Blockwise Linear Regression for Face Alignment

Enrique Sánchez-Lozano[1]
esanchez@gts.uvigo.es
Enrique Argones-Rúa[2]
eargones@gradiant.org
José Luis Alba-Castro[1]
jalba@gts.uvigo.es

[1] Multimedia Technologies Group
AtlantTIC Research Center
University of Vigo
Vigo, Spain

[2] Multimodal Information Area
Gradiant
Vigo, Spain

In this paper, we present a deeper analysis of linear regression as an efficient method for face alignment. Linear regression for face alignment [2] consists of learning a mapping matrix between image features and shape parameters displacements. Typically, shape parameters are projections of a set of points which follow a Point Distribution Model onto the subspace generated by applying Principal Component Analysis to a set of manually landmarked training images. Often, image features are extracted within patches around each point, and concatenated into a column vector $\mathbf{d}$. We call the features belonging to the i-th patch with $\mathbf{d}^i$. We can generate examples for each image j by systematically perturbing the ground-truth parameters $\mathbf{p}_0^j$ with $\delta\mathbf{p}$. If we rearrange the training features into the matrix $\mathbf{D}$, and their corresponding perturbations into the matrix $\mathbf{P}$, the mapping matrix is obtained through least squares:

$$\mathbf{R} = \min_{\mathbf{R}} \|\mathbf{P} - \mathbf{R}\mathbf{D}\|_F^2 = \mathbf{P}\mathbf{D}^T(\mathbf{D}\mathbf{D}^T)^{-1}. \quad (1)$$

A closer look to Eqn. (1) reveals what is inside $\mathbf{D}\mathbf{D}^T$. Let us consider each sample patch i, from the k-th training image, $\mathbf{d}_k^i$, as a contribution of three terms: the mean feature vector (intrinsic for all faces) $\mathbf{d}_0^i$, the subject-dependent features $\widetilde{\mathbf{d}}_k^i$, and a sample noise (which may be produced by the landmarking process) ε_k^i, which is assumed zero-mean gaussian. Then, $\mathbf{D}\mathbf{D}^T$ holds all the cross-products between patches along the training set. However, the main contribution of these products comes from the mean part. Fig. 1 (left) shows an example of this matrix. Let us have a closer look to the covariance matrix (right), which consists of the outer products of the subject-dependent features and the sample noise, since they can not be separated in practice. As can be seen, most of the information lie on the diagonal, although some remaining information is found outside it. Thus, we propose to find which patches are related, and cluster them, in order to remove all the remaining noise. Once patches are clustered, and the noise is removed, a "clean" mapping matrix is obtained.

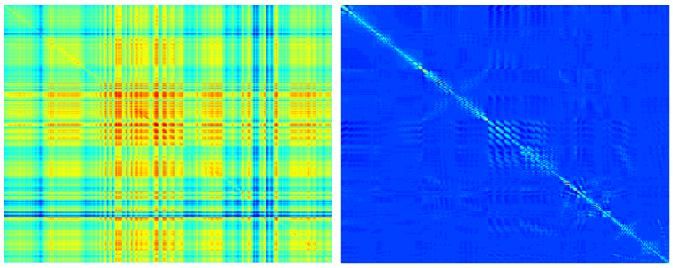

Figure 1: **Left:** Cross products between samples across the training set (best viewed in color). **Right:** Covariance matrix.

Correlation: For measuring the relation between two patches, we define the *modified Pearson correlation coefficient* as:

$$\gamma^{ij} = \frac{\|(\widetilde{\mathbf{D}}^i)(\widetilde{\mathbf{D}}^j)^T\|_F^2}{\|(\widetilde{\mathbf{D}}^i)(\widetilde{\mathbf{D}}^i)^T\|_F \|(\widetilde{\mathbf{D}}^j)(\widetilde{\mathbf{D}}^j)^T\|_F}, \quad (2)$$

where $\widetilde{\mathbf{D}}^i$ are the zero-mean features belonging to patch i. $\gamma^{ij} \in [0,1]$ is a measure of correlation between structures of independent variables. Two patches are related if $\gamma^{ij} \geq \theta$, where θ is a threshold.

Clustering: When two patches are related with a third one, but not with themselves, they must be clustered together. Two patches are clustered together if, and only if, there exists some patch that is related with both patches. If $\mathcal{P}$ denotes a patch, the clustering operation can be summarized as:

$$P^i \sim P^j \iff \{\exists P^h \in \mathcal{P} \,|\, \gamma^{ih} \geq \theta\,,\, \gamma^{ih} \geq \theta\}. \quad (3)$$

A fast way to see whether two patches are related or not, as well as whether two patches belong to the same cluster is through a binary matrix $\mathbf{Z}$, where the element (i,j) denotes whether two patches (i and j) meet $\gamma^{ij} \geq \theta$ (1), or not (0). An algorithm for clustering the data from the matrix $\mathbf{Z}$ and some examples can be found in the paper.

Regression: Once clusters are obtained, a regression matrix is learned for a modified version of each one. Instead of working with the whole matrices, we remove non-related products, by working with $z^{ij}\widetilde{\mathbf{D}}^i\widetilde{\mathbf{D}}^j$ (thorough mathematics can be found in the paper). Let $\widetilde{\mathcal{D}}_l$ be the features belonging to the l-th cluster, and $\widetilde{\mathcal{M}}_{ll,\theta} = \{z^{ij}\widetilde{\mathbf{D}}^i\widetilde{\mathbf{D}}^j | P^i, P^j \in \mathcal{C}_l\}$. Each regression matrix is then obtained as follows:

$$\mathcal{R}_l = \mathbf{P}(\widetilde{\mathcal{D}}_l)^T(\widetilde{\mathcal{M}}_{ll,\theta})^{-1}. \quad (4)$$

Fitting: Fitting is now calculated as follows. Consider the input data $\mathbf{d}$, which is divided into c clusters $\mathbf{d}_l$, and the mean feature vector $\mathbf{d}_0$, also divided into clusters $(\mathbf{d}_{10} \ldots \mathbf{d}_{c0})$. Shape parameters are now computed as

$$\delta\mathbf{p} = \sum_{l=1}^{c} \mathcal{R}_l(\mathbf{d}_l - \mathbf{d}_{l0}), \quad (5)$$

Experiments: We have trained our algorithm using a mixture of the Multi-PIE [3] and LFPW [1] databases, and tested it on the BioID [4]. We have measured the typical me_{17} error, by combining the eyes detection with the fitting done with our algorithm. Fig. 2 shows the results obtained for different values of θ. As can be seen, the best performance occurs when considering some relations among patches, although it may vary depending on the database.

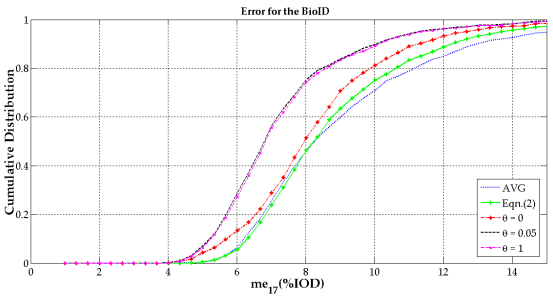

Figure 2: Fitting results obtained for the BioID database, varying θ.

[1] P.N. Belhumeur, D. Jacobs, D.J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*, 2011.

[2] T.F. Cootes and C.J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(6): 680–689, 2001.

[3] R. Gross, I. Matthews, J.F. Cohn, T. Kanade, and S. Baker. Multi-pie. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG'08)*, 2008.

[4] O. Jesorsky, K. Kirchberg, and R. Frischholz. Robust face detection using the hausdorff distance. In *Int'l Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA'01)*, pages 90–95, 2001.