

**Original citation:**

Bhalerao, Abhir, Zhang, Qiang, Helm, Emma and Hutchinson, Charles (2015) Active shape model unleashed with multi-scale local appearance. In: IEEE International Conference on Image Processing (ICIP), Quebec City, Canada, 27-30 Sep 2015. Published in: Proceedings of IEEE International Conference on Image Processing (ICIP) pp. 4664-4668.

**Permanent WRAP url:**

<http://wrap.warwick.ac.uk/69861>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

“© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting /republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [publications@warwick.ac.uk](mailto:publications@warwick.ac.uk)



<http://wrap.warwick.ac.uk>

# ACTIVE SHAPE MODEL UNLEASHED WITH MULTI-SCALE LOCAL APPEARANCE

Qiang Zhang<sup>1</sup>, Abhir Bhalerao<sup>1</sup>, Emma Helm<sup>2</sup>, Charles Hutchinson<sup>2</sup>

<sup>1</sup>Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK

<sup>2</sup>University Hospitals Coventry and Warwickshire, Coventry, CV2 2DX, UK

## ABSTRACT

We focus on optimising the Active Shape Model (ASM) with several extensions. The modification is threefold. First, we tackle the over-constraint problem and obtain an optimal shape with minimum energy considering both the shape prior and the salience of local features, based on statistical theory: a compact closed form solution to the optimal shape is deduced. Second, we enhance the ASM searching method by modelling and removing the variations of local appearance presented in the training data. Third, we speed up the convergence of shape fitting by integrating information from multi-scale local features simultaneously. Experiments show significant improvement brought by these modifications, i.e., optimal shape against standard relaxation methods dealing with inadequate training samples; enhanced searching method against standard gradient descent methods in searching accuracy; multi-scale local features against popular coarse-to-fine strategies in convergence speed.

**Index Terms**— Active shape model, over constraint, inverse gradient descent, multi-scale features

## 1. INTRODUCTION

Active Shape Models (ASMs) [1] have been successfully used as a robust tool for object segmentation. An ASM constrains the shape with a statistical shape prior learned from training examples, thus is robust to noise. The constraint is applied by abandoning the minor variation components and restricting the significant components within certain range. The argument is that the the minor components in the training data are more likely to be noise, and only shape variations within certain standard derivations (s.d.) covered in the training data are deemed plausible. However this is no longer true in the case of high dimensional training data sets with few training examples, which causes over constraint. This problem becomes more critical in medical image analysis due to inadequate training samples and varied pathologies. There have been strategies to relieve the over constraint, including loosening up the shape constraints [2,3], introducing additional flexibility [4], synthesising additional training samples [5], and more recently, modifying shape prior using manifold learning [6] or sparse composition [7]. In these methods there is no ob-

vious a priori optimal parameters and it is time consuming to tune the parameters to specific applications. Another strategy is to optimise an objective function [8–10]. In [9,10] Matthias adopted the density estimation theory [11] to calculating the information loss caused by constraint and constructed a energy function searching for a shape with minimum cost.

Another problem is that ASMs are prone to local minimum. A coarse-to-fine strategy is commonly employed in all recent publications [12], but the criterion of switching to next resolution has to be chosen appropriately to avoid trapped into local minima. Indeed, as shown below it is a trade-off between the failure rate and iteration number per level. In practice, the criterion is often chosen conservatively, and a seamless coarse-to-fine method could speed up the fitting process significantly.

The first part of this paper is inspired by [9,11]. We tackle the over-constraint problem by penalising the two constraint terms and obtain an optimal shape in a statistical sense with respect to the shape priori as well as the salience of local features. We further derive a closed form solution of the optimal shape under the assumption of the Gaussian distribution of variations. In section 3, we adopt the inverse gradient descent (IGD) for local feature searching and enhance its robustness by learning and excluding the variation prior of local appearances. In section 4, we integrate multi-scale features into the energy function of unleashed ASMs, which leads to a seamless, fast and robust coarse-to-fine fitting method. We evaluate the influence of these modifications by applying them separately. Experiment shows significant improvement of performance especially with very limited training data.

## 2. ACTIVE SHAPE MODEL UNLEASHED

A 2D point-distribution shape in an ASM is represented by  $\mathbf{s} = (\mathbf{x}_1, \dots, \mathbf{x}_N) = (x_1, y_1, \dots, x_N, y_N) \in \mathbb{R}^{2N}$ . The statistical model is built by applying PCA to the aligned training shapes,

$$\mathbf{b}_s = P_s^T(\mathbf{s} - \bar{\mathbf{s}}), \quad (1)$$

with  $\mathbf{b}_s$  being the shape parameters and  $P_s$  the eigenvectors matrix. The dimensionality is reduced by retaining  $t$  most significant eigenvectors associated with the eigenvalues  $\lambda_1, \dots, \lambda_t$ .  $t$  can be chosen as  $t = \min\{t' | \sum_{i=1}^{t'} \lambda_i / \sum_{i=1}^{2N} \lambda_i > 98\%\}$ .

Standard ASMs constrain the shape within either a hyperrectangle [1] or a hyperellipsoid [13]. This can cause over constraint with inadequate training data, which is often the case in medical data as no training set can cover all pathologies. The over constraint has two causes: first, it does not count the cost of truncating minor components, which might capture shape details although is not statistically significant; second, shape variations outside 3s.d. might still be reasonable variation. Choosing a larger  $t$  or more s.d. may relieve the problem [2], but an optimal solution is to find a shape with minimum information loss due to projection to the subspace and the constraint. This can be achieved by constructing an energy function accounting for these terms, i.e. penalising the cost of projection as well as the Mahalanobis deviations from the mean shape and from the local feature predictions:

$$E(\mathbf{s}) = E_{shape}^{DIFS}(\mathbf{s}) + E_{shape}^{DFFS}(\mathbf{s}) + \frac{\beta}{1-\beta} E_{im}(\mathbf{s}). \quad (2)$$

$\beta \in (0, 1)$  controls the extent of unleashing, which can be adjusted accordingly. With the assumption that the shape variations are multivariate Gaussian distributed, the shape energy terms [11] are as follow:

$$E_{shape}^{DIFS}(\mathbf{s}) = \frac{1}{2} \mathbf{b}_s^T \Lambda^{-1} \mathbf{b}_s, \quad (3)$$

namely *the distance in feature space* (DIFS) is the Mahalanobis distance from the mean shape,  $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_t\}$  is the variance matrix, whereas

$$E_{shape}^{DFFS}(\mathbf{s}) = \frac{1}{2\rho} (\|\mathbf{s} - \bar{\mathbf{s}}\|^2 - \|\mathbf{b}_s\|^2), \quad (4)$$

is the *distance from feature space* (DFFS) penalising the costs of the projection to the feature space (i.e. the information loss due to cutting-off the minor eigen values), with  $\rho = \sum_{i=t+1}^{2N} \frac{\lambda_i}{2N-t}$ . The image energy can be written as,

$$E_{im}(\mathbf{s}) = \frac{1}{2} (\mathbf{s} - \hat{\mathbf{s}})^T \Sigma^{-1} (\mathbf{s} - \hat{\mathbf{s}}), \quad (5)$$

which is the Mahalanobis distance from the observed landmarks  $\hat{\mathbf{s}} = (\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_N)$  deduced from local features.  $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$  is the variances, with  $\sigma_i^2 = (\sigma_{x,i}^2, \sigma_{y,i}^2)$  representing the salience of the  $i$ -th landmark in  $\hat{\mathbf{s}}$  (the smaller the more salient). The optimal shape  $\mathbf{s}^*$  with maximum likelihood [11] is the one with minimum energy,

$$\mathbf{s}^* = \arg \min E(\mathbf{s}). \quad (6)$$

Instead of using numerical searching methods [9], we deduce an closed form solution, which reduces the computation and guarantees the global minimum. Specifically, Eq. (6) is solved by  $\left. \frac{dE(\mathbf{s})}{d\mathbf{s}} \right|_{\mathbf{s}^*} = 0$ . From (1) we have  $d\mathbf{b}_s/d\mathbf{s} = P_s^T$ . We first calculate the differentiation of each term of  $E$ .

$$\begin{aligned} \frac{dE_{shape}^{DIFS}(\mathbf{s})}{d\mathbf{s}} &= P_s \Lambda^{-1} \mathbf{b}_s = P_s \Lambda^{-1} P_s^T (\mathbf{s} - \bar{\mathbf{s}}), \\ \frac{dE_{shape}^{DFFS}(\mathbf{s})}{d\mathbf{s}} &= \frac{1}{\rho} ((\mathbf{s} - \bar{\mathbf{s}}) - P_s \mathbf{b}_s) = \frac{1}{\rho} (I - P_s P_s^T) (\mathbf{s} - \bar{\mathbf{s}}), \\ \frac{dE_{im}(\mathbf{s})}{d\mathbf{s}} &= \Sigma^{-1} (\mathbf{s} - \hat{\mathbf{s}}). \end{aligned} \quad (7)$$

Substituting these into (6), the solution obtained is,

$$\mathbf{s}^* = B^{-1} C + \bar{\mathbf{s}}, \quad (8)$$

where  $B = (1 - \beta) \left( P_s \Lambda^{-1} P_s^T + \frac{1}{\rho} (I - P_s P_s^T) \right) + \beta \Sigma^{-1}$ , and  $C = \beta \Sigma^{-1} (\hat{\mathbf{s}} - \bar{\mathbf{s}})$ .

### 3. ACTIVE SHAPE MODEL SEARCHING

The shape model is driven by local features, appearance of which varies across the training samples. The performance can be improved by learning the appearance variations in the training set and excluding the differences by ‘projecting out’ these variations before matching features to their templates. We adopt inverse gradient descent (IGD) [14] previously used in active appearance model fitting [15] for this purpose.

Given a set of local 2D features  $A_{i,1}, A_{i,2}, \dots, A_{i,N} \in \mathbb{R}^2$  at the  $i$ -th landmark in  $N$  training samples, PCA is applied to obtain the mean  $\bar{A}_i$  and the eigenvectors matrix  $P_{A,i}$  spanning a linear subspace covering, say, 98% of the variation. For the concise of notation, we omit the landmark subscript  $i$  in this section. A new feature instance  $A$  thus can be approximated as,

$$A \approx \bar{A} + P_A \mathbf{b}_A. \quad (9)$$

$\mathbf{b}_A$  is the appearance parameters.

Given a new image and an appropriate initial location  $\mathbf{x} = (x, y)$  for a landmark, the searching task is to adjust  $\mathbf{x}$  and  $\mathbf{b}_A$  to minimise the difference between the appearance model and the image intensities it covers:

$$(\Delta \mathbf{x}, \mathbf{b}_A) = \arg \min (A(\mathbf{x} + \Delta \mathbf{x}) - (\bar{A} + P_A \mathbf{b}_A))^2. \quad (10)$$

The expression of the target function can be rewritten as,

$$\|A(\mathbf{x} + \Delta \mathbf{x}) - (\bar{A} + P_A \mathbf{b}_A)\|_{P_A}^2 + \|A(\mathbf{x} + \Delta \mathbf{x}) - (\bar{A} + P_A \mathbf{b}_A)\|_{\perp}^2. \quad (11)$$

where  $\|\cdot\|_{P_A}^2$  and  $\|\cdot\|_{\perp}^2$  denote the square of the Euclidean norm of the vector projected into the subspace spanned by  $P_A$  and its orthogonal space respectively. Note the minimum value of the first term is always zero. And in the second term, the projection of  $P_A$  in its orthogonal subspace is zero. As we are only interested in  $\Delta \mathbf{x}$ , it simplifies to:

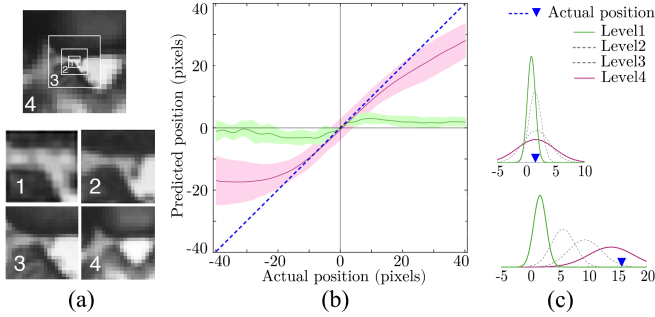
$$\Delta \mathbf{x} = \arg \min \|A(\mathbf{x} + \Delta \mathbf{x}) - \bar{A}\|_{\perp}^2, \quad (12)$$

which is a typical gradient descent problem. Now the prediction of  $\mathbf{x}$  is no longer affected by the appearance variations, which have been ‘projected out’. Applying the gradient descent directly is expensive as the numerical gradient of  $A$  needs to be calculated iteratively. This can be tackled with IGD, i.e., reverse the roles of  $A$  and  $A_0$ ,

$$\Delta \mathbf{x} = \arg \min \|\bar{A}(-\Delta \mathbf{x}) - A(\mathbf{x})\|_{\perp}^2. \quad (13)$$

The closed form solution is,

$$\Delta \mathbf{x} = \left( \frac{\partial \bar{A}_{\perp}}{\partial \mathbf{x}} \right)^+ (A(\mathbf{x}) - \bar{A}). \quad (14)$$



**Fig. 1.** (a) Four scale features at a landmark. (b) One standard error bar of predictions from finest level (green) and coarsest level (red) features. Blue dash line indicates the actual position. (c) Gaussian mixture model of four scale predictions with initial position at 2 (top) and 16 (bottom) pixels from the actual position.

where  $(\cdot)^+$  denotes the Moore-Penrose pseudo-inverse.  $\bar{A}_\perp$  is the projection of  $\bar{A}$  onto the orthogonal space of  $P_A$  and can be pre-calculated, see [14] for details.

**Variance:** The variance of prediction of  $\Delta\mathbf{x}$  should take two parts into account,

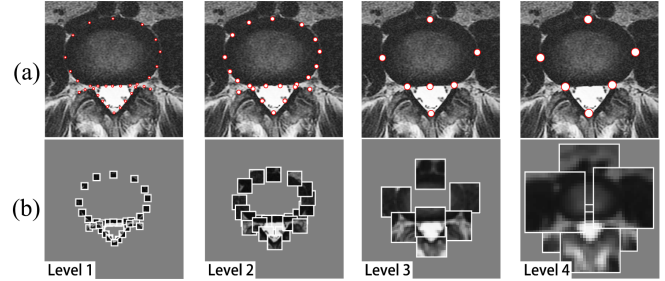
$$\sigma^2 = \sigma_{feature}^2 + \sigma_{diff}^2. \quad (15)$$

$\sigma_{feature}^2$  depends on the inherent feature saliency, which can be obtained by testing the feature searching method on the training data and calculating the mean variance.  $\sigma_{diff}^2$  can be estimated by calculating the difference between the observed feature and the template in the orthogonal space of  $P_A$ .

#### 4. RAPID AND ROBUST FITTING STRATEGY USING MULTI-SCALE LOCAL APPEARANCE

In this section, we introduce a seamless approach to speed up the ASM convergence and increase the robustness. Specifically, we generate an  $L$ -level Gaussian pyramid of the images. At each landmark, feature patches with same size in pixel are extracted from the pyramid, generating a set of profiles at multiple scales, see Fig. 1(a) as an example. Then at the  $i$ -th landmark, features at each level  $l \in \{1, \dots, L\}$  gives an prediction of the correction  $\Delta\mathbf{x}_{i,l}$  with variance  $\sigma_{i,l}^2$ , using the method described in section 3.

To evaluate the prediction of the multi-level features, an experiment on the feature example in Fig 1(a) is shown in Fig 1(b). The experiment is conducted by extracting the feature patches at each level with known displacements from the actual position and evaluating the precision of predictions. We can see that the prediction from smaller scale features has smaller variance and better linear approximation around zero, while larger scale features have larger linear range. This explains the feasibility of coarse-to-fine methods as larger scale features drag the shape rapidly to the object during initial iterations, while the features at smaller scales begin to converge



**Fig. 2.** (a) Multi-scale local features. Fewer patches are preserved at higher level. (b) The corresponding shape nodes.

and give a more precise prediction when near enough to the true position. However, it is difficult to decide whether it is close enough to change to the next level of detail during fitting as it is a trade-off between the efficiency and the risk of being trapped in a local minimum.

We propose therefore to incorporate multi-level predictions, in a form of a Gaussian mixture, together with the prior knowledge of shape to choose the best candidate prediction. The rationale is that the shape energy  $E_{shape}$  is more likely to be lower when for each landmark the best prediction is chosen. If we assume the errors of the prediction is Gaussian distributed, then the prediction of a feature location from all levels are actually a Gaussian mixture with expectation  $\{\Delta\mathbf{x}_{i,1}, \Delta\mathbf{x}_{i,2}, \dots, \Delta\mathbf{x}_{i,L}\}$  and variances of  $\{\sigma_{i,1}^2, \sigma_{i,2}^2, \dots, \sigma_{i,L}^2\}$ , as shown in Fig 1(c). Modifying the image energy term (5) taking all level predictions into account gives,

$$E_{im}(s) = \frac{1}{2} \sum_{l=1}^L (s - \hat{s}_l)^T \Sigma_l^{-1} (s - \hat{s}_l), \quad (16)$$

where  $\Sigma_l = \text{diag}(\sigma_{1,l}^2, \sigma_{2,l}^2, \dots, \sigma_{N,l}^2)$ ,  $\hat{s}_l = (\hat{\mathbf{x}}_{1,l}, \hat{\mathbf{x}}_{2,l}, \dots, \hat{\mathbf{x}}_{N,l})$ . The optimal shape  $s^*$  still has the same form as (8), with the two coefficients replaced by  $B = (1 - \beta)(P_s \Lambda^{-1} P_s^T + (\frac{1}{\rho^*} - P_s P_s^T)) + \beta \sum_{l=1}^L \Sigma_l^{-1}$  and  $C = \beta \sum_{l=1}^L \Sigma_l^{-1} (\hat{s}_l - \bar{s})$ .

In practice, larger scale feature patches at adjacent landmarks are more overlapping thus return similar prediction. Also lower precision at coarser levels is acceptable. So we can keep fewer feature patches at interval shape nodes at these levels, as shown in Fig. 2, and propagate the prediction to the landmarks nearby, which results in a more compact feature pyramid.

The implementation of the whole algorithm is as follow:

#### Training:

1. Train the shape prior using PCA and obtain  $\bar{s}$ ,  $P_s$ ,  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_t)$ ;
2. Build Gaussian pyramid, train the local appearance and obtain  $\bar{A}$ ,  $(\frac{\partial \bar{A}_\perp}{\partial \mathbf{x}})^+$  for each node (see Fig. 2) at each level;

#### Testing:

1. Generate the Gaussian pyramid of the testing image.
2. Initialise the shape  $s^{(0)}$ ;
3. For each node, calculate  $\Delta x_l, \sigma_l^2$  at each scale using (14);
4. Update the current shape  $\hat{s}_l^{(k+1)} = s^{(k)} + \Delta x_l$ ;
5. Calculate the optimal shape  $s^{(k+1)}$  using (8);
6. Repeat 3. 4. 5. until converged;

## 5. EXPERIMENTS AND RESULTS

As each of our proposed modifications can be applied independently, we focus on evaluating the improvement brought by each one, comparing the performance with and without. The experiments are performed on 56 axial lumbar vertebrae and 30 sagittal knee slices, sizes of both are  $512 \times 512$  pixels.

**1. Feature searching method with appearance modelling:** For each landmark, we randomly initialise the position within 5 pixels from the true feature and evaluate the searching method in section 3 comparing it with a standard gradient descent method. After the local appearance variations modelled and removed, the 1 s.d. error of local feature searching goes down from  $4.1 \pm 1.6$  to  $3.1 \pm 1.1$  pixels on vertebrae and from  $3.6 \pm 1.5$  to  $2.8 \pm 1.4$  pixels on knee data.

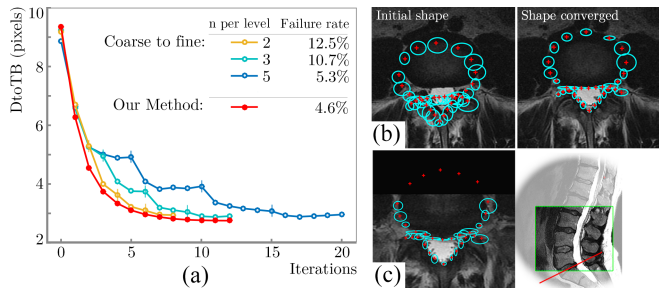
**2. Shape optimisation vs. shape relaxation dealing with over constraint:** We compare our energy minimising method with hyperrectangle [1] and hyperellipsoid [13] constraints in dealing with inadequate training data situation. All three methods are trained with different  $m$  data randomly sampled from the dataset and tested on the rest. The experimental optimal loosening parameters are chosen for each  $m$  (around 3 s.d. and 1.7 s.d. for hyperrectangle and hyperellipsoid constraints respectively), while a constant  $\beta = 0.3$  is chosen for our method. The Distance to ‘True Boundary’ (DtoTB) and Dice Similarity Coefficient (DSC) [16] to the ground truth are used as the metric of segmentation precision. Table 1 shows the experiment results. We can see that our method degrades less with reducing numbers of training samples, and performs better especially with very limited training.

**Table 1.** Performance of three methods on segmentation of vertebrae and knees, trained with reducing  $m$  samples.

	DtoTB (pixels)			DSC (%)		
	<i>Rect.</i> [1]	<i>Ellip.</i> [13]	<i>Our</i>	<i>Rect.</i> [1]	<i>Ellip.</i> [13]	<i>Our</i>
Vertebra						
$m = 40$	2.6 $\pm$ 1.2	2.6 $\pm$ 1.0	<b>2.5</b> $\pm$ 1.1	95.3 $\pm$ 2.1	<b>95.7</b> $\pm$ 1.8	95.6 $\pm$ 2.3
$m = 20$	3.1 $\pm$ 1.3	2.9 $\pm$ 1.1	<b>2.7</b> $\pm$ 1.0	94.3 $\pm$ 3.1	94.1 $\pm$ 2.2	<b>95.0</b> $\pm$ 3.5
$m = 10$	4.1 $\pm$ 1.9	4.2 $\pm$ 2.5	<b>3.1</b> $\pm$ 1.1	92.0 $\pm$ 3.0	91.8 $\pm$ 3.6	<b>93.9</b> $\pm$ 2.1
$m = 5$	7.7 $\pm$ 2.6	7.7 $\pm$ 2.9	<b>5.0</b> $\pm$ 1.7	84.1 $\pm$ 6.4	84.2 $\pm$ 6.7	<b>88.7</b> $\pm$ 4.0
Knee						
$m = 20$	2.7 $\pm$ 1.2	2.7 $\pm$ 1.3	<b>2.2</b> $\pm$ 0.7	97.4 $\pm$ 1.3	97.6 $\pm$ 1.2	<b>97.9</b> $\pm$ 0.7
$m = 10$	4.3 $\pm$ 2.7	3.5 $\pm$ 2.5	<b>2.7</b> $\pm$ 1.0	96.3 $\pm$ 3.2	96.7 $\pm$ 2.6	<b>97.6</b> $\pm$ 1.0
$m = 7$	5.5 $\pm$ 3.0	3.7 $\pm$ 2.5	<b>3.0</b> $\pm$ 1.3	94.3 $\pm$ 4.9	96.3 $\pm$ 2.5	<b>97.3</b> $\pm$ 1.1
$m = 5$	7.2 $\pm$ 3.2	4.9 $\pm$ 2.6	<b>4.0</b> $\pm$ 2.5	92.5 $\pm$ 4.3	94.9 $\pm$ 3.2	<b>96.6</b> $\pm$ 1.8

### 3. Multi-scale vs. coarse-to-fine in convergence speed:

We randomly initialise the shapes within 30 pixels from true



**Fig. 3.** (a) Convergence process and failure rate of multi-scale strategy compared with coarse-to-fine method with certain iterations per level. (b) Cyan ellipses show the predictions from local features with the sizes indicating the variances. Red crosses show the optimal shape. (c) A special case when part of the vertebra in the axial slice is missing.

position and compare our multi-scale method with a coarse-to-fine method using  $n = 2, 3, 5$  iterations per level. Fig. 3(a) shows the mean result of an intensive test trained on 20 samples. The corresponding failure rate (being trapped in local minima) is also given. It shows that with decreasing  $n$ , the failure rate increases. While our method converges fastest with lowest failure rate. The shape in our method is driven by multi-scale features simultaneously. Saliency of each scale is controlled by the variance of predictions. Fig. 3(b) gives an example of variances at the finest level before (left) and after (right) the convergence. The variations decrease significantly when the landmarks find the salient features, meanwhile the features whose boundary is blurred have higher variance, thus depend more on the shape prior. In a special case in Fig. 3(c) when part of the object is missing, these variances become infinite and the corresponding landmarks are completely based on other salient ones and the shape prior, which is similar to the partial ASM [17] in dealing with the missing boundary but is feature-driven thus more compact.

## 6. CONCLUSIONS

Several modifications of ASM have been presented. Experiments validated the improvement in three aspects.

(1) Higher precision in local feature searching. This is because variations of feature appearance presented in the training data are learned and modelled, and during feature searching these variations have been ‘projected out’ before matching to the feature template.

(2) Better performance with limited training samples. The shape is unleashed and optimal in a statistical sense taking both the shape prior and feature confidence into consideration thus the detail of salient features is better preserved.

(3) Seamless and rapid convergence process. Compared to coarse-to-fine methods, the multi-scale strategy converges faster with lower failure rate. This is because larger scale features keep the shape from local minima, while smaller scales take effect as soon as it gets into the convergence range.

## 7. REFERENCES

- [1] Cootes, Timothy F and Taylor, Christopher J and Cooper, David H and Graham, Jim, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [2] Milborrow, Stephen and Nicolls, Fred, "Locating facial features with an extended active shape model," in *Proceedings of the European Conference on Computer Vision*, pp. 504–513. Springer, 2008.
- [3] Shang, Y and Dossel, O, "Statistical 3D shape-model guided segmentation of cardiac images," in *Computers in Cardiology*. IEEE, 2004, pp. 553–556.
- [4] Heimann, Tobias and Münzing, Sascha and Meinzer, Hans-Peter and Wolf, Ivo, "A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation," in *Information Processing in Medical Imaging*. Springer, 2007, pp. 1–12.
- [5] Koikkalainen, Juha and Tolli, Tuomas and Lauerma, Kirsi and Antila, Kari and Mattila, Elina and Lilja, Mikko and Lotjonen, Jyrki, "Methods of artificial enlargement of the training set for statistical shape models," *IEEE Transactions on Medical Imaging*, vol. 27, no. 11, pp. 1643–1654, 2008.
- [6] Zhang, Wuxia and Yan, Pingkun and Li, Xuelong, "Estimating patient-specific shape prior for medical image segmentation," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2011, pp. 1451–1454.
- [7] Zhang, Shaoting and Zhan, Yiqiang and Dewan, Maneesh and Huang, Junzhou and Metaxas, Dimitris N and Zhou, Xiang Sean, "Towards robust and effective shape modeling: Sparse shape composition," *Medical Image Analysis*, vol. 16, no. 1, pp. 265–277, 2012.
- [8] Cristinacce, David and Cootes, Tim, "Automatic feature localisation with constrained local models," *Pattern Recognition*, vol. 41, no. 10, pp. 3054–3067, 2008.
- [9] Kirschner, Matthias and Wesarg, Stefan, "Active shape models unleashed," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2011, pp. 796211–796211.
- [10] Kirschner, Matthias and Becker, Meike and Wesarg, Stefan, "3D active shape model segmentation with non-linear shape priors," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 492–499. Springer, 2011.
- [11] Moghaddam, Baback and Pentland, Alex, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.
- [12] Heimann, Tobias and Meinzer, Hans-Peter, "Statistical shape models for 3D medical image segmentation: A review," *Medical Image Analysis*, vol. 13, no. 4, pp. 543–563, 2009.
- [13] Cerrolaza, Juan J and Villanueva, Arantxa and Cabeza, Rafael, "Shape Constraint Strategies: Novel Approaches and Comparative Robustness.," in *Proceedings of the European Conference on Computer Vision*, 2011, pp. 1–11.
- [14] Baker, Simon and Matthews, Iain, "Equivalence and efficiency of image alignment algorithms," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2001, vol. 1, pp. I–1090.
- [15] Matthews, Iain and Baker, Simon, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.
- [16] Popovic, Aleksandra and de la Fuente, Matías and Engelhardt, Martin and Radermacher, Klaus, "Statistical validation metric for accuracy assessment in medical image segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 2, no. 3-4, pp. 169–181, 2007.
- [17] Yan, Pingkun and Xu, Sheng and Turkbey, Baris and Kruecker, Jochen, "Discrete deformable model guided by partial active shape model for TRUS image segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, pp. 1158–1166, 2010.