Modeling Animations of American Sign Language Verbs through Motion-Capture of Native ASL Signers

Pengfei Lu
The City University of New York (CUNY)
Doctoral Program in Computer Science
CUNY Graduate Center
365 Fifth Ave, New York, NY 10016
+1-212-817-8190
pengfei.lu@qc.cuny.edu

ABSTRACT
Software to generate American Sign Language (ASL) automatically can provide benefits for deaf people with low English literacy. However, modern computational linguistic software cannot produce important aspects of ASL signs and verbs. Better models of spatially complex signs are needed. Our goals are: to create a linguistic resource of ASL signs via motion-capture data collection; to model the movement paths of inflecting/indicating verbs using machine learning and computational techniques; and to produce grammatical, natural looking and understandable animations of ASL. Our methods include linguistic annotation of the data and evaluation by native ASL signers. This summary also describes our research progress.

Keywords
Animation, American Sign Language, Accessibility Technology, Machine Learning, Motion Capture, Inflecting Verb, Indicating Verb.

1. MOTIVATIONS AND BACKGROUND

There are several accessibility motivations for generating ASL animations for people who are deaf. However, generating ASL software is quite challenging for several reasons:

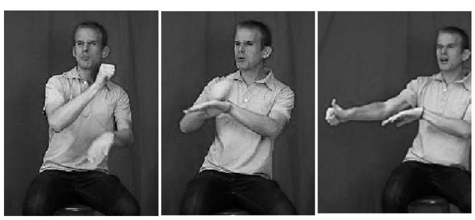
1. There is no standard written form for ASL, thus an ASL system cannot produce text output. Our software must produce an animation displaying a human character's signing. This lack of a writing system makes it difficult and expensive to collect a large corpus of ASL with sufficient details for computational research purposes.
2. Movement data is needed for animation research. Previous researchers have collected some video corpora of ASL that are annotated by ASL linguists [1]. However, it is challenging to record the signer's complex body joint movements, handshapes, facial expression, head tilt, and eye gaze from a video. Motion-capture technology is more reliable for this level of detail.
3. When signers use ASL, they use the space around them. The linguistic use of space is discussed below.

1.1 Linguistic Use of Space

Signers arrange invisible placeholders in the signing space to represent entities in the conversation [1] [2]. We shall refer to these placeholders as "tokens" [3]. Signers assign objects and persons under discussion to token locations, and they later refer to these entities by pointing to the token locations. The personal, possessive, and reflexive pronouns in ASL involve pointing to token locations.

Many verbs change their motion paths to indicate subject, object, or both. These verbs have been referred to as “inflecting verbs” [4], “indicating verbs” [3], and “agreeing verbs” [5]. We shall compromise by calling them “inflecting/indicating verbs.” Generally, the motion paths of inflecting/indicating verbs change so that their direction goes from the subject to the object. However, their paths are more complex than that. Each verb has a standard motion path that is affected by the subject and object locations in space around the signer. In Figure 1(a), three images are shown of the beginning, middle, and end of the movement of the ASL verb sign "blame"; the subject of the verb has been set up on the left side, and the object, on the right. In Figure 1(b), three images of the performance of the verb "blame" are shown again, but this time, the position of the subject and the object are reversed.

(a) Mary(token location on the left) blames John(token location on the right).



(b) John(token location on the right) blames Mary(token location on the left).

Figure 1. Subject and object location information incorporated into the movement path of an ASL inflecting/indicating verb, “blame.”

There are also other verb classes in ASL we are not focusing on in this research: plain verbs and classifier predicates. The motion paths of plain verbs do not change based on the locations in space established for their subject/object. Other ASL signs are produced using handshapes that indicate certain semantic features of entities being depicted in a 3D scene under discussion. The handshapes are called classifiers [6] and the signs containing them are called classifier predicates [7]. The motion of classifier predicates change in a 3D representative way and relate to where and how the real-world entity being discussed actually moves. Researchers have modeled and generated some animations of ASL classifier predicates [8] [9], but they are not the focus of our current work.

1.2 Prior Work

It has been experimentally determined that native ASL signers achieve better comprehension scores when viewing ASL animations in which the virtual human character uses token locations [10]. To generate such animations, researchers could benefit from studying token location data in collections of ASL sentences performed by human signers. Unfortunately, researchers building collections of ASL sentences for research purposes have not recorded 3D coordinates of token locations. Current ASL generation technology cannot predict how to dynamically modify the movement of signs based on these spatial associations.

There have been several research projects studying sign languages animations. The European eSIGN project [11] and ViSiCAST project [12] generate some British Sign Language (BSL) verb performances by a virtual human. BSL is a distinct language from ASL but shares some linguistic properties. However the European eSIGN and ViSiCAST projects cannot produce the complex inflecting/indicating verb performances discussed above. They only model a few verbs that move in a line from one location in space to another. SignSmith Studio is a commercial sign language animation program by VCom3D, Inc. [13]. This product provides users a dictionary of signs, a fingerspelling generator, and limited control of eye-gaze and facial expression. The product can replace each word of a scripted English sentence with a corresponding ASL sign without morphological modifications added to the animation, which must be later added by a human editor. Using Sign Smith Studio, usually only people that are fluent in ASL can manually create content that incorporates the signs, facial expressions, mouthing, and eye gaze of real ASL signing.

2. RESEARCH GOALS

The goal of our research is three-fold:

1. To create a permanent linguistic research data resource of digital 3D movement information of ASL body movement and handshapes collected from native signers that can be studied by future linguists.
2. To analyze the 3D spatial motion paths of the signer's hand and to uncover patterns of inflecting/indicating verbs in how those paths relate to the content of the sentences using computational techniques.
3. To produce animation of a character performing ASL based on our data analysis and to ask native ASL signers to evaluate the quality of animation during the development.

To achieve the goals, the first stage of our research is to accurately and efficiently record 3D motion-capture data from ASL signers. We are using a customized configuration of commercial motion-capture devices: Animazoo IGS-190 (spandex bodysuit with inertial magnetic sensors), Intersense IS-900 (microphone sensors using acoustic triangulation for head tracking), Applied Science Labs H6 eye-tracker (a near-eye camera), and a pair of Immersion CyberGlovesff. By collecting movement data from native ASL signers, we will build more accurate computer models for ASL synthesis.

We want to learn how the token locations affect sign movement and how inflecting/indicating verbs change their movements based on how tokens have been set up. We are collecting a multi -sentence corpus in which signers set up tokens as they sign and also collect repetitions of some ASL inflecting/indicating verb performances. We are designing experiments to record different performances of some inflecting/indicating verb performances with different given arrangements of tokens in the signing space. We plan to record the shapes of each hand, 3D orientation of the palm, and 3D coordinates of hands. Based on our collected data, we will design algorithms to calculate realistic movements for a signer.

We will experiment with machine learning techniques to learn and build the mathematical models of inflecting/indicating verbs of when tokens are set up and where they are placed.

We are considering using some of the motion-capture learning techniques like those of [14] [15] [16] to learn the verb performances. These models will be encoded in our ASL generation software to produce animations.

Native ASL signers will evaluate the animations we produce using our mathematical models. We will make use of earlier experiences [10] with recruiting and conducting experiments with native ASL signers. In past studies, we have asked participants to view the animations; then, we asked them 10-point Likert-scale subjective questions. Signers were asked to rate the grammaticality, naturalness, and understandability of the animations on these scales. We have also asked signers comprehension questions about the content of the animation – presenting answer choices in form of images to be circled. A native ASL signer had conducted our evaluations of ASL animations to ensure that a conversational ASL experimental environment is maintained during the study.

3. CURRENT PROGRESS

In the past year, we have set up the motion-capture equipment (with a body suit, cybergloves, overhead acoustic sensors for head tracking, high-definition cameras, eye-tracker). We have created manuals and protocols for equipment and annotation. We have also designed and evaluated a new calibration protocol for motion- capture gloves, which is designed to make the process more efficient and to be accessible for participants who are deaf and use ASL [17]. We have started recording ASL multi-sentence passages from subjects. Deaf ASL signers have been asked to perform ASL sentences at the lab while wearing the motion-capture equipment discussed in section 2 and while being videotaped. We have begun to linguistically annotate the data collected and use it to animate virtual human characters. Currently, we are planning how to mathematically analyze the data to build models of signs.

4. CONTRIBUTIONS

There are societal benefits from this research for the one half million deaf people in United States for whom ASL is their primary means of communication. Our research addresses challenges in creating computer-generated ASL animation. This technology can be used to make more information and accessible to deaf Americans. Our future goal is for content developers using our software to be able to script fluent sentences of ASL. This technology could also be used in software for automatically translating from English to ASL.

The corpus we create will include digitally recorded 3D movement data from signers using the motion-capture equipment, the video recording of signers performing ASL while wearing the motion capture equipment, and the linguistic annotation added by native ASL signers. This ASL data will be retained after our project and studied by the future researchers. Learning how to generate ASL signs whose movements are affected by the arrangement of entities in space will be a significant advancement for ASL linguistic understanding and animation generation.

REFERENCES

[1]C. Neidle, J. Kegl, D. MacLaughlin, B. Bahan and R. Lee. 2000. The Syntax of American Sign Language. Functional Categories & Hierarchical Structure. Cambridge, MA. MIT Press.

[2] Meier, Richard P. 1990. Person Deixis in American Sign Language. In: Susan D. Fischer & Patricia Siple (eds.), Theoretical Issues in Sign Language Research. Vol. 1: Linguistics. Chicago: University of Chicago Press, pp. 175-190.

[3]S. Liddell. 2003. Grammar, Gesture, and Meaning in American Sign Language. UK: Cambridge University Press.

[4] C. Padden. 1988. Interaction of Morphology and Syntax in American Sign Language. Outstanding Dissertations in Linguistics, Series IV. New York: Garland Press.

[5] K. Cormier. 2002. Grammaticalization of Indexic Signs: How American Sign Language Expresses Numerosity. Ph.D. Dissertation, University of Texas at Austin.

[6] N. Frishberg. 1975. Arbitrariness and Iconicity: Historical Change in American Sign Language. Language 51:676-710.

[7] S. Liddell. 1977. An Investigation into the Syntactic Structure of American Sign Language. Ph.D. Dissertation, University of California, San Diego.

[8] M. Huenerfauth. 2006. Generating American Sign Language Classifier Predicates For English-To-ASL Machine Translation. Doctoral Dissertation, Computer and Information Science, University of Pennsylvania.

[9] M. Huenerfauth, L. Zhou, E. Gu, J. Allbeck. 2008. Evaluation of American Sign Language Generation by Native ASL Signers. ACM Transactions on Accessible Computing. Volume 1,Number 1, Article 3.

[10] M. Huenerfauth. 2009. Improving Spatial Reference in American Sign Language Animation through Data Collection from Native ASL Signers. In Proceedings of UAHCI'09.

[11] R. Elliott, J. R. W. Glauert, J. R. Kennaway, I. Marshall, E. Safar. 2008. Linguistic Modelling and Language-Processing Technologies for Avatar-based sSgn Language Presentation, Universal Access in the Information Society, v.6 n.4, pp.375-391.

[12] R. Elliott, J.R.W. Glauert, J.R. Kennaway, I. Marshall. 2000. The Development of Language Processing Support for the ViSiCAST Project. ASSETS 2000, Arlington, VA, USA

[13] Vcom3D Sign Smith™. http://www.vcom3d.com/signsmith.php

[14] C. Rose, M. F. Cohen and B. Bodenheimer. 1998. Verbs and Adverbs: Multidimensional Motion Interpolation. In IEEE Computer Graphics and Applications, Vol. 18(5), pp. 32-40.

[15] P.-P. Sloan, C. F. Rose, and Michael F. Cohen. 2001. Shape by Example. In Proceedings of 2001 Symposium on Interactive 3D Graphics, pp. 135-144.

[16] S.I. Park, H.J. Shin, S.Y. Shin. 2002. On-line Locomotion Generation Based on Motion Blending, Proceedings of the ACM SIGGRAPH/Eurographics symposium on Computer animation, San Antonio, Texas, USA.

[17] P. Lu and M. Huenerfauth. 2009. Accessible Motion-Capture Glove Calibration Protocol for Recording Sign Language Data from Deaf Subjects. In Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2009), Pittsburgh, PA, USA.