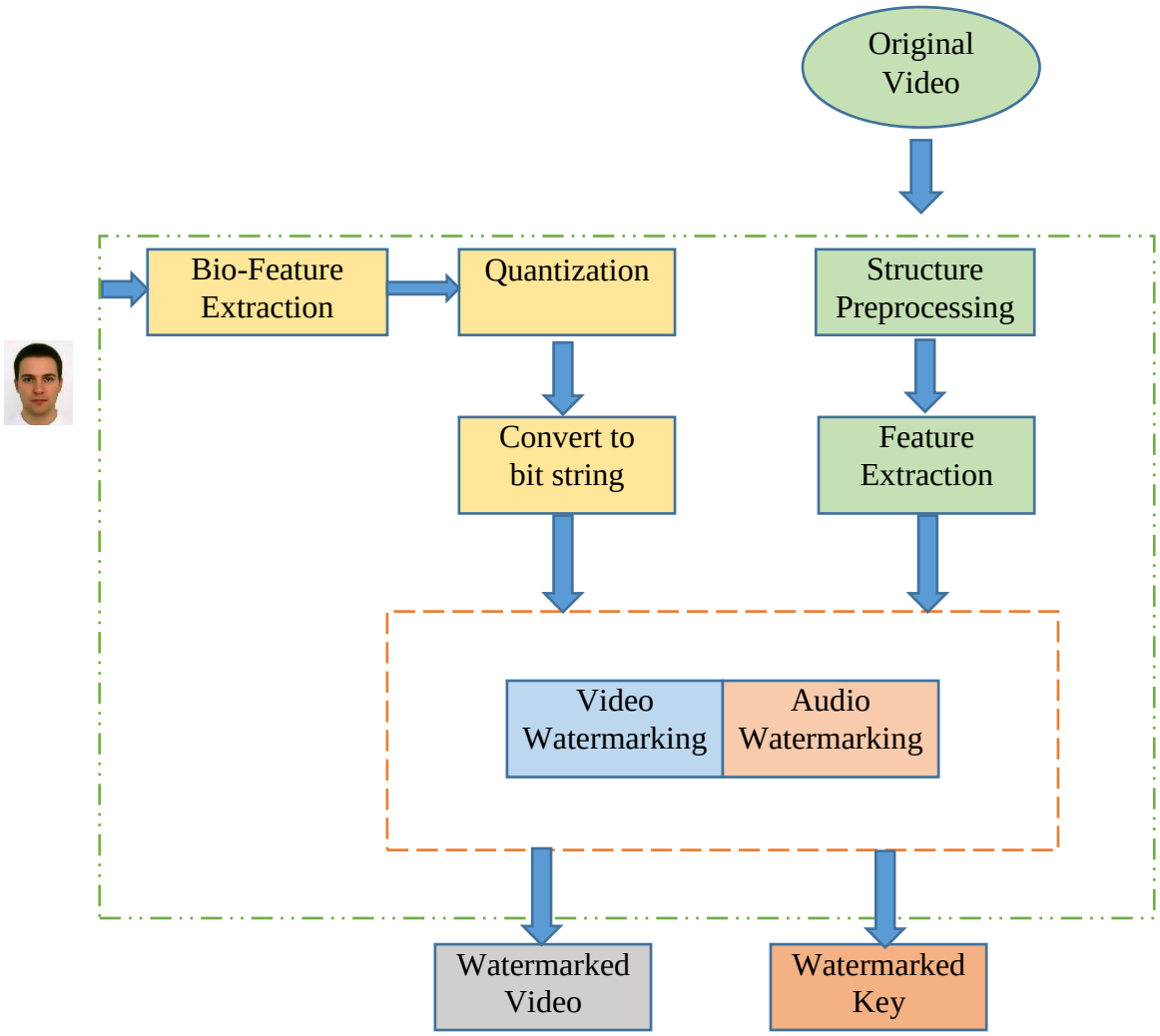

Fig. 3 Architecture of the biometric watermarking system

number) by the equation (6) which is suitable for converting into bit string. The one output feature vector $FE = (FE_1, FE_2,..., FE_{15})$ is considered as the final face biometric extraction vector of the user. The Round(x) function in (6) is a function returning the nearest integer of the input x.

$$FE_i = \text{Round}(F_{1i} * 10^{q_i}) \quad (6)$$

With the previous example, the feature vector with represents for the user will be the four-dimension integer vector FE = (76, 2, 59877, 3).

D. Convert to bit string

The output of the quantization step is a vector of integer numbers $FE = (FE_1, FE_2, ..., FE_{15})$ with $FE_i \in \mathbb{Z}$ The purpose of this component is to transform the vector FE into a bit string. The idea here is to transform each FE_i into binary number with the same length l and then concatenate the result of all FE_i. The length l is determined by the maximum value of all FE_i is calculated by (7).

$$l = \lfloor \log_2 Max\ (FE_i) \rfloor + 1 \quad (7)$$

If the length of binary number of any FE_i is less than l bits, then 0 will be added at the left hand side to achieve sufficient l bit length.

E. Structure Preprocessing and Feature Extraction

In this step, the approach of Jokay[6] is applied on the MP4 video data structures. Jokay defines a GOPs (groups of pictures) contains several types of I, B, and P frames. Each frame consists of slices of macroblocks, which are four 8x8 matrices and a MP4 video stream contains list of GOPs. This component will first transform the MP4 data into data channel, and decompose into audio and video packets which are suitable for inserting each bit of embedded information using the invisible watermark later.

F. Audio and Video Watermarking

Audio watermarking is extremely efficient because of the unpredictable and characteristic redundancy features which ensure the optimal environment for hiding embedded data. For audio watermarking, least significant bit (LSB) [9] is utilized. This is one of the simplest and earliest way to embed information into a media file especially audio. The bit of the embedded data changes the LSB of the target audio stream. One advantage of this approach is the large capacity which means large amount of data could be embedded without affecting the sound quality. On the other hand, embedded data by LSB cannot prevent from modification attack such as compression and format conversion.

For Video watermarking, the Frequency Domain Watermarking (TDW) approach is selected, and Discrete Cosine Transform (DCT) [7,10] is applied. The DCT breaks a video frame into different frequency bands for easily embedding information into the middle frequency bands. The middle frequency bands are picked because of low affection to the media quality and resistance to the compression and noise attacks. For each video frame which are embedded information, a bit of data would be mixed in a region characterized by color pixels in that frame.

All coordinates to embed bits of information will be randomly selected and saved into a watermarked key utilized for extracting copyright information later. After that, inverse DCT transformation will re-produce the video frame and export the watermarked video.

IV. EXPERIMENT AND EVALUATION

The system is deployed for biometric watermarking on Samsung Note 5 with Android 6.0 Marshmallow, CPU 8 x 1.5 GHz, Ram 4GB in order to apply for watermarking eLearning videos. The experiment shows that with an average 80-minute length video and frame size is 320x240 pixel, time for processing the invisible watermark takes about 120-150 minutes. Some major concerns are also retrieved by the experiment are a) the video length is proportional to the number of frames and also proportional to the time consuming of watermarking, b) The larger the size of video frame, the longer it takes to transform video into pictures for processing, therefore, increasing the time consuming and finally, c) the proportion of video frames needed to be embedded determine the time for watermarking.

ACKNOWLEDGMENT

This research is funded by Vietnam National University - Ho Chi Minh City (VNU-HCM) under grant number T-KHMT-2015-28.

REFERENCES

[1] A. Santa, et al, "An image adaptive, wavelet-based watermarking of digital images," *Journal of Computational and Applied Mathematics* 210.1 (2007): 13-21.

[2] D. Maltoni, et al, *Handbook of fingerprint recognition*, Springer Science & Business Media, 2009.

[3] K. Baek, et al, "PCA vs. ICA: A Comparison on the FERET Data Set," *JCIS*. 2002.

[4] L. Baisa and R. Manthalkar, "An overview of transform domain robust digital image watermarking algorithms," *Journal of Emerging Trends in Computing and Information Sciences* 2.1 (2010): 37-42

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience* 3.1 (1991): 71-86.

[6] M. Jókay, "The design of a steganographic system based on the internal MP4 file structures," *development* 17 (2012): 18.

[7] N. Meghanathan and L. Nayak, "Steganalysis algorithms for detecting the hidden information in image, audio and video cover media," *international journal of Network Security & Its application (IJNSA)* 2.1 (2010): 43-55.

[8] P. Rini, "Review of robust video watermarking techniques." *IJCA Special Issue on Computational Science* 3 (2011): 90-95.

[9] S.K. Bandyopadhyay and B. G. Banik, "Multi-Level Steganographic Algorithm for Audio Steganography using LSB Modification and Parity Encoding Technique," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* 1.2 (2012).

[10] S. Sinha, et al, "Digital video watermarking using discrete wavelet transform and principal component analysis," *International Journal of Wisdom Based Computing* 1.2 (2011): 7-12.

[11] T. Jayamalar and V. Radha, "Survey on digital video watermarking techniques and attacks on watermarks," *International Journal of Engineering Science and Technology* 2.12 (2010): 6963-6967.

[12] Thanki, M. Rohit, R. Kher and D. Vyas, "Robustness of Correlation Based Watermarking Techniques Using WGN against Different Order Statistics Filters," *International Journal of Computer Science and Telecommunications* 2.4 (2011): 45-49.

[13] Xia, X. Gen, C. Boncelet and G. Arce, "Wavelet transform based watermark for digital images," *Optics Express* 3.12 (1998): 497-511.

[14] Y. Yusof and O. Khalifa, "Digital watermarking for digital images using wavelet transform," Telecommunications and Malaysia International Conference on Communications, 2007. ICT-MICC 2007. IEEE International Conference on. IEEE, 2007