# 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

**Abstract**

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) as well as its unbinarized version PCT (Pseudo Census Transform) are used as features due to their robustness to illumination changes. To avoid the overfitting problem 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 pointwise ranking model to pairwise ranking model. Experiments show that the proposed model improves both robustness and accuracy for face alignment.

## 1 Introduction

Facial image analysis has found its application in various fields including security, entertainment, multimedia indexing, human-machine interaction, etc. Essentially, as the first step for facial image analysis, face image registration (a.k.a. face alignment) has a large impact on the robustness and quality of the later processes. Face alignment has been studied for several decades, yet it is still an open problem which suffers from the confounding factors of intrinsic and extrinsic imaging conditions. Due to these challenges, face alignment is still an interesting research problem and receives increasing attention. In particular, deformable model based face alignment became very popular since the invention of the Active Shape Model (ASM) [3] and the Active Appearance Model (AAM) [4]. Numerous successful application systems have been developed based on the deformable model.

As one of the early deformable model, the ASM models the distribution of the target's shape and profile texture. An important extension of the ASM is the AAM [4], in which the texture inside the shape convex hull models the appearance of the face. The model combines constraints on both shape and texture by learning generative statistical models. However, generative appearance modeling in the AAM suffers from generalization problem as claimed and demonstrated in [11].


http://dx.doi.org/10.5244/C.26.118

The generalization problem of the deformable model fitting has been intensively studied in the community. Most solutions tend to build discriminative appearance models to replace the generative model. For example, Tresadern *et al*. [18] learn a discriminative appearance model via boosted regression. While Liu [11] proposes a Boosted Appearance Model (BAM) based on boosting weak classifiers using Haar features. The resulting BAM 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 [7], by using the PCT feature for boosting a more robust BAM against illumination changes.

Boosting discriminative models based on classification has its own drawback as the positive and negative samples are highly imbalanced. Furthermore, the learned score function does not guarantee smoothness and concavity in the neighborhood of the real solution. Optimizing such a score function with local optimizer is prone to local maxima. In [19, 21], 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.

In this work, we compare two ranking-based appearance models. The first ranking appearance model learns a ranking function via pairwise ordinal classification as proposed in [19]. However, we apply the pairwise RankSVM [8] over the PCT features to build weak rankers, and the final strong ranking function is combined by boosting weak rankers. The second model is proposed in which we formulate the ranking problem with regression trees. The gradient boosted regression trees (GBRT) are used to learn the appearance model as it achieves top results in the domain of web-search ranking [12]. To overcome the drawbacks of GBRT, *e.g*. prone to overfitting and slow convergence rate, we train Random Forests (RF) and use its outputs as the initial estimation for GBRT learning. In the second model, both the PCT features and the MCT features are used for appearance representation, as a derivative-free local optimizer is applied for face alignment. Experimental results show that the regression trees-based RAM achieves superior results than the pairwise ordinal classification model. The initialization step for GBRT learning results in a very robust face alignment, which improves the performance about $23.5\% - 35.6\%$ on different data-sets compared to the model based on pairwise ordinal classification.

The rest of the paper is organized as follows. We describe the face model representation in Section 2. The details of learning two RAMs are presented in Section 3. The experimental setup and results are discussed in Section 4, and we give concluding remarks in Section 5.

# 2 Face Model

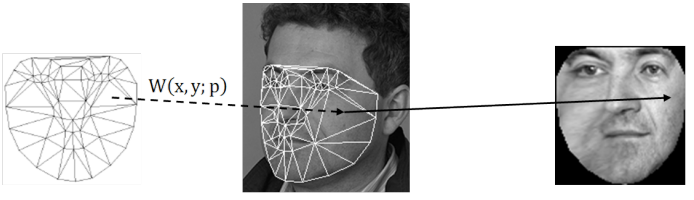
The presented models include a shape model and an appearance model. As with the conventional ASM and AAM, we use a generative shape model. However, the appearance model is constructed independently from the shape model in a discriminative manner.

## 2.1 Shape Model

A generative linear shape model is used to describe the distribution of the shape of faces:

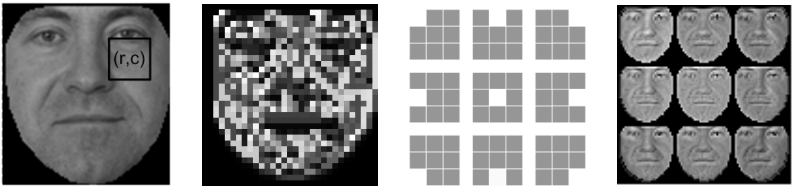
$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i.$$

This model is known as the Point Distribution Model (PDM) which was broadly applied




(a) (b) (c)

Figure 1: Shape model and warping function. (a) The mean shape $\mathbf{s}_0$. (b) A face image superimposed with a shape $\mathbf{s}(\mathbf{p})$. (c) A face image warped to the mean shape $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$.




(a) (b) (c) (d)

Figure 2: (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.

in statistical deformable models [3, 4]. Here $\mathbf{s}$ is a shape vector represented by a set of landmarks by stacking the coordinates of each. The landmarks, $\mathbf{x}_i = (x_i, y_i)_{i=1,\ldots,l}$, are placed over the fiducial facial feature points and the face contours. $\mathbf{s}_0$ defines the mean shape, and $\mathbf{s}_i$ represents the $i$-th shape basis. $\mathbf{p} = [p_1, \ldots, p_n]^\top$ is the shape parameter vector. The mean shape and the shape basis are learned from an annotated training set via Principal Component Analysis (PCA) after a normalization step using the Procrustes analysis.

## 2.2 Feature Representation for Appearance Model

In the AAM [4], the texture of a face image is represented as the image intensity inside the convex hull of the face shape. There are many methods for extracting the texture, among which piecewise affine warping is the simplest one. This approach applies the Delaunay triangulation on the mean shape $\mathbf{s}_0$ and the shape $\mathbf{s}$ and obtains a base mesh (Figure 1(a)) and an instance mesh (Figure 1(b)). A non-linear mapping function $\mathbf{W}(\mathbf{x};\mathbf{p})$ is defined which maps pixel $\mathbf{x}$ in the instance mesh to the base mesh. A shape-free image $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$ (Figure 1(c)) is obtained by warping a face image $\mathbf{I}$ with such a warping. We extract local features from the shape-free images to build our appearance models.

### 2.2.1 MCT Features

The local illumination variations are also modeled in the AAM. Scott *et al*. [16] enhanced the appearance model by using local image structures to further mitigate the impact of illumination variations. In this work we apply the Modified Census Transform (MCT) to extract local image structures. The MCT was originally proposed in [5] for developing a rapid and robust face detection algorithm. It is a non-parametric transform inspired by the Census Transform, which was first introduced by Zabih and Woodfill [20] for texture analysis. The transform is defined as a set of $3 \times 3$ kernels which captures the local spatial structure of an image. It compares the pixel intensities between all the pixels of the $3 \times 3$ neighborhood and the mean intensity of all the pixels of the neighborhood. More formally, we define $\bar{I}(x)$ as the average of the pixel intensities in a $3 \times 3$ local spatial neighborhood $N(x)$ of the pixel $x$. The MCT

generates an ordered bit string indicating which pixels in $N(x)$ have an intensity lower than $\bar{I}(x)$. Let $\zeta(\bar{I}(x), I(y)) = 1$ if $\bar{I}(x) < I(y)$ be the comparison function and $\otimes$ be the concatenation operator, then the transform is defined as: $\Gamma(x) = \otimes_{y \in N} \zeta(\bar{I}(x), I(y))$. Figure 2(b) shows a sample output of the MCT. It has been proven that the transform is fast and very robust to illumination changes.

### 2.2.2 PCT Features

Inspired by the MCT, an unbinarized version of census transform is proposed to build robust discriminative appearance models for face alignment in [7]. The unbinarized census transform is named as the pseudo census transform (PCT). 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 the local mean. We fix $K = 9$ as in [7]. The PCT feature $\varphi$ is obtained by ordering the $K$ filter responses of a filter bank plotted in Figure 2(c) at position $(r, c)$. Figure 2(c) plots the $K$ filter masks, where white corresponds to the positive element and gray corresponds to the negative elements. Figure 2(d) shows the PCT-filter responses of a shape-free image. 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:

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

# 3 Learning Ranking Appearance Models

## 3.1 Pairwise Ordinal Classification-based RAM

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

$$F(\mathbf{p}_2) > F(\mathbf{p}_1) \Longleftrightarrow \mathbf{p}_2 \succ \mathbf{p}_1, \tag{2}$$

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. Equation 2 ensures the learned ranking model is a unimodal objective function with its maximum located exactly at $\mathbf{p}_0$.

Gentleboost is appied for boosting weak rankers as in [19]. Equation 2 suggests that the ranking function $F$ can be formulated as a classification problem. More precisely, if we define a classifier $H(\mathbf{p}_1, \mathbf{p}_2) = \text{sign}[F(\mathbf{p}_2) - F(\mathbf{p}_1)]$, *i.e.* $H(\mathbf{p}_1, \mathbf{p}_2) = +1$ 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 that is defined as:

$$f_m(\mathbf{p}) = \frac{1}{\pi} atan(\mathbf{w}^{m\top} S(\varphi^m) - t^m). \tag{3}$$

Note that $f_m(\mathbf{p})$ is continuous within $(-0.5, 0.5)$, the $atan()$ function is used to ensure both discriminability and derivability. Please also note that the PCT is used as feature in this model as we want to have a differentiable ranking function for gradient-based optimization. The $S(\cdot)$ is a sigmoid function, which normalizes the raw PCT feature values into a range of $(0, 1)$ before the linear projection defined by a projection vector $\mathbf{w}^m$ learned with

**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

$$f_m = \arg\min_f \sum_\ell w_\ell (z_\ell - h(\mathbf{x}_\ell))^2 \tag{5}$$

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

RankSVM [8]. The threshold $t^m$ needs to be determined during boosting. The strong ranking function is again assumed to be an additive model:

$$F(\mathbf{p}) \doteq \sum_{m=1}^{M} f_m(\mathbf{p}). \tag{4}$$

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 images, 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 sample $V$ 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. 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 as $z_\ell = +1$. The boosting procedure is summarized in Algorithm 1. Equation 5 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 Equation 4 in the sense of gradient ascent.

## 3.2 Pointwise Regression Trees-based RAM

The pointwise ranking function learning is a popular trend and achieved remarkable results in information retrieval domain [12]. The approaches based on gradient boosting regression trees enjoyed their great success for learning ranking function with pointwise data [10, 12].

### 3.2.1 Gradient Boosted Regression Trees

Gradient Boosted Regression Trees (GBRT) [5] is a machine learning technique which is based on tree averaging. It iteratively adds shallow trees with biased estimation. Each iteration focuses on the data that is responsible for the current remaining regression error. We denote $T(\mathbf{x}_i)$ as the current prediction of sample $\mathbf{x}_i$, and $y_i$ as the corresponding ground truth response. We adopt square loss: $L = \frac{1}{2}\sum_{i=1}^{n}(T(\mathbf{x}_i) - y_i)^2$ as the loss function as it is widely used in solving regression problems. The GBRT performs gradient descent to minimize the loss function in the data space $\mathbf{x}_1, \ldots, \mathbf{x}_n$. During each iteration the current prediction $T(\mathbf{x}_i)$ is updated with a gradient descent step:

$$T(\mathbf{x}_i) \leftarrow T(\mathbf{x}_i) - \alpha \frac{\partial L}{\partial T(\mathbf{x}_i)}, \tag{6}$$

**Algorithm 2:** Random Forests initialized Gradient Boosted Regression Trees

**Data**: $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, Parameters: $\alpha, M_B, d, K_{RF}, M_{RF}$
**Result**: The initialized Gradient Boosted Regression Trees $T$

$F \leftarrow RandomForest(D, K_{RF}, M_{RF})$
Initialization: $r_i = y_i - F(\mathbf{x}_i), i = 1, \ldots, n$
**for** $t = 1$ **to** $M_B$ **do**
    Find the $h_t$ with the CART according to Equation 7;
    Update residue $r_i \leftarrow r_i - \alpha h_t(\mathbf{x}_i)$, $i = 1, \ldots, n$;
**end**
$T(\cdot) = F(\cdot) + \alpha \sum_{t=1}^{M_B} h_t(\cdot)$
**return** $T(\cdot)$

where $\alpha > 0$ denotes the learning rate. Thus a new tree $h_t(\cdot)$ is chosen with its responses most highly correlated with the negative gradient $-\frac{\partial L}{\partial T(\mathbf{x}_i)}$ over the data distribution:

$$h_t \approx \arg\min_{h \in T_d} \sum_{i=1}^{n} (h_t(\mathbf{x}_i) - r_i)^2, where : r_i = \frac{\partial L}{\partial T(\mathbf{x}_i)}. \quad (7)$$

As $L$ is the squared loss, the gradient for a sample $\mathbf{x}_i$ becomes the residual from the previous iteration, *i.e.* $r_i = y_i - T(\mathbf{x}_i)$. The standard CART (Classification and Regression Trees) [1] is applied to find a solution to Equation 7. The parameter $d$ denotes the tree-depth.

The GBRT has a weakness which lies in the inherent trade-off between the step-size and early stopping. To obtain the true global minimum, the step-size needs to be very small and the number of iterations becomes very large. This results in a large number of regression trees which essentially decreases the efficiency of the model fitting. To tackle this problem, we try to initialize the GBRT learning with a reasonable start point which is close enough to the global minimum. We borrow the idea in [17], in which the Random Forests (RF) [2] method is applied for initialization. The Random Forests are considered to be a good choice as they are insensitive to parameter choices and offer low bias estimation as each of the tree is fully grown. One difference between RF and GBRT is that, in RF only $K$ uniformly chosen features are evaluated to find the best splitting point for each split. Furthermore, unlike the sequential tree construction in the GBRT, the construction of a single tree in RF is independent from earlier trees, thus the algorithm is easily parallelizable. Only two parameters need to be tuned. $M_{RF}$ specifies the number of trees in the forest and $K$ determines the number of features that each node considers to find the best split. As suggested in the original paper, we set $K = \sqrt{f}$, where $f$ is the number of features.

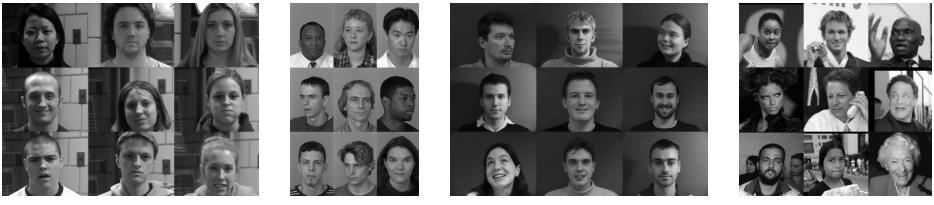
### 3.2.2 Initialized GBRT-based Ranking Model

The original GBRT is initialized with the average of the ground truth response, *i.e.* $T_0(x_i) = \bar{y}$, where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Consequently, the initial residual is $r_i = y_i - \bar{y}$. To initialize the GBRT with a better guess more close to the global minimum, the responses of the RF are used as the initial point for GBRT. We denote this initialized GBRT as iGBRT. Algorithm 2 details the steps in iGBRT. The output of the final boosted classifier is actually the response of RF combined with the boosted regression trees.

We apply iGBRT to learn a discriminative score function for face alignment. Basically, the ideal score function should return higher values if the shape parameter is closer to ground truth than the others. We use the data sampling process as in Section 3.1 for obtaining the

Table 1: Summary of the data-set.

| | FRGC | FERET | IMM | LFW |
|---|---|---|---|---|
| Images | 589 | 200 | 240 | 500 |
| Subjects | 200 | 200 | 40 | 500 |
| Variation | Expression, lighting | Pose | Pose, expression, lighting | All |
| Set 1 | 200 | 200 | | |
| Set 2 | 389 | | | |
| Set 3 | | | 240 | |
| Set 4 | | | | 500 |



(a) (b) (c) (d)

Figure 3: Example of the face data-set: (a) FRGC v2.0 database, (b) FERET database, (c) IMM database, and (d) LFW database.

training data. However, instead of assigning ordinal class labels to the ordered pairs, we assign ranking labels to each of the data point. That means we assign $y_i \in \{1,\ldots,V+1\}$, where the data generated using the ground truth is assigned with the highest value, *i.e.* $V+1$. The other samples in the same direction are assigned with $V+1-\nu, \nu = 1,\ldots,V$.

Face alignment is equivalent to maximizing the score for the regression trees with the constraint of the shape prior. We define the cost function as follows:

$$O(\mathbf{p}) = -T(\mathbf{p}) + \beta \sum_{i=0}^{n} \frac{p_i^2}{\lambda_i}, \tag{8}$$

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 learned objective function using regression trees, we use the Nelder-Mead simplex method [13] to minimize Equation 8 which only requires the evaluation of the cost function.

# 4 Experiments

The images for evaluating the proposed method are collected from multiple publicly available databases, including the FRGC v2.0 database [14], the FERET database [15], the IMM database [17], and the Labeled Faces in the Wild (LFW) database [9]. The collected images (see examples in Figure 3) are partitioned distinctively into four subsets. Table 1 lists the properties of each database and partition. We use Set 1 as training set and test the model fitting on all four data-sets. 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.

In the first experiment, we evaluate the pairwise RAM using the PCT as feature. We denote this method as PCT-SVM-RAM and compare it with the method in [7], which we name as PCT-SVM-BAM. Using Set 1, we train a shape model with 15 components preserving

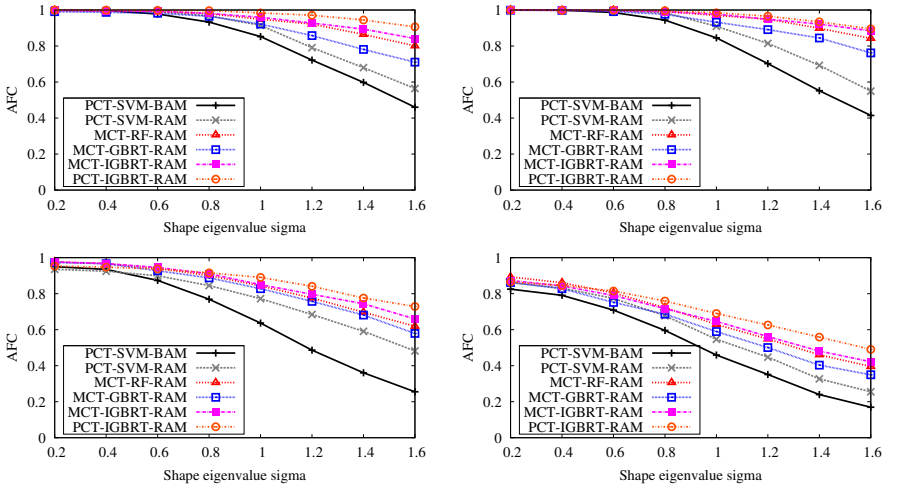
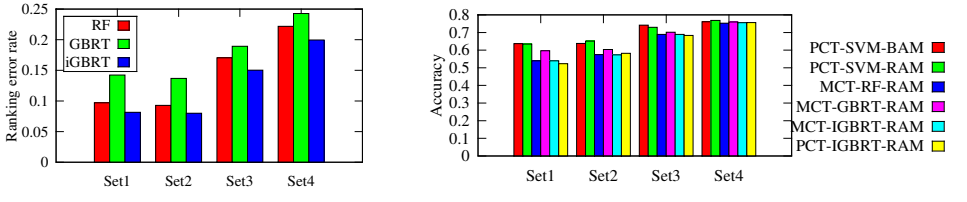



Figure 4: Alignment results on Set 1, Set 2 (first row) and Set 3, Set 4 (second row).




(a) (b)

Figure 5: (a) Ranking error rates of different regression trees-based RAM, (b) Average face alignment accuracy (in pixels) of the converged trials.

95% of shape variations. The size of shape-free images is $30 \times 30$ pixels. For each image we select $U = 10$ random 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 set includes $N = 24000$ $(400 \times 10 \times 6)$ ordinal pairs. The resulting ranking appearance model learns 100 weak rankers.

In testing, we randomly perturb ground truth landmarks at different 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 the 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. The same termination condition is applied for the fitting procedure as in [7].

The AFC curves for PCT-SVM-RAM and PCT-SVM-BAM in Figure 4 show that the pairwise RAM significantly improves the robustness of face alignment compared to the BAM. Especially, when we observe the AFC rates at $1.6\sigma$ ($\sigma$ is the shape eigenvalue) noise level, PCT-SVM-RAM outperforms PCT-SVM-BAM by around $8.5\% - 22.7\%$ on different data-sets. The most noticeable performance gain is achieved for the test on Set 3, which implies that the PCT-SVM-RAM has much better generalization ability than the PCT-SVM-

Table 2: Computational cost (ms in average) and fitting performance on Set 3

| Models | MCT-GBRT-RAM | MCT-RF-RAM | MCT-IGBRT-RAM | PCT-IGBRT-RAM |
|---|---|---|---|---|
| Fitting cost | 15.67ms | 29.88ms | 34.72ms | 256.95ms |
| AFC @ $1.6\sigma$ | 57.90% | 61.9% | 65.94% | 72.89% |

BAM on unseen data of unseen subjects. Improving alignment on Set 4 is difficult probably due to the limitation of the shape model learned on the Set 1.

The second experiment evaluates the pointwise RAM with regression trees and MCT used as feature. To assess the effectiveness of the RF initialized GBRT, we first compare the ranking performance for different models based on regression trees, *i.e.* RF, GBRT and iGBRT. The training data for building the regression trees is prepared according to Section 3.2.2. We again use Set 1 as training set to extract the training samples. As with the first experiment, we extract in total 7 samples in each direction including the ground truth. We set $M_{RF} = 100$ and $M_B = 100$. For GBRT and iGBRT, we set the tree-depth $d = 4$ and the learning rate $\alpha = 0.05$. The testing data is extracted in the same scheme on all four data-sets. The ranking results are plotted in Figure 5(a). From the plot we can observe that for all data-sets, the ranking error rates of RF are always lower than GBRT. The bagging technique and low bias regression makes RF resist to overfitting. The iGBRT outperforms RF and GBRT consistently on all data-sets.

The superior performance of iGBRT is proven again in the face alignment experiments. The RAM based on Random Forest (MCT-RF-RAM) shows large improvement compared to the pairwise RAM (PCT-SVM-RAM). The most significant improvement is observed on Set 2, where MCT-RF-RAM obtains around 30% performance gain over PCT-SVM-RAM. It is found that MCT-RF-RAM already boosts the fitting performance to a large extend. The introduction of iGBRT-based RAM increases the robustness further. In order to show that iGBRT also works for the PCT feature based representation, we train regression trees on top of PCT features after RankSVM scores are obtained. Results (PCT-IGBRT-RAM) that are plotted in Figure 4 show further improvement over MCT-IGBRT-RAM. This proves that the unbinarized census transform provides additional discriminative information for training the regression trees-based RAM. Finally, in Figure 5(b), we show the face alignment accuracy (pixel in average) of the converged trials with different models. The iGBRT-based RAM almost always outperforms the other models.

We analyse the computational cost for fitting different models in Table 2. We run the fitting experiments on a machine with Intel Xeon CPU (2.93GHZ) in an unparallelized C++ implementation. The second row in Table 2 lists the average fitting time (in millisecond) on the images in Set 3. The third row shows their AFC rates at $1.6\sigma$ noise level. We observe that although PCT-IGBRT-RAM achieves better results than MCT-IGBRT-RAM, the computational cost for each fitting is much higher due to the projection step using RankSVM.

# 5 Conclusions

We investigate deformable appearance models for face alignment based on learning a ranking function. Two different learning schemes for the ranking problem are compared in this work. The first one considers ranking as ordinal classification problem. The PCT feature and RankSVM are used to build weak rankers and a strong ranking function is learned via boosting regression stumps. The second proposed ranking model is based on the gradient boosted regression trees. The random forests technique is used to initialize the GBRT training iter-

ations. The initialization provides the GBRT with an initial estimation with low bias and requires less iterations to converge to the global minimum. We conducted experiments on four different data-sets. The results show that the regression trees-based RAM significantly improves the robustness and accuracy in terms of face alignment. Our best proposed model (PCT-IGBRT-RAM) boosts the alignment performance about $23.5\% - 35.6\%$ on different data-sets compared to the model based on pairwise ordinal classification (PCT-SVM-RAM).

# 6 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] L. Breiman. *Classification and regression trees*. Chapman & HallCRC, 1984.

[2] L. Breiman. Random forests. In *Machine Learning*, pages 5–32, 2001.

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

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

[5] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.

[6] 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.

[7] 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.

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

[9] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik 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.

[10] P. Li, C. Burges, and Q. Wu. Learning to rank using classification and gradient boosting. In *Proc. of the Intl. Conf. on Advances in Neural Information Processing Systems (NIPS)*, 2007.

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

[12] A. Mohan, Z. Chen, and K.Q. Weinberger. Web-search ranking with initialized gradient boosted regression trees. *JMLR Workshop and Conference Proceedings: Proceedings of the Yahoo! Learning to Rank Challenge*, 14:77–89, June 2011.

[13] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

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

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

[16] I. M. Scott, T. F. Cootes, and C. J. Taylor. Improving appearance model matching using local image structure. In *Infomation Processing in Medical Imaging*, pages 258–269. Springer-Verlag, 2003.

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

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

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

[20] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proc. of the European Conference on Computer Vision*, pages 151–158. Springer-Verlag, 1994.

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