# A Robust Shape Model for Multi-view Car Alignment

Yan Li Leon Gu * Takeo Kanade
Carnegie Mellon University
{yanli,gu,tk}@cs.cmu.edu

## Abstract

*We present a robust shape model for localizing a set of feature points on a 2D image. Previous shape alignment models assume Gaussian observation noise and attempt to fit a regularized shape using all the observed data. However, such an assumption is vulnerable to gross feature detection errors resulted from partial occlusions or spurious background features. We address this problem by using a hypothesis-and-test approach. First, a Bayesian inference algorithm is developed to generate object shape and pose hypotheses from randomly sampled partial shapes - subsets of feature points. The hypotheses are then evaluated to find the one that minimizes the shape prediction error. The proposed model can effectively handle outliers and recover the object shape. We evaluate our approach on a challenging dataset which contains over 2,000 multi-view car images and spans a wide variety of types, lightings, background scenes, and partial occlusions. Experimental results demonstrate favorable improvements over previous methods on both accuracy and robustness.*

## 1. Introduction

Deformable shape matching has been studied extensively in the past two decades with the emphasis on the alignment of human faces and anatomical structures. Representative work include Snakes [14], Active Shape Model [4] and Active Appearance Model [3], Bayesian shape model [24, 13], nonlinear shape models [4, 19, 25], view-based [6] and three dimensional models [1, 12], and models for weak initialization [16, 17, 23].

A common assumption in these models is that the observation noise is Gaussian distributed. However, in real-world images shape observations are usually corrupted by large-scale measurement errors which are in gross disagreement with the true underlying shape. Such errors, usually called outliers, are caused by the failures of the appearance model or undesirable conditions such as shadows and occlusions. Due to the iterative nature of most algorithms, these gross errors may become arbitrarily large and therefore cannot be "averaged out", as is typically done in the least-squares framework. Rogers and Graham [18] attempt to address this problem by use of M-estimators. However, M-estimators tend to suffer from local optima, and pose parameters have been ignored in their model.

Another limitation in the previous models is the sensitivity to initialization. The objective functions are usually highly nonlinear and a suboptimal initialization may cause the model to get stuck at local minimums. Previous work attempt to tackle this problem by sampling - starting from multiple initializations and choosing the optimal resulting shape [17, 23]. However, each individual sample was evaluated and matched in a least-squares fashion, so that the alignment process could still fail in the presence of outliers. It is also unclear how many samples are sufficient to achieve the best solution.

In this paper, we address these two problems in a hypothesis-and-testing framework. Our key insight is the following: since object shape typically resides in a low-dimensional subspace, the degree-of-freedom of a shape model is considerably less than the number of the observed features; therefore, a small subset of "good" features are sufficient to "jump start" the matching and produce a reasonable estimate. We adopt the random sample consensus (RANSAC) paradigm of Fischler and Bolles [9]. In particular, a Bayesian inference algorithm is developed for a generating shape and pose hypothesis from a randomly sampled subset of features; each hypothesis is matched against the full observation by a robust measure to identify the optimal one; and the hypothesis is further refined by incorporating more inliers into the corresponding subset.

We apply the approach to multi-view car alignment - identifying detailed car shapes from different viewpoints. The task is challenging because car images are often subject to significant amount of *occlusions*, and detecting individual parts are difficult. Combining our alignment model with a random forest [2] based detector we develop a robust, fully automatic car alignment system.

*Partial support provided by National Science Foundation (NSF) Grant IIS-0713406.

## 2. Problem Formulation

Consider the shape of a deformable object which consists of a set of 2D landmark points. Let $Y = (u_1, v_1, \ldots, u_N, v_N)^T$ denote the locations of the points observed from an input image. The observation contains not only noises, but also gross *outliers*. Our goal is to estimate the true underlying shape from such observation, and identify the outliers.

Instead of using the whole observation $Y$ for estimation, we will first use a randomly selected subset of $Y$, denoted by $Y_p$, to generate a shape hypothesis. The subset of points are postulated as inliers which, by assumption, satisfy the underlying noise model,

$$Y_p = M_p \mathcal{T}_\Theta(S) + \eta. \tag{1}$$

The vector $S$ denotes the normalized true shape which we refer to as *canonical shape*. It is transformed onto the image plane by $\mathcal{T}_\Theta(S) = sRS + t$ with rotation $R$, scale $s$ and translation $t$. $M_p$ is a $2M \times 2N$ indicator matrix which specifies the subset. Observation noise $\eta \sim \mathcal{N}(0, \Sigma)$ is assumed to be independent for individual points. One should note that large-scale measurement errors will not conform with the Gaussian noise assumption, therefore the model (1) applies only to $Y_p$.

The canonical shape $S$ is parameterized by a probabilistic PCA model [22, 24],

$$S = \mu + \Phi b + \epsilon \tag{2}$$

with the mean shape $\mu$, the low-dimensional eigen subspace spanned by $\Phi$, and the shape deformation parameter $b$. Each element of $b$ controls the magnitude of deformation along the corresponding axis in the subspace. A diagonal prior

$$b \sim \mathcal{N}(0, \Lambda) \tag{3}$$

is put on $b$, where $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_r\}$, and $\lambda_i$'s are eigenvalues. The shape noise $\epsilon$ is chosen to be isotropic, $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, and its variance $\sigma^2 = \frac{1}{2N-r}\Sigma_{i=r+1}^{2N}\lambda_i$ is determined by the residual, off-eigenspace shape energy.

Combining (1)$\sim$(3), we have established a hierarchical probabilistic model that can be used for generating hypotheses. Specifically, our problem is to estimate shape deformation $b$ and pose $\Theta = \{R, s, t\}$ from a partial shape $Y_p$, *i.e.*, find the MAP $\{b^*, \Theta^*\} = \underset{b,\Theta}{\text{argmax}}\, p(b, \Theta | Y_p)$. This is a typical missing data problem that can be solve by Expectation-Maximization as described in Sec. 3. Given a hypothesis of $b$ and $\Theta$, we can easily hallucinate the rest part of the shape

$$Y_h = M_h(sRS + t), \tag{4}$$

where $M_h$ is a binary matrix that indicates the remaining set of points. The hallucinate shape $Y_h$ is then used to test the hypothesis. Sec. 4 explains the details.

## 3. Generating Hypotheses from Partial Shapes

In this section, we develop a *Bayesian Partial Shape Inference* (BPSI) algorithm to estimate the model parameter $\Pi = \{b, \Theta\}$ iteratively. Detailed algorithm is shown in Alg. 1.

The inference can be performed by the standard EM algorithm. In the *E-step*, we compute the posterior of $S$ given the partial observation $Y_p$ and $\Pi^{(t-1)}$. Note that $S$ represents an "augmented" shape which can be decomposed into the partial observation $S_p$ and the hallucinated shape $S_h$. In particular, the posterior means are given by Eqn. 6 and 7. It shows that the hallucinated shape $S_h$ is generated completely from the shape prior, while $S_p$ subsumes two sources of information: one arises from the observation $Y_p$; the other reflects the subspace constraint on $b$. $S_p$ is essentially a weighted average of the observation and prior and the weight is determined by the two sources of noise.

In the *M-step*, we optimize $\Pi^{(t)}$ which maximize the expectation of the complete log-likelihood $\log p(Y_p, S|\Pi^{(t)})$ over the posterior of $S$ obtained in E-step. It shows that $b$ and $\Theta$ can be optimized independently.

One important parameter that has yet to be defined is the observation variance $\Sigma = \text{diag}\{\rho_1^2, \rho_1^2, \ldots, \rho_M^2, \rho_M^2\}$. Since shape alignment can be viewed as an iterative model fitting process, the observation noise can be estimated from the last iteration. In our implementation, $\rho_i$ is defined as the prediction residual

$$\rho_i^2 = ||Y^{(t-1)} - \mathcal{T}_{\Theta^{(t-1)}}(S^{(t-1)})|| \tag{5}$$

where $||\cdot,\cdot||$ denotes the Euclidean distance, and $\mathcal{T}_\Theta$ is the rigid transform which brings the canonical shape $S$ to the observation space by $\Theta$.

### 3.1. Discussion

The BPSI algorithm provides us some insights to the noise-presenting shape model. However, from the optimization point of view, the objective function and search method remain obscure. In this section, we re-examine the BPSI algorithm and focus on its optimization method.

In Step 6, we first compute the posterior mean of $p(b|S)$ which can be viewed as the probabilistic version of PCA projection. In addition to the subspace projection performed in PCA, BPSI applies an inhomogeneous shrinkage on each subspace dimension. The shrinkage parameter is defined by

$$\gamma_i = \frac{\lambda_i}{\lambda_i + \sigma^2} \quad (i = 1, \ldots, r) \tag{9}$$

Recall that $\sigma^2$ is the average of the remaining eigen-values. Since $b$ captures significant amount of variance (98% in our implementation), $\sigma^2$ has a very small value (*i.e.*, $\sigma^2 \approx 0$ and $\gamma_i \approx 1$). This implies that the PCA projection and reconstruction in Step 6 would not alter $S_p$ substantially.

**Algorithm 1** Bayesian Partial Shape Inference (**BPSI**)

**Input:** Partial observation $Y_p$. $b'$ and $\Theta'$ from last iteration.
**Output:** Updated $b$ and $\Theta$.

1: Initialize $b = b'$ and $\Theta = \Theta'$
2: **for** $t$ = 1 to $T$ **do**
3: **E-Step:**
4: Update $S_p$ by blending, and $S_h$ by reconstruction

$$\tilde{S}_p \leftarrow W_1 \mathcal{T}_\Theta^{-1}(Y_p) + W_2(\Phi b + \mu)_p \quad (6)$$
$$\tilde{S}_h \leftarrow (\Phi b + \mu)_h \quad (7)$$

where

$$W_1 = s^2\sigma^2(s^2\sigma^2 I + \Sigma)^{-1}$$
$$W_2 = I - W_1$$

5: **M-Step:**
6: Estimate shape

$$b \leftarrow \Lambda(\Lambda + \sigma^2 I)^{-1}\Phi^t(\tilde{S} - \mu)$$
$$S_p \leftarrow (\Phi b + \mu)_p$$

7: Estimate pose (Procrustes analysis [11])

$$\Theta \leftarrow \arg\min_\Theta ||Y_p - \mathcal{T}_\Theta(S_p)|| \quad (8)$$
$$M \triangleq (S_p - \bar{S}_p)(Y_p - \bar{Y}_p)^t, \quad [U, W, V] \triangleq \text{SVD}(M)$$
$$R = VU^t, \quad s = \text{tr}(W)/\text{tr}(M), \quad t = -sR\bar{S}_p + \bar{Y}_p$$

8: **end for**

Based on this observation, we can plug Eqn. 6 into Eqn. 8

$$\begin{aligned} Y_p - \mathcal{T}_\Theta(S_p) &\approx Y_p - \mathcal{T}_\Theta\left[W_1\mathcal{T}_\Theta^{-1}(Y_p) + W_2(\Phi b + \mu)_p\right] \\ &= Y_p - [W_1 Y_p + W_2\mathcal{T}_\Theta(\Phi b + \mu)_p] \\ &= (I - W_1)Y_p - W_2\mathcal{T}_\Theta(\Phi b + \mu)_p \\ &= W_2\left[Y_p - \mathcal{T}_\Theta(\Phi b + \mu)_p\right] \end{aligned}$$

It shows that Step 7 in BPSI solves a weighted least-squares problem

$$\min_\Theta \sum_{i=1}^{M} w_i(\rho_i)||Y_i - \mathcal{T}_\Theta(S_i)|| \quad (10)$$

The weight $w_i$ is a quasiconvex function of $\rho_i$

$$w_i(\rho_i) = \left(\frac{\rho_i^2}{s^2\sigma^2 + \rho_i^2}\right)^2 \quad (11)$$

Fig. 1 shows the profile of the weight function. Recall that $\rho_i$ is defined as the prediction residual from the previous step (Eqn. 5). Thus, *the BPSI algorithm minimizes the sum of square errors via the iterative reweighted least-squares* (IRLS).

$$\min \sum_{i=1}^{M} w_i(\rho_i^{(t-1)})\rho_i^2(b, \Theta) \quad (12)$$

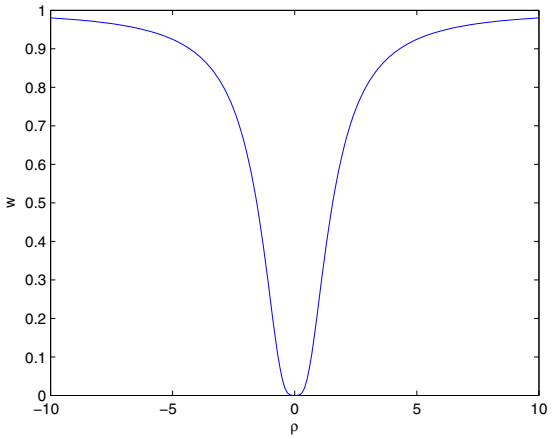


Figure 1. Graphic representation of the quasiconvex weight function.

## 4. Testing Hypotheses by RANSAC

If we have *a priori* knowledge about $Y_p$ such that it contains only inliers, the partial inference algorithm provides a principled framework for shape and pose estimation. However, a random feature subset may potentially contain outliers and the fitted parameters can become arbitrary. In this section, we adopt the random sample consensus (RANSAC) paradigm of Fischler and Bolles [9] to generate a large number of hypotheses and identify the optimal feature subset.

In the RANSAC framework, a minimal subset of features are used to estimate the model parameters. Specifically, our model requires six parameters (which capture 98% variance) to describe the shape $b$, and four parameters (scale/rotation/translation) to represent the pose $\Theta$. Since each 2D point provides two constraints on the parameters, five points are sufficient to form a *proposal subset* $Y_p$.

Ideally every possible subset would be considered, but this is usually computationally infeasible. Fischler and Bolles [9] proposed that the number $m$ of subsets can be chosen sufficiently high to achieve statistical significance. Assuming that the whole set of points may contain up to a fraction $\gamma$ of outliers, one can determine $m$ by

$$m = \frac{\log(1 - P)}{\log(1 - (1 - \gamma)^p)} \quad (13)$$

where $p = 5$ is the number of features in one subset. $P$ is the expected chance that at least one of the proposal subsets is good. In our implementation, we assume $\gamma = 40\%$ and require $P = 0.99$, thus $m = 57$.

Given the proposal subsets $Y_p^{(k)}$ $(k = 1, \ldots, K)$, the resulting shape $b$ can be obtained by the least median of squares (LMedS) estimator [20]

$$\min_k \text{Med}_i\ r_i^2\left(Y_{\bar{p}}^{(k)}, Y_h^{(k)}\right)$$

Figure 2. The partial shape $Y_p$ (red dots) is used to hallucinate the remaining shape $Y_h$ (gray dots). The marginal variance of the hallucinated points can be calculated and shown here in ellipses.

where $r_i$ is the residual between the $i$-th corresponding point of $Y_{\bar{p}}$ and $Y_h$.

In the traditional RANSAC literature, one usually assumes no *a priori* knowledge about the target model and the voting inliers are assumed to be *iid*. For instance, in the line fitting example, any two points can determine a model and the residual is simply the Euclidean distance from a voting sample to the fitted line. However, in a deformable shape alignment task varying amounts of residuals should be accommodated to deal with the inherent shape variation.

Note that the the hallucinated shape $Y_h$ is generated from $b$ through the canonical shape $S$. By propagating the information in $b$, we obtain the prior distribution of $Y_h$

$$\begin{aligned} \mathrm{E}[Y_h] &= M_h(sR\mu + t) \\ \mathrm{Var}[Y_h] &= s^2 M_h R(\Phi\Lambda\Phi^t + \sigma^2 I)R^t M_h^t \end{aligned}$$

In general, the points in $Y_h$ are correlated, thus the LMedS estimator cannot be applied directly. To remedy this problem, we make an independent assumption and use the marginal variance $\Sigma_i$ of each point to compute the residual

$$r_i^2(Y_{\bar{p}}, Y_h) = [Y_{\bar{p}}(i) - Y_h(i)]^t \, \Sigma_i^{-1} \, [Y_{\bar{p}}(i) - Y_h(i)] \quad (14)$$

$r_i$ is essentially the Mahalanobis distance between $Y_{\bar{p}}(i)$ and $Y_h(i)$. Fig. 2 illustrated the inhomogeneous prior variance exhibited in $Y_h$.

Although the LMedS estimator is highly resistent to outliers, it has a relatively low statistical efficiency and the estimate tends to be variable [21]. A post-processing must be employed to incorporate more inliers and re-estimate the model. Alg. 2 summarizes the complete hypothesis-and-test algorithm.

**Algorithm 2** Robust Shape Alignment

**Input:** Observation $Y$. $b'$ and $\Theta'$ from last iteration.
**Output:** Regularized $Y$. Updated $b$ and $\Theta$.

1: Generate random subsets $Y_p^{(1)}, Y_p^{(2)}, \ldots, Y_p^{(K)}$
2: **for** $k$ = 1 to $K$ **do**
3: $\quad \left[b^{(k)}, \Theta^{(k)}\right] \leftarrow$ **BPSI** $\left(Y_p^{(k)}, b', \Theta'\right)$
4: $\quad$ Hallucination: $Y_h^{(k)} \leftarrow M_h \mathcal{T}_{\Theta^{(k)}}(\Phi b^{(k)} + \mu)$
5: $\quad \epsilon^{(k)} \leftarrow \underset{i}{\text{Median}}\ r_i^2 \left(Y_{\bar{p}}^{(k)}, Y_h^{(k)}\right)$
6: **end for**
7: $\hat{k} \leftarrow \arg\min_k \epsilon^{(k)}$
8: Include more inliers to $Y_p^{(\hat{k})}$ and run BPSI to refine $b$ and $\Theta$
9: $Y \leftarrow \mathcal{T}_\Theta(\Phi b + \mu)$

## 5. Experiments

### 5.1. The Dataset

We evaluate our model on the MIT StreetScene dataset [1]. This dataset contains over 3,000 street scene images which were originally created for the task of object recognition and scene understanding under uncontrolled environment. We labeled 3,433 cars which span a wide variety of types, sizes, background scenes, lighting conditions, and partial occlusions. All the shapes are normalized to roughly the size 250x130 by the Generalized Procrustean Analysis [8]. The labeled data were manually classified into three views: 1,400 half-front view, 803 profile view and 1,230 half-back view. We randomly select 400 images from each view for training, and the rest for testing. For the occluded landmarks, we place their label at the most probable locations, but the corresponding local patches are excluded during training the appearance model.

### 5.2. Learning the Discriminative Appearance Model

The goal of appearance model is to provide an initial shape for the alignment algorithm. Due to the background clutter and substantial variations in color and pose, it is very challenging to capture the local appearance. To address the problem, we take a discriminative approach and learn the appearance density from the data.

We generate training samples from the labeled car images of three different views. The car shape is represented by 14, 10, and 14 landmarks respectively. For each landmark, we extract a 40x40 image patch as the positive sample. Negative samples of the same size are extracted uniformly around three concentric-circles centered at the landmark with 5 pixels apart. 36 negative samples are collected

[1] http://cbcl.mit.edu/software-datasets/streetscenes/

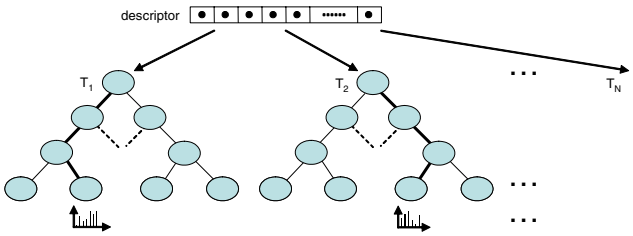


Figure 3. Random forest for posterior estimation. The descriptor is dropped to $N$ decision trees. The final posterior is the average over all the resulting histograms reached by the input descriptor.

for each landmark.

Local patches are further described by the Histogram of Oriented Gradients (HOG) descriptor [7]. The HOG descriptors are computed over dense and overlapping grids of spatial blocks, with image gradient features extracted at 9 orientations and gathered into a 576-dimensional feature vector (we use 8x8 cells, and 2x2 blocks).

The extracted descriptors are fed to a Random Forest [2] for discriminative learning. A random forest is essentially an ensemble of decision trees which are induced by bootstrapped data. Specifically, we adopt the *Extremely Randomized Trees* of Geurts *et al*. [10] for training. The random forest consists of $N$ randomly generated decision trees, each of which is trained by 5000 bootstrapped samples. At each non-terminal node, two random dimensions, denoted by $i$ and $j$, are chosen from the descriptor $d$. The splitting measure at that node is specified as

$$B(d) = \begin{cases} 1, & \text{if } d(i) < d(j) \\ 0, & \text{otherwise} \end{cases}$$

where $B(d)$ indicates the branch that $d$ should continue. At each terminal node, we save a normalized histogram that counts the frequency of each class reaching the node. Our random forest representation is similar to the feature classification trees by Lepetit *et al*. [15]. However, our task is to estimate the posterior of the landmark given the observed patch rather than classify it into different categories.

Since the decision trees are generated randomly, we can even combine all the landmarks into one random forest. In this case, each landmark represents a distinct class, while all the negative samples from different landmarks are combined into one single negative class. The resulting random forest is shown in Fig. 3. Given an input descriptor $d$, the posterior that it belongs to landmark $l_i$ is given by

$$p(l_i|d) = \frac{1}{N}\sum_{j=1}^{N} p_j(l_i|d) \qquad (15)$$

where $p_j(l_i|d)$ is the posterior returned by tree $T_j$.

The proposed random forest model offers two benefits: First, training and testing the model are extremely efficient as we only need to examine a subset of randomly selected feature dimensions. In addition, by combining all the landmarks and training the forest jointly, the model implicitly captures the image context information, thus being able to distinguish between neighboring landmarks. Fig. 4 illustrates the random forest result.

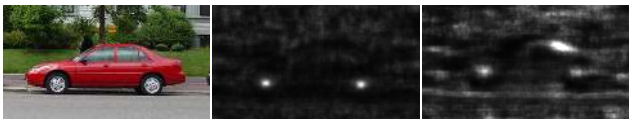


Figure 4. (a) A normalized image. We apply the trained random forest on the entire image. Posterior maps are shown for the wheels (b) and the top-right corner (c).

### 5.3. Performance Evaluation

We compare our approach with Active Shape Model (ASM) [5] and Bayesian Tangent Shape Model (BTSM) [24]. We initialize the car shape by a randomly perturbed mean shape, and the same initialization is applied to all three algorithms. Fig. 5 illustrates two example images. In the first example (top row), the appearance model is distracted by some bogus background features. ASM attempts to compensate the errors with large pose change, but at the expense of diverging the good features from their true locations. BTSM generates smoother results by assigning different weights on the observation and the shape prior. However, the errors are too large to be accommodated by its Gaussian noise model. Our approach successfully detects the outliers (colored in black) and excludes them from the parameter estimation. The second example shows a typical image with partial occlusion (bottom row). Again, the fitting is improved because the occluded features are automatically identified. Fig. 5(e) shows the random hypotheses generated by RANSAC. Although they are all car-like, our algorithm successfully identifies the optimal one which enjoys maximum agreement from the observation and the trained shape model.

Fig. 6 shows the root mean square error (RMSE) with respect to the labeled ground truth. We observe consistent improvement on the proposed model over the other approaches in all three views. A further investigation shows that our approach achieves comparable result as BTSM on "good" test images, but performs significantly better on the images with gross errors. Given that the error is averaged over 2,000 images, the pixel-level improvement is substantial for the alignment task.

To investigate the robustness of our algorithm to random initialization, we vary the noise level when perturbing the initial shape. Fig. 7 show the RMSE for different noise levels at each view. As expected, the performance of our


Figure 5. (a) The observed shape. (b) ASM. (c) BTSM. (d) Our approach (solid colored points represent the partial shape that generates the optimal hypothesis; white ones are the inliers included in the refinement step; and blacks ones are the outliers rejected by the model). (e) Random shape hypotheses generated by RANSAC. Top row shows an example with spurious background features; and bottom row shows an example with partial occlusion.

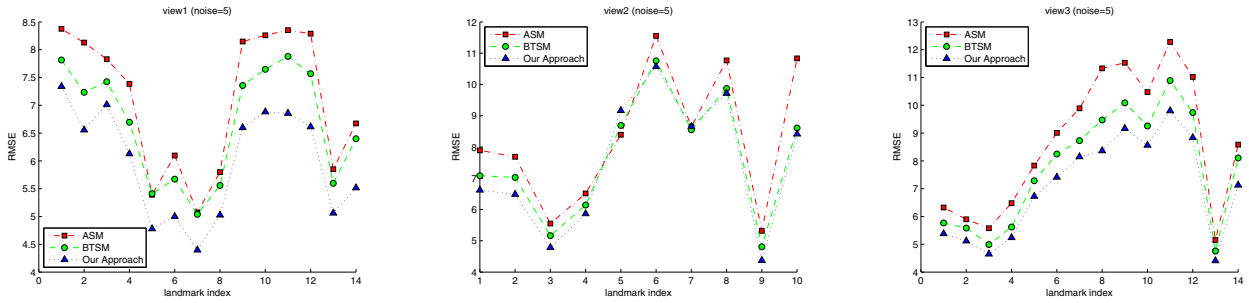


Figure 6. Test errors for ASM, BTSM and our approach. The initial shape is set to be the mean shape plus 5 pixels random noise on each landmark. For each test image, we use the same initialization for all three methods. The RMSE of each landmark is shown for different views: half-front view (left pane), profile view (middle pane), and half-back view (right pane).

alignment model drops as the noise level increases. However, the average error increases less than 1 pixel even when 20 pixels random shift is added to the initial shape. This is because our algorithm relies on a minimal subset of features to generate a hypothesis, therefore can recover the meaningful shape in a couple of iterations. Traditional approaches are more likely to fail in this case because shape observation is contaminated by more outliers.

Fig. 6 shows the landmark-wise average error over the entire test dataset. To investigate the error distribution, we need to make a side by side comparison for each example. We focus on the half-frontal view which contains 1,400 images. For each example, we run BTSM and Robust alignment respectively, using the same initialization. In Fig. 8, we use the sorted error of BTSM as reference and plot the corresponding error of the proposed method. A cubic curve is also fitted on the blue plot to provide a global illustration of the error distribution. As we can see, the two methods are comparable on the first 600 or so examples, while robust method overtakes BTSM in the remaining ones. Further inspections show that many of those difficult examples correspond to occlusion images.

Fig. 9 shows some alignment results by our approach. We demonstrate car images with various viewpoints, lightings, occlusion patterns, and cluttered background.

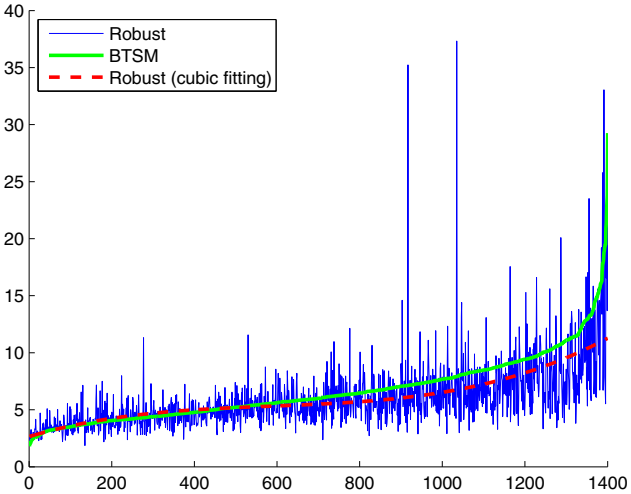


Figure 8. Side-by-side comparison of BTSM and our approach.

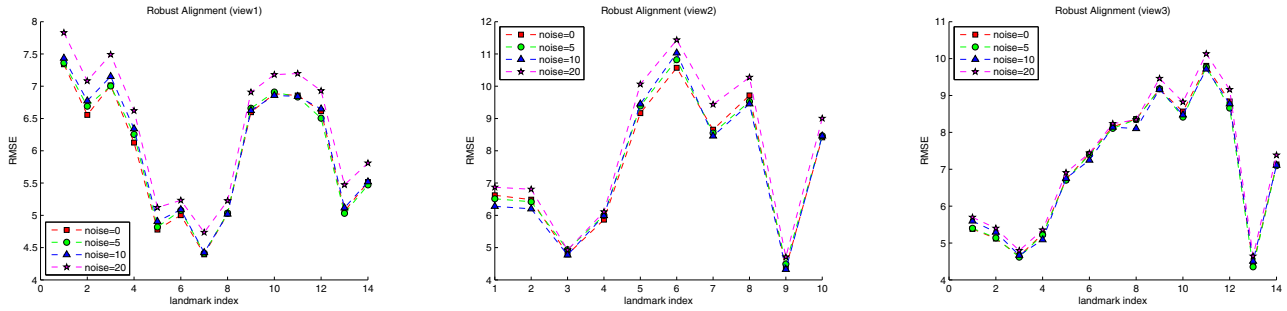


Figure 7. Test error for our approach using different initializations. The initial shape (mean shape) is perturbed by different levels of noise from 0 to 20 pixels.

## 6. Conclusions

We have described a RANSAC-based approach for robust object alignment, and applied it to a challenging multiple-view car alignment task. It is encouraging to see that the approach is capable of dealing with large measurement errors such as occlusions. The current algorithm takes locally detected feature point as input. However, there are great potentials for extending the RANSAC framework to operate over multiple, globally detected feature points. We plan to explore this approach in the future work.

## References


[1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH*, pages 187–194, 1999. 1

[2] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001. 1, 5

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *PAMI*, 23(6):681–685, 2001. 1

[4] T. F. Cootes and C. J. Taylor. A mixture model for representing shape variation. *Image and Vision Computing*, pages 110–119, 1997. 1

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape model – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan 1995. 5

[6] T. F. Cootes, K. Walker, and C. J. Taylor. View-based active appearance models. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000. 1

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005. 5

[8] I. Dryden and K. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998. 4

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. pages 381–395, 1981. 1, 3

[10] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63:3–42, 2006. 5

[11] J. Gower and G. Dijksterhuis. *Procrustes Problems*. Oxford University Press, 2004. 3

[12] L. Gu and T. Kanade. 3d alignment of face in a single image. In *Proceedings of Computer Vision and Pattern Recognition*, 2006. 1

[13] L. Gu and T. Kanade. A generative shape regularization model for robust face alignment. In *Proceedings of The 10th European Conference on Computer Vision*, 2008. 1

[14] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1988. 1

[15] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In *Proceedings of Computer Vision and Pattern Recognition*, 2005. 5

[16] L. Liang, R. Xiao, F. Wen, and J. Sun. Face alignment via component-based discriminative search. In *Proceedings of European Conference on Computer Vision*, 2008. 1

[17] C. Liu, H. Shum, and C. Zhang. Hierarchical shape modeling for automatic face localization. In *Proceedings of European Conference on Computer Vision*, 2002. 1

[18] M. Rogers and J. Graham. Robust active shape model search. In *Proceedings of European Conference on Computer Vision*, 2002. 1

[19] S. Romdhani, S. Gong, and A. Psarrou. A multi-view nonlinear active shape model using kernel PCA. In *BMVC*, 1999. 1

[20] P. J. Rousseeuw. *Robust regression and outlier detection*. Wiley, New York, 1987. 3

[21] C. Steward. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999. 4

[22] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 61:611–622, 1999. 2

[23] J. Tu, Z. Zhang, Z. Zeng, and T. Huang. Face localization via hierarchical CONDENSATION with Fisher boosting feature selection. In *Proceedings of Computer Vision and Pattern Recognition*, 2004. 1

[24] Y. Zhou, L. Gu, and H. J. Zhang. Bayesian tangent shape model: estimating shape and pose parameters via Bayesian inference. In *Proceedings of Computer Vision and Pattern Recognition*, 2003. 1, 2, 5

[25] Y. Zhou, W. Zhang, X. Tang, and H. Shum. A Bayesian mixture model for multi-view face alignment. In *Proceedings of Computer Vision and Pattern Recognition*, 2005. 1

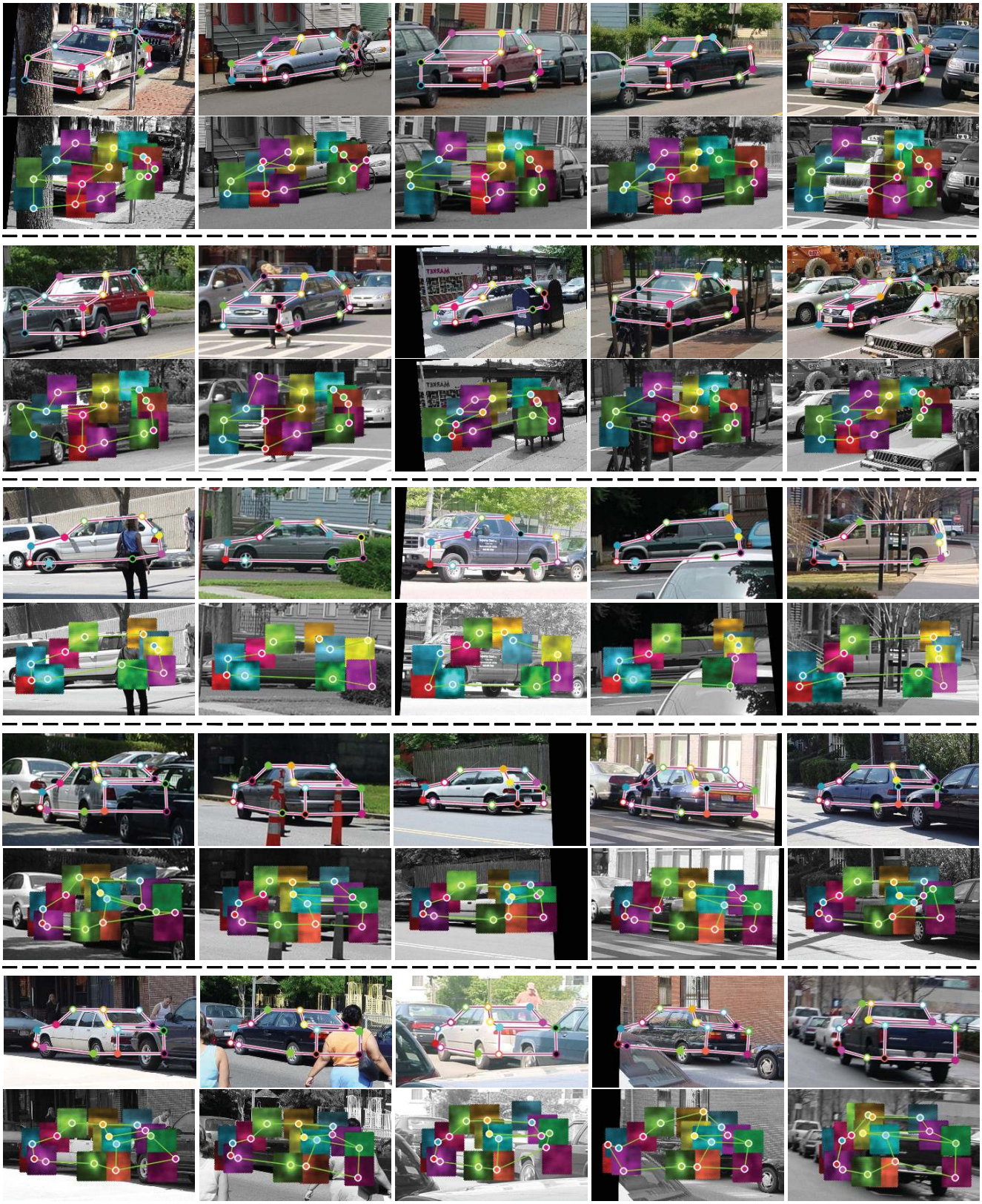

Figure 9. Alignment results by our approach. For each test image, we show the final result on the top and the observed shape at the bottom.