Recognizing cardiac magnetic resonance acquisition planes

Jan Margeta, Antonio Criminisi, Daniel C. Lee, Nicholas Ayache

Recognizing cardiac magnetic resonance acquisition planes

Jan Margeta[1]
jan@kardio.me
Antonio Criminisi[2]
antcrim@microsoft.com
Daniel C Lee[3]
dlee@northwestern.edu
Nicholas Ayache[1]
nicholas.ayache@inria.fr

[1] Asclepios, Inria Sophia Antipolis,
2004 Route des Lucioles
Sophia Antipolis, France

[2] Machine learning and perception group
Microsoft Research,
21 Station Rd, Cambridge, UK

[3] Feinberg Cardiovascular Research
Institute, Northwestern University,
Feinberg School of Medicine,
Chicago, USA

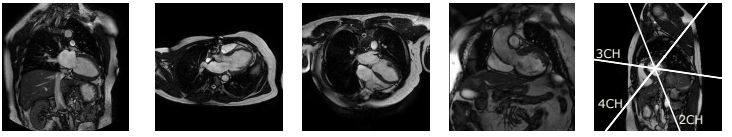
Abstract

In this paper we propose a method for automatic wrangling of missing or noisy acquisition plane information of cardiac magnetic resonance images in order to simplify case filtering and image lookup in large collections of cardiac data. To recognize standard cardiac planes we use features based on image miniatures combined with a decision forest classifier. We show that augmenting the dataset with a set of nondestructive transformations can improve classification accuracy. Our approach compares favorably to the state of the art while requiring fewer manual annotations.

1 Introduction

Cardiac magnetic resonance image acquisitions are mostly performed along several oblique axes with respect to the structures of the heart instead of traditional body planes (coronal, axial and sagital) to better visualize and quantify different aspects of the heart. Knowledge of these planes is also essential for data collection organization, selecting adapted image processing algorithms, grouping of related slices into volumetric image stacks, filtering of cases based on completeness for a clinical study and to respond with the most relevant acquisition plane in content based image retrieval.

This orientation information is often in some way encoded within the DICOM image tags: Series Description (0008,103E) and Protocol Name (0018,1030). The way to encode this view information is however not standardized, operator errors are present or such information is often completely missing. Searching through large databases to handpick desired views from cases in the collection is therefore tedious. Previous work in this field has concentrated mainly on real-time recognition of cardiac planes in echography acquisitions. These methods are based on appearance models, registration or part detection and require additional information to train view, part [6] or landmark specific detectors [10]. Therefore



(a) 2CH (b) 3CH (c) 4CH (d) LVOT (e) SAX

Figure 1: Examples of the main left ventricular long and short axis cardiac MR views.

any new view will require these extra annotations to be made. Otey *et al*. [5] on the other hand addressed this problem by using view labels only from gradient based image features.

1.1 Cardiac planes of acquisition

Optimal cardiac planes depend on global positioning of the heart in the thorax. This is more vertical in young individuals and more diaphragmatic in elderly. Imaging in standard cardiac planes ensures efficient coverage of relevant cardiac territories and enables comparisons across modalities, thus enhancing patient care and cardiovascular research. A good overview of standard cardiac acquisition planes can be found in [8].

Left ventricular long axis acquisition planes. These are usually acquired as 2D or cine 2D+t stack. The 2-chamber, 3-chamber, and 4-chamber views (figs. 1(a) to 1(c)) are used to visualize different regions of the left atrium, mitral valve apparatus, and left ventricle. The 3-chamber and left ventricular outflow tract (fig. 1(d)) views provide visualization of the aortic root from two orthogonal planes. The 4-chamber view enables visualization of the tricuspid valve and right atrium.

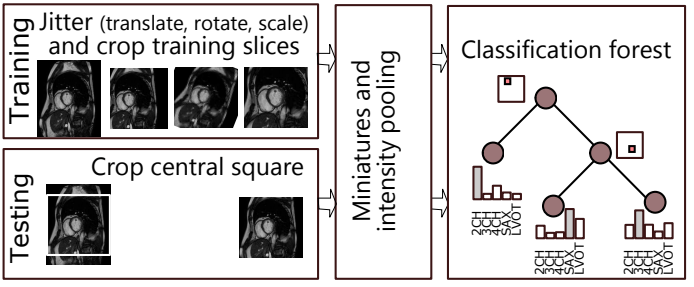
Short axis acquisition planes. Short axis slices (fig. 1(e)) are oriented parallel to the mitral valve ring. These are acquired regularly spaced from the cardiac base to the apex of the heart, often as a cine 3D+t stack. These views are excellent for reproducible volumetric measurements or radial cardiac motion analysis but their use is limited in atrio-ventricular interplay or valvular disease study.

2 Method

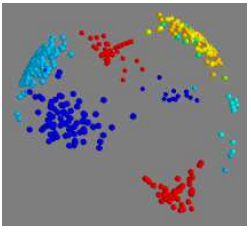
We propose automatic cardiac view recognition pipeline (Fig. 2(a)) that learns to recognize the acquisition planes directly from cardiac magnetic resonance images by combining image miniatures with classification forests. These are trained in a supervised fashion to automatically select the relevant features. To augment the training dataset we jitter the images with small translations, rotations and scales. Apart from the view annotation of the training set, no other information is necessary to classify the views.

2.1 Slice normals from DICOM

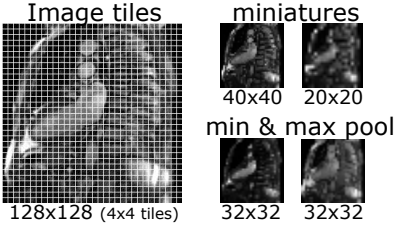
For images where the DICOM orientation (0020,0037) tag is present we can use it to predict the cardiac image acquisition plane. Similarly to [10] we extract image acquisition plane normal vectors as a cross-product of the two image orientation vectors specified in the tag and use this three-dimensional vector feature vector to describe each image (See Fig. 2(b)).



(a) Discriminative pixels from image miniatures are chosen from a random pool as features for a classification forest. We jitter the training dataset to improve robustness to differences in the acquisitions without the need for extra data.



(b) DICOM plane normals for different cardiac views.



(c) Image feature maps as downsampled images and tile intensity minima and maxima.

Figure 2: Our view recognition pipeline (above) and used features (below).

2.2 Image-based features

When the orientation tag is not present we rely on the image intensity information only. To capture differences between planes of acquisition, we extract a set of image based features. To allow use of this for method for all of the above mentioned applications, we use $2D$ image slices individually rather than $3D$ or $3D+t$ volumes.

Preprocessing Our only preprocessing steps are image central square extraction, resampling to 128x128 pixels and linear rescaling of the intensities to range between 0 and 255.

Pooled image miniatures Radiological images are mostly acquired with the object of interest in the center. Therefore some rough alignment of structures can be expected (See fig. 3). Image intensity samples at fixed locations can therefore provide strong clues about the position of different tissues (e.g. symmetric dark lungs or bright cavity in the center).

It has been shown before [9] that significantly down-sampled miniatures can be used

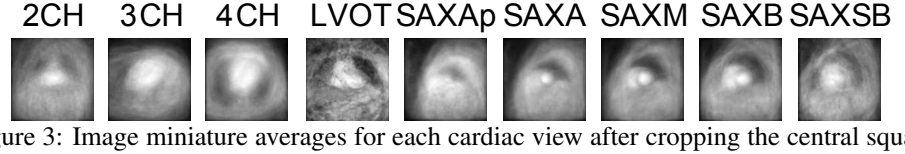


Figure 3: Image miniature averages for each cardiac view after cropping the central square. Short axis view means are displayed from the apex SAXAp to above base slices SAXSB. Distinct patterns of the bright blood pool or dark lungs can be observed for each view.

for such image recognition. In this paper we subsample the cropped centers to two fixed sizes (20x20 and 40x40 pixels). We also create another set of pooled image miniatures from each cropped center by dividing the image into non-overlapping 4x4 tiles and computing intensity minimum and maximum in each of these tiles (fig. 2(c)). As in [4], pooling can add invariance to small image translations and rotations (whose effect is within the tile size). We used directly pixel values from these miniatures sampled at random positions as features. Each image can be seen as a data point in this high-dimensional feature space.

2.3 Decision forest classifier

Random decision forest [1] is an ensemble classification method that consists of a set of binary decision trees. This method is computationally efficient and allows automatic selection of relevant features for the prediction. The tree structure is optimized by recursively partitioning the collection of data points X into the left or right branches such that points with different labels get separated while the same label points are grouped together. At each node of the tree a feature from a randomly drawn subset of all features is chosen such that impurity I in both branches is minimized. We weight samples from the under-represented classes more and we set sample weights inversely proportional to class probabilities and normalize them such that sum $w_{left} + w_{right}$ to equal to one at each node.

$$I(X,\theta) = w_{left}H(X_{left}) + w_{right}H(X_{right}) \tag{1}$$

H is weighted entropy and X_{left} and X_{right} are point subsets falling to either the left or the right branch, based on the tested feature value and threshold. Similarly, w_{left}, w_{right} are sums of sample weights at each branch and w_c is sum of weights for a particular class c.

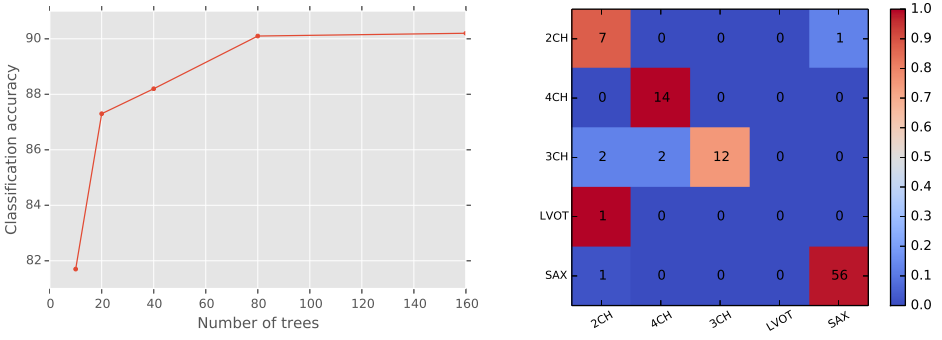
$$H(X) = -\sum_c w_c log(w_c) \tag{2}$$

Only a random subset of features (*i.e.* single pixel values at 64 different locations of the miniatures) is tested at each node of each tree and a simple threshold on this value is used to divide the data points into the left or the right partition. This helps to make the trees in the forest different from each other which leads to better generalization.

When classifying a new image, features chosen at the training stage are extracted and the image is passed through the decisions of the forest (fixed in the training phase) to reach a set of leaves. Class distributions of the reached leaves are averaged across the forest and the most probable label is selected as the image view.

2.4 Augmenting the dataset with jittering

Our dataset contains in total 960 image slices from 100 cardiac patients (SAX:540, 4CH: 140, 2CH: 112, 3CH: 107, LVOT: 9). This might seem rather modest compared to recent image recognition work *e.g.* [4] where deep models are trained with 15 million images. The object of interest is in general placed in the image center, however deviations exist. To account for these differences we expand the training set with extra images by applying a set of transformations of the original training images. Similarly to [4] these are translations (all shifts in x and y between -10 and 10 pixels for a 5x5 grid), but we also add rotations (angles between -10 and 10 degrees with 20 steps) and scales (1-1.4 zoom factor with 8 steps while keeping the same image size) resulting in a total of 51894 training images. Note that this extra expense is mainly at the training time. The test time is almost unaffected except that now a deeper forest can be learnt.



(a) The effect of the number of trees.

(b) Confusion matrix from one fold using image features only.

Figure 4: We selected the optimal parameters via cross validation. Despite the sample weighting the LVOT view recognition suffered from being underrepresented.

3 Validation

Datasets We validated this method on a dataset of 100 patients from multi-center study on post myocardial infarction hearts DETERMINE [3]. These consist of steady state free precession magnetic resonance acquisitions.

Results We ran a 25-fold cross validation by dividing the dataset on patient identifier to prevent biasing our results. This means that images from the same patient (despite being a different view) and therefore the same acquisition could never appear in the training and testing set at the same time. Here we present results for the merged short axis view labels by using features extracted exclusively either from image-based features or from DICOM slice orientation information.

	Image features only		DICOM orientation features only
	No jitter	With jitter	
Accuracy	$87.1 \pm 1.0\%$	$90.2 \pm 1.6\%$	$98.8 \pm 5.4\%$

Increasing the number of trees from 10 to 160 improves the prediction accuracy, however at the expense of increased training and classification computational cost . Beyond 80 trees the forest reaches a plateau with no further benefit (See Fig. 4(a)).

Conclusion

In this paper we presented an automatic cardiac view recognition technique from magnetic resonance images based on image appearance and DICOM plane orientation information. The DICOM orientation based features can be used for our population with confidence when this information is present. When this is not the case, the pure image information method can be used instead. It is achieving precision on par or better with pure image based results of [10] while being much simpler and not requiring to train any landmark detectors.

By jittering the dataset with extra translations, rotations and scales we observed noticeable benefits to view recognition. Extra jittering steps might be useful to improve robustness

to intensity differences between different acquisitions. If enough examples are collected, this method can be simply extended to other pathology specific views such as in congenital heart diseases and acquisition sequences different from SSFP.

We plan to use our method to organize collections of images, semantic image retrieval and parsing of medical literature.

Acknowledgments. This work was partly supported by Microsoft Research through its PhD Scholarship Programme and ERC grant MedYMA. We used data and infrastructure made available through the Cardiac Atlas Project (www.cardiacatlas.org) [2][3]. This work uses scikit-learn toolkit[7].

References

[1] Leo Breiman. Random forests-random features. pages 1–29, 1999.

[2] CG Fonseca, M Backhaus, DA Bluemke, RD Britten, JD Chung, BR Cowan, ID Dinov, JP Finn, PJ Hunter, AH Kadish, DC Lee, JAC Lima, P Medrano-Gracia, K Shivkumar, A Suinesiaputra, W Tao, and AA Young. The Cardiac Atlas Project- an Imaging Database for Computational Modeling and Statistical Atlases of the Heart. *Bioinformatics*, 27(16):2288–2295, 2011.

[3] Alan H Kadish, David Bello, J Paul Finn, Robert O Bonow, Andi Schaechter, Haris Subacius, Christine Albert, James P Daubert, Carissa G Fonseca, and Jeffrey J Goldberger. Rationale and design for the Defibrillators to Reduce Risk by Magnetic Resonance Imaging Evaluation (DETERMINE) trial. *Journal of cardiovascular electrophysiology*, 20(9):982–7, September 2009. ISSN 1540-8167.

[4] Alex Krizhevsky, I Sutskever, and GE Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, 2012.

[5] M Otey, Jinbo Bi, S Krishna, and Bharat Rao. Automatic view recognition for cardiac ultrasound images. In *International Workshop on Computer Vision for Intravascular and Intracardiac Imaging*, pages 187–194, 2006.

[6] J. H. Park, S. K. Zhou, C. Simopoulos, J. Otsuki, and D. Comaniciu. Automatic Cardiac View Classification of Echocardiogram. *IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007.

[7] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[8] A M Taylor and J Bogaert. Cardiovascular MR Imaging Planes and Segmentation. In Jan Bogaert, Steven Dymarkowski, Andrew M Taylor, and Vivek Muthurangu, editors, *Clinical Cardiac MRI SE - 333*, Medical Radiology, pages 93–107. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-23034-9.

[9] Antonio Torralba, Rob Fergus, and William T Freeman. 80 Million Tiny Images: a Large Data Set for Nonparametric Object and Scene Recognition. *IEEE transactions on pattern analysis and machine intelligence*, 30(11):1958–70, November 2008. ISSN 1939-3539.

[10] Y Zhou, Zhigang Peng, and XS Zhou. Automatic view classification for cardiac MRI. In *9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1771–1774, Barcelona, 2012. IEEE. ISBN 9781457718588.