SHAPE MODEL FITTING ALGORITHM WITHOUT POINT CORRESPONDENCE

Claudia Arellano and Rozenn Dahyot

School of Computer Science and Statistics
Trinity College Dublin, Ireland

ABSTRACT

In this paper, we present a Mean Shift algorithm that does not require point correspondence to fit shape models. The observed data and the shape model are represented as mixtures of Gaussians. Using a Bayesian framework, we propose to model the likelihood using the Euclidean distance between the two Gaussian mixture density functions while the latent variables are modelled with a Gaussian prior. We show the performance of our MS algorithm for fitting a 2D hand model and a 3D Morphable Model of faces to point clouds.

Index Terms— Mean Shift, Gaussian Mixture Models, Morphable Models, Shape Fitting

1. INTRODUCTION

Shape models are obtained by statistically capturing the shape variability from a set of training examples. Using Principal component analysis any shape in the class can be reconstructed using a small set of parameters. Algorithms for fitting shape models to a set of observations are often based on a Bayesian decision framework and the parameters of the model are estimated by maximizing its posterior probability given a set of observations (input data). The likelihood is expressed as the joint probability of each correspondence pair and this modeling assumes that the correspondences between observations and points on the model are known. In practice, however, it is difficult to get accurate correspondence to start the fitting process (initial mismatches affect the robustness and accuracy of the fitting algorithm).

In this paper we propose a Mean Shift algorithm that globally fits the model to the observations (as point clouds) without requiring point to point correspondences. The shape model and the observations are represented as independent mixtures of Gaussians and the likelihood is chosen proportional to the Euclidean distance between the two density functions. Our modelling does not require any kind of correspondence between the two data sets. The only pre-processing required is the affine alignment in between the two shapes that is solved automatically by using our algorithm for robust registration [1].

Thanks to Trinity College Dublin and The Government of Chile for funding.

2. RELATED WORK

Shape models have been widely used in computer vision and medical image analysis for detection, reconstruction and recognition of shapes. Cootes et al. [2] proposed the 2D Active Shape Model which has been widely used for feature detection and image segmentation. Blanz et al. [3] proposed a 3D Model of faces (a.k.a. The Morphable Model). The Morphable Model has been used as a prior for inferring 3D faces from a single image [4, 5] and also multiple-view images [6, 7]. The parameters of the model are estimated by maximizing the posterior probability in a Bayesian framework. The key problem for fitting a shape model is the need for correspondence between the observation and the model itself. For instance, the representation of a particular feature in the observation (i.e tip of the nose) must be in correspondence with the vertex representing the same feature in the Shape Model. Registration methods such as the iterative closest point (ICP) algorithm [8, 9, 10] have been proposed for solving the correspondence problem and the fitting process itself [11]. However, those approaches are sensitive to outliers and in practice, manual labelling to set initial correspondences is often required [12].

Our main contribution in this paper is a new algorithm that does not require correspondence for fitting a shape model to a point cloud. It is also based on a Bayesian framework but uses a robust likelihood [13] and a Gaussian prior (section 3.2). The resulting cost function can be optimised using a dedicated Mean Shift (MS) algorithm (section 3.3). Convergence of the MS algorithm to the global solution is improved using an annealing strategy [14] and it is well suited for parallel programming implementation [15]. We show experimentally (section 4) how robust and accurate our algorithm is for fitting a 2D hand model and a 3D Morphable Model of faces.

3. MEAN SHIFT FOR BAYESIAN SHAPE FITTING

3.1. Bayesian framework for model fitting

Let us define a shape model by its mean shape $\boldsymbol{\mu}$ and a set of J eigenvectors $\{\mathbf{v}_j\}_{j=1,..,J}$ associated with the eigenvalues $\{\sigma_j\}_{j=1,..,J}$ computed by Principal Component Analysis using a representative set of exemplar shapes. We assume any

shape from the same class can be well approximated as a linear combination of $(\boldsymbol{\mu}, \{\mathbf{v}_j\}_{j=1,..,J})$:

$$\mathbf{y}(\alpha) = \boldsymbol{\mu} + \sum_{j=1}^{J} \alpha_j \, \mathbf{v}_j \tag{1}$$

The reconstructed shape $\mathbf{y}(\alpha)$ depends on the latent parameters $\alpha = \{\alpha_j\}_{j=1,..,J}$. Given a set of observations $\mathcal{U} = \{\mathbf{u}_k\}_{k=1,\cdots,n}$, we aim at estimating the parameters $\{\alpha_j\}_{j=1,..,J}$ such that $\mathbf{y}(\alpha)$ best fit the observations. Using a Bayesian framework the parameters are estimated by maximising the posterior:

$$\hat{\alpha} = \arg\max_{\alpha} p(\mathcal{U}|\alpha)\, p(\alpha) \tag{2}$$

with $p(\alpha)$ the prior and $p(\mathcal{U}|\alpha)$ the likelihood. The prior is chosen here as a multivariate Gaussian and it is expressed as follows:

$$p(\alpha = (\alpha_1, \cdots, \alpha_J)) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{J} \frac{\alpha_j^2}{\sigma_j^2}\right) \tag{3}$$

The standard modelling of the likelihood corresponds to first setting correspondences between a vertex of model (e.g. $\mathbf{y}_i(\alpha)$) and a vertex in the observations (e.g.$\mathbf{u}_i$) : $(\mathbf{u}_i, \mathbf{y}_i(\alpha)), \forall i = 1, \cdots, m$ $(m < n)$. Assuming independence of the pairs, the standard likelihood is:

$$p(\mathbf{u} = (\mathbf{u}_1, \cdots, \mathbf{u}_m)|\alpha) \propto \prod_{i=1}^{m} \exp\left(-\frac{\|\mathbf{u}_i - \mathbf{y}_i(\alpha)\|^2}{2\sigma^2}\right) \tag{4}$$

Next, we propose to change this likelihood so that no correspondences are required between the data sets.

3.2. Modelling a robust Likelihood

The observations $\mathcal{U} = \{\mathbf{u}_k\}_{k=1,\cdots,n}$ and the model $\mathbf{y}(\alpha)$ are usually represented by vectors where the ordering of the vertices matters. Moreover, if the observations have n vertices, while the model is composed of m vertices comparing vectors of different dimensions gets even more difficult. So instead of representing the observations and the model by vectors, we propose to represent them by density functions, f_u and f_α, of a random variable $\mathbf{x} \in \mathbb{R}^D$ ($\mathbf{x}$ has the same dimension as the vertices of the model and the observations, i.e. $D = 2$ for 2D shape models, $D = 3$ for 3D shape models). Loosely speaking, f_u and f_α can be understood as infinite dimensional vectors and the likelihood can be expressed using the Euclidean distance between those two density functions:

$$p(\mathcal{U}|\alpha) = p(f_u|f_\alpha) \propto \exp\left(-\frac{(\|f_u - f_\alpha\|)^2}{2\sigma_d^2}\right) \tag{5}$$

The Euclidean distance is here defined by:

$$\begin{aligned}\|f_u - f_\alpha\|^2 &= \int_{\mathbb{R}^D} (f_u(\mathbf{x}) - f_\alpha(\mathbf{x}))^2 \, d\mathbf{x} \\ &= \int_{\mathbb{R}^D} (f_u(\mathbf{x})^2 - 2 f_u(\mathbf{x}) f_\alpha(\mathbf{x}) + f_\alpha(\mathbf{x})^2) d\mathbf{x}\end{aligned} \tag{6}$$

This Euclidian distance alone is known to be robust to estimate parameters [13]. Using equation (5) in equation (2), our parameter estimation is formulated as:

$$\hat{\alpha} = \arg\min_{\alpha} \left\{ -\log p(f_\alpha|f_u) \simeq \frac{\|f_u - f_\alpha\|^2}{2\sigma_d^2} + \sum_{j=1}^{J} \frac{\alpha_j^2}{2\sigma_j^2} \right\} \tag{7}$$

The variance σ_d^2 is set experimentally and allows us to control the influence of the likelihood with the prior.

3.3. Mean Shift Algorithm for Gaussian Mixture

We model $f_u(\mathbf{x})$ and $f_\alpha(\mathbf{x})$ as a multi-Gaussian density function where each Gaussian kernel is centered on each point of the data sets as follows:

$$\begin{aligned} f_u(\mathbf{x}) &= \sum_{k=1}^{n} \frac{\pi_k}{\left(\sqrt{2\pi} h_k\right)^D} \exp\left(\frac{-\|\mathbf{x} - \mathbf{u}_k\|^2}{2(h_k)^2}\right) \\ f_\alpha(\mathbf{x}) &= \sum_{i=1}^{m} \frac{\pi'_i}{\left(\sqrt{2\pi} h'_i\right)^D} \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}_i(\alpha)\|^2}{2(h'_i)^2}\right) \end{aligned} \tag{8}$$

with $D = \dim(\mathbf{x})$. The Euclidean distance defined equation (6) can be computed explicitly [1]:

$$\begin{aligned} &\|f_u - f_\alpha\|^2 = \\ &\sum_{k=1}^{n}\sum_{p=1}^{n} \underbrace{\frac{\pi_k \pi_p}{\left(\sqrt{2\pi((h_k)^2 + (h_p)^2)}\right)^D} \exp\left(-\frac{\|\mathbf{u}_k - \mathbf{u}_j\|^2}{2((h_k)^2 + (h_p)^2)}\right)}_{E_1(\mathbf{u}_k, \mathbf{u}_p)} \\ &-2\sum_{i=1}^{m}\sum_{k=1}^{n} \underbrace{\frac{\pi'_i \pi_k}{\left(\sqrt{2\pi((h'_i)^2 + (h_k)^2)}\right)^D} \exp\left(-\frac{\|\mathbf{y}_i(\alpha) - \mathbf{u}_k\|^2}{2((h'_i)^2 + (h_k)^2)}\right)}_{E_2(\mathbf{y}_i(\alpha), \mathbf{u}_k)} \\ &+\sum_{i=1}^{m}\sum_{p=1}^{m} \underbrace{\frac{\pi'_i \pi'_p}{\left(\sqrt{2\pi((h'_i)^2 + (h'_p)^2)}\right)^D} \exp\left(-\frac{\|\mathbf{y}_i(\alpha) - \mathbf{y}_p(\alpha)\|^2}{2((h'_i)^2 + (h'_p)^2)}\right)}_{E_3(\mathbf{y}_i(\alpha), \mathbf{y}_p(\alpha))} \end{aligned} \tag{9}$$

The estimate $\hat{\alpha}$ is then found by minimising $E(\alpha)$ defined by:

$$\begin{aligned} E(\alpha) = &\sum_{k=1}^{n}\sum_{p=1}^{n} E_1(\mathbf{u}_k, \mathbf{u}_p)) - 2\sum_{i=1}^{m}\sum_{k=1}^{n} E_2(\mathbf{y}_i(\alpha), \mathbf{u}_k) \\ &+ \sum_{i=1}^{m}\sum_{p=1}^{m} E_3(\mathbf{y}_i(\alpha), \mathbf{y}_p(\alpha)) + \lambda\sigma_d^2 \sum_{j=1}^{J} \frac{\alpha_j^2}{2\sigma_j^2} \end{aligned} \tag{10}$$

where we introduce a parameter λ to let the user control the influence of the likelihood with the prior. The Mean Shift Algorithm is then computed by differentiating the Energy function E with respect to α and equalling the results to zero.

Algorithm 1 Estimation of α

Input: $\alpha_j^{(0)} = 0, \forall j, \quad h_{min}, h_{max}, e_o, \ \lambda = 0.05$ and $\beta = 0.8$
  $h = h_{max}$
  repeat
    $\sigma_d^2 = \|f_u - f_\alpha\|^2$
    repeat
      Compute $A(\alpha^{(t)})$ and $\mathbf{b}(\alpha^{(t)})$ from eq. (12) and (13)
      $\alpha^{(t+1)} = A(\alpha^{(t)})^{-1}\mathbf{b}(\alpha^{(t)})$
    until $|\alpha^{(t+1)} - \alpha^{(t)}| \leq e_o$
    $h \leftarrow \beta h$
  until $h \leq h_{min}$

Starting from an initial guess $\alpha^{(t)}$, the update is computed by:

$$\alpha^{(t+1)} = \mathrm{A}(\alpha^{(t)})^{-1}\mathbf{b}(\alpha^{(t)}) \tag{11}$$

with A a $J \times J$ matrix defined as:

$$A_{j,s}(\alpha) = \begin{cases} L_{j,s}(\alpha), & \text{if } j \neq s \\ L_{j,s}(\alpha) + \frac{\lambda\sigma_d^2}{\sigma_j^2}, & \text{if } j = s \end{cases} \tag{12}$$

and $L_{j,s}$ defined by:

$$\begin{aligned} L_{j,s}(\alpha) \quad &= 2\sum_{i=1}^{m}\sum_{k=1}^{n}\frac{E_2(\mathbf{y}_i(\alpha),\mathbf{u}_k)}{((h_i')^2+(h_k)^2)}(\mathbf{v}_{ji}\mathbf{v}_{si}^T) \\ &\quad - \sum_{i=1}^{m}\sum_{p=1}^{m}\frac{E_3(\mathbf{y}_i(\alpha),\mathbf{y}_p(\alpha))}{((h_i')^2+(h_p')^2)}(\mathbf{v}_{ji}-\mathbf{v}_{jp})(\mathbf{v}_{si}-\mathbf{v}_{sp})^T \end{aligned}$$

The vector $\mathbf{b}$ is defined as:

$$\begin{aligned} \mathbf{b}_j(\alpha) \quad &= 2\sum_{i=1}^{m}\sum_{k=1}^{n}\frac{E_2(\mathbf{y}_i(\alpha),\mathbf{u}_k)}{((h_i')^2+(h_k)^2)}(\mathbf{u}_k - \mathbf{y}_i(0))\mathbf{v}_{ji}^T \\ &\quad + \sum_{i=1}^{m}\sum_{p=1}^{m}\frac{E_3(\mathbf{y}_i(\alpha),\mathbf{y}_p(\alpha))}{((h_i')^2+(h_p')^2)}(\mathbf{y}_i(0)-\mathbf{y}_p(0))(\mathbf{v}_{ji}-\mathbf{v}_{jp})^T \end{aligned} \tag{13}$$

Note that when $\alpha_j = 0, \forall j$, the model $\mathbf{y}(0)$ corresponds to the average shape of the model $\boldsymbol{\mu}$. The MS algorithm is presented in algorithm 1 with its annealing strategy using the bandwidth as temperature. In this paper all the kernels are modelled using the same bandwidth h. Starting from a maximum value h_{max}, the bandwidth is decreased using a geometric rate β until the minimum value h_{min} is reached. Values of β and λ have been set to be the same for all experiments in section 4 and a natural starting guess of the model shape is $\mathbf{y}(0) = \boldsymbol{\mu}$ (i.e. $\alpha_j = 0, \forall j$). The Euclidean distance between density functions (i.e. likelihood) has already been shown to be more robust than the state of the art techniques for point cloud registration [1, 16]. We show next experimentally how the prior helps in making the estimation even more robust for Morphable model fitting.

4. EXPERIMENTAL RESULTS

In this section we report the results obtained when fitting a 2D model of a hand and the 3D Morphable model of faces to synthetic and real data sets. The alignment between the observation and the model in each experiment is done using our algorithm for rigid registration [1].

4.1. Fitting a 2D hand Model to a point cloud

We compute the hand model from the annotated data sets of 18 hands provided by Tim Cootes[1]. We generated 100 random shapes from the hand model using $J = 10$ eigenvectors. Starting from an initial guess $\alpha^{(o)}$ we achieve convergence towards the right solution for all the target shapes (see figure 1a)). The same results are obtained when adding noise to the observation sets demonstrating the robustness of the algorithm to outliers (figure 1b). The use of the prior in our modelling provides the information needed for preserving the shape of the object during the optimization. When it is not used ($\lambda = 0$) the structure of the shape is lost and the algorithm gets stuck in local solutions (Figure 1c) and d)).

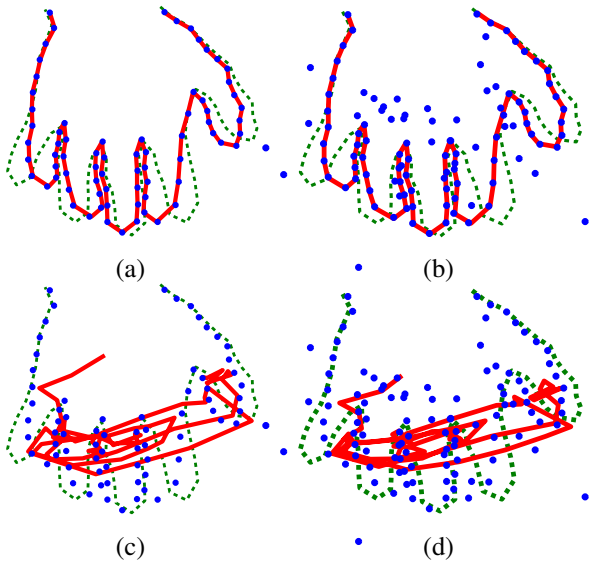
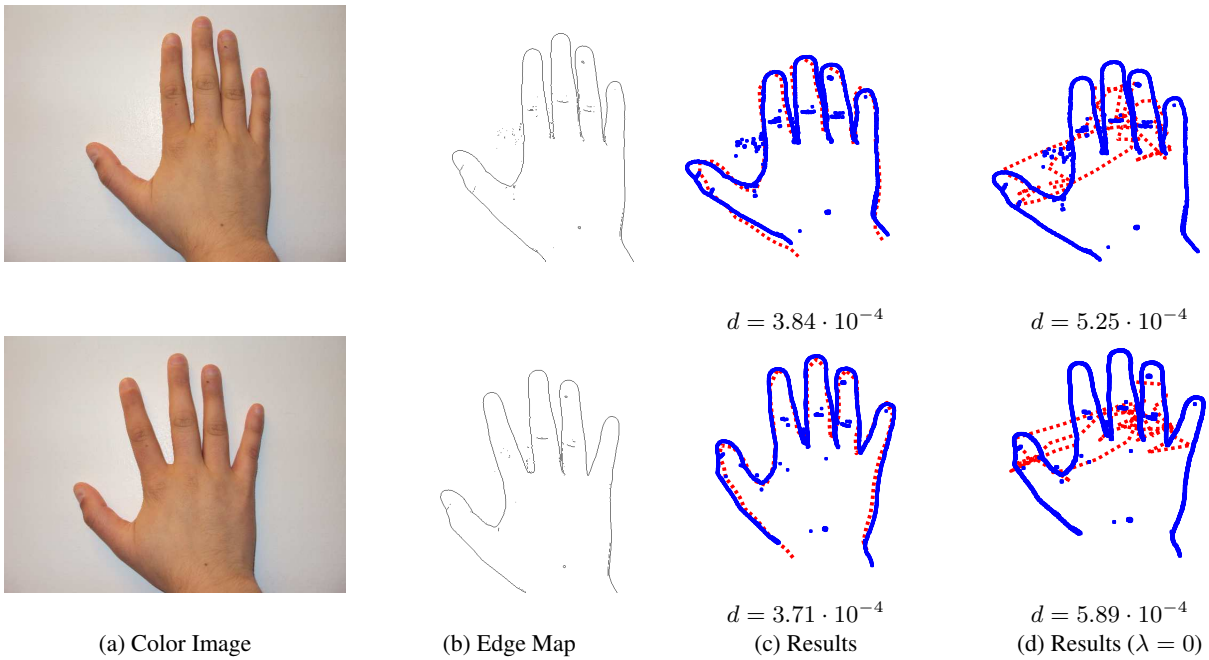


Fig. 1. Results obtained when using synthetic hands generated from the model (a) and when adding random noise (b). Figures (c) and (d) show the estimated shape when the prior information is not considered in the algorithm ($\lambda = 0$). In all figures: observation (blue dots), initial guess (green dash) and estimated solution (red line). Setting: $h_{max} = 50$, $h_{min} = 5$.

In figure 2 we show our results when fitting the 2D hand model to point clouds obtained from images. The number of points on each observed hand are 2421 (top hand) and 2329 (bottom hand). The 2D hand model only contains 72 vertices.

[1] http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/data/hand_data.html



(a) Color Image (b) Edge Map (c) Results (d) Results ($\lambda = 0$)

Fig. 2. Hand model fitting to a point cloud obtained from 2D images. Column (a) and (b) show the colour image and edge map used for the experiments. In column (c) the estimated hands (red dash) and in (d) the solution when the prior is not used in the algorithm ($\lambda = 0$). For the two experiments we compute the Euclidean distance d between the model and the observations. When using the prior information the algorithm minimize the Euclidean distance better (results in c) than when the prior is not used (d).

Since there is no need for one to one correspondence between the data sets, all the observations are considered during optimization. Figure 2 (column (c)) shows our estimated hands (red dash) and the observations (blue dots). An additional experiment was performed using the algorithm without prior (figure 2 column (d)). In this case the algorithm does not converge to the global solution and gets stuck in a local solution that misrepresents the shape contained on the data set.

4.2. 3D Morphable Model fitting

In this section, we use the 3D shape Face Model[2] provided by Basel University [17]. As a first experiment, we generate synthetic faces by selecting random αs. These synthetic faces are used as observations (Target face) for which we know the ground truth α_{GT}. Figure 3 shows an example of the convergence of our algorithm towards the expected solution. In the top row we have: target face, starting guess, result obtained and the error plot between the target face and our estimated face. At the bottom of figure 3, we show the estimated values for α (red diamond), the ground truth (blue line) and the starting guess (green dash). As a second experiment, we record 3D scans of faces (observations) with a Kinect sensor (these new faces were not in the exemplars used to compute the Basel Morphable model). First, the observation is aligned

[2] http://faces.cs.unibas.ch/bfm/.

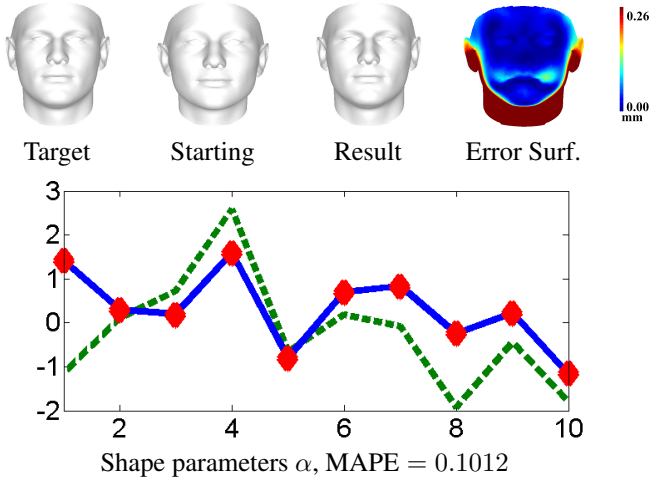


Fig. 3. 3D face fitting. Top row: observation (target), initial guess (starting), estimated face (result) and the error surface in between the estimated and the target face. Bottom row: parameters α corresponding to the estimated face (red diamond), the ground truth (blue line) and the starting guess (green dash). Setting used:$h_{min} = 2.5mm$, $h_{max} = 1.5cm$.

to the mean face [1]. Then the coefficients α are estimated. The estimated reconstruction with our algorithm is shown in figure 4 and compared to the reconstruction when using the

standard ICP algorithm for correspondence and modelling the likelihood as the joint probability of each observation. Note that manual labelling was required for using the ICP algorithm while our algorithm is fully automatic. Error surfaces computed between the estimates and the laser scan are also shown in figure 4. The average error for the reconstructed face using our method is $1.5mm$ with a standard deviation of $2.9mm$ which is comparable with results reported in literature [18].

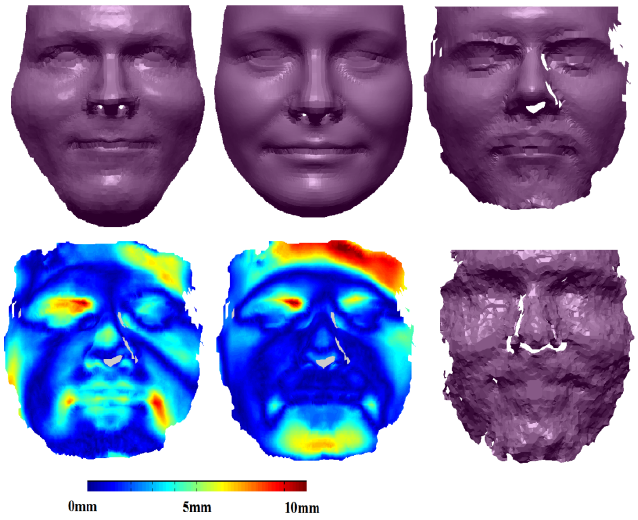


Fig. 4. 3D Face Fitting to kinect data. Reconstruction using the standard ICP-Newton algorithm (top left), our reconstruction (top middle), laser scan (top right), error surface between ICP estimate and laser scan (bottom left), error surface between our estimate and laser scan (bottom middle) and the observation captured using the Kinect (bottom right).

5. CONCLUSION

We have presented a Mean Shift Algorithm that globally fits a shape model to a point cloud in a Bayesian framework. The algorithm does not require any kind of correspondence between the model and the observation. The likelihood is defined over the Euclidean distance between two density functions modelled (as Multi-Gaussian) using the shape model and the observation respectively. Results when using our MS algorithm for fitting 2D and 3D models to real and synthetic data are reported showing the applicability and robustness of our algorithm.

6. REFERENCES

[1] C. Arellano and R. Dahyot, "Mean shift algorithm for robust rigid registration between gaussian mixture models," *20th European Signal Processing Conference (Eusipco)*, 2012.

[2] T. Cootes, C. Taylor, D.H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38 – 59, 1995.

[3] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," *Proceedings of the annual conference on Computer graphics and interactive techniques, SIGGRAPH*, pp. 187–194, 1999.

[4] S. Romdhani, V. Blanz, and T. Vetter, "Face identification by fitting a 3d morphable model using linear shape and texture error functions," *European Conference on Computer Vision, ECCV*, pp. 3–19, 2002.

[5] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1063–1074, 2003.

[6] B. Amberg, A. Blake, A. Fitzgibbon, S. Romdhani, and T. Vetter, "Reconstructing high quality face-surfaces using model based stereo," *IEEE International Conference on Computer Vision, ICCV*, pp. 1–8, 2007.

[7] X. Wang, W. Liang, and L. Zhang, "Morphable face reconstruction with multiple views," *International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC*, pp. 250–253, 2010.

[8] Paul B and N McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 239–256, 1992.

[9] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. Journal Comput. Vision*, vol. 13, pp. 119–152, 1994.

[10] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," *International Conference on 3-D Digital Imaging and Modeling*, 2001.

[11] D.C Schneider and P. Eisert, "Algorithms for automatic and robust registration of 3d head scans," *Journal of Virtual Reality and Broadcasting*, vol. 7, 2010.

[12] T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Realtime performance-based facial animation," *ACM Trans. Graph.*, vol. 30, pp. 1–10, 2011.

[13] D.W. Scott, "Parametric statistical modeling by minimum integrated square error," *Technometrics*, vol. 43, pp. 274–285, 2001.

[14] C. Shen, M.J. Brooks, and A. van den Hengel, "Fast global kernel density mode seeking: Applications to localization and tracking," *IEEE Transactions on Image Processing*, vol. 16, 2007.

[15] B. Srinivasan, H. Qi, and R. Duraiswami, "Gpuml: Graphical processors for speeding up kernel machines," *Workshop on High Performance Analytics - Algorithms, Implementations, and Applications, Siam Conference on Data Mining*, 2010.

[16] B. Jian and B. Vemuri, "Robust point set registration using gaussian mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[17] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3d face model for pose and illumination invariant face recognition," *IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 296–301, 2009.

[18] M. Zollhöfer, M. Martinek, G. Greiner, M. Stamminger, and J. Süßmuth, "Automatic reconstruction of personalized avatars from 3d face scans," vol. 22, no. 3-4, pp. 195–202, 2011.