

Discriminative Joint Context for Automatic Landmark Set Detection from a Single Cardiac MR Long Axis Slice

Xiaoguang Lu¹, Bogdan Georgescu¹, Arne Littmann²,
Edgar Mueller², and Dorin Comaniciu¹

¹Siemens Corporate Research, Princeton, NJ, USA

²Magnetic Resonance, Siemens Healthcare, Erlangen, Germany

Abstract. Cardiac magnetic resonance (MR) imaging has advanced to become a powerful diagnostic tool in clinical practice. Automatic detection of anatomic landmarks from MR images is important for structural and functional analysis of the heart. Learning-based object detection methods have demonstrated their capabilities to handle large variations of the object by exploring a local region, context, around the target. Conventional context is associated with each individual landmark to encode local shape and appearance evidence. We extend this concept to a landmark *set*, where multiple landmarks have connections at the semantic level, e.g., landmarks belonging to the same anatomy. We propose a joint context approach to construct contextual regions between landmarks. A discriminative model is learned to utilize inter-landmark features for landmark set detection as an entirety. This helps resolve ambiguities of individual landmark detection results. A probabilistic boosting tree is used to learn a discriminative model based on contextual features. We adopt a marginal space learning strategy to efficiently learn and search a high dimensional parameter space. A fully automatic system is developed to detect the set of three landmarks of the left ventricle, the apex and the two basal annulus points, from a single cardiac MR long axis image. We test the proposed approach on a database of 795 long axis images from 116 patients. A 4-fold cross validation results show that about 15% reduction of the errors is obtained by integrating joint context into a conventional landmark detection system.

1 Introduction

In cardiology, precise information on both the dimensions and functions of the left ventricle (LV) and other heart chambers is essential in clinical applications for diagnosis, prognostic, and therapeutic decisions. Magnetic Resonance (MR) imagery provides a complete morphological LV characterization. The precision on the measures extracted from MR images has been demonstrated and makes MR imagery a standard for LV analysis [1]. Although cardiac MR imaging technologies have rapidly advanced [2, 3], due to considerable amount of available

data, analysis such as segmentation of cardiac images for functionality quantification is time consuming and error-prone for human operators. Automated analysis tools are in great need.

In MR scans, long axis slices are not only used as scout images for acquisition planning, but also are complementary to the short axis stack [4]. Long axis slices captures left ventricle’s shape information and can also be used to correct misregistration of the short axis stack. Long axis acquisitions can be an image sequence or a single slice during acquisition planning, when the short stack and temporal information may not be available. Anatomic landmarks can be used for higher level segmentation, such as initialization of deformable model based approaches [5], and for accelerating acquisition time by facilitating the fully automatic planning of Cardiac MR examinations [6]. Therefore, we are focused on detecting a set of three landmarks from a single cardiac MR long axis slice, containing the apex and two basal annulus points of LV as shown in Fig. 1.

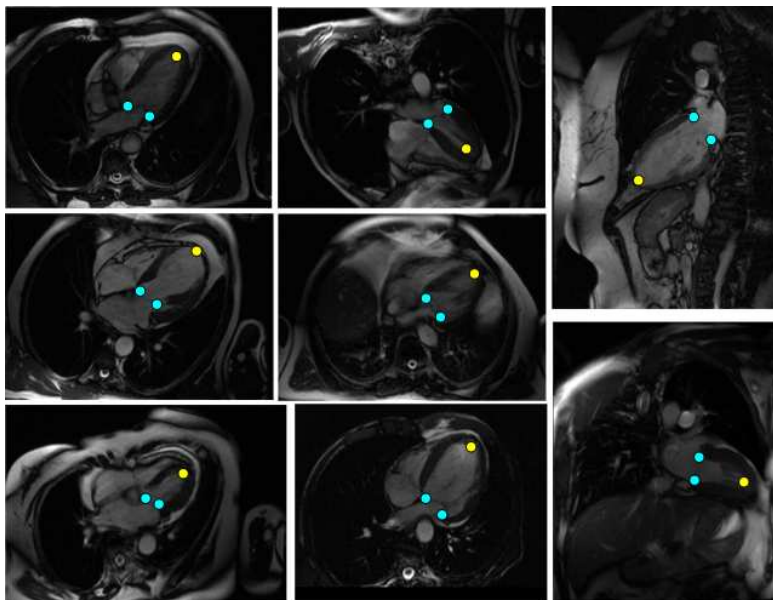


Fig. 1. Examples of cardiac long axis images and three landmarks of interest, which are the apex (yellow) and two basal annulus points (cyan) of LV.

Learning based object detection approaches have been demonstrated successful in many applications [7, 8]. However, they still encounter challenges in a cluttered environment, such as landmark detection in MR long axis slices, due to large anatomy shape and appearance variations across populations along with different acquisition parameters. A number of different anatomies other than the heart also appear in the same slice. For the same patient, time sampling

across the entire heart beat cycle, with ED and ES as two ends, also leads to significantly different myocardium contour shape changes. These variations and ambiguities result in challenges for each individual landmark detector to identify correct landmarks. Cardiac MR long axis image examples along with the three target landmarks in our experiments are shown in Fig. 1.

Context of a landmark is considered as its local evidence such as shape and appearance. Each individual landmark has limited local evidence to identify. However, a set of landmarks are not independent to each other. Correlation in shape and appearance among landmarks can be crucial to identify a landmark set as an entirety. For example, each basal annulus point of LV has limited context, but joint context of the two basal annulus points contains the base plane region, which has a more discriminative power than each individual annulus point to distinguish from other anatomic structures. For those anatomic landmarks that have semantic connections, such as the two basal annulus points and the apex of the left ventricle, joint contextual information (around LV region) captures the correlation of the shape and appearance constructed from the landmark set, which includes more evidence and helps resolve the ambiguities in detecting each landmark individually. Therefore, we propose a joint context based approach under a learning-based object detection framework to automatically identify a landmark set. We define a mapping calculated from a landmark set to derive contextual region, where features are automatically learned to build a discriminative model/classifier. We apply joint context to two scenarios: one for the two basal annulus points, and the other for the set of the apex and the base plane (defined by the two basal annulus points).

2 Methodology

2.1 Overall Workflow

For a cardiac MR long axis image, our joint context based landmark set detection framework includes the following steps (see Fig. 2):

- (1). Apply an apex detector and select the top M candidates.
- (2). Apply a base plane detector and select the top N candidates.
- (3). For each possible candidate pair of <apex, base plane> from the candidate pool (M apex candidates and N base plane candidates), construct joint context.
- (4). Apply the joint context classifier to determine the best pair of <apex, base plane> with the highest probability.
- (5). A local search can be applied to further improve the estimation accuracy of the detected landmarks.

2.2 Joint Context

We define a context operator C to represent context of an object O , whose parameters are denoted by θ , i.e., $C(O|\theta)$. For concise representation purposes, we use

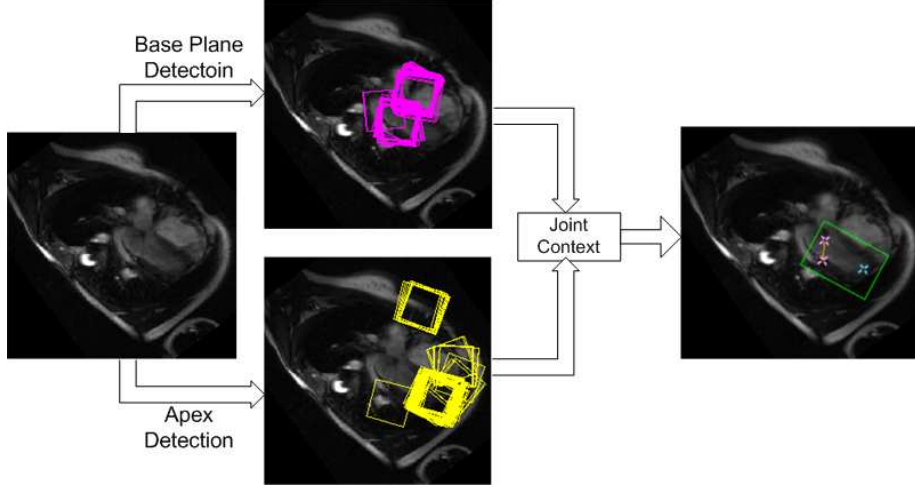


Fig. 2. Joint context based detection workflow.

$C(\theta)$. The operator C is applied to extract features (context information) from contextual appearance. For example, a series of Haar-wavelet like features [7] can be computed and selected by C .

Joint context is defined across a set of landmarks. For two objects O_1 and O_2 , which are represented by their respective parameters θ_1 and θ_2 , the joint context (JC) is defined as:

$$JC = C(f(\theta_1, \theta_2)). \quad (1)$$

JC is represented as appearance and encodes the shape by computing geometric relationship through the mapping f , which derives a geometric region based on θ_1 and θ_2 .

2.3 Joint Context Based Landmark Set Detection

Context Construction We associate a two-dimensional bounding box with each target landmark object and their derived context. Each bounding box is specified by a five-parameter set θ , containing two positions $\langle x, y \rangle$, one orientation $\langle \phi \rangle$, and two scales $\langle sx, sy \rangle$. The original landmark detection task can then be formulated into estimating this set of parameters. Although only positions are used as the output, orientation and scales are useful in encoding proper and consistent context learned during offline training process, where a set of contextual models/classifiers are obtained.

To learn contextual models, we collect a set of cardiac long axis images and annotate the landmark positions. Based on this annotated training set, we build a contextual model for each target object and a joint contextual model for the pair of $\langle \text{apex, base plane} \rangle$. Let $\langle x_a, y_a \rangle$, $\langle x_{b1}, y_{b1} \rangle$, and $\langle x_{b2}, y_{b2} \rangle$

denote the positions of the apex, and two basal annulus points, respectively. The joint contextual parameter sets derived from this landmark set are constructed as shown in Table 1.

Table 1. Context construction (mapping f in Eq. (1)). $\langle x_a, y_a \rangle$, $\langle x_{b1}, y_{b1} \rangle$, and $\langle x_{b2}, y_{b2} \rangle$ denote the positions of the apex, and two basal annulus points, respectively. Base Plane (BP) context is obtained from two basal annulus points. Joint context of $\langle \text{Apex, Base Plane} \rangle$ depends on the apex and the two basal annulus points.

	Positions	Orientation
Apex	$x_a = x_a$ $y_a = y_a$	$phi_a = \arctan \frac{(y_a - (y_{b1} + y_{b2})/2)}{(x_a - (x_{b1} + x_{b2})/2)}$;
Joint context BP	$x_b = (x_{b1} + x_{b2})/2$ $y_b = (y_{b1} + y_{b2})/2$	Orthogonal to the line segment connecting the two basal annulus points, and pointing toward the apex
Joint context $\langle \text{Apex, BP} \rangle$	$x_{jc} = (x_a + x_b)/2$ $y_{jc} = (y_a + y_b)/2$	$\phi_{jc} = \arctan \frac{y_a - (y_{b1} + y_{b2})/2}{x_a - (x_{b1} + x_{b2})/2}$

	Scales
Apex	$sx_a = \sqrt{(y_{b2} - y_{b1})^2 + (x_{b2} - x_{b1})^2} * 1.8^{-1}$ $sy_a = sx_a$
Joint context BP	$sx_b = \sqrt{(y_{b2} - y_{b1})^2 + (x_{b2} - x_{b1})^2} * 1.8$ $sy_b = sx_b$
Joint context $\langle \text{Apex, BP} \rangle$	$sx_{jc} = \sqrt{(y_a - y_b)^2 + (x_a - x_b)^2} * 1.5$ $sy_{jc} = \sqrt{(y_{b2} - y_{b1})^2 + (x_{b2} - x_{b1})^2} * 1.8$

2.4 Learning Discriminative Context

To utilize context for object detection, we build a discriminative model to differentiate the true object from background by calculating the probability of given context (of a candidate) being a target object, denoted as $P(O|C)$. Our joint context based landmark set detection approach follows the database-driven knowledge-based framework, which has been demonstrated effective in object detection and medical applications [7, 8]. Landmark detection is formulated as a two-category classification problem, i.e., true object vs. background. Discriminative features from context are extracted and learned by a machine algorithm based on the experts' annotations, resulting in a probabilistic model for each landmark context or a joint context of a landmark set. The online detection algorithm searches through multiple hypotheses in the parameter space to identify the ones with high probabilities.

Context-based landmark detection is to estimate the parameter set, θ , of an object target from a given image. There are five parameters for each context in our framework, including 2 positions (x, y) , 1 orientation (ϕ) , and 2 scales

along each axis (sx, sy). Because exhaustively searching in this five-dimensional parameter space is expensive for online applications, we adopt a search strategy similar to the marginal space learning based approach, which was proposed by Zheng et al. [9] for 3D object detection. These detectors are trained using positive samples based on the position, orientation, and size of the annotated object context, while the negative set is generated by extracting sub-images that are further from the positive samples in the parameter space.

For each learning/classification task, we use a probabilistic boosting tree (PBT) proposed by Tu et al. [10] as the classifier. The classifier is a tree-based structure with which the posterior probabilities of the presence of the landmark of interest are calculated from candidate context in given images. Following the marginal space learning strategy, we designed a series of classifiers that estimate parameters at a number of sequential stages in the order of complexity, i.e., position, orientation, and scale. Different stages utilize different features computed from image intensities. Multiple hypotheses are maintained between algorithm stages, which quickly removes false hypotheses at the earlier stages while propagates the right hypothesis to the final stage. Only one hypothesis is selected as the final detection result.

A set of discriminative features are selected to distinguish the positive target from negatives from a large pool of features at each stage. For the classifiers at the position stage, we choose Haar wavelet-like features [7], which are efficiently calculated using integral image-based techniques. For the classifiers at the orientation and scale search stages, steerable features [9] are applied, because their computation does not require image rotation and re-scaling, which are computationally expensive.

3 Experiments

We collected 253 long axis sequences from 116 patients (ages range from 11 to 72 years old), from which slices were selected to cover a large range of dynamic heart motion from end diastole to end systole. In total, there are 795 long axis slices (images) in our database. For each slice, three landmarks of LV (two basal annulus points and the apex) were manually annotated by experts and used as groundtruth for evaluation.

For each long axis slice, we applied the proposed joint context based algorithm to detect the landmark set, i.e., two basal annulus points and the apex, in a fully automatic fashion. We computed Euclidean distance between the detected landmark position and its corresponding groundtruth as the detection error for each landmark. The average error of all three landmarks was used as the metric to evaluate the system performance.

A 4-fold cross-validation scheme was applied for evaluation. The entire dataset of 795 images was randomly partitioned into four quarters. For each fold evaluation, three quarters were combined for training and the remaining one quarter was used as unseen data for testing. This procedure was repeated four time so that each image has been used once for testing. Examples of the detection results

are shown in Fig. 3. Performance is summarized based on all 4 folds and provided in Table 2. Both individual landmark detectors and joint context based approach are evaluated under the same experimental protocol. It shows that the overall mean errors are reduced by about 15 percent with the proposed joint context based approach integrated. While the median error is not reduced as much as the mean, this indicates that the proposed approach significantly reduces the outliers generated by the individual landmark detectors. Cardiac MR images in a large population present a large variations of appearance intensities along with the anatomy shape changes across the heart beat cycle, leading to difficulties for accurate identification. Large errors still occur when all individual detectors fail to locate correct landmarks, resulting in zero correct candidate for joint context. On the average, it took about 1.5 seconds to detect the three landmarks from a 400×400 image on a duo core 2.8GHz CPU.

Table 2. Statistics of the average error of all three landmarks (two basal annulus points and the apex) in a 4-fold cross validation. Errors are measured in *mm*.

	Mean	Std	Median
Individual landmark detector	9.4	15.9	5.4
Joint context based landmark set detection	8.0	13.9	5.2

4 Conclusions

We have proposed a joint context based approach that is integrated into a learning-based object detection framework. We have applied joint context and developed a fully automatic system to detect a landmark set from cardiac MR long axis images. The target landmark set contains the apex and the two basal annulus points. The proposed joint context based approach is not dependent on any specific learning algorithms. To account for both scout and diagnosis images, we are focused on a single MR long axis slice. However, for image sequences, the cue of temporal coherence can be integrated to further improve the landmark set detection accuracy.

Acknowledgements

We would like to thank Haibin Ling for helpful discussion and Yefeng Zheng for providing marginal space learning implementation.

References

1. Cousty, J., Najman, L., Couprie, M., Clement-Guinaudeau, S., Goissen, T., Garot, J.: Automated, accurate and fast segmentation of 4D cardiac MR images. In: Functional Imaging and Modeling of the Heart, LNCS 4466. (2007) 474–483

2. Frangi, A., Niessen, W., Viergever, M.: 3D modeling for functional analysis of cardiac images: a review. *IEEE Trans. on Medical Imaging* **20**(1) (2001) 225
3. Finn, J.P., Nael, K., Deshpande, V., Ratib, O., Laub, G.: Cardiac MR imaging: State of the technology. *Radiology* **241**(2) (2006) 338–354
4. Koikkalainen, J., Pollari, M., Lotjonen, J., Kivisto, S., Lauerma, K.: Segmentation of cardiac structures simultaneously from short- and long-axis MR images. In: *Proc. MICCAI*. (2004) 427–434
5. Jolly, M.P.: Automatic segmentation of the left ventricle in cardiac MR and CT images. *International Journal of Computer Vision* **70**(2) (2006) 151–163
6. Young, A., Cowan, B., Thrupp, S., Hedley, W., Dell’Italia, L.: Left ventricular mass and volume: Fast calculation with guide-point modeling on MR images. *Radiology* **216**(2) (2000) 597–602
7. Viola, P., Jones, M.J.: Robust real-time face detection. *International Journal of Computer Vision* **57**(2) (2004) 137–154
8. Georgescu, B., Zhou, X., Comaniciu, D., Gupta, A.: Database-guided segmentation of anatomical structures with complex appearance. In: *Proc. IEEE CVPR*. (2005)
9. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features. In: *Proc. ICCV*. (2007)
10. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: *Proc. ICCV*. (2005) 1589–1596

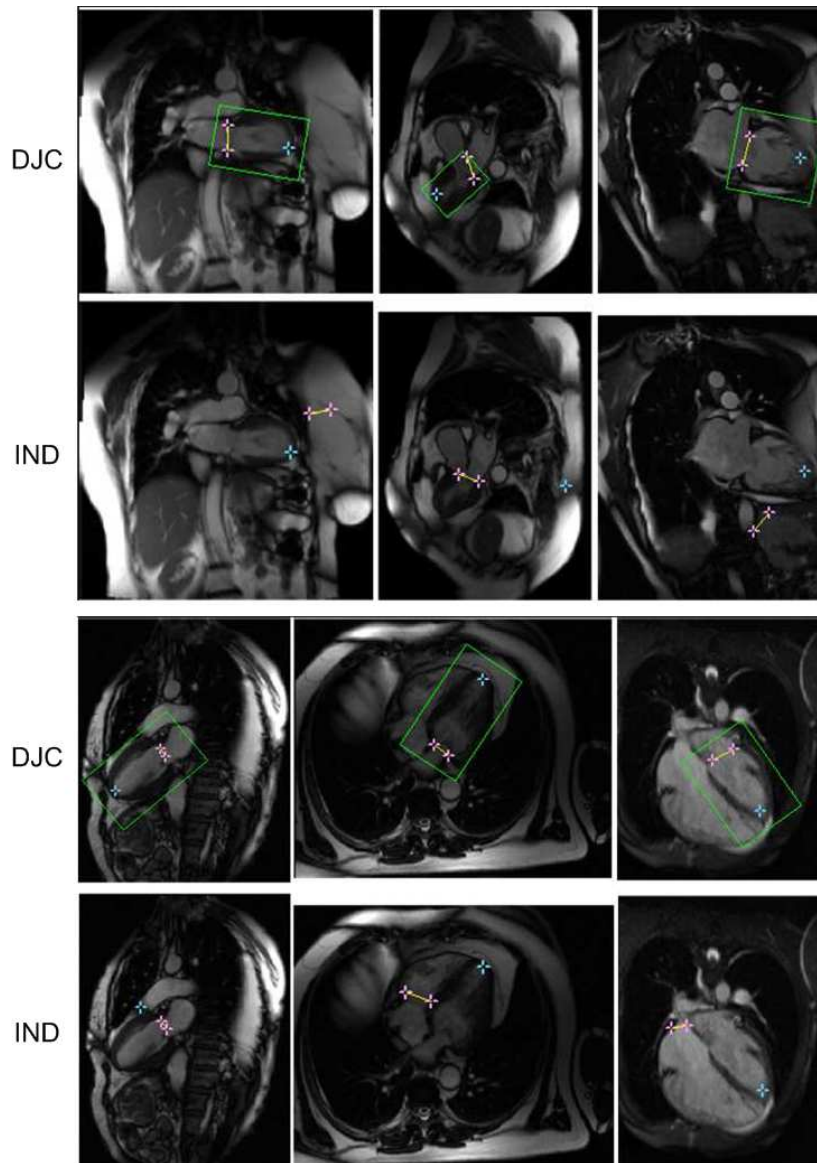


Fig. 3. Bottom ('IND' group): detection results of the base plane detector and the apex detector individually. Top ('DJC' group): landmark set detection results after applying the discriminative joint context (DJC) based approach. The green region represents joint context that provides the highest likelihood value of a pair of <apex, base plane>. Pink: detected two basal annulus points; blue: detected apex.