VoiceMarc3D: Software Specifications and Implementation Design

R. C. Motwani, M. C. Motwani, S. M. Dascalu, and F. C. Harris, Jr.
Computer Science and Engineering Department
University of Nevada, Reno
Reno, NV 89557

Abstract

With the advent of 3D virtual worlds and a plethora of online 3D model warehouses, copyright protection of 3D graphics has been of prime concern lately. Copyright issues are crucial in media arts since digital artwork can easily be duplicated, illegally distributed, and modified. Watermarking techniques are used for copyright protection and to prevent unauthorized access of digital content. This paper details the specifications and implementation design of *VoiceMarc3D*, a prototype watermarking system which is developed for 3D multimedia. *VoiceMarc3D* generates a biometric watermark from the voice of the consumer and embeds it into the graphics content. Comparison of the embedded watermark with a newly acquired voice sample from the consumer verifies the authenticity of the consumer for legitimate access and usage of the 3D artwork. The system borrows techniques from state-of-the-art speaker verification algorithms to provide a decision on granting or denying access to the consumer for a copyright protected 3D model. The scope of this implementation is limited to watermarking of 3D mesh models.

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. With an increased number of teenagers going online to socialize and connect with the world, sites like *Facebook* and *Entropia Universe* are becoming increasingly popular. Since users want to be ensured that they retain copyright for any content they create and post online, the need for protecting content owners against digital piracy and copyright infringement has been of increasing concern. Online archives and marketplaces for 3D models are subject to similar concerns. Artists spend a great deal of time on creating the digital artwork and many have had their work stolen and claimed by others, with little way of proving the original owner of the digital 3D artwork. Digitizing multimedia makes it difficult to manage, protect, and track, thereby making it vulnerable to plagiarism, malicious modifications, and digital piracy. Multimedia owners address this problem by using digital watermarking techniques. Watermarking allows for tracing the source of illegal copies of digital artwork, detecting modifications to the original work, as well as managing digital copyrights by assisting with access control decisions.

Watermarking techniques embed identifying data into the digital content in such a way that the embedded information is perceptually invisible to the human eye, but is detectable by software. Traditional watermarking techniques embed text, random numbers, binary digits, cryptographic keys, copyright ownership messages, logos, images, or digital media content-based information into the 3D multimedia [1, 2, 3, 4]. This paper presents a voice biometric based 3D watermarking system termed as *VoiceMarc3D*. *VoiceMarc3D* implements a novel watermarking scheme [5] that uses the consumers voice to create a biometric watermark for insertion into the 3D model. The watermark is generated by borrowing techniques from the state-of-the art speaker verification algorithms that use a Gaussian Mixture Model (GMM) representation of the consumer's voice feature vectors to represent the Mel Frequency Cepstral Coding (MFCC) coefficients of the voice signal [6, 7]. This voice-based watermark is used to provide an access control decision to grant a legitmate consumer access to the 3D graphics and deny illegimate consumers from accessing the digital 3D graphics content.

This prototypical system is a novel contribution to the 3D watermarking domain. Thus far, no research work has been published that employs voice biometric based watermarks in 3D domain. Furthermore, no 3D multimedia watermarking schemes to date employ biometric watermarks. In addition, no commercial products [8, 9] exist in the 3D multimedia industry that use biometric based watermarks. The presented system takes the existing watermarking schemes to the next level of security by inserting biometric data into the 3D

content. Every individual has various unique physical characteristics such as fingerprints, palm prints, face, iris, retina, voice, and handwritten signature. Biometric watermarking exploits these physical characteristics to establish a biometric template that links to the identity of an individual and assists in verifying the legitimacy of the individual. Biometric data is more advantageous to use as a watermark because it is difficult to duplicate and individuals are hesitant in sharing their biometrics on peer-to-peer networks (fearing misuse of personal biometric data by strangers) to allow illegitimate users to access the copyright protected graphics.

The following sections of this paper outline the system specifications and design of the implementation, use case modeling, system GUI screen shots, directions for future work, and conclusions.

2 Requirements Specification

Following standard software engineering guidelines [10], the main functional and nonfunctional requirements of *VoiceMarc3D* are presented below.

Functional Requirements: Each functional requirement has been assigned a reference number, consisting of characters - DT, PR, for traceability to other system artifacts.

Data Requirements

1. Graphics File Format (F_DT_01): The system must support the Open File Format (*.off*) for 3D mesh models.
2. Characteristics and quality of the recorded voice (F_DT_02): A 2 channels stereo sound signal with a bit rate of 1411kbps, sample size of 16 bits and an audio sampling rate of 44 KHz is captured in *.wav* format.

Process Requirements

1. Imperceptible watermark (F_PR_01): The system embeds a voice print of the user into a 3D model such that watermarked 3D mesh model looks like the original model and it should not reveal any clues of the presence of the watermark.
2. Semi-Blind watermarking technique (F_PR_02): The system does not require the original unmarked model to extract the watermark from the watermarked media but makes use of a key, that stores the locations and original values of the vertices that are modified by the watermarking scheme, in addition to the original watermark.
3. Semi-Fragile watermark (F_PR_03): A semi-fragile watermark is inserted into the host model and can withstand certain attacks (noise, cropping, smoothing) but not all. The watermarking scheme does not support 3D content editability, mesh subdivision, decimation and remeshing operations.
4. Text and speaker dependent voice print (F_PR_04): The voice capture process is text dependent as all the speakers have to speak a predetermined text. The speech waveform is then converted to a parametric representation i.e. feature vectors. Extraction of features is done such that they are primarily a function of speaker.
5. Variable speaking rate (F_PR_05): The voice print extracted from the recording has to be independent of the length of duration of recorded voice sample (talking speed of the speaker). The system achieves insensitivity to speaking rate by preprocessing the acquired voice sample to eliminate silence thereby accommodating slow, moderate, and fast speakers which generate voice samples of different duration.
6. Constant size voice print (F_PR_06): Since the voice print is derived from the Gaussian mixture model of the mel-frequency cepstral co-efficients of the voice signal, the voice print always assumes a fixed size of values irrespective of who the speaker is.
7. Access control decision (F_PR_07): The system will determine if a user for a 3D model is authentic or not. The system provides a decision on whether to accept or reject the user based on the comparison of the voice sample provided by the user during authentication with the voice print generated during user enrollment.

Non-Functional Requirements
Hardware and Software Requirements: The hardware and software requirements of the system are outlined in Figure 1.

3 System Design

Figure 2 demonstrates a high-level layered architecture of the system. There are six functional layers - Voice Acquisition, Watermark Generation, Watermark Embedding, Voice Parametrization, Watermark Detection, and Access Control. The system is divided into two sub-systems - Enrollment and Authentication subsystems.

Hardware	Minimum Specifications	Recommended Specifications
Processor	Intel Pentium II processor	Intel Pentium III 500 MHz processor or higher
RAM	64 MB of RAM installed	128 MB or more of RAM
HDD space	20 MB of available hard-disk space	Large-capacity hard disk
Adapter Card	16 MB display adapter card with 3D accelerator	32 MB display adapter card
Microphone	External USB or Miniplug input	Built-in PC/Laptop Mic

Software	Company	Version	Description
MATLAB	Mathworks	7.5	Mathworks is the leading global provider of software for technical computing and Model-Based Design. MATLAB is high-performance language for technical computing that integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.
Windows	Microsoft	95 or above	Microsoft Windows 95/98/Me/NT/2000/XP operating system version s are supported.

Figure 1: Hardware and Software Requirements

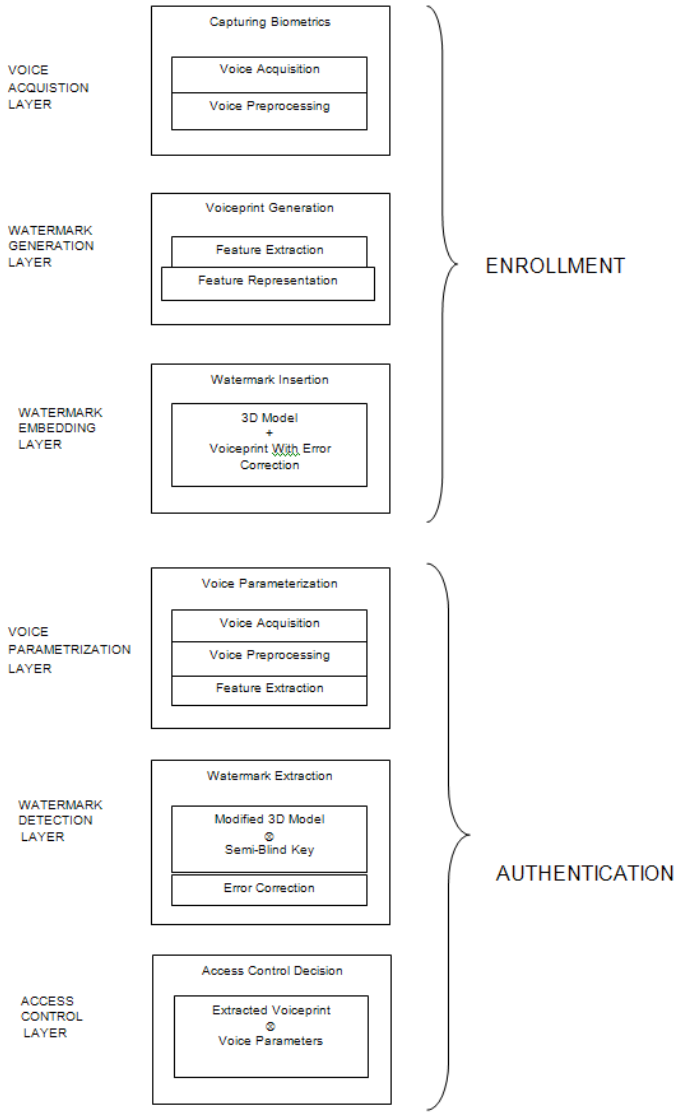

Figure 2: Architectural Diagram

3.1 Enrollment Sub-System

This sub-system is responsible for marking the 3D model with the identity of the consumer.

- Voice Acquisition Layer - This module is responsible for acquiring the voice samples of an individual and preprocessing the waveform to eliminate silence areas.
- Watermark Generation Layer - This module is responsible of generating a voice print from the speech waveform by extracting feature vectors and representing those features using a GMM.
- Watermark Embedding Layer - This module generates a watermarked model by embedding the voice print as a watermark into the 3D graphic model after employing error correcting codes to safeguard the voice print from modifications made to the watermarked 3D model such as noise, cropping, and smoothing.

3.2 Authentication Sub-System

This sub-system is responsible for determining the legitimacy of a consumer to access the watermarked 3D model.

- Voice Parametrization Layer - This module extracts the MFCC features from the newly acquired pre-processed voice sample for comparison against the voice print embedded into the watermarked 3D mesh model.
- Watermark Detection Layer - This module extracts the watermark and attempts to correct any modifications that the voice print may have been subject to by an attack.
- Access Control Layer - This module is responsible for authentication of the user by comparing the user's extracted voice features against the extracted voice print. This layer either grants or denies access to the graphics based on the obtained likelihood measure between voice features and the GMM-based voice print.

3.3 System Assumptions

The system operation is based on the following assumptions: i) The biometric trait is present in the consumer (which means that the consumer has a voice); ii) The consumer is willing to offer his/her voice biometric samples to the system for legitimate access control of watermarked 3D models; iii) The consumer is cooperative and utters the predetermined phrase while providing voice samples during both enrollment and authentication phases. The system does not check the valid-

ity of the spoken phrase should the consumer intentionally change the spoken text; iv) The voice samples are acquired in a quite environment; v) The acquired voice samples represent a single speaker audio stream. The system does not support multi-speaker streams. vi) The consumer will not share voice recordings of the spoken phrase with others. The security of the system is based on this assumption so that prerecorded voice samples of legitimate consumer do not circumvent the system. vii) The system ignores variations such as different pitch that alters the speaking manner, noisy voice samples due to the environment or communication channel (voice acquired over telephone), speaking under stress or during sickness (cough, cold, fatigue), and attempted mimicry; viii) The file formats that will be used are *.wav* for sound file input, *.off* format for 3D mesh models, ascii *.txt* representation for the biometric voice print; ix) The input *.off* file size is restricted to the range of 15KB - 2000KB. Most 3D files smaller than 15KB are unable to accommodate a watermark that houses a voice print (minimum size 1.17KB). 3D files over 2000KB require longer processing times (over a couple of minutes) on a desktop PC.

Figure 3: Use Case Diagram

4 Use Cases

The system has 10 use cases (see Figure 3). Each use case is assigned a reference number, starting with the characters UC, for requirements traceability matrix. The system GUI is shown in Figures 6 and 7.

1. AcquireVoiceSamples (UC01): The user provides the system with voice samples in *.wav* format for a predefined utterance. The user presses the *Provide 3 Voice Samples* button to upload the *.wav* files into the system.
2. SignalPreprocessing (UC02): The system eliminates silent regions from the waveform when the user presses the *Voice Signal Pre-Processing* button. This step enables the voice print formulation process to be independent of the pace at which the predetermined phrase was spoken during the voice acquisition phase. The display area is updated with the representation of processed signal.
3. FeatureExtraction (UC03): The user presses the *MFCC Feature Extraction* button to extract features representing the identity of the user.
4. SpeakerModelling (UC04): The user presses the *GMM Speaker Model* button to instruct the system to use the extracted features in the previous step for generating a GMM that constitutes the user's voice print.
5. GenerateWatermark (UC05): The user presses *Generate Watermark* button to encode the voice print using error correction routines in order to protect its data from attacks.
6. Input3DModel (UC06): The user presses the *Input 3D Mesh Model* button to upload the 3D model for watermarking. The display area is updated with a rendering of the graphic file.
7. WatermarkInsertion (UC07): The user presses the *Watermark Insertion* button to enable the system to inserts the watermark into the previously selected 3D model.
8. Display3DModel (UC08): The user pushes the *DRM Protected 3D Model* button to display the watermarked model in the bottom right axis on the interface.
9. ProvideVoiceSample (UC09): The user presses the *Provide Voice Sample* button to upload a voice sample with the predetermined utterance. The system pre-processes this voice sample and extracts user specific features from it.
10. AccessControlDecision (UC010): The user presses the *Access Control Decision* button for the system to provided a decision regarding granting or denying access to the graphics file. The access control decision is displayed in the center bottom area of the authentication interface panel.

5 Requirements Traceability Matrix

Figure 4 shows the mapping between the functional requirements and the corresponding use cases which implement the requirement. This traceability matrix is used to check if the requirements are being met.

		Use cases									
		UC01	UC02	UC03	UC04	UC05	UC06	UC07	UC08	UC09	UC10
Requirements	F_DT_01						X				
	F_DT_02	X									
	F_PR_01								X		
	F_PR_02							X			X
	F_PR_03					X					X
	F_PR_04			X	X						
	F_PR_05		X								
	F_PR_06				X						
	F_PR_07										X

Figure 4: Requirements Traceability Matrix

6 User Interface

Figure 5 illustrates the main GUI of the system. The GUI provides an interface with buttons and axes to interact with the user and display intermediate phases of the biometric watermarking process. The GUI is divided into two sections to represent the two phases of the system - Enrollment and Authentication. For the enrollment phase, the end user only needs to interact with the system using the first button (*Provide 3 Voice Samples*) which prompt the user to upload his voice samples, the *Input 3D Mesh Model* button which prompts the user to upload the 3D model that needs to be watermarked, and the last button (*DRM Protected 3D Model*) which inserts the voice biometric watermark into the uploaded 3D model and provides a rendering of the watermarked 3D model in the right bottom display area. The remaining intermediate buttons have been provided to enable - i) users to understand the various steps involved in the biometric watermarking process and ii) visualize the intermediate steps of signal pre-processing, MFCC feature extraction and GMM speaker modeling. The right panel does not provide an interface for the intermediate steps of watermark extraction solely for the purpose of giving users a feel of the simplicity of interacting with the system during the authentication phase. Figure 6 demonstrates the 3D model [11] watermarked using the voice signal plotted in red.

Figures 7 and 8 illustrate the system's access control decision for a given voice sample and watermarked 3D model.

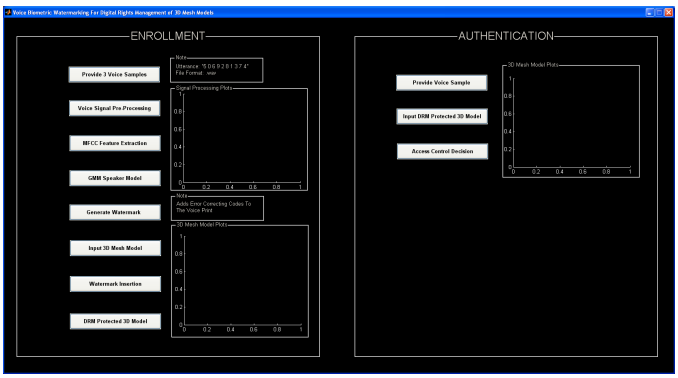

Figure 5: Main GUI - The left panel provides an interface for user enrollment. The right panel provides an interface for user authentication.

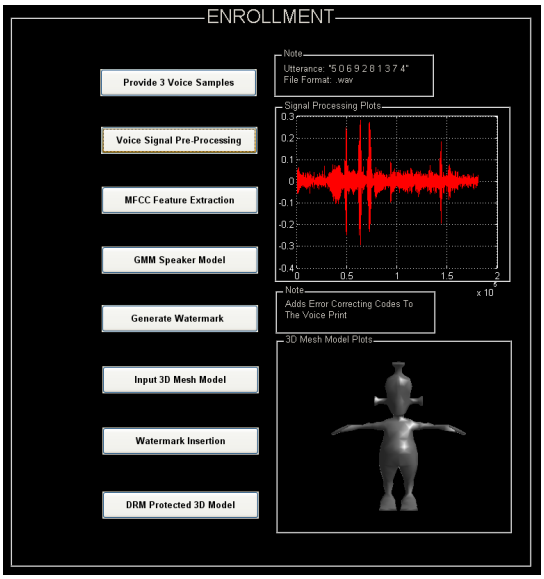

Figure 6: Enrollment (Left Panel of Main GUI) - The axis on the upper left corner shows a plot of the voice signal used for generating the watermark. The bottom right display area shows the watermarked 3D model.

7 Conclusion and Future Work

This prototype has been developed to demonstrate the proof-of-concept for protecting copyrights of graphic artists using a consumer's biometric trait. The consumer provides his or her voice recording during the enrollment phase and a voice print is extracted from the digitally recorded voice signal. A fragile invisible watermarking technique embeds the voice print into a 3D computer graphic model. Any slight modification to the watermarked model, whether casual or malicious, that attempt to remove the watermark from the 3D model prompts the authentication phase of the system to deny the consumer access to the model. The system thereby sieves pirate consumers from legit-

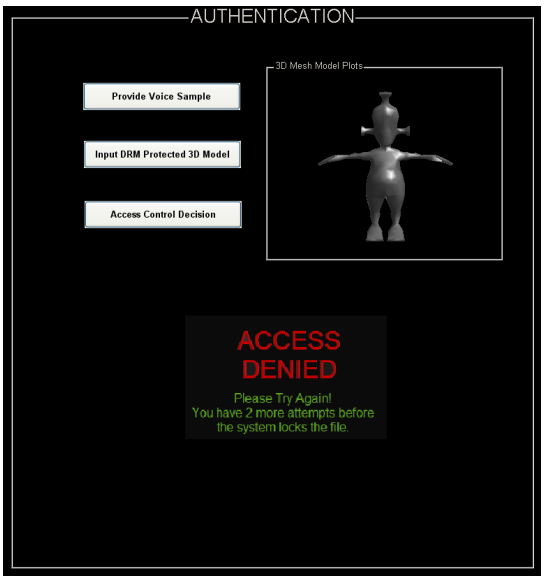

Figure 7: Authentication (Right Panel of Main GUI) - The user has only 3 verification attempts to gain access to the 3D mesh model. The system is guaranteed to grant access to a legitimate user within the three trials, since the False Reject Rate for the system is 0.333.

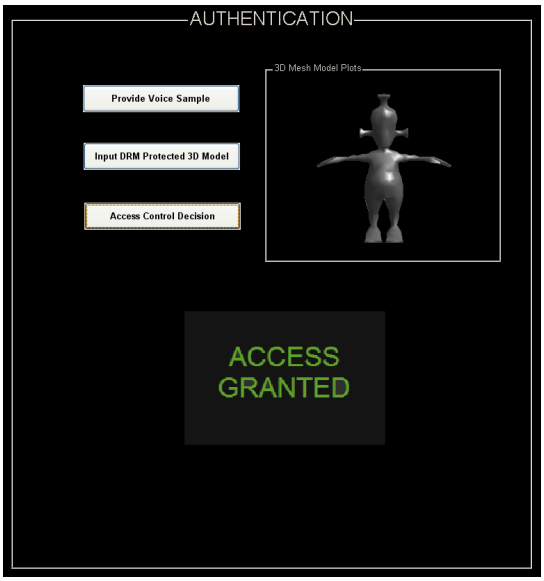

Figure 8: Access Control Decision - A legitimate user is granted access to the watermarked 3D model upon providing a voice sample.

imate consumers. Future work entails providing support for other popular 3D graphic file formats other than the *.off* format such as *.max, .3ds. .blend, .lwo, .md2, .md3*, and *.x*, and to employ various other biometrics such as fingerprint images and facial features to improve the system accuracy in terms of False Accept Rate (illegitimate user is falsely granted access by the system since biometric traits are not distinct over the population) and False Reject Rates (legitimate user is denied access due to variations in the user's biometric trait acquired at different instants of time).

References

[1] D. Han, X. Yang, and C. Zhang. A Novel Robust 3D Mesh Watermarking Ensuring the Human Visual System. In *Proceedings of Second International Workshop on Knowledge Discovery and Data Mining*, pages 705–709, Jan. 2009.

[2] P. Alface and B.Macq. Blind Watermarking of 3D Meshes using Robust Feature Points Detection. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 693–696, September 2005.

[3] M. Motwani, N. Beke, A. Bhoite, P. Apte, and F. Harris Jr. Adaptive Fuzzy Watermarking for 3D Models. In *Proceedings of International Conference on Computational Intelligence and Multimedia Applications*, volume 4, pages 49–53, 2007.

[4] O. Benedens. Watermarking of 3D Polygon based Models with Robustness against Mesh Simplification. In *Proceedings of SPIE Security and Watermarking of Multimedia Contents*, volume 3657, pages 329–340, 1999.

[5] R. Motwani. *A Voice-Based Biometric Watermarking Scheme For Digital Rights Management of 3D Mesh Models*. PhD thesis, University of Nevada, Reno, 2010.

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

[7] R. Motwani, S. Dascalu, and F. Harris. A Voice Biometric Watermark for 3D Models. In *Proceedings of IEEE International Conference on Computer Engineering and Technology (ICCET)*, pages 702–712, 2010.

[8] Side Effects Software Inc. Houdini 3D Animation Tools. http://www.sidefx.com/index.php?option=com_content&task=view&id=589&Itemid=221. Last Accessed Jun 17, 2010.

[9] Alpha Tec Ltd. Watermarking Software. http://www.alphatecltd.com. Last Accessed Jun 17, 2010.

[10] I. Sommerville. *Software Engineering*. Addison-Wesley, 2006.

[11] G. Peyre. Toolbox graph - A Toolbox to Perform Computations on Graph. Last Accessed Jun 17, 2010.