The Best of Both Worlds: Combining 3D Deformable Models with Active Shape Models

Christian Vogler
Gallaudet University
Christian.Vogler@gallaudet.edu

Zhiguo Li
Rutgers University
zhli@cs.rutgers.edu

Atul Kanaujia
Rutgers University
kanaujia@cs.rutgers.edu

Siome Goldenstein
Universidade Estadual de Campinas
siome@ic.unicamp.br

Dimitris Metaxas
Rutgers University
dnm@cs.rutgers.edu

Abstract

Reliable 3D tracking is still a difficult task. Most parametrized 3D deformable models rely on the accurate extraction of image features for updating their parameters, and are prone to failures when the underlying feature distribution assumptions are invalid. Active Shape Models (ASMs), on the other hand, are based on learning, and thus require fewer reliable local image features than parametrized 3D models, but fail easily when they encounter a situation for which they were not trained.

In this paper, we develop an integrated framework that combines the strengths of both 3D deformable models and ASMs. The 3D model governs the overall shape, orientation and location, and provides the basis for statistical inference on both the image features and the parameters. The ASMs, in contrast, provide the majority of reliable 2D image features over time, and aid in recovering from drift and total occlusions. The framework dynamically selects among different ASMs to compensate for large viewpoint changes due to head rotations. This integration allows the robust tracking of faces and the estimation of both their rigid and nonrigid motions. We demonstrate the strength of the framework in experiments that include automated 3D model fitting and facial expression tracking for a variety of applications, including sign language.

1. Introduction

Reliable tracking of videos based on 3D deformable models is a difficult task. Parametrized 3D deformable models rely on extracting image features that are subsequently used to obtain good estimates for each of the model's parameters. Most existing methods for image feature extraction, in turn, are based on deterministic feature tracking (e.g. KLT) or tracking feature distributions [12, 2, 9]. However, these methods are not based on learning and perform local computations for the tracking of each feature. If the assumptions that underlie the feature extraction methods are invalid — because of changes in illumination, noise, occlusions, and so on — the feature correspondences from frame to frame are not computed correctly. Such errors accumulate, and cause the 3D model to drift over time, which results in incorrect tracking of the video.

Active Shape Models (ASMs), on the other hand, yield good estimates without requiring as many reliable local image features as parametrized 3D models. They are based on learning, and tend to give more reliable results than feature trackers for poses that match closely the ones on which they have been trained. They also are less vulnerable to drifting, because they locate the features directly in the image. This latter ability is invaluable in recovering from occlusions or other situations that would cause a non-learning based system to lose track. Conversely, ASMs can fail spectacularly when they encounter a situation for which they were not trained. In addition, ASMs are linear models in principle, whereas face shape variations lie in a nonlinear space. However, switching among multiple linear ASMs can approximate this nonlinear space.

Parametrized 3D deformable models allow us to apply unsupervised techniques on the data that drive the tracking system, such as parameter-space outlier rejection [19], occlusion handling [8], and statistical inference to feed a Bayesian filter, such as the Kalman filter [10]. Moreover, the information provided by a 3D deformable model is a relatively low-dimensional parameter vector, where the information describing motion and deformation is already parsed in a meaningful way — making recognition and analysis considerably easier.

In this paper, we bring together the best of both worlds. We develop a framework for robust 3D face tracking that

978-1-4244-1631-8/07/$25.00 ©2007 IEEE

combines the strengths of both 3D deformable models and ASMs. In particular we use ASMs to track reliably 2D image features in a video sequence. In order to track large rotations we train multiple ASMs that correspond to frontal and several side head poses. The 2D features from the ASMs are then used for parameter estimation of the 3D deformable model. In addition we stochastically couple 3D deformable models and ASMs by using the 3D pose of the deformable model both to switch among the different ASMs, and to deal with occlusions. This integration allows robust tracking of faces and the estimation of both their rigid and nonrigid motions. We demonstrate the strength of the framework in experiments that include automated 3D model fitting and facial expression tracking across a variety of applications, including American Sign Language.

2. Related work

Linear vector spaces [1] have been very popular in modeling shape and texture. It is very convenient to apply learning techniques on this representation. Analysis by synthesis together with linear vector space representation is frequently used to solve vision and graphics problems. Examples include Active Shape Models [5], Active Appearance Models [17], hybrid 2D/3D Active Appearance Models [21], Eigenfaces [14], multidimensional morphable models [13], model-based structure from motion [7], and 3D morphable models [3, 4]. Many of these approaches have been applied to face modeling.

Blanz *et al*. [3] learned 3D shape and texture models from a set of laser scan data. Although it this approach works well, it is not suitable for real-time 3D face tracking, because it has too many parameters. In contrast, we build 3D deformable face models with relatively few parameters that encompass our prior knowledge about the range of possible facial expressions, instead of learning the 3D face shape and texture model from 3D data. In this sense, our work is an improvement over [6].

3. Overview of 3D Deformable Models

Three-dimensional deformable models belong to the general family of parametrized models, where a small numbers of parameters, denoted by the vector $\vec{q}$, control the shape, orientation, and position of the model [6, 15, 16, 20, 19]. For any given point p_i on the model surface, its 3D position $\vec{p}_{3d,i}$ is determined by a function defined over the parameter vector space:

$$\vec{p}_{3d,i} = F_i(\vec{q}). \tag{1}$$

Given a projection function, such as the one defined by a perspective camera, the 2D position of each model point is

$$\vec{p}_{2d,i} = Proj_i\left(\vec{p}_{3d,i}\right), \tag{2}$$

where $Proj_i$ is the projection function for the given point p_i. Conceptually, tracking a 3D deformable model consists of adjusting $\vec{q}$ such that the projected model points $\vec{p}_{2d,i}$ best match the image being tracked.

Traditionally, the first step in tracking consists of finding correspondences between the tracked image and the model points via standard computer vision techniques, such as point tracking, edge detection, and optical flow. The displacements between the actual projected model points $\vec{p}_{2d,i}$ and the identified correspondences are called *image forces*, and denoted by $\vec{f}_{2d,i}$. Adjusting $\vec{q}$ is done via gradient-descent optimization. To this end, the image forces projected into parameter space, resulting in *generalized forces* $\vec{f}_{g,i}$ that act on $\vec{q}$:

$$\vec{f}_{g,i} = \mathbf{B}_i^\top \vec{f}_{2d,i}, \tag{3}$$

where

$$\mathbf{B}_i^\top = \frac{\partial Proj_i}{\partial \vec{p}_{3d,i}} \frac{\partial F_i}{\partial \vec{q}} \tag{4}$$

is the projected Jacobian of the model point p_i with respect to the parameter vector. The generalized forces themselves are summed up into a single quantity

$$\vec{f}_g = \sum_i \vec{f}_{g,i}. \tag{5}$$

We then use Euler integration to solve the dynamical system

$$\dot{\vec{q}} = \mathbf{K}\vec{q} + \vec{f}_g, \tag{6}$$

where $\mathbf{K}$ is a stiffness matrix.

Goldenstein *et al*. showed in [9] that the image forces $\vec{f}_{2d,i}$ and generalized forces $\vec{f}_g$ in these equations can be replaced with affine forms that represent probability distributions, and furthermore that with sufficiently many image forces, the generalized force $\vec{f}_g$ converges to a Gaussian distribution. This result allows us to apply statistical inference at both Equations 3 and 5, upon which we elaborate in Section 5.3. This kind of statistical inference would not be possible if we used ASMs alone, without 3D models, and contributes substantially to the robustness of the combined 3D model and ASM tracking.

4. Overview of Active Shape Models

Active Shape Models (ASMs) [5] are landmark-based models that attempt to learn a statistical distribution over variations in shapes for a given class of objects. Shape statistics derived from training images are used to constrain the ASM search. An ASM consists of two parts: a global shape model and a set of local profile models. The global shape model captures shape variations, whereas the local profile models capture the grayscale levels around each

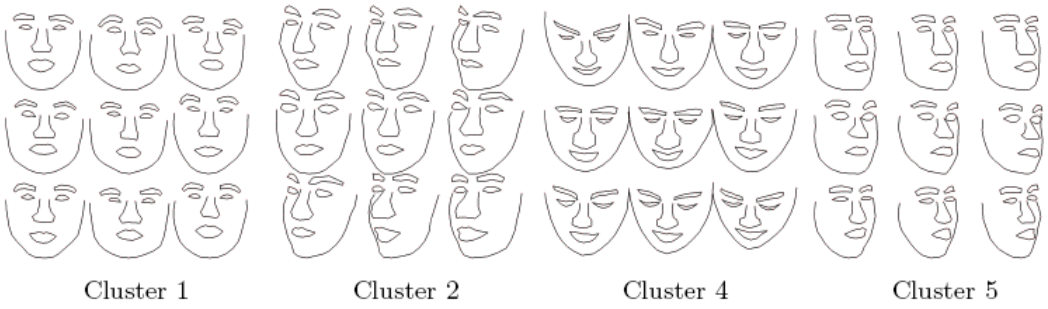

Figure 1. Example of multi-view ASM shapes

Figure 2. Example of multi-view ASM search results

landmark point and are used for selecting the best candidate image points.

Suppose that we have m aligned training face images using Procrustes Analysis [11], and for each training image we have n landmark points. We can represent the training images as a set of m $2n$-dimensional vectors $\vec{x}_j \in \{\vec{x}_1, \ldots, \vec{x}_m\}$. The unbiased covariance matrix for m training images is

$$\mathbf{\Lambda} = \frac{1}{m-1} \sum_{i=1}^{m} \left(\vec{x}_i - \bar{\vec{x}}\right) \left(\vec{x}_i - \bar{\vec{x}}\right)^{\top}. \tag{7}$$

The eigenvectors $\mathbf{\Psi}$ of the covariance matrix span a linear subspace where valid face shapes lie in some particular region of that space.

$$\mathbf{\Psi} = \left[\vec{\psi}_0, \vec{\psi}_1, \ldots, \vec{\psi}_{2n-1}\right] \tag{8}$$

The eigenvectors corresponding to the first few largest eigenvalues represent the major modes of shape variations. Any valid face shape can be represented as a linear combination of the basis vectors

$$\begin{aligned} \vec{x} &= \bar{\vec{x}} + \delta\vec{x} && (9) \\ &= \bar{\vec{x}} + \sum_{i=0}^{n-1} \left(b_{2i}\vec{\psi}_{2i} + b_{2i+1}\vec{\psi}_{2i+1}\right), && (10) \\ &= \bar{\vec{x}} + \mathbf{\Psi}\delta\vec{b} && (11) \end{aligned}$$

where $\delta\vec{b}$ is the coordinate difference between the average shape $\bar{\vec{x}}$ and the current shape $\vec{x}$. During the ASM search, we project the candidate results into this space and let the final ASM search result be the nearest valid face shape. Given a set of shape variations, the shape parameters can be estimated using linear least squares as:

$$\delta\vec{b} = (\Psi^T\Psi)^{-1}\Psi^T\delta\vec{x} \tag{12}$$

which, as Ψ is orthonormal, simplifies to:

$$\delta\vec{b} = \Psi^T\delta\vec{x}. \tag{13}$$

ASMs are viewpoint-dependent and, in principle, model only linear shape variations. For example, a single ASM is able to cope with shape variations from only a narrow range of face poses (turning and nodding of roughly 20 degree). We overcome this limitation by learning multiple ASM models for different viewpoints (Figure 1). We switch among them during tracking based on the 3D deformable model estimates of the head pose. Under these circumstances, the 2D ASM search works very well, as shown in Figure 2.

5. ASM and 3D Deformable Model Integration

3D deformable models and 2D ASMs complement each other. Although they are both parametrized models, with $\vec{q}$ taking on the role of the parameters in the case of 3D deformable models (Equation 1), and $\vec{b}$ taking on this role in the case of 2D ASMs (Equation 11), they function in very different ways.

The 3D deformable face model parametrization is crafted by the designer to meet the needs of the application. Each parameter stands for a specific high-level model action, such as eyebrow movements, lip parting, and so on. This representation also facilitates retargeting the application to different users with different face shapes, without having to change the interpretation of the parameters.

Human-designed parameters, however, are an idealization and thus an approximation to the actual deformations happening on a video. If a tracking cue, such as a point tracker, depends on being initialized with accurate positions from the model on every frame, any inaccuracies lead to drift over time as approximation errors cascade.

The parametrization of ASMs, on the other hand, is constructed from training data. Although it has no discernible high-level semantics and is not easily retargetable, it allows good correspondences between the model and images sufficiently similar to training examples. In addition, the ASM search matches shapes, instead of tracking iteratively from frame to frame, which allows it to locate features, even if the ASM lost track in a previous frame. However, the search is susceptible to getting stuck in local minima if the starting position is too far from the target; it cannot handle large head rotations; and it tends to yield contorted facial expressions when the tracked image does not fit the training data.

These considerations show that 3D deformable models and ASMs can compensate for each other's weaknesses: ASMs help with correcting drift and locating facial features, whereas the 3D model helps with a good initial position for the ASM, selects the correct ASM depending on the head pose, and provides mechanisms for statistical inference.

5.1. Overview of Combined Tracking Algorithm

The premise of the tracking algorithm, shown in Algorithm 1, is that the ASM is initialized with the current shape and position of the 3D model and then is tracked to the next frame, so as to establish the image displacements $\vec{f}_{2d,i}$ for Equation 5. The 3D model parameters are then adjusted to match the tracked ASM as closely as possible, and the cycle starts anew.

To establish the displacements between the ASM points $\vec{x}_i$ and 3D deformable model points $\vec{p}_{2d,i}$, the system needs to know which ASM points correspond to which 3D model points. These correspondences need to be defined manually, once per 3D model topology, and are shown in Figure 3.

We now give a detailed explanation of Algorithm 1.

5.2. Explanation of the Tracking Algorithm

Step 1 matches the 3D model to the initial image. It is based on establishing correspondences between selected 3D model points and the image. If we have a well-trained ASM for the person to whom we fit the 3D model, this step can be automated (see also Section 6.1). Otherwise, the correspondences need to be defined manually. Once they are defined, we integrate Equation 6 via 3 on the model's shape, position and orientation parameters, which results in an adaptation of the 3D model to the image and person being tracked.

Step 3 selects the best ASM given the current model orientation. Recall that ASMs are a view-based method and thus sensitive to changes in the subject's orientation. We train multiple ASMs for different view angles — left, front, right, down, and up —, which are selected based on the 3D model's head pose estimate.

Algorithm 1 Tracking algorithm for 3D models coupled with ASMs

1: Fit 3D model to initial frame
2: **while** there are frames to track **do**
3: Select ASM asm based on 3D model orientation
4: $\vec{x}_{asm} \leftarrow \vec{p}_{2d}$ {Equations 2, 11: Update ASM points from 3D model}
5: Adjust $\vec{b}_{asm}$ and $\vec{x}_{asm}$ to return ASM to valid state {Equations 11, 13}
6: Based on ASM intensity profile model, search next frame for new $\vec{b}_{asm}$, $\vec{x}_{asm}$ {Track ASM}
7: $\vec{f}_{2d,i} \leftarrow \vec{p}_{2d,i} - \vec{x}_{asm,i}$ {2D displacements between 3D model and tracked ASM}
8: Reject bad $\vec{f}_{2d,i}$ via statistical inference {Compensates for jumpy ASMs}
9: $\vec{f}_{g,i} \leftarrow \mathbf{B}_i^\top \vec{f}_{2d,i}$ {Equations 3, 4}
10: Reject bad $\vec{f}_{g,i}$ via statistical inference {Compensates for occlusions, bad correspondences}
11: Integrate $\dot{\vec{q}} = \mathbf{K}\vec{q} + \sum_i \vec{f}_{g,i}$ {Equations 5, 6}
12: **end while**

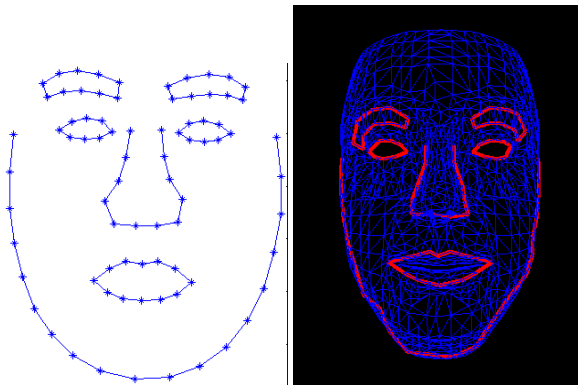

Figure 3. ASM to 3D model correspondences. The thick red lines on the right show the 3D model nodes that correspond to the ASM nodes on the left.

Steps 4 and 5 provide the initial guess for the ASM search described in Section 4 (Figure 4 left). Step 5 is necessary, because the new positions $\vec{x}_i$ are unlikely to fit the constraints of the eigenspace $\mathbf{\Psi}$, so they need to be projected into $\mathbf{\Psi}$. This projection obtains the new parameter vector $\vec{b}$ that makes the shape fit as closely as possible to the initial guess (Figure 4 center). The ASM search then finds the matching image features (Figure 4 right).

Step 6 tracks the ASM to the new frame, based on the initial guess from steps 4 and 5. This step results in new positions for the ASM's nodes $\vec{x}_i$ that are consistent with the eigenspace. The displacement between these ASM nodes and the projected 3D model nodes constitutes the image

Figure 4. 3D model-guided ASM initialization. Left to right: Tracked model position from previous frame. Initial ASM guess in next frame, based on model position from previous frame. Final ASM shape in next frame.

forces $\vec{f}_{2d,i}$ described in Equation 3.

Step 9 calculates the generalized forces, which describe the effect that the displacement between the current frame and the next frame has on the model parameters. These are then integrated in step 11, which yields the model parameters for the next frame and finds the best possible match between the 3D model and the next image frame.

We now turn to an explanation of the statistical inference in steps 8 and 10.

5.3. Statistical Inference: Unsupervised meets supervised

When an image exhibits patterns incompatible with what the ASM has learned, the resulting shape from the ASM search is likely to be contorted, or can even jump off track altogether. Although these contortions often are temporary, and eventually a correct ASM configuration is found again, any data over this period of time would be wildly inaccurate. Hence, we need to detect such contortions.

The 3D deformable model framework shines at this point. The statistical inference in step 8 of the parameters distributions give us an estimate of the expected distribution of the model points [9, 10] over time. By projecting these into 2D space, we obtain a probability distribution of where each model point is expected to be over time in each image, as described in [19]. Any ASM point that falls too far outside these distributions is discarded. This approach is fast and detects when the entire ASM estimate is off-base.

The statistical inference in step 10, on the other hand, looks for more subtle problems. Sometimes, there are occlusions, or image artifacts, that compromise the ASM search in only a few parts of the image. As a result, the ASM estimate becomes subtly incorrect in a few places. In this situation, 2D-based statistics are not fine-grained enough. Using a robust statistical estimate of the probability distribution of the generalized forces $\vec{f}_{g,i}$ in parameter space, we instead reject outliers based on how well a single force $\vec{f}_{g,i}$ matches this distribution [19].

Both steps 8 and 10 rely on obtaining and propagating probability distributions through the 3D model equations. They couple the supervised learned behavior of the ASM component with the data-driven unsupervised outlier rejection performed by the 3D deformable model component. This coupling is the main contribution of this paper, and achieves impressive results on sequences that could not be tracked before by either method alone, see Section 6.2.

6. Experiments

We ran a variety of experiments to validate the integration of ASMs with 3D deformable models. These included automated fitting and shape adaptation, and used a combination of individually unreliable estimates from ASMs and other cues to track 3D models under difficult circumstances.

6.1. Automated Calibration and Model Adaptation

The first step in tracking consists of automated calibration and adapting the model's shape to the subject. Assuming that the first frame of the video sequence is a frontal face view, we use the Viola-Jones face detector [18] to obtain a bounding box around the face for an initial estimate of its location. Then, ASMs can locate the frontal face feature points accurately. According to the predetermined correspondences between the 2D ASM and the 3D face model from Figure 3, we obtain two sets of coordinates (2D and 3D). We then iteratively perform the following two steps:

1. Approximation of rigid model parameters: With the 2D coordinates from ASM and the 3D coordinates of a generic face model, we use camera calibration techniques to determine the rotation and translation of the model. We have 79 2D-3D data pairs to recover the projection matrix in a robust manner. This matrix can be easily decomposed into the product of an upper triangular matrix K, which contains the internal parameters, an orthogonal matrix for rotation, and a translation part:

$$P_{3\times4} = K_{3\times3} R_{3\times3} \left[I_{3\times3} T_{C3\times1} \right]_{3\times4} .$$

2. Fine-grained recovery of rigid and shape parameters: The recovered rigid parameters allow us to initialize the 3D model for a first coarse alignment with the image. We then use the 2D correspondences from the ASM in the same way as we do for tracking, so as to find the best match between the image and the 3D model. We project the 3D model points into the image plane, calculate the displacements between them and the ASM correspondences, and adjust the model's shape parameters as in steps 3–11 in Algorithm 1 (see [9] for an in-depth discussion of fitting). Figure 5 shows the results of the fitting experiments.

6.2. Statistical inference with 3D models and ASMs

In order to test the effect of combining ASMs with 3D model-based statistical inference, we tracked a particularly difficult sign language sequence taken from the National Center for Gesture and Sign Language Resources at Boston

Figure 5. Automated model fitting and adaptation. Left to right: Before model fitting, ASM feature detection, Adjustment of rigid model parameters, Final fit after adjustment of nonrigid shape parameters.

University[1]. The difficulties stem from the numerous occlusions and fast movements. None of the ASMs were trained on this particular subject, so we expected to see numerous "jumps" — as described in Section 5.3. In addition, we integrated contributions from a KLT tracker and an edge tracker, neither of which provided entirely reliable results, either. The integration of these cues was based on a maximum likelihood estimator for the Gaussian probability distributions of the generalized forces, as mentioned at the end of Section 3, and is described in detail in [9].

Figure 6 shows how step 8 in Algorithm 1 manages to compensate for jumpy ASM behavior by excluding those points that stray too far from the predicted path of the 3D model over several frames. The results in Figure 7 are surprisingly good, considering how unreliable all the individual cues are. In several instances the ASM component pulls the mask back on track. It also reinforces the claim in [8] that the statistical inference in step 10 in Algorithm 1 compensates for occlusions.

7. Conclusions and future work

In this paper, we have shown that the combination of two distinct approaches, 2D ASMs and 3D deformable models, can prevail in situations where neither worked on its own. 3D deformable models can deal with unknown situations without training, but they use an inductive approach, and are prone to drifting. They do, however, allow the application of powerful statistical inference mechanisms. In contrast, ASMs allow automated fitting, prevent drifting, and allow recovery from unforeseen situations, such as occlusions.

However, ASMs are good only on what they learn. As their parameters are principal components, they are not meaningful to the human eye, and are unstable in the presence of outliers in the training set. Moreover, they are not retargetable to different configurations, where new data have to be annotated all over.

Improvements to the statistical inference mechanism are a promising avenue for future work. Although the method described in this paper is a good start, it fails when so many nodes are rejected that only a few remain. In this case, the estimate of the generalized force becomes inaccurate, and can pull the model off track. Future work needs to come up with methods to compensate for this situation.

[1]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.

Acknowledgments

The research in this paper was supported by NSF CNS-0427267 and CNS-0428231, research scientist funds by the Gallaudet Research Institute, CNPq PQ-301278/2004-0, and FAPESP 07/50040-8.

References

[1] D. Beymer and T. Poggio. Image representations for visual learning. *Science*, 272(5270):1905–1909, 1996.

[2] A. Blake and M. Isard. *Active Contours: The Application of Techniques from Graphics,Vision,Control Theory and Statistics to Visual Tracking of Shapes in Motion*. Springer-Verlag, 1998.

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

[4] M. Brand. Morphable 3D models from video. In *CVPR*, 2001.

[5] T. Cootes and C. Taylor. Active shape models — their training and application. *CVIU*, 61(1):38–59, 1995.

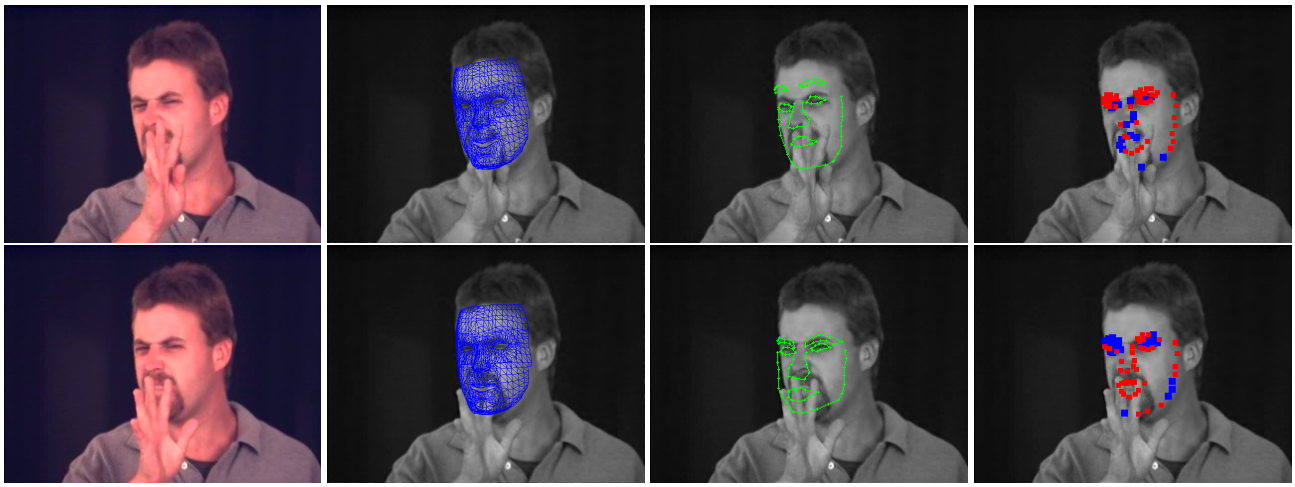

Figure 6. Tracking example where ASMs encounter difficulties, and are only partially reliable. The ASMs shown in this figure have never been trained on this subject. Left to right: Original frame, Tracked 3D model, ASM search result, Statistical rejection of unreliable ASM features, using projected regions of confidence (in red, blue points are accepted).

Figure 7. Results of sign language tracking sequence using ASM, KLT and edge tracking.

[6] D. de Carlo and D. Metaxas. Optical flow constraints on deformable models with applications to face tracking. *IJCV*, 38(2):99–127, July 2000.

[7] M. Dimitrijevic, S. Ilic, and P. Fua. Accurate face models from uncalibrated and ill-lit video sequences. In *CVPR*, pages 1034–1041, 2004.

[8] S. Goldenstein and C. Vogler. When occlusions are outliers. In *Workshop on the 25 Years of RANSAC (in conjunction with CVPR)*, 2006.

[9] S. Goldenstein, C. Vogler, and D. Metaxas. Statistical Cue Integration in DAG Deformable Models. *PAMI*, 25(7):801–813, 2003.

[10] S. Goldenstein, C. Vogler, and D. Metaxas. 3D facial tracking from corrupted movie sequences. In *CVPR*, 2004.

[11] C. Goodall. Procrustes methods in the statistical analysis of shape. In *Journal of the Royal Statistical Society B*, volume 53, pages 285–339, 1991.

[12] M. Isard and A. Blake. Condensation: conditional density propagation for visual tracking. 29(1):5–28, 1998.

[13] M. Jones and T. Poggio. Multidimensional morphable models. In *ICCV*, pages 683–688, 1998.

[14] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face images using flexible models. *PAMI*, 19(7):743–756, 1997.

[15] F. Pighin, R. Szeliski, and D. Salesin. Modeling and animating realistic faces from images. *IJCV*, 50(2):143–169, 2002.

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

[17] T.F.Cootes, G. Edwards, and C.J.Taylor. Active appearance models. *PAMI*, 23(6):681–685, 2001.

[18] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[19] C. Vogler, S. Goldenstein, J. Stolfi, V. Pavlovic, and D. Metaxas. Outlier rejection in high-dimensional deformable models. *IVC*, 25(3):274–284, 2007.

[20] L. Wang, W. Hu, and T. Tan. Face tracking using motion-guided dynamic template matching. In *ACCV*, 2002.

[21] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *CVPR*, 2004.