Face Pose Estimation in Uncontrolled Environments

Jania Aghajanian
j.aghajanian@cs.ucl.ac.uk
Simon J.D. Prince
s.prince@cs.ucl.ac.uk

Department of Computer Science
University College London
Gower Street
London
WC1E 6BT, UK
http://pvl.cs.ucl.ac.uk

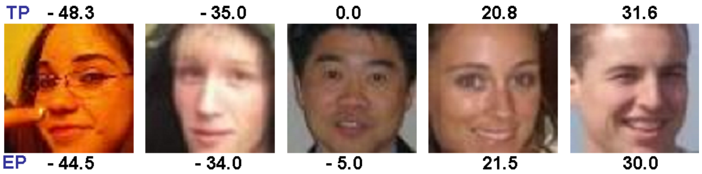

Figure 1: Example images from our pose estimation results. The value above the image is the true pose (human estimate) and the value below the image is the estimated pose using our method.

Automatic estimation of head pose from a face image is a sub-problem of human face analysis with widespread applications such as gaze direction detection, human computer interaction or video teleconferencing. It can also be integrated in a multi-view face detection and recognition system.

Current methods on face pose estimation from a 2D image can be divided into two groups. The first group : geometric shape or template based methods (e.g. [3]) use a set of landmarks such as the relative position of the eyes, mouth, etc. or a template such as an Active Shape Model (ASM) to estimate pose. The second group: manifold learning methods (e.g. [1]) use linear/non-linear embedding methods to learn a lower dimensional space in which they estimate pose.

One limitation of current methods is that most of them estimate pose in a limited range and treat pose estimation as a classification problem by assigning the face to one of many discrete poses [3]. However pose estimation is truly a regression problem. Another drawback of current methods is that they have mainly been tested on faces taken in controlled environments i.e. with solid background and small or no variation in illumination and expression.

In this paper we propose a probabilistic framework for continuous pose estimation. We use a general image representation that does not rely on locating facial features. This representation is inspired by recent successes of *patch-based* methods which have shown to be highly effective for other areas of computer vision such as texture generation [2]. We use this representation in a generative model for automatic estimation of head pose in "real world" images ranging from $-90°$ to $90°$.

Our approach breaks the test image into a non-overlapping regular grid of patches. Each is treated separately and provides independent information about the true pose. There is also a predefined library of object instances. The library can be considered as a palette from which image patches can be taken. We exploit the relationship between the patches in the test image and the patches in the library to estimate the face pose.

In *inference* (see Figure 2), the test image patch is approximated by a patch from the library $\mathscr{L}$ (Fig. 2b,c). The particular library patch chosen can be thought of as having a different affinity with each pose. These affinities were learned during a training period and are embodied in a set of parameters $\mathbf{W}$ (Fig.2d). The relative affinity of the chosen library patch for each pose is used to determine a posterior probability over pose (Fig. 2e).

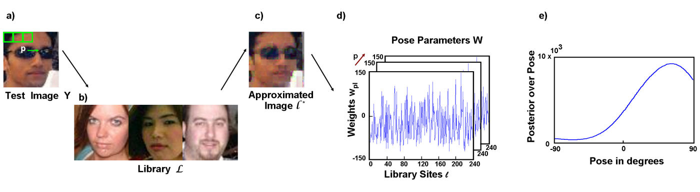

Figure 2: Patch-based probabilistic inference.

The output of our algorithm is a posterior probability over the pose parameter β. We calculate this using Bayes' rule

$$Pr(\beta|\mathbf{Y},\mathbf{W}) = \frac{\prod_{p=1}^{P} Pr(\mathbf{y}_p, l^*|\beta, \mathbf{w}_{p\bullet})Pr(\beta)}{Pr(\mathbf{Y})} \tag{1}$$

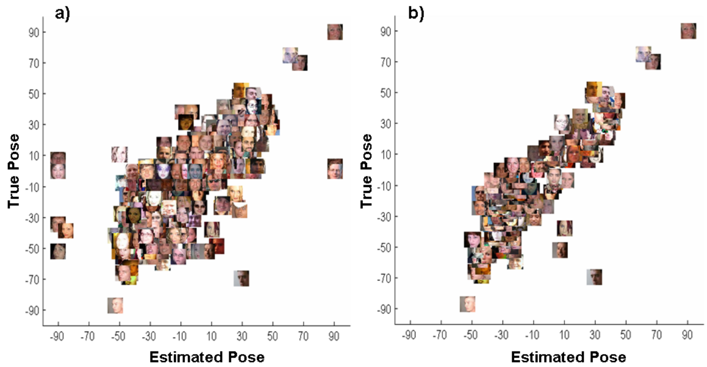

Figure 3: The scatter plot of the results (a) for all of the test data. We achieve a PCC of 0.76 and a MAE of 13.21 degrees and (b) for a subset of the test data uniformly sampled at each true pose. We achieve a PCC of 0.88 and a MAE of 11.72 degrees. The x-axis represents the estimated pose and the y-axis represents the true pose.

where we have assumed that the test patches $\mathbf{y}_p$ are independent. The term β represents pose, $\mathbf{w}_{p\bullet}$ are the parameters of the model and the variable l^* is the site in the library that most closely matches the test patch $\mathbf{y}_p$. The notation $\bullet$ indicates all of the values that an index can take, so $\mathbf{w}_{p\bullet}$ denotes the parameter vector associated with the p^{th} patch in the test image and all of the sites in the library. To find the site l^* in the library we assume that the test patch is a Gaussian corruption of the library patch and use

$$l^* = \arg\max_{l} \mathscr{G}_{\mathbf{y}_p}[\mathscr{L}_l; \sigma^2\mathbf{I}] \tag{2}$$

where $\mathscr{L}_l$ is the patch from site l of the library $\mathscr{L}$.

The likelihood term in Equation 1 has the form of a multinomial distribution, and we choose a uniform prior over pose. The parameters $\mathbf{W}$ are learned from training examples using multi-class logistic regression which maps the continuous pose space to the discrete space of the library sites. In practice we used radial basis functions (RBF) of pose. To find the best pose we do a one dimensional line search on pose varying from $-90°$ to $90°$ and estimate the pose by maximizing the energy function that is set to be the posterior probability over pose i.e. $Pr(\beta|\mathbf{Y},\mathbf{W})$ in Equation 1.

We harvested a large database (tens of thousands) of "real world" face images from the web. These images were captured in uncontrolled environments and exhibit wide variation in illumination, scale, expression and pose varying from $-90°$ to $90°$. We test our algorithm for pose estimation on this challenging database. Four human subjects were asked to label this database for pose. The labelled poses are averaged over the subjects to obtain an average human estimate to compare against. The performance of our algorithm is evaluated by Pearson correlation coefficient (PCC) and absolute error averaged across all test images (MAE). Figure 3 shows the scatter plot of the results. We achieve a PCC of 0.76 and a MAE of 13.21 on 1000 test images (Fig. 3a). We achieve a higher PCC of 0.88 for the uniformly sampled test set with a lower MAE of 11.72 (Fig. 3).

We have proposed a probabilistic method that exploits large databases for pose estimation. Our model uses generic patch-based representation. Therefore it can be used for regression problems on other object classes without major alteration. In future work we intend to investigate this representation for localization and segmentation.

[1] VN Balasubramanian, J. Ye, and S. Panchanathan. Biased Manifold Embedding: A Framework for Person-Independent Head Pose Estimation. In *CVPR*, pages 1–7, 2007.

[2] A.A. Efros and W.T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH*, pages 341–346, 2001.

[3] N. Kruger, M. Potzsch, T. Maurer, and M. Rinne. Estimation of face position and pose with labeled graphs. In *BMVC*, pages 735–743, 1996.