Adaptable landmark localisation: applying model transfer learning to a shape model matching system

C. Lindner[1], D. Waring[2], B. Thiruvenkatachari[2], K. O'Brien[2], and T.F. Cootes[1]

[1] Centre for Imaging Sciences, The University of Manchester, UK
[2] School of Dentistry, The University of Manchester, UK

Abstract. We address the challenge of model transfer learning for a shape model matching (SMM) system. The goal is to adapt an existing SMM system to work effectively with new data without rebuilding the system from scratch.
Recently, several SMM systems have been proposed that combine the outcome of a Random Forest (RF) regression step with shape constraints. These methods have been shown to lead to accurate and robust results when applied to the localisation of landmarks annotating skeletal structures in radiographs. However, as these methods contain a supervised learning component, their performance heavily depends on the data that was used to train the system, limiting their applicability to a new dataset with different properties.
Here we show how to *tune* an existing SMM system by both updating the RFs with new samples and re-estimating the shape model. We demonstrate the effectiveness of tuning a cephalometric SMM system to replicate the annotation style of a new observer.
Our results demonstrate that tuning an existing system leads to significant improvements in performance on new data, up to the extent of performing a well as a system that was fully rebuilt using samples from the new dataset. The proposed approach is fast and does not require access to the original training data.

Keywords: Model Transfer Learning, Random Forests, Landmark Localisation, Statistical Shape Models, Machine Learning, Model Tuning

1 Introduction

Shape model matching (SMM) plays an important role in a range of medical imaging application areas in both research and clinical practice – including disease diagnosis, treatment planning and assessment of treatment response or progression of disease. Recent work has shown that SMM systems which combine Random Forest (RF) [1] regression with constraints from a linear shape model lead to accurate and robust results across application areas [2–4, 8, 11, 14].

Transfer learning in the context of machine learning describes the ability of a system to apply knowledge learned in a previous task to a new task in a related domain with some commonality [10]. In model transfer learning, the goal is to fine-tune a pre-trained system to new data without access to the original training data [13]. When the new data arrives sequentially then this is also referred to as online transfer learning [15].

An early example of applying online transfer learning in the context of RF regression was given in [12] where On-line Hough Forests were used for object detection and tracking. In other work, transfer learning was applied in the context of RF classification [5, 13].

Our motivation to consider model transfer learning for SMM systems arose from a collaborative project aimed at introducing automated cephalometric SMM systems in clinical practice. Even though definitions exist for the positions of cephalometric landmarks, the actual annotations often are very subjective and based on years of training and experience. While in the long term the goal would be to achieve consistency across surgeries by having a standardised automated system to identify the landmark positions [9], introducing any form of automated systems in clinical practice tends to require a transitional phase. Thus, in the medium term, to get clinicians accustomed to the automation of annotations (a.k.a. tracings), the goal is to have SMM systems that could imitate the individual tracing style of any clinician. Due to the methodological, computational and time requirements for developing such a system, it would not be feasible to do this from scratch for every clinician/surgery. We therefore aim to develop an *adaptable* system that can be quickly and easily refined locally.

Contributions: In this paper, we propose to apply model transfer learning to RF regression-based SMM systems by tuning an existing pre-trained SMM system to new data without access to the original training data. We describe simple but effective RF regression and shape model update schemes, and apply the latter to tuning a cephalometric SMM system to replicate the annotation style of a new observer. We demonstrate that the tuned SMM system significantly improves performance on new data.

For the experiments below, we apply model transfer learning ("tuning") to Random Forest regression-voting Constrained Local Models (RFRV-CLM) [7], and refer to the resulting SMM systems as *tuned SMM systems*.

2 Methods

2.1 Random Forest regression-voting Constrained Local Models

RFRV-CLMs combine a linear shape model with a set of local models which use RFs to vote for the likely position of each landmark point. Full details are given in [7, 8], here we summarise the approach. The shape model and local models are constructed from an annotated training set. Each set of landmark points is encoded as a shape vector, $\mathbf{x}$, by concatenating the n point co-ordinates.

Training the shape model: We align the shapes then build a shape model of the form

$$\mathbf{x} = T(\bar{\mathbf{x}} + \mathbf{P}\mathbf{b} + \mathbf{r}; \boldsymbol{\theta}) \tag{1}$$

where $\bar{\mathbf{x}}$ is the mean shape, $\mathbf{P}$ are the first t eigenvectors of the covariance matrix, with eigenvalues Λ, which define the shape modes, $\mathbf{b}$ the shape parameters, $\mathbf{r}$ allows small deviations from the model, and $T(.;\boldsymbol{\theta})$ applies a global similarity transformation (with parameters $\boldsymbol{\theta}$) to map from a reference frame to the image frame.

Training the local models: For training the RF regressors, the region of interest of the image that captures all landmark points of the object is re-sampled into a standardised reference frame. For every landmark point $\mathbf{p}$ in $\mathbf{x}$ and every image, image patches $\mathbf{v}_i$ are sampled at a set of random displacements $\delta\mathbf{p}_i$ from the true position of the point in the reference frame (i. e. $\mathbf{v}_i$ is centred at $\mathbf{p}+\delta\mathbf{p}_i$). A set of trees are trained on these patches to predict the displacement $\delta\mathbf{p}_i$ from features in the patch. Each tree leaf stores the mean offset $\mathbf{d}$ and the standard deviation σ of the displacements of all training samples arriving at that leaf.

Search using RF regression-voting CLMs: For a new image, given an initial estimate of the pose of the object, the region of interest of the image is re-sampled into the reference frame. Local image patches (centred on $\mathbf{q}$) are then sparsely sampled in the area around the initial estimate of the landmark's position. The relevant features are extracted from each patch and fed into the trees of the RF to make predictions ($\mathbf{q} + \mathbf{d}$) on the true position of the landmark, resulting in a 2D histogram of votes V_ℓ for every landmark point ℓ. All predictions are made using a single weighted vote per tree.

To match the shape model to the new image, the goal is to seek the shape and pose parameters $\{\mathbf{b}, \boldsymbol{\theta}\}$ that maximise the number of votes over all landmarks

$$Q(\{\mathbf{b}, \boldsymbol{\theta}\}) = \Sigma_{\ell=1}^{n} V_\ell(T(\bar{\mathbf{x}}_\ell + \mathbf{P}_\ell\mathbf{b} + \mathbf{r}_\ell; \boldsymbol{\theta})). \tag{2}$$

2.2 Tuning the shape model

Suppose that we have the shape model parameters $(\bar{\mathbf{x}}, \mathbf{P}, \lambda_j)$, and N new shape examples $\{\mathbf{y}_k\}$. If the shape model is not a good representation of the new shapes, we can *tune* the model so it better matches the shapes of the tuning dataset.

Updating the mean shape: If there are small systematic differences in the way the landmark points are defined then a simple approach is to replace the model mean with a mean estimated from the tuning data. If $\boldsymbol{\theta}_k$ is a vector of the parameters of the similarity transformation $T(\mathbf{x}; \boldsymbol{\theta})$ which best matches the model mean to the target shape $\mathbf{y}_k$, i. e. the minimiser of

$$Q(\boldsymbol{\theta}_k) = |T(\bar{\mathbf{x}}; \boldsymbol{\theta}_k) - \mathbf{y}_k|^2 \tag{3}$$

then we can compute a new estimate of the mean as

$$\bar{\mathbf{x}}' = \frac{1}{N} \sum_k T^{-1}(\mathbf{y}_k; \boldsymbol{\theta}_k). \tag{4}$$

Updating the mean shape and the shape model modes: If the tuning dataset is likely to include significant variations not exhibited in the original training set then the modes of the original shape model will not be able to represent the tuning data well. In this case, we need to update both the mean shape and the modes of shape variation. There are sophisticated techniques available for updating eigen-space models [6]. Here we describe a simplified approach.

We first assume that the original mean is suitable for defining the reference frame for the new model. We map each new tuning example into this frame by applying $\mathbf{y}_k \leftarrow T^{-1}(\mathbf{y}_k; \boldsymbol{\theta}_k)$ where $\boldsymbol{\theta}_k$ are the pose parameters which minimise Equation (3).

The covariance of the tuning data about the origin, $\mathbf{S}_{yy}$, and the covariance of the tuning data about the mean ($\bar{\mathbf{y}}$), $\mathbf{S}$, are given by

$$\mathbf{S}_{yy} = \frac{1}{N} \sum_k \mathbf{y}_k \mathbf{y}_k^T \qquad \text{and} \qquad \mathbf{S} = \mathbf{S}_{yy} - \bar{\mathbf{y}}\bar{\mathbf{y}}^T \tag{5}$$

If we had access to the original training data then we could simply add it to the sums in Equation (5). If we, however, only have access to the shape model parameters, we can reconstruct the covariance about the origin for the training dataset using

$$\mathbf{S}_{xx} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T + \bar{\mathbf{x}}\bar{\mathbf{x}}^T \tag{6}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues.

We can then create a merged shape model by combining the means and covariances from the original training data and the new tuning data via

$$\bar{\mathbf{x}}_m = (1-\beta)\bar{\mathbf{x}} + \beta\bar{\mathbf{y}} \qquad\qquad \mathbf{S}_{mm} = (1-\beta)\mathbf{S}_{xx} + \beta\mathbf{S}_{yy} \tag{7}$$

where $\beta \in [0, 1]$ indicates the relative weight on the tuning data.

The new modes of shape variation are the eigenvectors of $(\mathbf{S}_{mm} - \bar{\mathbf{x}}_m\bar{\mathbf{x}}_m^T)$.

2.3 Tuning the RF trees

To tune the RF trees, for each image of the tuning dataset and corresponding manual annotations $\mathbf{x}$, we re-sample the image into the reference frame. For every point $\mathbf{p}$ in $\mathbf{x}$, N_s random displacements ($\delta\mathbf{p}_i$) are generated and image patches $\mathbf{v}_i$ are sampled (centred at $\mathbf{p} + \delta\mathbf{p}_i$). Every image patch $\mathbf{v}_i$ is fed into each tree and a record is made of which of the samples arrive at each leaf.

To update the displacement offset $\mathbf{d}$ at a leaf, using the n samples, $\{\delta\mathbf{p}_i\}$, that arrived at that leaf, we use

$$\mathbf{d} \leftarrow (1-\alpha)\mathbf{d} + \alpha\overline{\delta\mathbf{p}_i} \text{ where } \overline{\delta\mathbf{p}_i} = \frac{1}{n}\sum_i \delta\mathbf{p}_i \tag{8}$$

The parameter $\alpha \in [0,1]$ indicates how much attention to pay to the tuning data. Similarly, to update the standard deviation σ that is stored at a leaf use

$$\sigma^2 \leftarrow (1-\alpha)\sigma^2 + \alpha\sigma_t^2 \qquad \text{with} \qquad \sigma_t^2 = \frac{1}{n}\sum_i \|\delta\mathbf{p}_i - \overline{\delta\mathbf{p}_i}\|^2. \tag{9}$$

3 Experiments

We tested the system on a set of 289 lateral cephalograms. All images were provided by the Central Manchester University Hospital NHS Foundation Trust (CMFT) and were collected under relevant ethical approvals. All cephalograms were annotated ("traced") by two orthodontists with 54 cephalometric landmarks as in Figure 1, resulting in two sets of manual annotations per image.

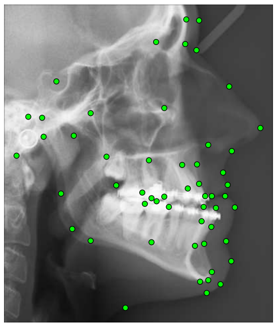


Fig. 1. Cephalometric manual annotation example.

Study design for performance evaluation: We used the dataset to investigate the ability of a tuned SMM system to adapt to a new annotation style. We ran a series of systematic two-fold cross-validation experiments as outlined in Figure 2.

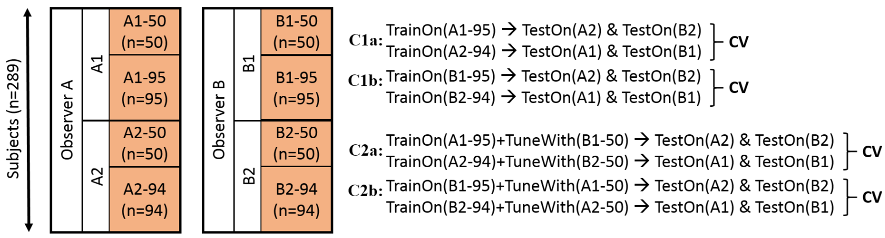
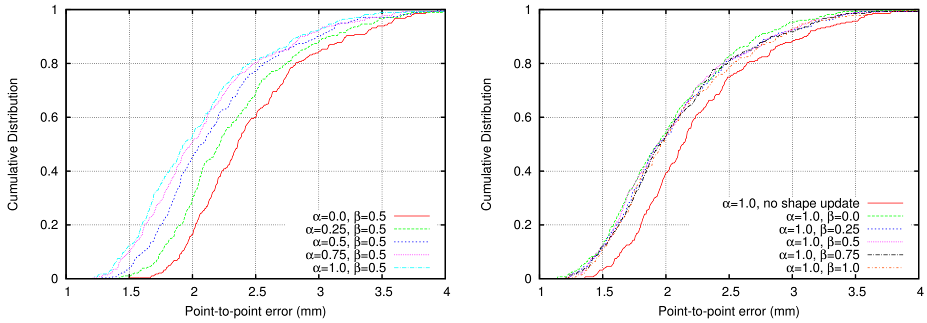


Fig. 2. Cross-validation (CV) study design.

For all experiments, we used a single-stage (coarse) RFRV-CLM with 10 trees as described in [8], and ran five search iterations starting the search from the mean shape at the correct pose.

3.1 Parameter optimisation experiments

To identify how much attention to pay to the new data when tuning an existing SMM system, we ran parameter optimisation experiments varying the values for



(a) Ceph. [train on A, tune with B] (b) Ceph. [train on A, tune with B]

Fig. 3. Parameter optimisation (applying the tuned SMM systems to images of the tuning dataset): analysing the impact on performance of the proportion of new data considered when tuning (a) the trees and (b) the shape model. Note that β=0.0 refers to only updating the mean shape but not the shape model modes.

α (tree tuning) and β (shape model tuning). For all experiments we set N_s, the number of sampled patches per point and image, to 200.

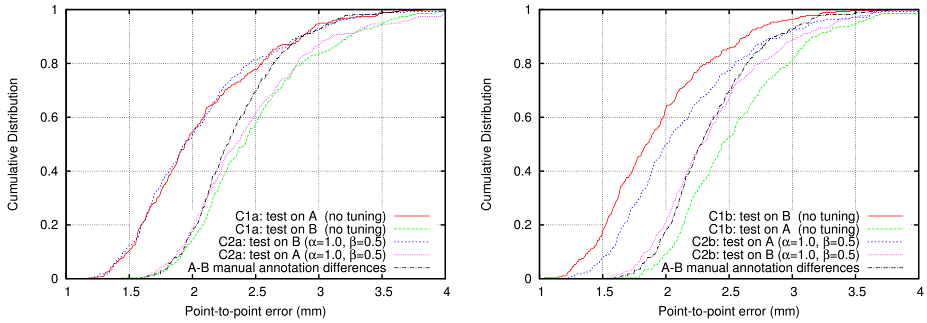
Figure 3 shows the results of applying the tuned SMM systems to new images in the tuning dataset. The best results are obtained when only considering the new data to tune the trees ($\alpha = 1.0$). The shape model tuning results demonstrate that the tuned SMM systems benefit from updating the shape model in addition to tuning the RF trees. Updating the RF trees has more impact on the overall performance improvement than updating the shape model.

3.2 Performance evaluation

To estimate the overall gain in performance by tuning an existing SMM system we compare the performance of the tuned systems with *pure* systems that were fully trained from scratch (i. e. before tuning).

Figure 4 shows the results in both directions for the cephalometric dataset. These demonstrate that overall the pure systems trained on B (C1b) perform slightly better on the training dataset and slightly worse on the tuning dataset, compared to the pure systems trained on A (C1a). The latter are less accurate but generalise better, perhaps because the manual annotations of A are less consistent than the manual annotations of B (since RFRV-CLMs replicate the annotation quality of the training data [9]).

In both directions, tuning the systems (with $\alpha = 1.0$ and $\beta = 0.5$) significantly improves their performance on the tuning dataset. In contrast, the performance of the tuned systems drops on the training dataset with the amount of this performance drop being reversed compared to the performance improvement on the tuning dataset. In Figure 4, the tuned systems C2a perform equally well on the tuning dataset to the pure systems C1a on the training dataset, and the tuned systems C2b perform significantly better on the tuning dataset than



(a) Train on A, tune with B.

(b) Train on B, tune with A

Fig. 4. Comparing the original vs tuned SMM systems (see Figure 2)

the pure systems C1b perform on the tuning dataset but not as well as the pure systems C1b on the training dataset. This leaves the impression that the pattern of performance improvement/drop by tuning is different for both directions but a closer look shows that the amount of improvement/drop is similar and independent of which dataset (A vs B) was chosen for training and tuning.

Figure 4 (red and blue curves) also demonstrates that both the pure and the tuned SMM systems are more accurate on the training and tuning dataset, respectively, than is the agreement between the two manual annotation sets.

For comparison we tried building models from scratch with the training+tuning data. This leads to worse results compared the tuning approach we propose (testing on A [training data = C2b training+tuning data], median error: 2.2mm vs 2mm; testing on B [training data = C2a training+tuning data], median: 2.1mm vs 1.9mm). This is probably because training on all data blurs the differences between the annotators, rather than tuning to the style of the target annotator.

4 Discussion and Conclusions

We have proposed to apply model transfer learning to RF regression-based SMM systems by tuning both the trees and the underlying shape model.

Our results show that tuning to an observer leads to significant improvements in performance (i.e. reducing the discrepancy between the system and the observer on new images). Other experiments (omit for space) show that the same technique can be used to tune the system to a new dataset with different image properties (such as a diseased bone). It took less than 5 minutes to tune a the system, compared to many hours to build a system in the first place. This is of benefit for a range of scenarios such as when there is no access to the methodology required to train an SMM system from scratch, when there are computational or time constraints with regards to generating an SMM system, when the size of the tuning dataset is significantly smaller than the training dataset, or when additional training data might become available at a later stage (which can then be used to tune the system).

Even though in this work model transfer learning was specifically applied to RFRV-CLMs, the proposed and effective RF update scheme can be applied in combination with any shape constraints.

Acknowledgements: C. Lindner is funded by the Engineering and Physical Sciences Research Council, UK (EP/M012611/1).

References

1. L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.
2. C. Chen, W. Xie, J. Franke, P. Grutzner, L.-P. Nolte, and G. Zheng. Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements. *Medical Image Analysis*, 18(3):487–499, 2014.
3. R. Donner, B. Menze, H. Bischof, and G. Langs. Fast anatomical structure localization using top-down image patch regression. In *Proc. MICCAI 2012 - Workshop MCV*, volume 7766 of *LNCS*, pages 133–141. Springer, 2013.
4. Y. Gao, Y. Shao, J. Lian, A. Wang, R. Chen, and D. Shen. Accurate Segmentation of CT Male Pelvic Organs via Regression-Based Deformable Models and Multi-Task Random Forests. *IEEE TMI*, 35(6):1532–1543, 2016.
5. N. Goussies, S. Ubalde, and M. Mejail. Transfer Learning Decision Forests for Gesture Recognition. *Journal of Machine Learning Research*, 15:3847–3870, 2014.
6. P. Hall, D. Marshall, and R. Martin. Merging and splitting eigenspace models. *IEEE TPAMI*, 22(9):1042–1049, 2000.
7. C. Lindner, P. Bromiley, M. Ionita, and T. Cootes. Robust and Accurate Shape Model Matching using Random Forest Regression-Voting. *IEEE TPAMI*, 37(9):1862–1874, 2015.
8. C. Lindner, S. Thiagarajah, M. Wilkinson, The arcOGEN Consortium, G. Wallis, and T. Cootes. Fully Automatic Segmentation of the Proximal Femur Using Random Forest Regression Voting. *IEEE TMI*, 32(8):1462–1472, 2013.
9. C. Lindner, C.-W. Wang, C.-T. Huang, C.-H. Li, S.-W. Chang, and T. Cootes. Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms. *Scientific Reports*, 6(Art. Nr. 33581), 2016.
10. S. Pan and Q. Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.
11. M. Roberts, , T. Cootes, and J. Adams. Automatic Location of Vertebrae on DXA Images using Random Forest Regression. In *Proc. MICCAI 2012 - Part III*, volume 7512 of *LNCS*, pages 361–368. Springer, 2012.
12. S. Schulter, C. Leistner, P. Roth, L. Gool, and H. Bischof. On-line Hough Forests. BMVC, 2011.
13. N. Segev, M. Harel, S. Mannor, K. Crammer, and R. El-Yaniv. Learn on Source, Refine on Target: A Model Transfer Learning Framework with Random Forests. *IEEE TPAMI*, doi:10.1109/TPAMI.2016.2618118, 2016 (ePub).
14. D. Štern, T. Ebner, and M. Urschler. From local to global random regression forests: Exploring anatomical landmark localization. In *Proc. MICCAI 2016 - Part II*, volume 9901 of *LNCS*, pages 221–229. Springer, 2016.
15. P. Zhao, S. Hoi, J. Wang, and B. Li. Online Transfer Learning. *Artificial Intelligence*, 216:76–102, 2014.