Ranking Prior Likelihood Distributions for Bayesian Shape Localization Framework*

Shuicheng Yan[†], Mingjing Li[‡], Hongjiang Zhang[‡], Qiansheng Cheng[†]
[†] School of Mathematical Sciences, Peking University, 100871, China
[‡] Microsoft Research Asia, Beijing Sigma Center, Beijing, 100080, China

Abstract

In this paper, we formulate the shape localization problem in the Bayesian framework. In the learning stage, we propose the Constrained RankBoost *approach to model the likelihood of local features associated with the key points of an object, like face, while preserve the prior ranking order between the ground truth position of a key point and its neighbors; in the inferring stage, a simple efficient iterative algorithm is proposed to uncover the MAP shape by locally modeling the likelihood distribution around each key point via our proposed* variational Locally Weighted Learning *(VLWL) method. Our proposed framework has the following benefits: 1) compared to the classical PCA models, the likelihood presented by the ranking prior likelihood model has more discriminating power as to the optimal position and its neighbors, especially in the problem with ambiguity between the optimal positions and their neighbors; 2) the VLWL method guarantees that the posterior probability of the derived shape increases monotonously; and 3) the above two methods are both based on accurate probability formulation, which spontaneously leads to a robust confidence measure for the discovered shape. Moreover, we present a theoretical analysis for the convergence of the Constrained RankBoost. Extensive experiments compared with the Active Shape Models demonstrate the accuracy, robustness, and stability of our proposed framework.*

1. Introduction

Accurate localization of representative points of a face is essential to many face analysis and synthesis problems, such as 3-D modeling[16], face synthesis and recognition [11][12][14]. The geometry structure, *i.e.* shape, normalized texture patch, and local features associated with the key points, provide important clues for the face interpretation. The traditional algorithms, like Active Shape Model (ASM) [4] and its variations [8][9][15], Active Appearance Models (AAM) [4] and its variations [17], make use of part or all of these clues for shape localization.

In this paper, we focus on exploring the shape localization problem by only using the shape information and local features due to the observation that the local features are more robust to illumination and expression variations than the normalized texture patch. The problem is formulated in a Bayesian framework with two novel approaches for learning and inferring. The framework is motivated by the following observations: 1) the local features associated with the neighborhood points of a key point are often similar to, thus somehow ambiguous with that associated with the key point; moreover, the most representative features are not always the best discriminative features. Therefore the traditional Principle Components Analysis is insufficient to present discriminative likelihood for differentiating a key point from its neighbors; and 2) the traditional shape localization algorithms using only local features, like ASM, search for the optimal shape without explicit objective function. These algorithms can not guarantee that the searched shape has monotonously increasing posterior probability in each step in the sense of Bayesian modeling. They provide neither a robust confidence evaluation for the searched shape nor a stop criterion for the entire searching process in a principled manner.

In this work, first, a semi-supervised learning algorithm, *Constrained RankBoost*, is proposed to build the likelihood model that ensures the ground truth position will more likely have a higher likelihood than its neighbors. It aims at providing more discriminative likelihood for the key points and their neighbors. Second, we present a simple effective iterative algorithm for the optimization of the objective functions by dynamically locally learning the likelihood distribution around each key point using our proposed *Variational Locally Weighted Learning* method. It guarantees that the objective function increases monotonously. These two approaches are both based on accurate probability formulation, which naturally leads to a robust confidence measure for the searched shape. It can be used as stopping condition for the inferring process. Moreover, we present a theoretical proof for the convergence of our proposed Constrained RankBoost which can be applied to the general classification and ranking problems.

The rest of this paper is organized as follows. In Section 2, the shape localization problem is formulated in a Bayesian framework. In Section 3, the Constrained RankBoost method is introduced to build the ranking prior likelihood model. We systematically introduce the efficient iterative optimization method for the posterior probability function

* The work was performed at Microsoft Research Asia.

and give a summary for the whole inferring process in Section 4. Experimental results are presented in Section 5. Finally, the concluding remarks are given in Section 6.

2. Bayesian Shape Localization Formulation

Suppose a set of labeled samples $\{I, S\}$ are given, where the shape $S = ((x_1, y_1), ..., (x_K, y_K)) \in \mathbb{R}^{2K}$ is a sequence of K labeled key points representing the object, such as a face, in image I. The task of shape localization is to infer the optimal shape with the maximal posterior probability in an image by learning from the training samples, *i.e.*

$$S^* = \arg\max_{s \in \mathbb{S}_s} p(S \mid I) \quad (1)$$

where $\mathbb{S}_s$ is the learnt traditional tangent shape space [5][6][10][18] by Principle Components Analysis [4]. The shape S can be decomposed into two parts: intrinsic shape space parameters and the geometrical transformation, *i.e.*

$$S = Ts_c(\bar{S} + Us) \quad (2)$$

where shape parameter $s \in \mathbb{S}_s$, $\bar{S}$ is the average shape and U is the first k leading eigenvectors; $Ts_c(*)$ is the 2-D geometrical transformation function based on four parameters: scale (r), rotation (θ), and translations (T_x, T_y).

From the Bayesian rule and the assumption that the local features around different key points are independent to each other, Eq. (1) can be reformulated as:

$$\begin{aligned} S^* &= \arg\max_{s \in \mathbb{S}_s} P(I \mid S) P(S) \\ &= \arg\max_{s \in \mathbb{S}_s} \prod_{i=1}^{K} P_i^{n_i}(I \mid S_i) \times P(S) \end{aligned} \quad (3)$$

where $P_i^{n_i}(I \mid S_i)$ is the likelihood of the *i*-th point S_i with n_i being the normal direction to the contour.

From Eq. (3), the accurate inferring of the optimal shape consists of two fundamental problems: 1) accurate likelihood modeling for local features; and 2) proper optimization method for the objective function.

For the first problem, the classical way is to model the likelihood by Principle Components Analysis or Mixture Gaussian Models [9]. However, it is often the case that the points around a ground truth position have similar local features especially in low-resolution images. The ambiguities between a ground truth position and its neighbors lead to that the principle components are insufficient to formulate the difference between the optimal position and its neighbors; moreover, the most *representative* features are not always the best *discriminative* features. These considerations motivate us to explore a supervised learning algorithm for the likelihood modeling. In the following, we propose the Constrained RankBoosting algorithm to model the likelihood while preserving the prior likelihood ranking order between the optimal positions and their neighbors.

For the second problem, the objective function can not be optimized using gradient descent methods directly since the distribution of $P_i^{n_i}(I \mid S_i)$ is hard to model in advance due to the randomicity of n_i. Liu [13] proposed a hierarchical data-driven Markov Chain Monte Carlo (HDDMCMC) approach to infer the optimal shape; however, it converges very slowly and hence is not suitable for real world applications that require real-time performance. In this paper, we present a simple efficient iterative approach to optimize the objective function by online locally modeling the likelihood distribution $P_i^{n_i}(I \mid S_i)$ via our proposed variational Locally Weighted Learning algorithm.

The ranking prior likelihood models and the optimization strategy are both based on the accurate probability models and they are integrated for accurate inferring of the optimal shape in the Bayesian shape localization framework.

3. Ranking Prior Likelihood Models

As discussed in Section 2, the ambiguity between the ground truth position and its neighbors requires that the local likelihood model should be able to correctly rank the likelihoods of these ambiguous positions. A natural way is to take into account the ranking priors of these positions in the construction of the local likelihood model. Consequently, we propose a semi-supervised approach to learn the ranking prior likelihood models that not only characterize the local features of a ground truth position, but also preserve the likelihood ranking order between the ground truth position and its neighbors using the Constrained RankBoost algorithm.

3.1. Constrained RankBoost

RankBoost [3][7] is a variation of the classical boosting algorithm which produces highly accurate prediction rule by combining many "weak" rules that may be only moderately accurate. It aims at providing high accurate ranking evaluation and was originally used in the applications like web page ranking.

In the ranking order problem, we are given a set of samples χ ={x}, where x is the feature vector, and a set of crucial pairs with ranking order priors $\Omega = \{ (x, x') : R(x, x') > 0\}$, where $R(x, x') > 0$ means sample x has higher confidence than sample x' and $R(x, x')$ is normalized to satisfy $\sum_{(x,x') \in \Omega} R(x, x') = 1$; meanwhile, a set of "weak" ranking evaluation functions are presented as $F = \{ f_i, i = 1, \cdots, L\}$, where L is the number of the

IEEE COMPUTER SOCIETY

(Constrained) RankBoost

1) Initiate the weights over the crucial pairs in $\Omega : D_1(x,x') = R(x,x')$

2). For t = 1,…,T:

For each ranking function f_i, the error is evaluated with respect to weights D_t,

$$\varepsilon_j = \sum_{(x,x')\in\Omega} D_t(x,x')(f_j(x) - f_j(x'))$$

Choose the function $h_t = f_i$ with the largest absolute value $|\varepsilon_j|$; Set $\alpha_t = \frac{1}{2}\ln\left(\frac{1+\varepsilon_j}{1-\varepsilon_j}\right)$.

Constrained RankBoost:

(If $|\varepsilon_j| > 1 - e^{-2K_0}$, set $\alpha_t = sign(\varepsilon_j)\, K_0$)

Update the weights:

$$D_{t+1}(x,x') = \frac{D_t(x,x')\exp\{\alpha_t(h_t(x') - h_t(x))\}}{Z_t}$$

where Z_t is a normalization factor, making D_{t+1} be a distribution.

3). Output the final ranking function:

$$H(x) = \sum_{i=1}^{T} \alpha_t h_t(x)$$

Figure 1. (Constrained) *RankBoost* Algorithm

functions and f_i satisfies $0 \le f_i \le 1$ for all $x \in \chi$. The task of *RankBoost* is to search for the confidence evaluation function $H : \chi \to \mathbb{R}$ with "minimal" weighted number of the incorrectly-ranked crucial pairs, *i.e. the ranking loss:*

$$H^* = \arg\min_{H\in\mathbb{S}_F} \sum_{(x,x')\in\Omega} R(x,x')\left\lceil \tilde{H}(x,x') \right\rceil \quad (4)$$

where $\tilde{H}(x,x') = H(x) - H(x')$; $\lceil z \rceil = 1$ means x and x' are mis-ranked, *i.e.* $z < 0$, else 0; $\mathbb{S}_F$ is the confidence evaluation function space spanned by the functions in F. The pseudo code of the classical *RankBoost* algorithm is listed in Figure 1.

The RankBoost is a procrustes, local optimal algorithm. In each step, the large value of α_t for local pursing may make H(x) escape from the optimal combination, which is often observed in our modeling of the likelihood of local features. We present an improved RankBoost by banding the coefficient $|\alpha_t|$ with K_0 when $|\alpha_t|$ is big enough, and we call it *Constrained RankBoost*. The convergence is proved as below:

Theorem 1: Assuming that h_t is selected with the largest absolute value of the error ε_j, and

$$\alpha_t = \begin{cases} \frac{1}{2}\ln\left(\frac{1+\varepsilon_j}{1-\varepsilon_j}\right) & when\ |\varepsilon_j| \le 1 - e^{-2K_0} \\ sign(\varepsilon_j)K_0 & else \end{cases}$$ in each step,

then the rank loss of the final function

$$\sum_{(x,x')\in\Omega} R(x,x')\left\lceil \tilde{H}(x,x') \right\rceil \le \prod_{t=1}^{T} Z_t \text{ and } Z_t \le 1.$$

Proof. According to the updating rule of $D_t(x,x')$, we have that

$$D_{T+1}(x,x') = \frac{D(x,x')\exp\{\alpha_t(H(x') - H(x))\}}{\prod_t Z_t}.$$

As $\lceil x \rceil \le e^x$, therefore

$$\sum_{(x,x')\in\Omega} R(x,x')\left\lceil \tilde{H}(x,x') \right\rceil \le \sum_{(x,x')\in\Omega} R(x,x')\exp\{H(x) - H(x')\}$$

$$= \sum_{(x,x')\in\Omega} D_{T+1}(x,x')\prod_{t=1}^{T} Z_t = \prod_{t=1}^{T} Z_t$$

In step t, if $|\varepsilon_j| \le 1 - e^{-2K_0}$, then $Z_t \le \sqrt{1-\varepsilon_j^2}$ [1];

Otherwise, $|\varepsilon_j| > 1 - e^{-2K_0}$, then

$$Z_t \le \frac{1-\varepsilon_j}{2}e^{\alpha_t} + \frac{1+\varepsilon_j}{2}e^{-\alpha_t}$$

$$< \frac{e^{-2K_0 + sign(\varepsilon_j)\alpha_t}}{2} + e^{-sign(\varepsilon_j)\alpha_t}$$

$$= \tfrac{3}{2}e^{-K_0} < 1 \ (when\ K_0\ is\ big\ enough)$$ □

The Constrained RankBoost avoids the strong oscillation of the coefficient in the selected weak ranking evaluation function and makes the algorithm converge more robustly. Moreover, it is a general problem in the boosting algorithms and our proposed algorithm can be generalized to the other Boosting algorithm for robust convergence.

3.2. Ranking Prior Likelihood Models

For the likelihood $P_i^{n_i}(I \mid S_i)$, there exists ranking orders between the optimal position and its neighbors, although they are ambiguous in many cases. We present the Ranking Prior Likelihood Model to formulate the likelihood output via our proposed Constrained RankBoost. In the shape localization problem, the feature vector *x* is the local features sampled around the key point S_i, and the ranking order priors $R(x,x')$ are given by setting the ranking order between the ground truth position and its neighborhood points. Notice that the projection to princi-

IEEE COMPUTER SOCIETY

ple components of the local features from ground truth position presents fundamental ranking evaluation; we construct the "weak" ranking evaluation function set using different combinations of these principle components. Let L' denote the number of the principle components. There are $(2^{L'}-1)$ kinds of different combinations:

$$f_j(x) = \exp\{-\frac{1}{2}\sum_{k=1}^{l}\frac{\tilde{x}_{jk}^2}{\lambda_{jk}}\} \quad (5)$$

where $\tilde{x}_{jk}$ is the projection of x to the *jk*-th principle component of the local features from the ground truth positions; λ_{jk} is the *jk*-th largest eigenvalue; $1 \le j1 < j2 < \cdots < jl \le L$. Since the ranking order of candidate pairs is only determined by the sign of their likelihood difference, we normalize H(x) such that $\int H(x)dx = 1$.

By taking into account the prior information in the learning stage, the ranking prior likelihood model is able to present more accurate likelihoods for the candidate positions than the traditional PCA model does. We conducted experiments to compare the performance of the new models using Constrained RankBoost with that using classical RankBoost and the traditional PCA models by systematically comparing the likelihood of each point of the ground truth shape with that of its neighbors on 100 testing samples. It is observed that our proposed Constrained RankBoost outperforms the Classical RankBoost, and the classical RankBoost performs better than traditional PCA Models. The statistics results are listed in Table 1.

Table 1. The percentages of right-ranked pairs of ground truth position and its neighbors compared between models constructed by Constrained RankBoost, classical RankBoost and PCA.

Approach	Constrained RankBoost	Classical RankBoost	PCA Models
Accuracy	89%	84%	76%

4. Locally Weighted Learning for Optimal Shape Inferring

As analyzed in Section 2, it is difficult to directly undertake the optimization for the complex global structure of function $P_i^{n_i}(I \mid S_i)$. An intuitive way is to locally learn the distribution of $P_i^{n_i}(I \mid S_i)$, which fascinates discovering the local optimal shape. In this section, we present a simple efficient iterative approach to maximize the posterior probability of the observed shape by locally modeling the likelihood of $P_i^{n_i}(I \mid S_i)$ via a variational *locally weighted learning* (LWL) [1][2] approach. The LWL dynamically models the complex function using simple local models, with no necessity to find an appropriate structure for a global model. We simplify this idea and present a variational Locally Weighted Learning (VLWL) method to locally model the complex function $P_i^{n_i}(I \mid S_i)$ using classical semi-Gaussian functions. Consequently, the optimum of the shape localization problem in the neighborhood of the original shape can be obtained by using this local model, namely *adaptive local likelihood distribution model*. In the following subsections, we begin with the introduction of this new model, and then present a new optimization framework for the shape localization problem.

4.1. Adaptive Local Likelihood Distribution Model

There is no close form solution for the objective function (3). A natural way is to search for the solution in an iterative manner. Denote S^k as the resulted shape from the (*k-1*)-th iteration, the task of each step is to find the optimal shape in the neighborhood of S^k using local optimization approach. Following the ideas of VLWL approach, $P_i^{n_i}(I \mid S_i)$ can be locally approximated around S_i^k using the confidences presented by the *i*-th ranking prior likelihood model. An efficient way is to fit the distribution using semi-Gaussian models. As described later, these simple models fascinates the objective function be optimized increasingly. These models can be constructed by two steps: (1) local neighbor selection; and (2) local distribution model construction.

In the first step, the local neighbors of S_i^k are sampled following the distribution:

$$p(x,y) = \begin{cases} 1/\pi\theta^2 & \text{when } \|(x,y) - S_i^k\| \le \theta \\ 0 & \text{else} \end{cases} \quad (6)$$

where θ is the coefficient determining the sampling range around the point S_i^k. Let $\{A_m^{(k,i)}\}$ denote the sample set, the confidence for each sample can be obtained as:

$$Cf_i(A_m^{(k,i)}) = H_i(x(A_m^{(k,i)}, n_i^k)) \quad (7)$$

where $x(A_m^{(k,i)}, n_i^k)$ is the local features around $A_m^{(k,i)}$ and $H_i(\cdot)$ is the learnt ranking prior local likelihood evaluation function for the *i*-th point.

In the second step, the confidence distribution in the neighborhood of S_i^k is approximately modeled from the samples $\{A_m^{(k,i)}\}$ and their confidences $\{Cf_i(A_m^{(k,i)})\}$ by the semi-Gaussian Model:

$$p_{N(S_i^k)}^{n_i^k}(I \mid S_i) = C \cdot N(\mu_i^k, \Sigma_i^k) \quad (8)$$

IEEE COMPUTER SOCIETY

where parameter C, μ_i^k and Σ_i^k can be easily learnt via least square method.

4.2. Shape parameter Inferring

From Eq. (3), we can write the object function for shape localization problem as follows:

$$F(S,I)=\prod_{i=1}^{K} P_i^{n_i}(I \mid S_i)\times P(S) \qquad (9)$$

In the k-th iteration, $F(S,I)$ can be optimized in the neighborhood of S^k in term of the *adaptive local likelihood distribution models*, then the object function is changed to:

$$F_{N(S^k)}(S,I)=\prod_{i=1}^{K} P_{N(S_i^k)}^{n_i^k}(I \mid S_i)\times P(S) \qquad (10)$$

Thus, the local optimum of the shape localization problem can be derived from the energy function:

$$\begin{aligned} S^* &= \arg\min_{S\in\mathbb{S}_s} \sum_{i=1}^{K} En_{N(S_i^k)}^{n_i^k}(I;S_i)+En(S) \\ &= \arg\min_{S\in\mathbb{S}_s} \sum_{i=1}^{K} \left\| Ts_c(M_i^s)-\mu_i^k \right\|_{\Sigma_i^k}^2 + \sum_{j=1}^{K'} \frac{s_j^2}{\lambda_j} \end{aligned} \qquad (11)$$

where $\mathrm{En}(\cdot)$ is the corresponding energy function of the distribution $\mathrm{p}(\cdot)$, namely $p(x)\propto \exp\{\text{-}En(x)\}$; K' is the dimension of $\mathbb{S}_s$; λ_j is the j-th largest eigenvalue of the covariance matrix from the training shapes; s is the corresponding shape parameter; $M^S=\bar{S}+Us$; and Ts_c is the geometric transformation function based on the transformation parameter $c=(r,\theta,T_x,T_y)$:

$$Ts_c(x,y)=Tr\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}T_x\\T_y\end{pmatrix}=\begin{pmatrix}a & -b\\ b & a\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}T_x\\T_y\end{pmatrix} \qquad (12)$$

where $a=r\cos\theta, b=r\sin\theta$.

In Eq. (11), the optimization function is multinomial but has no close form solution. As discussed in [9], the solution can be approximated iteratively using a two-step optimization method as following:

Transformation parameter estimation with given s**:** Given the shape parameter s, M^s and $En(S)$ are constant. In this case, we only need to minimize the following energy function:

$$En(c)=\sum_{i=1}^{K}\left\| \begin{pmatrix}a & -b\\ b & a\end{pmatrix} M_i^s+\begin{pmatrix}T_x\\T_y\end{pmatrix}-\mu_i^k \right\|_{\Sigma_i^n}^2 \qquad (13)$$

In order to obtain the optimal transformation parameter, we set the partial derivative of $En(c)$ to zero. That is, the optimal parameters are obtained by solving the following linear functions:

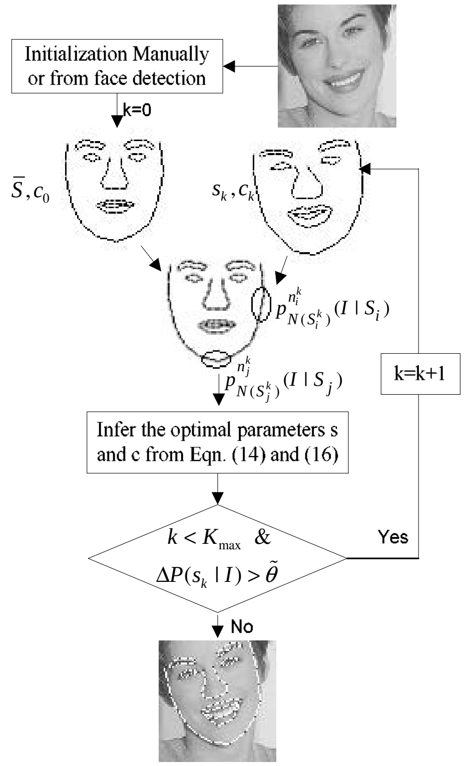

Figure 2. The inferring process for Bayesian shape localization.

$$\begin{aligned} &\sum_{k=1}^{4}\left[\sum_{i=1}^{K}\left(\frac{\partial Ts_c(M_i^s)}{\partial c_l}\right)^T \Sigma_i^k \left(\frac{\partial Ts_c(M_i^s)}{\partial c_l}\right)\right] c_k \\ &= \sum_{i=1}^{K}\left(\frac{\partial Ts_c(M_i^s)}{\partial c_l}\right)^T \Sigma_i^k\, \mu_i^k \quad l=1,\cdots,4 \end{aligned} \qquad (14)$$

Shape parameter estimation with given c**:** Given the transformation parameter c, the energy function is changed to:

$$\sum_{i=1}^{K}\left\| Tr(\bar{S}_i+U_i s)+\begin{pmatrix}T_x\\T_y\end{pmatrix}-\mu_i^k \right\|_{\Sigma_i^n}^2 + \sum_{j=1}^{K'}\frac{s_j^2}{\lambda_j} \qquad (15)$$

where U_i is a matrix consisting of the (2i-1)-th and 2i-*th* row of U. Using the same approach as above, s can be obtained by solving the following linear functions:

$$\begin{aligned} &\left[\sum_{i=1}^{K}(TrU_i)^T \Sigma_i^k (TrU_i)+ 1/\lambda_j\right] s \\ &= \sum_{i=1}^{K}(TrU_i)^T \Sigma_i^k(\mu_i^k - Tr\bar{S}_i - T) \end{aligned} \qquad (16)$$

where $T=(T_x,T_y)^T$.

We summarize the entire inferring process in Figure 2, where $K_{\max}$ is the manually defined maximal iteration number and $\Delta P(S_k \mid I)=p(S_{k+1} \mid I)-p(S_k \mid I)$ and $\tilde{\theta}$ is the least posterior probability increasing value each step must achieve.

IEEE COMPUTER SOCIETY

5. Experimental Results

The experiments have been conducted on a data set consisting of 500 frontal face images, in which each face area is about 150*150 pixels and many faces have ambiguous key points as those shown in Figure 8. All faces were manually labeled with 83 key points. 400 of them were randomly selected for model construction, and the rest 100 for testing.

For comparison, ASM and the proposed shape localization framework, referred as RPBF (**R**anking **P**rior likelihood Distributions for **B**ayesian Shape Localization **F**ramework) were trained on the same data set, in a four-level image pyramid (Resolution is reduced 1/2 level by level). All results were obtained by maximally searching five times per layer.

The most commonly used criterion to evaluate the searched shape is the average point-point or point-curve distance between the searched shape and the ground truth. In all our experiments, the results were evaluated using the average point-point distance.

5.1. Accuracy evaluation

The average point-point distance between a searched shape and the manually labeled shape of ASM and RPBF are compared and showed in Figure 3. The vertical axis represents the distribution of point-point distance. It shows that most results from using RPBF have smaller point-point error than those from ASM.

We have also explored the performance when only using the ranking prior local confidence model or our proposed shape parameter estimation approach. As shown in Table 2, the performance was improved when only using one of our proposed approaches; moreover, our proposed framework integrated the two approaches and significantly improved the performance of the shape localization compared to the other three algorithms.

5.2. Algorithm robustness and stability

In Figure 4, the statistics of the average point-point distance of each step of search are compared between ASM and RPBF. The statistics are obtained from 100 testing images with initial displacements about 20 pixels from the ground truth. The standard deviation of point-point distance of each step is also plotted as bars in both mean lines, respectively.

The capture range of in-plane rotation angle is an important criterion to evaluate shape localization algorithms. Figure 5 demonstrates that RPBF can capture a larger percentage of the rotation cases in the range of [-40°, 40°].

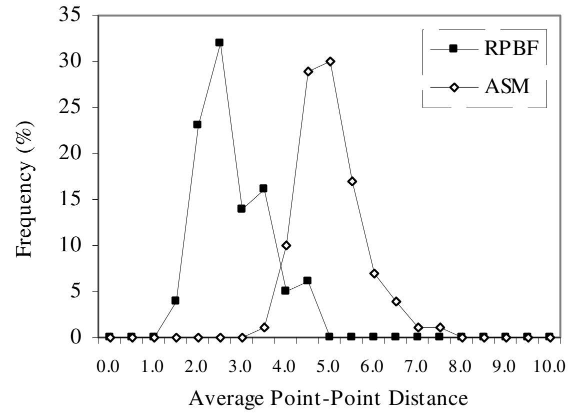

Figure 3. Shape localization accuracy comparison: ASM vs. RPBF. Note that most results of RPBF have smaller point-point distances than those from ASM.

We measure the stability by the standard deviation of different results from different initializations with similar point-point distances between the initial shapes and the ground truths. The results from using RPBF and ASM are compared in Figure 6, where the vertical axis represents the average standard deviation of the results obtained from 9 different initializations. The result conveys that RPBF is more stable in initialization than conventional ASM. Figure 7 shows the average error and standard deviation of the localization results from ASM and RPBF in terms of the initial errors in the x direction. It indicates that RPBF has a little larger capture range to initial error in x direction and much better accuracy in all conditions than ASM.

A comparative example of RPBF and ASM is presented in Figure 8, in which ASM fails to accurately locate the contour points for ambiguous furrow on the face; however, RPBF accurately locates the contour. Figure 9 shows some more experimental results by using RPBF computed on the images with diverse challenging conditions for conventional ASM algorithm, such as low resolution, small faces, strong expression and in-plane rotation.

RPBF is a fast algorithm. It costs only 80 ms per iteration (on P4 1.8G computer with 512M memory) although it is slower than the classical ASM for the locally learning of the local likelihood distribution. It takes about 8 iterations to converge in average. As discussed above, it outperforms much better than ASM in accuracy, stability and robustness.

Table 2. Mean value of the average point-point distance in #1: ASM, #2: Ranking prior likelihood model + conventional shape parameter estimation approach; #3: Conventional PCA model + our proposed shape parameter estimation approach, #4: Our proposed shape localization framework.

Algorithm	#1	#2	#3	#4
Average Error	4.99	4.14	4.01	2.77

IEEE COMPUTER SOCIETY

6. Conclusions and future work

In this paper, the shape localization problem is formulated in a Bayesian framework and two novel methods are proposed for highly confident inferring of the optimal shape. The likelihood of the local features associated with the key point is modeled using a Constrained RankBoost method which is introduced to ensure the ground truth position has higher likelihood than its neighbors. This model is learned in a semi-supervised manner and presents discriminative likelihood output for the local features. On the other hand, the optimal shape is inferred in an iterative method that locally models the likelihood distribution around each key point via the Variational Locally Weighted Learning method and simplifies the task into a general multinomial optimization problem in each step.

A large part of our work focus on the highly confident likelihood modeling. We propose the Constrained RankBoost method for the ranking prior likelihood modeling. It prevents the coefficients of the selected weaker functions from strongly oscillating and makes sure that the algorithm can converge robustly. We present a theoretical analysis on the convergence of the Constrained RankBoost algorithm. It is a general approach for classification and ranking problems. Moreover, the coefficients constraints analysis can be generalized to other Boosting algorithms for roust convergence.

Our proposed framework of face shape localization can be easily extended to multi-view face shape localization problem by properly modeling the multi-view shape space in linear or nonlinear manner. On the other hand, the key points used in our experiments are manually selected; the automatic feature point selection algorithm is helpful for compact shape space and robust local likelihood modeling. We are currently exploring these extensions in both theory and practice.

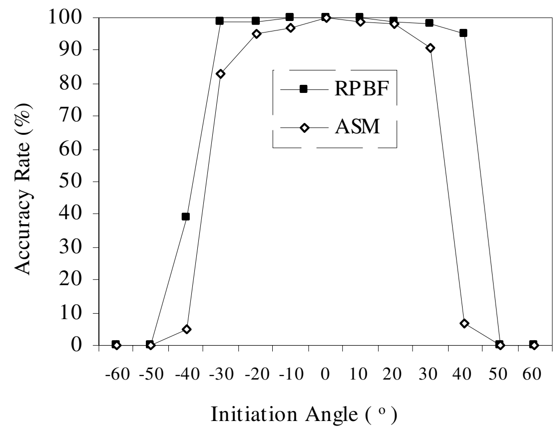

Figure 5. The capture range comparison of in-plane rotation angle: ASM vs. RPBF.

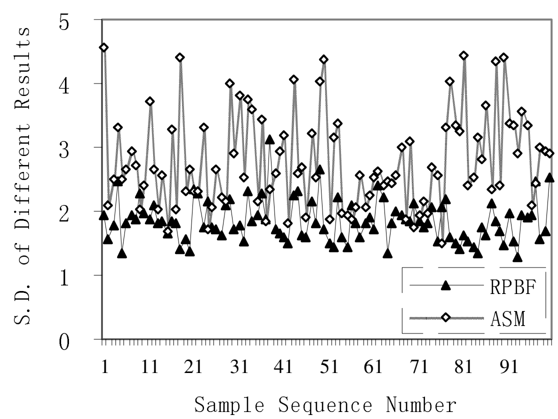

Figure 6. Standard deviation of results from different initialization compared between ASM and RPBF.

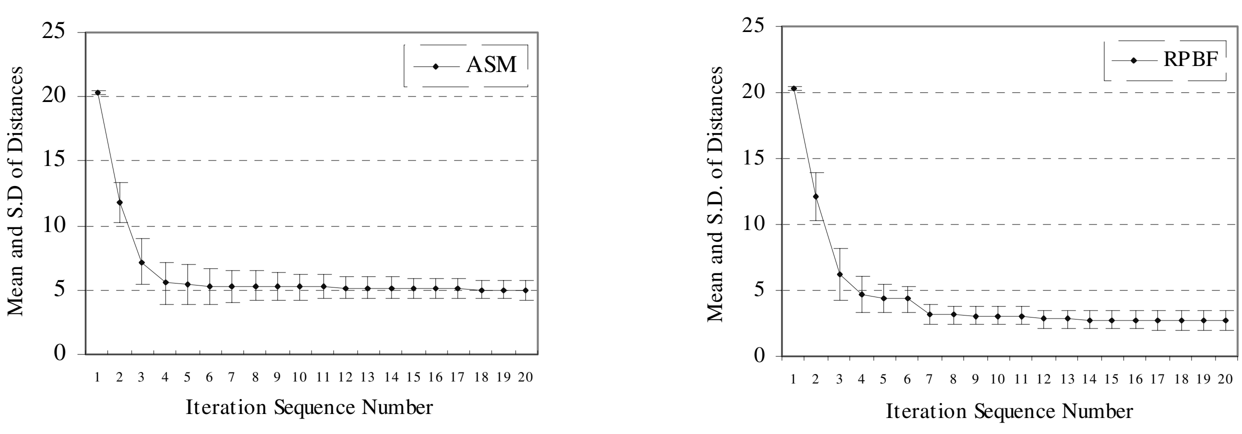

Figure 4. Mean error (curve) and standard deviation (bars) of the point-point distances between the searched results and the ground truths as a function of iteration number for the ASM (left) and RPBF (right) methods.

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 **IEEE**

IEEE COMPUTER SOCIETY

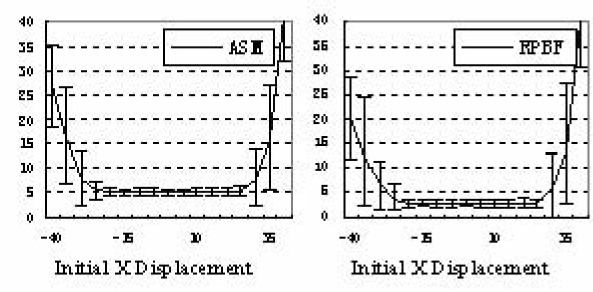

Figure 7. Mean error and standard deviation of the localization results from ASM (left) and RPBF (right) in terms of localization errors in the x direction.

Figure 8. A case ASM (Left) fails for the ambiguous furrow in the face, RPBF (Right) performs well.

Figure 9. Some results of RPBF from the images with challenging conditions: low resolution, small face，strong expression and in-plane rotation.

References

[1]. C. G. Atkeson, A.W. Moore, and S. Schaal. “Locally Weighted Learning”. Artificial Intelligence Review, 1997, 11:11-73.

[2]. C. G. Atkeson, A. W. Moore, and Schaal, S. “Locally Weighted Learning for Control”. Artificial Intelligence Review, 1997, 11:75-113.

[3]. W. W. Cohen, Robert E. Schapire, Yoram Singer. “Learning to Order Things”. JAIR 10 1999, 243-270.

[4]. T.F. Cootes and C.J. Taylor. Technical Report: Statistical models of appearance for computer vision. http://www.isbe.man.ac.uk/~bim/refs.html, 2001.

[5]. I. L. Dryden. “General shape and registration analysis”. In Stochastic Geometry: Likelihood and Computation, 1997.

[6]. I. L. Dryden. “Statistical shape analysis in high-level vision”. IMA Workshop on Image Analysis and High Level Vision Modeling, 2001.

[7]. Y. Freund, R. Iyer, R. E. Schapire and Y. Singer. “An efficient boosting algorithm for combining preferences”. Extended abstract appeared in Machine Learning: Proceedings of the Fifteenth International Conference, 1998.

[8]. B. van Ginneken, A. F. Frangi, J. J. Staal, B.. M. ter Haar Romeny, M. A. Viergever. “A nonlinear gray-level appearance model improves active shape model segmentation”. IEEE Workshop on Mathematical Models in Biomedical Image Analysis, MMBIA 2001, pp. 205-212.

[9]. A. Hill, T.F. Cootes, C.J. Taylor. “Active Shape Models and the shape approximation problem”. Image and Vision Computing.14 (8), Aug. 1996, pp 601-608.

[10]. J. T. Kent and K. V. Mardia. Shape, “Procrustes tangent projections and bilateral symmetry”. Biometrica, 2001.

[11]. A. Lanitis, C.J. Taylor, T.F. Cootes, “Automatic Interpretation and Coding of Face Images Using Flexible Models”. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19 No.7 pp.743-756, July 1997.

[12]. Y. Li, S. Gong, and H. Liddell. “Modeling faces dynamically across views and over time”. In *Proceedings of IEEE* ICCV01, Vancouver, Canada, July 2001.

[13]. C. Liu, H. Shum and C. Zhang. “Hierarchical Shape Modeling for Automatic Face Localization”. In Proceedings of the ECCV2002, May, 2002.

[14]. B. Moghaddam and A. Pentland. “Probabilistic visual learning for object representation”. IEEE Transactions on PAMI, July 1997, 7:696–710.

[15]. M. Rogers and J. Graham. “Robust Active Shape Model Search”. In Proceedings of the 7th European Conference on Computer Vision. Copenhagen, Denmark, May 2002, vol. 4 pp. 517-530.

[16]. Sami Romdhani, Volker Blanz and Thomas Vetter. “Face identification by matching a 3D Morphable model using linear shape and texture error functions”. In Proceedings of the ECCV2002, vol.4, pp. 3-19.

[17]. S.C. Yan, C. Liu, S.Z. Li, L. Zhu, Z, H.J. Zhang, H. Shum, Q.S. Cheng. "Texture-Constrained Active Shape Models". In *Proceedings of the First International Workshop on Generative-Model-Based Vision in ECCV02*. Copenhagen, Denmark. May, 2002.

[18]. Y. Zhou, L. Gu, Hongjiang Zhang. “A Unified Bayesian Treatment for Shape Registration in Tangent Space”, To Appear in IEEE Computer Society Conference on. Computer Vision and Pattern Recognition, June, 2003.

0-7695-1950-4/03 $17.00 © 2003 **IEEE**