# Analysis of Facial Expressions in American Sign Language

*Christian Vogler*

Gallaudet Research Institute
Gallaudet University
800 Florida Ave. NE
Washington, DC 20002, USA
Christian.Vogler@gallaudet.edu

*Siome Goldenstein*

Instituto de Computação
Universidade Estadual de Campinas
Caixa~Postal 6176
Campinas, SP Brazil
siome@ic.unicamp.br

## Abstract

In the age of speech and voice recognition technologies, sign language recognition is an essential part of ensuring equal access for deaf people. To date, sign language recognition research has mostly ignored facial expressions that arise as part of a natural sign language discourse, even though they carry important grammatical and prosodic information. One reason is that tracking the motion and dynamics of expressions in human faces from video is a hard task, especially with the high number of occlusions from the signers' hands.
In this paper, we present a 3D deformable model tracking system to address this problem. We apply it to sequences of native signers, taken from the National Center of Sign Language and Gesture Resources (NCSLGR), with a special emphasis on outlier rejection methods to handle occlusions. Our experiments validate the output of the face tracker against expert human annotations of the NCSLGR corpus, demonstrate the promise of our proposed face tracking framework for sign language data, and reveal that the tracking framework picks up properties that ideally complement human annotations for linguistic research.

## 1 Introduction

With the growing popularity of speech and voice recognition technologies, it is only a matter of time before they pose serious accessibility problems to deaf people who use signed languages as their primary mode of communication. Sign language recognition technologies are, therefore, an essential component of accessible human-computer interaction (HCI). Research into recognizing the manual components of sign languages has yielded promising results (Bauer and Kraiss, 2002; Vogler and Metaxas, 2004), but to date, all this research has virtually ignored the facial expressions that arise as part of a natural sign language discourse, even though they carry important grammatical and prosodic information (Neidle et al., 2000). In addition, linguistic research into signed languages can greatly benefit from face tracking and recognition systems, as they can assist with the tedious task of annotating data.
One reason why facial expressions have been ignored to date is that the challenges in incorporating them into a recognition system are so substantial. Tracking human faces from video is a very difficult problem, even more so because 3D information integral to signed languages needs to be recovered, such as head tilting and side-to-side movements. In addition, the subtle facial movements that arise in signed languages require a much greater degree of precision than many past tracking approaches have been able to provide. Another complicating factor is that the system must be able to track and recognize the facial expressions exactly as they would occur in natural settings. This requirement precludes many controlled laboratory conditions that would otherwise simplify the task, because they would alter the appearance of the signs. Foremost among them is a lack of control over the movements that the subjects exercise. As a result, we have to deal with extreme head movements from side to side, frequent self-occlusions of the face by the signer's hands, as well as possible partial obstructions of the signer's facial features by hair.
In this paper, we present a 3D deformable model tracking system and apply it to American Sign Language (ASL) sequences of native signers, taken from the National Center of Sign Language and Gesture Resources (NCSLGR) (Neidle and Sclaroff, 2002). With deformable models, we restrict the family of possible solutions. Instead of estimating the positions of every relevant feature of the face, we constrain the changes in the image to changes in the values of a 3D parameterized model that describes, in the case of face tracking, the subject's face.

These parameters model both the rigid motion (orientation and translation), as well as nonrigid deformations, such as eyebrow, lip and jaw movement (Section 3).
Deformable model tracking is an inductive procedure (Section 3.1). The basic underlying assumption is that we have the parameters that register the 3D model to the image in the previous frame. Then, based on changes in the new frame, the algorithm locally searches for the new configuration of the parameters that aligns the model to the new image. Since a human face has locally distinct photometric properties, we use various techniques to find features that define local correspondences between the images and the model: point trackers for image corners, edge trackers for image borders, and optical flow for areas that have texture. A statistical framework then merges these correspondences (Section 3.2).
In their basic formulation, deformable models are sensitive to outliers; that is, some of the image features may provide wildly inaccurate estimates and throw off the tracking process as a whole. Getting rid of the outliers is essential, especially in the case of sign language sequences, because self-occlusions of the face by the hands generate large numbers of them (Section 3.3).
Originally, these methods were all geared toward tracking the face from noisy images and had nothing to do with sign language. In this paper we show in experiments on the NCSLGR data (Section 4) that the same methods can be used to deal with uncertain information on the subject's face in sign language recognition applications — especially occlusions —, and that 3D deformable model tracking holds great promise for both sign language linguistic research and recognition in the context of HCI. In particular, the experiments validate the output of the face tracker (Section 4.1) against expert human annotations of the NCSLGR corpus (Section 4.2). In addition, plots of the extracted face trajectories exhibit several interesting properties that are not picked up in the annotations, showing that linguistic research and computer science research into signed languages complement each other (Section 4.3).

## 2 Related Work

There are many approaches to deformable model tracking, such as snakes (Kass et al., 1988), active shapes (Cootes and Taylor, 1995), active appearance models (Cootes et al., 2001), and active contours (Blake and Isard, 1999). Other methods include a local-global hybrid approach (Metaxas, 1996), and PCA (principal component analysis) decomposition (Blanz and Vetter, 1999). PCA decompositions are an extremely powerful way to fit, track and even recognize objects (Murase and Nayar, 1995; Blanz and Vetter, 1999; Pighin et al., 1999; Romdhani and Vetter, 2003; Dimitrijevic et al., 2004), at the cost of requiring a large database of examples. Additionally, a powerful deformable volumetric model has been used for fast face tracking (Tao and Huang, 2002) and subtle motion capture (Wen and Huang, 2003). Predictive filters (Goldenstein, 2004), such as the Kalman filter, can add reliability to these tracking methods (Goldenstein et al., 2004a).
In a broader sense, the approaches can be divided into two categories: the ones that use machine learning techniques to train a a model on a set of data before they can be put to work, and the ones that do not. The former category often requires carefully constructed training examples, which are not always available in preexisting data sets. The latter category is more flexible in this respect; however, it still requires that models are fitted to their respective starting frames, which can be a time-consuming task. A possible step toward automation of the fitting task consists of combining stereo with shading (Samaras et al., 2000), and using anthropometric databases to generate biometrically accurate models (DeCarlo et al., 1998).
The deformable model tracking framework that we present here requires no training. As a result, it is particularly suitable for the analysis of sign language data that were not collected with training in mind. Parameterizing muscle actuator groups (Essa. and Pentland, 1997) provides another way to recognize dynamic facial expressions without training. A contrasting approach to analyzing facial expressions as part of language recognition is based on training active appearance models (Canzler and Kraiss, 2004).
Recognizing facial expressions is just one part of a comprehensive sign language recognition framework. Recent work has focused on modeling the language in terms of its constituent parts (phonemes) (Bauer and Kraiss, 2002; Vogler and Metaxas, 2004) and first steps toward signer-independent recognition (Fang et al., 2001; Zieren and Kraiss, 2005).

## 3 Deformable Models

For sign language recognition and analysis, we need to track a moving face. Deformable models provide a way to reduce the dimensionality of the tracking problem. If we were to work with a free-form 3D mesh, we would somehow need to obtain the position of every single node on the mesh. In contrast, within a deformable model, the

position of every node on the mesh is specified through a function of a small set of parameters — the parameter vector **q**. The components of this vector can represent such things as the position and orientation of the model in 3D space, the degree of eyebrow raising, mouth aperture, and so on. In Figure 1, we can see how a few parameters that affect local regions of a base mesh can generate deformations in a 3D face model.

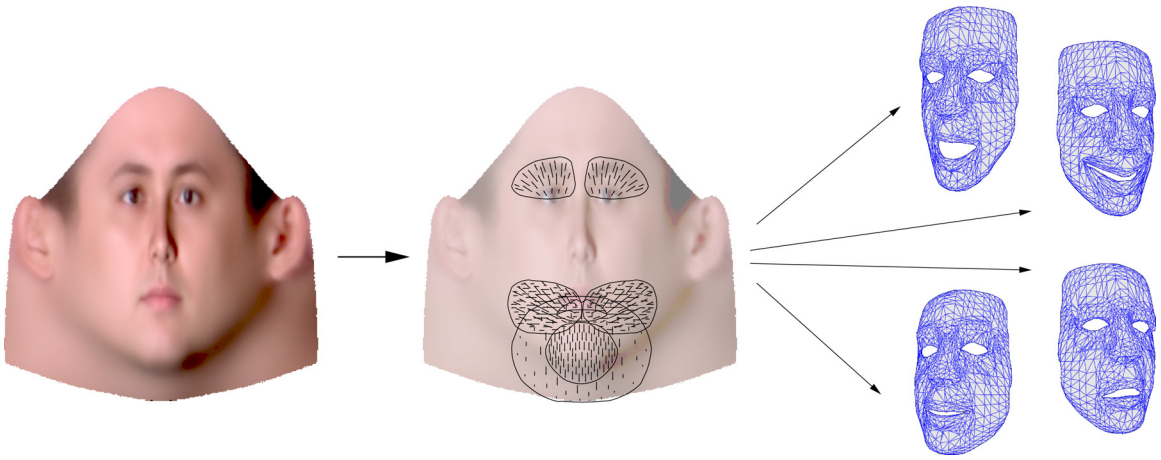

Figure 1: With a set of parameters, and localized areas of effect, we can create a 3D deformable face model.

## 3.1 Mathematics of Deformable Models

Mathematically, there are many ways to represent deformable models. A simple, yet powerful, approach consists of using a triangulated 3D mesh, with a function for each of its nodes. This function evaluates that node's position for any given value of the parameters, such as the eyebrows, face orientation, and the jaw opening. We use a directed acyclic graph representation (Goldenstein et al., 2003), and an adaptive resolution method (Goldenstein et al., 2004b) to construct this mesh.

Tracking is an inductive process. For the initial frame, we have the parameters that put the face in its proper place in the image, with the correct facial expressions. For every new frame, we need to adjust the parameters, such that they best describe the model for this frame. The basic idea is to select distinctive image features, such as edges and corners, and track them over time using standard 2D image processing methods. The result are (potentially inaccurate) positions of where the model nodes should be in the new frame. We then find the 2D displacements between the current model node positions — through a projection of the model onto the image — and the tracked features, and express them as vectors, called $\mathbf{f}_i$. With them we perform a local nonlinear optimization on the parameter vector **q**

$$\dot{\vec{q}} = \sum_i B_i^\top \vec{f_i}, \tag{1}$$

where $B_i$ is the Jacobian

$$B_i = \frac{dp_i}{d\vec{q}} = \begin{bmatrix} | & | & & | \\ \frac{\partial p_i}{\partial q_1} & \frac{\partial p_i}{\partial q_2} & \cdots & \frac{\partial p_i}{\partial q_n} \\ | & | & & | \end{bmatrix}, \tag{2}$$

to find the value of **q** that minimizes the 2D displacements. The simplest way to accomplish this task is through an iterative gradient descent method, using Equation 1 to find the appropriate direction of the next step (Goldenstein et al., 2003).

## 3.2 Statistical Extensions

One problem with using features from different image processing methods (such as corners, edges, optical flow) is that there is no straightforward way to integrate the information from different sources ("cues"). However, we proved in earlier work that the distribution of **q** can be properly approximated as a multivariate Gaussian (Goldenstein et al., 2003), by assigning regions of confidence of the 2D displacements $\mathbf{f}_i$ , and representing them as *affine forms* (Stolfi and Figueiredo, 1997), which are symmetric convex regions. This statistical representation of **q** results in a maximum likelihood approach to the fusion of the information from different cues (Goldenstein et al., 2003), and in a confidence measure of the distribution of the final model parameters, represented as a Gaussian probability distribution. Knowing this distribution is important for two reasons. First, it provides the groundwork for using a Kalman filter to stabilize the tracking (Goldenstein et al., 2004a), and second, it provides the necessary information for identifying outliers among the features.

## 3.3 Occlusion Handling via Outlier Rejection

As we mentioned before, in sign language, the hand frequently occludes the face (see Figure 2). This scenario has to be handled gracefully without ever causing catastrophic results, such as a loss of track. From a high-level point of view, an occlusion means that we possess no information on the regions of the face that are hidden, and hence any image features that are affected by these regions must be treated as suspect. Identifying these regions themselves is not easy. However, the suspect feature points typically express behavior that is inconsistent with the overall flow of the model; that is, they become outliers.

Since there is no algebraic model for the deformable model tracking problem, we cannot directly apply some of the traditional outlier rejection techniques, such as RANSAC (Fischler and Bolles, 1981). Instead, we project all 2D displacements into the parameter space, find a Gaussian approximation to their distribution, and reject the correspondences that are incompatible with this Gaussian model. Because standard maximum likelihood estimators for the mean and covariance matrix of the Gaussian model may mask the outliers, we use a robust estimator instead. This estimator is called the MCD (minimum covariance determinant) (Rousseeuw and Driessen, 1999). Conceptually, it finds the subset of features that minimize the spread (i.e., the covariance of the Gaussian model), and applying it to the deformable model tracking framework has shown to be very effective (Goldenstein et al., 2005). The breakdown point of MCD depends on the number of features that are chosen as representative of the Gaussian model. This parameter is configurable by the application, so we can tailor it to the needs of sign language tracking, depending on how many outliers we expect.

# 4 Experiments and Discussion

In the past, we validated the efficacy of statistical cue integration and outlier rejection within the tracking framework extensively (Goldenstein et al., 2003; Goldenstein et al., 2005). In addition, the suitability of the framework for recognizing facial expressions was demonstrated in stress detection experiments (Dinges et al., 2005), where the computer achieved a detection accuracy of 75%-88% across 60 subjects in double-blind experiments. Because of the high speed of the hand and facial movements, as well as the frequent self occlusions that we discussed earlier, tracking sign language sequences is substantially harder than the previous stress detection task, and even harder than tracking corrupted videos. To make matters worse, validation of sign language sequences is a hard task itself, because we cannot use any kind of intrusive system to measure the subject's head and facial movements. Doing so would interfere with the subject's production of signs and substantially alter the appearance of the utterances, thus violating the prerequisite of having natural utterances from native signers.

As a result, the best that we can do is comparing the output of the tracking process with the annotations provided by expert human transcribers. To this end, we tracked sequences taken from the NCSLGR corpus. This corpus consists of a large number of videos of native signers taken in a carefully constructed laboratory setting to ensure that these sequences were representative of ASL, along with annotations. One video that we tracked in particular shows a story about a close call with a deer on a highway (Figure 2). This sequence, just like the others, was shot in full color at a resolution of 640×480 at 60 frames per second. The annotations consisted of information on head rotation and tilts, along with eye blink information and glosses[1] .

[1]Glosses are representations of the signs by their closest English equivalent in all capital letters

Figure 2: Excerpt from the tracked sign language sequence ("Close Call," taken from NCSLGR), showing the sign for "REMEMBER." Occlusions of the face, as shown in the middle image, are frequent and greatly complicate the tracking problem.

## 4.1 Tracking Results

To test the hypothesis that our methods to deal with noisy images also help with sign language tracking, especially self-occlusions, we tracked video clips from two different subjects. For each subject, the shape of the generic 3D face model needs to be fitted to the characteristics of that person's face. We performed the fitting in a semi-automated manner by choosing a frame with a frontal view, and selecting correspondences between the nodes on the model and the image. We then integrated the system according to Equation 1 from Section 3.1, with the shape characteristics functioning as the model parameters, and the model-image correspondences acting as the $\mathbf{f}_i$ in this equation. The integration process subsequently yielded a version of the model that was adjusted to the particular subject's face. Although a laborious task, it needs to be done only once per subject. Afterward, the model can be applied to the same subject shown in arbitrary poses on arbitrary video clips.

The face models consisted of 1101 nodes and 2000 triangles. Tracking the video sequences with these models, using gradient descent with 600 iterations per frame, took 0.3 seconds per frame on an AMD Opteron 246 workstation running a 64-bit version of Linux. This number includes time spent on image processing, which still contains much potential for performance optimization.

Figure 3 shows some representative tracking results. Note that particularly in the bottom example, there is a strong self-occlusion of the face, yet it has no effect on the alignment of the face model. To see why this is so, we need to look at the behavior of the outlier rejector, which we described briefly in Section 3.3. In the following discussion, we focus on the effect of outlier rejection on point features, but a similar argument also applies to other cues, such as edges.

Figure 4 shows three snapshots of the point tracker, with the tracked features divided into acceptable points and outliers, at the frame immediately before the occlusion, during the occlusion, and immediately afterward. Many points are already dropped by the point tracking algorithm, because it cannot find any acceptable matches for them during the occlusion, but some points "survive" and are pulled along the contour of the hand (Figure 4, center). These points then stay on the hand contour and are pulled downward (Figure 4, right). They are now gross outliers, and if they were not recognized as such, they would pull the face model downward, thereby destroying the alignment between the model and the video.

Intuitively, the effect of robust parameter space outlier rejection is to select a relatively small subset of features that forms the most likely hypothesis for the trajectory of the model parameters, and discarding any point that does not support this hypothesis. Under these circumstances, the points on the hand contour in the center and right parts of the figure are clear outliers, because they induce a downward motion that is incompatible with the general sideways and tilting motion of the subject's face, as well as other points on the face that have provided unreliable information.

Dealing with partial occlusions in this manner works well, as long as the percentage of gross outliers does not rise above the breakdown point of the MCD estimator. We showed in earlier work that the optimum trade-off between robustness and efficiency mandates a breakdown point of 75–80% (Goldenstein et al., 2005), which roughly corresponds to a maximum allowable outlier percentage of 25%[2] . Unfortunately, in the case of a near-total occlusion of the face, this condition no longer holds, because nearly all of the salient image features are obstructed, and the system is prone to losing track. This problem will need to be addressed by future work; possibly by

[2]The exact number is dependent on the dimension of the parameter space; higher dimensions reduce the percentage.

identifying the occluding region and discarding anything that falls into it. The parameter space outlier rejection method then may be able to get rid of the remaining outliers that slip through the cracks.

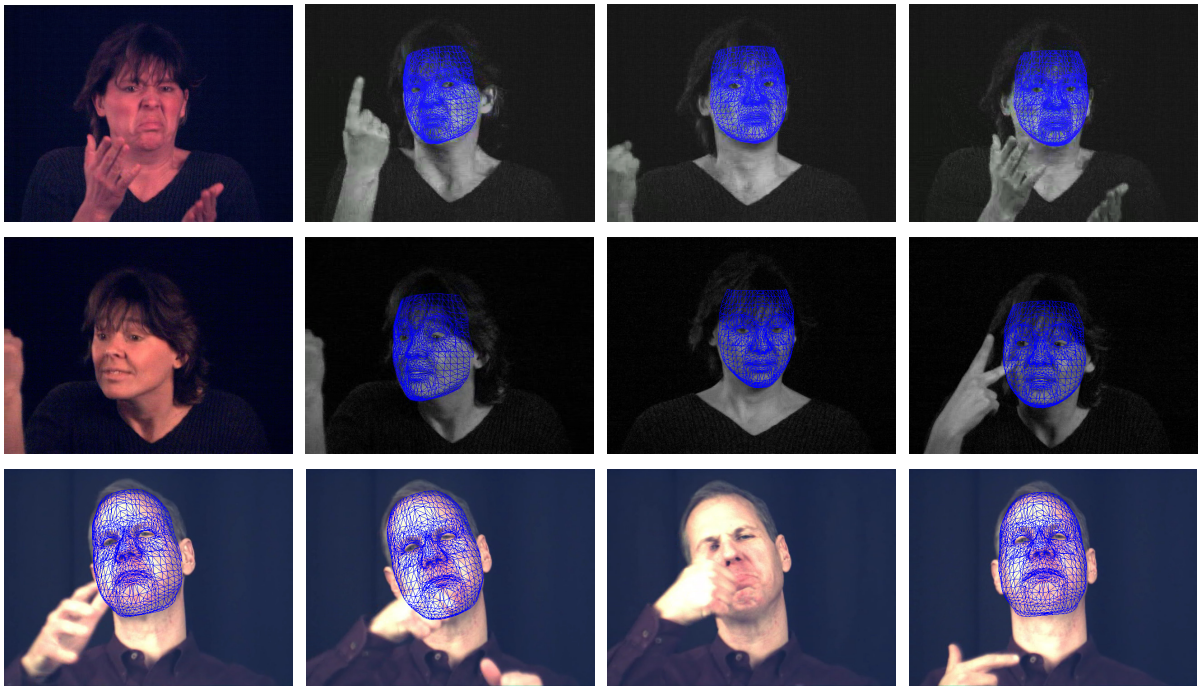

Figure 3: Three tracking examples from two different native signers with the tracked 3D model overlaid. The self-occlusions of the face by the hands have no effect on the tracking.

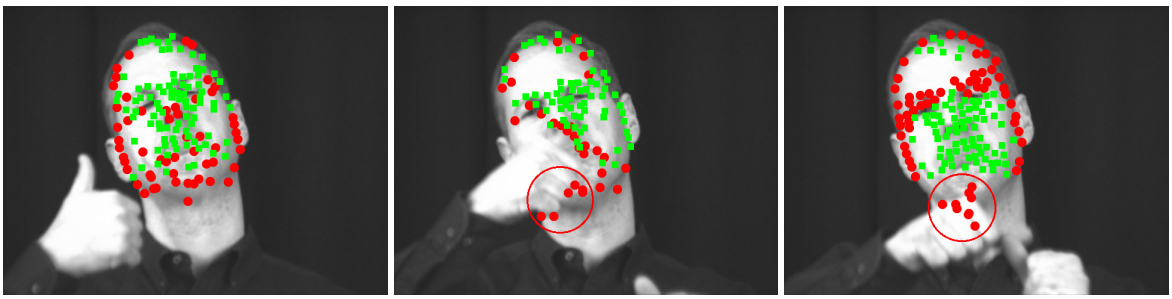

Figure 4: Outlier rejection behavior during occlusion. Outliers are plotted as red circles, whereas valid points are plotted as green boxes. Left: Immediately before occlusion. Center: During occlusion. Right: Immediately after occlusion. Some points are erroneously "acquired" by the hand, but rejected as outliers (large circle center and right).

## 4.2 Data Analysis and Comparison with Corpus Annotations

The sequence of parameter vectors **q**, which we obtain through the tracking process, could conceivably be used as features for a recognition algorithm, but to compare it to the expert human annotations, we first need to convert the head rotation information contained in this vector to Euler angles. In Euler nomenclature, the side-to-side head movement corresponds to the heading (yaw), the forward-backward tile corresponds to the pitch, and the side-to-side tilt corresponds to the roll (cf. the mention of the NCSLGR corpus and annotations above). A representative plot of these three head angles, in comparison to the annotations, is shown in Figure 5.

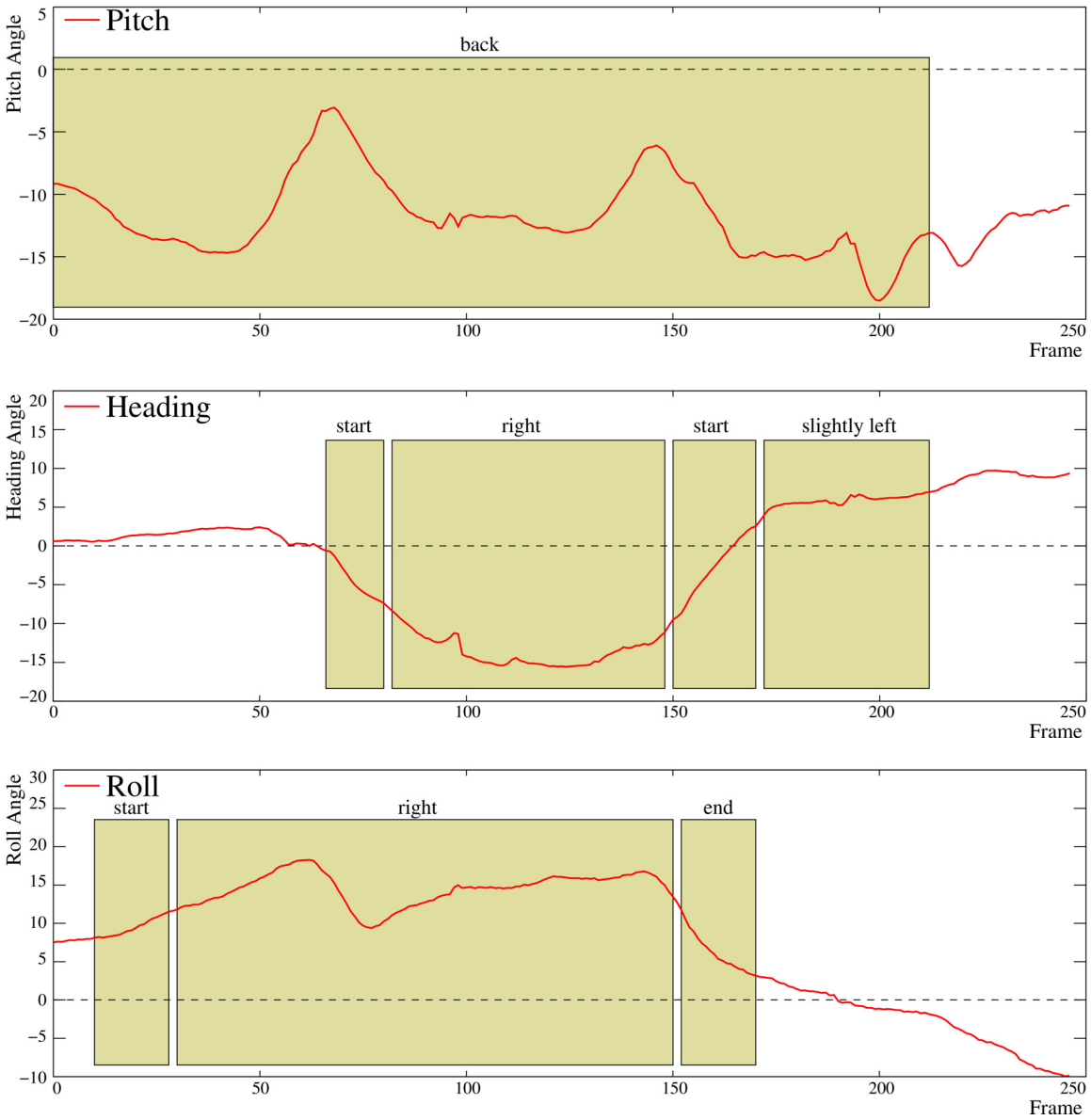


Figure 5: Plot of the head's Euler angles compared to the expert human annotations of the NCSLGR (labeled boxes) for the beginning of the "Close Call" sequence. "Start" and "end" correspond to anticipatory head movements. In the top figure, "right" and "slightly left" indicate the respective head turns; in the middle figure, "back" indicates a head tilt backward; and in the bottom figure, "right" indicates a head tilt to the right.

These plots exhibit qualitative agreement between the tracked head parameters and the human annotations. There are slight discrepancies in the boundaries, but these are difficult for humans to catch accurately. In general, the human annotations have quantitative counterparts in the plots, such as positive versus negative angles, and magnitude of angles, which holds great promise for future recognition work. Anticipatory head movements, characterized by "start" and "end," are an exception, because they are impossible to quantify from the angles. In fact, with respect to head movements these two labels are highly subjective for humans (although much less so for other types of movements) [3] . Essentially, they mark the head moving into position, immediately before and after a movement starts being perceived as significant by the human annotators.

[3]Carol Neidle, personal communication.

Also of note is that there is remarkably little jitter in the plots, considering that the extracted angles were not smoothed in any way. A large part of this absence of jitter is due to the outlier rejection method, which, as previously stated, eliminates features that do not coincide with the overall tracking hypothesis. The plots, however, exhibit clear peaks and valleys in areas that are perceived as a single entity by the human annotators, such as the two valleys in the head pitch plot. These highlight how human annotators make decisions about what to annotate in particular, and what they concentrate on.

### 4.3 Implications for Linguistic and Recognition Research

The differences between the human annotations and the plots lead to several interesting consequences and problems for both linguistic and recognition research. On the linguistics side, fundamentally, the human annotators have to deal with a small discrete set of labels to describe continuous phenomena, such as the degree of head tilt, which drops information on the exact kinematics. As a consequence, the annotations often focus on the markings that the annotators perceive as important for their particular research project, whereas in reality there are complex interactions with the movements accompanying lexical items (i.e. individual signs).

Although humans could attempt to pick up the finer details of the movements, and to transcribe them, as well, doing so would require an inordinate amount of time and effort, especially because humans have a hard time judging 3D angles on 2D video. Moreover, doing so would not solve the basic problem that the annotators have only a discrete set of labels at their disposal. Thus, for a linguistics researcher wishing to investigate the interplay of overlapping head movements, the tracking framework can provide an invaluable research tool. The main hurdle to providing this tool lies in devising a suitable user interface. At present, effective 3D tracking requires considerable computer science expertise, and it is not clear how to overcome this hurdle in the short term. For the time being, collaboration between sign language linguists and computer scientists remains the most promising avenue of action.

On the computer science side, the clear correspondence between the head angles and the head rotation and tilt labels holds great promise for future systems that recognize the nonmanual markings of signed languages. It is important not to get hung up on the question of identifying the anticipatory head movements (i.e., the portions labeled with “start” and “end”), which only exist to provide the human annotators with a greater choice of discrete labels. Rather, the first step for a recognition system should be the segmentation of the head movements into meaningful parts. The fluctuations in the angles may complicate this task. For instance, given the task of extracting the segment for the head tilt backward in Figure 5, in accordance with the human annotations, if a recognizer naively thresholded the angles to label the degree of tilt, it would incorrectly interpret the sequence as alternating between slight and full tilts backward, whereas conceptually, it consists of a single, long full tilt. Clearly, a recognizer will need to look at the global behavior of the head, not only localized angles, to overcome this problem.

## 5 Conclusions and Outlook

The methods that make face tracking robust against noisy images carry over well to meet the demands of face tracking in the context of sign language analysis and recognition. Outlier rejection, in particular, makes the tracking framework resistant to partial face occlusions, as long as enough valid image features remain to keep the percentage of outliers below the breakdown point of robust statistical estimators. Total occlusions, however, require a different approach, and are a high priority for future research. The goal will not necessarily be accurate tracking during full occlusions, but rather graceful recovery from such events.

This tracking framework can be used as a basis for assisting sign language linguistic research. The ability to extract and plot the trajectories of various facial parameters may well prove invaluable for research into sign language prosody. In addition, it could help with the verification of annotations.

Moreover, tracking the human face is an important first step into augmenting a sign language recognition system with facial expressions. The next step is tracking a large number of sequences, so that the extracted parameters can be used in machine learning algorithms.

### Acknowledgments

The research in this paper was supported by NSF CNS-0427267, research scientist funds by the Gallaudet Research Institute, NASA Cooperative Agreements 9-58 with the National Space Biomedical Research Institute, FAEPEX-Unicamp, and FAPESP. Carol Neidle provided helpful advice and discussion on the NCSLGR annotations vis-a-vis

the tracking results. Lana Cook and Ben Bahan were the subjects in the video sequences that we discussed in this paper.

## References

Bauer, B. and Kraiss, K.-F. (2002). Video-based sign recognition using self-organizing subunits. In *International Conference on Pattern Recognition*.

Blake, A. and Isard, M. (1999). *Active Contours : The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*. Springer Verlag.

Blanz, V. and Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *SIGGRAPH*, pages 187–194.

Canzler, U. and Kraiss, K.-F. (2004). Person-adaptive facial feature analysis for an advanced wheelchair user-interface. In Drews, P., editor, *Conference on Mechatronics & Robotics*, volume 3, pages 871–876. Sascha Eysoldt Verlag.

Cootes, T., Edwards, G., and Taylor, C. (2001). Active appearance models. *IEEE PAMI*, 23(6):681–685.

Cootes, T. and Taylor, C. (1995). Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59.

DeCarlo, D., Metaxas, D., and Stone, M. (1998). An anthropometric face model using variational techniques. In *Proceedings of the SIGGRAPH*, pages 67–74.

Dimitrijevic, M., Ilic, S., and Fua, P. (2004). Accurate face models from uncalibrated and ill-lit video sequences. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 1034–1041.

Dinges, D., Rider, R., Dorrian, J., Rogers, E. M. N., Cizman, Z., Goldenstein, S., Vogler, C., Venkataraman, S., and Metaxas, D. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, Space and Environmental Medicine (in press)*.

Essa., I. and Pentland, A. (1997). Coding, analysis, interpretation and recognition of facial expressions. *IEEE PAMI*, 19(7).

Fang, G., Gao, W., Chen, X., Wang, C., and Ma, J. (2001). Signer-independent continuous sign language recognition based on SRN/HMM. In Wachsmuth, I. and Sowa, T., editors, *Gesture and Sign Language in Human-Computer Interaction. International Gesture Workshop*, volume 2298 of *Lecture Notes in Artificial Intelligence*, pages 76–85. Springer.

Fischler, M. and Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Goldenstein, S. (2004). A gentle introduction to predictive filters. *Revista de Informatica Teórica e Aplicada*, XI(1):61–89.

Goldenstein, S., Vogler, C., and Metaxas, D. (2003). Statistical Cue Integration in DAG Deformable Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):801–813.

Goldenstein, S., Vogler, C., and Metaxas, D. (2004a). 3D facial tracking from corrupted movie sequences. In *Proceedings of IEEE Computer Vision and Pattern Recognition*.

Goldenstein, S., Vogler, C., Stolfi, J., Pavlovic, V., and Metaxas, D. (2005). Outlier rejection in high-dimensional deformable models. *Image and Vision Computing, to appear*.

Goldenstein, S., Vogler, C., and Velho, L. (2004b). Adaptive deformable models. In *Proceedings of SIBGRAPI*, pages 380–387.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active Contour Models. *International Journal of Computer Vision*, 1:321–331.

Metaxas, D. (1996). *Physics-based Deformable Models: Applications to Computer Vision, Graphics and Medical Imaging*. Kluwer Academic Publishers.

Murase, H. and Nayar, S. K. (1995). Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14:5–24.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., and Lee, R. G. (2000). *The syntax of American Sign Language*. Language, Speech, and Communication. MIT Press, Cambridge, Massachusetts.

Neidle, C. and Sclaroff, S. (2002). Data collected at the National Center for Sign Language and Gesture Resources, Boston University, under the supervision of C. Neidle and S. Sclaroff. Available online at `http://www.bu.edu/asllrp/ncslgr.html`.

Pighin, F., Szeliski, R., and Salesin, D. (1999). Resynthesizing facial animation through 3D model-based tracking. In *Proceedings of International Conference of Computer Vision*, pages 143–150.

Romdhani, A. and Vetter, T. (2003). Efficient, robust and accurate fitting of a 3D morphable model. In *Proceedings of International Conference of Computer Vision*, pages 59–66.

Rousseeuw, P. J. and Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223.

Samaras, D., Metaxas, D., Fua, P., and Leclerc, Y. (2000). Variable albedo surface reconstruction from stereo and shape from shading. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 480–487.

Stolfi, J. and Figueiredo, L. (1997). *Self-Validated Numerical Methods and Applications*. $21^{O}$ Colóquio Brasileiro de Matemática, IMPA.

Tao, H. and Huang, T. (2002). Visual Estimation and Compression of Facial Motion Parameters: Elements of a 3D Model-Based Video Coding System. *International Journal of Computer Vision*, 50(2):111–125.

Vogler, C. and Metaxas, D. (2004). Handshapes and movements: Multiple-channel ASL recognition. In et al., G. V., editor, *Proceedings of the Gesture Workshop*, volume 2915 of *Lecture Notes in Artificial Intelligence*, pages 247–258. Springer.

Wen, Z. and Huang, T. (2003). Capturing subtle facial motions in 3D face tracking. In *Proceedings of International Conference of Computer Vision*, pages 1343–1350.

Zieren, J. and Kraiss, K.-F. (2005). Robust person-independent visual sign language recognition. In *Proceedings of the 2nd Iberian Conference on Pattern Recognition and Image Analysis IbPRIA*, Volume Lecture Notes in Computer Science.