Multi-Atlas Segmentation with Augmented Features for Cardiac MR Images

Wenjia Bai[a,*], Wenzhe Shi[a], Christian Ledig[a], Daniel Rueckert[a]

[a]*Biomedical Image Analysis Group, Department of Computing, Imperial College London, United Kingdom*

Abstract

Multi-atlas segmentation infers the target image segmentation by combining prior anatomical knowledge encoded in multiple atlases. It has been quite successfully applied to medical image segmentation in the recent years, resulting in highly accurate and robust segmentation for many anatomical structures. However, to guide the label fusion process, most existing multi-atlas segmentation methods only utilise the intensity information within a small patch during the label fusion process and may neglect other useful information such as gradient and contextual information (the appearance of surrounding regions). This paper proposes to combine the intensity, gradient and contextual information into an augmented feature vector and incorporate it into multi-atlas segmentation. Also, it explores the alternative to the K nearest neighbour (KNN) classifier in performing multi-atlas label fusion, by using the support vector machine (SVM) for label fusion instead. Experimental results on a short-axis cardiac MR data set of 83 subjects have demonstrated that the accuracy of multi-atlas segmentation can be significantly improved by using the augmented feature vector. The mean Dice metric of the proposed segmentation framework is 0.81 for the left ventricular myocardium on this data set, compared to 0.79 given by the conventional multi-atlas patch-based segmentation (Coupé et al., 2011; Rousseau et al., 2011). A major contribution of this paper is that it demonstrates that the performance of non-local patch-based segmentation can be improved by using augmented features.

Keywords: multi-atlas segmentation, patch-based segmentation, cardiac image segmentation, augmented features

1. Introduction

Multi-atlas segmentation warps a number of atlas images to the target image by image registration and then infers the target segmentation from the warped atlas images and corresponding label maps. It incorporates prior knowledge and anatomical variability from the atlas population and has been successfully applied to segmentation of a variety of medical images, including brain, cardiac and abdominal images (Aljabar et al., 2009; Asman and Landman, 2013; Cardoso et al., 2013; Eskildsen et al., 2012; Heckemann et al., 2006; Hu et al., 2012; Iglesias et al., 2013; Jia et al., 2011; Lotjonen et al., 2010; van Rikxoort et al., 2010; Wang et al., 2011a; Wolz et al., 2013).

The label at a voxel in the target image is determined by fusing the labels of the corresponding voxels in the atlas label maps. One type of label fusion methods is weighted voting, where the label of each atlas voxel contributes to the final result with a weight (Artaechevarria et al., 2009; Bai et al., 2013a; Langerak et al., 2010; Išgum et al., 2009; Sabuncu et al., 2010; Wang et al., 2011b). To account for potential target-to-atlas image registration errors, a search window or volume is introduced so that a number of atlas voxels in the search volume are evaluated in label fusion, with those looking more similar to the target voxel contributing a higher weight. Many label fusion techniques determine the weight by evaluating the intensity similarity between the image patch centred at the target voxel and the image patch centred at the atlas voxel and this forms the patch-based segmentation (PB) (Coupé et al., 2011; Rousseau et al., 2011). It has been quite successful in practice since it allows the weight to vary spatially and to represent the local similarity between the target image and the atlas image. Other ways to determine the weight have also been proposed. For example, Wang et al. proposed to consider the similarity not only between the target patch and each atlas patch, but also between each pair of atlas patches in order to reduce correlated labelling errors (Wang et al., 2013). Other works proposed to determine the weight by decomposing the target patch using a sparse representation of the atlas patches (Tong et al., 2013; Wang et al., 2014; Zhang et al., 2012). Apart from weighted voting, another type of label fusion methods is statistical label fusion, such as STAPLE (Warfield et al., 2004) or non-local STAPLE (Asman and Landman, 2013). STAPLE estimates the label probability by combining the atlas label information and the rater performance matrix in an expectation-maximisation (EM) framework (Warfield et al., 2004). Non-local STAPLE further improves the statistical fusion method by incorporating the intensity information using a non-local correspondence model (Asman and Landman, 2013).

*Corresponding author. Address: Department of Computing, Huxley Building, Imperial College London, London, SW7 2AZ, United Kingdom. Telephone: +44 (0)20759 48183.

Email address: w.bai@imperial.ac.uk (Wenjia Bai)

Table 1: Comparison of multi-atlas segmentation methods in terms of features and classification algorithms being used. (Note that this is not a comprehensive review and only a few examples from the literature are listed here.) Most existing methods utilise intensity information (patch voxel intensities) only and use K nearest neighbours (KNN) for classification. This paper proposes to use the augmented features (combination of intensity, gradient and contextual features) for classification. Using the augmented features, two methods (PBAF and SVMAF) are proposed, which respectively use KNN or SVM for classification.

Feature	Classifier	Literature	Label fusion weights
None	KNN	Heckemann et al. (2006)	Equal for all samples
Intensity	KNN	Artaechevarria et al. (2009)	Patch intensity similarity (MSD/NCC/MI)
		Sabuncu et al. (2010)	Voxel intensity similarity
		Coupé et al. (2011)	Patch intensity similarity (MSD)
		Rousseau et al. (2011)	Patch intensity similarity (MSD)
		Zhang et al. (2012)	Patch sparse representation
		Wang et al. (2013)	Minimising correlated labelling errors
		Tong et al. (2013)	Patch sparse representation
Augmented	KNN	Proposed 1: PBAF	Feature vector similarity
Augmented	SVM	Proposed 2: SVMAF	Feature vector similarity

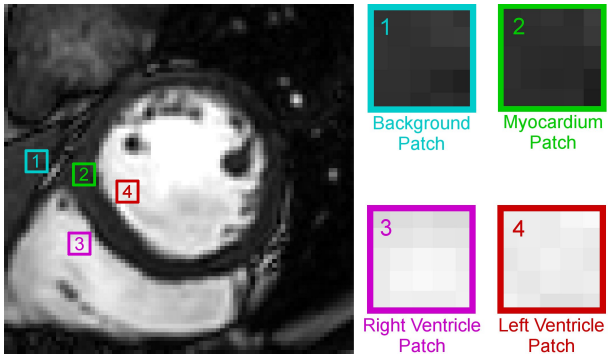


Figure 1: In a cardiac MR image, a patch in the myocardium (green) may look similar to a patch in the background (blue) in terms of voxel intensities. Similarly, a patch in the left ventricle (red) may look similar to a patch in the right ventricle (purple).

1.1. Motivation

From the perspective of machine learning, multi-atlas segmentation can be viewed as a voxel-wise classification problem. At each voxel, the testing sample is the target patch centred at this voxel and the training set consists of all the atlas patches within the search volume. Though there are many ways proposed to perform multi-atlas label fusion, most existing methods utilise patch intensity information only. The training and testing samples are both characterised by a low dimensional feature vector, i.e. the intensities of the voxels inside the patch, without considering other information. The classifier is normally a K nearest neighbour (KNN) classifier, with the weights often determined by intensity similarity between the target patch and the atlas patches. Table 1 compares some multi-atlas segmentation methods in terms of features and classification algorithms being used.

In some cases, however, patch intensities may be highly ambiguous. For example, in a cardiac MR image, a patch in the myocardium may look similar to a patch in the background though they represent different tissue classes (see Figure 1) and a patch in the left ventricle may look similar to a patch in the right ventricle. If we compute the patch similarity only based on patch voxel intensities, we might map a background patch to a myocardium patch with a high weight and this would degrade the accuracy of label fusion. In order to describe two patches more discriminatively, we can incorporate more information into the feature set, for example, by using gradient or contextual information. This means that the dimension of the feature vector is increased. With the increase of the feature space dimensionality, samples from different classes that used to be clustered together in the low dimensional space may be dispersed in the high dimensional space and it would therefore become easier to learn a classifier that can discriminate between them. This is the main motivation for incorporating gradient and contextual information and augmenting the feature vector for multi-atlas segmentation.

Most existing multi-atlas segmentation methods use KNN for voxel-wise classification and many different ways have been proposed to determine the sample weights, as listed in Table 1. Another aspect that we aim to explore in this paper is that whether KNN can be replaced by other more sophisticated classifiers in multi-atlas segmentation. Two algorithms are proposed and compared, namely patch-based segmentation with augmented features (PBAF) and multi-atlas SVM segmentation with augmented features (SVMAF). They both use the same augmented feature vector, which is aware of gradient and contextual information. The difference is that the first algorithm uses the KNN classifier, just as the conventional patch-based segmentation method (PB) does, whereas the second algorithm uses the SVM classifier.

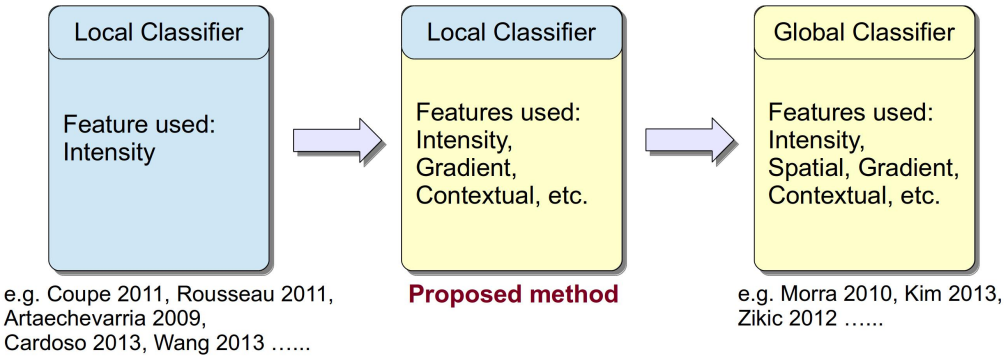


Figure 2: The proposed method is a link between segmentation methods using local classification and those using global classification.

1.2. Related Work

The proposed methods can be regarded as a link between segmentation methods using local classification and those using global classification, as illustrated in Figure 2. Conventional patch-based segmentation methods (Coupé et al., 2011; Rousseau et al., 2011) perform local classification, training a KNN classifier at each voxel. They utilise voxel intensities inside a patch as a feature to discriminate patches from different tissue classes. On the contrary, segmentation methods using a global classifier (Morra et al., 2010; Kim et al., 2013; Zikic et al., 2012) normally involve more features, including not only intensity, but also spatial, gradient and contextual features. These additional features are essential for a single global classifier to work and adapt to each local region. The proposed methods fill in the gap between the two classes of methods. It trains local classifiers at each local region and incorporates more information into the features, such as gradient and contextual features.

The proposed methods fall into the category of atlas-based segmentation. Apart from atlas-based segmentation (Zhuang et al., 2010), other methods used for cardiac image segmentation include, but are not limited to, pixel intensity classification (Jolly, 2006; Lynch et al., 2006), guide-point modelling (Young et al., 2000), active contours (Jolly, 2006), deformable model adaptation (Ecabert et al., 2008), level sets (Feng et al., 2013), active shape model (ASM) and active appearance model (AAM) (Zhang et al., 2010), random forests (Margeta et al., 2011) etc. For a comprehensive review on short-axis cardiac MR image segmentation, please refer to (Petitjean and Dacher, 2011).

The proposed methods were tested on a short-axis cardiac MR data set of 83 subjects for myocardium segmentation. Experimental results show that by incorporating more information and using extra features, the segmentation accuracy can be significantly increased, compared to the conventional non-local patch-based segmentation method which only explores intensity information. We have also found that using a more sophisticated classifier such as SVM can perform slightly better than weighted KNN. However, the SVM classifier also involves more expensive computation.

2. Methods

2.1. Framework

In multi-atlas segmentation, the target image is registered with a number of atlas images using either affine or non-rigid registrations. Using the resulting transformations, the atlas images and label maps are warped to the target image space. The segmentation, or the target label map, is then inferred from the warped atlas label maps.

For each voxel x in the target image, a feature vector f_x can be constructed, which characterises the local image features. This vector forms the testing sample in our problem. The training samples come from the warped atlas images. Feature vectors are constructed in the warped atlas images for all the voxels in the search volume $S(x)$ around x. The training set is formed of pairs of feature vector and corresponding label,

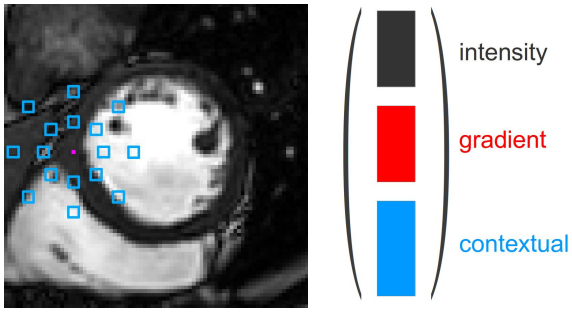
$$D_x = \{(f_{n,y}, l_{n,y}) | n = 1, \cdots, N,\ y \in S(x)\} \tag{1}$$

where n denotes the atlas index, N denotes the number of atlases, y denotes a voxel in the search volume $S(x)$, $f_{n,y}$ denotes the feature vector at voxel y in atlas n and $l_{n,y}$ denotes the corresponding label. Determining the label at x amounts to a supervised learning problem with the training set D_x and the testing sample f_x.

2.2. Augmented Feature Vector

For most existing multi-atlas segmentation methods, the feature vector f_x is formed of the voxel intensities in the image patch centred at x. The dimension of f_x is equal to the number of voxels of the patch. A major contribution of this work is that we increase the dimension of f_x to include not only intensity but also gradient and contextual features.

For the gradient feature, the gradient orientations of all the voxels in the image patch are used. The gradient orientation is defined as the image gradient normalised by its magnitude. Compared to the image gradient, the gradient orientation is less sensitive to change of luminance. It has been successfully used to recognise local patterns in computer vision (Tzimiropoulos et al., 2012).



(a) Contextual feature (b) Augmented feature

Figure 3: Contextual feature is characterised by a number of regions in the surroundings of the patch under consideration. It is defined by the intensity difference between the patch centre (purple) and the median intensity in each sample region (blue). The augmented feature vector is the concatenation of the intensity, gradient and contextual features.

The contextual feature is used to describe the relative relation of the image patch with regard to its surroundings. For each image patch, a number of regions are sampled in its surroundings, radiating from the patch centre with equal degree intervals and a certain distance, as shown in Figure 3(a). The intensity difference between the image patch centre and the sample region is computed,

$$d_{x,C} = \mu_C - I_x \tag{2}$$

where I_x denotes the intensity at the patch centre, C denotes one of the sample regions and μ_C denotes the median intensity in C. The intensity difference is a continuous description of the context. Apart from this, a compact and binary description of the context, the BRIEF descriptor (Calonder et al., 2012), is also computed,

$$b_{x,C} = \begin{cases} 1 & \text{if } I_x < \mu_C, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The contextual feature we use is formed of both the continuous and the binary descriptors for all the sample regions.

It has to be noted although a similar contextual feature has been used for image segmentation (Zikic et al., 2012; Kim et al., 2013), the key difference is that in both (Zikic et al., 2012) and (Kim et al., 2013), the contextual feature was used to train a global classifier for the whole image, whereas in the proposed method, it is used for label fusion at each voxel, which means training a local classifier at each voxel.

Finally, the augmented feature vector f_x is the concatenation of the intensity, gradient and contextual features, as shown in Figure 3(b). Since each kind of feature has a different value range, the feature vector is scaled before classifier training to avoid features with greater value ranges dominating the classification. Except for the BRIEF descriptor of the contextual feature, all the elements of the feature vector in the training set are scaled to be in the range [-1, +1]. The same scaling is applied to the feature vector of the testing sample.

2.3. Patch Pre-Selection

Similar to (Coupé et al., 2011), we perform patch preselection in order to save computation. For all the atlas patches in the search volume, the mean squared difference (MSD) of voxel intensity between the atlas patch and the target patch is computed. Atlas patches with the M lowest MSD are selected. Evaluation of the feature vector and the subsequent classification will only be performed on the set of pre-selected M patches. By doing this, we are essentially pruning the set of training samples by computing a crude distance (MSD) which is fast to compute, before computing the more expensive distance (distance between the augmented feature vectors) which involves evaluating the gradient and contextual features. This does not only save time for feature evaluation, but will also accelerate the classifier training process.

2.4. Patch-Based Segmentation with Augmented Features

In this section, we propose the patch-based segmentation algorithm with augmented features, "PBAF", which incorporates the augmented feature vector into patch-based segmentation. From the perspective of machine learning, patch-based segmentation (Coupé et al., 2011) is essentially a weighted K nearest neighbours (KNN) classifier. In conventional patch-based segmentation, the vote for each tissue class k at voxel x is determined as,

$$P_x(k) = \frac{1}{Z} \sum_{n=1}^{N} \sum_{y \in S(x)} e^{-\frac{\|f_x - f_{n,y}\|^2}{h}} \cdot \delta_{l_{n,y},k} \tag{4}$$

where $P_x(k)$ denotes the vote for class k, Z denotes the normalisation factor so that $\sum_k P_x(k) = 1$, h denotes the bandwidth for the Gaussian kernel and $\delta_{l_{n,y},k}$ denotes the Kronecker operator, which is equal to one when $l_{n,y} = k$ and equal to zero otherwise. The resulting label at voxel x is determined by the label class with the highest vote,

$$l_x = \arg \max_{k=1,2,\cdots,K} P_x(k) \tag{5}$$

where K denotes the number of label classes.

Eq. (4) is essentially a kernel density estimator, which estimates the probability to be class k at point f_x based on its neighbours $\{f_{n,y}\}$ and using the Gaussian distribution function as the kernel or the Parzen window. It follows that the label fusion process, Eq. (5), is a Bayes classifier, which determines the label by maximum likelihood estimation.

KNN has been demonstrated to work well for patch-based segmentation and it does not require any explicit training. When the target image and the atlas image is registered, many atlas feature vectors $\{f_{n,y}\}$ are close to the target feature vector f_x in the feature space. As a result, the kernel density estimator, Eq. (4), can generate a good approximation to the real probability, which leads to good performance of the maximum likelihood classification.

The proposed algorithm is similar to the conventional patch-based segmentation. However, it performs label fusion using the augmented feature vector in Eq. (4), instead of using the

patch intensity feature. By adding the gradient and contextual features, the feature vector may have a stronger discriminative power so that patches from different label classes can be separated in the high dimensional feature space. Algorithm 1 explains the procedure of the algorithm.

Algorithm 1: Patch-Based Segmentation with Augmented Features (PBAF)

for *each voxel x* **do**
  initialise the vote vector $v_x = 0$
  pre-select M atlas patches
  for *atlas patch i* **do**
    evaluate the augmented features f_i
    accumulate the vote: $v_x(l_i) = v_x(l_i) + e^{-\frac{\|f_x - f_i\|^2}{h}}$
  end
  for *class k* **do**
    $P_x(k) = \frac{v_x(k)}{\sum_k v_x(k)}$
  end
  determine the label: $l_x = \arg\max_k P_x(k)$
end

2.5. SVM Segmentation with Augmented Features

In this section, we combine the augmented feature vector with the support vector machine (SVM) classifier and propose the SVM segmentation algorithm with augmented features, "SVMAF". The motivation is that we would like to explore whether learning a more sophisticated classifier such as SVM can perform better than KNN in the scenario of multi-atlas segmentation.

A SVM classifier minimises classification error by finding optimal separating hyperplanes (Boser et al., 1992; Cortes and Vapnik, 1995; Chang and Lin, 2011). Given a training set of feature-label pairs $\{f_i, g_i\}$, where f_i denotes the feature vector and $g_i \in \{+1, -1\}$ denotes whether the sample belongs to the class or not, the cost function of the binary SVM is formulated as,

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + C \sum_i \xi_i \\ \text{subject to} \quad & g_i \cdot (w^T \phi(f_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0. \end{aligned} \tag{6}$$

where w and b denote parameters of the separating hyperplanes, ξ_i denotes the error penalty for slack variables (those located on the wrong side of the margin of the hyperplanes), C denotes the regularisation parameter which controls the trade-off between the margin and the error penalty, $\phi(f_i)$ denotes the non-linear transformation function for the input vectors. The two separating hyperplanes are respectively defined by $w^T \phi(f_i) + b = 1$ and $w^T \phi(f_i) + b = -1$. The margin is the region between the two hyperplanes, whose width is determined by $\frac{2}{\|w\|}$. By optimising the constrained problem, SVM tries to find two hyperplanes that separate the data with no points in-between and meanwhile maximise the margin between the hyperplanes.

We use the radial basis function (RBF) kernel for SVM, which is formulated as,

$$K(f_i, f_j) \equiv \phi(f_i)^T \phi(f_j) = e^{-\gamma \|f_i - f_j\|^2} \tag{7}$$

For a multi-class classification problem, in our problem K classes, $\frac{K(K-1)}{2}$ binary classifiers are constructed, each one trained to classify data between two classes. Given a new testing sample, each binary classifier will give a vote and the final result will be determined as the class with the highest number of votes.

For this algorithm, since training a SVM classifier for each voxel in the target image can be computationally prohibitive if the image size is very large, we accelerate the algorithm by placing equally spaced control points in the image space and training the SVM classifier only at each control point. At the testing stage, each target voxel is classified by the SVMs located on its neighbouring control points. This is similar to trilinear interpolation for an image. The classification results of the SVMs are combined by weighted voting, the weight determined by the distance from the target voxel to the control point,

$$v_x(k) = \sum_{z \in N(x)} w_{x,z} \cdot \delta_{l_{x,z},k} \tag{8}$$

$$w_{x,z} = (1 - \frac{z_1 - x_1}{s_1}) \cdot (1 - \frac{z_2 - x_2}{s_2}) \cdot (1 - \frac{z_3 - x_3}{s_3}) \tag{9}$$

where $N(x)$ denotes the set of the neighbouring control points near x, $l_{x,z}$ denotes the classification result for voxel x given by the SVM located at the control point z, s denotes the control point spacing, the subscript of x denotes each dimension of the 3D coordinate system. The label at voxel x is determined by the class with the highest vote. Algorithm 2 explains the procedure of the SVMAF algorithm.

Algorithm 2: SVM Segmentation with Augmented Features (SVMAF)

for *each control point z* **do**
  pre-select M atlas patches
  for *atlas patch i* **do**
    evaluate the augmented features f_i
  end
  train the SVM_z at z using $\{(f_i, l_i) | i = 1, 2, \cdots, M\}$
end
for *each voxel x* **do**
  evaluate the augmented features f_x
  for *each control point z near x* **do**
    determine the label $l_{x,z}$ by applying SVM_z to f_x
  end
  determine the label l_x by combining $l_{x,z}$
end

3. Experimental Results

3.1. Data Set

In order to validate the proposed segmentation algorithms, we tested them on the cardiac training data set from MIC-

CAI 2013 SATA Segmentation Challenge[1]. The data set was provided by the Cardiac Atlas Project (CAP), which is a collaborative database contributed by multiple institutions (Fonseca et al., 2011). The data used for the challenge were randomly selected from the DETERMINE (Defibrillators to Reduce Risk by Magnetic Resonance Imaging Evaluation) cohort (Kadish et al., 2009). The DETERMINE study comprises of patients with coronary artery disease and regional wall motion abnormalities due to prior myocardial infarction. MR scanner types include GE Medical Systems (Signa 1.5T), Philips Medical Systems (Achieva 1.5T, 3.0T and Intera 1.5T) and Siemens (Avanto 1.5T, Espree 1.5T and Symphony 1.5T). The cardiac MR images were acquired using the steady-state free precession (SSFP) pulse sequence. MR parameters vary between cases. Short-axis images are acquired over multiple breath-holds. Typical short-axis slice parameters are either 6 mm slice thickness with 4 mm gap or 8 mm slice thickness with 2 mm gap. Slice size ranges from 138×192 to 512×512 pixels and in-plane resolution ranges from 0.7 mm to 2.0 mm. The number of slices ranges from 8 to 15. Each cine image sequence has about 19 to 30 frames. The provided training data set consists of short-axis cine MR images for 83 subjects. The segmentation of the myocardium was carried out for all the temporal frames with the guide-point modelling approach in which most of input was manual (Young et al., 2000; Li et al., 2010).

We perform leave-one-out cross-validation on the data set. When each subject is segmented, the other 82 subjects are considered as the atlas set. Using the atlas selection procedure as proposed in (Aljabar et al., 2009), 40 out of the 82 atlases most similar to the target image in terms of NMI are selected for non-rigid image registration in order to save computation and to remove dissimilar atlases for label fusion.

3.2. Pre-Processing: Spatial and Temporal Alignment

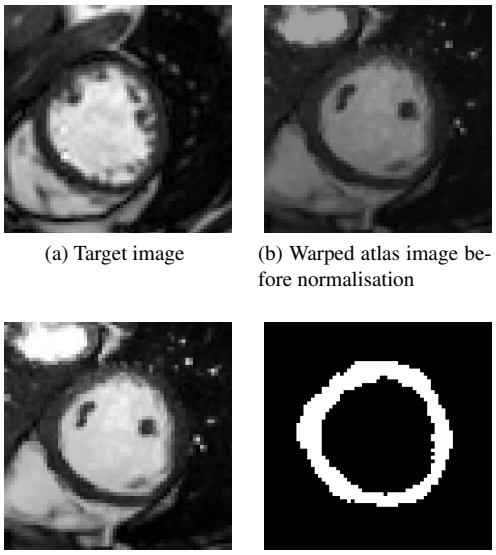
Since the target image sequence and the atlas image sequence may contain different numbers of frames, temporal alignment is performed, which establishes the temporal relationship between the target frame and the atlas frame by piecewise linear transformation,

$$T(i) = \begin{cases} i \cdot \frac{ES_a}{ES_t} & \text{if } 0 \le i \le ES_t \\ (i - ES_t) \cdot \frac{F_a - ES_a}{F_t - ES_t} + ES_a & \text{if } ES_t < i \le F_t \end{cases}$$

where i denotes the target frame index, $T(i)$ denotes the atlas frame index that it corresponds to, the index of the end-diastolic (ED) frame is normally zero and the index of the end-systolic (ES) frame is denoted by ES, F denotes the total number of frames and the subscripts t and a respectively denote target and atlas. Note that piecewise linear mapping is only a simple model to establish the temporal relationship between target sequence and atlas sequence and more sophisticated non-linear models may be used for this purpose. However, the focus of this paper is on the multi-atlas segmentation algorithm. Therefore, we do not discuss these possibilities further.

Once the temporal mapping is established, each frame of the target image sequence is spatially aligned with the corresponding atlas frame. Five landmarks are manually picked in the target image and in the atlas image both at the ED frame, which takes about 15 seconds per subject. Landmark-based registration is used to initialise the B-spline non-rigid image registration, which uses normalised mutual information (NMI) as the similarity metric (Bai et al., 2013a; Rueckert et al., 1999). Since the short-axis image slices are acquired over multiple breath-holds, there may be inconsistency in the breath-hold position, which can lead to inter-slice shift in neighbouring image slices (Suinesiaputra et al., 2014). To compensate for the potential inter-slice shift, we perform 2D slice-to-slice image registration, based on an initial 3D affine transformation provided by the landmarks.

Image registration takes about 5 to 10 minutes per frame, using the IRTK software package (Rueckert et al., 1999). Using the resulting transformation, the atlas image and label map are warped to the target image space. Since the target and atlas images may have a different intensity distribution due to the difference of MR scanners and acquisition protocols, intensity normalisation is performed prior to label fusion which maps the histogram of the atlas image to the histogram of the target image (Nyul et al., 2000). Figure 4 displays an example of the target image, the warped atlas image before and after intensity normalisation and the corresponding label map.



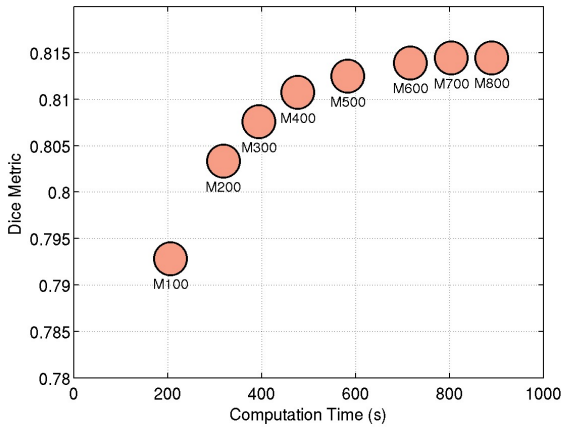
(a) Target image (b) Warped atlas image before normalisation (c) Warped atlas image after normalisation (d) Warped label map

Figure 4: An example of the target image, the warped atlas image before and after intensity normalisation and the warped atlas label map.

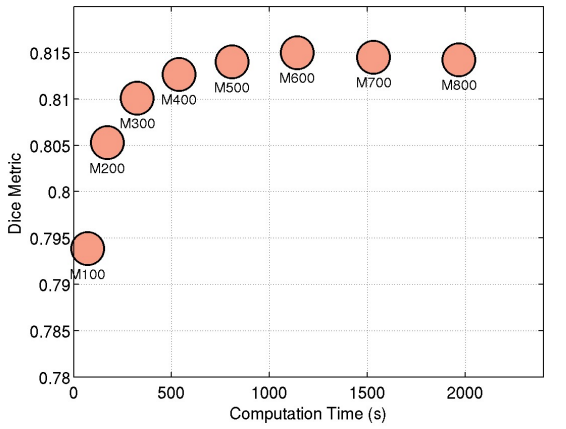
[1]The data set is available upon request from the challenge website `https://masi.vuse.vanderbilt.edu/workshop2013/index.php/Segmentation_Challenge_Details` (the challenge data set) and from the CAP website `http://www.cardiacatlas.org` (a even larger data set). (Last accessed: 2 Sep 2014)

Table 2: Dimension of the augmented feature vector.

Type	No. features	Detail
Intensity	49	Voxel intensities in the 7×7×1 patch
Gradient	147	Gradient orientations at the patch voxels
Contextual	32	Intensity difference
	32	BRIEF descriptor
Total	260	



(a) PBAF



(b) SVMAF

Figure 5: The impact of patch pre-selection on the computation time and the segmentation accuracy of the PBAF and SVMAF algorithms. The subscript below the circle symbol denotes the number of pre-selected patches M.

3.3. Parameter Settings

A subset of 10 subjects was used as a testing set for parameter tuning. Segmentation was performed for the ED frames of these subjects and the performance was evaluated. First of all, we tuned the parameters of patch size and search volume and found that the conventional patch-based segmentation (PB) worked quite well using a patch size of 7×7×1 voxels and a search volume of 5×5×1 voxels. The reason that we used a 2D patch and search volume is that the cardiac image set is of anisotropic spacing and has a large slice thickness of 8 mm, whereas the in-slice resolution is about 1.5 mm. Then we set the same patch size and search volume for the proposed PBAF and SVMAF algorithms so that we can compare their performance to PB. The main difference between the conventional PB and the proposed PBAF is the different feature vector being used, switching from the patch intensity feature to the augmented features, whereas the main difference between PBAF and SVMAF is the different classifier, switching from KNN to SVM.

For the contextual feature, we draw sample regions of size 3×3×1 voxels, radiating from the patch centre at every 45° and respectively at the radius of 4, 8, 16 and 32 voxels. As a result, the dimension of both the continuous and binary contextual feature is 32. The dimension of the augmented feature vector is 260, as explained in Table 2.

Since the number of features is substantially increased as compared to the conventional patch-based segmentation, we perform patch pre-selection and the augmented feature vector is only evaluated for those selected patches. This strategy does not only save computation for evaluating the feature vector, but also accelerates the training process for the SVM classifiers. Figures 5(a) and (b) illustrate the impact of patch pre-selection on the speed and the segmentation accuracy of the PBAF and SVMAF algorithms. Given 40 atlases and a 5×5×1 search volume, the total number of available patches is 1000. We tested the number of pre-selected patches $M \in \{100, 200, 300, \ldots, 800\}$. The segmentation accuracy is evaluated using the mean Dice metric at the ED frame across the 10 subjects.

For PBAF, as Figure 5(a) shows, the mean Dice metric gradually increases as M increases. The improvement becomes trivial when $M \geq 600$. Also, more patches result in a longer computation time. For SVMAF, Figure 5(b) shows a similar trend. Interestingly, when $M \geq 600$ and more patches are used to train the SVM, the segmentation accuracy even decreases slightly. This is probably because that patch pre-selection is performed based on the ranking of the MSD between the target patch and the atlas patch. Therefore, when more patches are introduced,

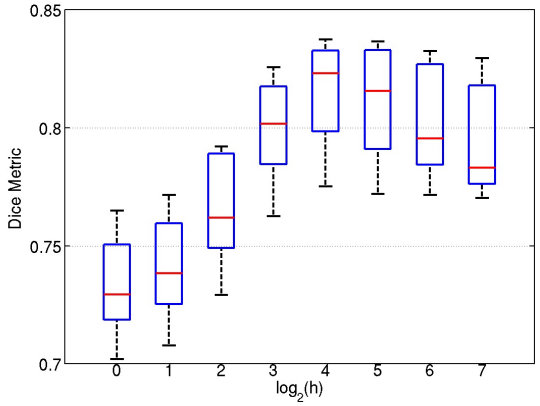


Figure 6: The impact of the parameter h on the segmentation accuracy of the PBAF algorithm. The central mark of the box is the median and the edges are the 25th and 75th percentiles.

it is likely that they are more and more dissimilar to the target patch and as a result degrade the segmentation accuracy. Another interesting observation is that the segmentation accuracy of the two algorithms are quite close, although SVM is more computationally expensive. We will set $M = 600$ for both PBAF and SVMAF algorithms in the following experiments.

3.3.1. Specific Parameters for PBAF

For the PBAF segmentation algorithm, a key parameter is the h parameter in the Gaussian kernel of Eq. (4). We tested a series of values for h: $\{2^0, 2^1, 2^2, \ldots, 2^7\}$ and found $h = 2^4$ worked well on the 10 subjects. We will set $h = 2^4$ for PBAF in the following experiments. Figure 6 shows the impact of h on the segmentation accuracy.

3.3.2. Specific Parameters for SVMAF

For the SVMAF segmentation algorithm, γ and C are the two key parameters. We tested different values for γ and C using grid search as recommended in (Chang and Lin, 2011): $\gamma \in \{2^{-11}, 2^{-10}, 2^{-9}, \ldots, 2^{-1}\}$ and $C \in \{2^{-5}, 2^{-4}, 2^{-3}, \ldots, 2^{1}\}$. We found that $\gamma = 2^{-8}$ and $C = 2^{-1}$ achieved a high segmentation accuracy on the set of 10 subjects. We will use this parameter setting for SVMAF in the following experiments. Figure 7 shows the impact of C and γ on the segmentation accuracy.

The control point spacing of the SVM classifiers is another important parameter for SVMAF. The denser the control points are, the more localised the resulting classifiers are and possibly the more accurate the segmentation is. However, it also results in much higher computational cost. We tested different control point spacings: 5×5×1, 4×4×1, 3×3×1 and 2×2×1. Again, since the cardiac image set is of anisotropic spacing, we did not downsample along the long-axis direction. Figure 8 shows the impact of the control point spacing on the speed and the segmentation accuracy. We found that 3×3×1 achieves a good balance between speed and accuracy. We will use this value in the experiments.

Using the above parameters, we then performed multi-atlas segmentation for all the temporal frames and for all of the 83

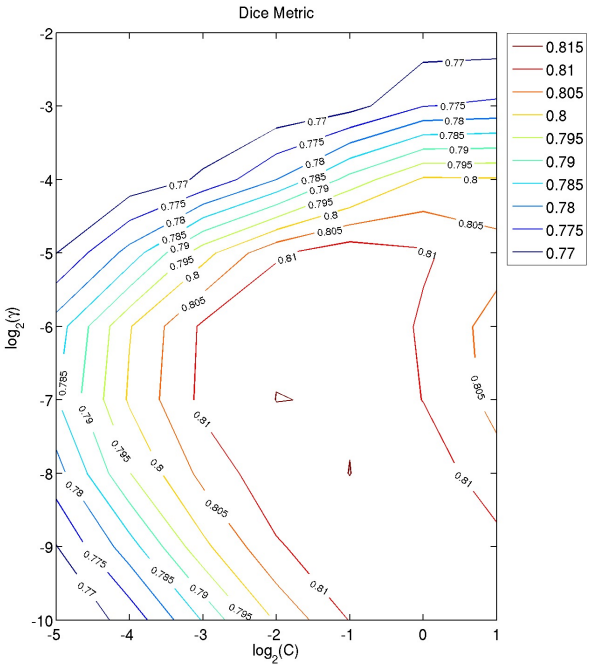


Figure 7: The impact of the parameters γ and C on the segmentation accuracy of the SVMAF algorithm.

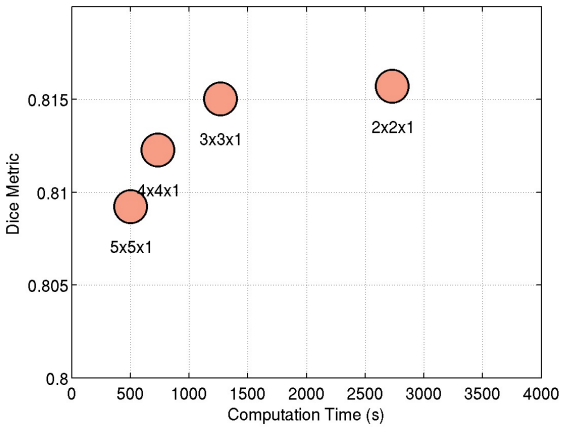


Figure 8: The impact of the control point spacing s on the computation time and the segmentation accuracy of the SVMAF algorithm.

Table 3: Comparison of the segmentation accuracy in terms of the mean Dice metric and the mean Jaccard metric for majority vote (MV), patch-based segmentation (PB), patch-based segmentation with augmented features (PBAF), SVM segmentation with augmented features (SVMAF) and joint label fusion (JLF).

	MV	PB	PBAF	SVMAF	JLF
Dice	0.764	0.794	0.812	0.815	0.815
Jaccard	0.625	0.661	0.686	0.689	0.690

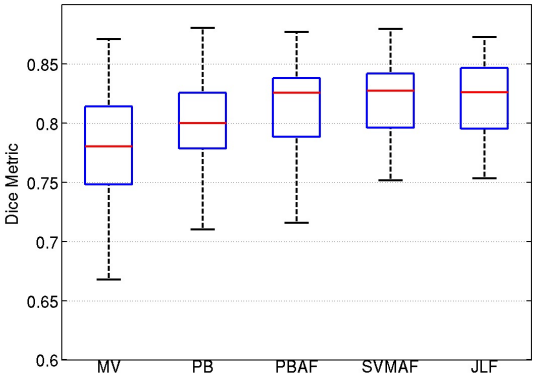


Figure 9: Comparison of the segmentation accuracy for different algorithms in terms of the Dice metric. The central mark of the box is the median and the edges are the 25th and 75th percentiles.

subjects. Segmentation takes about 10 minutes per frame for PBAF and about 20 minutes per frame for SVMAF using a single 2.6 GHz CPU on a workstation.

3.4. Segmentation Accuracy

In this experiment, we compare the segmentation accuracy of the proposed algorithms to the conventional patch-based segmentation algorithm (PB) (Coupé et al., 2011) as well as the state-of-the-art joint label fusion algorithm (JLF) (Wang et al., 2013). The segmentation accuracy is evaluated using the mean Dice metric and the mean Jaccard metric of the myocardium across all the temporal frames and across all the 83 subjects.

Table 3 lists the results. Apart from majority vote, all the other algorithms use the same patch size (7×7×1) and the same search volume (5×5×1). The parameters for PB, PBAF and SVMAF were set as discussed in the previous section. There are two parameters in the joint label fusion algorithm (JLF), namely α and β. We tuned $\alpha \in [0.05, 0.1, 0.15]$ and $\beta \in [0.5, 1, 1.5, 2, 2.5]$ in a cross-validation test and set $\alpha = 0.1$ and $\beta = 1.5$ which performed best in the test.

The majority vote algorithm (MV) serves as the baseline performance measure. The patch-based segmentation utilises a local search volume and the patch intensity feature in label fusion. It performs significantly better than majority vote ($p < 0.001$). By utilising even more features including gradient and contextual features, both PBAF and SVMAF perform better than PB ($p < 0.001$). By using a more sophisticated local classifier, SVMAF performs even better than PBAF ($p < 0.001$). Finally, the performance of SVMAF is comparable to the state-of-the-art JLF algorithm ($p > 0.1$). Figure 9 compares the segmentation performance in a box-plot. It also shows that both PBAF and SVMAF perform better than the conventional PB segmentation. SVMAF yields the highest median value. However, PBAF, SVMAF and JLF are all very close in terms of median values.

Figure 10 shows an example of the segmentation results for different segmentation algorithms. It shows that all the algorithms perform relatively well on the mid-ventricular slice. However, on the basal slice, PB produces a broken contour and on the apical slice, PB misses a large part of the myocardium. By incorporating the augmented features, both PBAF and SVMAF produce more continuous and better connected contours at the basal and apical slices. The segmentations are comparable to the state-of-the-art joint label fusion segmentation. Visually, SVMAF produces the best segmentation because the resulting myocardial contours are continuous on both apical and basal slices.

3.5. Clinical Measurement Errors

The end-diastolic volume (EDV), end-systolic volume (ESV) and the left ventricular mass (LVM) are clinically important indices for ventricular function assessment. Measurement errors of these clinical indices are also an important way to assess the segmentation accuracy (Suinesiaputra et al., 2014). We measure the EDV, ESV and LVM based on the segmentation and compute the measurement error against the ground truth measurement. The left ventricular mass (LVM) is calculated from the end-diastolic myocardial volume and using 1.05 gram/cm^3 as the density (Grothues et al., 2002).

The absolute measurement error is computed as,

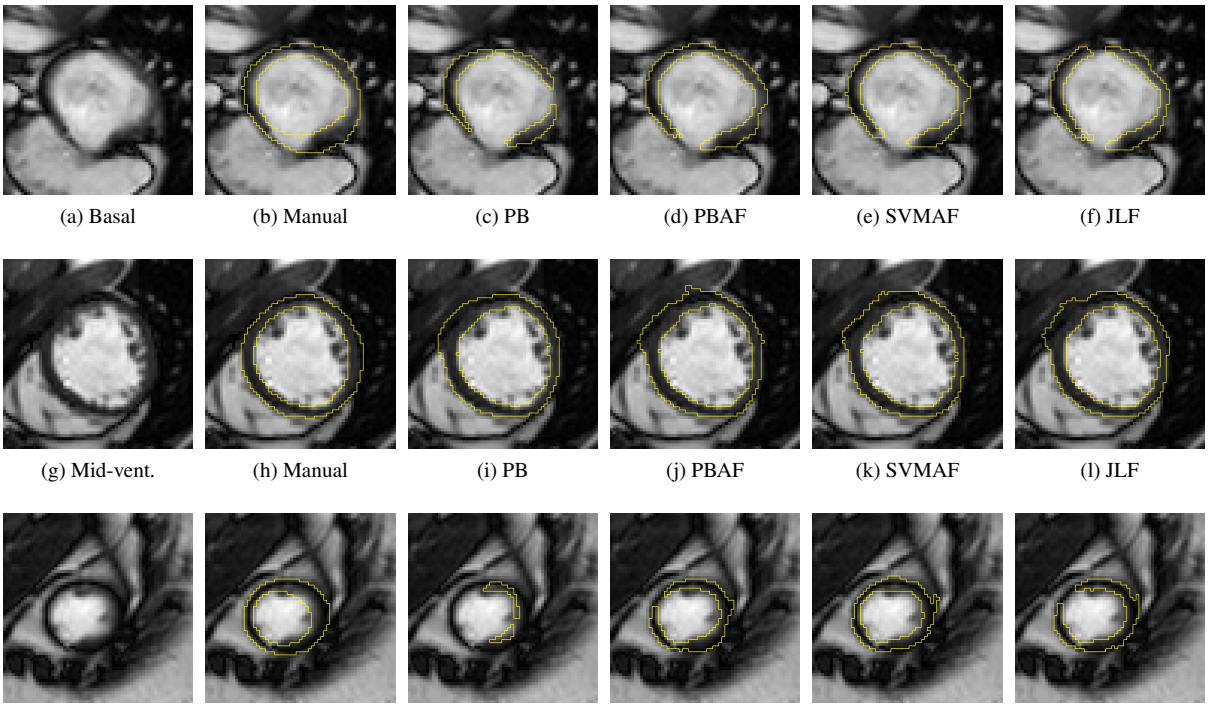
$$e_{\text{abs}} = |a - a_t| \tag{10}$$

where e_{abs} denotes the absolute error, a and a_t respectively denote the automated measurement and the ground truth.

Table 4 reports the mean and the standard deviation of the absolute measurement errors for different algorithms. As it shows, both PBAF and SVMAF reduce the measurement errors compared to the conventional PB. In addition, the performance of SVMAF is comparable to the state-of-the-art JLF.

3.6. Impact of Additional Features

In this experiment, we evaluate how segmentation performance is affected by incorporating additional features for the patches. The patch-based segmentation result is used as the baseline in the comparison. Then we add contextual and gradient information separately into the algorithm, i.e. extending the feature vector in Eq. (4). Finally, we use all the augmented features for patch-based segmentation which forms the proposed PBAF algorithm. Table 5 lists the segmentation accuracy for the myocardium. It shows that by incorporating additional features, the performance of patch-based segmentation can be improved. Incorporating continuous contextual feature and BRIEF contextual feature results in very similar improvement. When all the



(a) Basal (b) Manual (c) PB (d) PBAF (e) SVMAF (f) JLF

(g) Mid-vent. (h) Manual (i) PB (j) PBAF (k) SVMAF (l) JLF

(m) Apical (n) Manual (o) PB (p) PBAF (q) SVMAF (r) JLF

Figure 10: Comparison of the segmentation results for different algorithms. Top to bottom: the basal, mid-ventricular and apical slices from one of the subjects. Left to right: original images, manual labelling, patch-based segmentation (PB), PBAF, SVMAF and joint label fusion (JLF).

Table 4: Comparison of the absolute measurement errors of the end-diastolic volume (EDV), end-systolic volume (ESV) and left ventricular mass (LVM) for different algorithms. The mean and the standard deviation (Std.) of the absolute errors are reported.

		MV	PB	PBAF	SVMAF	JLF
EDV (mL)	Mean Error	12.5	10.1	9.2	8.9	8.8
	Std.	11.6	9.1	8.7	8.2	8.6
ESV (mL)	Mean Error	14.5	10.9	9.4	9.3	9.7
	Std.	14.5	11.1	9.4	9.9	10.3
LVM (gram)	Mean Error	17.8	14.0	11.8	11.9	11.5
	Std.	17.1	14.8	12.8	12.4	11.9

Table 5: Impact of incorporating more features on segmentation accuracy, in terms of the mean Dice metric and the mean Jaccard metric, compared to conventional PB. Columns 2 to 4 denote the metric when continuous context feature (C.Context), BRIEF context feature and gradient feature are respectively added to patch-based segmentation. The last column denotes the metric when all the augmented features are used.

	PB	PB+C.Context	PB+BRIEF	PB+Gradient	PB+All
Dice	0.794	0.800	0.800	0.810	0.812
Jaccard	0.661	0.669	0.668	0.683	0.686

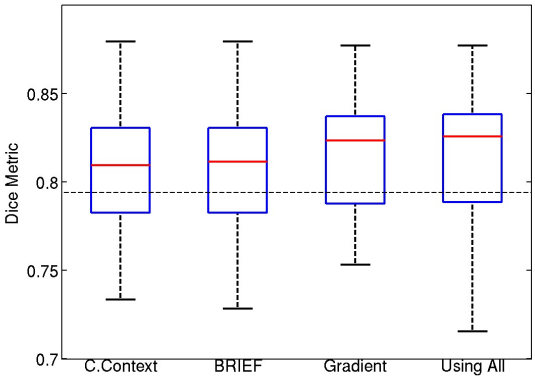


Figure 11: Impact of incorporating more features on the segmentation accuracy. The dashed line denotes the mean Dice metric for conventional PB. Columns 1 to 3 denote the Dice metric when continuous context feature (C.Context), BRIEF context feature and gradient feature are respectively added to patch-based segmentation. The last column denotes the Dice metric when all the augmented features are used.

Table 6: Segmentation performance on the challenge testing set. The columns list the proposed PBAF and SVMAF methods without and with graph-cuts (GC) post-processing.

	Without graph-cuts		With graph-cuts	
	PBAF	SVMAF	PBAF_GC	SVMAF_GC
Dice	0.800	0.804	0.807	0.807

augmented features are used, the proposed algorithm performs best. Figure 11 compares the impact of different features on the segmentation accuracy in a box-plot.

3.7. Impact of SVM

The proposed methods PBAF and SVMAF use the same feature vector but different classifiers, namely the KNN classifier and the SVM classifier. In this experiment, we evaluate how the SVM classifier has an impact on the segmentation performance, compared to the KNN classifier. Figure 12 compares the mean Dice metric of the two methods when different numbers of atlases are used. It shows that SVMAF performs consistently better than PBAF. When only a few atlases are used, the advantage is most significant. When more atlases are used, the difference between the two methods becomes marginal. This is probably because that when more atlases are used, there will be more training samples (atlas patches) distributed near the testing sample (target patch) in the feature space. In this scenario, both the KNN classifier and the SVM classifier can perform quite well and therefore their difference becomes marginal. However, in real scenarios, where the number of atlases are often limited, SVMAF can offer a practical improvement.

3.8. SATA Testing Set

All the previous experiments were performed on the CAP training set of MICCAI 2013 SATA Segmentation Challenge,

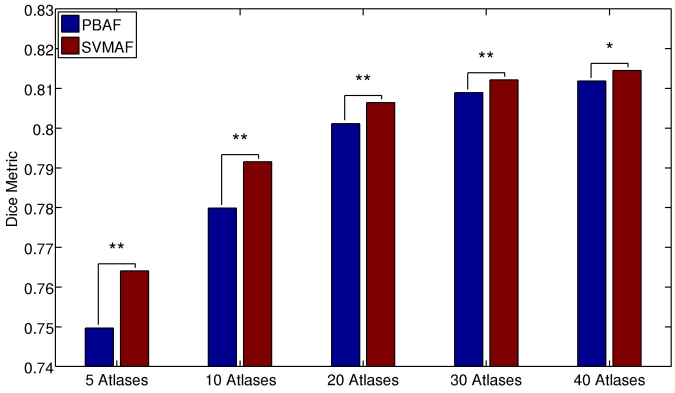


Figure 12: Comparison of the mean Dice metric for PBAF and SVMAF when different numbers of atlases are used. The Dice metric of the two methods are compared using paired t-test (one-sided). The symbol "$*$" denotes that $p < 0.01$ and "$**$" denotes that $p < 0.001$.

which consists of 83 subjects with ground truth segmentations. To compare performance against the state-of-the-art methods, we also tested the proposed methods on the CAP testing set of the segmentation challenge, which consists of 72 subjects, whose ground truth segmentations are unknown to the challenge participants. We submitted our segmentations to the challenge website. The results were evaluated by the website and published on its leaderboard[2]. We also list our results here in Table 6. The columns list performance of the proposed PBAF and SVMAF methods without and with graph-cuts (GC) post-processing. For post-processing, spatio-temporal graph-cuts is applied to the label probability map with a smoothness term linking both the spatial and the temporal neighbours (Bai et al., 2013b; Boykov and Jolly, 2001). Since the focus of this paper is on label fusion, we do not elaborate the post-processing step here. As the leaderboard on the challenge website shows, the proposed methods outperform or are comparable to other state-of-the-art methods.

4. Discussion

In this paper, we incorporate contextual and gradient features into the multi-atlas patch-based segmentation method and improve the segmentation accuracy. It is interesting to see that the gradient feature has a great impact on the improvement (Figure 11). This is possibly due to the fact that in the augmented feature vector, there are more dimensions for the gradient feature than the other features (Table 2). Introducing a weighting term for each individual components of the augmented feature vector (intensity, gradient and context) may result in a better combination of these features. The possibility of optimising the combination of different features can be an interesting direction for future research.

[2] `http://masi.vuse.vanderbilt.edu/submission/leaderboard.html#capfree`. (Last accessed: 2 Sep 2014)

In machine learning, good features are essential for classification performance. The proposed algorithms use hand-crafted features for classification. However, these features may not be optimal. Recently, it has been proposed that representation learning via a deep neural network can be used to generate good features for image segmentation and registration (Liao et al., 2013; Wu et al., 2013). Combining representation learning and multi-atlas image segmentation is another interesting direction to explore in the future.

One aspect that we have explored in this paper is whether it is plausible to use the SVM for local classification. The experimental results have shown that by training the SVM at sparsely located control points, it becomes computationally affordable to perform segmentation using local SVMs. We have found that in the comparison, the SVMAF algorithm performs well both visually and quantitatively. The reason that we choose the SVM classifier is that it is a non-linear classification algorithm and is easy to use. Also, it has been well adapted to many classification problems in the past. However, SVM is only one of the potential choices and other classification algorithms can also be explored for their applicability to multi-atlas segmentation problems in the future.

Recently we have noticed that in (Hao et al., 2012, 2014), Hao et al. explored very similar ideas to this work. However, they used a quite different feature set and applied the method to a different application, brain image segmentation. Both Hao's work and this work demonstrate the benefit of using augmented features for multi-atlas segmentation and the improvement due to using a more sophisticated classifier such as SVM. Apart from that, this work also demonstrates when SVM would outperform a simple classifier such as KNN. When the number of atlases is limited, SVM can offer a practical improvement. However, when more atlases are used, KNN can work almost as well as SVM but with a lower computational cost.

We have found that joint label fusion (JLF) proposed by Wang et al. (Wang et al., 2013) achieves very good segmentation performance similar to our methods. Although JLF only utilises intensity feature, it computes the weights for label fusion using an intelligent way to reduce the correlated labelling error. Our methods perform as well as JLF, however, by using more features instead of by using a more intelligent way for weighting. It would be interesting to combine the advantages of the two approaches. However, we have come across some difficulties in combining the two in a trial experiment.

In particular, JLF only uses the best matched patch in each atlas. If there are 40 atlases, the number of atlas patches that it uses is then n = 40. When JLF optimises its cost function, it boils down to computing the inverse of the patch correlation matrix (dimension: n by n) at each voxel. It would be computationally prohibitive if n is too large, i.e. too many atlas patches are used. On the contrary, in our methods, we use 600 atlas patches and therefore could not afford the computational cost for computing the matrix inverse of this size at each target voxel. As an alternative, we tried selecting only the 40 best matched atlas patches, in terms of best matched augmented features. This is computationally feasible, however, the segmentation performance is comparable to JLF but not significantly improved against it. It would be interesting to carry out further research in the future to understand how to best integrate the augmented feature vector into the JLF framework.

Finally, a bottleneck of the proposed multi-atlas segmentation framework is the expensive computational cost, which is a combination of the costs from target-to-atlas image registration and from label fusion. In our experiments, non-rigid image registration is performed in parallel for the selected 40 atlases using multiple CPUs of the workstation and it takes 5 to 10 minutes per temporal frame. If a cine image sequence consists of 30 temporal frames, the image registration would take nearly 300 minutes. PBAF and SVMAF take 10 minutes and 20 minutes respectively and they are performed in parallel for all the temporal frames. So combining image registration and label fusion, it takes about 320 minutes or over 5 hours per subject.

However, the computational time can be reduced in several ways. First of all, we can perform affine image registration instead of non-rigid image registration, which would be much faster. On our data set, affine registration would only take about 30 minutes for all the 30 temporal frames. The adverse effect is that it may sacrifice the segmentation accuracy slightly and may require a larger search volume for the subsequent label fusion step, which increases the cost for label fusion. Second, it is possible to implement image registration (Modat et al., 2010) or label fusion on GPU, which may increase the speed by the order of magnitude. Third, if accurate cardiac motion tracking is available, we only need to segment a single temporal frame and propagate it to all the other frames, which may be much faster.

In our previous paper (Bai et al., 2013a), we performed registration refinement to improve the image registration accuracy. However, we did not perform registration refinement in this work mainly for two reasons. One reason is that registration refinement would significantly increase the computational cost. Registration refinement incorporates intermediate label fusion results and iteratively performs image registration between the target image and each of the atlas image. The computational cost for image registration will increase by the factor of the number of iterations. The other reason is that the current paper focuses on the label fusion step, such as features and classifiers used for fusion. Therefore, we did not perform the iterative registration refinement here.

5. Conclusions

To conclude, we have proposed two novel segmentation algorithms, namely PBAF and SVMAF, which incorporate gradient and contextual information into multi-atlas label fusion. Experimental results have shown that both algorithms yield significant improvement in terms of segmentation accuracy over conventional patch-based segmentation. This study has shown that non-local patch-based segmentation, as a classification problem, can benefit from adding a careful selection of discriminative image features into segmentation.

Acknowledgements

The authors would like to thank Dr. Anil Rao, Dr. Kanwal Bhatia, Dr. Claire Donoghue, Haiyan Wang, Zehan Wang and Tong Tong for fruitful discussion. We would like to thank the anonymous reviewers for their helpful and constructive comments.

References

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 46 (3), 726–738.

Artaechevarria, X., Muñoz-Barrutia, A., Ortiz-de Solorzano, C., 2009. Combination strategies in multi-atlas image segmentation: Application to brain MR data. IEEE Transactions on Medical Imaging 28 (8), 1266–1277.

Asman, A., Landman, B., 2013. Non-local statistical label fusion for multi-atlas segmentation. Medical Image Analysis 17 (2), 194–208.

Bai, W., Shi, W., O'Regan, D., Tong, T., Wang, H., Jamil-Copley, S., Peters, N., Rueckert, D., 2013a. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: Application to cardiac MR images. IEEE Transactions on Medical Imaging 32 (7), 1302–1315.

Bai, W., Shi, W., Peters, N., Rueckert, D., 2013b. Patch-based label fusion with spatio-temporal graph cuts for cardiac MR images. In: MICCAI 2013 SATA Workshop. pp. 1–4.

Boser, B., Guyon, I., Vapnik, V., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. ACM, pp. 144–152.

Boykov, Y., Jolly, M.-P., 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In: ICCV 2001. Vol. 1. IEEE, pp. 105–112.

Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P., 2012. BRIEF: Computing a local binary descriptor very fast. IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (7), 1281–1298.

Cardoso, M., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N., Ourselin, S., 2013. STEPS: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation. Medical Image Analysis 17 (6), 671–684.

Chang, C., Lin, C., 2011. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2 (3), Article No. 27.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20 (3), 273–297.

Coupé, P., Manjón, J., Fonov, V., Pruessner, J., Robles, M., Collins, D., 2011. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. NeuroImage 54 (2), 940–954.

Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M., Vembar, M., Olszewski, M., Subramanyan, K., Lavi, G., Weese, J., 2008. Automatic model-based segmentation of the heart in CT images. IEEE Transactions on Medical Imaging 27 (9), 1189–1201.

Eskildsen, S., Coupé, P., Fonov, V., Manjón, J., Leung, K., Guizard, N., Wassef, S., Østergaard, L., Collins, D., 2012. BEaST: Brain extraction based on nonlocal segmentation technique. NeuroImage 59 (3), 2362–2373.

Feng, C., Li, C., Zhao, D., Davatzikos, C., Litt, H., 2013. Segmentation of the left ventricle using distance regularized two-layer level set approach. In: MICCAI 2013. Springer, pp. 477–484.

Fonseca, C., Backhaus, M., Bluemke, D., Britten, R., Chung, J., Cowan, B., Dinov, I., Finn, J., Hunter, P., Kadish, A., et al., 2011. The Cardiac Atlas Project - An imaging database for computational modeling and statistical atlases of the heart. Bioinformatics 27 (16), 2288–2295.

Grothues, F., Smith, G., Moon, J., Bellenger, N., Collins, P., Klein, H., Pennell, D., 2002. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. The American Journal of Cardiology 90 (1), 29–34.

Hao, Y., Liu, J., Duan, Y., Zhang, X., Yu, C., Jiang, T., Fan, Y., 2012. Local label learning (L3) for multi-atlas based segmentation. In: SPIE Medical Imaging 2012. p. 83142E.

Hao, Y., Wang, T., Zhang, X., Duan, Y., Yu, C., Jiang, T., Fan, Y., 2014. Local label learning (LLL) for subcortical structure segmentation: Application to hippocampus segmentation. Human Brain Mapping 35 (6), 2674–2697.

Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33 (1), 115–126.

Hu, S., Coupé, P., Pruessner, J., Collins, D., 2012. Nonlocal regularization for active appearance model: Application to medial temporal lobe segmentation. Human Brain Mapping, DOI: 10.1002/hbm.22183.

Iglesias, J., Sabuncu, M., Van Leemput, K., 2013. A unified framework for cross-modality multi-atlas segmentation of brain MRI. Medical Image Analysis 17 (8), 1181–1191.

Išgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M., van Ginneken, B., 2009. Multi-atlas-based segmentation with local decision fusion - Application to cardiac and aortic segmentation in CT scans. IEEE Transactions on Medical Imaging 28 (7), 1000–1010.

Jia, H., Yap, P., Shen, D., 2011. Iterative multi-atlas-based multi-image segmentation with tree-based registration. NeuroImage 59, 422–430.

Jolly, M., 2006. Automatic segmentation of the left ventricle in cardiac MR and CT images. International Journal of Computer Vision 70 (2), 151–163.

Kadish, A., Bello, D., Finn, J., Bonow, R., Schaechter, A., Subacius, H., Albert, C., Daubert, J., Fonseca, C., Goldberger, J., 2009. Rationale and design for the defibrillators to reduce risk by magnetic resonance imaging evaluation (DETERMINE) trial. Journal of Cardiovascular Electrophysiology 20 (9), 982–987.

Kim, M., Wu, G., Li, W., Wang, L., Son, Y., Cho, Z., Shen, D., 2013. Automatic hippocampus segmentation of 7.0 Tesla MR images by combining multiple atlases and auto-context models. NeuroImage 83, 335–345.

Langerak, T., van der Heide, U., Kotte, A., Viergever, M., van Vulpen, M., Pluim, J., 2010. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Transactions on Medical Imaging 29 (12), 2000–2008.

Li, B., Liu, Y., Occleshaw, C., Cowan, B., Young, A., 2010. In-line automated tracking for ventricular function with magnetic resonance imaging. JACC: Cardiovascular Imaging 3 (8), 860–866.

Liao, S., Gao, Y., Oto, A., Shen, D., 2013. Representation learning: A unified deep learning framework for automatic prostate MR segmentation. In: MICCAI 2013. Springer, pp. 254–261.

Lotjonen, J., Wolz, R., Koikkalainen, J., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage 49 (3), 2352–2365.

Lynch, M., Ghita, O., Whelan, P., 2006. Automatic segmentation of the left ventricle cavity and myocardium in MRI data. Computers in Biology and Medicine 36 (4), 389–407.

Margeta, J., Geremia, E., Criminisi, A., Ayache, N., 2011. Layered spatio-temporal forests for left ventricle segmentation from 4D cardiac MRI data. In: STACOM Workshop. Springer, pp. 109–119.

Modat, M., Ridgway, G., Taylor, Z., Lehmann, M., Barnes, J., Hawkes, D., Fox, N., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine 98 (3), 278–284.

Morra, J., Tu, Z., Apostolova, L., Green, A., Toga, A., Thompson, P., 2010. Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. IEEE Transactions on Medical Imaging 29 (1), 30–43.

Nyul, L., Udupa, J., Zhang, X., 2000. New variants of a method of MRI scale standardization. IEEE Transactions on Medical Imaging 19 (2), 143–150.

Petitjean, C., Dacher, J., 2011. A review of segmentation methods in short axis cardiac MR images. Medical Image Analysis 15, 169–184.

Rousseau, F., Habas, P., Studholme, C., 2011. A supervised patch-based approach for human brain labeling. IEEE Transactions on Medical Imaging 30 (10), 1852–1862.

Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging 18 (8), 712–721.

Sabuncu, M., Yeo, B., Van Leemput, K., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. IEEE Transactions on Medical Imaging 29 (10), 1714–1729.

Suinesiaputra, A., Cowan, B., Al-Agamy, A., Elattar, M., Ayache, N., Fahmy, A., Khalifa, A., Medrano-Gracia, P., Jolly, M., Kadish, A., Lee, D., Margeta, J., Warfield, S., Young, A., 2014. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. Medical Image Analysis 18 (1), 50–62.

Tong, T., Wolz, R., Coupé, P., Hajnal, J., Rueckert, D., 2013. Segmentation of MR images via discriminative dictionary learning and sparse coding: Appli-

cation to hippocampus labeling. NeuroImage 76, 11–23.
Tzimiropoulos, G., Zafeiriou, S., Pantic, M., 2012. Subspace learning from image gradient orientations. IEEE Transactions on PAMI 34 (12), 2454–2466.
van Rikxoort, E., Išgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M., Pluim, J., van Ginneken, B., 2010. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. Medical Image Analysis 14 (1), 39–49.
Wang, H., Das, S., Suh, J., Altinay, M., Pluta, J., Craige, C., Avants, B., Yushkevich, P., 2011a. A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation. NeuroImage 55 (3), 968–985.
Wang, H., Suh, J., Das, S., Pluta, J., Altinay, M., Yushkevich, P., 2011b. Regression-based label fusion for multi-atlas segmentation. In: CVPR 2011. pp. 1113–1120.
Wang, H., Suh, J., Das, S., Pluta, J., Craige, C., Yushkevich, P., 2013. Multi-atlas segmentation with joint label fusion. IEEE Transactions on PAMI 35 (3), 611–623.
Wang, L., Shi, F., Gao, Y., Li, G., Gilmore, J., Lin, W., Shen, D., 2014. Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain MR image segmentation. NeuroImage 89, 152–164.
Warfield, S., Zou, K., Wells, W., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging 23 (7), 903–921.
Wolz, R., Chengwen, C., Misawa, K., Fujiwara, M., Mori, K., Rueckert, D., 2013. Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE Transactions on Medical Imaging 32 (9), 1723–1730.
Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S., Shen, D., 2013. Unsupervised deep feature learning for deformable registration of MR brain images. In: MICCAI 2013. Springer, pp. 649–656.
Young, A., Cowan, B., Thrupp, S., Hedley, W., DellItalia, L., 2000. Left ventricular mass and volume: Fast calculation with guide-point modeling on MR images. Radiology 216 (2), 597–602.
Zhang, D., Guo, Q., Wu, G., Shen, D., 2012. Sparse patch-based label fusion for multi-atlas segmentation. In: MICCAI Workshop: Multimodal Brain Image Analysis. Springer, pp. 94–102.
Zhang, H., Wahle, A., Johnson, R., Scholz, T., Sonka, M., 2010. 4-D cardiac MR image analysis: left and right ventricular morphology and function. IEEE Transactions on Medical Imaging 29 (2), 350–364.
Zhuang, X., Rhode, K., Razavi, R., Hawkes, D., Ourselin, S., 2010. A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Transactions on Medical Imaging 29 (9), 1612–1625.
Zikic, D., Glocker, B., Konukoglu, E., Shotton, J., Criminisi, A., Ye, D., Demiralp, C., Thomas, O., Das, T., Jena, R., Price, S., 2012. Context-sensitive classification forests for segmentation of brain tumor tissues. MICCAI Workshop: Challenge on Multimodal Brain Tumor Segmentation, 1–9.