Model-driven segmentation of X-ray left ventricular angiograms

Oost, C.R.

好奇心
'curiosity'

Chapter 4

The Effect of the Composition of the Training Set

This chapter was adapted from:
Automated Contour Detection in Cardiac MRI Using Active Appearance Models: The Effect of the Composition of the Training Set
E. Angelié, E.R. Oost, D. Hendriksen, B.P.F. Lelieveldt, R.J. van der Geest, and J.H.C. Reiber
Investigative Radiology, vol. 42, no. 10, pp. 697-703, 2007.

Abstract

This chapter aims to define the characteristics of an optimal training set for the automated segmentation of short-axis left ventricular magnetic resonance (MR) images in clinical practice, using an Active Appearance Model (AAM). We investigated the segmentation accuracy by varying the size and composition of the training set (i.e., the ratio between pathologic and normal ventricle images, and the vendor dependence). The accuracy was assessed using the degree of similarity and the difference in ejection fraction between automatically detected and manually drawn contours. Including more images in the training set results in a better accuracy of the detected contours, with optimum results achieved when including 200 to 250 images. Using AAM-based contour detection with a mixed model of 80% normal and 20% pathologic cases provides good segmentation accuracy in clinical routine. Finally, this work shows that it is essential to define different AAM models for images from different MRI systems. A model defined on a sufficient number of images with the correct distribution of image characteristics achieves good results in clinical routine.

4.1 Introduction

Cardiac magnetic resonance (MR) imaging is playing an increasingly important role in anatomic and functional assessment of the cardiovascular system. An accurate delineation of the endocardial (endo) and epicardial (epi) boundaries is important to quantify left ventricular (LV) function. Manual segmentation requires expert knowledge and is a time-consuming procedure, which limits the routine clinical use of cardiovascular MR. Moreover, manual segmentation is observer-dependent and therefore is associated with considerable inter- and intra-observer variability [1]. Various automated contour detection techniques have been developed to overcome the disadvantages of manual contour drawing, but clinically available systems still require too much user-interaction.

An automated contour detection method should incorporate a priori knowledge, including information about the cardiac shape as well as information about the image characteristics, which depend on the pulse sequence and the MR hardware used (MR vendors, coils, etc). The widely recognized effectiveness of statistical models stems from their ability to segment images of anatomic structures by exploiting constraints derived from the image data together with a priori knowledge about the location, size, and shape of the structures of interest. These constraints are derived from training data using manually drawn contours. Active Appearance Models have been introduced as a powerful technique for modeling images of anatomic objects and has been successfully used in a variety of automated medical image segmentation applications [2-4].

The AAM segmentation procedure consists of 2 different phases, a training phase and a matching phase:

- For training an AAM, a data set of manually annotated example images is used in which all expert drawn contours should have the same point distribution. Using principal component analysis (PCA) a statistical model is constructed, representing the observed shape variations in the training data. After extracting a shape-free pixel intensity patch for all example images, PCA is applied on these texture vectors, resulting in a model describing the observed pixel intensity variations. Concatenation of both statistical models and another PCA results in an AAM. Additionally, this model learns the relationship between model parameters and the residual errors, induced by known perturbations on single model parameters.

- The matching (detection) phase attempts to find the best fit of the model to the data in a new image. Matching to an image involves finding the model and pose parameters, which minimize the difference between the image and a synthesized model example, projected onto the image.

An elaborate description of model training and matching can be found in [2] and [5]. Despite promising preliminary results in cardiovascular MR [6-9], the validation and definition of optimal settings of such an algorithm are still very challenging. Whereas in the cardiac MR case, large variations in shape characteristics are seen due to the spectrum of pathologies [10], the assessment of the current segmentation techniques remains still narrowed on image data sets obtained from healthy subjects and/or small populations of images [1,5,8,11]. Moreover, in clinical practice a wide range of acquisition protocols coexists, resulting in MR images with large texture variations [12,13]. The evaluation of AAM-based techniques has not yet provided conclusions on the definition of the optimal constitution of the training set. For instance, it has not been studied whether image data from multiple pathologies should be used to define a model, or whether image data from different vendors should be included in a single model, or vendor-specific models should be used.

Therefore, the purpose of this study was to define the composition of the training data set, from which an AAM is constructed that performs optimally in clinical practice. This model should provide good segmentation results for images acquired by scanners of any vendor, it should be able to cope with the entire range of relevant pathologies as seen in short-axis MR examinations, and it should be able to deal with possible poor image quality, induced either by acquisition or by pathology. Three different questions were addressed in this work:

- What is the optimal number of images to be included in the training set?

- What is the optimal mixture of images from healthy and pathologic cases in the training set?

- Is it necessary to construct separate models for different vendors, or is it sufficient to construct a model based on data from multiple vendors?

4.2 Materials and Methods

4.2.1 Study Population

Clinical short-axis MR imaging studies were obtained from 8 institutions using MR equipment from 3 different MR vendors. All examinations were performed using a steady-state free-precession (SSFP, 256 x 256 field of view) imaging protocol on a 1.5 Tesla MR system. The inclusion of data from different institutions guarantees sufficient wide range of variation in imaging protocols and patient population. MR images of 207 LV short-axis examinations (105 from vendor 1, 35 from vendor 2, and 67 from vendor 3) formed the database we used in this study. To differentiate between normal and pathologic cases, we used the criteria proposed by Rominger *et al.* [14], defining an ejection fraction (EF) between 54% and 75% combined with an LV mass between 79 and 137 gram, as normal [15]. The histograms presented in Figure 4.1 illustrate the distribution of EF and LV mass of the population studied. Following these criteria, 98 subjects were considered normal and 109 pathologic.

4.2.2 Slice Labeling and Manual Contour Tracking

Endocardial and epicardial contours were manually traced in the end-diastolic (ED) and end-systolic (ES) phases excluding papillary muscle and trabeculations from the myocardial wall. A reference point was placed in each image at the posterior junction of the right ventricle free wall with the septum, which was used to establish registration between images. For the AAM experiments the images at the most basal and most apical slice level were excluded to avoid extreme shape and texture variations in the AAM model.

4.2.3 Assessment of the Segmentation Quality

To evaluate the quality of the AAM image segmentation, 2 different metrics were used: the degree of similarity and the EF calculation. The degree of similarity is defined as the percentage of points that is similar between 2 contours [16]:

$$\mathrm{S} = \frac{\sum_{n=0}^{N} p_n(d)}{N} \quad where\ p_n(d) = \begin{cases} 1 & if \quad d \le T \\ 0 & if \quad d > T \end{cases} \tag{4.1}$$

where d is the distance between each pair of corresponding points on the manually drawn contour and the automatically detected contour, N is the number of sample points per contour (N = 25 for the presented experiments), and T is a distance threshold [17]. Pairs of corresponding points are assumed to be similar if the distance does not exceed a certain threshold value T (For the presented experiments T was set to T = 2 mm).

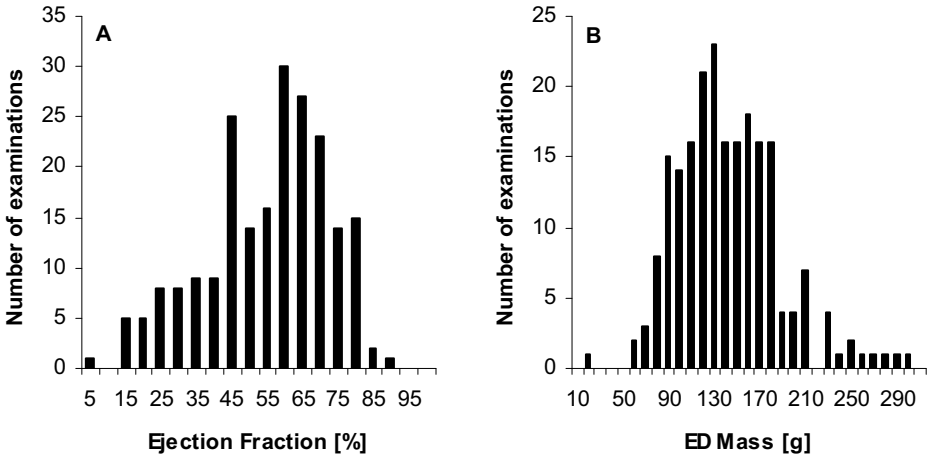

Figure 4.1: Histograms related to the population of MR examinations included in this study. Figure A displays the ejection fraction distribution in the study population. Figure B displays the distribution of ED mass within the study population. These 2 graphs illustrate that the study population contains sufficiently different image characteristics to give a proper representation of the data seen in clinical routine.

The ejection fraction is an important clinical parameter. Evaluation of the difference between the manually derived EF and the automatically derived EF should indicate the clinical relevance of the contour detection using AAMs [18].

4.2.4 Inter-Observer Study

To rate the quality and clinical relevance of the automatic segmentation results obtained by the AAM, we produced inter-observer variability measures based on manual image analysis for comparison. Two observers independently segmented 24 randomly chosen MR examinations manually (50% pathologic and 50% normal, distributed equally between different vendors). The difference in EF and in degree of similarity were computed using only the contours in the ED and ES phase, and slices comprising the section between apex and base. These differences were used as a gold standard in the following studies, presenting the results in a more clinical context.

4.2.5 Optimal Number of Training Images

This study aims at the assessment of the minimal number of data needed to train a model to give good segmentation results. The 2D AAM algorithm used a model shape defined on 25 points equidistantly sampled for both the epicardial and endocardial contours. Ninety-nine percent of variation in shape, texture, and appearance were kept in the defined models to guarantee a proper description of

the variation observed in the training data set [4]. A stepwise increase (from 23 to 298) of the number of randomly chosen training data from a set of normal examinations from a single type of MR scanner was used to define different models that were used to automatically segment an independent set of 194 images with similar characteristics. We analyzed the measurements using a regression analysis.

4.2.6 Impact of the Normal versus Pathologic Ratio

In clinical practice, shape characteristics vary mainly with the pathologies. Therefore, in this study we analyzed the impact on the accuracy of the segmentation when varying the distribution between pathologic (P) and normal (N) examinations in the training data. A fixed number of 180 images were included in the training set, which approximates the number that was defined by the outcome of the previous experiment. Three experiments were realized. The first one consisted of defining a model on a 50% N – 50% P distribution training data. The accuracy of segmentation using such a model was tested on 3 different matching sets (one described with a 50% N – 50% P distribution, the other with a 80% N – 20% P distribution, and the last with a 20% N – 80% P distribution). The second and third experiments consisted of repeating the first experiment with models defined on a 80% N – 20% P and 20% N – 80% P distribution training set, respectively.

4.2.7 Impact of the Distribution of Acquisition Systems

The main cause of texture variation can be attributed to the MR system or to the pulse sequence used [19,20]. Therefore, we also studied the impact on the segmentation accuracy of including images from different vendors in the training set. We created 4 different models: 3 were vendor specific and 1 was created on a mixed population. We performed the automatic segmentation on different sets of images corresponding to different vendors. The “mixed model” was trained on a population of images where all 3 vendors were equally represented. Based on the limited availability of MR images, the models were trained on a set of 76 images defined using 50% N and 50% P examinations and matched on an independent set of 76 images showing the same image characteristics.

4.3 Results

4.3.1 Inter-Observer Study

The inter-observer variability in EF, based on manual contour tracing, was 4.5% ± 2.8%, which is in agreement with previously published results [21]. The degree of similarity between contours drawn by the 2 experts is summarized in Table 4.1. These values will be used as reference values for the degree of similarity and EF.

	ED epi	ED endo	ES epi	ES endo
Degree of Similarity	84 %	84 %	81 %	74 %

Table 4.1: Inter-observer variability in the degree of similarity between two observers.

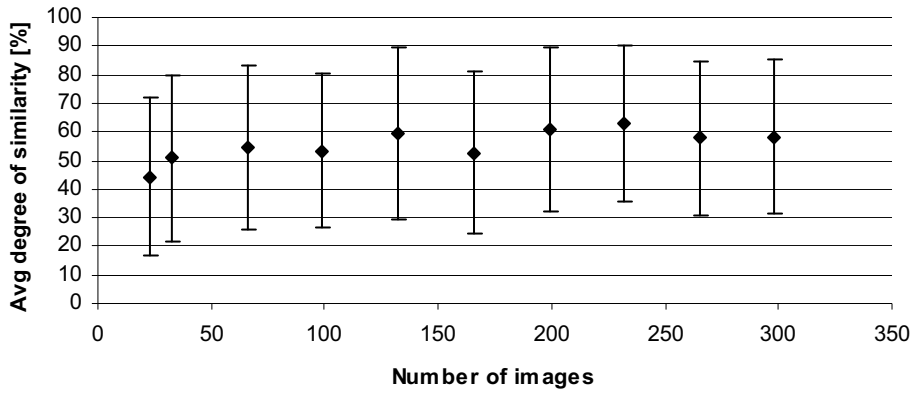

Figure 4.2: Influence of the number of images included in the training on the averaged degree of similarity (Averaged over results for the ED and ES phases, for the endo and epicardial contours).

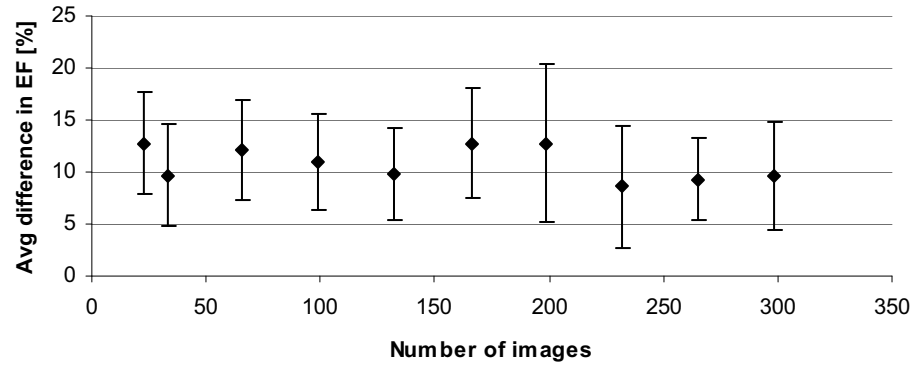

Figure 4.3: Average difference between and standard deviation of the EF calculated using the automatically and manually drawn contours versus the number of images in the training set.

4.3.2 Optimal Number of Training Images

Figures 4.2 and 4.3 illustrate how the quality of AAM contour detection is influenced by the number of images included in the training set. As expected, the

degree of similarity significantly increases with an increasing number of images in the training set ($P < 0.05$ in all the regressions except the variation in ED endocardial contours). The best estimation from Figures 4.2 and 4.3 of the minimal number of images to be included in the training set was around 200 to 250 images. This was supported by the data given in Table 4.2, which shows optimal values for ED at 199 training images and for ES at 232 training images, while including more training images results in only marginally different values. Table 4.2 also points out that the accuracy of the endo contour segmentation is better than the one of the epi contour segmentation and that the quality of the automated segmentation is better for ED than for ES images. In agreement with these observations, Figure 4.3 shows a slight but non-significant ($P = 0.72$) decrease in the deviation in EF with the increase in the number of images in the training set.

	ED Phase		ES Phase		Average
nr of images	endo	epi	endo	epi	
23	74.26	51.03	35.69	16.53	44.38
33	67.59	56.35	44.32	34.75	50.75
66	72.64	62.35	46.22	36.68	54.47
99	71.99	59.14	45.02	36.52	53.17
132	76.91	66.43	50.45	44.56	59.59
166	70.42	60.26	43.92	36.35	52.74
199	79.88	71.35	50.59	42.12	60.99
232	78.05	70.99	55.05	47.54	62.91
265	70.71	63.05	51.75	45.7	57.80
298	76.59	67.6	47.59	41.09	58.22

Table 4.2: Influence of number of images included in the training set on the degree of similarity for the endocardial and epicardial contours in the ED and ES phases.

4.3.3 Impact of the Normal versus Pathologic Ratio

Figures 4.4 and 4.5 display the impact of using a model defined on a set of images for the automatic segmentation of images with different characteristics. Figure 4.4 shows that the highest segmentation accuracy is obtained using the 80% N – 20% P model regardless of the segmentation population. The overall best average degree of similarity was observed when applying the 80% N – 20% P model to 80% N – 20% P data. Similarly, Figure 4.5 shows that the minimum difference between EF is obtained when using a model that describe a mixed population of 80% normals and 20% patient studies.

4.3.4 Impact of the Distribution of Acquisition Systems

Tables 4.3 and 4.4 illustrate the impact of using models from different scanners on the accuracy of the segmentation. In particular, this study demonstrates that the highest accuracy of segmentation (70% in the ED phase and 50% in the ES phase) is obtained using a model defined on a population with the same texture characteristics as the matching population of image. In terms of segmentation performance, a mixed model resulted in lower accuracies (maximally 54% in ED phase and 41% in ES phase). These observations were convincing for vendor 1 and vendor 2 data, however, considerably lower degrees of similarity were found for vendor 3 data, in which rather poor image quality was observed.

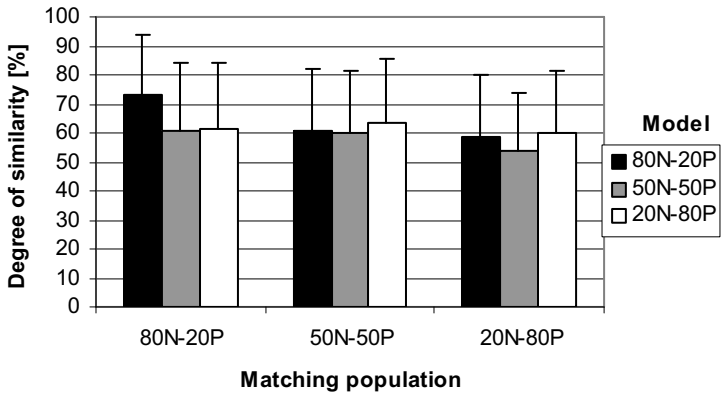

Figure 4.4: Average degree of similarity between automated segmentation results and manually defined gold standard (ED and ES phase, endocardial and epicardial contour) when using different models on different matching populations.

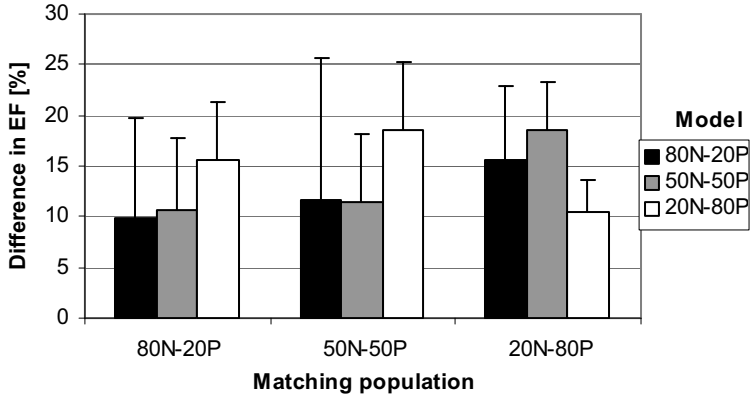

Figure 4.5: Average difference in ejection fraction between automatically detected and manually drawn contours when using different models on different matching populations.

ED degree of similarity	Matching population		
Model	Vendor-1	Vendor-2	Vendor-3
Vendor-1	70±27	57±28	43±28
Vendor-2	60±28	58±33	39±31
Vendor-3	52±22	44±21	34±26
Mixed	54±29	50±29	50±29
Vendor - Mixed	16%	8%	-16%

Table 4.3: Average degree of similarity (%) and standard deviation observed in the segmentation of the ED phase using models trained on images from different vendors. The segmentation accuracy is lower when vendor 3 images/model are concerned.

ES degree of similarity	Matching population		
Model	Vendor-1	Vendor-2	Vendor-3
Vendor-1	50±31	34±30	22±32
Vendor-2	43±29	42±33	21±31
Vendor-3	40±31	26±30	15±30
Mixed	41±30	32±30	27±27
Vendor - Mixed	9%	10%	-12%

Table 4.4: Average degree of similarity (%) and standard deviation observed in the segmentation of the ES phase using models trained on images from different vendors. The segmentation accuracy is lower when vendor 3 images/model are concerned.

4.4 Discussion

Domain knowledge about the geometrical properties of cardiac structures is an important feature for segmentation in medical images. So far, a strong focus was put on the development of new segmentation methods using statistical models. By fitting a model to image data, cardiac surface positions can be predicted with a high accuracy [6,8,20,22,23]. Potentially such a model can be used to automate a large range of diagnostic and therapeutic applications in cardiac medicine [24]. For the application of such a model in the context of automatic segmentation of cardiac image data, practical issues need to be addressed. The goal of this study was to analyze whether AAM-based segmentation could be used in clinical routine. For this, we analyzed the impact of the definition of the training set on the accuracy of the automatic segmentation.

4.4.1 Optimal Number of Training Images

When comparing a model that was trained on 23 images with a model that was trained on approximately tenfold more data, an improvement of the degree of similarity of 6% and 20% could be observed for ED endo en ED epi (both at n=199), respectively. Similar experiments resulted in an increase of the degree of similarity of 19% for ES endo and 31% for ES epi (both at n=232), respectively. This indicates that increasing the number of data included in the training set improves the AAM contour detection. Given that including even more images (n > 199 for ED, n > 232 for ES) does not substantially improve or deteriorate the degree of similarity, a minimal amount of training data should be determined at approximately 200 to 250 images. This is supported by the results for the difference in ejection fraction (Figure 4.3), showing the best results for the three models with the highest amount of training images (n=232, n=265 and n=298).
In addition, the accuracy measured with the degree of similarity in the endo contour detection is higher than in the epi contour detection. This could be explained by the intensity gradient between external tissue and the myocardial wall being weaker than the gradient between the blood pool and the myocardial wall. Thus, the border of the endo contour has a clearer definition, and as a consequence is delineated more robustly than the epi border. The promising endo contour detection results suggest that the presence of papillary muscles or trabeculations does not affect the performance of the AAM. This was supported by visual inspection. Furthermore, the non-significant variation in the degree of similarity measured in the endo contour detection showed that the endo contour detection is less dependent on the number of images included in the training set than the epi contour detection.
The measurement in this study displayed significant increase in the degree of similarity with the number of images defining the model and non-significant decrease in the difference in EF between automatically detected and ground truth contours. From this, we conclude that including numerous images in the training set to define the AAM's model does not matter when the clinical study focuses only on EF measurements, whereas it does have a big impact when analyzing border displacement measurements (wall thickness measurement).
The design of analyzing the optimal number of images to be included in an AAM had some limitations. To facilitate the process of training a model, we narrowed the analysis to only examination of healthy ventricles in the training set. Thus, we limited the variation in LV shape, and possibly artificially reduced the minimally required number of images. This experiment stressed that an AAM model can be described using at least 200 to 250 images from normal examinations, and we expect that more images should be included in the training set for covering all shape change variations in routine clinical practice.

4.4.2 Impact of the Normal versus Pathologic Ratio

A difference in degree of similarity up to 15% was noticed when matching a model describing a particular shape variation on images with different characteristics, and

matching a model defined on the same image characteristics. This experiment stressed the importance of using a suitable model for a population of images used. The accuracy seemed to be reduced when using an equally mixed model.
Although there is a significant improvement of the accuracy of the contour detection when optimizing the population distribution in the training set, the difference in EF measurements (10%) still remained high compared with the inter-observer variability (4.5%). The results show a noticeable discrepancy between EF calculated from detected contours and from reference contours, which can be the consequence of either an average poor detection or the presence of few contour outliers. In fact, the EF is sensitive to a single contour detection failure for a single image, whereas the degree of similarity measurement still remains high. The high degree of similarity of the corresponding contours (~70%) associated with a standard deviation (~20%) leads to the conclusion that the presence of few outliers or detection failure did not affect the degree of similarity because of the larger number of samples included in this measure compared with the EF quantification quantity (the degree of similarity relies on the number of contour data, whereas the EF relies on the number of examinations).
Regarding short-axis MR segmentation, using a model of 80% normals and 20% patients appeared to be the best choice.

4.4.3 Impact of the Distribution of Acquisition Systems

Several inherent problems appear in cardiac MR image segmentation. The change in image contrast [20], the non-uniform nature of the MR signal intensity introduced by noise, physiological factors, and non-uniform radio frequency fields [19] are major challenges when designing and implementing a reliable automated contour detection algorithm.
It was observed that the data from vendor 3 was of relatively poor image quality, mainly due to radio frequency pulse inhomogeneity artifacts. It is expected that Active Appearance Models in which vendor 3 training data was incorporated, were possibly deteriorated. When mutually comparing the results for vendor 1 and vendor 2, it is proven that vendor specific models provide better results than when, for example, a vendor 2 model is used to analyze vendor 1 data, or vice versa. Differences in performance can amount to 10% degree of similarity.
Given the poor image quality of the vendor 3 data, it is difficult to assess the performance of the mixed model. However, averaging the values in the bottom rows of Tables 4.3 and 4.4, results in a positive score for the vendor specific Active Appearance Model results. These findings, combined with the results of the mutual comparison of vendor 1 and vendor 2, shows the need for the application of vendor specific AAMs in clinical practice. In general, poor quality images should evidently not be incorporated in the training data set of an Active Appearance Model.
Due to the availability of image data, this study was designed on only 76 MR images as a training set. Therefore, the overall performance in terms of degree of similarity values is lower than the values reported in the other two experiments. These findings are in correspondence with the results presented in section 4.3.2 and discussed in section 4.4.1.

4.5 Conclusions

It was demonstrated that AAM-based contour detection can be used in cardiac MR imaging studies in clinical practice. Defining an appropriate training set of at least 200 to 250 data sets is a crucial step towards obtaining high quality results of the AAM-based segmentation. Furthermore, the best training set distribution of images from normal and pathologic ventricles seems to be 80–20%. Finally, in case MR scanners from multiple vendors are used, it is essential to define different models for each of the vendors. The inclusion of low quality images in the training set should be avoided.

References

[1] C. Corsi, C. Lamberti, A. Sarti, G. Saracino, T. Shiota, and J. D. Thomas, "A semi-automatic method for left ventricle volume estimate: an in vivo validation study," *Computers in Cardiology*, vol. 28, pp. 109-112, 2001.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *Proceedings of the European Conference on Computer Vision*, H. Burkhardt and B. Neumann, Eds., vol. 2, Berlin: Springer Verlag, 1998, pp. 484-498.

[3] E. Oost, G. Koning, M. Sonka, P. V. Oemrawsingh, J. H. C. Reiber, and B.P.F. Lelieveldt, "Automated contour detection in X-ray left ventricular angiograms using multiview active appearance models and dynamic programming," *IEEE Transactions on Medical Imaging*. vol. 25, no. 9, pp. 1158-1170, 2006.

[4] T. F. Cootes and C. J. Taylor, "Anatomical statistical models and their role in feature extraction," *British Journal of Radiology*. vol. 77, pp. 133-139, 2004.

[5] M. B. Stegmann, B.K. Ersboll, and R. Larsen, "FAME—a flexible appearance modeling environment," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1319-1331, 2003.

[6] M. Üzümcü, R. J. van der Geest, M. Sonka, H. J. Lamb, J. H. C. Reiber, and B. P. F. Lelieveldt, "Multiview active appearancemodels for simultaneous segmentation of cardiac 2-and 4-chamber long-axis magnetic resonance images," *Investigative Radiology*, vol. 40, no. 4, pp. 195-203, 2005.

[7] M. Üzümcü, R. J. van der Geest, C. Swingen, J. H. C. Reiber, and B. P. F. Lelieveldt, "Time continuous tracking and segmentation of cardiovascular magnetic resonance images using multidimensional dynamic programming," *Investigative Radiology*, vol. 41, no. 1, pp. 52-62, 2006.

[8] R. J. van der Geest, B. P. F. Lelieveldt, E. Angelié, M. Danilouchkine, C. Swingen, M. Sonka, and J. H. C. Reiber, "Evaluation of a new method for automated detection of left ventricular boundaries in time serie of magnetic resonance images using active appearance motion model," *Journal of Cardiovascular Magnetic Resonance*, vol. 6, no. 3, pp. 609-617, 2004.

[9] S. C. Mitchell, B. P. F. Lelieveldt, R. J. Van der Geest, H. G. Bosch, J. H. C. Reiber, and M. Sonka, "Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 5, pp. 415-423, 2001.

[10] M. F. Santarelli, V. Positano, C. Michelassi, M. Lombardi, and L. Landini, "Automated cardiac MR image segmentation: theory and measurement evaluation," *Medical Engineering and Physics*, vol. 25, no. 2, pp. 149-159, 2003.

[11] R. Beichel, H. Bischof, F. Leberl, and M. Sonka, "Robust active appearance models and their application to medical image analysis," *IEEE Transactions on Medical Imaging*, vol. 24, no. 9, pp. 1151-1169, 2005.

[12] H. J. Michaely, K. Nael, S. O. Schoenberg, G. Laub, M. F. Reiser, J. P. Finn, and S. G. Ruehm, "Analysis of cardiac function— comparison between 1.5 Tesla and 3.0 Tesla cardiac cine magnetic resonance imaging: preliminary experience," *Investigative Radiology*, vol. 41, no. 2, pp.133-140, 2006

[13] M. Fenchel, V. S. Deshpande, K. Nael, J. P. Finn, S. Miller, S. Ruehm, and G. Laub, "Cardiac cine imaging at 3Tesla: initial experience with a 32-element body-array coil," *Investigative Radiology*, vol. 41, no. 8, pp.601-608, 2006.

[14] M. B. Rominger, G. F. Bachmann, W. Pabst, and W. S. Rau, "Right ventricular volumes and ejection fraction with fast cine MR imaging in breath-hold technique: applicability, normal values from 52 volunteers, and evaluation of 325 adult cardiac patients," *Journal of Magnetic Resonance Imaging*, vol. 10, no. 6, pp. 908-918, 1999.

[15] L. E. Hudsmith, S. E. Petersen, D. J. Tyler, J. M. Francis, A. S. H. Cheng, K. Clarke, J. B. Selvanayagam, M. D. Robson, and S. Neubauer, "Determination of cardiac volumes and mass with FLASH and SSFP cina sequences at 1.5 vs. 3 Tesla: a validation study," *Journal of Magnetic Resonance Imaging*, vol. 24, no. 2, pp. 312-318, 2006.

[16] E. Angelié, J. J. M. Westenberg, R. J. van der Geest, M. Danilouchkine, G. Koning, and J. H. C. Reiber, "How to define the quality of a segmentation: application for the assessment of left ventricular function from short-axis cardiac MR data sets," [abstract] *EuroCMR*, 2002.

[17] P. R. Detmer, G. Bashein, R. W. Martin, "Matched filter identification of left-ventricular endocardial borders in transesophageal echograms," *IEEE Transactions on Medical Imaging*, vol. 9, no. 4, pp. 396-404, 1990.

[18] C. M. Swingen, R. T. Seethamraju, M. Jerosh-Herold, "Feedback-assisted three-dimensional reconstruction of the left ventricle with MRI," *Journal of Magnetic Resonance Imaging*, vol. 17, no. 5, pp. 528-537, 2003.

[19] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger, "MRI segmentation: methods and applications," *Magnetic Resonance Imaging*, vol. 13, no. 3, pp. 343-368, 2004.

[20] I. M. Scott, T. F. Cootes, and C. J. Taylor, "Improving appearance model matching using local image structure," *Proceedings of Information Processing in Medical Imaging*, C. J. Taylor and J. A. Noble Eds., vol. 2732, pp. 258-269, 2003.

[21] R. J. M. van Geuns, T. Baks, E. H. B. M. Gronenschild, J. M. M. Aben, P. A. Wielopolski, F. Cademartiri, and P. J. de Feyter, "Automatic quantitative left ventricular analysis of cine MR images by using three-dimensional information for contour detection," *Radiology*. vol. 240, pp. 215-221, 2006

[22] X. P. Zhu, C. E. Hutchinson, J. M. Hawnaur, T. F. Cootes, C. J. Taylor, and I. Isherwood, "Magnetic resonance image synthesis using a flexible model," *British Journal of Radiology*, vol. 67, no. 802, pp.976-982, 1994.

[23] Z. Zhou, J. You, P. A. Heng, and D. Xia, "Cardiac MR image segmentation and left ventricle surface reconstruction based on level set method," *Studies in Health Technology and Informatics*, vol. 111, pp. 629-632, 2005.

[24] C. Lorenz and J. von Berg, "A comprehensive shape model of the heart," *Medical Image Analysis*, vol. 10, no. 4, pp, 657-670, 2006.