An Approach to Modeling Facial Expressions Used in American Sign Language

Brock Craft, Damien Hinkle, Eric Sedgwick, Karen Alkoby, Mary Jo Davidson, Roymieco Carter, Juliet Christopher, Jacob Furst, Brian Konie, Glenn Lancaster, Steve Luecking, Ashley Morris, John McDonald, Noriko Tomuro, Jorge Toro, Rosalee Wolfe

School of Computer Science, Telecommunications and Information Systems
DePaul University
243 South Wabash Avenue Chicago, IL 60604-2301
asl.cs.depaul.edu

Abstract

American Sign Language (ASL) is the natural and living language of the Deaf Community in North America. In addition to hand gestures, facial expressions are a key component of communicating in ASL. We present a method for reproducing facial expressions through computer graphic animation. Directions for further research are suggested.

Keywords: Animation, American Sign Language, Facial Expressions

1. Introduction

American Sign Language (ASL) is the third most widely used natural language in the United States and is the primary means of communication among the North American Deaf Community [1][2]. It is a natural, living language composed of its own linguistic characteristics, including its own distinct vocabulary, grammar, and syntax [3] [4]. Contrary to popular notions amongst English speakers, ASL is not simply a sophisticated gestural representation of spoken English. This misapprehension is common because most people outside the deaf community have little knowledge of ASL. In this way, deafness presents a barrier not only of sound, but also of language.

To facilitate communication with the speaking community, deaf people primarily rely on human interpreters who are professionally trained in understanding the unique challenges of interpreting from spoken language to ASL. When ASL Interpreters translate the sounds of spoken English into ASL, they do not merely render English words with their hands. Rather, they must take into account all of the linguistic characteristics of ASL and communicate the spoken information in a way that is meaningful and accurate to its particular characteristics, using hand signs, fingerspelling, and facial expressions. The shortage of hearing people who understand ASL and the limited availability of ASL interpreters creates barriers between the deaf and hearing communities.

English is a second language for most deaf people. The average reading ability of deaf adults in the US is between the third and fourth grade level [5], because the linguistic differences between these two languages are significant and complex. When an interpreter is not available, written English provides only limited assistance in communication. English subtitles and closed-captioning, which are useful for translating foreign languages for hearing audiences, are not as effective as an ASL interpreter because the captions are not in a deaf person's native language. In addition, they are only available in predetermined, controlled circumstances, wherein the text has been prepared beforehand. They are not context-sensitive, not portable, and often, not accurate.

We believe that this circumstance calls for an electronic translating system between these two languages, a *personal digital translator,* which would be an effective and practical tool for bridging the formidable communication barriers between them. In much the same way that portable foreign language translators provide assistance for English speakers, a similar device would convey written or spoken English visually, in ASL. Such a system should be easy to use, portable, affordable, and would provide effective interpretation in unanticipated situations.

2. ASL facial expressions

In order for such a solution to be viable, it will have to communicate more than hand gestures alone. This is because in addition to the gestural communications of hand signs and fingerspelling, facial expressions are a key component of ASL. Speaking individuals are used to understanding facial expressions as ancillary components to meaning. While it is possible to understand the meaning of an English sentence without seeing facial expressions, this is less the case for ASL. In ASL, facial expression is a key component of grammar that involves far more than merely emotional disposition or level of intensity. Unlike spoken language, the gestural complexity of ASL necessitates the use of the head and facial features as an intrinsic part of the language. Facial expressions are used in conjunction with word signs and fingerspelling to communicate specific vocabulary, questions, intensity, and subtleties of meaning [4].

There are also large numbers of hand signs that are formed with the face, its features and expression [6]. For example, the words AMAZE and WORRY require the use of facial expressions [7]. To pose a yes/no question, deaf people raise their eyebrows. Also, questions such as who, what, when, where, and why (wh- questions) are formed by furrowing the eyebrows. In this context failing to use the eyebrows would be the spoken equivalent of not asking a question at all. In the development of this proposed system, therefore, we have undertaken the modeling of facial expressions in ASL. This involves the use of computer animation and models to represent the facial structure and its expressions. Two methods have emerged in the course of this research.

3. Modeling facial expressions

Computer graphic models of the human form in three dimensions are based upon a mesh of mathematical points, or *vertices*, that are used to form the surfaces of the body. Kinematics, the movement of these vertices, accomplishes the movement of the human model and changes in the position of the body parts relative to one another. The movement can be accomplished by specifying the appropriate coordinates overtly, as from a database, or by calculating new positions mathematically, based on changes in the locations of vertices in other parts of the model [8]. For example, since movements of certain objects can be tied to and correlated with movements of other parts of the human model, it is possible to move the position of an arm by moving the position of an invisible object within the arm. An unseen bone can be made to change the position of the arm that it resides within, relative to the rest of the body. In the same way, unseen muscles can be modeled to alter the positions of body parts and surfaces [9].

This is particularly useful where modeling of the facial features is concerned because of the varied and numerous positions that the face can assume. In order to make a facial expression, there are 17 pairs of muscles in the face that change position beneath the skin, which causes a change in the surface of the skin [10]. In the same way, modeling of the facial expressions of a computer-generated model can be accomplished by manipulating unseen objects representing the facial muscles. Attaching the surface of the facial skin to these muscles causes the skin to form various expressions as the muscles are moved. While this technique of modeling expressions has been studied extensively, knowledge in this domain has not been applied specifically to solving the problems presented in building an interpreter for ASL. In this context, the muscle movements could be based upon a database of facial expressions that would represent the relationships of the expressions to the vocabulary, grammar, and syntax of ASL.

However, this approach has a significant drawback in that it requires a large number of mathematical calculations in order to determine the position of the skin surfaces based upon the movements of the muscles underneath. Further an extensive amount of designer time is required to model the shape and behavior of facial muscles. While we concede that it is likely that such a technique will need to be utilized, other alternatives are clearly warranted, especially for the more subtle changes in expression.

4. Mapping facial expressions

Toward this end, we are exploring another method of changing facial features on human models. This technique takes advantage of the fact that the coloration of the facial skin can be accomplished by applying a two-dimensional *texture map* over the three-dimensional contours of the model's facial structure [11]. Rather than changing the position of the skin itself, it is possible to change the coloration of the skin to produce certain facial expressions. This is much cheaper, computationally, than moving the locations of the three-dimensional vertices especially when depicting subtle variations in geometry [12]. In addition, different variations in the image map can be stored and linked to a database that specifies which images are appropriate to use for a given set of linguistic requirements. The images can be retrieved as needed and applied to the facial structure to represent the appropriate expression.

Because the human model is a three-dimensional object and the image map is two-dimensional, the positions of landmarks in the two-dimensional model must be positioned carefully so that they will be located in the proper position when the image map is applied [13]. In the same way a mercator projection distorts the continents on a map of the Earth, the texture map distorts familiar facial landmarks with odd effects, as in the middle image of Figure 1. In this image, coloration has been provided for the cheeks and the orbits of the eyes. Also, the eyebrows have been painted into the image. The result is that the coloration of the texture map simulates the geometry of the eyebrows.

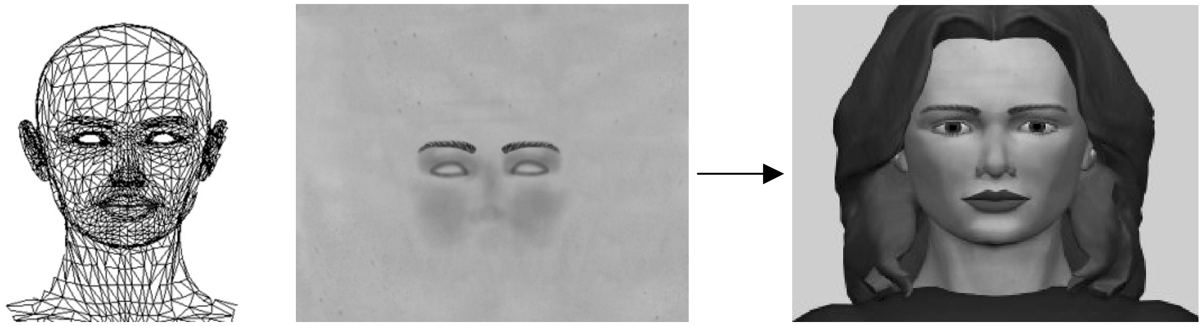

Figure 1.

In areas where the face is relatively flat and where movements are subtle, it is possible to change the facial expression of the model by painting features such as the eyebrows in slightly different positions. Wrinkles can also be effectively created in areas where the skin folds only slightly by darkening the skin tones along the wrinkle lines. This technique has proven very effective in the areas between the eyebrows and in the wrinkles that form on the forehead when the eyebrows are raised.

Our research of available literature on facial expressions used in ASL reveals that there are many relevant facial expressions that can be created using this technique [7][4][1]. However, there are no current sources that catalog all of the facial expressions in ASL. To find accurate depictions of ASL facial expressions, we examined artists' renderings of facial expressions found in portraiture texts. Based on this research, we have begun to construct a database that will contain the appropriate relationships between lexical and grammatical elements in ASL and their correlated facial expressions. By associating the appropriate image maps with these elements, it will be possible to create the facial expression for a given linguistic situation with the most computationally economical technique.

Particular to this technique, the wh- questions and the yes/no questions, which require eyebrow movement, can be formed by changing the texture map of the face. However, difficulties arise in determining how much to change features so that they remain realistic and credible when they are applied to the three-dimensional model. It is hard enough to create a credible life drawing of human expression. It is prohibitively difficult to draw on a distorted canvas with the hope that the expression will "look right" when it is texture mapped onto a three-dimension object. To cope with this problem, we developed a *registration technique* to correlate artist's renderings of facial expressions to the initially distorted texture map By digitally scanning existing artist's renderings of facial expressions we were able to superimpose those images onto the image map that is applied to the human model, using Adobe Photoshop image processing software. This superimposition provides guidance in appropriately repositioning the eyebrow features of the image map, depending on the expression that is required, in the case of Figure 3, a wh- question:

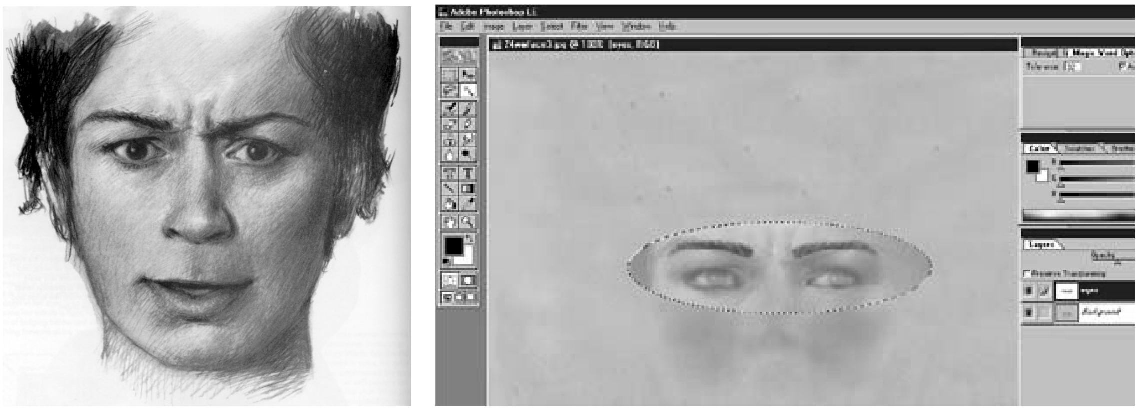

Figure 3.

When the results of this repositioning of the eyebrows are applied to the three-dimensional model, a relatively realistic rendering of the facial expression can be accomplished. Compare Figure 4, which shows the model with its original, neutral texture map, with Figure 5, which shows our results in using this technique. The geometry is identical in Figures 4 and 5. Only the texture map has changed:

Figure 4.

Figure 5.

Notice that the facial features of the original artist's rendering in Figure 3 are quite different from the features of the synthetic model in Figure 4. This method will allow us to transcribe a large number of facial expressions from life drawings to three-dimensional models.

5. Directions for research

We are now in the process of creating and animating facial expressions that involve the upper part of the face. However, for the lower part of the face, it is clear that we will need to address the issue of modeling musculature. In addition, user testing is required assess the overall effectiveness of this technique.

References:

[1] Sternberg, M.: The American Sign Language Dictionary, *Multicom* (CD-ROM), 1996.

[2] Deafworld web site: dww.deafworldweb.org/int/us/hm as of October, 4 2000.

[3] Valli, C. and Luca, C.: *Linguistics of American Sign Language*, Galluadet University Press, 1993.

[4] Baker-Shenk, S. and Cockely, D.: American Sign Language, A Teacher's Resource Text on Grammar and Culture (the Green Book), Gallaudet University Press, 1980.

[5] Holt, J.: Demographic, Stanford Achievement Test – 8th Edition for Deaf and Hard of Hearing Students: Reading Comprehension Subgroup Results. http://www.gallaudet.edu/~cadsweb/sat-read.html. As of October 2000.

[6] Bridges, B. and Metzger, M. *Deaf Tend Your: Non-Manual Signals in American Sign Language*. Silver Spring, Maryland: Calliope Press, 1996.

[7] Tennant, R. and Brown, M. G., *The American Sign Language Handshape Dictionary* Washginton, D.C.: Clerc Books, 1998.

[8] Foley, J., van Dam, A., Feiner, S., and Hughes, J. *Computer Graphics: Principles and Practice.* 2nd ed. Reading, Massachusetts: Addison-Wesley, 1990.

[9] Doenges, P., Capin, T. K., Lavagetto, F. Ostermann, J., Pandzic, I. S., and Petajan, E. D. MPEG-4: Audio/Video & Synthetic Graphics/Audio for Mixed Media, *Image Communications Journal, Vol.5, No.4, May 1997.* http://miralabwww.unige.ch/ARTICLES/ICJ97.html

[10] from the web pages of the Biology department and the University of Indianapolis. http://biology.uindy.edu/Biol504/HUMANSTRATEGY/6skull.htm. As of October 2000.

[11] Pighin, F., Hecker, J., Lischinski D., Szeliski, R., and Salesin, D. Synthesizing Realistic Facial Expressions from Photographs. *Proceedings of SIGGRAPH 98,* 75-84.

[12] Blinn, J. Simulation of Wrinkled Surfaces, *Proceedings of SIGGRAPH 78*, 286-292.

[13] Wolfe, R. *3D Graphics: A Visual Approach.* Oxford: Oxford University Press, 1999.