# Face Alignment Using a Ranking Model based on Regression Trees

Hua Gao[1]
gao@kit.edu
Hazım Kemal Ekenel[1,2]
ekenel@kit.edu
Rainer Stiefelhagen[1]
rainer.stiefelhagen@kit.edu

[1] Institute for Anthropomatics
Karlsruhe Institute of Technology
Karlsruhe, Germany

[2] Faculty of Computer and Informatics
Istanbul Technical University
Istanbul, Turkey

In this work, we exploit the regression trees-based ranking model, which has been successfully applied in the domain of web-search ranking, to build appearance models for face alignment. The model is an ensemble of regression trees which is learned with gradient boosting. The MCT (Modified Census Transform) [1] as well as its unbinarized version PCT (Pseudo Census Transform) [2] are used as features due to their robustness to illumination changes. To avoid the overfitting problem and ensure quick convergence in gradient boosting, we use random trees to initialize the boosting. The Nelder Mead's simplex method is applied for fitting the learned model. We compare the proposed regression trees-based point-wise ranking model to pairwise ranking model. Experiments show that the proposed model improves both robustness and accuracy for face alignment.

Classification-based boosted appearance model using the PCT features has been proposed in [2]. However, the model has its own drawback as the positive and negative training samples are highly imbalanced. Furthermore, the learned score function does not guarantee smoothness and concavity in the neighborhood of real solution. Optimizing such a score function with a local optimizer is prone to local maxima. In [4, 5], Ranking-based Appearance Models (RAM) are investigated by boosting the score function in a pairwise ordinal classification way. This model ensures that the score function returns a higher value if the current alignment is closer to the ground truth than the others in the shape parameter space. A local optimizer benefits from such a model as the gradient of the learned score function is constrained to the same direction towards the ground truth.

Based on this idea, we propose and compare two ranking-based appearance models in this work. Both models use a generative shape model assuming 2D 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 non-linear 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 (Figure 1(a)) $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$ is obtained after warping a face image $\mathbf{I}$ given shape parameter $\mathbf{p}$.

The first ranking appearance model learns a ranking function via pairwise ordinal classification as proposed in [4]. However, in this work, we apply the pairwise RankSVM [3] on the PCT features to build weak rankers (Eq. 1):

$$f_m(\mathbf{p}) = \frac{1}{\pi} atan(\mathbf{w}^{m\top} S(\varphi^m) - t^m). \tag{1}$$

A PCT feature vector $\varphi^m$ is extracted at a particular location in the masked shape-free image. The function $atan()$ is used to ensure both discriminability and derivability. The $S(\cdot)$ is a sigmoid function, which normalizes the values in a raw PCT feature into a range of $(0,1)$. The projection vector $\mathbf{w}^m$ is learned with RankSVM. The threshold $t^m$ is determined with weighted least square estimation. The final strong ranking function (Eq. 2) is combined by boosting weak rankers with gentleboost:

$$F(\mathbf{p}) \doteq \textstyle\sum_{m=1}^{M} f_m(\mathbf{p}). \tag{2}$$

Fitting a learned model to a novel image is done by maximizing this score function with respect to the shape parameters using gradient ascent method.

Point-wise ordinal regression based on gradient boosted regression trees (GBRT) has gained much attention for solving ranking problem in information retrieval domain. We apply GBRT in the second ranking appearance model. The GBRT iteratively fits regression trees of certain depth to regression residues, which results in less biased estimation. As the learned ranking function is not differentiable anymore due to the hierarchical tree structure, we apply both PCT and MCT as our feature representation. Usually a small shrinkage is used for training GBRT to avoid overfitting. However, setting small shrinkage leads to slow convergence which results in enormous number of trees in the strong ranking function. This problem can be solved by initializing the GBRT properly so that the initial estimation is already close to the regression target.

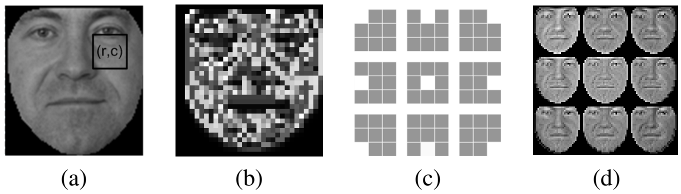

Figure 1: (a) A shape-free face image; (b) MCT output of a shape-free image; (c) 9 PCT filter masks; (d) PCT-filter responses of a shape-free image.

We use random forest as an initial estimation for the iterative GBRT training. The random forest is a combination of bagging and random feature selection. This leads to a low variance estimation and less sensitive to noise and outliers. We denote this initialized GBRT as iGBRT. The output of the final boosted ranking score function is actually the response of RF combined with the boosted regression trees:

$$T(\mathbf{p}) = F(\mathbf{p}) + \alpha \sum_{t=1}^{M_B} h_t(\mathbf{p}). \tag{3}$$

Where $F(\cdot)$ is the initial estimation from random forest. $h_t(\cdot)$ is a boosted weak regression tree. $\alpha$ is the shrinkage and $M_B$ is the total number of boosted trees.

Face alignment is equivalent to maximizing Equation 3 with the constraint of the shape prior. We define the cost function as follows:

$$O(\mathbf{p}) = -T(\mathbf{p}) + \beta \sum_{i=0}^{n} \frac{p_i^2}{\lambda_i}, \tag{4}$$

where $\beta$ is the parameter that we estimated from the training data. $\lambda_i$ is the eigenvalue corresponding to shape parameter $p_i$. As it is difficult to derive the analytical gradient for the cost function, we use the Nelder-Mead simplex method to minimize Equation 4 which only requires the evaluation of the cost function.

The details of the learning the ranking appearance models are described in the paper, including data sampling and parameter setting for training the models. We evaluated the alignment on different data sets and observed that ranking-based appearance models improve the fitting performance over the classification-based model. The iGBRT model outperforms the pairwise ranking model as well as the original GBRT-based model.

[1] 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.

[2] H. Gao, H. K. Ekenel, M. Fischer, and R. Stiefelhagen. Boosting pseudo census transform features for face alignment. In *Proc. of BMVC, Dundee, UK*, 2011.

[3] R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. *Advances in Large Margin Classifiers*, pages 115–132, 2000.

[4] H. Wu, X. Liu, and G. Doretto. Face alignment via boosted ranking model. In *Proc. of CVPR*, 2008.

[5] J. Zhang, S. K. Zhou, D. Comaniciu, and L. McMillan. Discriminative learning for deformable shape segmentation: A comparative study. In *Proc. of ECCV*, 2008.