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

Abstract

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 [8, 9]. 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 dimensionality of the configuration space of the proposed feature.

1 Introduction

Deformable face model fitting is essentially an image registration problem. After deformation, the facial features in a face image are aligned with the model. Numerous works have been conducted to solve the face alignment problem, on account of its importance in a wide range of applications, such as analysis of expression, pose, gender, age, and identity of human faces. Nevertheless, it remains to be a challenging problem in the computer vision community due to the variation factors such as illumination, expression, occlusion and image quality. These factors make it difficult for face models to generalize to unseen data.

One of the early face alignment approaches using a deformable face model is the Active Shape Model (ASM) [2], where the model fits to the data in a way consistent with a training set. There are several extension of the ASM. One noticeable work is the Bayesian formulation of the ASM [15], where a Bayesian inference solution and an expectation maximization (EM) based method is used for estimating the maximum a posteriori probability (MAP).

Another popular extension of the ASM is the Active Appearance Model (AAM) [3, 10], in which the appearance of the face is also considered. The model combines constraints

BMVC 2011 http://dx.doi.org/10.5244/C.25.54

on both shape and texture by learning statistical generative models for the shape of a face and the appearance of a face. Shape is represented by landmark positions (see Figure 1 (b)), whereas the appearance is represented by pixel intensities in the shape-free face image (see Figure 1 (c)). The fitting of the AAM is defined by solving a least mean square error (LMSE) problem, where difference between the warped image and the model appearance is minimized. Efficient optimization algorithms such as the Inverse Compositional (IC) and Simultaneously Inverse Compositional (SIC) methods have been proposed by Baker and Matthews [1], which enable fast face alignment for real-time applications. However, the alignment performance degrades quickly when generic AAMs are trained instead of person specific AAMs [6]. The generalization issue is caused by generative appearance modeling and the LMSE optimization schema as claimed in [8].

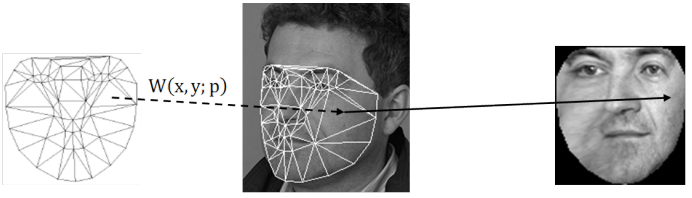
In order to tackle this generalization problem, Liu proposed the Boosted Appearance Model (BAM) [8], in which a shape representation similar to the AAM is used, whereas the appearance is represented by a set of discriminative features, trained to form a boosted classifier. The discriminative appearance models is able to distinguish between correct and incorrect alignment. The BAM fitting can be done by iteratively updating the landmark positions according to gradient ascent on the corresponding classifier score function. It has been shown that the BAM improves the generalization capabilities of the AAM.

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 BAM 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*. [5], in which the modified census transformation (MCT) is applied for face detection. The face detector based on the MCT feature yields better detection performance in addition to its fast training and detection speed compared to the state-of-the-art approach [14]. The MCT feature, however, 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 [4]. Multi-scale PCT features are also investigated. We evaluated our PCT-based BAM fitting on four different datasets. Our proposed approach achieved slightly better performance on seen data compared to the Haar-based BAM. However, results on the unseen data showed that the PCT-based BAM outperforms the Haar-based BAM significantly in terms of the Average Frequency of Convergence (AFC), which indicates that our approach generizes better on unseen data.

The rest of the paper is organized as follows. We present our discriminative appearance model using pseudo census transformation in detail in Section 2. The algorithm for fitting our proposed face model is presented in Section 3. The experimental setup and results will be discussed in Section 4, and we conclude with future work in Section 5.

2 Face Model

Generative face models, such as the AAMs, are usally represented by the combination of a generative appearance and shape model. In contrast, we introduce a discriminative face model in this section, in which the appearance is represented with a set of boosted discrim-



(a) (b) (c)

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

inative features. The following subsections describe the features we used and the details of feature selection for face alignment.

2.1 Shape Model

We use a generative shape model to describe the distribution of the shape of faces. The shape of a face is represented by a set of l 2D-landmarks, defined by their image coordinates $\mathbf{x}_i = (x_i, y_i)_{i=1,\ldots,l}$. The coordinates of the landmarks are stacked to form the shape vector $\mathbf{s} = [x_1, y_1, x_2, y_2, \ldots, x_l, y_l]^\top$. Assuming the face shapes lie in a linear subspace, we represent a novel shape $\mathbf{s}$ with a linear combination of shape basis:

$$\mathbf{s} = \mathbf{s}_0 + \Sigma_{i=1}^{n} p_i \mathbf{s}_i, \tag{1}$$

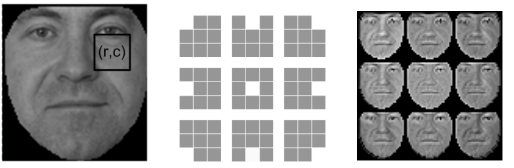
where $\mathbf{s}_0$ is the mean shape, $\mathbf{s}_i$ is the i-th shape basis, and $\mathbf{p} = [p_1, p_2, \ldots, p_n]^\top$ is the shape parameter. The mean shape and the shape basis can be learned from a labeled training set of face images via Principal Component Analysis (PCA). This model is known as the point distribution models which has been used in many of previous works [2, 3, 10].

With Delauney triangulation, the mean shape $\mathbf{s}_0$ (Figure 1 (a)) and the shape $\mathbf{s}$ (Figure 1 (b)) are triangulated to a base mesh and an instance face 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(c)) is obtained by warping a face image $\mathbf{I}$ to the coordinate of the mean shape.

2.2 Appearance Model

The appearance model is a collection of m features computed over the shape-free face image $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$. In [8], the rectangular Haar features were adopted. Haar features are know as good local features for general object detection [14]. One drawback of the Haar features is that the configuration space is extremely large, which makes the selection procedure very slow. In [5], Fröba *et al*. found out that the feature extracted by the modified census transformation (MCT) outperforms the Haar features in face detection, and especially, due to the small configuration space of the MCT feature, the detector can be trained very fast. Inspired by their work, we propose to select the unbinarized census transform (which we call pseudo census tranform (PCT)) feature for our appearance model [1]. The PCT feature $\varphi = (\varphi_1, \ldots, \varphi_K)^\top$ is a K dimensional vector 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

[1] The MCT features can not be applied for deriving the fitting algorithm since it is a binary pattern.



(a) (b) (c)

Figure 2: (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 (2); (c) PCT-filter responses of a shape-free image.

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 2(b) at a specific position (r,c). The mask of the first filter can be defined as follows:

$$\mathbf{T}_0 = \begin{pmatrix} 8/9 & -1/9 & -1/9 \\ -1/9 & -1/9 & -1/9 \\ -1/9 & -1/9 & -1/9 \end{pmatrix} \tag{2}$$

The rest of the filter masks can be defined accordingly by shifting the position of the value $8/9$ in the matrix (see Figure 2 (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 $\mathbf{A}_{i=1,\ldots,K}$ 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:

$$\varphi_i = \mathbf{A}_i^\top \mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p})) = \mathbf{T}_i * \mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p})), i=1,\ldots,K. \tag{3}$$

2.3 Learning Alignment

We formulate the problem of learning an alignment score function to perform the fitting of the face model in this section. More precisely, for a given image, let us suppose that $\mathbf{p}$ is the shape parameter that represents the current alignment of the shape model, with the face in the image: 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 corresponding to the correct alignment. Mathematically, if $\mathbf{p}^*$ is the shape parameter representing the correct alignment, F has to be such that

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} F(\mathbf{p}) \tag{4}$$

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:

$$F(\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))) = \Sigma_{m=1}^{M} f_m(\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))) \tag{5}$$

where $f_m(\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p})))$ is a function operating on one PCT feature of $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$. Given this formulation of the appearance model, machine learning tools such as boosting become a natural choice to learn such a model. Note that $f_m(\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p})))$ in (5) can be viewed as a weak classifier operating on $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$. For simplification of the notation, we will denote the weak classifier and the strong classifier as $f_m(\mathbf{p})$ and $F(\mathbf{p})$ respectively.

(a) (b) (c)

Figure 3: Boosted PCT feature locations. (a) Boosted PCT feature locations in PCT-BAM, (b) Boosted PCT feature locations in the original scale in MSPCT-BAM, (c) Boosted PCT feature locations in the half scale in MSPCT-BAM.

Our weak classifier using the PCT features is defined as follows:

$$f_m(\mathbf{p}) = \frac{\pi}{2} atan(\Sigma_{i=1}^{K} w_i^m S(\mathbf{A}_i^{m\top} \mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))) + b^m), \tag{6}$$

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 normalizes the raw PCT feature values into a range of $(0,1)$ before a linear projection. The projection vector $\mathbf{w}^m$ and bias b^m are learned on the training data with linear support vector machines (SVM). The individual SVM cost parameter C for each feature location is searched with cross validation.

The GentleBoost algorithm [4] is used to boost the weak classifiers as suggested in [8] for two reasons. On one hand, it is a soft classfier with continuous output which enables us to derive gradient ascent algorithm for maximizing the strong classifier function. On the other hand, the GentleBoost outperforms other boosting algorihms since it is more robust against noisy data. Figure 3(a) plots the top 40 locations of the boosted features overlaid on a 30×30 mask image. The gray pixels inside the mask indicates the location of the boosted feature. Note that the boosted PCT features are mainly located around the natural facial features, i.e. the eyes, nose and mouth region. The features extracted at those locations contribute the most to the face alignment.

The PCT features extracted on the images at different scales 2^{-j} might contribute additional discriminative information for face alignment, where j is the level index in a multi-scale image pyramid. We also boost PCT features on different scales ($j = 0,1,2,3$) of the shape-free images. The location of the boosted features in the original scale ($j = 0$) and the half scaled image ($j = 1$) are displayed in Figure 3(b) and (c) respectively. These feature locations are boosted together with all scales. We found that actually there are no features boosted at the scale level 2 and 3, because the images are too small to obtain useful features. Hereafter, we refer to the single scale face model as PCT-BAM (PCT-based boosted appearance model) and the multiple scale face model as MSPCT-BAM.

3 Face Alignment

In order to align a PCT-BAM with the face in a given image $\mathbf{I}$, we assume that the model is currently aligned with a shape parameter $\mathbf{p}$ (at the i-th iteration). In order to achieve the optimal alignment one may perform a simple gradient ascent on the score function F, and therefore update the shape parameter as follows

$$\mathbf{p} = \mathbf{p} + \nu \frac{dF}{d\mathbf{p}}, \tag{7}$$

where ν is a suitable constant. From (3), (5), and (6) one can see that the derivative of F with respect to $\mathbf{p}$ is

$$\frac{dF}{d\mathbf{p}} = \frac{2}{\pi}\Sigma_{m=1}^{M}\frac{\alpha\Sigma_{i=1}^{K}w_i^m S(\varphi_i^m)S(1-\varphi_i^m)[\nabla\mathbf{I}\frac{\partial\mathbf{W}}{\partial\mathbf{p}}]^\top\mathbf{A}_i^m}{1+[\Sigma_{i=1}^{K}w_i S(\varphi_i^m)+b^m]^2}, \tag{8}$$

where $\nabla\mathbf{I}$ is the gradient of the image evaluated at $\mathbf{W}(\mathbf{x};\mathbf{p})$, and $\frac{\partial\mathbf{W}}{\partial\mathbf{p}}$ is the Jacobian of the warp evaluated at $\mathbf{p}$. The detailed fitting steps is summarized in Algorithm 1.

Algorithm 1 The face alignment algorithm of PCT-BAM

Input:

Input image $\mathbf{I}$, initial shape parameters $\mathbf{p}$, pre-computed Jacobian $\frac{\partial\mathbf{W}}{\partial\mathbf{p}}$, the shape model $\{\mathbf{s}_i; i=0,1,\ldots,n\}$ and the appearance model.

Output:

Shape parameters $\mathbf{p}$.

0. Compute the 2D gradient of the image $\mathbf{I}$.

repeat

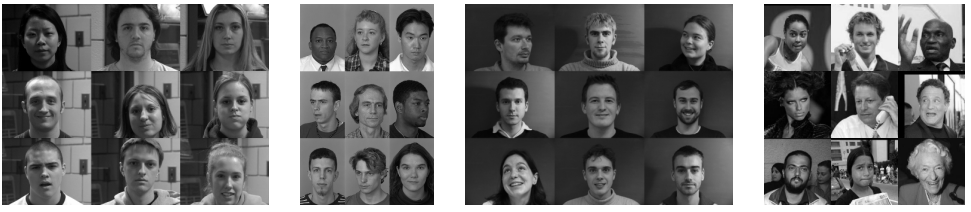
1: Compute $\mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$ by warping image $\mathbf{I}$ with $\mathbf{W}(\mathbf{x};\mathbf{p})$.
2: For each weak classifier compute the feature: $e_m = \Sigma_{i=1}^{K}w_i S(\varphi_i^m)+b^m; m=1,2,\ldots,M$.
3: Interpolate the gradient of image $\mathbf{I}$ at $\mathbf{W}(\mathbf{x};\mathbf{p})$ with bilinear interpolation.
4: Compute the steepest descent images $SD = \nabla\mathbf{I}\frac{\partial\mathbf{W}}{\partial\mathbf{p}}$.
5: Compute the PCT feature from each column of SD and project with $\mathbf{w}^m$: $\mathbf{d}_m = \alpha\Sigma_{i=1}^{K}w_i^m S(\varphi_i^m)S(1-\varphi_i^m)[\nabla\mathbf{I}\frac{\partial\mathbf{W}}{\partial\mathbf{p}}]^\top\mathbf{A}_i^m; m=1,2,\ldots,M$.
6: Compute $\Delta\mathbf{p}$ using $\Delta\mathbf{p} = \nu\frac{2}{\pi}\Sigma_{m=1}^{M}\frac{\mathbf{d}_m}{1+e_m^2}$.
7: Update $\mathbf{p} = \mathbf{p}+\Delta\mathbf{p}$.

until $\|\Sigma_{i=1}^{n}\Delta ps_i\| \le \tau$.

4 Experiments

4.1 Evaluation Dataset and Procedure

The dataset for evaluation in this work contains 1529 images. These images are collected from multiple publicly available databases, including the FRGC v2.0 database [12], the FERET database [11], the IMM database [13], and the Labeled Faces in the Wild (LFW) database [7]. Figure 4 shows sample images from these four databases. The collected images are partitioned distinctively into four subsets. Set 1 includes 400 images (one image per subject), where 200 images are from the FRGC database and the other 200 images are from the FERET database. Set 1 is used as the training set. Set 2 includes 389 images from the same subjects but different images as the FRGC database in Set 1. Set 3 includes 240 images from 40 subjects in the IMM database that were never used in the training. Set 4 includes randomly selected 500 images of 500 subjects from the LFW database. This partition 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. Set 4 is a particular challenging dataset since it is collected from the Internet. The images were captured under cluttered background and various real-world illumination environments using different types of cameras. There are 58 manually labeled landmarks for each of the



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

Figure 4: Example of the face dataset: (a) FRGC v2.0 database, (b) FERET database, (c) IMM database, and (d) LFW database.

1529 images. The images are down-sampled such that the facial width is roughly 40 pixels across the set in order to speed up the training process.

We compare our proposed PCT-BAM and MSPCT-BAM to the Haar feature-based BAM (Haar-BAM). We do not compare our model against AAM-based methods, as it has been shown in [8] that the Haar-BAM outperforms them. We train the two models with Set 1 by taking the shape-free images extracted with ground truth landmarks at the positive samples, and the negative samples are generated by perturbing the shape parameter of the ground truth shapes uniformly in a range of the corresponding eigenvalues. We generate 10 negative samples for each image and resulting 4000 negative samples in total. The shape model has 15 shape bases and it is the same for all the models which preserve 95% of shape varations. We use the same mean shape size as in [8], that means the size of shape-free images is 30×30 pixels. The resulting appearance models are such that the MSPCT-BAM and the Haar-BAM have 50 weak classifiers, and the single scale PCT-BAM has 43 weak classifiers, as it can only boost 43 weak classifiers on Set 1.

The false alarm rate (FAR) of the strong classifiers of the three models are plotted in Figure 5. The FAR is plotted as a function of the number of weak classifiers, when the miss-detection rate on the training set is set to 0%. The plot shows that both PCT-based models converge faster than the Haar-BAM. In particluar, for 50 weak classifiers the FAR's of MSPCT-BAM and Haar-BAM are 1.12% and 5.47%, respectively.

A faster convergence means that it is less likely to have local maximum on a classification score surface. Figure 6(a) shows that for a given image, a concave surface of classification scores can be observed while perturbing the shape parameters along two shape bases. The concavity property of the score surface ensures that the gradient ascent algorithm can perform well. The perturbation range is set to be 1.6 times the eigenvalue of these two bases. When the perturbation is at the maximal amount for two bases, the corresponding four perturbed landmarks are plotted at Figure 6(b). To see the properties of score surfaces, more surfaces are plotted as images in Figure 6(c), where the intensity corresponds to the classification score. Each sub-image is generated in the same way as Figure 6(a). For most cases, we see the intensity changes from high to low when the pixel deviates from the center, i.e., the alignment gets less accurate. This monotonic surface is important for a successful face alignment algorithm.

4.2 Experimental Results

In the evaluation, we use the randomly perturbed ground truth landmarks to initialize each alignment. In order to perform a statistical evaluation of the result, we repeat the random perturbation multiple times on each test image. The initial position of the landmarks is

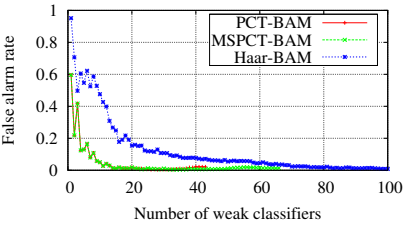
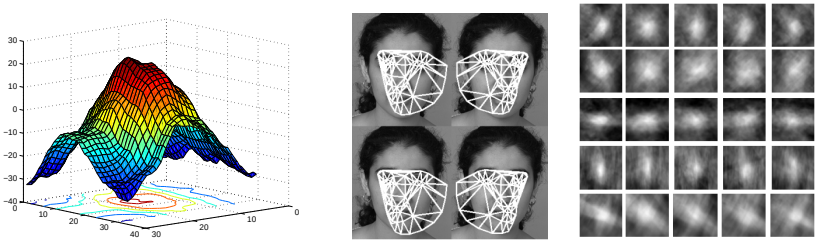


Figure 5: False alarm rate of the strong classifiers when the miss-detection rate on the training set is set to 0%.



(a) (b) (c)

Figure 6: (a) The classification score surface while perturbing the shape parameters in the neighborhood of the ground truth along the 4th and 5th shape basis. (b) The four perturbed facial landmarks when the perturbation is at the four corners of the surface on the left. (c) The classification score surface of 5 facial images (one by each column) while perturbing the shape parameters along pairs of shape bases (from top to bottom (p_1, p_2), (p_2, p_3), (p_3, p_4), (p_4, p_5),(p_5, p_6)).

generated by perturbing the shape parameter with independent Gaussian noise with variances multiple of the corresponding eigenvalues. We consider as converged if the Root Mean Square Error (RMSE) between the aligned landmarks and the ground truth is less than one pixel. For the converged trails, we use two metrics to measure the robustness and accuracy of the alignment. The Average Frequency of Convergence (AFC) which assesses the robustness of the alignment is calculated as the number of converged trials divided by the total number of trials. The second metric is the histogram of the RMSE (HRMSE) of the converged trials, which measures how close the aligned landmarks are to the ground truth.

The evaluation of PCT-BAM, MSPCT-BAM and Haar-BAM is conducted under the same conditions. All algorithms are initialized with the same set of randomly perturbed landmarks. All algorithms have the same constant ν in Equation 7, and also the same termination condition. That is, if the number of iterations is larger than 55 or the RMSE between consecutive iterations is less than 0.025 pixels. Figure 7 plots the AFC of the PCT-BAM, MSPCT-BAM and Haar-BAM against the level of the initial landmarks perturbation, computed over Set 1, 2, 3, and 4, respectively. For each perturbation value, we randomly perturb each image of each set five times.

The AFC plots in Figure 7 show that MSPCT-BAM based alignment achieved comparable results on the seen data (Set 1 and 2). The robustness of the MSPCT-BAM based alignment is slightly better than Haar-BAM based alignment with increasing perturbing variance. However, in the experiments on unseen data (Set 3 and 4), PCT-based (both PCT-BAM

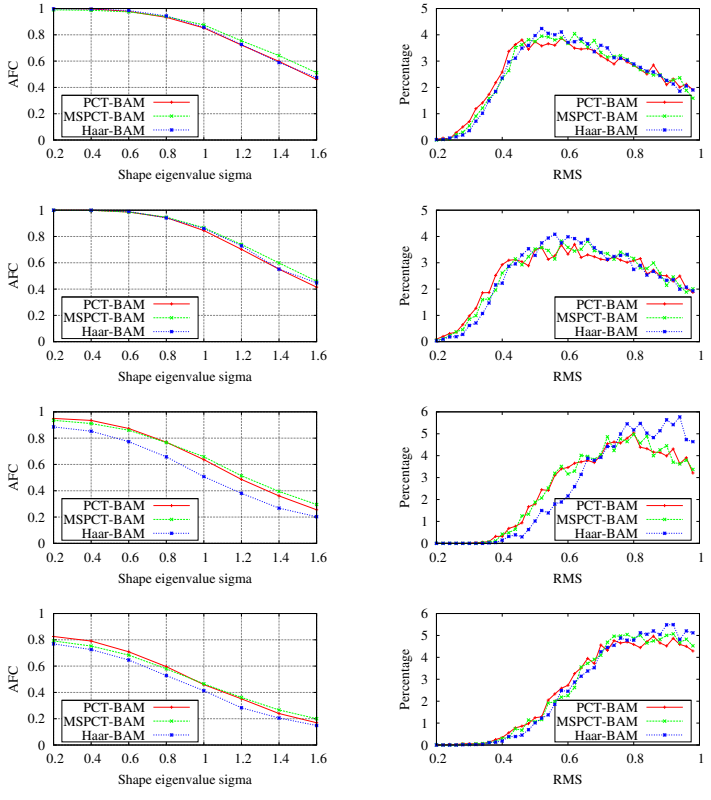


Figure 7: Alignment results of three algorithms on Set 1, 2, 3, and 4. From top to bottom, each row is the result for one set. Left column is the AFC; right column is the HRMSE.

and MSPCT-BAM) alignment outperforms the Haar-BAM-based alignment significantly as plotted on the third and fourth row in Figure 7. On Set 3, 95% of the perturbed samples with a perturbation range of 0.2 times the shape eigenvalue converge in the PCT-BAM fitting, while the convergence rate of Haar-BAM is 88%. When the perturbation range increases to 1.6 times the shape eigenvalue, the AFC value of MSPCT-BAM is 9% higher than Haar-BAM. On the most challenging testing set (Set 4), in which the imaging conditions are totally different from each other, the performance of both algorithms degrades a lot. However, the performance drop of PCT-BAM (17%) is less than that of Haar-BAM (23%) at the first perturbation index. Additional PCT features selected on other scales also improve the robustness of aligment with large perturbation as can be observed consistently through all the experiments. The accuracies of the two methods are comparable as we can see from the HRMSE plots. (MS)PCT-BAM is slightly superior to Haar-BAM again on the unseen data as displayed in Figure 7. Overall, our PCT-based alignment has better generalization capability than the Haar-BAM-based alignment.

The reason for the performance gain on unseen data is probably because the responses of the PCT filter are somewhat similar to the Laplacian filter, which is a highpass filter. The 3×3 filter mask in the center of Figure 2(b) is indeed a discretized Laplacian filter. And these filter responses are less sensitive to illumination changes, which make the PCT-based approach generalize better on unseen data with mismatched illumination conditions.

5 Conclusions

We have introduced the pseudo census transformation-based boosted appearance model (PCT-BAM), a new discriminative appearance model suitable to perform face alignment. The adopted PCT feature has much less parameters to be boosted which enables extremely fast model training compared to the training procedure of the Haar-BAM. We compared the proposed PCT-based alignment to the Haar-BAM on seen data and unseen data. Our experimental results on seen data are comparably better. However, our PCT-BAM model shows significant performance improvement on unseen data, which means the proposed model has better generalization properties on unseen data. Additional PCT features selected on other scales also improve the robustness of alignment with large perturbation as can be observed consistently through all the experiments. As future work, PCT features on different scales can be boosted separately to enable hierarchical alignment.

References

[1] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *Int. J. Computer Vision*, 56(3):221–255, 2004.

[2] T. F. Cootes and C. J. Taylor. Active shape models. In *Proc. of 3rd British Machine Vision Conference*, pages 266–275, 1992.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. of* 5^{th} *European Conference on Computer Vision*, volume 2, pages 484–498, 1998.

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

[5] 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.

[6] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(12):1080–1093, 2005.

[7] 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.

[8] 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.

[9] X. Liu. Discriminative face alignment. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31:1941–1954, 2009.

[10] I. Matthews and S. Baker. Active appearance models revisited. *Int. J. of Computer Vision*, 60(2): 135–164, 2004.

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

[12] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 947–954, 2005.

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

[14] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. of Computer Vision*, 57(2): 137–154, 2004.

[15] Y. Zhou, L. Gu, and H. Zhang. Bayesian tangent shape model: Estimating shape and pose parameters via bayesian inference. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 109–116, 2003.