VOICE BIOMETRIC WATERMARKING OF 3D MODELS

Rakhi C. Motwani, Sergiu M. Dascalu and Frederick C. Harris Jr.
Department of Computer Science and Engineering
University of Nevada, Reno
Reno NV USA 89507
Email: {rakhi,dascalus,fredh@cse.unr.edu}

Abstract

Digital watermarking provides a solution to piracy and copyright protection of 3D multimedia content by embedding a hidden piece of information in the original content. This hidden piece of information travels with the 3D multimedia in the distribution chain and assists in verifying legitimate ownership and tracing piracy. This paper proposes a voice-based biometric watermark for 3D graphics for the purpose of owner identification, traitor tracing and access control schemes. A voice print is generated using a Gaussian mixture model representation of Mel-frequency cepstral coding coefficients of an individual's speech. This voice print is inserted as the watermark. The proposed technique generates a semi-fragile watermark, which is tolerant to a designated class of transformations. Experimental results indicate that the biometric watermark is resistant to cropping, low levels of Gaussian noise addition, and is intolerant to mesh smoothing attacks.

1. Introduction

Strong media coverage for virtual universes like *Second Life* and 3D online games such as *World of Warcraft*, has motivated users to discover new online spaces for playing, communicating, and entertainment. Since users want to be ensured that they retain copyright for any graphics they create and post online, the need for digital rights management for protecting artwork owners against digital piracy and copyright infringement has been of increasing concern. Online marketplaces of 3D models are subject to similar concerns. Research in digital watermarking has significantly progressed during the past two decades to address copyright issues. Typically, bit pattern representations of a logo image or text which identifies copyright information, random number sequence is inserted as watermark, or content-based information is used to derive the watermark. However, researchers have lately begun to explore the effectiveness of employing biometrics in digital watermarking schemes. The next section reviews some such techniques.

This paper explains a technique for inserting the voice biometric of the artist(or consumer) as the watermark for the purpose of owner identification, access control and traitor tracing. Since biometric traits represent the identity of an individual, the potential for conflicts is lessened while establishing ownership identification or identifying the person responsible for piracy. Biometric watermarks when used within a digital rights management framework can also assist in controlling access of digital content to the legitimate user(who is verified by the biometric watermark [1]). The different host mediums used thus far for biometric watermarking by published literature are digital documents, images, audio and video. To date biometric watermarks have not been investigated for 3D models. The novelty of this paper lies in embedding voice biometrics into 3D models for copyright protection.

2. Related Work

Since biometric watermarks have not been investigated for 3D graphics yet, this section reviews literature to explore the purpose of utilizing biometric watermarks for other digital content types. The authors in [2] utilize biometric watermarks for digital images and biometric images such as face or fingerprint image, to establish the authenticity of the biometric trait and the user. In [3] the biometric watermark assists in verifying the integrity of the host image by detecting tampering of the host data. In [4] document images are secured by an image watermark generated from the author's digitized handwritten signature. An iris biometric watermark is adopted by [5] for document images in order to protect the document and assist in owner identification.

In [6] the authors discuss a copyright protection scheme for images using fingerprint images. Researchers in [7] and [8] employ fingerprint and iris biometrics as watermarks for audio and video digital media to formulate reliable identification of the user of the digital media. As biometrics possess exclusive characteristics that can be hardly counterfeited, conflicts related to the intellectual property rights protection can be potentially discouraged.

The only published work related to employing voice biometrics as a watermark is [9]. The technique embeds mel-frequency cepstrum coding (MFCC) coefficients of an individual's voice into the face image of the same individual. However, the watermark formulation assumes that the extracted MFCC features are invariant for an individual. A

978-1-4244-6349-7/10/$26.00 ©2010 IEEE

statistical modeling component is required to incorporate the variability in the voice samples originating from the same individual owing to envrionmental, improper sensor interaction, sickness(cough, cold) and emotional stress factors. The main contribution of this paper is voice print formulation by borrowing techniques from acoustic biometric systems ([10],[11]) and embedding this voice print as a watermark in 3D models.

3. Approach

A voice signal conveys an individual's physiological characteristics [12] such as the vocal chords, glottis, and vocal tract dimensions. MFCC [13] and GMM [10] are by far the most prevalent techniques used to represent a voice signal for feature extraction and feature representation in state-of-the art speaker recognition systems. Therefore, the proposed approach adopts the MFCC features and employs GMM to generate the speaker model.

The watermark embedding process is responsible for generating a voice print and inserting it as a watermark into the 3D model, as shown in Fig. 1. The watermark retrieval

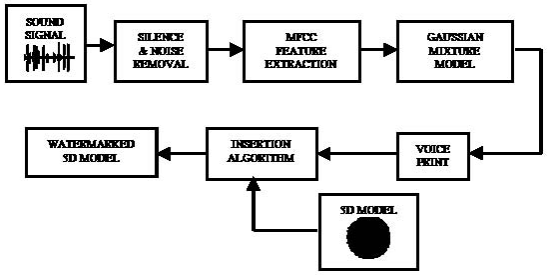

Figure 1. Block Diagram for Watermark Embedding

process extracts the embedded voice print and correlates it with the original voice print, as depicted in Fig. 2.

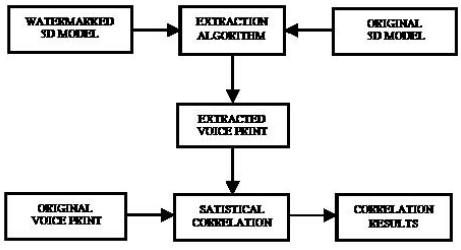

Figure 2. Block Diagram for Watermark Retrieval

3.1. Watermark Embedding

Step 1. Voice Acquisition of Sound Signal - The first block in Fig. 1 acquires the voice of an individual using a PC microphone at a sampling rate of 44 kHz. Recording length ranges from 10 to 15 seconds. Fig. 3 illustrates the digital representation of an acquired voice sample.

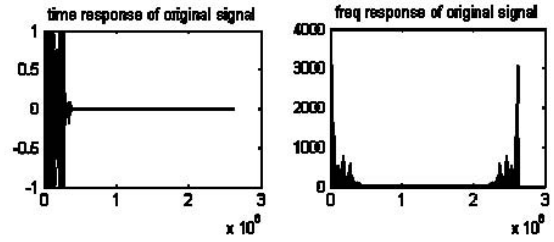

Figure 3. Digital Voice Signal

Step 2. Silence & Noise Removal - Recognizing silence in the speech and removing the silence reduces processing load. Zero-crossing rate (ZCR) parameter is used to eliminate silent portions of the signal [14]. Speech contains noise which is always external signals that may interfere with the sound. Since noise falls into the high frequency range of the signal spectrum, a low-pass filter is used to reduce the noise. Fig. 4 illustrates the preprocessed signal.

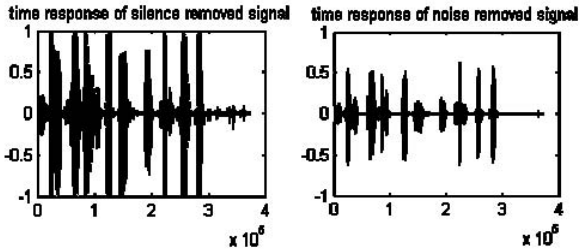

Figure 4. Silence and Noise Removed Voice Signal

Step 3. MFCC Feature Extraction - This module converts the speech waveform to a parametric representation. The first step is windowing which allows for short-term spectral analysis that decomposes the signal into frames (length 2048 samples) that are extracted by a Hamming window as shown in Fig. 5.

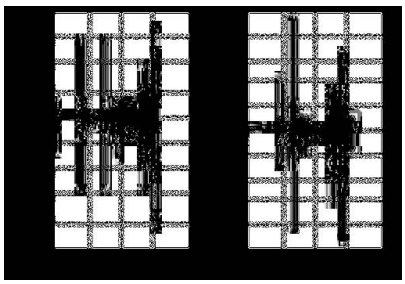

Figure 5. Frames Before & After Windowing (Frame1:Red, Frame2:Blue, Frame3:Green)

Then the DFT stage transfers each frame of the signal to the frequency domain (Fig. 6).

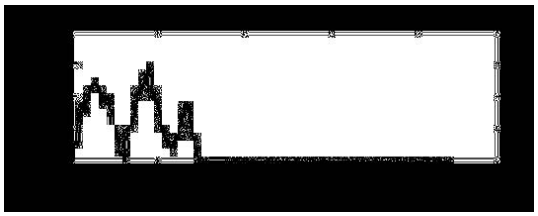

Figure 6. DFT Spectrum Magnitude and Phase

A Mel-warped filter bank is applied to the DFT spectrum to simulate the critical band filters of the hearing mechanism. The filters are evenly spaced on the *mel* scale, and are triangular shaped (Fig. 7).

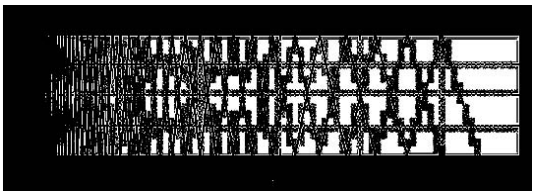

Figure 7. Mel Filter Bank

The energies of the resultant spectrum are then transfered to the log scale, as depicted by Fig. 8.

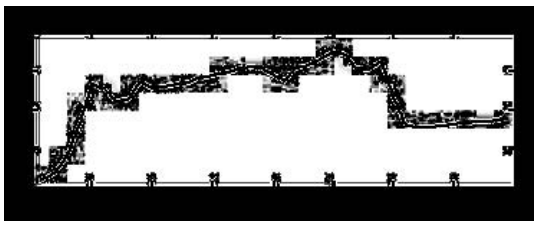

Figure 8. Log Energies For Each Filter

An inverse DFT transfers the log spectrum to cepstrum (Fig. 9) which represents MFCC coefficients.

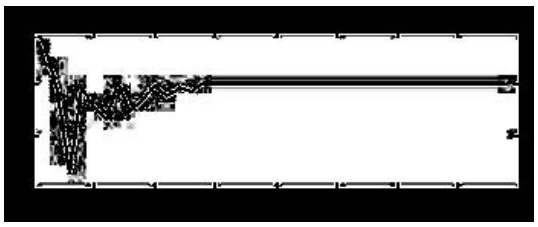

Figure 9. Cepstrum

Each frame of the speech data is then represented by a 12-dimensional [15] feature vector consisting of MFCC coefficients i.e., $x = x_1, x_2, x_3, \ldots, x_{12}$.

Step 4. Feature Representation using Gaussian Mixture Model(GMM) - GMM is a weighted mixture of a series of Gaussian distributions over the space of the MFCC coefficient data, as demonstrated in Fig. 10. Each multivariate

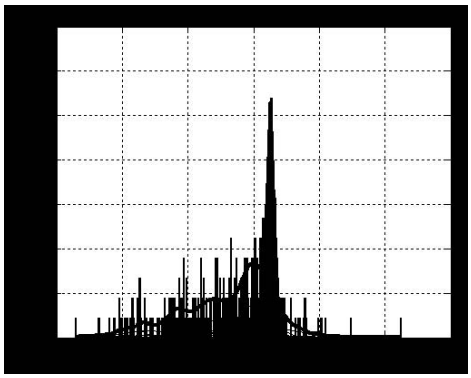

Figure 10. Gaussian Mixture Model for one MFCC

Gaussian in the mixture model is parameterized by a mean vector $\vec{\mu}$ and covariance matrix Σ for the N-dimensional feature vector x_n, where N denotes the number of MFCC coefficients. A model for M Gaussian mixtures (M denotes the order of the GMM), with each mixture having weight w_k is given by the equation:

$$p(x_n) = \sum_{k=1}^{M} w_k p(x_n|k) \tag{1}$$

where,

$$p(x_n|k) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_n-\vec{\mu})^T \Sigma^{-1}(x_n-\vec{\mu})} \tag{2}$$

represents the Gaussian distribution.
Expectation Maximization [16] optimization technique is used to iteratively estimate parameters for each Gaussian in the mixture model. The generated voice print is the aggregate of the parameter values $\lambda = \{w_k, \mu_k, \Sigma_k\}$ of the Gaussian mixture model representation, where $k = 1, \ldots, M$ denotes the order of the GMM. This voice print is used as the watermark.

Step 5. Insertion Algorithm - The steps outlining this algorithm are: *i)Normalize and Shifting* - The 3D model to be watermarked is initially normalized and shifted. The center of gravity of the 3D model to be watermarked is determined and shifted to origin (if it is already not at the origin) of rectangular co-ordinate system. Also the co-ordinates of vertices in model are normalized to lie in between -1 and +1. These steps ensure robustness of system to translation and scaling attacks on a model. *ii)Selection of Vertices* - The insertion algorithm selects vertices based on curvature and bumpiness properties of surfaces. The curvature of surface corresponding to each

vertex is computed from the spatial domain representation of the model [17]. To compute the bumpiness of surfaces [18], the 3D model is transformed into wavelet domain using Cohen-Daubechies-Feauveau CDF (2, 2) wavelet transform. *iii)Inserting Watermark* - A sequence of 40 bits is derived from each value of the voice print, since this is the maximum amount of information that can be inserted in the vertex without causing visible distortion. The vertex is represented in IEEE double precision floating point form as a string of binary bits. It is composed of 64 bits, divided into a 52 bit mantissa M, 11 bit exponent E, and sign bit S, as shown in Fig. 11. The truncated voice print

Figure 11. Modifying 40 Least Significant Bits of Mantissa

is inserted into the 40 least significant bits (LSB) of the vertex mantissa by performing Boolean *OR* operation. This ensures that the watermark is additive. An inverse wavelet transform is applied to the 3D model to get the watermarked model. Furthermore, this model can be re-shifted to the initial location in space and also the co-ordinates can be rescaled. This watermarked model is ready for distribution.

3.2. Watermark Retrieval

The original 3D model is required during watermark extraction (Fig. 2) to identify which vertices were modified. The values of vertices which have been modified are selected, and the 40 LSB of the mantissa of original vertex values are subtracted from the 40 LSB of the mantissa of selected vertices to retrieve the embedded watermark. Similarity measure of the extracted and original voice prints is computed by the Mahalanobis distance metric which takes into account the covariances Σ of each of the dimension of the voice print vector. Eq. 3 gives the extent of similarity between the embedded watermark x and the recovered watermark y, both of size n.

$$d_{mahalanobis}(x,y) = (x-y)^T \Sigma^{-1}(x-y) \qquad (3)$$

4. Experimental Results

Common attacks on watermarked 3D models are mesh smoothing, additive noise and cropping, as shown in Fig. 12. *MATLAB* is used to noisify 3D vertex coordinates using Gaussian noise(mean 0, variance 0.5). A noise level of 100% equates to adding noise to all vertices of the 3D model, while 10% level means noise is added to only $\frac{1}{10th}$ of the total count of vertices. *MeshLab* tool is used to apply Laplacian mesh smoothing which moves a mesh point to the centroid of surrounding mesh points which are topologically connected. Cropping attacks are simulated in either x,y or z dimension using *MATLAB*.

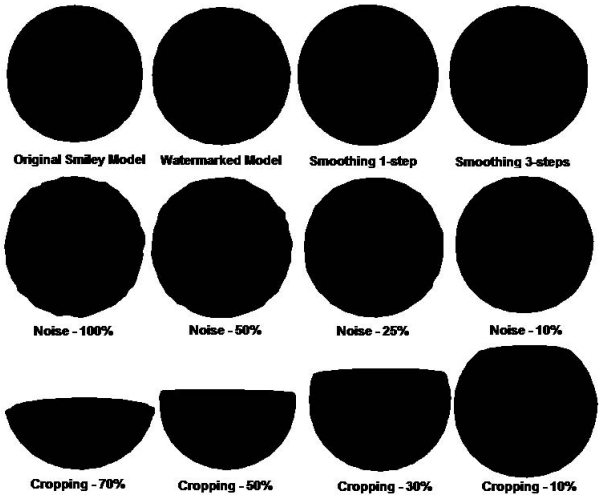

Figure 12. Attacks on *Smiley* 3D Model

Table 1 reflects results of various experiments. Hausdorff distance is a measure of computing the perceptible differences between the original and watermarked models. The correlation results ($d_{mahalanobis}(x,y)$ * 100%) below a threshold of 80% indicate a destroyed watermark.

Experiments	*Smiley*	*Nefertiti*	*Dinopet*
# of Vertices in Model	1026	654	4500
# of Watermarked Vertices	94	169	1910
Order of GMM (M)	3	7	16
Hausdorff distance	0.035447	0.041107	0.035880
No Attacks	100%	100%	100%
Noise (100%)	49.66%	22.77%	23.29%
Noise (10%)	88.92%	85.08%	82.84%
Smoothing (3 steps)	44.11%	46.38%	44.91%
Smoothing (1 step)	51.19%	53.50%	54.83%
Cropping (50%)	99.38%	97.41%	98.93%
Cropping (10%)	100%	97.41%	99.70%

Table 1. Correlation Results For The Embedded and Original Voice Biometric Watermark

The False Reject Rate (FRR) is obtained by using 4 voice samples of the same individual from the *XM2VT* dataset (remaining 4 voice samples are used for training the GMM speaker model). The False Accept Rate (FAR) is obtained by comparing voice prints of 100 test subjects from the *XM2VT* database against the extracted voice print. The FRR and FAR plots are shown in Fig. 13 and 14. The FRR for certain subjects is above 30% and can be improved by increasing the number of training samples used to generate the GMM speaker model for the test subject.

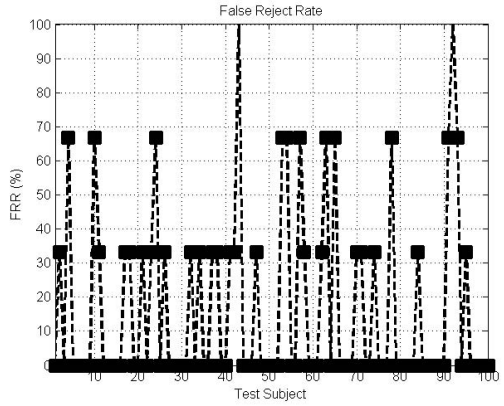

Figure 13. False Reject Rate For Voice Biometric

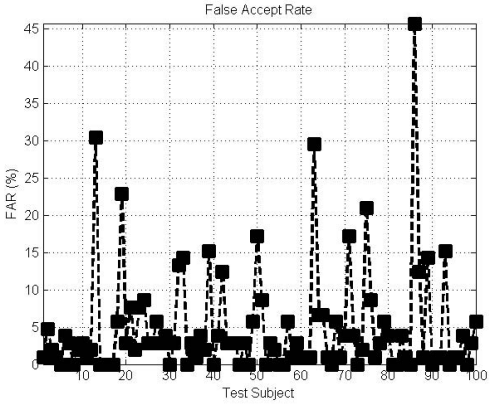

Figure 14. False Accept Rate For Voice Biometric

5. Conclusion

This paper proposes a novel watermarking scheme for copyright protection of 3D models based on voice biometrics. The voice print formulation technique is borrowed from state-of-the-art speaker verification systems. The impact of watermarking attacks on the biometric watermark is evaluated and for attacks that result in a correlation value above a threshold of 80%, the extracted GMM speaker model identifies the correct individual with an average False Accept Rate of 5% and False Reject Rate of 30%.

References

[1] R. Motwani, F. Harris, and K. Bekris, "A proposed digital rights management system for 3D graphics using biometric watermarks," in *Proceesdings of IEEE CCNC Digital Rights Management Workshop*, January 2010.

[2] C. Vielhauer and R. Steinmetz, "Approaches to biometric watermarks for owner authentication," in *Security and Watermarking of Multimedia Contents III*, vol. 4314, 2001, pp. 209–219.

[3] T. Satonaka, "Biometric watermarking based on face recognition," in *Security and Watermarking of Multimedia Contents IV*, vol. 4675, no. 1, 2002, pp. 641–651.

[4] A. Namboodiri and A. Jain, "Multimedia document authentication using on-line signatures as watermarks," in *Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, no. 1, 2004, pp. 653–662.

[5] G. Feng and Q. Lin, "Iris feature based watermarking algorithm for personal identification," in *MIPPR Remote Sensing and GIS Data Processing and Applications*, vol. 6790, no. 1, 2007, p. 679045.

[6] S.-L. Hsieh, H.-C. Huang, and I.-J. Tsai, "A copyright protection scheme for gray-level images using human fingerprint," in *Information Technology: New Generations,*, April 2006, pp. 482–489.

[7] K. Khan, L. Xie, and J. Zhang, "Robust hiding of fingerprint-biometric data into audio signals," in *ICB07*, 2007, pp. 702–712.

[8] S. Jung, D. Lee, S. Lee, and J. Paik, "Biometric data-based robust watermarking scheme of video streams," in *6th International Conference on Information, Communications and Signal Processing*, Dec. 2007, pp. 1–5.

[9] M. Vatsa, R.Singh, and A. Noore, "Feature based rdwt watermarking for multimodal biometric system," *Image Vision Comput.*, vol. 27, no. 3, pp. 293–304, 2009.

[10] D. Reynolds and R. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing,*, vol. 3, no. 1, pp. 72–83, Jan 1995.

[11] M. Chenafa, D. Istrate, V. Vrabie, and M. Herbin, "Biometric system based on voice recognition using multiclassifiers," *First European Workshop on Biometrics and Identity Management,*, pp. 206–215, 2008.

[12] J. Harrington and S. Cassidy, *Techniques in Speech Acoustics*. Springer, 1999.

[13] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing,*, vol. 28, no. 4, pp. 357–366, Aug 1980.

[14] F. Gouyon, F. Pachet, and O. Delerue, "On the use of zero-crossing rate for an application of classification of percussive sounds," in *Proceedings of the COST G-6 Conference on Digital Audio Effects*, 2000.

[15] D. Jurafsky and J. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall Series in Artificial Intelligence, 2008.

[16] A. Dempster, N. Laird, , and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, p. 138, 1977.

[17] R. Motwani and F. Harris, "Robust 3D watermarking using vertex smoothness measure," in *Proceedings of the 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition*, July 2009.

[18] M. Motwani and F. Harris, "Adaptive fuzzy watermarking for 3D models," in *Proceedings of IEEE International Conference on Computational Intelligence and Multimedia Applications 2007*, vol. 13-15, no. 4, Dec 2007, pp. 49 – 53.