Face Alignment Through 2.5D Active Appearance Models

Pedro Martins
pedromartins@isr.uc.pt
Rui Caseiro
ruicaseiro@isr.uc.pt
Jorge Batista
batista@isr.uc.pt

Institute of Systems and Robotics
University of Coimbra
Coimbra, Portugal

This work addresses the fitting of 3D deformable face models from a single view through 2.5D Active Appearance Models (AAM) [4]. The main contribution of this paper is the use of 2.5D AAM that combines a 3D *metric* Point Distribution Model (PDM) and a 2D appearance model whose control points are defined by *full perspective* projections of the PDM. The advantage is that, assuming a calibrated camera, 3D metric shapes can be retrieved from single view images.

1 2.5D Parametric Models

The shape of a non-rigid object can be expressed by a linear combination of a set of n basis shapes (stored in a matrix Φ) i.e. a PDM. A 3D v-point shape is defined by $s = (X_1,\dots,X_v,Y_1,\dots,Y_v,Z_1,\dots,Z_v)^T$ and the 3D PDM, including the full pose variation, is given by

$$s = s_0 + \sum_{i=1}^{n} p_i\phi_i + \sum_{j=1}^{6} q_j\psi_j + \underbrace{\int_0^{t-1}\sum_{j=1}^{6} q_j\psi_j\partial t}_{s_\psi} \tag{1}$$

where $\mathbf{p}$ are the shape parameters, $\mathbf{q} = [w_x, w_y, w_z, t_x, t_y, t_z]^T$ are the pose parameters and s_ψ is the contribution of pose increments over time t. $\psi_1,\dots\psi_6$ are a special set of eigenvectors that are expressed w.r.t. the base mesh, s_0, and derived from first order approximations of the Rodrigues rotation formula (becoming only valid for small changes in pose).

Using a *full perspective* camera, the 3D shape s is projected into the image space as $\mathbf{x_p} = \mathbf{K}\left[\begin{array}{c|c}\mathbf{R}_0 & \mathbf{t}_0\end{array}\right]s$ where $\mathbf{K}$ is the camera matrix, assumed to be know and $\mathbf{R}_0|\mathbf{t}_0$ is the *Base Pose* - the head reference.

The 2D appearance model is built by texture-warp all the training images into a common reference using a warping function $\mathbf{W}$. The warp $\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})$ is a piecewise affine warp and is a function of the shape and pose parameters that defines the 2D texture control points by means of the perspective projection of the mesh s. The appearance model given by $\mathbf{A}(\mathbf{x_p}) = \mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2}\lambda_i\mathbf{A}_i(\mathbf{x_p})$, $\mathbf{x_p}\in s_{0\mathbf{p}}$ where λ is a m dimensional vector of appearance parameters. Two extra eigen images are used to model illumination gain and offset.

2 Model Fitting

Fitting the 2.5D AAM consists in solving

$$\sum_{\mathbf{x_p}\in s_{0\mathbf{p}}}\left[\mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2}\lambda_i\mathbf{A}_i(\mathbf{x_p}) - \mathbf{I}(\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q}))\right]^2 \tag{2}$$

simultaneously for $\mathbf{p}$, $\mathbf{q}$ and λ respectively. $\mathbf{I}(\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q}))$ represents the input image $\mathbf{I}(\mathbf{x_p})$ warped by $\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})$. Since the Inverse Compositional approach was proved in [2] to be invalid for the 2.5D AAM, the additive formulation proposed by Lucas-Kanade[3] was adopted. Eq.2 can be solved by the Simultaneous Forwards Additive (SFA) using additive updates to the parameters as

$$\Delta\mathbf{r} = \mathbf{H}_{\text{sfa}}^{-1}\sum_{\mathbf{x_p}\in s_{0\mathbf{p}}}\mathbf{SD}(\mathbf{x_p})_{\text{sfa}}^T\mathbf{E}(\mathbf{x_p})_{\text{sfa}} \tag{3}$$

where $\mathbf{r} = [\mathbf{p}^T\ \mathbf{q}^T\ \lambda^T]^T$, $\mathbf{H}_{\text{sfa}}$ is the Hessian matrix,

$$\mathbf{SD}(\mathbf{x_p})_{\text{sfa}} = \left[\nabla\mathbf{I}\frac{\partial\mathbf{W}}{\partial\mathbf{p}_1}\ \dots\ \nabla\mathbf{I}\frac{\partial\mathbf{W}}{\partial\mathbf{p}_n}\ \nabla\mathbf{I}\frac{\partial\mathbf{W}}{\partial\mathbf{q}_1}\ \dots\nabla\mathbf{I}\frac{\partial\mathbf{W}}{\partial\mathbf{q}_6}\ \mathbf{A}_1(\mathbf{x_p})\ \dots\ \mathbf{A}_{m+2}(\mathbf{x_p})\right] \tag{4}$$

are the Steepest Descent images that depend on $\mathbf{p}$ and $\mathbf{q}$ (by the Jacobian of the Warp $\frac{\partial\mathbf{W}}{\partial\mathbf{p}}$ and $\frac{\partial\mathbf{W}}{\partial\mathbf{q}}$) and $\mathbf{E}(\mathbf{x_p})_{\text{sfa}}$ is the Error image.

The Normalization Forwards Additive (NFA) algorithm solves eq.2 by projecting out the appearance images $\mathbf{A}_i(\mathbf{x_p})$ from the error image. searching only for the shape and pose parameters. The parameters update is given by $\left[\begin{array}{c}\Delta\mathbf{p}\\ \Delta\mathbf{q}\end{array}\right] = \mathbf{H}_{\text{nfa}}^{-1}\sum_{\mathbf{x_p}\in s_{0\mathbf{p}}}\mathbf{SD}(\mathbf{x_p})_{\text{nfa}}^T\mathbf{E}(\mathbf{x_p})_{\text{nfa}}$, where the $\mathbf{SD}(\mathbf{x_p})_{\text{nfa}}$ are far less dimensional than the equivalent for SFA algorithm.

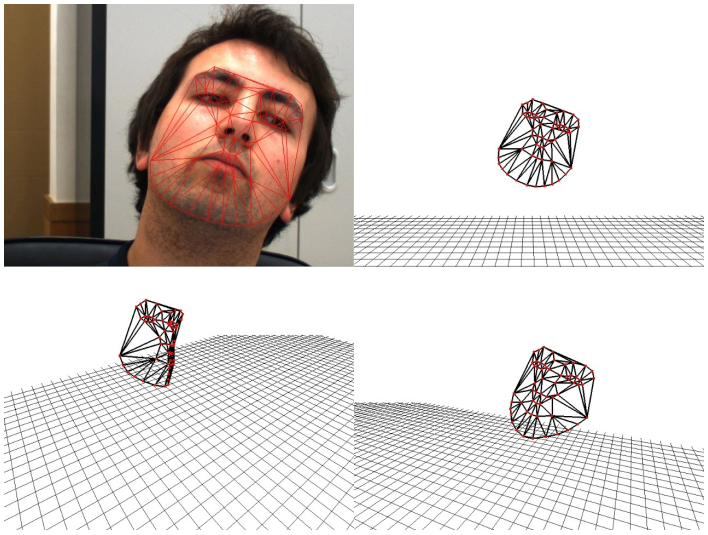

Figure 1: Example of 2.5D AAM fitting.

Some computational load can be reduced by eliminating the need to recompute image gradients at each iteration. Following the idea proposed by [1] we can use the approximation $\nabla\mathbf{I}(\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})) \approx \nabla\mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2}\lambda_i\nabla\mathbf{A}_i(\mathbf{x_p})$ which, besides providing extra computation efficiency (the gradients of the template in eq.4 can be precomputed), it has the great advantage of providing better stability to noise sensitivity since it avoids the reevaluation of the gradients in the input image.

Self-occlusion can be modeled as outlier pixels in the appearance model and handled by robust fitting methods. Robust fitting seek to minimize $\sum_{\mathbf{x_p}}\rho\left(\mathbf{E}(\mathbf{x_p})_{\text{sfa}}^2,\sigma\right)$ where $\rho(.)$ is a robust function and σ is the scale parameter. Using Back-face Culling invisible triangles by the camera can be dropped, setting them as outliers and not taking them into consideration in the fitting process.

2.1 The Jacobian of the Warp

The Jacobian for the shape parameters can be decomposed by the chain rule as $\frac{\partial\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})}{\partial\mathbf{p}} = \sum_{k=1}^{v}\left[\frac{\partial\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})}{\partial\mathbf{x}_k}\frac{\partial\mathbf{x}_k}{\partial\mathbf{p}} + \frac{\partial\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})}{\partial\mathbf{y}_k}\frac{\partial\mathbf{y}_k}{\partial\mathbf{p}}\right]$. The $\frac{\partial\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})}{\partial\mathbf{x}_k}$ and $\frac{\partial\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})}{\partial\mathbf{y}_k}$ components are given by $(1-\alpha-\beta,0)$ and $(0,1-\alpha-\beta)$ respectively, where α and β are barycentric coordinates of the projected base mesh $s_{0\mathbf{p}}$. They depend only on the configuration of the base mesh and thus can be precomputed and efficiently stored as sparse matrices. The same approach is taken to evaluate the Jacobian of the warp for the pose parameters $\frac{\partial\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q})}{\partial\mathbf{q}}$. The remaining terms $\frac{\partial\mathbf{x}_k}{\partial\mathbf{p}}$, $\frac{\partial\mathbf{y}_k}{\partial\mathbf{p}}$ and $\frac{\partial\mathbf{x}_k}{\partial\mathbf{q}}$, $\frac{\partial\mathbf{y}_k}{\partial\mathbf{q}}$ are all *scalars* given by

$$\left[\begin{array}{c}w\mathbf{x}_k\\ w\mathbf{y}_k\\ w\end{array}\right] = \mathbf{K}\left[\begin{array}{c|c}\mathbf{R}_0 & \mathbf{t}_0\end{array}\right]\left[\begin{array}{c}s_{0_i}^{x_k} + p_i\phi_i^{x_k}\\ s_{0_i}^{y_k} + p_i\phi_i^{y_k}\\ s_0^{z_k} + p_i\phi_i^{z_k}\\ 1\end{array}\right]$$

$i = 1,\dots,n$ parameters; $k = 1,\dots,v$ landmarks

$\frac{\partial\mathbf{x}_k}{\partial\mathbf{p}} = \frac{\partial}{\partial\mathbf{p}}\left(\frac{w\mathbf{x}_k}{w}\right)$ and $\frac{\partial\mathbf{y}_k}{\partial\mathbf{p}} = \frac{\partial}{\partial\mathbf{p}}\left(\frac{w\mathbf{y}_k}{w}\right)$

$$\left[\begin{array}{c}w\mathbf{x}_k\\ w\mathbf{y}_k\\ w\end{array}\right] = \mathbf{K}\left[\begin{array}{c|c}\mathbf{R}_0 & \mathbf{t}_0\end{array}\right]\left[\begin{array}{c}s_{0_i}^{x_k} + q_i\psi_i^{x_k} + s_\psi^{x_k}\\ s_{0_i}^{y_k} + q_i\psi_i^{y_k} + s_\psi^{y_k}\\ s_0^{z_k} + q_i\psi_i^{z_k} + s_\psi^{z_k}\\ 1\end{array}\right]$$

$i = 1,\dots,6$ parameters; $k = 1,\dots,v$ landmarks

$\frac{\partial\mathbf{x}_k}{\partial\mathbf{q}} = \frac{\partial}{\partial\mathbf{q}}\left(\frac{w\mathbf{x}_k}{w}\right)$ and $\frac{\partial\mathbf{y}_k}{\partial\mathbf{q}} = \frac{\partial}{\partial\mathbf{q}}\left(\frac{w\mathbf{y}_k}{w}\right)$. (5)

[1] G.Hager and P.Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1025–39, October 1998.

[2] I.Matthews, J.Xiao, and S.Baker. 2d vs. 3d deformable face models: Representational power, construction, and real-time fitting. *International Journal of Computer Vision*, 75(1):93–113, October 2007.

[3] S.Baker and I.Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(1):221–255, March 2004.

[4] T.F.Cootes, G.J.Edwards, and C.J.Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, June 2001.