The 9th International Conference on Cognitive Science

Emotion expression of avatar through eye behaviors, lip synchronization and MPEG4 in virtual reality based on Xface toolkit: Present and future

Ahmad Hoirul Basori, Itimad Raheem Ali

UTM VicubeLab, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
Interactive Media and Human Interface Lab, Department of Informatics, Faculty of Information Technology, Institut Teknologi Sepuluh November, Sarabaya, Indonesia

Abstract

Eye movement combined with lip synchronization, eye movements, and emotional facial expression revealed an interesting research field that gives information about verbal and nonverbal behaviors occurring in the human body. Most of the previous researchers focused on eyes gazes, lip synching and emotion expression which are the most important features that can transfer nonverbal information to enhance, understand or express emotion. In this paper, the recent advances in 3D facial expression are introduced focusing on the presentation of Xface platform toolkit that developed a 3D talking avatars synthesis by implementing text-to-speech engine (TTS) to depict the basic lip shapes necessary for each phonemes to convey the dialogue. This work is believed to give the future direction that can lead into new research issue in facial animation.

© 2013 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and/or peer-review under responsibility of the Universiti Malaysia Sarawak.

Keywords: Eye movements; Lip Synchronization; Emotional Facial Expression.

1. Introduction

Animated virtual humans have widely been used in many applications in recent years, such as movies, games and embodied conversational agents (ECAs). Facial behavior is very important to increase the reality of the character. Most of the latest research on facial animation has led to human models that explore expressiveness, communication and interactivity; much of these researches are focused on ECA development [1].

The combination between eye movement, lip synchronization, gesture, facial expression and body orientation all together give information about human ideas, consideration and knowledge of the human emotions. While the emotions play an essential role in collective interactions performs important organize and beneficial functions within human body to facilitate making decision and perception. Terms like “emotional design” refer to the integration of emotions in the design of computer systems with attempting to more reality in understanding and use speech representation. The gaze and saccade are coupled with human award process therefore it acts like "a window to the mind". Synthesizing eye movement is blended to the reality and believability of virtual agent and the gaze conveys more nonverbal information and emotional intentions when the human speaks. The researcher tries to realize the
behavior of embodied expression using many systems, one of these is BEAT (The Behavior Expression Animation Toolkit), which allows the virtual character to speak input text written by the animator and obtaining nonverbal expressive behaviors and synthesized speech [2].

Realistic facial animation is still challenging tasks despite of extensive research because description of emotion is an important process in human’s intelligence. However, it is important to generate virtual characters that can produce realistic utterances with facial animations. This combination is closed relationship between lip movement and face expressions. For example, the virtual character talking while at the same time looking to another object. Thus, interesting speaking and gazing are prominent research issues in animation research. This is a significant challenge as nonverbal behavior can be complex.

2. Related Works

Building virtual characters require a combination of many researchers’ skills. The combination of eye gaze with lip synchronization, facial expression and gestures require different skills from their researchers. Recently, there has been the need to share issues and expertise to improve the component of virtual characters and to achieve realistic motion. The face is considered as the most important feature of human communication tools that can transfer verbal and non-verbal information.

Previous researches have introduced new tools like Xface developed by Koray Balci, it is an Open Source Project and SMIL-Agent Scripting Language to create and animate Embodied Conversational Agents (ECAs) by a set of tools that is easy to use and extend [3].

Queiroz et al developed a usable, extendable and robust facial animation platform, capable of easily animating MPEG-4 parameterized face through high-level description of facial actions and behaviors by providing interaction between the user and a virtual character, as shown in Fig. 1 [1].

Attracted with human to human interaction, Bailly examined audio and visual face-to-face interaction between human-human and a human- virtual conversational agent with the aim on mutual gaze patterns during interactions by new technological devices and at measuring the selection impact of award states and communicative functions [4].

![Fig. 1. Overall architecture diagram of framework [1].](image-url)
2. Facial Animation And Emotional Facial Expression

Emotional facial expression is an excellent way to express human perception, intent and other verbal and nonverbal expressions in communication. Computer science becomes more interesting with computer vision and graphics to analyze and synthesize facial expression automatically. The realistic emotive speech’s synthesis remains indefinite and little was known about speech gestures with facial expressions, despite the previous efforts that made it possible to develop the coding system. The research leads to develop methodologies for the formation of an emotive 3D humanoid audio visual character. The research is believed to advance the interaction of human computer to more reality to human-human interaction. Actually, the technology is distinct to conduct research on emotional expression during speech and eye gaze [5]. The previous efforts were made possible in the development of the psychology of coding systems to produce emotive speech and facial expression by describing facial actions and its behaviors. Facial Action Coding System (FACS) was the most exhaustive and informed of the MPEG4 facial animation parameters that enabled automated synthesis of facial actions and expression [6], as shown in Figure 2. It is eligible to make virtual character similar to human communication capable of showing emotive speech and facial expressions.

The aim of most researchers, for a long time has been to describe appropriate and effective human emotion and facial expression. Darwin worked with emotion, motivational, behavioral or personality, sensations and cognitive processes. Ekman et al, proposed a theory that has an influence on psychological research on facial expression. It is a popular and simplest approach to describe emotional state [8]. And in 1978, Suwa et al presented an introductory examination on automatic facial analysis from a sequencing image [9].

Parameterization the facial expression by Ekman used a set of Action Units (AU)[8][10]. This Action Units (AU) defines facial expression in small regions. Most previous approaches are based on 2D motion that didn't provide any control to the emotion but Cao et al proposed another approach of a data driven for animating 3D face model, make organization to gain an efficient data structure, to allow a search algorithm to locate appropriate movement and implement a Support Vector Machine (SVM) that automatically detected the emotion of arbitrary input utterance. The difficult question to be answered is how we create virtual human capable to show the emotional during the conduct gaze and speech behaviors [11]. Brent Lance and Stacy Marsella defined a gaze manner as changes in physical parameters’ movements, such as the head speed in a gaze shift. The challenge is in developing a model of gaze manner because the gaze is not simple. Gaze is a complex behavioral that includes eye movements, posture, head movements, and all these components should be observed [12]. To be believable, ECAs should reflect multimodal human nature conversation that contains verbal and nonverbal expression. This is challenge research, as nonverbal behavior can be extremely complex, such as personality, sex, emotion [7][13][14].

The exciting part in facial expressions is diverse due to skillful promotions in related work areas such as face detection, face recognition. Facial expression analysis is used in various applications and can be visualized in future and interesting research in different areas[15], such as face image compression and synthetic face animation [16]. The recognition of facial expression transacts with the classification of facial motions and features into groups that are simply based on visual information, human emotions are detected through many stages such as; voice, gestures, pose, gaze and expressions. On the other hand, facial expressions have many sources, one of them emotions; emotion recognition interprets to attempt and often demands understanding of a given status [17]. A robust expression transfer (NET) model presents an animation method to transfer facial expressions extracted from video to the facial sketches [18] and fully automatic, robust and fast system which generate a 3D face model from an
image of a face and allow to synthesis expressions [19] with the possibility of using effective virtual communication [20].

3. Facial Generation For Emotion Expression

Human behavior understanding has attracted a great deal of interest over the past two decades, mainly because it spans a variety of fields such as; computer technology, medicine and security. Facial expression is the most convincing, essence of the environment for humans to communicate emotions, to clarify and give highlights to express intentions and to regulate interactions with other people. Most of the available datasets of expressive faces contained only posed affective displays, mainly of the prototypical expressions of six basic emotions (i.e. happiness, sadness, surprise, anger and disgust). This paper focused on the presentation of Koray Balcı platform independent toolkit for developing 3D talking agents, namely Xface. The toolkit includes three main features:

- The core Xface library is for developers who want to embed 3D facial animation to their software as well as researchers who want to focus on related topics without the hassle of implementing a full framework from scratch.
- XfaceEd editor provides an easy to use interface to generate MPEG-4 ready meshes from static 3D models.
- XfacePlayer is a sample application that demonstrates the toolkit in action. It supports SMIL-Agent scripts, MPEG-4 FAPs as input.
- XfaceClient provides possibility of remote control for XfacePlayer over TCP/IP.

The development of Xface project is a set of tools that are easy to use and extendable, that opens to the researcher and developers prospects for expansion in new research, as shown in Figure 3.

![Avatar expression process of Xface](image)
Table 1. Xface Emotional expression results.

<table>
<thead>
<tr>
<th>Expression Type</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice with Normal Expression</td>
<td><img src="image1" alt="Image" /></td>
</tr>
<tr>
<td>Alice with Fear Expression</td>
<td><img src="image2" alt="Image" /></td>
</tr>
<tr>
<td>Alice with Angry Expression</td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>Alice with Sad Expression</td>
<td><img src="image4" alt="Image" /></td>
</tr>
<tr>
<td>Alice with Open Smile Expression</td>
<td><img src="image5" alt="Image" /></td>
</tr>
<tr>
<td>Alice with Closed Smile Expression</td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>Alice with Surprise Expression</td>
<td><img src="image7" alt="Image" /></td>
</tr>
<tr>
<td>Alice with Disgust Expression</td>
<td><img src="image8" alt="Image" /></td>
</tr>
<tr>
<td>Alice with right Blink</td>
<td><img src="image9" alt="Image" /></td>
</tr>
<tr>
<td>Alice with left Blink</td>
<td><img src="image10" alt="Image" /></td>
</tr>
<tr>
<td>Alice with Looking Down</td>
<td><img src="image11" alt="Image" /></td>
</tr>
</tbody>
</table>

Table 2. Xface shapes for the phoneme results.

<table>
<thead>
<tr>
<th>Phoneme Type</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>aah</td>
<td><img src="image12" alt="Image" /></td>
</tr>
<tr>
<td>m, b, p</td>
<td><img src="image13" alt="Image" /></td>
</tr>
<tr>
<td>r</td>
<td><img src="image14" alt="Image" /></td>
</tr>
<tr>
<td>big aah</td>
<td><img src="image15" alt="Image" /></td>
</tr>
<tr>
<td>ch, j, sh</td>
<td><img src="image16" alt="Image" /></td>
</tr>
<tr>
<td>d, s, t</td>
<td><img src="image17" alt="Image" /></td>
</tr>
<tr>
<td>ooh,q</td>
<td><img src="image18" alt="Image" /></td>
</tr>
<tr>
<td>ee</td>
<td><img src="image19" alt="Image" /></td>
</tr>
<tr>
<td>eeh</td>
<td><img src="image20" alt="Image" /></td>
</tr>
<tr>
<td>f, v</td>
<td><img src="image21" alt="Image" /></td>
</tr>
<tr>
<td>th</td>
<td><img src="image22" alt="Image" /></td>
</tr>
<tr>
<td>i</td>
<td><img src="image23" alt="Image" /></td>
</tr>
<tr>
<td>k</td>
<td><img src="image24" alt="Image" /></td>
</tr>
<tr>
<td>n</td>
<td><img src="image25" alt="Image" /></td>
</tr>
<tr>
<td>w</td>
<td><img src="image26" alt="Image" /></td>
</tr>
<tr>
<td>oh</td>
<td><img src="image27" alt="Image" /></td>
</tr>
</tbody>
</table>
The principle of the core library is to have an interface to simplify the use as an application. It is an extendable library that provides opportunities for the researchers according to their area. The core library is responsible to load the face models and its FP and FAPU information ordering to create facial animation. The system was implemented to the keyframe animation and it has a key model for each emotion and viseme with separate rendering module with the possibility to add various advanced rendering techniques.

Moreover, the first step of XfaceEd relies on the creation of the static 3D face mesh in a 3D modeling package. XfaceEd simplifies the creation of speaking agent by defining the FAPU, FP regions, weights and parameters for manipulating and animating the static face models.

Finally, XfacePlayer is an application that shows the implementation of a face player using Xface library. Table 1 shows Xface emotional expression results. Presentably, it uses SDL library and Boost library to manage the creation of the face.

4. Xface with Speech Synthesis

The focus of animation system is to develop characters that can react and perform conversation in a natural environment with the greatest amount of realism; it can produce arbitrary utterances with reasonable looking facial movements. This system uses a set of visemes that are activated by a text- to- speech engine (TTS). The TTS engine covert an utterance in text format into a series of phonemes and poses. In addition, the system blends a set of the Facial Action Coding System (FACS) units used to express emotion and display facial movements, and speech synthesis is implemented by using text-to-speech systems. The representation of the phoneme is difficult according to a specific language and differs from one language to another; Xface represents the Italian language, as shown in the table 2.

5. Discussion and Conclusion

A lot of existing 3D corresponding algorithms are computationally expensive, so it is considered as a major challenge in 3D face tracking. 3D emotional facial expression analysis is still at its early stage. Xface toolkit is a suitable system for human-computer interaction giving a proper preparation of an emotional speech database with evaluating performance. It relies on MPEG-4 Face Animation (FA) standard. Xface presents an interactive facial animation framework by considering emotional facial expressions, synchronized speech, and the generation of eye behaviour, where the users define the characters’ actions by high-level descriptions. It followed the MPEG-4 facial animation standard for generating animation. Consequently, the generated animations can be used for different face models. The real face acquired from advanced technologies such as 3D laser scan is expected to give great improvement on interactive avatar inside virtual environment.

Tracking lip synchronization still remain as a research challenge because the complexity of speech synthesis and lip movement. The eyes behaviour based on emotion condition also can be considered as future challenge on facial animation research.

Future researches aim to improve the module of emotional facial expression since it is considered as a challenge to generate real time emotion module and support visemes from other languages. Xface results show the promise to allow the integration of other researches in the area by providing interactive control of virtual character such as; Embodied Conversational Agents and Games.

Acknowledgements

This research is supported by the Ministry of Science and Technology (MOSTI) and collaboration with Research Management Center (RMC), Universiti Teknologi Malaysia (UTM). This paper is financial supported by E-Science Grant Vot. No.: R.J130000.7911.4S026
References