Data-Driven Feature Learning for Myocardial Segmentation of CP-BOLD MRI

Anirban Mukhopadhyay[1], Ilkay Oksuz[1(✉)], Marco Bevilacqua[1], Rohan Dharmakumar[2], and Sotirios A. Tsaftaris[1,3]

[1] IMT Institute for Advanced Studies Lucca, Lucca, Italy
ilkay.oksuz@imtlucca.it
[2] Biomedical Imaging Research Institute, Cedars-Sinai Medical, Los Angeles, CA, USA
[3] Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA

Abstract. Cardiac Phase-resolved Blood Oxygen-Level-Dependent (CP-BOLD) MR is capable of diagnosing an ongoing ischemia by detecting changes in myocardial intensity patterns at rest without any contrast and stress agents. Visualizing and detecting these changes require significant post-processing, including myocardial segmentation for isolating the myocardium. But, changes in myocardial intensity pattern and myocardial shape due to the heart's motion challenge automated standard CINE MR myocardial segmentation techniques resulting in a significant drop of segmentation accuracy. We hypothesize that the main reason behind this phenomenon is the lack of discernible features. In this paper, a multi scale discriminative dictionary learning approach is proposed for supervised learning and sparse representation of the myocardium, to improve the myocardial feature selection. The technique is validated on a challenging dataset of CP-BOLD MR and standard CINE MR acquired in baseline and ischemic condition across 10 canine subjects. The proposed method significantly outperforms standard cardiac segmentation techniques, including segmentation via registration, level sets and supervised methods for myocardial segmentation.

Keywords: Dictionary learning · CP-BOLD MR · CINE MR · Segmentation

1 Introduction

CP-BOLD MR is a truly noninvasive (without contrast or stress agents and ionizing radiation) method for early diagnosis of ongoing ischemia. CP-BOLD identifies the ischemic myocardium by examining changes in myocardial signal intensity patterns as a function of cardiac phase [14]. However, visualizing and quantifying such changes requires significant post-processing, including

The first two authors contributed equally to this work.

H. van Assen et al. (Eds.): FIMH 2015, LNCS 9126, pp. 189–197, 2015.
DOI: 10.1007/978-3-319-20309-6_22

myocardial segmentation to isolate the myocardium from the rest of the anatomy. In particular, although CP-BOLD is a cine type acquisition, automated myocardial segmentation and registration algorithms developed for standard CINE under-perform, due to the spatio-temporal intensity variations of the myocardial BOLD effect [9], an example of which is shown in Fig. 1. Thus, in CP-BOLD in addition to violations of shape invariance (as with standard CINE MRI) the principal assumption of appearance invariance (consistent intensity) is violated as well.

As a result, no automated CP-BOLD MR myocardial segmentation algorithms exist, and semi-automated methods based on tracking are currently employed [13]. We hypothesize that it is due to the lack of appropriate features, which are invariant yet unique and descriptive under the particular type of appearance and shape deformation observed in CP-BOLD images. Rather than relying on low-level features used often for myocardial segmentation of standard CINE MR which are inconsistent for CP-BOLD MR, a more generalized feature learning method should be developed to accommodate the myocardial BOLD effect while still being reliable in the CINE MR case.

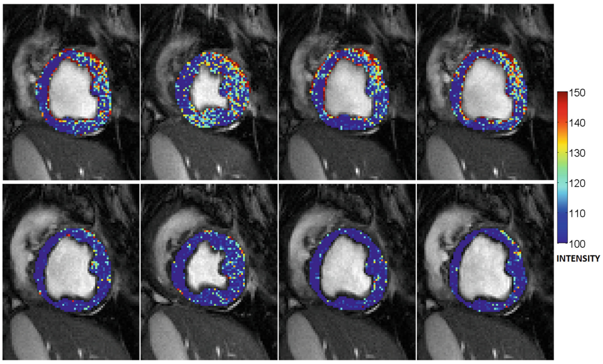

Fig. 1. Exemplary cardiac phases of CP-BOLD MR (top row) and standard CINE MR (bottom row) obtained from the same subject under baseline conditions (absence of ischemia) where the myocardium is color coded to underline the challenge of appearance variation in CP-BOLD MR which is minimal in the case of standard CINE MR (Color figure online).

We adopt a patch-based discriminative dictionary learning technique (which has been used also in echocardiography [6]) to learn features from previously segmented data in a fully supervised manner. The motivation behind the choice of a sparse dictionary is to employ a compact and high-fidelity low-dimensional subspace representation which is able to extract semantic information of the myocardium as well [16]. The key observation behind this strategy is that, though the patch intensity level varies significantly across the cardiac cycle, sparse representations based on learnt dictionaries are invariant across the cardiac cycle, as well as unique and robust. Briefly described, during training two separate dictionaries are learnt at multiple scales for the myocardium and background. In this regard, we also introduce a discriminative initialization step (discarding patches with high values in intra-class Gram matrix) to promote diversity in initialization, and a discriminative pruning step (discarding training patches with high values in inter-class Gram matrix) to further boost the discriminative abilities of the dictionaries. During testing, multiscale sparse features are used.

The main contributions of the paper are twofold. First, we experimentally demonstrate that BOLD contrast significantly affects the accuracy of segmentation algorithms (including segmentation via registration of an atlas, level sets, supervised classifier-based and other dictionary-based methods) which instead perform well in standard CINE MR. Second, to address our hypothesis we design a set of compact features using Multi-Scale Discriminative Dictionary Learning, which can effectively represent the myocardium in CP-BOLD MR. The method has been evaluated on canine subjects, which makes the problem even more challenging (lower accuracy is expected) due to the smaller size of myocardium. The remainder of the paper is organized as follows: Sect. 2 discusses related work, Sect. 3 presents the proposed method, whereas results are described in Sect. 4. Finally, Sect. 5 offers discussions and conclusion.

2 Related Work

Automated myocardial segmentation for standard CINE MR is a well studied problem [10]. Most of these algorithms can be broadly classified into three categories based on whether the methodology is segmentation-only, level set or Atlas-based segmentation with inherent registration. Recently, Atlas-based segmentation techniques have received significant attention. The myocardial segmentation masks available from other subject(s) are generally propagated to unseen data in Atlas-based techniques [2] using non-rigid registration algorithms, e.g., diffeomorphic demons (dDemons) [15], FFD-MI [5] or probabilistic label fusion [2]. Level set class of techniques uses a non-parametric way for segmenting myocardium with weak prior knowledge [3,7].

Segmentation-only class of techniques mainly focuses on feature-based representation of the myocardium. Texture information is generally considered as an effective feature representation of the myocardium for standard CINE MR images [17]. The patch-based static discriminative dictionary learning technique (SJTAD) [11] and Multi-scale Appearance Dictionary Learning technique [6] have achieved high accuracy and are considered as state-of-the-art mechanisms for supervised learning of discernible myocardial features from previously segmented data. In this paper, we follow the segmentation-only approach with the major feature of considering multi-scale appearance and texture information as the input of a discriminative dictionary learning procedure.

3 Method

General image segmentation strategies are developed on the assumption that both appearance and shape do not vary considerably across the images of a given sequence. Cardiac motion affects the shape invariance assumption, and varying CP-BOLD signal intensities violate the appearance invariance assumption as well. To overcome this issue, dictionary learning techniques can be leveraged to learn better representative features. To this end, we propose a Multi-Scale Discriminative Dictionary Learning (MSDDL) method (detailed in Algorithm 1). The features learnt via dictionary learning are tested in a rudimentary classification scheme solely for the purpose of comparing to other methods.

Feature generation with Multi-scale Discriminative Dictionary Learning (MSDDL): Given some sequences of training images and corresponding ground truth labels (i.e. masks), we can obtain two sets of matrices, $\{Y_k^B\}_{k=1}^K$ and $\{Y_k^M\}_{k=1}^K$, where the matrix Y_k^B contains the background information at a particular scale k (each scale is characterized by a different patch size), and Y_k^M is the corresponding matrix referring to the myocardium. Information is collected from image patches: squared patches are sampled around each pixel of the training images. More precisely, the i-th column of the matrix Y_k^B (and similarly for the matrix Y_k^M) is obtained by concatenating the normalized patch vector of pixel intensities at scale k, taken around the i-th pixel in the background, along with the Gabor and HOG features of the same patch. Our MSDDL method takes as input these two sets of training matrices, to learn, at each scale k, two dictionaries, D_k^B and D_k^M, and two sparse feature matrices, X_k^B and X_k^M. E.g. , the i-th column of the matrix X_k^B, $x_{k,i}^B$, is considered as the discriminative feature vector for the particular pixel corresponding to the i-th column in Y_j^B. Dictionaries and sparse features are trained via the well known K-SVD algorithm [1]. One main modification to K-SVD is the use of the "intra-class Gram matrix" to promote diversity in the initialization step. The idea is to have a subset of patches as much diverse as possible to train dictionaries and sparse features. For a given class considered (let us say background) and a given scale k, we can define the intra-class Gram matrix as $G_k^B = (Y_k^B)^T Y_k^B$. To ensure a proper discriminative initialization, patches that correspond to high values in the Gram matrix are discarded from the training before performing K-SVD. Notably, we sort the training patches w.r.t. the sum of their related coefficients in the Gram Matrix, and we prune them by choosing a certain percentage.

A second proposed modification relates to a pruning step, which is performed after K-SVD. In this case, at each scale k, an "inter-class Gram matrix" is computed ($G_k^{BM} = (D_k^B)^T D_k^M$): the atoms of each dictionary are sorted according to their cumulative coefficients in G^{BM}, and a chosen percentage of them is discarded to ensure mutual exclusiveness between the dictionaries of the two different classes. The philosophy behind this operation is similar to the one of the discriminative dictionary learning algorithm proposed in [8], where the norm of the inter-class Gram matrix appears in the optimization formulation as a constraint to be minimized. By pruning the undesired dictionary atoms all at one time, we actually adopt a greedier and low-complexity approach to the same problem. Moreover, we believe that, instead of globally minimizing the Gram matrix norm, directly removing the most "problematic" patches, which create ambiguity between background and myocardium, is more effective in our case.

Algorithm 1. Multi-scale Discriminative Dictionary Learning (MSDDL)

Input: Multi-scale training patches for background and the myocardium: $\{Y_k^B\}_{k=1}^K$ and $\{Y_k^M\}_{k=1}^K$

Output: Multi-scale dictionaries for background and the myocardium: $\{D_k^B\}_{k=1}^K$ and $\{D_k^M\}_{k=1}^K$

for $k = 1...K$ **do**
  for C={B,M} **do**
    Evaluate Y_k^C
    Compute the intra-class Gram matrix G_k^C
    Discard atoms with high values in G_k^C
    Learn dictionary and sparse feature matrix with the K-SVD algorithm

$$\underset{D_k^C, X_k^C}{\text{minimize}} \|Y_k^C - D_k^C X_k^C\|_2^2 \quad \text{s. t.} \quad \|x_{k,i}^C\|_0 \leq L$$

  Compute the inter-class Gram matrix G_k^{BM}
  Discard from D_k^B and D_k^M atoms with high values in G_k^{BM}

Building a Rudimentary Classifier for Segmentation: When considering the same patch-based approach in a segmentation problem, we have a set of test matrices $\{\hat{Y}_k\}_{k=1}^K$, obtained by sampling patches at multiple scales from the test image, and concatenating intensity values of these patches, along with Gabor and HOG features. The goal is to assign to each pixel of the test image a label, i.e. establish if the pixel is included in the background or the myocardial region. To perform this classification, we use the multi-scale dictionaries, $\{D_k^B\}_{k=1}^K$ and $\{D_k^M\}_{k=1}^K$, previously learnt with MSDDL. The Orthogonal Matching Pursuit (OMP) algorithm [12] is used to compute, at each scale k, the two sparse feature matrices $\hat{X}_k^B$ and $\hat{X}_k^M$. A certain patch at scale k, $\hat{y}_{k,i}$ will be assigned to the class that gives the smallest dictionary approximation error. More precisely, if $\|\hat{y}_{k,i} - D_k^B \hat{y}_{k,i}^B\|_2$ is larger than $\|\hat{y}_{k,i} - D_k^M \hat{y}_{k,i}^M\|_2$, at scale k the patch is assigned to the background; otherwise, it is considered belonging to the myocardial region. In this study, we employed a simple majority voting across all scales to obtain the final classification for each pixel of the test image.

4 Results

This section offers a qualitative and quantitative assessment of our proposed method w.r.t. state-of-the-art methods, to demonstrate its effectiveness for myocardial segmentation. It is particularly important to note that our method significantly outperforms all methods from current literature in both baseline and ischemia cases of CP-BOLD MR, whereas yields state-of-the-art results for both baseline and ischemia cases of standard CINE MR.

4.1 Data Preparation and Parameter Settings

2D short-axis images of the whole cardiac cycle were acquired at baseline and severe ischemia (inflicted as stenosis of the left-anterior descending coronary artery (LAD)) on a 1.5T Espree (Siemens Healthcare) in the same 10 canines along mid ventricle using both standard CINE and a flow and motion compensated CP-BOLD acquisition within few minutes of each other. All quantitative experiments are performed in a strict leave-one-subject-out cross-validation setting.

As for the parameters of MSDDL, in this paper we have empirically chosen a dictionary of 1000 atoms for foreground and background respectively, a sparsity of 4, a number of scales $K = 3$, and $9 \times 9, 11 \times 11$ and 13×13 as patch sizes. We tested the parameter sensitivity within a reasonable range, but a detailed performance chart is beyond the scope of this paper.

4.2 Visual Comparison of the Discriminativeness of the Learnt Dictionaries and Features

The feature patches learnt by MSDDL are discriminative enough for representing the myocardium separately from the background. In particular a set of feature patches of size 11×11 (without HOG and Gabor) learnt for the myocardium and background are shown in Fig. 2 to illustrate the discriminativeness of the learnt feature patches.

The motivations behind choosing each step of the proposed MSDDL strategy and the effectiveness of the features learnt by this technique are highlighted in Fig. 3, where the Cosine Similarity metric [4] is used to determine the most similar patches to a given patch in the MSDDL feature space. When selecting a patch inside the myocardium, without texture and Gram filtering, similar patches are found outside the myocardium too. Adding texture improves somewhat localization, but when considering also Gram filtering, the discriminative strengths of the approach are more evident, since few similar patches are found only within the myocardium. Similar observations hold also for the case of images from standard CINE as well (not shown for brevity).

4.3 Quantitative Comparison

As segmentation quality metric, the Dice coefficient, which measures the overlap between ground truth segmentation masks and those obtained by the

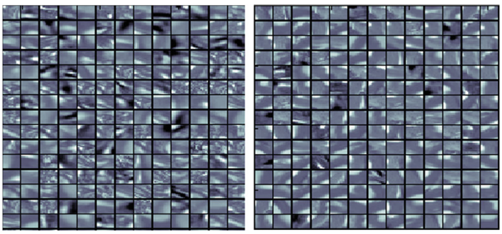

Fig. 2. Exemplar set of dictionary atoms (without HOG and Gabor) for Background (left) and Myocardium (right) learnt from patches of size 11×11 on CP-BOLD MR.

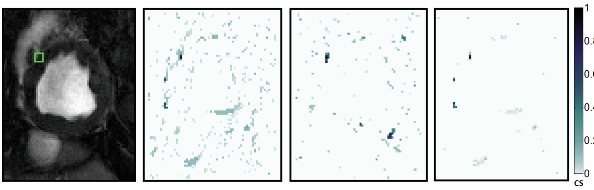

Fig. 3. Cosine Similarity (CS) between the learnt features showing the advantage of adding texture and Gram filtering. Test patch denoted by a green square in the raw image (first column), MSDDL only on appearance (second column), with texture (third column), and with proposed Gram filtering (final column) (Color figure online).

algorithm(s), is employed. For our implementation of Atlas-based segmentation methods, the registration algorithms dDemons [15] and FFD-MI [5] were used to propagate the segmentation mask of the end-diastole image from all other subjects to the end-diastole image of the test subject, followed by a majority voting to obtain the final myocardial segmentation. For level-set class of methods, a hybrid approach of [3] for endocardium and [7] for epicardium is used. For supervised classifier-based methods, namely Appearance Classification using Random Forest (ACRF) and Texture-Appearance Classification using Random Forest (TACRF) we used random forests as classifiers to get segmentation labels from different features. To provide more context we compare our approach with dictionary-based methods, SJTAD and RDDL. SJTAD is an implementation of the method in [11], whereas for RDDL we used the discriminative dictionary learning of [8] within the same classification framework that we described in Sect. 3. Finally to showcase the strengths of our design choices we considered two additional variants of MSDDL, one without Gram filtering (MSDDL No GF) and one without texture information as well (MSDDL No GF No Texture). Note that the former is similar to [6] without level-set refinement.

As Table 1 shows, overall, when standard CINE acquisition is used, most algorithms perform adequately and the presence of ischemia slightly reduces performance. However, when BOLD contrast is present, other approaches fail to accommodate changes in appearance due to contrast, but MSDDL obtains consistent performance. Specifically, Atlas-based methods are shown to perform well in standard CINE cases but poorly in CP-BOLD. ACRF and TACRF, instead, show very low performance in both standard CINE MR and CP-BOLD MR. Among dictionary-based techniques, SJTAD performs well in standard CINE MR, but underperforms in CP-BOLD MR. Our MSDDL method outperforms all approaches. When comparing it with its variants, it shows that both texture and appearance are important and that the pruning steps based on the Gram matrix are extremely beneficial. Even when we replaced our dictionary learning algorithm with RDDL, an algorithm that forces discrimination by explicitly penalizing the inter-class Gram matrix norm, the results are unimpressive. These findings are also statistically significant using a paired t-test between the results of MSDDL and the second-best performing one, i.e. SJTAD [11]. For both baseline

Table 1. Dice coefficient (mean(std)) for segmentation accuracy in %.

Methods	Baseline		Ischemia	
	Standard CINE	CP-BOLD	Standard CINE	CP-BOLD
Atlas-based methods				
dDemons [15]	60(8)	55(8)	56(6)	49(7)
FFD-MI [5]	60(3)	54(8)	54(8)	45(6)
Level set-based methods				
CVL [3,7]	50(8)	43(11)	45(9)	37(10)
Supervised classifier-based methods				
ACRF	57(3)	25(2)	52(3)	21(2)
TACRF	65(2)	29(3)	59(1)	24(2)
Dictionary-based methods				
SJTAD [11]	71(2)	32(3)	66(3)	23(4)
RDDL [8]	42(15)	50(20)	48(13)	61(12)
MSDDL No GF No Texture	52(8)	51(7)	45(4)	51(6)
MSDDL No GF	62(5)	52(4)	53(5)	57(7)
MSDDL	75(3)†	75(2)⋆	72(2)‡	71(2)⋆

and ischemia cases of CP-BOLD MR, MSDDL shows improved performance compared to SJTAD ($\star, p < 0.001$). In the case of standard CINE MR although differences appear small they are still statistically significant, i.e. ($\dagger, p < 0.05$) and ($\ddagger, p < 0.01$) for baseline and ischemia respectively.

5 Discussions and Conclusion

Rethinking the assumptions underlying the design of analysis algorithms for standard CINE MR is critical for successfully developing the appropriate analytical tools necessary to meet the new challenges posed by myocardial CP-BOLD MR. In particular, this study pin-pointed the challenges the BOLD effect poses on these assumptions made when segmenting the myocardium and quantitatively analyzed the adverse effect on algorithmic performance. In addition, in this study we showed that by learning appropriate features to best represent texture and appearance in CP-BOLD, it is possible to improve the performance of automated algorithms for myocardial segmentation. This study also showed overall low performance of state-of-the-algorithms even for standard CINE MR in canine subjects, which can be attributed to the small size of the myocardium. The proposed algorithm does not exploit the temporal information across cardiac phases and doing so should increase performance in future extensions. Finally, such post-processing tools are expected to be instrumental in advancing the utility of cardiac CP-BOLD MR towards effective clinical translation.

Acknowledgments. This work was supported by the National Institutes of Health under Grant 2R01HL091989-05.

References

1. Aharon, M., et al.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE TSP **54**(11), 4311–4322 (2006)
2. Bai, W., et al.: A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. IEEE TMI **32**(7), 1302–1315 (2013)
3. Chan, T.F., et al.: Active contours without edges. IEEE TIP **10**(2), 266–277 (2001)
4. Chang, L.-H., et al.: Achievable angles between two compressed sparse vectors under norm/distance constraints imposed by the restricted isometry property: a plane geometry approach. IEEE T Inf. Theory **59**(4), 2059–2081 (2013)
5. Glocker, B., et al.: Dense image registration through MRFs and efficient linear programming. MIA **12**(6), 731–741 (2008)
6. Huang, X., et al.: Contour tracking in echocardiographic sequences via sparse representation and dictionary learning. MIA **18**, 253–271 (2014)
7. Li, C., et al.: Distance regularized level set evolution and its application to image segmentation. IEEE TIP **19**(12), 3243–3254 (2010)
8. Ramirez, I., et al.: Classification and clustering via dictionary learning with structured incoherence and shared features. In: IEEE CVPR, pp. 3501–3508 (2010)
9. Rusu, C., et al.: Synthetic generation of myocardial bloodoxygen-level-dependent MRI time series via structural sparse decomposition modeling. IEEE TMI **7**(33), 1422–1433 (2014)
10. Tavakoli, V., et al.: A survey of shape-based registration and segmentation techniques for cardiac images. CVIU **117**, 966–989 (2013)
11. Tong, T., et al.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage **76**, 11–23 (2013)
12. Tropp, J.A., et al.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE T Inf. Theory **53**(12), 4655–4666 (2007)
13. Tsaftaris, S.A., et al.: A dynamic programming solution to tracking and elastically matching left ventricular walls in cardiac CINE MRI. In: IEEE ICIP, pp. 2980–2983 (2008)
14. Tsaftaris, S.A., et al.: Detecting myocardial ischemia at rest with cardiac phaseresolved blood oxygen leveldependent cardiovascular magnetic resonance. Circ.: Cardiovasc. Imaging **6**(2), 311–319 (2013)
15. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Non-parametric diffeomorphic image registration with the demons algorithm. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 319–326. Springer, Heidelberg (2007)
16. Wright, J., et al.: Sparse representation for computer vision and pattern recognition. Proc. IEEE **98**(6), 1031–1044 (2010)
17. Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S.: Direct estimation of cardiac bi-ventricular volumes with regression forests. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 586–593. Springer, Heidelberg (2014)