Let the Shape Speak - Discriminative Face Alignment using Conjugate Priors

Pedro Martins
http://www.isr.uc.pt/~pedromartins

Rui Caseiro
http://www.isr.uc.pt/~ruicaseiro

João F. Henriques
http://www.isr.uc.pt/~henriques

Jorge Batista
http://www.isr.uc.pt/~batista

Institute of Systems and Robotics,
University of Coimbra
Portugal

This work presents a novel Bayesian formulation for aligning faces in unseen images. Our approach is closely related to Constrained Local Models (CLM) [2] and Active Shape Models (ASM) [6], where an ensemble of local feature detectors are constrained to lie within the subspace spanned by a Point Distribution Model (PDM).

Fitting a model to an image typically involves two steps: a local search using a detector, obtaining response maps for each landmark (likelihood term) and a global optimization that finds the PDM parameters that jointly maximize all the detections. The global optimization can be seen as a Bayesian inference problem, where the posterior distribution of the PDM parameters (and pose) can be inferred in a *maximum a posteriori* (MAP) sense. We present a novel Bayesian global optimization strategy, where the prior is used to encode the dynamic transitions of the PDM parameters. Using recursive Bayesian estimation we model the prior distribution of the data as being Gaussian. The mean and covariance were assumed to be unknown and treated as random variables.

The Shape Model: The shape of a PDM is represented by the 2D locations of a mesh $\mathbf{s} = (x_1, y_1, \ldots, x_v, y_v)^T$ (v landmarks). Applying PCA on training examples, results in the parametric model $\mathbf{s} = \mathbf{s}_0 + \Phi\mathbf{b} + \Psi\mathbf{q}$, where $\mathbf{s}_0$ is the mean shape, Φ is the shape subspace matrix (n eigenvectors), $\mathbf{b}$ is a vector of shape parameters, $\mathbf{q}$ the pose parameters vector and Ψ holds four special eigenvectors that linearly model the 2D pose [4].

Goal: Given a $2v$ vector of observed positions $\mathbf{y}$, the goal is to find the optimal set of parameters $\mathbf{b}$ that maximizes the posterior probability of being aligned. Using an Bayesian approach, the shape parameters are

$$p(\mathbf{b}|\mathbf{y}) \propto \left(\prod_{i=1}^{v} p(\mathbf{y}_i|\mathbf{b})\right) p(\mathbf{b}|\mathbf{b}_{k-1}) \tag{1}$$

where $\mathbf{y}_i$ is the i^{th} landmark coordinates and $\mathbf{b}_{k-1}$ is the previous optimal estimate of $\mathbf{b}$. The prior encodes how the shape/pose parameters change.

The Likelihood Term: is the following convex energy function:

$$p(\mathbf{y}|\mathbf{b}_k) \propto \exp\left(-\frac{1}{2}\underbrace{(\mathbf{y}-(\mathbf{s}_0+\Phi\mathbf{b}))}_{\Delta\mathbf{y}}{}^T\Sigma_{\mathbf{y}}^{-1}(\mathbf{y}-(\mathbf{s}_0+\Phi\mathbf{b}))\right) \tag{2}$$

where $\Delta\mathbf{y}$ is the difference between the observed and the mean shape and $\Sigma_{\mathbf{y}}$ is the uncertainty of the spatial localization of the landmarks ($2v \times 2v$ block diagonal covariance matrix).

The response maps can be nonparametrically approximated by using a Kernel Density Estimator (KDE) [5]. Maximizing over the KDE is typically performed by the mean-shift algorithm. Let $\mathbf{z}_i = (x_i, y_i)$ be a candidate to the i^{th} landmark, being $\mathbf{y}_i^c$ the current landmark estimate, $\Omega_{\mathbf{y}_i^c}$ a $L \times L$ patch centered at $\mathbf{y}_i^c$, $\mathbf{I}$ the target image and $p_i(\mathbf{z}_i)$ the probability $\mathbf{z}_i$ is aligned. The i^{th} mean-shift landmark update and its uncertainty are

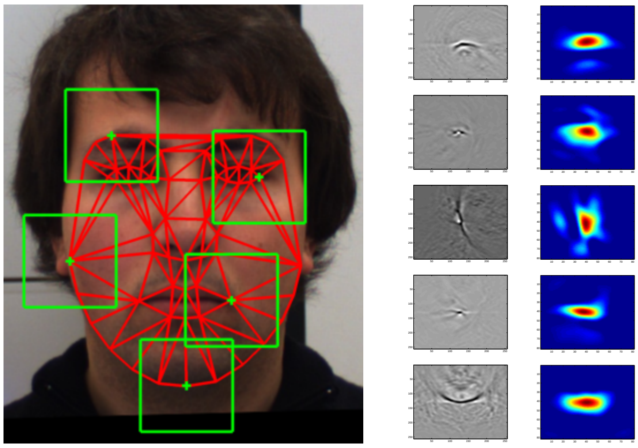
$$\mathbf{y}_i^{\mathrm{KDE}(\tau+1)} \leftarrow \frac{\sum_{\mathbf{z}_i\in\Omega_{\mathbf{y}_i^c}} \mathbf{z}_i\, p_i(\mathbf{z}_i)\, \mathcal{N}(\mathbf{y}_i^{\mathrm{KDE}(\tau)}|\mathbf{z}_i, \sigma_{h_j}^2\mathbf{I}_2)}{\sum_{\mathbf{z}_i\in\Omega_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i)\, \mathcal{N}(\mathbf{y}_i^{\mathrm{KDE}(\tau)}|\mathbf{z}_i, \sigma_{h_j}^2\mathbf{I}_2)}, \tag{3}$$

$$\Sigma_{\mathbf{y}_i}^{\mathrm{KDE}} = \frac{1}{d-1}\sum_{\mathbf{z}_i\in\Omega_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i)(\mathbf{z}_i-\mathbf{y}_i^{\mathrm{KDE}})(\mathbf{z}_i-\mathbf{y}_i^{\mathrm{KDE}})^T, \quad d = \sum_{\mathbf{z}_i\in\Omega_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i), \tag{4}$$

with $\mathbf{I}_2$ a 2D identity matrix and $\sigma_{h_j}^2$ the decreasing bandwidth.

The Prior Term: $p(\mathbf{b}_k|\mathbf{b}_{k-1}) \propto \mathcal{N}(\mathbf{b}_k|\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}})$ follows a Gaussian distribution. Mean $\mu_{\mathbf{b}}$ and covariance $\Sigma_{\mathbf{b}}$ of the data are assumed to be unknown and modeled as random variables [1]. Recursive Bayesian estimation can be applied to infer the parameters of the prior distribution. Defining $\mathbf{b}$ as an observable vector, the joint posterior can be written as

$$p(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}|\mathbf{b}) \propto p(\mathbf{b}|\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}})p(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}). \tag{5}$$



(a) Local search regions. (b) Detectors [3] (c) Responses $p_i(\mathbf{z}_i)$

Figure 1: The Bayesian global optimization strategy jointly combines all detectors scores (MAP sense), explicitly modelling the prior distribution.

The joint prior $p(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}})$ follows a normal-inverse Wishart distribution, assuming $p(\mu_{\mathbf{b}}|\Sigma_{\mathbf{b}})$ a Gaussian (the conjugate prior for a Gaussian with known mean is an inverse Wishart). The joint posterior density $p(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}|\mathbf{b})$ follows an normal-inverse Wishart distribution with hyperparameters [1]:

$$\upsilon_k = \upsilon_{k-1} + m, \quad \kappa_k = \kappa_{k-1} + m \tag{6}$$

$$\theta_k = \frac{\kappa_{k-1}}{\kappa_{k-1}+m}\theta_{k-1} + \frac{m}{\kappa_{k-1}+m}\overline{\mathbf{b}} \tag{7}$$

$$\Lambda_k = \Lambda_{k-1} + \frac{\kappa_{k-1}m}{\kappa_{k-1}+m}(\overline{\mathbf{b}}-\theta_{k-1})(\overline{\mathbf{b}}-\theta_{k-1})^T \tag{8}$$

where θ_{k-1} is the prior mean, κ_{k-1} is the number of prior measurements, $\overline{\mathbf{b}}$ the mean of the new samples, m number of samples, υ_{k-1} and Λ_{k-1} are the degrees of freedom and scale matrix for the inv-Wishart distribution.

Marginalizing $p(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}|\mathbf{b})$ with respect to $\Sigma_{\mathbf{b}}$ gives the marginal posterior distribution for the mean $p(\mu_{\mathbf{b}}|\mathbf{b})$, that follows a multivariate Student-t distribution. Using the expectation of $p(\mu_{\mathbf{b}}|\mathbf{b})$ as the update at instance k we get $\mu_{\mathbf{b}_k} = E(\mu_{\mathbf{b}}|\mathbf{b}) = \theta_k$. Similarly, marginalizing $p(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}|\mathbf{b})$ with respect to $\mu_{\mathbf{b}}$ gives $p(\Sigma_{\mathbf{b}}|\mathbf{b})$ that follows an inverse Wishart distribution. By the expectation of $p(\Sigma_{\mathbf{b}}|\mathbf{b})$ we get $\Sigma_{\mathbf{b}_k} = E(\Sigma_{\mathbf{b}}|\mathbf{b}) = (\upsilon_k - n - 1)^{-1}\Lambda_k$.

Global Alignment (MAP): The recursive posterior distribution is Gaussian, and takes the form of $p(\mathbf{b}_k|\mathbf{y}_k, \ldots, \mathbf{y}_0) \propto \mathcal{N}(\mathbf{b}_k|\mu_k, \Sigma_k)$ with

$$\Sigma_k = \left((\Sigma_{\mathbf{b}_k} + \Sigma_{k-1})^{-1} + \Phi^T\sum_{m=1}^{M}\left(\Sigma_{\mathbf{y}_{(m)}}^{-1}\right)\Phi\right)^{-1} \tag{9}$$

$$\mu_k = \Sigma_k\left(\Phi^T\sum_{m=1}^{M}\left(\Sigma_{\mathbf{y}_{(m)}}^{-1}\Delta\mathbf{y}_{(m)}\right) + (\Sigma_{\mathbf{b}_k} + \Sigma_{k-1})^{-1}\mu_{\mathbf{b}_k}\right) \tag{10}$$

where $\Delta\mathbf{y}_{(m)}$, $\Sigma_{\mathbf{y}_{(m)}}$ are the multiple likelihood observations.

[1] *Bayesian Data Analysis*. Chapman & Hall/CRC, 2nd edition, 2004.

[2] D.Cristinacce and T.F.Cootes. Automatic feature localisation with constrained local models. *Pattern Recognition*, 41(10):3054–3067, 2008.

[3] D.S.Bolme, J.R.Beveridge, B.A.Draper, and Y.M.Lui. Visual object tracking using adaptive correlation filters. In *IEEE CVPR*, 2010.

[4] I.Matthews and S.Baker. Active appearance models revisited. *IJCV*, 60(1): 135–164, 2004.

[5] J.Saragih, S.Lucey, and J.Cohn. Face alignment through subspace constrained mean-shifts. In *IEEE ICCV*, 2009.

[6] T.F.Cootes, C.J.Taylor, D.H.Cooper, and J.Graham. Active shape models-their training and application. *CVIU*, 61(1):38–59, 1995.