LPSM: Fitting Shape Model by Linear Programming

Jilin Tu, Brandon Laflen, Xiaoming Liu, Musodiq Bello, Jens Rittscher and Peter Tu
Visualization and Computer Vision Lab, General Electric, 1 Research Circle, Niskayuna, NY, USA
{tujilin, laflenb, liux, bello, rittche, tu}@crd.ge.com

***Abstract*— We propose a shape model fitting algorithm that uses linear programming optimization. Most shape model fitting approaches (such as ASM, AAM) are based on gradient-descent-like local search optimization and usually suffer from local minima. In contrast, linear programming (LP) techniques achieve globally optimal solution for linear problems. In [1], a linear programming scheme based on successive convexification was proposed for matching static object shape in images among cluttered background and achieved very good performance. In this paper, we rigorously derive the linear formulation of the shape model fitting problem in the LP scheme and propose an LP shape model fitting algorithm (LPSM). In the experiments, we compared the performance of our LPSM with the LP graph matching algorithm(LPGM), ASM, and a CONDENSATION based ASM algorithm on a test set of PUT database. The experiments show that LPSM can achieve higher shape fitting accuracy. We also evaluated its performance on the fitting of some real world face images collected from internet . The results show that LPSM can handle various appearance outliers and can avoid local minima problem very well, as the fitting is carried out by LP optimization with l_1 norm robust cost function.**

I. INTRODUCTION

Given pictures of objects of the same category (i.e., human faces) with manually labeled landmarks as training data, Cootes proposed to learn the intrinsic shape and appearance variations as statistics and proposed the Active Shape Model (ASM) [2] and Active Appearance Model(AAM) [3]. The ASM/AAM approaches received wide attention in the computer vision community and were extended in many ways for different applications.In brief summary, the shape model has been extended from PCA [2] to Kernel PCA [4], and to a hierarchical bayesian model [5]. The appearance likelihood model of ASM has been extended from the simple Gaussian distribution of normalized pixel values [2] to Mixture of Gaussians with discriminative features computed from image patch centering at the landmarks (Gabor wavelet [6],Ranking-Boost [7], Fisher-boost [8], etc.).

The fitting algorithm of the ASM models to the object in the image was extended from coarse-to-fine iterative optimization process constrained by PCA shape subspace [2] to Expectation-Maximization (EM) optimization [9] and sampling based approaches [5][8][10] in a Bayesian inference framework. These approaches are based on gradient descent optimization in essence and suffer from problems with local minima. Due to the high dimensionality of the image space and the highly cluttered image background, these approaches depend heavily on good initialization.

On the other hand, if we assume the object of interest do not have large intrinsic shape variation, or the shape variations can be characterized by some heuristic constraints, we can characterize the object shape by a graph template, and graph matching techniques can be employed to localize the object in a testing image. Graph matching has a rich history in computer vision and pattern recognition applications, dating back to the 1970's [11]. Recently, in [12], Jiang and Yu formulated the graph matching problem as a linear programming (LP) framework. The optimization was carried out in a successive convexification scheme. By successively convexifying the matching cost surface in a coarse to fine manner, the fitting of the template to the target image can be solved after it is formulated as a linear programming problem. As linear programming can achieve globally optimal solutions to linear objective functions subject to linear constraints, Jiang and Yu demonstrated that their graph matching algorithm can achieve good localization performance in images with a large search range and with cluttered background.

In this paper, we extend the LP graph matching (LPGM) technique in [12] to the object shape model fitting problem, and propose an LP shape model (LPSM) fitting algorithm. By incorporating a shape subspace model that spans the object shape variations, LPSM can achieve better shape fitting accuracy for the objects of a category with non-rigid shape deformations (i.e., human faces). By doing the optimization in an LP framework, LPSM can find globally optimal solution without relying on a good model initialization.

More specifically, we summarize the novelties of our paper as follows:
1) We rigorously derive the linear formulation of the shape model fitting problem, and propose to replace the gradient descent or other local optimization techniques) in ASM with a globally optimal linear programming (LP) technique, after the shape fitting problem is formulated in the LP framework. The proposed algorithm therefore does not rely on good model parameter initialization. 2) In order to get more accurate estimation of the scale and orientation of the shape model, we propose a refined unit circle linearization procedure by extending the LP framework in [12]. 3) For implementation, we also proposed to approximate the joint distribution of the allowable shape instances from real world by convex hulls in a number of 3 dimensional most correlated

This work was supported by award #2007-DE-BX-K191 from the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice.

subspaces to refine the search space in the LP formulation. 4) We propose to model the landmark appearance likelihood in Maximal Rejection Classifier[13] subspace that captures the discriminability between the appearances at the landmarks and its surrounding areas.

In section 2, we briefly summarizes the shape fitting problem in ASM literatures, and the graph matching problem in LP framework. Based on the analysis in section 2, we formulate the shape model fitting problem in LP framework in section 3, and provide linearization solutions for the refined estimation of the scale and rotation parameters. We then describe our MRC appearance likelihood model and provide some implementation details. We show the experimental results in section 4, and finally provide a discussion on the strengths of our work and some future works in section 5.

II. Mathematical Formulation

A. Shape model fitting

As shown in [2], a set of corresponding points on object shapes in the form $\vec{S} = [x_1, y_1, x_2, y_2, ..., x_N, y_N]'$ can be captured by a PCA model after Procrustes analysis, i.e., $\vec{S} = \bar{\vec{S}} + \mathbf{U}\vec{c}$. where $\vec{S}$ is the intrinsic shape, $\bar{\vec{S}}$ is the mean of the shapes, $\mathbf{U}$ is the shape basis matrix that characterizes the intrinsic shape variations, and $\vec{c}$ is the intrinsic shape parameter.

In [14], the extrinsic transformation of the object shape (in terms of scaling factor s, rotation angle θ and translation $t = [t_x, t_y]'$) is modeled as a global similarity transformation. Consider the landmark $S_i = [x_i, y_i]'$ on the intrinsic shape $\vec{S}$, the similarity transform is $S_i^{img} = \begin{bmatrix} 1+a & -b \\ b & 1+a \end{bmatrix} S_i + t,\ i = 1, ..., N$, where $a = s\cos\theta - 1$ and $b = s\sin\theta$. By rearranging this equation, the 2D similarity transformation for the object shape $\vec{S}$ can be formulated in a linear form: $\vec{S}^{img} = \vec{S} + [\vec{g}_1, \vec{g}_2, \vec{g}_3, \vec{g}_4]\vec{d} = \vec{S} + \mathbf{G}(\vec{S})\vec{d}$, where $\vec{g}_1 = [x_1, y_1, x_2, y_2, ..., x_N, y_N]' = \vec{S}$, $\vec{g}_2 = [-y_1, x_1, -y_2, x_2, ..., -y_N, x_N]' = \mathbf{R}\vec{S}$($\mathbf{R}$ is a diagonal block matrix with diagonal block being $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$), $\vec{g}_3 = [1, 0, 1, 0, ..., 1, 0]'$, $\vec{g}_4 = [0, 1, 0, 1, ..., 0, 1]'$ and $\mathbf{d} = [a, b, t_x, t_y]'$. Here we denote $\mathbf{G}(\vec{S}) = [\vec{S}, \mathbf{R}\vec{S}, \vec{g}_3, \vec{g}_4]$. And further derivation yields

$$\vec{S}^{img} = \bar{\vec{S}} + \mathbf{G}(\bar{\vec{S}})\vec{d} + \mathbf{U}\vec{c}(1+a) + \mathbf{R}\mathbf{U}\vec{c}b. \quad (1)$$

According to [15][14], $\mathbf{G}(\bar{\vec{S}})'\mathbf{U} = 0$ because the training shape data is normalized by generalized Procrustes analysis. Given the properties of $\mathbf{R}$, $\mathbf{G}(\bar{\vec{S}})$, and $\mathbf{G}(\bar{\vec{S}})'\mathbf{U} = \mathbf{0}$, it is straight forward to prove that $\mathbf{G}(\bar{\vec{S}})'\mathbf{R}\mathbf{U} = \mathbf{0}$.

Therefore, by defining $\mathbf{V}$ as the gram-schmidt orthogonalization of $[\mathbf{U}, \mathbf{R}\mathbf{U}]$ (in practice, we can let $\mathbf{V} \approx \mathbf{U}$ as we observe most of the shape variations represented in $\mathbf{R}\mathbf{U}$ can be captured by $\mathbf{U}$), we have $\vec{S}^{img} = \bar{\vec{S}} + [\mathbf{G}(\bar{\vec{S}}), \mathbf{V}] \begin{bmatrix} \vec{d} \\ \vec{w} \end{bmatrix}$ where $\mathbf{G}(\bar{\vec{S}}) \perp \mathbf{V}$, $\vec{d}$ defines the global affine transformation and $\vec{w}$ defines the generalized intrinsic shape parameter associated with $\mathbf{V}$.

Given an object of interest in an image, fitting the ASM model to the object involves minimizing the following cost function

$$C_{shape} = \|\vec{S}^{img} - \bar{\vec{S}} - [\mathbf{G}(\bar{\vec{S}}), \mathbf{V}] \begin{bmatrix} \vec{d} \\ \vec{w} \end{bmatrix}\|. \quad (2)$$

To ensure that the similarity transformation is Euclidean, we optimize the cost function with the constraint $(a+1)^2 + b^2 = s^2$, $(s > 0)$.

B. Graph matching by linear programming

In [1], the matching of two sets of points is formulated as a linear programming problem. Given a point set $\mathcal{M} = \{p\}$, we can pre-define a set of pair-wise neighboring model points $\mathcal{N} = \{p, q | p, q \in \mathcal{M}\}$. Let $f(p)$ be the target point in a query image matched to the model point p, the objective function for minimizing the matching cost is defined as follows

$$\min_f \{\sum_{p\in\mathcal{M}} c(p, f(p)) + \lambda \sum_{\{p,q\}\in\mathcal{N}} g(p, q, f(p), f(q))\}, \quad (3)$$

where $c()$ is the feature matching cost, which is minimized when p and $f(p)$ match correctly, and $g()$ is the spatial matching cost that can be minimized when the pairwise spatial connection (p, q) in the model is similar to that of $(f(p), f(q))$ in the target image. In particular, we can define $g(p, q, f(p), f(q)) = \|sR(p-q) - (f(p) - f(q))\|$ where s and R define the scaling and rotation respectively.

In [12], the rotation and scaling parameters are utilized in two separate terms applied to the model points and the target points respectively, and a matrix formulation is derived as follows

$$\min \epsilon(X, s, R) = tr(C'X) + \lambda|EMR - sEXT|, \quad (4)$$

subject to $X1_{n_t} = 1_{n_m}, X \in \{0, 1\}^{n_m \times n_t}$, $s > 0$ and $R'R = I$, where the following terms are defined:

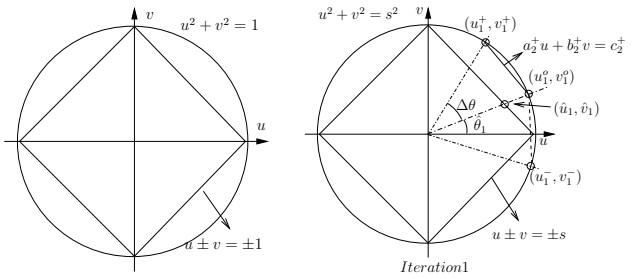
- $X_{n_m \times n_t}$ is a binary assignment matrix. $X(i, j) = 1$ defines a match between the i-th model point and the j-th target point. $X1_{n_t} = 1_{n_m}$ indicates each model point matches to only one target point.
- s is the global scaling parameter.
- R is the 2D rotation parameter.
- $M_{n_m \times 2}$ is the matrix for the 2D model points.
- $T_{n_t \times 2}$ is the matrix for the candidate 2D target points.
- $C_{n_m \times n_t}$ is the feature matching cost matrix, where $C(i, j)$ is the matching cost between the i-th model point and the j-th target point.
- $E_{n_e \times n_m}$ defines the graph edges of the neighboring model points. For edge e which connects the model point i to the point j, we have $E(e, i) = 1$ and $E(e, j) = -1$. $1 \le e \le n_e = |\mathcal{N}|$.

To remove the l_1 norm in Eq. 4, the cost function can be further linearized [12] as

$$\min \epsilon(X, s, R, Y, Z) = tr(C'X) + \lambda 1'_{n_e}(Y + Z)1_2 \quad (5)$$

subject to $Y - Z = EMR - sEXT$, $Y \ge 0$,$Z \ge 0$.

As illustrated in Figure 1-(a), the rotation matrix $R = [u, -v; v, u]$ is linearized by approximating the unit normal



(a) (b)

Fig. 1. The approximation of rotation on a unit circle. (a) The four line segments approximation in [1] ; (b) The finer line segment approximation of the unit circle at iteration 1.

constraint $u^2 + v^2 = 1$ with four line segment $u \pm v = \pm 1$, $|u| \leq 1$, $|v| \leq 1$.

The scaling parameter s is quantized into n_s variables, i.e., $0 < s_1 < s_2 < ... < s_{n_s}$, so that the assignment matrices $\{X_l\}$ are introduced, one for each scale n_l, $X = \sum_l X_l$. Since the binary assignment matrix is relaxed to permit real values in $[0, 1]$, the formulation is able to approximate any $s \in [s_1, s_{n_s}]$.

In [1] and [12], the linearization of the problem makes it possible to find the globally optimal model fit (for the re-formulated linear system) amid the clutter in the image.

III. Shape Model Fitting by Linear Programming

Based on Eq. 2 and Eq. 4, we can formulate shape model fitting in an LP framework by minimizing

$$\epsilon(X, \vec{d}, \vec{w}) = tr(C'X) + \lambda |E\Omega(\bar{\vec{S}} + [\mathbf{G}, \mathbf{V}] \begin{bmatrix} \vec{d} \\ \vec{w} \end{bmatrix}) - EXT| \quad (6)$$

subject to $X1_{n_t} = 1_{n_m}, X \in \{0,1\}^{n_m \times n_t}$, and $(\vec{d_1} + 1)^2 + \vec{d_2}^2 = s^2 (s > 0)$, where X,C,E, T, have been defined in Eq. 4; $\bar{\vec{S}}$, $\mathbf{G}$, $\mathbf{V}$, $\vec{d}$ and $\vec{w}$ defines the shape model in Eq. 2, and $\Omega(\cdot)$ is a matrix operator that reshapes the vector $[x_1, y_1, ...x_n, y_n]'$ into $\begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{bmatrix}'$. As the constraint on the scaling and rotation parameters is nonlinear, we propose our linearization procedure as follows.

A. Linearization

Noticing that the task of object shape fitting is usually carried out after object detection, and the input image can be normalized based on an approximate estimation of the object location, scale and orientation, we therefore do not put efforts in making LPSM scale and rotation invariant. Given the assumption, we simply assume the scaling factor is in a range of $[0.7\, 1.3]$. We also assume the object orientation angle is in a range of $[-30^o + 30^o]$.

In [12], the rotation matrix is modeled by variable u,v satisfying $u \pm v = \pm 1$, $|u| \leq 1$, $|v| \leq 1$. Four programs for Eq. 5 are carried out, one for each line segment as the constraint in the uv space. The optimal solution is obtained by the one that achieves the lowest matching cost at the converging stage.

Approximating the unit circle by four line segment introduces inaccuracy. For example, the magnitude of the rotation matrix represented uv in the line segment may be biased to $\cos(\frac{\pi}{4}) = 0.707 < 1$ in the worst situation. This causes an unnecessary bias to the estimation of the rotation and scaling factor. We propose to refine the approximation of the unit circle as the successive convexification is moving toward finer resolution. We define the rotation and scaling by the circle $u^2 + v^2 = s^2$ in the uv space, as shown in Fig. 1-(b). Assuming the optimal solution is found at $(\hat{u}_1, \hat{v}_1)$ on the line segment $u + v = s$ in iteration 1, the rotation angle can be computed as $\hat{\theta}_1 = \tan^{-1} \frac{\hat{v}_1}{\hat{u}_1}$. Assuming the angle estimation error is within the range of $\Delta\theta$, two line segments that approximate the arc within $\hat{\theta}_1 \pm \Delta\theta$ on the circle can be defined by the ending points (u_1^o, v_1^o) and $(u_1^\pm, v_1^\pm)$ with

$$\begin{aligned} u_1^o &= s\cos(\hat{\theta}_1), \qquad & u_1^\pm &= s\cos(\hat{\theta}_1 \pm \Delta\theta), \\ v_1^o &= s\sin(\hat{\theta}_1), \qquad & v_1^\pm &= s\sin(\hat{\theta}_1 \pm \Delta\theta). \end{aligned} \quad (7)$$

As successive convexfication is moving toward finer resolution, we can assume the error range for the angle estimation $\Delta\theta$ is gradually decreasing. The resulting line segments $a_i^\pm u + b_i^\pm v = c_i^\pm$ provide finer approximation to the circle $u^2 + v^2 = s^2$ in the uv space, which can be beneficial to the linear programming defined in Eq. 6.

B. The algorithm

We summarize the linear programming formulation of the shape model fitting problem at iteration i as follows:

$$\min \epsilon(X, \vec{d}, \vec{w}) = tr(C'X) + \lambda_e 1'_{n_e}(Y + Z)1_2 \quad (8)$$

subject to

- $Y - Z = E\Omega(\bar{\vec{S}} + [\mathbf{G}, \mathbf{V}] \begin{bmatrix} \mathbf{d} \\ \mathbf{w} \end{bmatrix}) - EXT$, $Y, Z \geq 0$;
- $X1_{n_t} = 1_{n_m}, X \in \{0,1\}^{n_m \times n_t}$.
- $u = d_1 - 1$, $v = d_2$, $a_i^\pm u + b_i^\pm v = c_i^\pm$, $0.7 < s < 1.3$;
- $u \in [u_i^o, u_i^\pm]$, $v \in [v_i^o, v_i^\pm]$, $|w| \leq 3\sigma_w$;

where σ_w is the standard deviation of the intrinsic shape parameters. As the translation of the model does not affect the cost, we set the exterior parameter $d_3 = d_4 = 0$. At each iteration, two linear programs are generated with line segment parameter $a_i^\pm$, $b_i^\pm$, $c_i^\pm$ computed according to Eq. 7, given the estimation of rotation angle $\hat{\theta}$ and assumed estimation error range $\Delta\theta$ at the previous iteration.

Therefore, taking advantage of the successive convexification technique in [1], the LPSM fitting algorithm is summarized in Algorithm 1.

IV. Implementation

The LPSM is implemented with a successive convexification scheme described in detail in [1]. In the first iteration, we let $\hat{\theta} = 0$ and $\Delta\theta = 30^o$. As the trust regions are getting smaller, we reduce the angle estimation range $\Delta\theta$ by $\frac{\pi}{M}$,$(M > 6)$, with M gradually increasing.

Algorithm 1: LPSM fitting algorithm

Data: Input image $\mathbf{I}$, local appearance model $p_j(\Gamma)$, mean shape $\bar{S}$, and shape basis matrix $[\mathbf{G}, \mathbf{V}]$

Result: the assignment matrix $X_{n_m \times n_t}$ and the shape parameter $[\vec{d}, \vec{w}]'$

Compute the appearance log-likelihood matrix C between the ASM landmark appearance model and the candidate target features points.
Initialize the trust region for each landmark to be the entire target image.
Compute the lower convex hull vertices of matching costs for each landmark within its trust region.
Let $\hat{\theta} = 0$ and $\Delta\theta = 30^o$, and solve two linear programming in Eq. 8.
Update the trust regions.
repeat
    find the linear program that has the lowest matching cost.
    Update the trust regions.
    Solve two programs defined by Eq. 8 given the previous estimation of the rotation angle $\hat{\theta}$ and $\Delta\theta$.
until *The trust region is small enough.*
return X,$\vec{d}$, $\vec{w}$.

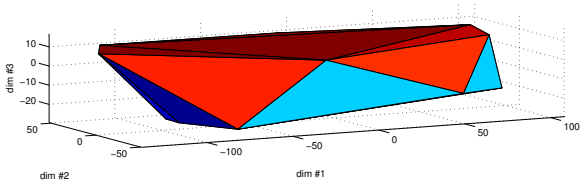
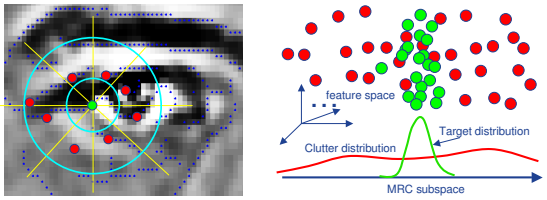


Fig. 2. Convex hull constraints for intrinsic shape space



(a) Training data (b) MRC intuition

Fig. 3. Training of MRC appearance likelihood model

A. Convex hull constraints in the intrinsic shape space

In Eq. 8, the constraints $|w| \leq 3\sigma_w$ limits the search of the intrinsic shape parameter in a hypercube in the parameter space. The hypercube however is not compact enough to describe the distribution of the allowable shape instances from real world. In [16], Li and Ito proposed to adding additional shape constraints by quantizing the shape space and modeling the allowable intrinsic shape parameter distributions by histograms in the top k most correlated dimensions in the parameter space. While the computation may become intractable if $k > 3$, the paper proposed an approximation solution. For each element of the shape parameters, it is tractable to build histogram to model the joint distribution between itself and the elements in two most correlated dimensions. The top two most correlated dimensions can be found by computing the correlation coefficient between the elements from the training data instances. Therefore, the joint distribution of the allowable shape instances in the shape space can be approximated by a number of 3-dimensional histogram tables.

Following similar intuition, we propose the region of the allowable shape instances in the shape space can be approximated by convex hulls in a number of 3 dimensional most correlated subspaces. We identify the 3-dimensional most correlated subspaces using a similar procedure in [16], and we then generate simplified convex hull based on the shape instances from the training set. We then add the hyperplane constraints defined by the convex hulls to the LP formulation so that the LP output is allowable shape instance in real world. Fig. 2 shows one typical convex hull in a 3 dimensional correlated space.

B. Model appearance likelihood in Maximal Rejection Classifier(MRC) subspace

Given a testing face image, we generate the candidate feature points by canny edge detection. The feature matching cost matrix C is then evaluated for each landmark using a landmark appearance likelihood model. We train the landmark appearance likelihood model as a Gaussian model to capture the landmark appearance statistics in the subspace learned by Maximal Rejection Classifier [13]. In order to reduce the illumination variations, the face images are preprocessed using DoG filtering similar to [17].

As shown in Fig. 3-(a), given a set of face images with landmark labels as the training set, we crop the image patches centering at each landmark (the green dot) as the *Target* class data. To acquire the *Clutter* class data, we first carry out canny edge detection (the blue dots) at the neighboring area of the landmark in each image; We evaluate the normalized correlation between the image patch on the landmark and the image patches centering on the canny edge points that are 4-10 pixels away from the landmark. We then acquire eight image patches centering on the edges (as indicated by the red dots) that have the maximal normalized correlation in the surrounding eight sub-regions. Fig. 3-(b) shows the intuition of Maximal Rejection Classifier. The MRC subspace minimizes the within class covariance for the $Target$ data, and spreads out the $Clutter$ data as much as possible. This intuition fits well with the landmark detection scenario.

C. Getting the fitted landmarks

LPSM solves Eq. 8 and obtains the solution X, $\vec{d}$ and $\vec{w}$. The fitted landmarks can be computed either as XT or as $S = \bar{S} + [U, V] \begin{bmatrix} \vec{d} \\ \vec{w} \end{bmatrix}$. XT is the landmark locations based on the actual feature assignment for each landmark according to the assignment matrix X. S is the intrinsic active shape model obtained through the linear programming optimization. In general, XT latches onto the actual image features (i.e., edges or corners), which can be noisy in the presence of outliers, but could be preferable when the shape space can not fully capture the observed shape variations and the canny edge detection is free of outliers; S is usually smoother since it is regularized by the shape subspace. In this paper, we utilize the landmark estimation S for LPSM.

	PCA	FDA	MRC
PE	0.348	0.141	0.138
$\bar{D}_{gt}(pixels)$	8.25	8.11	4.37

TABLE I

COMPARISON OF THE PCA, FDA AND MRC LIKELIHOOD MODEL.

V. EXPERIMENTS

We prepare our experimental evaluations on the PUT database[18]. The PUT database contains 2193 face images of 100 people with ground truth contour labels. For each person, 22 images are captured in different head poses and some scale variations and facial expression changes, in partially controlled illumination conditions over an uniformed background. Among the 22 face images for each person, we pick six images that sample across the range of the head poses, and obtain in total 600 face images with ground truth. While the original image data are of high resolution, we reduced the size of the image so that the distance between the eyes in the face is about 50-80 pixels, and we picked 68 landmarks from the ground truth contour model to characterize the facial shape in the data. We utilize the first 500 face images ($\sim$ 80 subjects) as the training set and we evaluate the fitting performance of the algorithms on the rest 100 face images ($\sim$ 20 subjects). The fitting accuracy is computed as average distances in pixels between the fitted landmarks and the ground truth shape contour. We implemented LPSM in Matlab environment, and we observe the computation takes about 1-2 minutes per image depending on how much Canny edge points are detected in the image.

A. Evaluation of the MRC likelihood model

We first evaluate the performance of the MRC appearance likelihood model. We train the likelihood models in PCA, FDA and MRC subspaces, respectively, for the image patches (of size 25×25) at the landmarks on the training data. Fig. 4 shows the negative log-likelihood surfaces of the PCA, FDA, and MRC model centering at the left eye corner of a subject. It shows that the peak of the MRC surface is more discriminative than PCA surface, and is smoother and contain less minima outliers than FDA surface. We then compute the minimal classification error based on testing image patches cropped at the landmarks in the testing set, and we also compute the average distances from the local minima of the cost surfaces to the ground truth landmark locations over all the landmarks for the testing data. Table I shows that, MRC produces similar performance as FDA in terms of classification error, but generate the minimal average localization error in terms of distance from the ground truth locations. Considering the fact that LPSM depends on the evaluation of appearance likelihood on canny edges, which are typically a few pixels away from the likelihood surface local minima, we consider the properties of MRC fits well to the design requirement for our LPSM algorithm.

B. Evaluation of LPSM

As a comparison, we compare the performance of our proposed **LPSM** against the following algorithms:

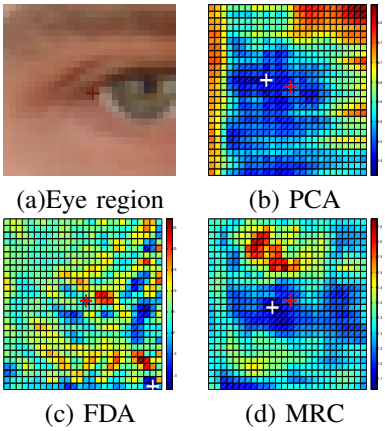


Fig. 4. The appearance negative log-likelihood surface at the eye corner. The ground truth location is marked by a red cross, and the minima of the surface is marked by a white dot. The minima location of MRC surface is the closest to the ground truth.

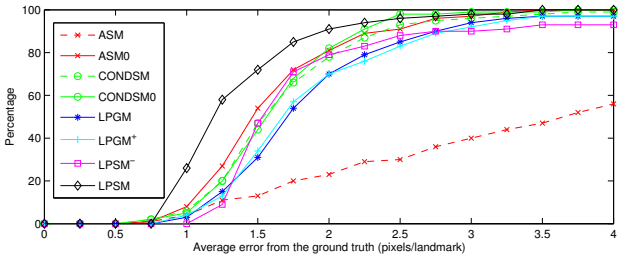


Fig. 5. Performance comparison between ASM, ASM0, CONDSM, CONDSM0, LPGM, LPGM^{+}, LPSM^{-} and LPSM. The fitting error is computed as average distances in pixels from the fitted landmarks to the ground truth shape contour with eye-to-eye distance normalized to 65 pixels.

LPSM^{-}: LPSM without the rotation parameter refinement.
LPGM: LPSM reduced to graph matching by degenerating the shape subspace of LPSM to the mean face. The fitting result is retrieved as XT.
LPGM^{+}: After fitting LPGM and obtain XT, we project XT into the PCA shape space and do reconstruction, so that the outliers in XT can be subdued.
ASM: The standard ASM shape fitting algorithm initialized with model centroid shifted away from the ground truth by 15% of the distance between the eyes. The ASM adopted a local search based optimization strategy closely related to the Iterative Closest Point (ICP) approach[19].
ASM0: ASM initialized with ground truth landmark labeling.
CONDSM: A CONDENSATION based ASM fitting algorithm[8], initialized with initial guess of the face centroid shifted by 15% of the between-eye distance from the ground truth. 100 samples were employed for the evaluation of distributions in the parameter space.
CONDSM0: CONDSM initialized with initial guess of face location centering at the ground truth.

Fig. 5 shows the cumulative distribution of the average landmark localization errors for ASM, ASM0, CONDSM, CONDSM0, LPGM, LPGM^{+}, LPSM^{-}, and LPSM. It shows that LPSM achieves the highest accuracy, as about

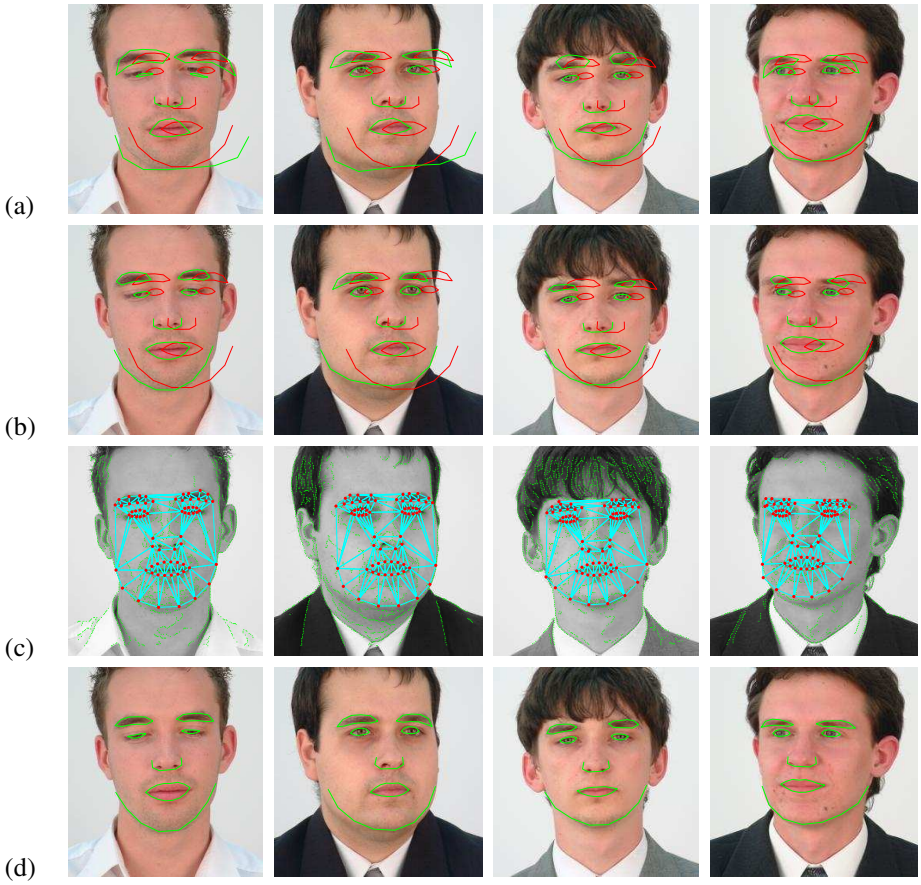


Fig. 6. Comparison of some fitting results on PUT database. (a) shows ASM results and (b) shows CONDSM results. The initial face shapes are in red, and the fitted face shapes are in green. (c) shows LPGM result and (d) shows our proposed LPSM result. In the first row, the feature points (canny edges) are shown in green, and the feature assignment are marked in red dots; The second rows show the fitted shape models.

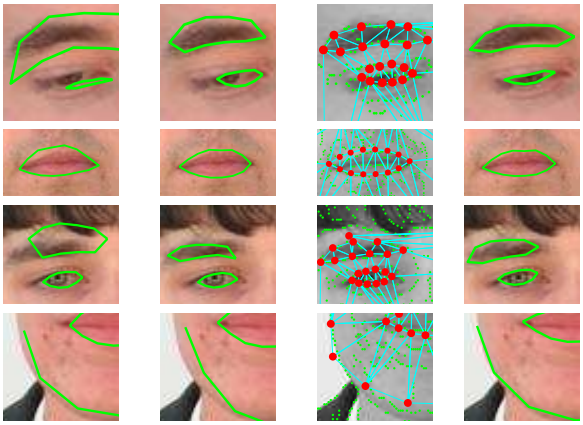
60% LPSM landmarks are within 1.25 pixels from the ground truth landmarks, while this percentage for the rest algorithms are less than 20%. A paired t-test on the results shows that the improvement is statistically significant(with $p < 0.0005$). Without the rotation parameter refinement, the performance of LPSM^- drops significantly due to inaccurate estimation of rotation parameters. LPGM and LPGM^+ can achieve pretty good localization with the mean face shape as graph template but can not reach as high fitting accuracy as LPSM because the fitting of the shape deformation is not constrained by a shape model. We also observe ASM0 and CONDSM0 achieves pretty high accuracy as their parameters are initialized properly, and that CONDSM can avoid local minima problem much better than ASM through sampling technique. However they can not achieve as high localization accuracy as LPSM.

In Fig. 6, we show some typical fitting results from the PUT database. Fig. 6-(a),(b) shows the initialized and fitted shape models in red and green for ASM and CONDSM, respectively. And Fig. 6-(c) and (d) shows the feature assignment XT for LPGM and the intrinsic shape S for LPSM. Comparison on the fitting details can be found in Fig. 5. LPSM produces fitting results that latch onto the canny edges along the facial feature contour, thus are more visually satisfactory.

We then applied LPSM fitting to random real world face images collected from CMU expression face database and MBGC database. With the same Canny edge detection parameters, we find LPSM achieves very promising fitting performance given the presence of strong variations of illumination, facial expression, skin color, gender of the faces in the images. Some of the results are shown in Fig. 8.

VI. Conclusions

In this paper, we proposed to do shape model fitting in images by finding the optimal solution using LP techniques (LPSM). LPSM conveniently integrates the top-down (subspace based shape and appearance models) and bottom-up (Canny edge detection) image analysis techniques based on the LP optimization framework with l_1 norm cost formulation, and can achieve very promising and robust shape fitting performance. The advantages of LPSM can be summarized as follows: 1) LPSM can fit object in images of strong background clutters and do not rely heavily on the initial guess of the model parameters. 2) By fitting the shape model to canny edges, LPSM can produce fitting contours that latches on the object contour edges and look visually satisfactory. The resulting fitting consistency on the object edge boundaries makes the fitting result promising for applications, such as subtle facial expression analysis.



(a) ASM (b) CONSM (c)LPGM (d) LPSM

Fig. 7. Comparison on the fitting details. LPSM fits the shape model to the canny edges and thus usually can produce results that are more visually satisfactory. Though the mean shape template for LPGM can not capture the facial shape variations, LPGM can do reasonably good feature assignment XT when the deformation is not substantial and the feature detection is free of outliers.

Fig. 8. Some LPSM fitting results for images from various face databases.

While LPGM can do reasonable shape fitting through feature assignment XT guided by a static shape model, LPSM incorporates a shape subspace model that explains away much of the intrinsic shape variations in the shape matching error, and thus guides the linear programming optimization toward more accurate feature assignment solution. LPSM can produce shape fitting results by feature assignment XT and by the intrinsic shape model S. Both may be useful depending on the different applications. For applications where the object in the images deforms in predictable ways, but may yield candidate features with strong outliers, shape fitting by S may be the better choice as it is regularized by the shape space in l_1 norm. For applications where the object in testing images deforms in ways that may not be captured in the training data, but produces good candidate features in the images, shape fitting by XT could be more desirable.

In this paper, we also proposed to model the landmark appearance likelihood in the Maximal Rejection Classifier (MRC) subspace trained on raw image patches for LPSM, considering the fact that the design of MRC matches nicely to the purpose of discriminating the landmark appearances against its surrounding regions in this scenario. We showed that modeling the landmark appearance likelihood in MRC subspace can achieve better performance than in PCA, and LDA subspaces. Due to its linear formulation, it can be evaluated efficiently. In the mean time, we believe the performance of our proposed model can be further improved by employing more powerful feature extraction methods (such as SIFT or HOG), and by utilizing more advanced learning techniques (such as Adaboosting). Finally, it is still an open question on how to make LPSM real-time. We believe parallel computing and coarse-to-fine image pyramid analysis techniques can be considered for improving the computational efficiency.

REFERENCES

[1] Jiang, H., Drew, M.S., Li, Z.n.: Matching by Linear Programming and Successive Convexication. IEEE TPAMI **29** (2007) 959–975

[2] Cootes, T., Taylor, C.: Active shape models. In: The 3rd BMVC. (1992) 266–275

[3] Cootes, T., Taylor, C.: Constrained active appearance models. Proceedings of International Conference on Computer Vision (2001) 748–754

[4] Romdhani, S., Gong, S., Psarrou, A.: A multi-view nonlinear active shape model using kernel pca. In: Proceedings of the 1999 British Machine Vision Conference. (1999) 483–492

[5] Liu, C., Shum, H.Y., Zhang, C.: Hierarchical shape modeling for automatic face localization. In: ECCV. Volume 2. (2002) 687–703

[6] Jiao, F., Li, S., Shum, H.Y., Schurmans, D.: Face alignment using statistical models and wavelet features. In: CVPR. (2003) 321–327

[7] Yan, S., Li, M., Zhang, H., Cheng, Q.: Ranking prior likelihood distributions for Bayesian shape localization framework. In: IEEE International Conference on Computer Vision. (2003) 51–58

[8] Tu, J., Zhang, Z., Zeng, Z., Huang, T.: Face localization via hierarchical condensation with fisher boosting feature selection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Volume 2. (2004) 719–724

[9] Zhou, Y., Gu, L., Zhang, H.: Bayesian tangent shape model: Estimating shape and pose parameters via Bayesian inference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Volume 1. (2003) 109–118

[10] Liang, L., Wen, F., Xu, Y.Q., Tang, X., Shum, H.Y.: Accurate face alignment using shape constrained markov network. CVPR **1** (2006) 1313–1319

[11] Conte, D., Foggia, P., Sansone, C., Vento, M.: Thirty years of graph matching in pattern recognition. (IJPRAI)

[12] Jiang, H., Yu, S.: Linear solution to scale and rotation invariant object matching. IEEE CVPR (2009) 2474–2481

[13] Elad, M., Hel-Or, Y., Keshet, R.: Pattern detection using a maximal rejection classifier. Pattern Recognition Letters **23** (2001) 1459–1471

[14] Matthews, I., Baker, S.: Active appearance models revisited. International Journal of Computer Vision **60** (2004) 135 – 164

[15] Zhou, Y., Gu, L., Zhang, H.: Bayesian tangent shape model: Estimating shape and pose parameters via Bayesian inference. In: Proc. IEEE Computer Vision and Pattern Recognition, Madison, WI. Volume 1. (2003) 109–116

[16] Li, Y., Ito, W.: Shape parameter optimization for adaboosted active shape model. Computer Vision, IEEE International Conference on **1** (2005) 251–258

[17] Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: FG07. (2007) 168–182

[18] Kasinski, A., Florek, A., Schmidt, A.: The PUT face database. Technical report, Poznan University of Technology, Poznan, Poland (2009)

[19] Ruto, A., Buxton, B.: Application of a robust and efficient icp algorithm for fitting a deformable 3d human torso model to noisy data. Digital Image Computing: Techniques and Applications **0** (2005) 62