Combining Local and Global Shape Models for Deformable Object Matching

Philip A Tresadern
philip.tresadern@manchester.ac.uk
Harish Bhaskar
harish.bhaskar@manchester.ac.uk
Steve A Adeshina
steve.adeshina@postgrad.manchester.ac.uk
Chris J Taylor
chris.taylor@manchester.ac.uk
Tim F Cootes
tim.cootes@manchester.ac.uk

Imaging Sciences and Biomedical Engineering,
School of Cancer and Imaging Sciences,
University of Manchester,
Manchester M13 9PT,
United Kingdom

Abstract

We describe a method for modelling and locating deformable objects using a combination of global and local shape models. An object is represented as a set of patches together with a geometric model of their relative positions. The geometry is modelled with a global pose and linear shape model, together with a Markov Random Field (MRF) model of local displacements from the global model. Matching to a new image involves an alternating scheme in which an MRF inference technique selects the best candidates for each point, which are then used to update the parameters of the global pose and shape model. A cascade of increasingly complex models is used to achieve robust matching to new images. We explore the effect of model parameters on system performance and show that the proposed method achieves better accuracy than other widely used methods on standard datasets.

1 Introduction

Object tracking and segmentation are central problems in Machine Vision, with applications including biometric authentication and medical imaging. However, these tasks are complicated by non-rigid variation within the class of objects we wish to track (*e.g.* faces, hands, hearts [25]). A popular approach to addressing this problem is to build a parametric *deformable* model of the object that encapsulates expected variations in shape and appearance. Fitting the model to an unseen query image then gives the motion parameters and/or spatial extent of the object in the image.

However, there exist competing approaches to modeling shape and appearance. Appearance can be modelled either at a discrete set of landmark features on the object, as in point-based methods such as the Active Shape Model (ASM) [9] or Pictoral Structure Matching (PSM) [11], or as a dense sampling of the image region corresponding to the object, as

BMVC 2009 doi:10.5244/C.23.95

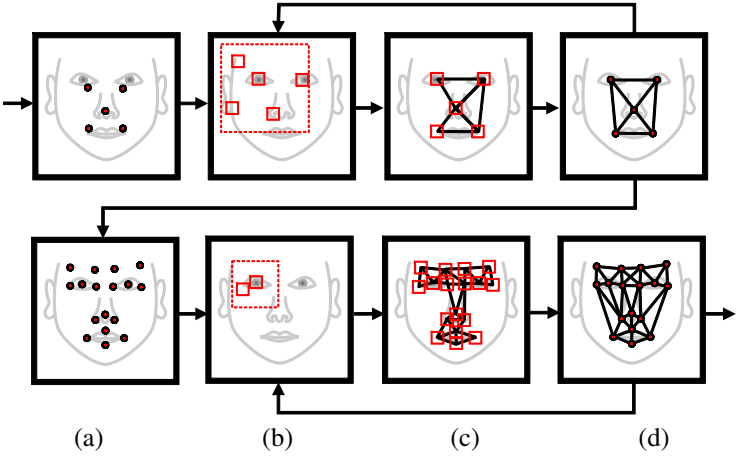


Figure 1: Schematic of a two-level cascade search: (a) initialize; (b) find candidate points; (c) select best set of candidates using MRF; (d) regularize using global model and either iterate, go to next level or finish.

in texture-based methods such as the Active Appearance Model (AAM) [3, 19]. Similarly, shape can be represented by a fully-connected (*global*) graphical model where every point has a statistical relationship with every other point, or a sparsely-connected (*local*) graphical model where each point depends only on a small subset of the other points (typically its immediate neighbours).

In this paper we investigate a method of combining a global shape model with a local model of displacements. The local models efficiently select good candidate points, giving robustness to false matches, while the global model regularizes the result to ensure a plausible final shape. We denote this method a *cascade of Combined Shape Models* (c-CSM) and demonstrate it on images of faces and hands.

1.1 Related Work

Under the Active Shape Model (ASM) [4] framework, each image in a training set is manually annotated with N 'landmark' points. A statistical model of each point's photometric properties (*i.e.* appearance) is then built from the training images. Similarly, a statistical model of the geometric relationships between points is learned from the data via Principal Component Analysis (PCA). When fitting the model to an unseen image with a coarse initialization, the algorithm localizes each landmark independently by searching for the 'best' (with respect to the appearance model) candidate location within a search region centred on its current position. This set of candidate regions is then regularized by projecting them onto the space of valid shapes as defined by the shape model. These search-regularize steps are iterated to convergence under a coarse-to-fine implementation to make the method robust.

The use of a learned statistical model ensures that shape is specific to the class of objects to be located yet is flexible enough to permit natural variation within that class. Although using the 'tangent space' has proven to improve performance [28], a Gaussian model cannot capture non-linear shape variations (*e.g.* as a result of 3D pose variation in face tracking). However, this limitation can be addressed using more complex models such as a mixture of

Gaussians [2], kernel PCA [23] or a Gaussian process latent variable model [13].

When fitting the model to data, it has been shown that applying non-linear optimization directly over the shape model parameters to minimize an error function, based on the learned appearance model, can be more robust than the ASM [8]. However, all local optimization schemes (including the ASM) risk falling into local minima. In recent years, sampling strategies have demonstrated some success in addressing this problem [17, 18, 22, 26] and are also well-suited to a hierarchical model fitting strategy [17].

One further problem with the ASM is that candidate landmark locations are selected independently *before* any prior knowledge over shape is exploited, potentially making the algorithm unstable. A better approach would take shape priors into account when selecting matches in order to favour sets of candidates that have the desired geometry, resulting in 'lighter' regularization and improved stability. This is very expensive for a fully connected model such as that generated by PCA (though Lekadir and Yang [15] show how tree search algorithms can be used to find approximally optimal solutions). Therefore, much interest has been directed toward shape matching using a sparsely connected (*local*) graphical model, also known as a Markov Random Field (MRF), that exploits reasonable assumptions of conditional independence between landmarks [6]. These can be solved by methods such as dynamic programming [10] or belief propagation [5], and under certain conditions can reduce complexity further such that it is computationally feasible to find the *global* optimum, as demonstrated by the PSM algorithm [11]. Matching models of parts+geometry using sparse MRF models of geometry is widespread in the object recognition community [12, 21], and is being applied to locating complex structures in medical images [9].

Such methods have recently been successfully applied to guide candidate selection within the ASM framework [1, 13, 16]. This permits solutions to deviate from the global shape model under suitable constraints, allowing a closer fit to the image data. A variant on this approach allows the global shape model to account for more of the variation, using the MRF to model only the deviations from the current estimate of shape [24]. In effect, this allows the pairwise relationships to vary with the global shape rather than remaining as fixed.

1.2 Contributions and Structure

The method proposed in this paper combines an MRF-based local shape model for guided candidate selection with a PCA-based global shape model for regularization. One novelty of the method is that it uses a two-stage *cascade* implementation for robustness: in the first step, we determine the global translation, orientation and scale using only the MRF to localize a salient subset of the landmark points; in the second step, we refine the previously localized points and localize a larger set of less salient landmarks in the model by allowing the global model to account for more of the shape variation. Our results suggest that using a cascade of increasing complexity can improve performance when compared to a single model.

Furthermore, when training the model we use each trained level of the cascade to update the predicted feature locations before training the following level. This ensures that we model the error distribution accurately rather than (incorrectly) assuming that error distributions are equal at all levels. Similarly, system parameters (such as the radius of the search region) are learned from training data rather than specified by hand.

We begin by describing our method, including the models of shape and appearance, in Section 2. We then conduct experiments on varied datasets (Section 3) that investigate the effect of key system parameters and demonstrate the efficacy of MRF-guided candidate selection when compared to other recently developed techniques. Section 4 concludes.

2 Method

In our proposed technique, we formulate the deformable object matching as a global shape alignment problem combined with MRF-based local modeling. We represent the model as a set of N points, $\mathbf{X} = \{\mathbf{x}_i = (u_i, v_i)\}$. Given a query image, $\mathbf{I}$, our aim is then to find the optimal set, $\mathbf{X}^*$, that maximizes the posterior,

$$p(\mathbf{X}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{X})p(\mathbf{X}). \tag{1}$$

Since the number of possible positions for each $\mathbf{x}_i$ is very large, however, considering the combinatorial number of all possible $\mathbf{X}$ is computationally intractable, even if we restrict the set of candidates for each $\mathbf{x}_i$ to some smaller number (*e.g.* by considering only locally optimal candidates within a region of interest). By making certain assumptions of conditional independence between features, however, we can approximate the joint prior with an MRF that reduces the complexity of the problem such that an approximate solution, $\mathbf{Y}$, can be found efficiently. We can then *regularize* this approximation to give a solution that is closer to the optimum, $\mathbf{X}^*$. We summarize this approach as follows:

1. Initialize point locations, $\mathbf{X}_0^*$
2. For $t = 1 \ldots T$
   (a) Select most promising candidates, $\mathbf{Y}_t = \arg\max_{\mathbf{Y}} p(\mathbf{I}|\mathbf{Y})p(\mathbf{Y}|\mathbf{X}_{t-1}^*)$
   (b) Regularize candidates in order to update points, $\mathbf{X}_t^* = \arg\max_{\mathbf{X}} p(\mathbf{X}|\mathbf{Y}_t)$

In the following sections, we describe these two steps in detail before outlining how they are integrated within a hierarchical 'cascade' framework.

2.1 MRF-guided Candidate Selection

The first step involves finding the set of candidates, $\mathbf{Y}_t$, that maximizes the posterior,

$$p(\mathbf{Y}|\mathbf{I}) = p(\mathbf{I}|\mathbf{Y})p(\mathbf{Y}|\mathbf{X}_{t-1}^*). \tag{2}$$

The first term in (2) is the likelihood that indicates how well the image data supports the hypothesized candidates. It is common to assume that the patch, Q_i, associated with each candidate, $\mathbf{y}_i$, does not overlap with any other patch such that

$$p(\mathbf{I}|\mathbf{Y}) = \prod p(Q_i|\mathbf{y}_i) \tag{3}$$

$$\Rightarrow -\log p(\mathbf{I}|\mathbf{Y}) = -\sum \log p(Q_i|\mathbf{y}_i) = \sum \phi(\mathbf{y}_i) \tag{4}$$

where $\phi(\cdot)$ is an error function that indicates goodness of fit with the image data. In our case, we use (negated) normalized correlation over the x- and y-gradient images for accurate feature localization. To make inference practical, we restrict the set of candidates considered for each $\mathbf{y}_i$ to the k_i lowest-scoring local minima of $\phi(\mathbf{y}_i)$ within a search region of size r_i.

The second term in (2) is the MRF prior, where the conditional dependence on $\mathbf{X}_{t-1}^*$ is introduced as a result of taking all measurements in a normalized (with respect to scale and orientation) coordinate frame defined by the current estimate of $\mathbf{X}^*$. By making certain assumptions of conditional independence between points, we can approximate the joint prior

with a more sparsely connected graph. In the case where we consider only dependencies between pairs of points,

$$p(\mathbf{Y}|\mathbf{X}^*_{t-1}) \approx \prod p(\mathbf{y}_i|\mathbf{y}_j,\mathbf{X}^*_{t-1}) \tag{5}$$

$$\Rightarrow -\log p(\mathbf{Y}|\mathbf{X}^*_{t-1}) \approx -\sum \log p(\mathbf{y}_i|\mathbf{y}_j,\mathbf{X}^*_{t-1}) = \sum \psi(\mathbf{y}_i,\mathbf{y}_j) \tag{6}$$

where $\psi(\cdot)$ is an error function that indicates goodness of fit between a pair of candidates corresponding to connected nodes in the graph (we have omitted the dependency on $\mathbf{X}^*_{t-1}$ for clarity). Typically, it is assumed that

$$p(\mathbf{y}_i|\mathbf{y}_j) \sim N(\mathbf{y}_i - \mathbf{y}_j; \mu_{ij}, \Sigma_{ij}) \tag{7}$$

where μ and Σ are the mean and covariance of the displacement distribution, learned from training data. As a result, $\psi(\cdot)$ is the Mahalanobis distance from the mean displacement. When combining global and local models, however, we instead model relationships between residuals:

$$p(\mathbf{y}_i|\mathbf{y}_j,\mathbf{X}^*) \sim N((\mathbf{y}_i - \mathbf{x}^*_i) - (\mathbf{y}_j - \mathbf{x}^*_j); \mu_{ij}, \Sigma_{ij}) \tag{8}$$

$$\sim N((\mathbf{y}_i - \mathbf{y}_j) - (\mathbf{x}^*_i - \mathbf{x}^*_j); \mu_{ij}, \Sigma_{ij}) \tag{9}$$

In the case of a rigid shape model, this is equivalent to modelling the raw displacements. When the shape is allowed to vary with respect to its coordinate frame, however, the distance $\mathbf{x}^*_i - \mathbf{x}^*_j$ varies and modifies the pairwise potential. In practice, we maximize (2) by minimizing an energy function,

$$E = \sum \phi(\mathbf{y}_i) + \lambda \sum \psi(\mathbf{y}_i,\mathbf{y}_j), \tag{10}$$

where λ is a parameter that weights the influence of the prior and likelihood terms. This energy function is minimized using inference algorithms such as dynamic programming (as used in this work), belief propagation [5] or tree re-weighted message passing [14].

2.2 PCA-based Regularization

Having selected a set of candidate feature locations, we regularize them by projecting onto a learned subspace of allowable solutions. Specifically, we align the set of $2N$-dimensional training vectors, $\mathbf{X} = (u_1,\ldots,u_N,v_1,\ldots,v_N)^{\mathrm{T}}$, using Procrustes analysis and then learn their underlying linear subspace via PCA [4]. As a result, any $\mathbf{X}$ can be described by a (similarity, in this case) transformation, S, and a vector of shape coefficients, $\mathbf{b}$:

$$\mathbf{X} = S(\overline{\mathbf{X}} + \mathbf{P}\mathbf{b} + \varepsilon) \tag{11}$$

where $\overline{\mathbf{X}}$ is the mean shape over the training data (in a normalized reference co-ordinate frame), $\mathbf{P}$ is a set of orthogonal modes of variation, and ε vector accounts for residual displacements associated with every feature point in the global shape. We then compute an optimal value of $\mathbf{X}$ by projecting the selected candidates onto this subspace:

$$\mathbf{X}^*_t = S_t(\overline{\mathbf{X}} + \mathbf{P}\mathbf{P}^{\mathrm{T}}(S^{-1}_{t-1}\mathbf{Y}_t - \overline{\mathbf{X}})) \tag{12}$$

where S_t is the estimated pose, S, at time t.

2.3 Cascade Implementation

We extend the proposed model by applying the search algorithm in a hierarchical fashion, first localizing a small subset of highly salient points which are then used as an initialization for a more complex model with a greater number of points and shape modes. Importantly, search parameters and the properties of the MRF are re-learned at each level such that the correct distribution parameters (*i.e.* all μ_{ij} and Σ_{ij}) are employed rather than (incorrectly) assuming constant values for all levels. Our motivation is to use the most salient points to efficiently and accurately estimate the global pose (translation, scale and orientation) of the object before estimating the locations of the remaining features with a more flexible global shape model. Given the cascade of combined global and MRF-based local shape models and a target image, the following algorithm (see also Figure 1) is used to localize the object of interest using C levels:

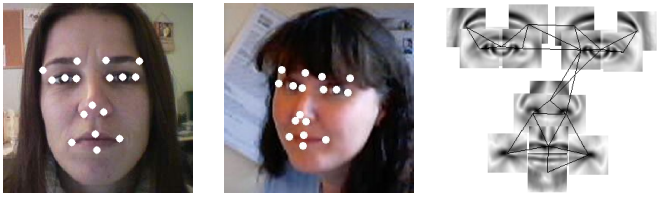
1. Initialize point locations, $\mathbf{X}^*_{0,0}$, using locations that are fixed with respect to face detector output.
2. For $c = 1 \ldots C$
   (a) Initialize $\mathbf{X}^*_{0,c}$ by fitting the shape model at level c to $\mathbf{X}^*_{T^{c-1},c-1}$ via weighted least-squares with a zero-mean Gaussian prior over shape parameters, $\mathbf{b}$.
   (b) Compute points $\mathbf{X}^*_{T^c,c}$ using the search algorithm with level-specific values for number of iterations (T^c), MRF parameters (λ^c and all μ^c_{ij} and Σ^c_{ij}) and search parameters (k^c_i and r^c_i). This involves searching in a radius r^c_i around the current estimate of each point position and finding the best k^c_i candidates for each. An MRF solver picks the best combination of candidates by minimising (10). These are then used to update the global pose and shape parameters.

As an example, in experiments with models of the face we use a two-level cascade with 7 points in the first level and 17 points in the second (see Section 3.1.1). The feature finder is iterated at each level until E reaches a minimum. Typically this involves only a few iterations, though this can be modified by imposing hard limits on the number of iterations and by modifying termination criteria.

2.4 Parameter Estimation

At each level of the cascade, we re-estimate the MRF parameters and search parameters from training data to reflect the increasing accuracy of our estimated solution. For example, due to high uncertainty in scale and orientation of the detected face (the MRF potentials are invariant to global translations), the spread of pairwise potentials at the first level is likely to be large. After applying one iteration of the search, however, we would expect the error distribution between the updated points and their true values to have a much smaller spread. Therefore, we re-estimate all μ_{ij} and Σ_{ij} at each level.

We also estimate appropriate values for all r_i (*i.e.* the search radius for point i) at each level because once we have accurately located a small subset of highly salient points at one level, we should not need to search as far for these points at the next. Therefore, we set r_i to the maximum distance between the estimated and true location for point i over the training set. In effect, this is a principled method of constraining points that we think have already been accurately localized.



(a) (b) (c)

Figure 2: (a,b) Example training images with landmark points; (c) diagram of the trained model.

Similarly, we estimate k_i (*i.e.* the number of local minima to consider for point i) by noting the maximum rank among the considered candidates of the true candidate (*i.e.* the closest candidate to the true location). In effect, we consider fewer candidates for highly discriminant landmarks that result in few spurious local minima, maintaining efficiency in a principled manner.

Finally, λ (*i.e.* the weighting between prior and likelihood) is estimated during the training phase via an exhaustive search over a fixed (typically $\sim$50) number of values, based on the mean error after applying the model to the training set of images.

3 Results

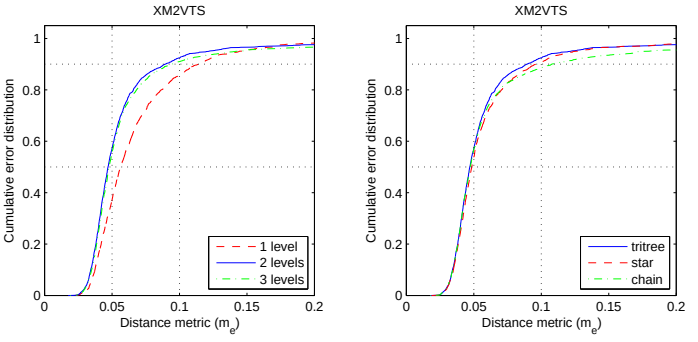
In this section we perform systematic experiments evaluating the influence of the number of levels of the cascade and the structure of the MRF prior. We then compare performance with the PSM [10] and CLM [7] algorithms before demonstrating the method on hand radiograph images. To evaluate performance for a given image we compute $d(\mathbf{x}_i^*, \mathbf{z}_i)$, the mean Euclidean distance between each estimated point, $\mathbf{x}_i^*$, and its hand-labelled counterpart, $\mathbf{z}_i$. This is then normalized with respect to a reference distance, d_{ref} (*e.g.* the distance between the eye centres) in order to make error invariant to scale:

$$m_e = \frac{1}{N} \sum_i \frac{d(\mathbf{x}_i^*, \mathbf{z}_i)}{d_{ref}}. \tag{13}$$

Results are presented in the form of a cumulative error distribution, as employed in other related studies [7].

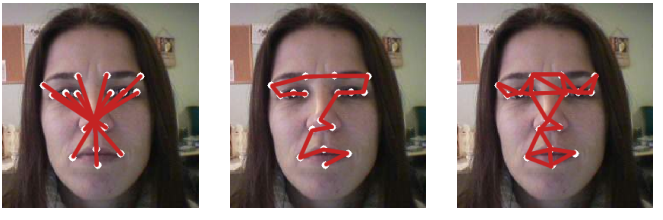
3.1 Faces

To train a face model, we use a dataset of 1052 face images collected from a number of sources under varying conditions (Figure 2). These faces are hand-annotated with 17 points of interest at landmarks on the eyes, eyebrows, nose and mouth. We evaluate the model's performance with two commonly used face image datasets: XM2VTS [20] consists of 2360 images (covering 295 subjects) of 720×576 pixels; BioID contains 1521 images (covering 23 subjects) of 384×286 pixels. Both datasets have the corresponding 17 landmark points hand-labelled on every face. All systems are initialized using the output of a Viola-Jones face detector [27]. Since the true feature locations were available, we eliminated failed detections from our evaluations by discarding examples where the true points did not fall within the limits of the face detector output.



(a) (b)

Figure 3: Feature localization performance with respect to: (a) number of cascade levels; (b) MRF structure.



(a) (b) (c)

Figure 4: Different MRF model structures: (a) star; (b) chain; (c) tri-tree.

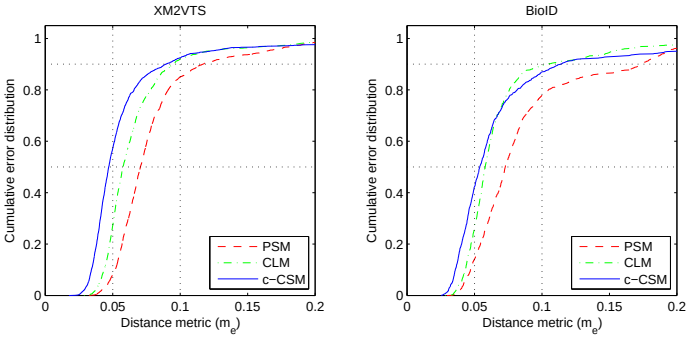
3.1.1 Effect of Cascade Structure

One of the novelties of c-CSM is its use of a cascade architecture that allows a coarse alignment between the model with the image data using only a small subset of highly salient features, followed by a more accurate alignment of all points in higher cascade levels. The following experiment examines the impact of increasing the number of levels in the cascade by comparing cascades with one level (17 points), two levels (7 and 17 points) and three levels (7, 12 and 17 points). The algorithm was run to convergence at each level of the cascade. The cumulative frequency curves of m_e for each cascade (Figure 3a) suggest that increasing complexity from one to two levels increases accuracy but increasing complexity beyond this does not improve performance further.

3.1.2 Effect of MRF Structure

We also investigated the effect of varying the MRF structure among configurations such as a star, a chain or a 'tri-tree'[1] (Figure 4) all of which can be solved using dynamic programming. The results of this experiment (Figure 3b) show that the 'tri-tree' structure is slightly most successful, followed by the chain and finally the star configuration. This is an intuitive result since the 'tri-tree' configuration includes more pairwise potentials (*i.e.* makes fewer

[1]Where all but the two root nodes have two parents such that the set of triplets that are mutually connected (*i.e.* that share an edge) form a tree.



(a) (b)

Figure 5: Feature localization performance for the PSM, CLM and c-CSM on: (a) XM2VTS; (b) BioID.

independence assumptions) than the star or chain representations, and is therefore a closer approximation to the joint prior [6].

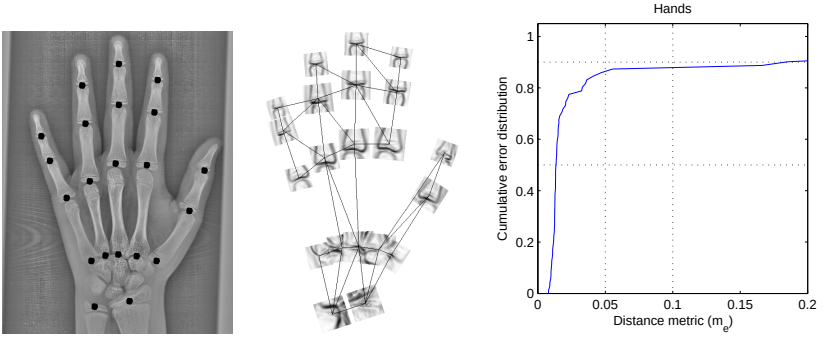
3.1.3 Comparison of c-CSM with Related Techniques

We now present results showing that our model compares favourably with the PSM [11] and the CLM [7] trained with the same data and using comparable model parameters. In our implementation, the PSM performs a global search over the sampled image region specified by the face detector output; the CLM is a constrained local optimization of the PSM output. The results show that that our proposed method demonstrates a reduction of approximately $0.01 \cdot d_{ref}$ in the median value of m_e for the XM2VTS dataset (Figure 5a). For the more challenging BioID dataset (Figure 5b), there is a small reduction in the median error though its significance is unclear. There are also slightly more failures in the BioID experiment, for example due to failed detections when subjects have their eyes closed. Accounting for missing candidates is the subject of ongoing work. We also note that these methods use simple correlation-based likelihood scores and would expect that more complex appearance models would improve localization accuracy further.

In terms of computational demand, the PSM and CLM approaches are more efficient ($\sim$200ms per BioID image compared to $\sim$300ms for the c-CSM) though one should note that the c-CSM also captures rotation in the plane (requiring repeated sub-sampling of the image) which our implementations of the PSM/CLM do not.

3.2 Hand Radiographs

The method is applicable to a range of objects. For instance, we have constructed a model to locate structures on hand radiographs. We have a challenging dataset of 2400×3000 pixel hand radiographs that exhibit large variation in hand pose. Using 72 of the images with 21 selected landmarks manually annotated (see Figure 6a for an example), we built a model (Figure 6b) with two levels of 12 and 21 points, respectively. We then tested the model on the remaining 70 images, giving a median m_e of approximately approximately $0.015 \cdot d_{ref}$ where d_{ref} was defined as the length of the fifth metacarpal. Again, we anticipate improved performance when we make use of more sophisticated feature detectors.



(a) (b) (c)

Figure 6: (a) Example hand radiograph image; (b) the trained model; (c) performance evaluation on 72 images.

4 Conclusion

We have investigated a novel framework for combining global shape models with MRF-based local models, using a hierarchical matching approach where lower levels of the cascade give a coarse alignment that is later refined by higher levels with a more flexible shape model. System parameters are estimated from training data rather than specified empirically, resulting in a method that outperforms the similar recent methods on standard datasets.

One open question concerns the optimal number of shape modes to use in the global shape model at each level in the cascade. Preliminary investigations suggest that this relationship, controlling the distribution of shape variation between the global and local models, may be complex.

Acknowledgements

This work was supported by European Funded Project FP7-2007-ICT-1: Mobile Biometrics (MoBio).

References

[1] G. Behiels, D. Vandermeulen, F. Maes, P. Suetens, and P. Dewaele. Active shape model-based segmentation of digital X-ray images. In *Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*, pages 128–137, 1999.

[2] T. Cootes and C. J. Taylor. A mixture model for representing shape variation. *Image Vision Comput.*, 17(8):567–574, 1999.

[3] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – Their training and application. *Comput. Vis. Image Und.*, 61(1):266–275, January 1995.

[5] J. Coughlan and S. Ferreira. Finding deformable shapes using loopy belief propagation. In *Proc. European Conf. on Computer Vision*, volume 3, pages 453–468, 2002.

[6] D. Crandall, P. Felzenszwalb, and D. Huttenlocher. Spatial priors for part-based recognition using statistical models. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, 2005.

[7] D. Cristinacce and T. F. Cootes. Automatic feature localisation with constrained local models. *Pattern Recogn.*, 41:3054–3067, 2008.

[8] D. Cristinacce, T. Cootes, and I. Scott. A multi-stage approach to facial feature detection. In *Proc. British Machine Vision Conf.*, 2004.

[9] R. Donner, B. Micusik, G. Langs, and H. Bischof. Sparse MRF appearance models for fast anatomical structure localisation. In *Proc. British Machine Vision Conf.*, volume 2, pages 1080–89, 2007.

[10] P. Felzenszwalb. Representation and detection of deformable shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(2), February 2005.

[11] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *Int. J. Comput. Vis.*, 61(1):55–79, January 2005.

[12] R. Fergus, P. Perona, and A. Zisserman. Weakly supervised scale-invariant learning of models for visual recognition. *Int. J. Comput. Vis.*, 71:273–303, 2007.

[13] Y. Huang, Q. Liu, and D. N. Metaxas. A component based deformable model for generalized face alignment. In *Proc. IEEE Int'l Conf. on Comp. Vis.*, pages 1–8, 2007.

[14] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1568–1583, 2006.

[15] K. Lekadir and G.-Z. Yang. Optimal features point selection and automatic initialisation in active shape model search. In *Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*, volume I, pages 434–441, 2008.

[16] L. Liang, F. Wen, X. Tang, and Y.-Q. Xu. An integrated model for accurate shape alignment. In *Proc. European Conf. on Computer Vision*, 2006.

[17] L. Liang, F. Wen, Y.-Q. Xu, X. Tang, and H.-Y. Shum. Accurate face alignment using shape constrained Markov network. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006.

[18] C. Liu, H.-Y. Shum, and C. Zhang. Hierarchical shape modelling for automatic face localization. In *Proc. European Conf. on Computer Vision*, 2002.

[19] I. Matthews and S. Baker. Active appearance models revisited. *Int. J. Comput. Vis.*, 26 (10):135–164, October 2004.

[20] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Int'l Conf. on Audio- and Video-Based Biometric Person Authentication*, 1999.

[21] J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors. *Towards Category-Level Object Recognition*. Springer-Verlag, 2006.

[22] W. Qu, X. Huang, and Y. Jia. Segmentation in noisy medical images using PCA model based particle filtering. In *SPIE Conf. on Medical Imaging*, 2008.

[23] S. Romdhani, S. Gong, and A. Psarrou. A multi-view nonlinear active shape model using kernal PCA. In *Proc. British Machine Vision Conf.*, 1999.

[24] J. Schmid and N. Magnenat-Thalmann. MRI bone segmentation using deformable models and shape priors. In *Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*, 2008.

[25] T. Schwarz, T. Heimann, I. Wolf, and H. P. Meinzer. 3D heart segmentation and volumetry using deformable shape models. *Computers in Cardiology*, 34:741–744, 2007.

[26] J. Tu, Z. Zhang, Z. Zeng, and T. Huang. Face localization via hierarchical condensation with fisher boosting feature selection. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 719–724, 2004.

[27] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vis.*, 57(2): 137–154, May 2004.

[28] Y. Zhou, L. Gu, and H.-J. Zhang. Bayesian Tangent Shape Model: Estimating shape and pose parameters via Bayesian inference. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2003.