When Occlusions are Outliers

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

Christian Vogler
Gallaudet University
Gallaudet Research Institute
Christian.Vogler@gallaudet.edu

Abstract

In many tracking applications, the deformable object of interest suffers from frequent occlusions. Traditional augmenting methods use templates and measures of similarity to recover from occlusions. In this paper, we break with these methods. Instead, we model bad image correspondences, which are induced by occlusions, as statistical outliers in the context of tracking high-dimensional deformable models. This interpretation allows us to use robust statistical estimators in the deformable model's parameter space to detect and eliminate such outliers. Because fast-moving occlusions can generate an excessively large outlier to inlier ratio in the occluded areas, we combine the robust statistical estimation with an initial rejection of correspondences based on the magnitude of the optical flow, a simple 2D criterion. To improve robustness even further, we have the final outlier rejection test take into account both the statistical distribution of the deformable model's parameters, and that different parameters are affected by different subsets of correspondences. We validate and demonstrate our technique on real sequences of American Sign Language, which exhibit frequent and extensive occlusions caused by fast movement of the subjects' hands.

1. Introduction

Object tracking is one of the core computer vision activities, and functions as a building block for many higher-level recognition and cognition applications. Every application has a different set of particular characteristics that can constrain or simplify the problem, so tracking is a highly specialized task, with a panoply of different approaches. Yet, all these approaches have in common that they stumble when the object of interest undergoes occlusions. In some applications, such as sign language recognition, occlusions come with the territory, and we cannot dismiss them as rare anomalies or pathological cases. In this paper, we defend the rather bold statement that a statistical representation of the objects with random variables, and detecting statistical outliers, gives rise to a fundamental theoretical framework for dealing with occlusions.

Detecting and recovering from occlusions is a daunting problem. The approach of choice in the literature is the use of *prototypes* [5, 10, 19], which are exemplars of the object in question. Along with measures of similarity, they help identify incompatible changes. In contrast, we view occlusions as 2D regions moving over the object of interest, such as a hand in front of a face, and we assume that we have a parameterized model for the object of interest. We decide on a frame-by-frame basis whether a previously valid tracked feature has to be discarded, using robust statistical estimators for the distribution of the features in parameter space.

The justification for this approach is that, generally, the valid features outnumber the invalid ones, because at any moment, previously valid features only become invalid along the fringe of the occlusion. Even so, fast moving occlusions with large jumps between frames exhibit large fringes, and the number of invalid features can become overwhelming. We address this problem by using an optical flow-based rejection criterion as a preprocessing step, which lowers the number of outliers sufficiently for robust statistical estimators to take over. Coupling outlier rejection and flow allows us to have both model and image information working in a complementary manner, exploiting both 3D [7] and 2D [20] aspects of the problem.

We demonstrate these ideas in the context of a 3D deformable model-based face tracking system (Section 3), and use it to track very difficult sequences from the National Center of Sign Language and Gesture Resources (NCSLGR) [14]. In them subjects tell stories in American Sign Language (ASL), which exhibit fast head translations and rotations, dynamic facial expressions, and abundant occlusions of the face by the hands.

In addition to the practical results (Section 5), we make four theoretical contributions, which all help us conquer occlusions: the concept of applying outlier detection to deal with occlusions (Section 4), the analysis and solution of the problem when insufficient data skew results (Section 3.3),

the statistical improvement of an existing parameter-space outlier rejection method [26] (Sections 4.2 and 4.3), and the combination of outlier rejection with detecting fast 2D motions via Bayesian optical flow (Section 4.4).

2. Related Work

Bayesian filters are the base of statistical tracking. If we model the random variables as Gaussians, the *Kalman filter* [13] and the *Unscented Kalman filter* [27] are popular choices. If the random variables are treated as nonparametric entities, the usual choice is the *particle filter* [8, 9].

Many tracking applications deal with occlusion through prior knowledge. A popular approach is to acquire templates of the subject at key locations [5, 10, 19], and to use measures of similarity to discard points that do not conform to the known model. Another possibility is to learn families of shapes to constrain the search for configurations [4].

Outliers are an important field in the statistics literature [16], and have also attracted a lot of attention in computer vision [2, 24]. Some methods, such as RANSAC [6], MLESAC [25], and IMPSAC [23], look how to identify and discard the outliers. On the other hand, M-estimators look for an optimal weighting of each element [2, 3, 12].

The theory of deformable models is powerful and has been used in a wide variety of applications in 2D [4, 11], and in 3D, such as fitting [1, 18] and tracking [7, 22, 28].

3. Parameterized Deformable Models

A parameterized deformable model can be represented by a set of points p_i, arranged in a polygonal mesh, that describes a discretization of the underlying continuous shape. Each point has a position and associated Jacobian

$$p_i = F_i\left(\vec{q}\right), \quad \text{and} \quad \mathbf{J}_i = \begin{bmatrix} | & & | \\ \frac{\partial p_i}{\partial q_1} & \cdots & \frac{\partial p_i}{\partial q_n} \\ | & & | \end{bmatrix}, \quad (1)$$

with respect to the underlying parameter vector $\vec{q} \in \mathbb{R}^n$.

To match a deformable model with some input media, such as a 2D or volumetric image, we search for the value of $\vec{q}$ that best aligns the model points p_i with the input media's points; in other words, deformable models can be viewed as an optimization problem. In tracking applications, we repeat this procedure recursively, taking advantage of temporal and spatial cohesion by using the estimate from the previous frame as a starting point for the next frame.

Thus, our goal is to minimize the alignment error between the predicted model positions $F_i(\vec{q})$ and the corresponding points in the input image. The traditional optimization procedure consists of a multivariate gradient descent, using the Jacobian as the guide:

$$\dot{\vec{q}} = \sum_i \vec{f}_{gi} = \sum_i \mathbf{B}_i^\top \vec{f}_i, \quad (2)$$

where $\mathbf{B}_i$ is the projection of the Jacobian $\mathbf{J}_i$ into the image plane, and $\vec{f}_i$, also called an *image force,* is the displacement between the desired location of point p_i and the current prediction based on F_i. These image forces are typically estimated by low-level computer vision algorithms, such as point trackers or optical flow measurements. $\vec{f}_{gi}$ is the *generalized force*, which describes how the displacement at point p_i affects the parameter vector $\vec{q}$. Solving the system consists of integrating Equation 2 repeatedly with small Euler steps until the new value of $\vec{q}$ converges.

3.1. Statistical Deformable Models

So far, we have described a deterministic deformable model framework, where $\vec{q}$ is assumed to be known precisely. In reality, because of measurement errors and uncertainties in the underlying low-level algorithms that estimate the image correspondences, $\vec{q}$ can only be approximated with a probability distribution. Instead of representing $\vec{f}_i$ as a fixed quantity, each low-level vision algorithm estimates the correspondences as random variables with an associated error distribution [7, 20]. Goldenstein and colleagues [7] proved that, under certain assumptions, the distribution of $\dot{\vec{q}}$ (Equation 2) is a multivariate Gaussian, and that different algorithms can then be combined optimally with a maximum likelihood estimator.

In this paper we exploit the representation of $\vec{f}_i$ as random variables to obtain an improved outlier rejection criterion in Section 4.3. However, before we can tie outlier rejection to occlusion handling, we need to review what it means for model parameters to be affected by image forces.

3.2. The Unobservability Phenomenon

In the deformable model framework, the low-level algorithms map the image forces $\vec{f}_i$ to their generalized-force counterparts $\vec{f}_{gi}$ through $\mathbf{B}_i^\top \vec{f}_i$ (see Equation 2). In parameter space, the effect of the correspondence $\vec{f}_i$ on the j^{th} parameter of the model is

$$\vec{f}_{gi,j} \quad = \quad \frac{\partial \vec{p}_i}{\partial q_j} \cdot \vec{f}_i, \quad (3)$$

where $\vec{p}_i$ is the model point p_i projected into image space.

If $\vec{f}_{gi,j} = 0$; that is, a particular parameter is unaffected by this image force, there are two possible explanations. Either q_j is already at its best possible value, or $\frac{\partial \vec{p}_i}{\partial q_j} = 0$. In the first case, we do not want the parameter to change, so a zero component in the generalized force is correct. In the second case, we simply cannot decide what to do, because the parameter q_j does not affect the point p_i. This parameter is denoted as *unobservable* at point p_i. Vogler and colleagues showed that unobservability has widespread repercussions for outlier detection and rejection [26]. In

particular, all statistical calculations and tests must be performed on subspaces of the generalized forces $\vec{f}_{gi}$, where the unobservable components have been dropped.

In the following, we will repeatedly refer to the subset of generalized forces for which a parameter q_j is observable:

$$S_j := \left\{ \vec{f}_{gi} \;\middle|\; \frac{\partial \vec{p}_i}{\partial q_j} \neq 0 \right\}. \tag{4}$$

3.3. Unobservability due to Occlusions

Surprisingly, all previous work in the field of 3D deformable models so far has missed the effect that unobservable parameters can have on the correctness of the final generalized force $\vec{f}_g$ in Equation 2. The point of this equation is to average the contributions of the individual generalized forces $\vec{f}_{gi}$ from each correspondence. This averaging works fine if each correspondence contributes equally to each parameter, but as soon as unobservability enters the picture, the result becomes skewed, because the unobservable components still contribute as zeroes in this equation. As a result, $\vec{f}_{gi}$ can end up pointing in an incorrect direction.

As a simplified example, consider the three generalized forces $[1\ 1]^\top$, $[1\ 1]^\top$, and $[1\ \cdot]^\top$, where the second parameter is unobservable in the last force. Summing up these forces according to Equation 2, with unobservable components contributing zeroes, yields the final vector $[3\ 2]^\top$. However, the observable parts of these forces hint that the first and second parameters should be affected equally, so the correct final vector is more likely to be $[3\ 3]^\top$, whose direction differs by 11.3 degrees from $[3\ 2]^\top$.

Under normal circumstances the skew is slight, but is exacerbated by the presence of occlusions, because a typical occlusion can cover large areas of the image, and can almost completely hide regions that are affected by localized parameters, such as the mouth in a deformable face. In such situations there are likely to be large disparities in the sizes of the observable force subsets S_j from Equation 4, resulting in a correspondingly larger skew. Thus, the parameters need to be weighted by their corresponding number of observable forces, and Equation 3 becomes

$$\vec{f}_{gi,j} = \frac{N}{|S_j|} \sum_i \frac{\partial \vec{p}_i}{\partial q_j} \vec{f}_i, \tag{5}$$

where N is the total number of forces available.

4. Occlusion Handling via Outlier Rejection

The averaging in Equations 3 and 5 implies that wildly incorrect correspondence estimates from the low-level image processing algorithms, due to occlusions, can throw off the tracking process arbitrarily. Thus, we need either a method to recover from losing track after occlusions, or a method to prevent the system from losing track altogether by detecting bad correspondences. Traditional occlusion handling methods, such as templates and measures of similarity across image patches, can serve in either role, but make assumptions about image brightness constancy, color constancy, and to a lesser degree, viewpoint invariance. In other words, their success depends very much on the specifics of the input images.

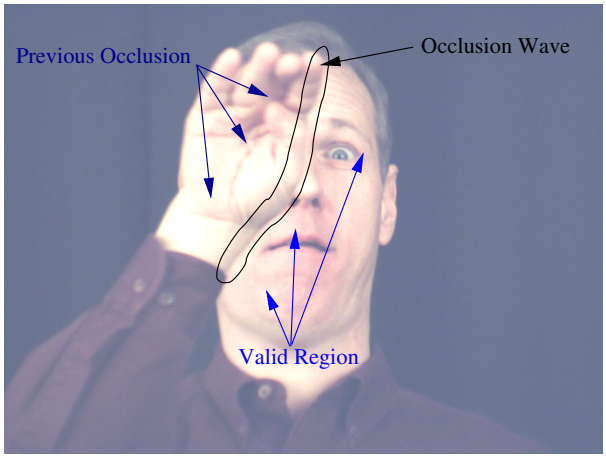

Figure 1. Occlusion "wave front." Only points in this area are newly occluded in this frame.

We now discuss a radically different take on the occlusion problem. It consists of treating bad correspondences, which are induced by occlusions, as statistical outliers. This idea does not depend on any specific characteristics of the input images at all, and emphasizes the prevention of losing track. There are two fundamental assumptions underlying this idea: First, we can estimate and represent the set of generalized forces $\vec{f}_{gi}$ as a probability distribution in parameter space. Second, at any point in time, only a fraction of all estimated correspondences are newly occluded in each frame, so that the number of outliers does not exceed the breakdown point of the statistical estimator. The validity of the second assumption is generally granted through the structure of typical occluding movements. As an example, consider a subject's hand occluding his face in Figure 1. The hand creates a moving 2D area of occlusion, which overlaps from frame to frame. Hence, only newly occluded points lie on the *wave front* of the occluding object, as illustrated in this figure. Nevertheless, there are cases when the second assumption is not valid. We address them in Section 4.4.

The main task is to determine which subset of correspondences is good. Although RANSAC [6] has been successfully applied in similar situations, it is not suitable for the type of high-dimensional deformable models we use in this paper. We now discuss the reasons, then describe and extend an approach based on robust statistical estimators.

4.1. RANSAC: Unsuitable for Deformable Models

The main stumbling block is that Equation 1 does not provide a closed-form solution for $\vec{q} = F_i^{-1}(p_i)$, thanks to highly nonlinear F_i. Thus, during a RANSAC trial [6], for finding the parameters $\vec{q}$, given a subset of the points $\{p_i\}$, we have nothing better than integrating Equation 2. This integration typically requires 600 iterations, and has to be repeated for each trial.

Therefore, in the case of our 3D deformable models, the RANSAC algorithm is computationally far too expensive. For instance, the minimum required number of image forces to estimate $\vec{q}$ correctly is 60–70 for a model with 12 parameters [7], although in experiments we still can obtain fair approximations by using only three times as many points as there are degrees of freedom; that is, $n = 12 \times 3 = 36$. In the presence of occlusions, an outlier percentage of at least 20% is reasonable, hence the probability of hitting a good point is $w = 0.8$. According to [6], we should run $3 \times w^{-n}$ trials with n sampled image forces each, so we end up with $3 \times 0.8^{-36} \approx 9000$ trials.

Running 600×9000 iterations on each frame to solve for $\vec{q}$, before we can apply the RANSAC consensus criterion to select the proper subset of image forces, is prohibitively expensive. Although it is possible to reduce the minimum number of forces through subspace decompositions (see Section 4.2), solving the system 100 or 100,000 times is equally impractical. We now turn to robust parameter space-based outlier rejection techniques, according to [26], and then improve on the rejection criterion by considering additional statistics.

4.2. Parameter Subspace-Based Outlier Rejection

As stated in the introduction to this section, we assume that the generalized forces $\vec{f}_{gi}$ obey a probability distribution, and in Section 3.2 we mentioned that the unobservability phenomenon forces us to calculate all statistics on subspaces of the $\vec{f}_{gi}$, where the unobservable components have been dropped. We now summarize the main results from [26].

If two parameters q_j and q_k can be observed through the same set of forces — that is, $S_j = S_k$ (see Equation 4) —, then we can treat them as correlated, otherwise we have to make the simplifying assumption that they are independent, because we cannot estimate cross-correlations between parameters that are affected by different numbers of points. The respective correlated components of $\vec{q}$ and $\vec{f}_{gi}$ can be grouped together. For the l^{th} such group, $\mathcal{P}_l(\{\vec{f}_{gi}\})$ is defined to be the set of forces that affect the parameter components in this group, projected into the subspace spanned by these components.

The mean of the probability distribution over the $\vec{f}_{gi}$ is

$$\vec{\mu} = \begin{bmatrix} \text{mean}\left(\mathcal{P}_1\left(\left\{\vec{f}_{gi}\right\}\right)\right) \\ \vdots \\ \text{mean}\left(\mathcal{P}_l\left(\left\{\vec{f}_{gi}\right\}\right)\right) \end{bmatrix}, \tag{6}$$

where l is the number of groups, and mean$(\cdot)$ is the mean from the robust minimum covariance determinant (MCD) estimator [16], and

$$\mathbf{\Lambda} = \begin{bmatrix} \boxed{\text{cov}\left(\mathcal{P}_1\left(\left\{\vec{f}_{gi}\right\}\right)\right)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boxed{\text{cov}\left(\mathcal{P}_l\left(\left\{\vec{f}_{gi}\right\}\right)\right)} \end{bmatrix}, \tag{7}$$

where cov$(\cdot)$ is the covariance from the MCD estimator.

The MCD estimator can be configured to have a breakdown point between 0–50%. It is similar to the RANSAC selection procedure in that it also searches for a subset of data points, but its selection criterion consists of minimizing the determinant of the covariance matrix, instead of consensus. We choose this estimator, because it has many desirable statistical and computational properties: it is robust with a configurable breakdown point, which is especially important to prevent the many outliers that arise during occlusions from masking one another; it is affine equivariant; it has reasonable statistical efficiency; and there exists a fast randomized algorithm for calculating a good approximation in arbitrarily high dimensions [17]. Furthermore, experimental results in [26] show that, even without any occlusions, tracking results using outlier rejection via MCD are significantly better than with nonrobust estimators.

The rejection criterion is based on a modified Mahalanobis distance. If a point p_i has k observable parameters, we can build a projection matrix P_i, with dimensions $k \times n$, composed of only 0s and 1s, that projects $\vec{f}_{gi}$, $\vec{\mu}$, and $\mathbf{\Lambda}$ from the n-dimensional parameter space into the k-dimensional observable subspace. Then $\vec{f}_{gi}$ is an outlier when

$$\left(\vec{f}_{gi} - \vec{\mu}\right)^\top P_i^\top \, P_i \Lambda^{-1} P_i^\top \, P_i \left(\vec{f}_{gi} - \vec{\mu}\right) > thresh, \tag{8}$$

where $thresh$ is the threshold determined by the well-known χ^2 probability distribution function:

$$thresh = \left(\chi_k^2\right)^{-1}(0.975). \tag{9}$$

This particular implementation of occlusion handling via outlier rejection requires that the generalized forces conform to a unimodal Gaussian distribution, for which MCD acts as a robust estimator. The principle itself, however, is far more general, as the statistical estimators in Equations 6 and 7 are arbitrary. The only conditions are that the statistics

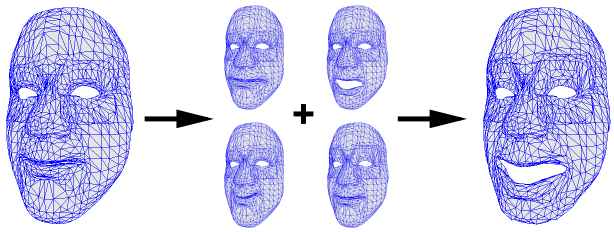

Figure 2. Sum of deformations over the base mesh.

and the rejection criterion can be computed efficiently. Consequently, the outlier rejection procedure is independent of the model parameters $\vec{q}$, and thus needs to be performed only once per frame, before the integration of Equation 2.

4.3. Uncertainties in Image Forces

The rejection criterion in Equation 8 was shown to work well in [26], but it has the defect that it ignores any knowledge about the probability distributions of the individual image forces $\vec{f}_i$, and by extension $\vec{f}_{gi}$. Recall that these estimates are provided by the low-level image processing algorithms, as described in Section 3.1. When it comes to occlusions, we need every bit of help that we can get, so we now propose an extension to this criterion, so as to take these probabilities into account.

Intuitively, if the difference between a generalized force and the MCD mean, $\vec{f}_{gi} - \vec{\mu}$, falls onto a high-variance axis of the force, then $\vec{f}_{gi}$ is less likely to be an outlier than if it falls onto a low-variance axis. If, as in [7], the uncertainties in the image forces $\vec{f}_i$ are represented by affine forms (i.e., regions of confidence) [21], with mean $\vec{f}_i$,

$$\hat{\vec{f}}_i = \vec{f}_i + \sum_k \vec{a}_{i,k}\varepsilon_{i,k}, \text{ where } \varepsilon_{i,k} \in [-1, 1], \qquad (10)$$

the projection of $\hat{\vec{f}}_i$ into parameter space, according to Equation 2, gives regions of confidence for the generalized force

$$\hat{\vec{f}}_{gi} \quad = \quad \mathbf{B}_i^\top \hat{\vec{f}}_i = \vec{f}_{gi} + \sum_k \vec{b}_{i,k}\varepsilon_{i,k}. \qquad (11)$$

Conceptually, a generalized force is an outlier if its region of confidence does not overlap with the bulk of the MCD mean and covariance. We can calculate the overlap by projecting the force onto $\vec{f}_{gi} - \vec{\mu}$ to obtain a 1D interval of confidence:

$$\hat{\vec{f}}_{axis,i} \quad = \quad \vec{f}_{g,i} + \sum_k \vec{d} \left| \frac{\vec{d}}{||\vec{d}||} \cdot \vec{b}_{i,k} \right| \varepsilon_{i,k}, \qquad (12)$$

$$\vec{d} \quad = \quad \vec{f}_{gi} - \vec{\mu}. \qquad (13)$$

Taking the absolute value of the dot product in this equation simplifies the task of finding the two endpoints of the interval, because we can simply let

$$\vec{e}_{i,1} = \hat{\vec{f}}_{axis,i}, \text{ where } \forall_k : \varepsilon_{i,k} = 1 \qquad (14)$$

$$\vec{e}_{i,2} = \hat{\vec{f}}_{axis,i}, \text{ where } \forall_k : \varepsilon_{i,k} = -1 \qquad (15)$$

$\vec{f}_{gi}$ is an outlier if, and only if, both interval endpoints exceed the threshold in Equation 8 in Section 4.2, where we substitute $\vec{f}_{gi}$ with $\vec{e}_{i,1}$ and $\vec{e}_{i,2}$.

4.4. Fast Moving Occlusions and Recovery

As we stated in the introduction to this section, we assume that only a fraction of all points are newly occluded in each frame. Otherwise, if the occluding object is moving too fast, the number of such points may exceed the statistical breakdown point of the robust parameter space outlier rejection technique. We can alleviate this problem by applying other techniques to detect occlusions, and running outlier rejection only on the points that remain afterward.

Bayesian optical flow [20], in particular, is an ideal match for outlier rejection. We calculate the flow over the image sequence, a low-level 2D operation, and threshold its magnitude to detect fast moving regions, shown in Figure 3. We first eliminate all 2D forces that touch the masked region, thereby lowering the total number of outliers, and then proceed with the algorithm from Sections 4.2–4.3.

Even though the thresholded Bayesian flow does not identify the entirety of the occluding object, it has a low incidence of false positives and catches many points on a fast-moving occlusion wave front (cf. Figure 1). In other words, it does best in exactly those situations that cause the most problems for outlier rejection.

4.5. Tracking after Outlier Rejection

To summarize, for each frame we find the model-image correspondences. We then reject a number of them via the optical flow criterion described in the previous section. We convert the remaining correspondences into generalized forces via a projection into parameter space, and subject them to the outlier rejection algorithms described in Sections 4.2–4.3. We then have a set of correspondences and associated generalized forces left that are known to be good. We plug these into Equation 2, and perform a gradient descent procedure to minimize the distance between the model points' positions and their correspondences, at the end of which the face model is aligned with the new frame.

5. Experiments

We applied our technique to several sequences taken from the National Center of Sign Language and Gesture Resources (NCSLGR) [14] using an all-purpose deformable face model with 1101 nodes, 2000 triangles, and 12 parameters that controlled both the rigid transformation and the

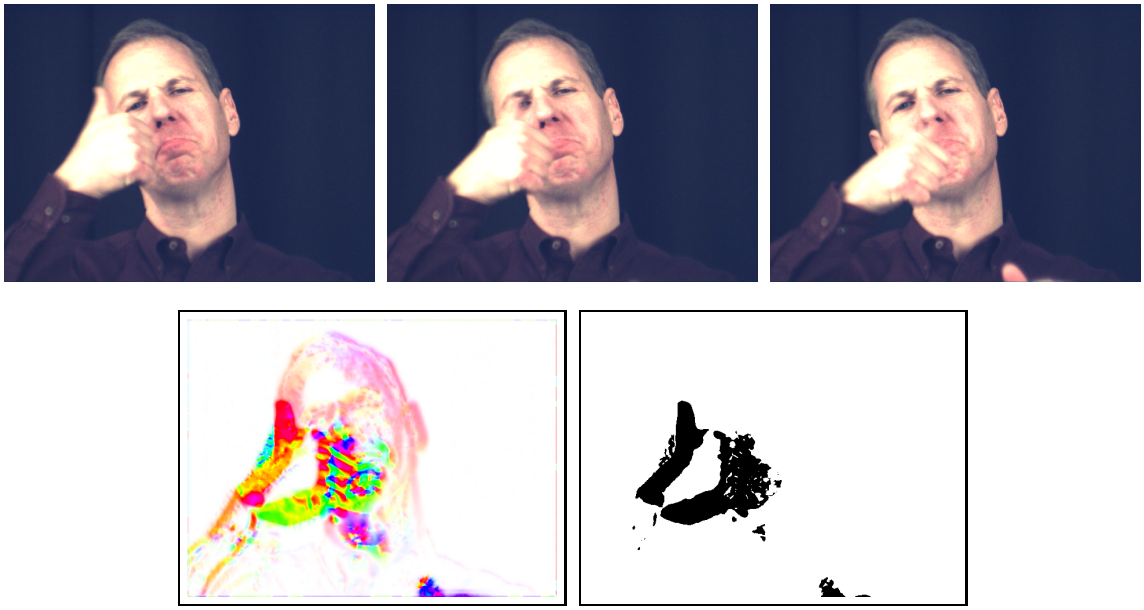

Figure 3. Bayesian flow and thresholded rejection mask. Points that fall into the black area are eliminated.

facial deformations (eyebrows, lip stretching, smiling, lip opening, jaw opening, etc.) of each of these nodes. Before we can track a new subject, the model's mesh first needs to be fitted to the face in an image at rest position, without the effect of deformations. This process needs to be only done once for every subject, through methods such as [1, 15].

In these experiments, we use a face model parameterized by 12 parameters. Seven parameters describe the rigid transformation (quaternion for orientation and a translation vector). One nonrigid parameter simultaneously simulates the effect of the *frontalis* (eyebrow raise) and corrugator (frowning) muscles. Two more simulate the *zygomatic major* muscle and the *risorius* muscle, responsible for cheek movements and smiling. The last two nonrigid parameters are responsible for mouth opening, and jaw movement. These deformations are smooth and applied individually over the base mesh, and for simplicity we add the effects of the different deformations on the mesh (Figure 2).

For the purposes of the validation experiments, we fitted the model in a semi-automated way: the user manually selected a few dozen model-image correspondences. Fitting then consisted of solving Equation 2, with the user-defined correspondences acting as the image forces $\vec{f}_i$ using a finite-element inspired set of shape deformations. Because the definitions of the facial expression deformations are independent of the base mesh, the model was ready for tracking immediately after fitting.

For all experiments, the low-level image processing was provided by a combination of a KLT feature tracker [19], and an edge tracker, with the maximum number of integrations of Equation 2 set to 600. The tracking speed was an average of 0.9s per frame on an AMD Athlon 64 3500+, with precomputed dense Bayesian flow, the calculation of which took around 0.15s per frame.

In Figure 4, we display a few snapshots of a sequence over 1000 frames long, where the subject signs "*I was travelling over the U.S. Route 80 on the highway across the country. I was driving. The road was dry and plain. I was driving and felt bored. So, I figured out what should I do.*" The accompanying video contains the full sequence, downsampled to half of its original frame rate.

The center right picture in this figure typifies the different contributions of this paper. The large combined number of purple and red dots across the arm and hand implies that more than half of all correspondences in this image were bad, due to occlusions, and would have overwhelmed even robust statistical estimators. With the optical flow preprocessing, however, the remaining number of outliers drops to approximately 25%; not too much of a challenge for the MCD estimator. The eyebrows and jaw are almost fully occluded, with very few available correspondences, so without the normalizing technique described in Section 3.3, the generalized force would undergo serious skewing. Finally, the points at the hairline come from the edge tracker, which provides reliable information for movements perpendicular to the edge, but is prone to detecting spurious movements along the edge. Without the statistical extensions to the outlier rejection criterion in Section 4.3, such points are invariably rejected, even though they contain useful information.

Figure 4. Snapshots of a tracking sequence. Top to bottom: original sequence, categorized image correspondences, final 3D model position. The correspondences constitute a mix of KLT and edge tracking results. In the center row, blue points denote accepted correspondences, red points denote rejected outliers, and purple points denote correspondences eliminated by optical flow thresholding.

6. Conclusions and Future Work

In this paper we have shown that outlier rejection is a viable approach to handling occlusions. The following contributions made it suitable for the task: extending the rejection criterion to take into account statistical correspondences, developing a solution for the case when there are large discrepancies in parameter observability, and using a 2D Bayesian optical flow algorithm to handle fast-moving occlusions, which violate the assumptions behind the basic outlier rejection framework. From an artificial intelligence interpretation, the outlier approach is *unsupervised,* whereas template-based approaches are *supervised.* In addition, the KLT feature tracker that we used in our experiments already contains a built-in template-based procedure to discard incorrect matches; yet it did not prevent the tracker from choosing many incorrect correspondences during occlusions. This result provides further evidence that statistical outlier rejection is competitive with template-based methods.

Sign language sequences pose the ultimate tracking challenge, because of the numerous quick movements, facial actions, and occlusions. Large 3D rotations are common, and make simple 2D tracking systems fail. The results in this paper are encouraging, but detailed mouth movements are still problematic. The reason is that outlier rejection does not help with recovery after occlusions, so future work needs to include recovery via prototypes and template matching. In addition, detailed movements require a large number of tracking features, yet these are not easy to find at the resolutions and view angles used in the NCSLGR sequences.

Acknowledgments

The research in this paper was supported by NASA Cooperative Agreements 9-58 with the National Space Biomedical Research Institute, CNPq PQ-301278/2004-0, FAEPEX-Unicamp 1679/04, FAPESP, research scientist funds by the Gallaudet Research Institute, and NSF CNS-0427267.

References

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

[2] H. Chen and P. Meer. Robust computer vision through kernel density estimation. In *Proc. of European Conference of Computer Vision*, pages 236–250, 2002.

[3] H. Chen and P. Meer. Robust regression with projection based m-estimators. In *Proc. of International Conference of Computer Vision*, pages 878–885, 2003.

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

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

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

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

[8] N. Gordon, D. Salmon, and A. Smith. A novel approach to nonlinear/nongaussian bayesian state estimation. *IEEE Proc. Radar Signal Processing*, (140):107–113, 1993.

[9] M. Isard and A. Blake. Condensation: conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[10] H. Jin, P. Favaro, and S. Soatto. Real-time feature tracking and outlier rejection with changes in illumination. In *Proc. of International Conference of Computer Vision*, 2001.

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

[12] R. Marona. Robust m-estimators of multivariate location and scatter. *Ann. Stat.*, 4:51–67, 1976.

[13] P. Maybeck. *Stochastic Models, Estimation, and Control*. Academic Press, 1979.

[14] C. Neidle and S. Sclaroff. Data collected at the National Center for Sign Language and Gesture Resources, Boston University. Available online at http://www.bu.edu/asllrp/ncslgr.html, 2002.

[15] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. Salesin. Synthesizing realistic facial expressions from photographs. In *Proceedings of the SIGGRAPH*, pages 75–84, 1998.

[16] P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.

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

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

[19] J. Shi and C. Tomasi. Good features to track. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[20] E. Simoncelli. *Handbook of Computer Vision and Applications*, volume II, chapter Bayesian Multi-scale Differential Optical Flow, pages 397–422. Acad. Press, 1999.

[21] J. Stolfi and L. Figueiredo. *Self-Validated Numerical Methods and Applications*. 21^{o} Colóquio Brasileiro de Matemática, IMPA, 1997.

[22] H. Tao and T. Huang. 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, 2002.

[23] P. Torr and C. Davidson. IMPSAC: A synthesis of importance sampling and random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3):354–365, 2003.

[24] P. Torr and D. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[25] P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[26] C. Vogler, S. Goldenstein, J. Stolfi, V. Pavlovic, and D. Metaxas. Outlier rejection in high-dimensional deformable models. *Image and Vision Computing*, 2006. In Press.

[27] E. A. Wan and R. van der Merwe. *Kalman Filtering and Neural Networks*, chapter Chapter 7 : The Unscented Kalman Filter, (50 pages). Wiley Publishing, 2001.

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