Motion-compensation of Cardiac Perfusion MRI using a Statistical Texture Ensemble

Mikkel B. Stegmann[1,2] and Henrik B. W. Larsson[2,3]

[1] Informatics and Mathematical Modelling, Technical University of Denmark, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark
mbs@imm.dtu.dk, http://www.imm.dtu.dk/~mbs/

[2] Danish Research Centre for Magnetic Resonance, H:S Hvidovre Hospital, Kettegaard Allé 30, DK-2650 Hvidovre, Denmark

[3] MR-Senteret, St. Olavs Hospital, Trondheim University, Olav Kyrres Gate 17, N-7006 Trondheim, Norway

Revised edition: July 2003

Abstract. This paper presents a novel method for segmentation of cardiac perfusion MRI. By performing complex analyses of variance and clustering in an annotated training set off-line, the presented method provides real-time segmentation in an on-line setting. This renders the method feasible for e.g. analysis of large image databases or for live non-rigid motion-compensation in modern MR scanners. Changes in image intensity during the bolus passage is modelled by an Active Appearance Model augmented with a cluster analysis of the training set and priors on pose and shape. Preliminary validation of the method is carried out using 250 MR perfusion images, acquired without breath-hold from five subjects. Quantitative and qualitative results show high accuracy, given the limited number of subjects.

1 Introduction

Within the last decade magnetic resonance imaging has been proven able to assess myocardial perfusion in an accurate and safe manner, see e.g. [9]. While scanning times have improved drastically, the amount of manual post-processing remains to render the method prohibitive to clinical practice. A major part of this manual labour is spent by marking up points of correspondence on the myocardium, thus enabling compensation of any motion during a perfusion sequence. This paper present a novel approach aiming at replacing the tedious and error prone labour with an automatic image analysis method, which provides a structured way of collecting and applying expert knowledge given by medical doctors into a learning-based framework.

The paper is organised as follows. Section 2 describes the data used for this study. Section 3 begins by introducing the foundations of this work, namely myocardial perfusion imaging and active appearance modelling, and concludes by describing the proposed method. Section 4 presents a preliminary experimental validation. Finally, Section 5 and 6 serve a discussion of the obtained results and draw some concluding remarks.

2 Data Material

The data material comprises 250 myocardial perfusion, short-axis, magnetic resonance images (MRI). For each of five subjects, 50 sequential images were acquired before, during and after the bolus of contrast. The used contrast agent was gadolinium diethylenetriaminopentaacetic acid (Gd-DTPA). Breath-hold was not used and the time-gap between images was approximately three seconds. Registration relative to the heart-cycle (end-diastole) was obtained using ECG-triggered acquisition from a whole-body MR unit, Siemens Vision, operating at 1.5 T. We used an inversion recovery turbo-FLASH (fast low-angle shot) MR-sequence. Matrix size was 128x128 pixels. Slice thickness was 10 mm. The endocardial and epicardial contours of the left ventricle (LV) were annotated in all images by manually placing 66 landmarks. To fix rotation around the LV long-axis, the right ventricle (RV) was annotated using 12 landmarks.

3 Methods

3.1 Myocardial Perfusion Imaging

Developments in MR-technology during the past decade have made it possible to acquire physiological information about dynamic processes in the human body. As an example of such, myocardial perfusion imaging encompasses assessment of myocardial perfusion at rest and during stress (e.g. pharmacological). By injecting a bolus of contrast the myocardial perfusion mechanism can be quantified, which is essential in ischemic heart diseases. As the contrast agent tags the blood stream and amplifies the MR signal, areas of the myocardium served by diseased arteries show a delayed and attenuated response. Acquisition is carried out dynamically and registered to the heart cycle using ECG-triggering. Images are typically acquired from one or more short-axis slices every n-th heartbeat, trading through-plane resolution for temporal resolution. If the time-window is sufficiently short (typically < 40 secs), breath-hold can be used to remove respiration artefacts. Another source of unwanted variation is erroneous ECG-triggering, giving an erroneous heart-phase and destroying trough-plane (z-axis) correspondence, due to the long-axis movement of the LV during the heart cycle.

3.2 Active Appearance Models

Active Appearance Models (AAMs) [7, 4] were introduced as a method for segmentation and interpretation of face images. This was carried out by building models based on a set of annotated images without any ordering. By being a generic approach, medical applications were soon to follow. These include segmentation of knee cartilage MRI [6], short-axis cardiac MRI [11, 14], metacarpal radiographs [15], diaphragm dome CT [1], echocardiogram time series [3], and corpus callosum in brain MRI [13].

Formally, AAMs establish a compact parameterisation of object variability, as learned from a representative training set. The modelled object properties are

usually shape and pixel intensities. The latter is henceforward denoted *texture*. From these quantities new images similar to the training set can be generated. Objects are defined by marking up each example with points of correspondence (i.e. landmarks) over the set either by hand, or by semi- to completely automated methods. Using a learning-based optimisation strategy, AAMs can be rapidly fitted to unseen images, thus providing image segmentation and analysis.

Variability is modelled by means of a Principal Component Analysis (PCA), i.e. an eigen analysis of the dispersions of shape and texture. Let there be given Q training examples for an object class, and let each example be represented by a set of N landmark points and M texture samples. The shape examples are aligned to a normalised common mean using a Generalised Procrustes Analysis. The texture examples are warped into correspondence using a piece-wise affine warp, normalised, and subsequently sampled from this *shape-free* reference. Typically, this geometrical reference frame is the Procrustes mean shape. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesised shape and texture and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding sample means. New instances are now generated by adjusting the PC scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{b}_s \quad , \quad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{b}_t \tag{1}$$

where $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_t$ are eigenvectors of the shape and texture dispersions estimated from the training set. To obtain a combined shape and texture parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined into

$$\mathbf{b} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \mathbf{\Phi}_s^{\mathrm{T}} (\mathbf{s} - \bar{\mathbf{s}}) \\ \mathbf{\Phi}_t^{\mathrm{T}} (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix}. \tag{2}$$

A suitable weighting between pixel distances and pixel intensities is carried out through the diagonal matrix $\mathbf{W}_s$. To recover any correlation between shape and texture the two eigenspaces are usually coupled through a third PC transform

$$\mathbf{b} = \mathbf{\Phi}_c \mathbf{c} = \begin{bmatrix} \mathbf{\Phi}_{c,s} \\ \mathbf{\Phi}_{c,t} \end{bmatrix} \mathbf{c} \tag{3}$$

obtaining the combined appearance model parameters, $\mathbf{c}$, that generate new object instances by

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{W}_s^{-1} \mathbf{\Phi}_{c,s} \mathbf{c} \quad , \quad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{\Phi}_{c,t} \mathbf{c}. \tag{4}$$

The object instance, $(\mathbf{s}, \mathbf{t})$, is synthesised into an image by warping the pixel intensities of $\mathbf{t}$ into the geometry of the shape $\mathbf{s}$ and applying the current pose parameters $\mathbf{p} = [\ t_x\ t_y\ s\ \theta\ ]^{\mathsf{T}}$ where t_x, t_y and θ denotes in-plane translation and rotation, and s denotes the shape size.

Given a suitable similarity measure the model is matched to an unseen image using an iterative updating scheme based on a fixed Jacobian estimate [5] or a principal component regression [4]. For further details on AAMs refer to [4–6].

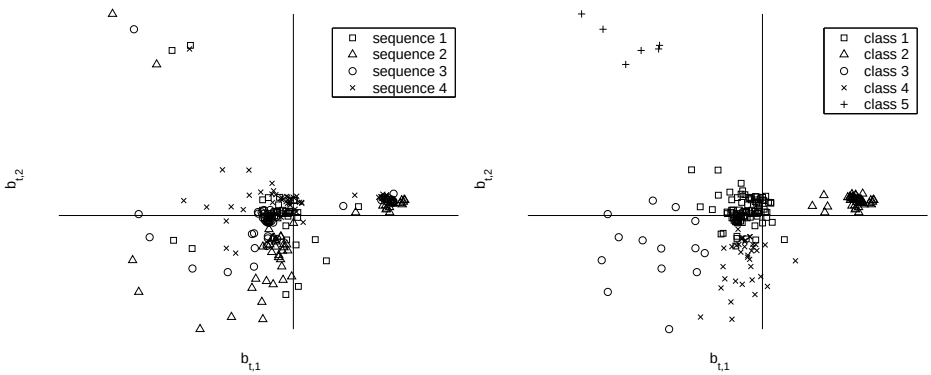

Fig. 1: First versus second principal component of 200 texture vectors from four perfusion sequences (left). Unsupervised classification result using five classes (right).

3.3 Modelling of Perfusion Time-series

Since perfusion MRI sequences differ in structure from the single-image oriented AAMs, this section will discuss the issue of data modelling.

Treating each perfusion sequence as one observation (as in [10]) is not feasible due to random fluctuations in pose and shape induced by the variation sources mentioned in Section 3.1. However, texture variability – decoupled from changes in shape and pose – can be modelled as one observation per sequence providing a more specific model. Unfortunately, to model this behaviour properly, far more training sequences would be required compared to taking each frame as an observation. Consequently, given the low number of sequences, we will treat each image in a sequence as an observation. Circumventing the need for large training sets unfortunately violates the basic assumption in AAMs, that the variation in texture is well modelled by a multivariate Gaussian. Due to the radical changes in intensity during contrast uptake this is clearly not the case. On a coarse level we can split the sequence into i) pre-contrast arrival, ii) contrast agent entering the RV, iii) LV, and iv) the myocardium.

Figure 1 (left) shows the two most significant texture parameters of the texture model shown in Eq. (1), $b_{t,1}$ versus $b_{t,2}$, from the first four sequences in the data set.[1] Here it is verified that the expected clustering is very conspicuous. Modelling this distribution of textures with a multivariate Gaussian gives rise to several problems. Most problematic is that the resulting model is not very specific and can easily generate textures that are not plausible to occur during a perfusion bolus passage.

[1] Since the signal variation of the LV and RV is very small prior to contrast arrival and thus leading to little contrast, the standardisation of texture vectors normally used in AAM would result in severe amplification of scanner noise. Hence, we have only removed the texture mean in all models presented in this paper.

3.4 Adding Cluster Awareness

In order to model the distribution of textures above we propose an unsupervised learning approach that models texture variation using an ensemble of linear subspaces. Alternatively, these subspaces could have been given by an operator, which identifies different phases of each bolus passage. However, to reduce, i) the tedious burden of training set generation, and ii) inter- and intra-observer variability, supervised learning was rejected. Further, since we believe that no canonical number of subspaces exists, we would like to evaluate different ensemble sizes, which would have been very tedious in the case of manual labelling.

Though machine learning literature offers an abundance of classification methods, it is generally agreed upon that no *silver bullet* exists. We have chosen a k-means classification [8] combined with a Monte Carlo simulation scheme where several classifications are carried out, based on different initial random class centres. The final classification is chosen using a minmax criterion, i.e. the classification having the smallest maximum distance to the nearest class centre.

The obtained classification using five classes ($k = 5$) of the data set is shown in Figure 1 (right). From this classification, a set of linear texture subspaces, $\{\mathbf{\Phi}_{t,i}\}_{i=1}^{k}$, is obtained directly by k separate texture PCAs. A corresponding set of texture parameter update matrices, $\{\mathbf{R}_{t,i}\}_{i=1}^{k}$, is obtained following the procedure in [5]. As changes in texture over the sequence is assumed to be uncorrelated with shape, building a joint shape model, $\mathbf{\Phi}_s$, from all sequences yields the best estimate of inter- and intra-subject shape variability. We call this joint model a *Cluster-aware AAM* (CAAM).

Fitting a CAAM to unseen images now involves choosing the appropriate texture subspace. As a reasonable choice for k is very low, model selection is performed by exhaustively trying all models and selecting the model producing the best fit, subject to a set of constraints given later in this paper. To increase performance during model fitting, model selection could be accomplished by a classification of the texture vector into the set of training classes.

To choose k prior knowledge can employed. However, being an optimisation problem in one positive integer variable, we would prefer a data-driven method. Here, the optimal k is estimated using cross-validation on the training set.

3.5 Estimating and Enforcing Pose and Shape Priors

The fact that changes in pose and shape are uncorrelated with the change of texture is highly useful for initialising and constraining the model fitting process in each frame. Further, it can validate the final segmentation results. Thus, if it is possible to obtain reliable estimates of the shape and pose in a subpart of the sequence these can be used in the remains of that sequence. This is the case in the latter part of a bolus passage where the contrast agent has been washed out of the RV and LV, only leaving the subtle changes stemming from the perfusion mechanism in the myocardium. Hence, we propose to estimate prior distributions of pose and shape from the latter part of a perfusion sequence of P frames.

Let κ, γ, D_{max} denote a set of user-selectable constants controlling the influence of the priors. Then, let $\mathbf{\Sigma}$ denote the dispersion matrix of the pose parameters and let $\boldsymbol{\sigma}$ denote the standard deviations of the shape parameters. How these two quantities are estimated is treated in Appendix A. Further, let $\mathbf{F}_t$ denote the t-th frame. Let the set of frames $\{\mathbf{F}_t\}_{t=1}^{S-1}$ denote the unstable period, and the frames $\{\mathbf{F}_t\}_{t=S}^{P}$ denote the stable period. Then, an algorithm for exploiting these priors can be formulated as:

Algorithm 1 Sequence Prior Augmented AAM Search

Require: S, κ, γ, D_{max}, $\mathbf{\Sigma}$ and $\boldsymbol{\sigma}$
1: $\overline{\mathbf{p}}$ = initialisation
2: $\overline{\mathbf{b}}_s$ = initialisation
3: **for** $t = P$ down to S **do**
4: $\{\mathbf{p}_t, \mathbf{b}_{s,t}\}$ = CAAM search started at $\{\overline{\mathbf{p}}, \overline{\mathbf{b}}_s\}$ in $\mathbf{F}_t$
5: $\overline{\mathbf{p}} = \frac{1}{P-t+1} \sum_{j=t}^{P} \mathbf{p}_j$ *)
6: $\overline{\mathbf{b}}_s = \frac{1}{P-t+1} \sum_{j=t}^{P} \mathbf{b}_{s,j}$ *)
7: **end for**
8: **for** $t = S-1$ down to 1 **do**
9: $\{\mathbf{p}_t, \mathbf{b}_{s,t}\}$ = Constrained CAAM search started at $\{\overline{\mathbf{p}}, \overline{\mathbf{b}}_s\}$ in $\mathbf{F}_t$
(constrain using κ, γ, D_{max}, $\mathbf{\Sigma}$, $\overline{\mathbf{b}}_s$, $\boldsymbol{\sigma}$)
10: **end for**
*) Notice that this is carried out most efficiently by using provisional means.
This circumvents the need for storing past observations when calculating the mean.

During CAAM search in the unstable period, pose and shape priors are used to stabilise parameter updates by limiting the maximal update step. To simplify notation, the time index t is omitted from this point on. Let Σ_{ij} denote the element in the i-th row and j-th column of $\mathbf{\Sigma}$, and let p_i denote the i-th element of $\mathbf{p}$. Pose parameter updates can now be constrained using the following simple clamping approach:

$$\delta p_i = \begin{cases} \text{sign}(\delta p_i)\kappa\sqrt{\Sigma_{ii}} & \text{if } |\delta p_i| > \kappa\sqrt{\Sigma_{ii}} \\ \delta p_j & \text{otherwise.} \end{cases} \tag{5}$$

The constant κ acts thus as a clamping constant given in units of standard deviations of pose variation as estimated from the training sequences. In all experiments we have used $\kappa = 0.5$. Likewise, we also exploit the prior knowledge of shape variation, $\boldsymbol{\sigma}$, (as obtained from the training set) in a similar clamping approach,

$$b_{s,i} = \begin{cases} \overline{b}_{s,i} + \text{sign}(b_{s,i} - \overline{b}_{s,i})\gamma\sigma_i & \text{if } |b_{s,i} - \overline{b}_{s,i}| > \gamma\sigma_i \\ b_{s,i} & \text{otherwise,} \end{cases} \tag{6}$$

where γ denotes the maximally accepted distance from the mean in units of standard deviations. We have used $\gamma = 3$ in our experiments.

After ended CAAM search in the unstable period, pose prior enforcement is determined by testing the Mahalanobis distance to the pose distribution:

$$\{\mathbf{p}, \mathbf{b}\} = \begin{cases} \{\overline{\mathbf{p}}, \overline{\mathbf{b}}\} & \text{if } D_{max}^2 < (\mathbf{p} - \overline{\mathbf{p}})^{\mathsf{T}} \mathbf{\Sigma}^{-1} (\mathbf{p} - \overline{\mathbf{p}}) \\ \{\mathbf{p}, \mathbf{b}\} & \text{otherwise} \end{cases} \tag{7}$$

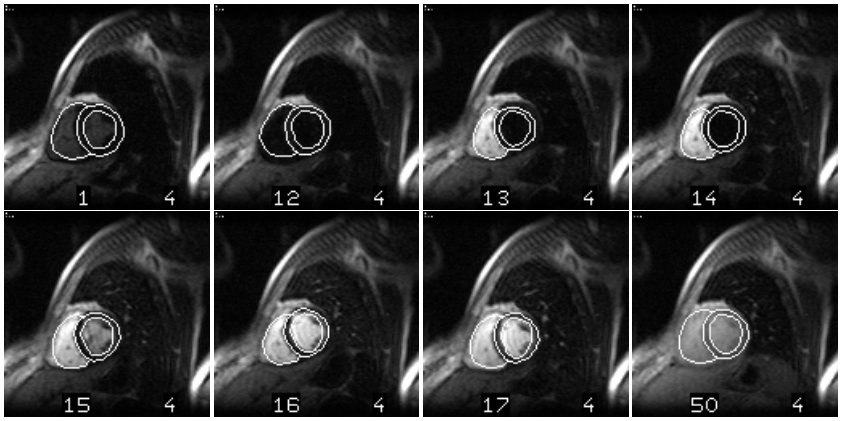

Fig. 2: Segmentation results before, during and after the bolus passage, $k = 3$.

Under the assumption of normal distributed pose parameters, we have chosen to use $D_{max} = 3$. Hence, implausible pose solutions are discarded and replaced with the maximum likelihood of the prior; the mean configuration.

4 Experimental Results

To evaluate the proposed method a Cluster-aware AAM was built using four sequences of 50 frames each. Our software was implemented in C++ (see [14]) and executed on a 1.2 GHz Athlon PC. The stable period was set manually as the last 25 frames of each sequence, i.e. $S = 26$. Using hold-out evaluation the model was tested on the remaining fifth sequence. The model was manually initialised in the P-th frame. Table 1 shows segmentation results for five different values of k. Double-mean landmark errors were calculated as the mean of all landmark points (for both RV and LV) and the mean over all frames. Not surprisingly, we found that the error was above average before and during the bolus passage. Segmentation results for $t = \{1, 12, \ldots, 17, 50\}$ are shown in Figure 2.

Table 1: Segmentation Results for Different Values of k

k	Pose rejects	Pt.pt.[1)] [pixels]	Pt.crv.[2)] [pixels]	Time [sec]
1	3	2.19±0.35	1.33±0.19	0.5
2	3	2.26±0.41	1.38±0.20	0.7
3	0	1.81±0.38	1.03±0.15	1.2
4	2	2.13±0.60	1.19±0.17	1.5
5	0	2.13±0.49	1.24±0.18	1.8

[1)]Pt.pt. measures the Euclidean distance between corresponding landmarks of the model and the ground truth. [2)]Pt.crv. measures the shortest distance to the ground truth curve in a neighbourhood of the corresponding landmark.

To hint the behaviour of the method given a large training set, contrary to the four sequences used in the above, we have performed a leave-all-in evalu-

ation where a CAAM was built using all 250 images from the five sequences, and the tested on the fifth sequence. Consequently, the model had a full representation of the test sequence, except for the regularisation of the texture and shape eigenspaces in an AAM [4]. The results given in Table 2 show a trend of decreasing landmark error with increasing class numbers. Though being rather positively biased, the results also hints the lower bound on the landmark error.

Table 2: Leave-all-in Segmentation Results

k	Pose rejects	Pt.pt. [pixels]	Pt.crv. [pixels]	Time [sec]
1	0	1.80±1.20	0.89±0.52	0.8
2	1	1.70±1.17	0.89±0.65	1.0
3	1	1.30±0.80	0.67±0.37	1.7
4	0	1.50±0.93	0.77±0.40	1.9
5	0	1.13±0.46	0.61±0.24	2.5

5 Discussion

As observed in the previous section, the addition of cluster-awareness seems to increase segmentation accuracy. However, as the set of training sequences is very limited, conclusions would always remain premature and fragile. By design cluster-awareness adds specificity in the texture, which is a very important property for a generative model. In other words, the chance of synthesising implausible images from a cardiac perfusion sequence is drastically reduced. However, the cost turns up as a need for more training examples to reliably estimate a cluster texture model.

Typically, cardiac perfusion images are acquired for several slices in the apex-basal direction. An extension of our approach is straightforward to implement by concatenating texture vectors to obtain one joint multi-slice texture model.

As touched upon earlier, a classification approach could be used instead of a brute-force evaluation of all models. This would lead to a segmentation time corresponding roughly to using one cluster. However, as the initial texture sample is only close to the correct subspace, misclassification is likely to happen. This would need to be dealt with, if a further increase in speed is required.

Compared to a recent approach to segmentation of perfusion cardiac MRI [12] our method offers a 20 times speed-up. Unfortunately, the approach described in [12] is restricted to breath-hold sequences, why a modification would be needed to enable a comparable study using non breath-hold perfusion sequences. Segmentation of perfusion MRI was also reported in [2] where a non-rigid transformation was used compared to our deformable approach.

To let our method generalise to new data, very few assumptions concerning the data content have been made. Except for a few scalar parameters (which are indices relating to the actual data), all values are estimated from training data, rather than being hard-coded into a computer framework. We believe this to be a very fruitful approach, as the method easily adapts to new expert knowledge

given by medical doctors. Knowledge, that typically already exists in the form of hand-annotated training data from previous studies.

Finally, contrary to most other segmentation methods and due to the inherent representation of texture vectors in AAMs, we mention that motion-compensated images are directly obtained by projecting each texture vector in the shape-free reference frame. Thereby a correspondence for each pixel over the complete perfusion sequence is obtained, ready to be fed into a perfusion model as e.g. [9]. In future studies, we aim at validating our method on much larger data sets. As a natural performance benchmark, we will compare estimated perfusion parameters as obtained from automatic versus hand-segmented sequences.

6 Conclusion

We have described a novel, data-driven method for motion-compensation of cardiac perfusion MRI. Preliminary validation of the method showed high segmentation accuracy, considering the small number of subjects available. We anticipate a substantial increase in accuracy when more training data becomes available. The running time of the method using a standard PC is below two seconds for a 50-frame perfusion sequence and can easily be sped up. Thus, the method provides means for segmentation in an on-line setting, e.g. analysis of large image databases or for live motion-compensation in MR scanners.

References

1. R. Beichel, S. Mitchell, E. Sorantin, F. Leberl, A. Goshtasby, and M. Sonka. Shape- and appearance-based segmentation of volumetric medical images. *IEEE International Conference on Image Processing*, 2:589–592, 2001.
2. L. M. Bidaut and J. P. Vallee. Automated registration of dynamic mr images for the quantification of myocardial perfusion. *Jour. Magn. Reson. Imaging*, 13(4):648–655, 2001.
3. J. G. Bosch, S. C. Mitchell, B. P. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. Reiber. Fully automated endocardial contour detection in time sequences of echocardiograms by three-dimensional active appearance models. *Medical Imaging 2002: Image Processing, San Diego CA, SPIE*, pages 452–462, 2002.
4. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. European Conf. on Computer Vision*, volume 2, pages 484–498. Springer, 1998.
5. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 23(6):681–685, 2001.
6. T. F. Cootes and C. J. Taylor. *Statistical Models of Appearance for Computer Vision.* Tech. Report. Feb 2000, University of Manchester, 2000.
7. G. J. Edwards, C. J. Taylor, and T. F. Cootes. Interpreting face images using active appearance models. In *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 300–5. IEEE Comput. Soc, 1998.
8. E. Forgey. Cluster analysis of multivariate data. *Biometrics*, 21:768, 1965.
9. Henrik B. W. Larsson, Thomas Fritz-Hansen, Egill Rostrup, Lars Søndergaard, Poul Ring, and Ole Henriksen. Myocardial perfusion modeling using MRI. *Magnetic Resonance in Medicine*, 35:716–726, 1996.

10. S. Mitchell, B. Lelieveldt, R. Geest, H. Bosch, J. Reiber, and M. Sonka. Time continuous segmentation of cardiac MR image sequences using active appearance motion models. In *Medical Imaging 2001: Image Processing, San Diego CA, SPIE*, volume 1, pages 249–256. SPIE, 2001.
11. S. Mitchell, B. Lelieveldt, R. Geest, J. Schaap, J. Reiber, and M. Sonka. Segmentation of cardiac MR images: An active appearance model approach. In *Medical Imaging 2000: Image Processing, San Diego CA, SPIE*, volume 1. SPIE, 2000.
12. L. Spreeuwers and M. Breeuwer. Automatic detection of the myocardial boundaries of the right and left ventricle in mr cardio perfusion scans. *Proceedings of SPIE - The International Society for Optical Engineering*, 4322(3):1207–1217, 2001.
13. M. B. Stegmann and R. H. Davies. Automated analysis of corpora callosa. Technical Report IMM-REP-2003-02, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, http://www.imm.dtu.dk/∼mbs/, mar 2003.
14. M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME – a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, 2003 (to appear).
15. M. B. Stegmann, R. Fisker, and B. K. Ersbøll. Extending and applying active appearance models for automated, high precision segmentation in different image modalities. In *Proc. 12th Scandinavian Conference on Image Analysis - SCIA 2001*, volume 1, pages 90–97, 2001.

A Estimation of $\boldsymbol{\Sigma}$ and $\boldsymbol{\sigma}$

The dispersion matrix of the pose parameters, $\boldsymbol{\Sigma}$, and the standard deviations of the shape parameters, $\boldsymbol{\sigma}$, can be estimated from the stable period. However, in order to obtain more reliable estimates we propose to use the training sequences, since much more samples are available. To avoid confusion of inter- and intra-sequence variability, we filter out sequence-specific information (such as mean shape size) prior to the estimation. Now, let R denote the number of sequences, each containing P frames. Further, let $\odot$ denote the Hadamard, i.e. element-wise, product. Then the estimation of $\boldsymbol{\Sigma}$ and σ can be specified as:

Algorithm 2 Estimation of Pose and Shape Variation

1: **for** $i = 1$ to R (for each training sequence) **do**
2: $\quad \mathbf{p}_t \leftarrow \mathbf{p}_t \odot (\frac{1}{\bar{s}}[\ 1\ 1\ 1\ \bar{s}\ ]^\mathsf{T}) \ \ \forall \ \ t$ — Normalise pose w.r.t. mean size
3: $\quad \overline{\mathbf{p}}_i = \frac{1}{P} \sum_{t=1}^{P} \mathbf{p}_t$ — Estimate pose mean
4: $\quad \boldsymbol{\Sigma}_i = \frac{1}{P-1} \sum_{t=1}^{P} (\mathbf{p}_t - \overline{\mathbf{p}}_i)(\mathbf{p}_t - \overline{\mathbf{p}}_i)^\mathsf{T}$ — Estimate pose dispersion matrix
5: $\quad \overline{\mathbf{b}}_{s,i} = \frac{1}{P} \sum_{t=1}^{P} \mathbf{b}_{s,t}$ — Estimate shape mean
6: $\quad \boldsymbol{\sigma}_i = \sqrt{\frac{1}{P-1} \sum_{t=1}^{P} (\mathbf{b}_{s,t} - \overline{\mathbf{b}}_{s,i}) \odot (\mathbf{b}_{s,t} - \overline{\mathbf{b}}_{s,i})}$ — Estimate shape std. dev.
7: **end for**
8: $\boldsymbol{\Sigma} = \frac{1}{R} \sum_{i=1}^{R} \boldsymbol{\Sigma}_i$ — Calculate pooled estimate for all sequences
9: $\boldsymbol{\sigma} = \frac{1}{R} \sum_{i=1}^{R} \boldsymbol{\sigma}_i$ — Calculate pooled estimate for all sequences

Further, to maintain simplicity, normalisation w.r.t. rotation has not been included in the above algorithm. However, if sequences differ in orientation, this should be taken into account when estimating t_x and t_y. Finally, during CAAM search, appropriate normalisation and de-normalisation using the stable period mean size (and orientation) must be carried out.