Dictionary Learning Based Image Descriptor for Myocardial Registration of CP-BOLD MR

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

[1]IMT Institute for Advanced Studies Lucca, [2]Biomedical Imaging Research Institute, Cedars-Sinai Medical, Los Angeles, [3]Department of Electrical Engineering and Computer Science, Northwestern University, Evanston

Abstract. Cardiac Phase-resolved Blood Oxygen-Level-Dependent (CP-BOLD) MRI is a new contrast agent- and stress-free imaging technique for the assessment of myocardial ischemia at rest. The precise registration among the cardiac phases in this cine type acquisition is essential for automating the analysis of images of this technique, since it can potentially lead to better specificity of ischemia detection. However, inconsistency in myocardial intensity patterns and the changes in myocardial shape due to the heart's motion lead to low registration performance for state-of-the-art methods. This low accuracy can be explained by the lack of distinguishable features in CP-BOLD and inappropriate metric definitions in current intensity-based registration frameworks. In this paper, the sparse representations, which are defined by a discriminative dictionary learning approach for source and target images, are used to improve myocardial registration. This method combines appearance with Gabor and HOG features in a dictionary learning framework to sparsely represent features in a low dimensional space. The sum of squared differences of these distinctive sparse representations are used to define a similarity term in the registration framework. The proposed descriptor is validated on a challenging dataset of CP-BOLD MR and standard CINE MR acquired in baseline and ischemic condition across 10 canines.

Keywords: Registration, Dictionary Learning, Similarity Metric, CP-BOLD MR, CINE MR.

1 Introduction

Nonrigid image registration is an essential step in medical imaging, including automatic segmentation, motion tracking and morphometric analysis [13]. However, since most of the proposed algorithms rely on a (dis)similarity metric build based on the assumptions of consistent intensity and local shape, images with pathologies and locally varying intensity may not be accurately aligned. One example of that is the registration of image sequences of Cardiac Phase-resolved Blood Oxygen-Level-Dependent (CP-BOLD) MR. CP-BOLD MR is a truly non-invasive method for early diagnosis of an ongoing ischemia, observing changes

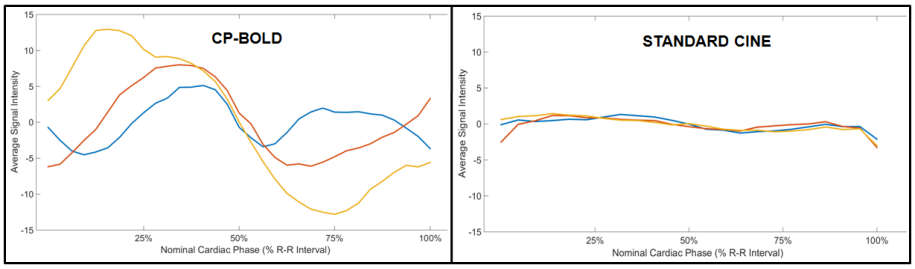

Fig. 1: Exemplary plots of time series extracted from the same corresponding regions in the same subject under baseline (absence of disease) conditions using CP-BOLD MR and standard Cine. Observe how in CP-BOLD, intensity varies with cardiac phase, but in standard CINE MR this variation is minimal.

in myocardial signal intensity patterns as a function of cardiac phase [17]. As Figure 1 illustrates, time series of intensity vary as a function of cardiac phase when BOLD effect is present –it appears maximal in systole and minimal in diastole. In disease this effect is not present. However, visualizing and quantifying such patterns requires significant post-processing, including myocardial registration to establish pixel-precise time series for identifying such patterns [12]. Such spatio-temporal intensity variations of the myocardial BOLD effect cause the methods developed for standard CINE MR registration to under-perform. Thus, in CP-BOLD in addition to violations of shape invariance (as with standard CINE MR) the principal assumption of appearance invariance (consistent intensity) is violated as well.

As a result, no CP-BOLD MR myocardial registration algorithms exist and due to this absence either segmental information [11] or synthetic data sets are used [12], to obtain pixel-wise time series. We assume that it is due to lack of proper similarity criteria. Rather than relying on low-level features used often for myocardial registration of standard CINE MR, a more distinguishing descriptor should be developed to accommodate the BOLD effect.

In this paper, we propose a feature-based descriptor as a similarity measure of the alignment to register the myocardium in the entire cardiac sequence of CP-BOLD. We adopt a patch-based discriminative dictionary learning technique [4] to learn features from data. Our motivation is to employ a compact and high-fidelity low-dimensional subspace representation, which is able to extract semantic information of the myocardium pixels [5]. We observe that although the patch intensity level varies significantly across the cardiac cycle, sparse representations based on learnt dictionaries are invariant, as well as unique and robust. The discriminative dictionary learning strategy is designed to facilitate this key observation regarding CP-BOLD.

During training, two dictionaries of patches for myocardium and background are learnt offline. To register two images with unknown myocardium masks, the sparse representations that are obtained on the basis of previously trained dictionaries for background and myocardium, are concatenated and considered

as the feature for that particular pixel. The similarity term evaluates the match of the sparse features at every iteration on a pixel level. The sum of squared differences of the sparse representations between the target image and warped source image are utilized as similarity criteria.

There are three major contributions of the paper. First of all, we propose a sparse representation-based image descriptor in a registration framework, for the first time to the best of our knowledge. Second, we experimentally validate the fact that BOLD contrast significantly affects the accuracy of registration algorithms (including intensity-based and feature-based methods), which instead perform well in standard CINE MR. Finally, we address the fundamental problem in handling BOLD contrast by designing a set of compact features using discriminative dictionary learning, which can effectively represent the myocardium in CP-BOLD MR.

The remainder of the paper is organized as follows: Section 2 discusses the background and Section 3 presents the method. Based on our experimental results (Section 4), we draw conclusions in Section 5.

2 Background

Automated myocardial registration for standard CINE MR is a well studied problem [15]. Most of these algorithms can be classified into two groups according to similarity criteria used: intensity-based or feature-based. General intensity-based registration algorithms can be summarized as an energy minimization procedure, where the energy functional is [13]:

$$E = \sum_{p \in \Omega} DS(I_s(p), I_t(p+u)) + \lambda E_R, \tag{1}$$

where Ω represents the entire image domain, and p denotes a pixel in the domain. Non-rigid registration consists of minimizing a dissimilarity measure DS between a source image I_s and a target image I_t, u denotes the displacements and E_R denotes the regularization term. In this work, we are particularly interested in the definition of the similarity measure. Sum of squared differences (SSD) and cross correlation (CC), are the earlier metrics utilized in registration. Recently, information theory-based approaches gained attention, e.g., derivatives of Mutual Information (MI), which is based on individual and joint gray level distributions [8].

When registration under inhomogeneity conditions is required, some have proposed modifications on regional intensity distortions (for brain MRI) [14] or spatially intensity variations [19]. Alternatively, feature-based approaches can be used. A recent example is DRAMMS [6], where the similarity is based on optimal Gabor attributes. Another approach, MIND [3], relies on regional information following the footsteps of self-similarity (a method utilized for image denoising) for multi-modality registration.

In this paper, we concentrate on developing a feature-based metric but also learning features instead of using fixed ones. We use sparse representation coefficients of patches, generated by a dictionary trained offline, to define a similarity

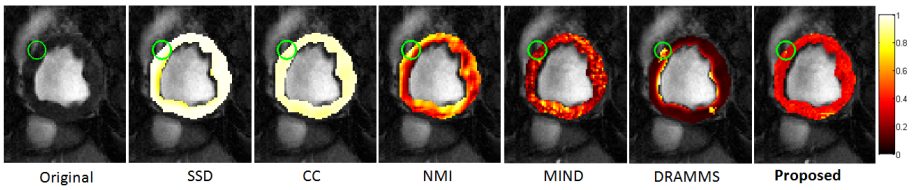

Fig. 2: Similarity of patches in two consecutive images. First image shows the test patch (green circle) and the remainder shows responses of each similarity metric inside the myocardium. All metrics are normalized and dissimilarity metrics are inverted.

measure of alignment. In this study, we compare our method with SSD and MI based Free Form Deformations (FFD) [10], optical flow based diffeomorphic demons (ddemons) [18] and symmetric diffeomorphic transformation with CC metric implemented in Advanced Normalization Toolkit (ANTs) [2]. To demonstrate that our proposed approach provides better localized matches, Figure 2 shows the values of matches using several criteria when taking a patch from one image and matching it to myocardial locations in another image.

3 Method

We leverage dictionary learning techniques to learn better representative features. Accordingly, we integrate a Dictionary Learning-based Image Descriptor (DLID) derived from training patches into a similarity term of our proposed registration framework. Features learnt via dictionary learning are used in an image registration framework to evaluate the performance of the proposed descriptor.

3.1 Using learnt Features in a Registration Framework

When registering a cardiac sequence $I_1, \ldots, I_t$, we aim to find a deformation that can register each image in the sequence to the first one. Here following the formulation of equation 1, we adopt a regularization in the form of

$$\underset{u}{\operatorname{argmin}} \sum_{p \in \Omega} S(I_1(p), I_t(p+u))^2 + \lambda \operatorname{tr}(\nabla u(p)^T \nabla u(p))^2, \tag{2}$$

where ∇u denotes the gradient of the displacement field. This function is minimized over u with Gauss-Newton optimization as described in [3].

We propose an appropriate similarity term S based on the sparse feature representation of image patches. Assuming two input images, considering I_1 as fixed and I_t as moving, we extract for each pixel location in both, patches, which we represent with appearance and texture features (HoG, Gabor). We create a sparse representation $\hat{X}_p$ for each pixel location for the two images to be registered. The Orthogonal Matching Pursuit (OMP) algorithm [16] is used to compute, two sparse feature matrices $\hat{X}^B$ and $\hat{X}^M$, both with n dimensions, based on previously computed dictionaries D^B and D^M (detailed below). At

Algorithm 1 Dictionary Learning

Input: Training patches for background and myocardium: Y^B and Y^M
Output: Dictionaries for background and myocardium: D^B and D^M

1: **for** C={B,M} **do**
2: Find intra-class Gram matrix G^C and discard atoms with high values
3: Learn dictionary and sparse feature matrix with the K-SVD algorithm

$$\underset{D^C, X^C}{\text{minimize}} \|Y^C - D^C X^C\|_2^2, \text{ subject to } \|x_i^C\|_0 \leq s$$

4: **end for**
5: Compute inter-class Gram matrix G^{BM}
6: Discard from D^B and D^M atoms with high values in G^{BM}

a certain pixel p of the image, a concatenation of these sparse representation vectors $\hat{X}_p = [\hat{X}^B; \hat{X}^M]$ are used to represent the image instead of the pixel level definitions. The proposed similarity term S at pixel p is defined as the ℓ^1 norm of the difference vector between the sparse representations of the warped source image and the target image as shown in equation 3.

$$S(I_1(p), I_t(p+u)) = \| \hat{X}_p^1 - \hat{X}_{p+u}^t \|_1 \tag{3}$$

3.2 Feature generation with Discriminative Dictionary Learning

Given some training images (e.g., sequences in the context of cine (BOLD) MRI) and corresponding ground truth labels (i.e., myocardial masks), we obtain two sets of matrices, Y^B and Y^M, where the matrix Y^B contains background information, and Y^M contains information of patches within the myocardium. Information is collected from image patches: $K \times K$ squared patches are sampled around each pixel in the training images. More precisely, the i-th column of the matrix Y^B (and similarly for the matrix Y^M) is obtained by concatenating the normalized patch vector of pixel intensities, taken around the i-th pixel in the background (or myocardium), along with Gabor and HOG features of the same patch. The dictionary learning method takes as input these two sets of training matrices, to learn, two dictionaries, D^B and D^M, with n number of atoms, and two sparse feature matrices, X^B and X^M, with sparsity s. The i-th column of the matrix X^B, x_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], in an optimization problem shown in Algorithm 1. During initialization we first find the "intra-class Gram matrix" to promote diversity. 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) we can define the intra-class Gram matrix as $G^B = (Y^B)^T Y^B$. 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, and K-SVD is initialized obtaining a random set of patches as initial atoms.

We also use pruning, inspired as a greedy approach of [9], which is performed after K-SVD to remove undesired (similar to other) atoms from each dictionary trained. In this case, an "inter-class Gram matrix" between dictionaries is computed ($G^{BM} = (D^B)^T D^M$), 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 (and better discrimination) between the different dictionaries. These modifications ensure that patches of different origin will have different support and that similar atoms are excluded.

4 Results

This section describes qualitatively and compares quantitatively our proposed dictionary learning-based descriptor with state-of-the-art approaches.

Data Preparation and Parameter Settings: 2D short-axis images of the whole cardiac cycle (2D+time, cine) 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 the mid ventricle using both standard CINE and a flow and motion compensated CP-BOLD acquisition within few minutes of each other [17]. All quantitative experiments are performed in a strict leave-one-subject-out cross-validation. Parameters and settings were optimized for each method used in comparison. For DLID, in this paper we have empirically chosen a dictionary of $n = 1000$ atoms for foreground and background respectively, a sparsity of $s = 4$, and as patch size K=9. The regularization weight (λ) is set to 0.8 to ensure smooth deformations.

Visual Evaluation: In an example sequence, we register each image in the sequence throughout the cardiac cycle to the first image using our approach. We take two orthogonal short axis profiles that intersect approximately at the center of the Left Ventricle, and in Figure 3 we show the temporal evolution of the profiles with and without registration (left-most and right-most horizontal and vertical profile, respectively). The proposed shows clearly defined structure and the ability to correct for cardiac motion. Notice that BOLD intensity variation is subtle and not perceptible in these images (ie., is not a global change).

Quantitative Comparison: Using again the same process, in a strict-leave-one-out fashion we want to investigate the effect of different similarity metrics in recovering cardiac motion. To evaluate performance, we use again manual delineations of the myocardium provided by experts, and train dictionaries on a set of images and test on one subject. For validation, via segmentation, the myocardial mask from the source image was propagated to the target using the deformation field found with the algorithms, and its overlap with the ground truth mask of the fixed is measured using the Dice overlap metric [7]. Note that these masks are unknown to the algorithms and are used only for comparison.

Our findings in Table 1, show that using discriminative features and our similarity term significantly improve the performance for CP-BOLD cardiac se-

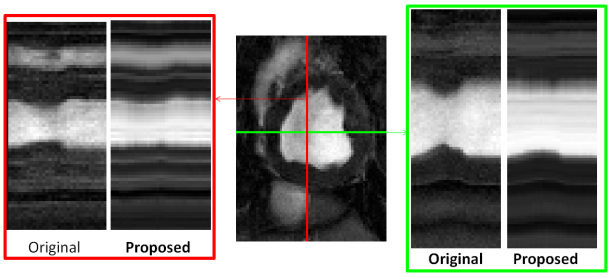

Fig. 3: Temporal evolution of two orthogonal short axis profiles (red and green line) intersecting approximately at the center of the left ventricle, without registration (original) and with registering every image in the sequence with the first image (proposed).

Table 1: Dice overlap comparison for different similarity metrics

	Baseline		Ischemia	
Methods	Standard Cine	CP-BOLD	Standard Cine	CP-BOLD
ANTs (CC metric) [2]	0.60 ∓ 0.11	0.55 ∓ 0.10	0.55 ∓ 0.15	0.51 ∓ 0.12
dDemons [18]	0.59 ∓ 0.11	0.51 ∓ 0.16	0.58 ∓ 0.13	0.45 ∓ 0.13
DRAMMS [6]	0.67 ∓ 0.09	0.61 ∓ 0.07	0.59 ∓ 0.10	0.54 ∓ 0.06
FFD-SSD [10]	0.49 ∓ 0.07	0.45 ∓ 0.16	0.48 ∓ 0.14	0.39 ∓ 0.13
FFD-MI [10]	0.54 ∓ 0.12	0.48 ∓ 0.08	0.53 ∓ 0.06	0.38 ∓ 0.07
MIND [3]	0.62 ∓ 0.07	0.62 ∓ 0.12	0.61 ∓ 0.15	0.53 ∓ 0.09
Proposed without sparsity	0.55 ∓ 0.08	0.52 ∓ 0.11	0.45 ∓ 0.09	0.42 ∓ 0.12
Proposed	0.63 ∓ 0.07	0.66 ∓ 0.09	0.58 ∓ 0.07	0.60 ∓ 0.13

quence registration either under baseline or ischemia conditions w.r.t. other approaches. To highlight the unique challenge of BOLD, we also include results based on standard CINE. Our proposed method, although not its main focus, performs as good as other algorithms even in this case. To emphasize the importance of sparsity and learning we also use directly the ℓ^2 norm between input patches, instead of spare representations. Lower performance in ischemia for all algorithms could be attributed to changes in myocardial contractility.

5 Discussions and Conclusion

We propose a new dictionary learning-based image descriptor (DLID) for myocardial registration. The experiments clearly underline the need for a new representation in image registration. Their integration into analytical tools are necessary to meet new challenges posed by myocardial CP-BOLD MR. In particular, this study pin-pointed the challenges the BOLD effect poses on common assumptions made when registering the myocardium and quantitatively analyzed the performance of the descriptor both under baseline and ischemia conditions. Moreover, in this study we showed that by learning appropriate features to best represent texture and appearance in CP-BOLD, it is possible to obtain better correspondences for the entire cardiac sequence. The proposed method can be

utilized for other challenges, where spatio-temporal intensity as a biomarker of disease, especially in the presence of motion. One limitation is computational time, since calculating sparse representations is the bottleneck of the problem. The successful application of this post-processing tools are foreseen to be critical in the clinical translation of cardiac CP-BOLD MR.

References

1. Aharon et al., K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, IEEE TSP, 54 (11): 4311-4322, 2006.
2. Avants, et al., Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, MIA 12(1), 26-41, 2008.
3. Heinrich et al., MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration, MIA, 16(7), 1423-1435, 2012.
4. Huang et al., Contour tracking in echocardiographic sequences via sparse representation and dictionary learning, MIA, 18: 253-271, 2014.
5. Mukhopadhyay et al., Data-Driven Feature Learning for Myocardial Segmentation of CP-BOLD MRI. FIMH, 2015
6. Ou et al., DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting, MIA 15(4), 622-639, 2011.
7. Ou et al., Validation of DRAMMS among 12 popular methods in cross-subject cardiac MRI registration.,WBIR, 209-219, 2012.
8. Pluim et al.,Mutual-information-based registration of medical images: a survey, TMI 22(8), 986-1004, 2003.
9. Ramirez et al., Classification and clustering via dictionary learning with structured incoherence and shared features. IEEE CVPR, 3501-3508, 2010.
10. Rueckert et al., Nonrigid registration using free-form deformations: application to breast MR images, TMI 18(8), 712-721, 1999.
11. Rusu et al., Structured Dictionaries for Ischemia Estimation in Cardiac BOLD MRI at Rest,MICCAI 562-569, 2014.
12. Rusu et al., Synthetic generation of myocardial bloodoxygen-level-dependent MRI time series via structural sparse decomposition modeling, IEEE TMI, 7 (33), 1422-1433, 2014.
13. Sotiras et al., Deformable Medical Image Registration: A Survey Medical Imaging, IEEE TMI, 7 (32), 1153-1190, 2013.
14. Studholme et al., Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change, IEEE TMI, 5 (25), 626-639, 2006.
15. Tavakoli et al., A Survey of shape-based registration and segmentation techniques for cardiac images, CVIU, 117, 966-989, 2013.
16. Tropp et al., Signal recovery from random measurements via orthogonal matching pursuit. IEEE T Information Theory, 53(12), 4655-4666, 2007.
17. Tsaftaris et al., Detecting myocardial ischemia at rest with cardiac phaseresolved blood oxygen leveldependent cardiovascular magnetic resonance. Circulation: Cardiovascular Imaging, 6(2), 311-319, 2013.
18. Vercauteren et al., Non-parametric diffeomorphic image registration with the demons algorithm, MICCAI, 319-326, 2007.
19. Zhuang et al., A nonrigid registration framework using spatially encoded mutual information and free-form deformations, IEEE TMI,10 (30), 1819-1828, 2011.