Boosting Pseudo Census Transform Features for Face Alignment

Hua Gao
gao@kit.edu

Hazım Kemal Ekenel
ekenel@kit.edu

Mika Fischer
mika.fischer@kit.edu

Rainer Stiefelhagen
rainer.stiefelhagen@kit.edu

Institute for Anthropomatics
Karlsruhe Institute of Technology
Karlsruhe, Germany

Face alignment using deformable face model has attracted broad interest in recent years for its wide range of applications in facial analysis. Previous work has shown that discriminative deformable models have better generalization capacity compared to generative models [3]. In this paper, we present a new discriminative face model based on boosting pseudo census transform features. This feature is considered to be less sensitive to illumination changes, which yields a more robust alignment algorithm. The alignment is based on maximizing the scores of boosted strong classifier, which indicate whether the current alignment is a correct or incorrect one. The proposed approach has been evaluated extensively on several databases. The experimental results show that our approach generalizes better on unseen data compared to the Haar feature-based approach. Moreover, its training procedure is much faster due to the low dimensional configuration space of the proposed feature.

In [3], a discriminative appearance model is built based on boosting Haar features. However, as we know that the number of Haar features to be boosted is extremly large since the dimension of the parameter space is high. Training a model using Haar features requires to boost more than one hundred thousand rectangular features within the mean shape, which results in a very inefficient training procedure. To avoid this, we propose to use a local feature with less configurable parameters for boosting, which enables the training procedure extremely fast. The local feature is inspired by the work of Fröba *et al.* [2], in which the modified census transformation (MCT) is applied for face detection. However, the MCT feature is a binarized pattern which is not suitable for deriving an analytical optimization algorithm. In this work, we used the unbinarized census transform feature, which we call pseudo census transform (PCT). The PCT feature is projected discriminatively to a scalar indicating the correctness of face alignment. We boost the scalar values using GentleBoost [1]. Multi-scale PCT features are also investigated.

In the proposed face model, we use a 2D generative shape model to describe the distribution of the face shapes. Assuming the face shapes lie in a linear subspace, we represent a novel shape $\mathbf{s}$ with a linear combination of shape basis: $\mathbf{s} = \mathbf{s}_0 + \Sigma_{i=1}^{n} p_i \mathbf{s}_i$, where $\mathbf{s}_0$ is the mean shape, $\mathbf{s}_i$ is the i-th shape basis, and $\mathbf{p} = [p_1, p_2, \ldots, p_n]^\top$ is the shape parameter. A nonlinear mapping function $\mathbf{W}(\mathbf{x};\mathbf{p})$ is defined which maps pixel $\mathbf{x}$ defined in an instance shape to the mean shape. A shape-free image $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$ is obtained after warping a face image $\mathbf{I}$.

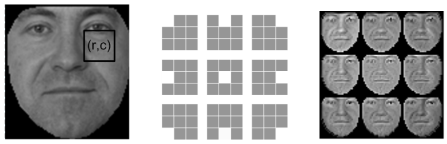
The appearance model is a collection of m PCT features computed over the shape-free face image $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$. The PCT feature is a K dimensional vector $\varphi = (\varphi_1, \ldots, \varphi_K)^\top$, which contains the pixel values in a $\sqrt{K} \times \sqrt{K}$ neighborhood centered at $\mathbf{x} = (r, c)$, and subtracted with local mean. For simplicity, we used a fixed K ($K = 9$) in this work. The PCT feature φ is then obtained by ordering the K filter responses of a filter bank plotted in Figure 1(b) at a specific position (r, c). The mask of the first filter can be defined as follows:

$$\mathbf{T}_0 = \begin{pmatrix} 8/9 & -1/9 & -1/9 \\ -1/9 & -1/9 & -1/9 \\ -1/9 & -1/9 & -1/9 \end{pmatrix} \qquad (1)$$

Note that the responses of the filters are equivalent to the PCT feature values. This enables us to define K image templates $\mathbf{A}_{i=1,\ldots,K}$ with the filter mask placed at position $\mathbf{x} = (r, c)$ for one PCT feature. The inner product between the template and the warped image is equivalent to computing the filter responses:

$$\varphi_i = \mathbf{A}_i^\top \mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p})) = \mathbf{T}_i * \mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p})), i = 1, \ldots, K. \qquad (2)$$

We are interested in learning from data a score function F, such that, when maximized with respect to $\mathbf{p}$, it will return the shape parameter



(a) (b) (c)

Figure 1: (a) The parametrization of a weak classifier, i.e. center of the PCT filter positioned at (r, c); (b) K PCT filter masks ($K = 9$), the top left filter mask correspond to the filter kernel defined in Equation (1); (c) PCT-filter responses of a shape-free image.

corresponding to the correct alignment. Mathematically, if $\mathbf{p}^*$ is the shape parameter representing the correct alignment, F has to be such that

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} F(\mathbf{p}) \qquad (3)$$

With this formulation, the appearance model is actually a two-class classifier. In particular, we use a linear combination of several PCT features to define the appearance model:

$$F(\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))) = \Sigma_{m=1}^{M} f_m(\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))) \qquad (4)$$

where $f_m(\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p})))$ is a function operating on one PCT feature of the shape-free face image $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$, which is considered as a weak classifier. We use the GentleBoost algorithm to boost M of the weak classifiers. For simplicity the weak classifier is denoted as $f_m(\mathbf{p})$, and it is defined as follows:

$$f_m(\mathbf{p}) = \frac{\pi}{2} atan(\Sigma_{i=1}^{K} w_i^m S(\mathbf{A}_i^{m\top} \mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))) + b^m), \qquad (5)$$

where $\mathbf{A}_i^m$ is the i-th template defined at the m-th position (r_m, c_m). Since the classifier response $f_m(\mathbf{p})$ is continuous within -1 and 1, the $atan()$ function is used to ensure both discriminability and derivability. $S(\bullet)$ is a sigmoid function defined as $S(t) = \frac{1}{1+e^{-\alpha t}}$, where α is a scale parameter. The sigmoid function maps the raw PCT feature values into a range of $(0, 1)$ before the linear projection defined by the projection vector $\mathbf{w}^m$ and bias b^m. The projection vector $\mathbf{w}^m$ and bias b^m are learned on the training data with binary linear support vector machines (SVM).

To obtain more discriminative information for face alignment we also boost PCT features on images at different scales. The optimization of Equation 3 is solved by using the gradient ascent method. The details of the optimization alogrithm is decribed in the paper. The reason why it is solvable using the gradient ascent method is also stated in the experiment part. Evaluation on different data sets shows the improved robustness on unseen data.

We conclude that the proposed PCT feature enables extremely fast appearance model training compared to the training procedure of using Haar feature. Moreover, as the PCT feature is less sensitive to illumination changes, the discriminative appearance models based on it have better generalization capacity.

[1] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 38(2): 337–374, 2000.

[2] B. Fröba and A. Ernst. Face detection with the modified census transform. In *Proc. of* 6^{th} *Int. Conf. on Automatic Face and Gesture Recognition*, pages 91–96, 2004.

[3] X. Liu. Generic face alignment using boosted appearance model. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'07)*, pages 1–8, 2007.