21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

A Ranking Model for Face Alignment with Pseudo Census Transform

Hua Gao[1], Hazım Kemal Ekenel[1,2], Rainer Stiefelhagen[1]
[1]*Institute for Anthropomatics, Karlsruhe Institute of Technology, Germany*
[2]*Faculty of Computer and Informatics, Istanbul Technical University, Turkey*
{*gao,ekenel,rainer.stiefelhagen*}*@kit.edu*

Abstract

We extend the PCT (Pseudo Census Transform)-based appearance model [3] to ranking-based appearance model for face alignment. The PCT-based weak ranking function is learned using RankSVM, and the ranking appearance model (RAM) is constructed in a boosting manner. Experiments show that the PCT-based RAM is more robust and generalize better than the PCT-based boosted appearance model (BAM). The PCT-RAM achieves about 23% *improvement when tested on unseen data. We also investigate different sampling strategies for the learning to rank problem and find out that random permutation achieves similar results as using adjacent ordering pairs. The alignment results do not decrease significantly when only one ordinal pair is used for each direction.*

1 Introduction

Face image registration is an essential step for further facial analysis such as identification, expression recognition, age estimation, etc. Among various techniques, alignment using deformable model has been attracted the researchers since the invention of Active Shape Model (ASM) [2] and Active Appearance Model (AAM) [1]. Numerous successful application systems have been developed based on the deformable model. Despite of that, issue of generalization on unseen data or unmatched condition is still an open problem, due to variation factors such as illumination, expression, occlusion and image quality.

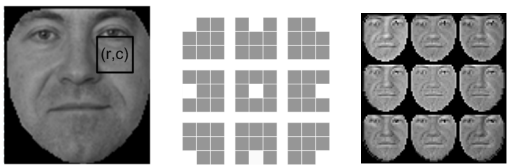
As one of the early deformable models, the ASM models the distribution of target's shape and profile texture. An important extension of the ASM is the AAM [1], in which the texture inside the shape convex hull is modeled as appearance of face. The model combines constraints on both shape and texture by learning generative statistical models. However, as claimed and demonstrated in [7], AAM suffers from generalization problem due to generative appearance modeling.

Several attempts have been proposed to tackle the generalization problem. Most solutions tend to build discriminative appearance models to replace the generative model. For example, [11] learn discriminative appearance model via boosted regression. While in [7], the author proposed a boosted appearance model (BAM) based on boosting weak classifiers using Haar feature. The resulting discriminative model is able to distinguish between correct and incorrect alignment. Fitting a BAM is done by maximizing the strong classifier score function subject to the model shape parameters. This model is further extended in [3], in which the PCT feature is used for boosting a more robust appearance model against illumination changes.

Boosting discriminative models based on classification has its own drawback as the positive and negative samples are highly imbalanced. Furthermore, the resulting score function does not guarantee of smoothness and concaveness in the neighborhood of the real solution. Optimizing such a score function with local optimizer is prone to local maxima. In [12, 13], ranking based appearance models are investigated by boosting the score function in an ordinal regression way. This model ensures that the score function returns higher value, if the current alignment is closer to the ground truth than the others in the shape parameter space. Local optimizer benefits from such model as the gradient of the learned score function is constrained to be the same as the direction towards the ground truth.

In this work, we learn a ranking-based appearance model using the PCT features due to its robustness to illumination changes and fast training procedure. Different sampling strategies are investigated for the learning to rank problem. We find out that random permutation performs similarly to the method using adjacent ordering pairs. Experiments show that face alignment using the PCT-based RAM is more robust than the PCT-based BAM. It is also interesting to observe that boosting with only one random ordinal sample per direction already performs as good as using all adjacent pairs.

978-4-9906441-1-6



(a) (b) (c)

Figure 1: (a) A shape-free face image; (b) 9 PCT filter masks, the top left filter mask corresponds to the kernel defined in Eq. (1); (c) PCT-filter responses of a shape-free image.

2 Methodology

2.1 Shape Model

We use a linear shape model to describe the distribution of the shape of faces: $\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i$, where s is a shape vector represented with a set of landmarks by stacking the coordinates of each. $\mathbf{s}_0$ 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. The mean shape and the shape basis can be learned from a labeled training set of face images via Principal Component Analysis (PCA).

2.2 Appearance Model based on PCT Feature

With Delauney triangulation, the mean shape $\mathbf{s}_0$ and the shape $\mathbf{s}$ are triangulated to a base mesh and an instance mesh. A non-linear mapping function $\mathbf{W}(\mathbf{x}; \mathbf{p})$ is defined with a piece-wise affine warping, which maps pixel $\mathbf{x}$ defined in the instance shape to the mean shape. A shape-free image $\mathbf{I}(\mathbf{W}(\mathbf{x}; \mathbf{p}))$ (Figure 1(a)) is obtained by warping a face image $\mathbf{I}$ with such a warping.

The appearance model is a collection of m features computed over the shape-free face image $\mathbf{I}(\mathbf{W}(\mathbf{x}; \mathbf{p}))$. We use the PCT feature [3] for building our appearance model due to its robustness to lighting variations. The PCT feature $\varphi = (\varphi_1, \ldots, \varphi_K)^\top$ is a K dimensional vector extracted from the pixel values in a $\sqrt{K} \times \sqrt{K}$ neighborhood centered at $\mathbf{x} = (r, c)$, and subtracted with local mean. We used a fixed K ($K = 9$) as in [3]. The PCT feature φ is obtained by ordering the K filter responses of a filter bank plotted in Figure 1(b) at position (r, c). The mask of the first filter is defined as:

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

The rest of the filter masks are defined accordingly by shifting the position of the value $8/9$ in the matrix (see Figure 1 (b), white corresponds to the positive element and gray corresponds to the negative elements). Note that the responses of the filters are equivalent to the PCT feature values. This enables us to define K image templates $\mathbf{A}_{k=1,\ldots,K}$ 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:

$$\varphi_k = \mathbf{A}_k^\top \mathbf{I}(\mathbf{W}(\mathbf{x}; \mathbf{p})) = \mathbf{T}_k * \mathbf{I}(\mathbf{W}(\mathbf{x}; \mathbf{p})), k = 1, \ldots, K. \quad (2)$$

2.3 Ranking Model for Face Alignment

A ranking model is considered to be a good option to learn a local maximum free objective function. The model suppose to return higher value, if the corresponding shape parameter is closer to the ground truth than the other one:

$$F(\mathbf{p}_2) > F(\mathbf{p}_1) \Longleftrightarrow \mathbf{p}_2 \succ \mathbf{p}_1, \quad (3)$$

where $\mathbf{p}_2 \succ \mathbf{p}_1$ means $\mathbf{p}_2$ is superior to $\mathbf{p}_1$ or $\|\mathbf{p}_2 - \mathbf{p}_0\| < \|\mathbf{p}_1 - \mathbf{p}_0\|$, $\mathbf{p}_0$ corresponds to the shape parameter of ground truth. Eq. (3) ensures that the learned ranking model is a unimodal objective function with its maximum located exactly at $\mathbf{p}_0$.

In [3], linear SVMs are applied as weak classifiers. To learn individual weak rankers, we adopt RankSVM [4] with linear kernel. We used the implementation in [6] for training weak rankers. The RankSVM is formulated as a margin bound ordinal regression problem, which tries to maximize the soft margin with constrained quadratic programming:

$$\min \|\vec{w}\|^2 + C \sum_{\ell=1}^{N} \xi_\ell \quad (4)$$

$$\text{s.t. } \xi_\ell \geq 0, z_\ell \left\langle \vec{w}, \vec{x_\ell}^{(1)} - \vec{x_\ell}^{(2)} \right\rangle > 1 - \xi_\ell, \quad (5)$$

where z_ℓ is a label defined as:

$$z_\ell = \begin{cases} +1 & \vec{x_\ell}^{(1)} \succ \vec{x_\ell}^{(2)} \\ -1 & \vec{x_\ell}^{(1)} \prec \vec{x_\ell}^{(2)} \end{cases}$$

As with in [12], we applied Gentleboost for boosting weak rankers. Eq. (3) suggested that the ranking function F can be formulated as a classification problem. More precisely, if we define a classifier $H(\mathbf{p}_1, \mathbf{p}_2) = \text{sign}[F(\mathbf{p}_2) - F(\mathbf{p}_1)]$, then $H(\mathbf{p}_1, \mathbf{p}_2) = +1$ if $\mathbf{p}_2 \succ \mathbf{p}_1$, else $H(\mathbf{p}_1, \mathbf{p}_2) = -1$. Note that here we ignore the tie case. The classifier H implies whether or not switching from $\mathbf{p}_1$ to $\mathbf{p}_2$ constitutes an alignment improvement. In the boosting framework, we assume H to be an additive model: $H = \sum_{m=1}^{M} h(\mathbf{p}_1, \mathbf{p}_2)$, where $h_m(\mathbf{p}_1, \mathbf{p}_2) = f_m(\mathbf{p}_2) - f_m(\mathbf{p}_1)$. f_m is the m-th weak ranking function, which is defined as:

$$f_m(\mathbf{p}) = \frac{1}{\pi} atan(\mathbf{w}^{m\top} S(\varphi^m) - t^m). \quad (6)$$

Since the weak ranking function $f_m(\mathbf{p})$ is continuous within $(-0.5, 0.5)$, the $atan()$ function is used to ensure both discriminability and derivability. The $S()$ is

Algorithm 1: PCT-RAM Learning

Data: Training samples, with labels $\{z_\ell = +1\}$
Result: The alignment score function F

Initialize the weights $w_\ell = \frac{1}{N}$ and the score function $F = 0$
foreach *m=1,..., M* **do**
  Fit f_m with weighted least squares, such that

$$f_m = \arg\min_f \sum\nolimits_\ell w_\ell (z_\ell - h(\mathbf{x}_\ell))^2 \quad (8)$$

  where $h(\mathbf{x}_\ell) = f(x_\ell^{(1)}) - f(x_\ell^{(2)})$
  $F \longleftarrow F + f_m$
  $w_\ell \longleftarrow w_\ell \exp(-z_\ell h_m(\mathbf{x}_\ell))$
  Normalize the weights such that $\sum_\ell w_\ell = 1$
return $F = \sum_{m=1}^{M} f_m$

a sigmoid function, which normalizes the raw PCT feature values into a range of $(0, 1)$. The linear projection vector $\mathbf{w}^m$ is learned with RankSVM. The threshold t^m needs to be determined during boosting. The strong ranking function is again assumed to be an additive model:

$$F(\mathbf{p}) \doteq \sum\nolimits_{m=1}^{M} f_m(\mathbf{p}). \quad (7)$$

To learn the strong ranking function F, we sample ordering pairs from a training dataset containing D facial images with annotated landmarks. For each of the training image, we randomly perturb the ground truth $\mathbf{p}_i$ in U different directions $\{\Delta\mathbf{p}_{iu}\}_{u=1,\ldots,U}$. In each direction, we evenly sampled V shape parameters $\{\mathbf{p}_i + v \times \Delta\mathbf{p}_{iu}\}_{v=1,\ldots,V}$. For each direction, we can generate V ordinal adjacent pairs using the samples including the ground truth. In total $N = D \times U \times V$ ordinal pairs are generated from the training set. We denote each of the pairs as $\{\mathbf{x}_\ell = (x_\ell^{(1)}, x_\ell^{(2)}\}_{\ell=1,\ldots,N}$, where $x_\ell^{(1)} \succ x_\ell^{(2)}$ and their corresponding label $z_\ell = +1$. In addition to use adjacent ordinal pairs as training samples, we also employ another sampling strategy by randomly generating R ordinal pairs out of $V + 1$ samples in each direction. The boosting procedure is summarized in Algorithm 1. Eq. (8) denotes that in each iteration a weak ranking function f_m is found by fitting weighted least squares. Fitting the learned model to a novel image is done by maximizing the score function (Eq. 7) in the sense of gradient ascent.

3 Experiments

The images for evaluating the proposed method are collected from multiple publicly available databases, including the FRGC v2.0 database [8], the FERET database [9], the IMM database [10], and the Labeled Faces in the Wild (LFW) database [5]. The collected images are partitioned distinctively into four subsets. Table 1 lists the properties of each database and partition. We use Set 1 as training set to build a ranking appearance model and test the model fitting on all four datasets. This setting ensures that we have two levels of generalization to be tested, i.e., Set 2 is tested as the unseen data of seen subjects; Set 3 and 4 are tested as the unseen data of unseen subjects. There are 58 manually labeled landmarks for each of the 1529 images. The images are down-sampled such that the facial width is roughly 40 pixels across the set.

Table 1: Summary of the dataset.

	FRGC	FERET	IMM	LFW
Images	589	200	240	500
Subjects	200	200	40	500
Variation	Fron., expr.	Pose	Pose, expr.	All
Set 1	200	200		
Set 2	389			
Set 3			240	
Set 4				500

For the first experiment, we compare the proposed PCT-RAM to PCT-BAM [3]. Using Set 1, we train a shape model with 15 components preserving 95% of shape variations. The size of shape-free images has 30×30 pixels. For each image we select $U = 10$ randomly perturbed directions and in each direction $V = 6$ positions are evenly sampled. Including the position at ground truth, in total 6 adjacent ordinal pairs can be generated. The overall training sample includes $N = 24000$ $(400 \times 10 \times 6)$ ordinal pairs. The resulting ranking appearance model learn 100 weak rankers.

In testing, we randomly perturb ground truth landmarks at different Gaussian noise levels for initializing each alignment. We repeat the random perturbation for each noise level multiple times on each test image in order to perform a statistical evaluation of the result. A fitting is considered as converged if the Root Mean Square Error (RMSE) between the aligned landmarks and the ground truth is less than one pixel. The Average Frequency of Convergence (AFC) is used as an evaluation metric, which assesses the robustness of the alignment. The metric AFC is calculated as the number of converged trials divided by the total number of trials. We apply the same termination condition for the fitting procedure as in [3].

Figure 2 plots the AFC rates of the PCT-RAM, PCT-BAM and MS-PCT-RAM (multi-scale model as in [3]) with respect to different levels of initial landmarks perturbation, computed over Set 1, 2, 3, and 4, respectively. Improvement on fitting robustness is clearly observed in these plots. The AFC rates increase about $8.5\%-22.7\%$ on different datasets at 1.6σ noise level. The most noticeable performance gain is the test on the Set 3, which

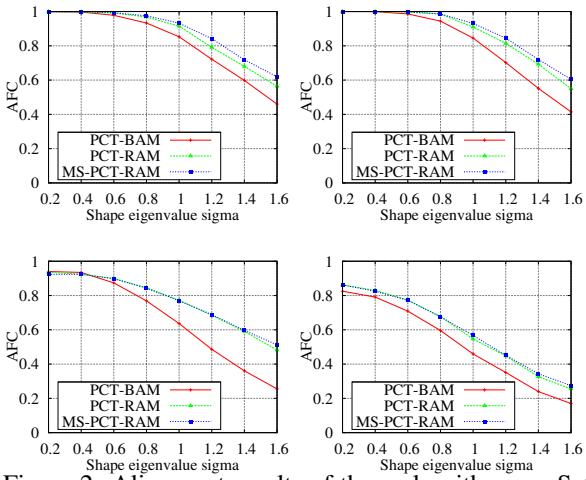


Figure 2: Alignment results of three algorithms on Set 1, Set 2 (first row) and Set 3, Set 4 (second row).

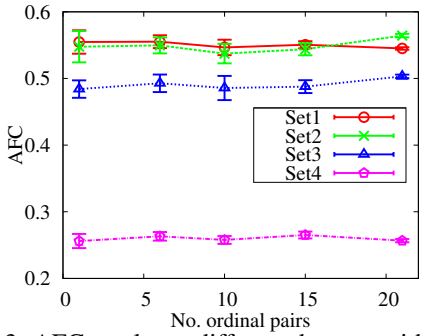


Figure 3: AFC results on different datasets with increasing number of randomly sampled ordinal pairs.

implies that the PCT-RAM has much better generalization ability than the PCT-BAM on unseen data of unseen subjects. To improve alignment on Set 4 is difficult probably due to the limitation of the shape model learned on the Set 1. The MS-PCT-RAM improves the alignment further as it learns additional features on multiple scales of the shape-free image.

In the second experiment, we found out that using random permutation of ordinal pairs as training samples for boosting works almost as good as using adjacent pairs. Figure 3 plots the AFC results at 1.6σ noise level with the models trained with $R = \{1, 5, 10, 15, 21\}$ random ordinal pairs per direction. The fitting performance of a model trained with only one ordinal pair per direction does not degenerate much. In fact, the mean AFC rate varies slightly with increasing number of ordinal training pairs used. However, the variance of the AFC rates decrease when R increases.

4 Conclusions

We investigated a deformable appearance model for face alignment based on boosting a strong ranking function. The function is an additive model which is composed of a set of weak rankers based on PCT features and are learned using RankSVM. Having the learned appearance model, we align the face model to novel face images by maximizing the ranking function with a local optimizer. We conducted experiments on four different datasets and the results show that our method is superior to the PCT-BAM. We further learn the ranking function based on random sampling of ordinal pairs. Experiments show that the robustness of fitting is in the same order as the methods of using adjacent pairs.

5 Acknowledgments

This study is funded by OSEO, French State agency for innovation, as part of the Quaero program; and the "Concept for the Future" of Karlsruhe Institute of Technology within the framework of the German Excellence Initiative.

References

[1] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. of ECCV*, volume 2, pages 484–498, 1998.

[2] T. F. Cootes and C. J. Taylor. Active shape models. In *Proc. of BMVC*, pages 266–275, 1992.

[3] 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.

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

[5] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[6] T. Joachims. Optimizing Search Engines Using Clickthrough Data. In *Proc. of SIGKDD*, pages 133–142, 2002.

[7] X. Liu. Generic face alignment using boosted appearance model. In *Proc. of CVPR*, pages 1–8, 2007.

[8] P. Phillips et al. Overview of the face recognition grand challenge. In *Proc. of CVPR*, pages 947–954, 2005.

[9] P. Phillips, H. Moon, P. Rauss, and S. Rizvi. The feret evaluation methodology for face recognition algorithms. *IEEE Trans. on PAMI*, 22(10):1090–1104, 2000.

[10] M. Stegmann, B. Ersboll, and R. Larsen. FAME - a flexible appearance modeling environment. *IEEE Trans. on Medical Imaging*, 22(10):1319–1331, 2003.

[11] P. A. Tresadern, P. Sauer, and T. F. Cootes. Additive update predictors in active appearance models. In *Proc. of BMVC, Aberystwyth, UK*, 2010.

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

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