Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI☆

Xiahai Zhuang [a,b,*], Juan Shen [c]

[a] *SJTU-CU ICRC, Department of Engineering Mechanics, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, P R China*
[b] *Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration (CISSE), Shanghai 200240, P R China*
[c] *Department of Statistics, School of Management, Fudan University, China*

ABSTRACT

A whole heart segmentation (WHS) method is presented for cardiac MRI. This segmentation method employs multi-modality atlases from MRI and CT and adopts a new label fusion algorithm which is based on the proposed multi-scale patch (MSP) strategy and a new global atlas ranking scheme. MSP, developed from the scale-space theory, uses the information of multi-scale images and provides different levels of the structural information of images for multi-level local atlas ranking. Both the local and global atlas ranking steps use the information theoretic measures to compute the similarity between the target image and the atlases from multiple modalities. The proposed segmentation scheme was evaluated on a set of data involving 20 cardiac MRI and 20 CT images. Our proposed algorithm demonstrated a promising performance, yielding a mean WHS Dice score of 0.899 ± 0.0340, Jaccard index of 0.818 ± 0.0549, and surface distance error of 1.09 ± 1.11 mm for the 20 MRI data. The average runtime for the proposed label fusion was 12.58 min.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

According to the World Health Organization, an estimated 17.3 million people died from cardiovascular diseases (CVDs) in 2008, accounting for 30% of deaths around the world (Organization, 2011). Early diagnosis and treatment play a vital role in reducing the mortality and morbidity of CVDs. Recently advances in medical imaging and computing technology and their application to the clinics have shown great potential towards achieving this goal. Cardiac magnetic resonance imaging (MRI) is a unique non-ionizing radiation technique which provides clear view of heart's anatomy. To enable the development of novel clinical applications and thus improve cardiology, accurate and automatic extraction of the anatomical information becomes particularly important.

Whole heart segmentation (WHS) aims to extract the substructures of the heart, commonly including the four chamber blood cavities, the left ventricular myocardium, and sometimes the great vessels as well if they are of interest (Zheng et al., 2008; Peters et al., 2009; Zhuang, 2013). The results of WHS have a number of potential clinical applications. For instance, it can be directly used to calculate the functional indices of the heart such as the ejection fraction and myocardial mass; it can provide the initial geometric information of the heart for surgical guidance such as in radio-frequency ablation of the left atrium. It is also anticipated that the functional analysis of the whole heart has the potential of detecting subtle functional abnormalities or changes of the heart (Vasan et al., 1996). This is critical for the early diagnosis of patients who otherwise have normal systolic function of the ventricles but are suspected to have abnormal function in other regions. Two examples of WHS are provided in Fig. 1.

Obtaining fully automatic WHS is arduous due to the three major challenges: (1) the large shape variations of the cardiac anatomy, (2) the indistinct boundaries between substructures of the heart in cardiac MRI images, and (3) the low image quality (Zhuang, 2013). Zheng et al. (2008) developed a statistical shape model (SSM) and learning-based method to hierarchically detect boundary landmarks of the heart, based on the steerable features and marginal space learning. The SSM was used to regularize the resulting shape of the WHS defined by the detected landmarks. Peters et al. (2009) and Ecabert et al. (2011) developed a deformable model-based method for the WHS of both cardiac CT and MRI images. The large shape variations were tackled by the piecewise affine parametric adaptation and the boundary was detected by the simulated search of the optimal response of edges. Kirisli et al. (2010) constructed eight atlases and performed a multicenter, multi-vendor evaluation study on the WHS of CT data. The
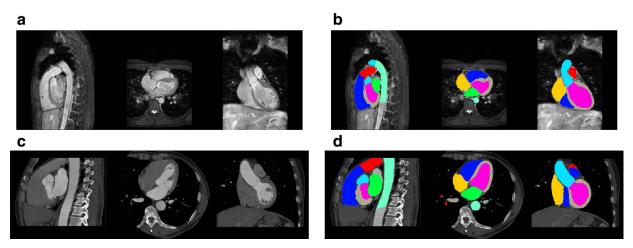
Fig. 1. Examples of cardiac images and whole heart segmentation (WHS) results: (a) and (b) are a cardiac MRI image and its corresponding manual WHS result, (c) and (d) are a cardiac CT image and the corresponding manual WHS result.

WHS performance was validated using a leave-one-out validation strategy on 8 datasets with manual segmentation. Zhuang et al. (2010) built a mean cardiac MRI atlas from 10 healthy subjects and developed a comprehensive registration algorithm for the atlas-based WHS of MRI. The authors employed the mean atlas to segment a test set involving nine different pathologies, relying on the locally affine registration method (LARM) to tackle the large shape variability of the cardiac anatomy. They also showed that the conventional multi-atlas segmentation (MAS) scheme did not performed better than the single mean atlas segmentation. Recently, Zhuang et al. (2015) developed a MAS scheme for the WHS of CT images, which adopted a new atlas ranking algorithm based on conditional entropy. The results showed that both the atlas ranking and label fusion could greatly affect the performance of a MAS scheme.

Previous studies show that the MAS reaches optimal performance by fusing a certain number of selected atlases using an effective atlas ranking method (Aljabar et al., 2009; Zhuang et al., 2015). The global atlas selection idea has been extended to local atlas selection, based on local substructures or organs (Shi et al., 2010; van Rikxoort et al., 2010; Wolz et al., 2013). Recently, the idea has been further extended to the pixel level, to provide atlas selection for each location (Wolz et al., 2013; Tong et al., 2014; 2015). Currently, the atlas selection strategy generally requires the users to determine the number of atlases selected for label fusion, either on the global level or local level. In this manner, the selection is equivalent to applying a binary threshold on the contribution of atlases during label fusion.

Furthermore, to implement the multi-level atlas selection, one needs to define the masks of local regions according to the definition of anatomical structures (Wolz et al., 2013). This limits the flexibility of choosing the number of levels and consequently reduces the applicability of the method.

Another general limitation of the conventional MAS is that the atlases are constructed from one single modality. Since the number of atlases determines the potential optimal performance of MAS, including multi-modality atlases can be beneficial to the applications when different modality atlases are available. Iglesias et al. (2013) developed a MAS method where the atlas-to-target image registration and label fusion are solved simultaneously. Their proposed method was applied to cross-modality atlas-based segmentation, i.e. the proton density brain MRI atlases were used to segment the T1-weighted MRI and vice versa. Wang et al. (2015) proposed a learning-based multi-source atlas segmentation method. In their work, each atlas had intensity images from three modalities, including T1-weighted, T2-weighted and fractional anisotropy MRI, and the atlases were used to segment the target images from one modality such as T1 MRI. This is also different from the segmentation using multi-modality atlases, where one atlas is constructed solely from one imaging modality and different atlases can be built from different modalities.

Recently, a number of works adopt learning-based methods to predict the label of a target patch (Zhang et al., 2012; 2011; Tong et al., 2014; 2015; Bai et al., 2015). In these methods, the known patches, with gold standard labeling, from the atlases are used as training data and the target patch is considered as the test data. The voting weights of the known patches are implicitly implemented in the learning-and-prediction algorithm, such as the linear sparse encode using dictionary learning framework or the nonlinear kernel support vector machine method (Awate and Whitaker, 2014; Tong et al., 2014; 2015; Bai et al., 2015). These approaches generally adopt the online learning scheme and a large number of atlases for training, and the performance is comparable to the state-of-the-art joint label fusion method.

To address the challenges of WHS and aforementioned limitations in current MAS research, we propose a *m*ulti-*m*odality *m*ulti-atlas segmentation (M^3AS) framework for WHS of cardiac MRI. M^3AS adopts a new multi-scale patch (MSP) strategy, based on the multi-scale theory (Lindeberg, 1998), to obtain hierarchical local atlas ranking. The multi-scale space theory can handle different-level information within a limited window and has been applied to feature extraction/detection and image matching (Lindeberg, 1998; 2011; Holden et al., 2004; Leutenegger et al., 2011; Wu et al., 2015; Lowe, 2004; Wu et al., 2014). We develop the MSP to compute the patches from different scale spaces to represent the different levels of structural information, with low scale capturing local fine structure and high scale suppressing fine structure but providing global structural information of the image. This is different from the conventional patch-based methods which only compute the local structural information within the patch. To avoid increasing the computational complexity, we adopt the multi-resolution implementation and couple it with the MSP where the high-scale patch can be efficiently computed using a low-resolution image space.

The main contributions of this work are summarized as follows: (1) the MSP for patch-based hierarchical local atlas ranking; (2) a framework for M^3AS label fusion which evaluates both global image similarity and local pixel/patch similarity using information theoretic measures; (3) the non-binary global atlas ranking scheme which is formulated using a truncated Gaussian kernel regression model and does not require explicit atlas selection; (4) a validation

study of WHS from cardiac MRI where comparisons between different patch parameterization schemes and label fusion methods are provided.

The remainder of the paper is organized as follows: Section 2 describes the methodologies of this work. Section 3 introduces the experiments and results, followed by the discussion and conclusion in Section 4.

2. Method

This section elaborates on the methodologies of this paper. First, Section 2.1 presents the framework of MAS and local weighted label fusion (LWF). Here, we propose a local similarity metric for multi-modality images on the pixel level and introduce three joint label fusion (JLF) methods which are used for comparisons in the experiments. Then, Section 2.2 describes the proposed MSP for multi-level hierarchical local atlas ranking. Finally, Section 2.3 provides the framework of the proposed M^{3}AS, where the label fusion is based on the MSP and a global atlas ranking scheme.

2.1. Multi-atlas segmentation and local weighted label fusion

Let I be the target image to be segmented, $\{(A_a, L_a)|a = 1, \ldots, N\}$ be the set of atlases where A_a and L_a are respectively the intensity image and corresponding segmentation label image of the ath atlas. For each atlas, MAS performs an atlas-to-target registration to derive the set of warped atlases, $\{(\mathcal{A}_a, \mathcal{L}_a)|a = 1, \ldots, N\}$, where $\mathcal{A}_a$ is the warped atlas intensity image and $\mathcal{L}_a$ is the corresponding segmentation result.

MAS generally selects a subset of N' ($N' \leq N$) atlases based on an atlas ranking and selection criterion, and the segmentation result is derived by a label fusion step (Rohlfing et al., 2004; Rohlfing and Maurer, 2005; Aljabar et al., 2009; van Rikxoort et al., 2010; Zhuang et al., 2015). The optimal value of N' is unknown and differs from case to case. Hence, it is *manually determined* based on previous experience and then *fixed for all the target cases* of an application (Bai et al., 2013; Zhuang et al., 2015).

The majority vote fusion (MVF) (Kittler et al., 1998) counts the number of atlases which provide the same label for a pixel c whose label is to be decided. The label obtaining the largest number of votes is then selected as the estimated label:

$$L(c) = \underset{l \in \{l_1 \ldots l_K\}}{\mathrm{argmax}} \sum_{a=1}^{N'} \delta(\mathcal{L}_a(c), l), \tag{1}$$

where $\{l_1 \ldots l_K\}$ is the set of K labels; $\delta(x, y)$ is the Kronecker delta function which returns 1 when $x = y$ and returns 0 otherwise. In MVF, the votes from all the atlases are equally treated, regardless their different segmentation qualities. This equality can however affect the performance of label fusion, particularly in the challenging cases, where only a small number of *good* atlases can generate correct labels.

Artaechevarria et al. (2009) proposed the LWF where the contribution (vote) of an atlas to the label fusion is weighted according to the performance of the atlas segmentation,

$$L(c) = \underset{l \in \{l_1 \ldots l_K\}}{\mathrm{argmax}} \sum_{a=1}^{N'} w_a(c) \delta(\mathcal{L}_a(c), l), \tag{2}$$

where $w_a(c)$ is the weighting function for the ath atlas on pixel c. This idea of LWF has been implemented in many different fashions, either explicitly or implicitly (Heckemann et al., 2006; Isgum et al., 2009; Coupe et al., 2010; Langerak et al., 2010; Sabuncu et al., 2010; Wolz et al., 2013; Eskildsen et al., 2012; Wang et al., 2013; Bai et al., 2013; Hu et al., 2014).

In patch-based LWF, one computes the local appearance similarity, such as the intensity difference, within a small *patch* between the target and atlas images to estimate the local weights (Sabuncu et al., 2010; Bai et al., 2013; Wang et al., 2013). These methods can also adopt a patch-search process to correct local misalignments, particularly when the nonrigid algorithm is not adopted for the atlas-to-target registration. For estimating the local weights, Wang et al. (2013) argued that different atlases may produce similar labeling errors, which may induce erroneous labeling in local patches due to the accumulated false votes. They proposed the JLF, which estimates local weights by minimizing the expectation of labeling errors, which are computed based on the intensity difference between the atlas and the target image.

Intensity difference is generally reliable in single-modality image segmentation or in the applications where the assumption holds after applying intensity normalization to the images. While to tackle multi-modality images such as cardiac CT and MRI, the information theoretic measures are needed (Bai et al., 2013; Wang et al., 2013; Iglesias et al., 2013; Awate and Whitaker, 2014).

2.1.1. Joint label fusion

In this section, we extend the JLF for the application of multi-modality atlases whose performance are compared with our proposed method in the experiments. JLF estimates the local weight of a patch by minimizing the expectation of labeling errors on pixel c (Wang et al., 2013):

$$\begin{aligned} \mathbf{w}_c^\star &= \underset{\mathbf{w}_c}{\mathrm{argmin}} \, E\left[\delta(L_{\mathrm{gt}}(c), L(c))^2 | I, \mathcal{A}_1, \ldots, \mathcal{L}_N\right] \\ &= \underset{\mathbf{w}_c}{\mathrm{argmin}} \sum_{a=1}^{N} \sum_{b=1}^{N} w_a(c) w_b(c) \mathbf{M}_c(a, b) \\ &= \underset{\mathbf{w}_c}{\mathrm{argmin}} \, \mathbf{w}_c^t \mathbf{M}_c \mathbf{w}_c, \end{aligned} \tag{3}$$

where L_{gt} is the unknown ground truth; $\delta(L_{\mathrm{gt}}(c), L(c))$ denotes the labeling error on c; $\mathbf{w}_c = [w_1(c), \ldots, w_N(c)]$ is the vector of local weights and $\sum_{a=1}^{N} w_a(c) = 1$; $\mathbf{M}_c$ is a pairwise dependency matrix, whose element $\mathbf{M}_c(a, b)$ denotes the expected label error of the two atlases a and b and represents the likelihood of the two atlas both providing wrong labeling on position c. For simplicity, we use $\mathcal{E}_a(c) = \delta(L_{\mathrm{gt}}(c), \mathcal{L}_a(c))$ to denote the labeling error of atlas a on c:

$$\mathbf{M}_c(a, b) = E[\mathcal{E}_a(c)\mathcal{E}_b(c) | I, \mathcal{A}_1, \ldots, \mathcal{L}_N]. \tag{4}$$

The dependency matrix $\mathbf{M}_c$ can be estimated using intensity similarity measures, such as intensity difference or correlation coefficient between patches (Wang et al., 2013). In M^{3}AS, new measures which can provide local similarity or dissimilarity at a location for images from different modalities are needed. In this work, we develop the following three measures to compute $\mathbf{M}_c$ for JLF-based M^{3}AS:

- JLF$_{\mathrm{in}}$ is based on the original JLF method by Wang et al. (2013), where the intensity of a patch is first normalized by the mean and standard deviation of the intensity values of the patch. The dependency matrix $\mathbf{M}_c$ is then computed based on the intensity normalized patches $\{\mathcal{P}\}$,

$$\mathbf{M}_c(a, b) = \sum_{x \in \mathcal{P}} |\mathcal{P}_I(x) - \mathcal{P}_a(x)| \times |\mathcal{P}_I(x) - \mathcal{P}_b(x)|. \tag{5}$$

- JLF$_{\mathrm{ei}}$ first computes the entropy images of the target image and the atlas intensity images; then, the target and atlas images are replaced by their corresponding entropy images for the computation of $\mathbf{M}_c$; finally, the joint label fusion for each pixel is evaluated in a similar manner as JLF$_{\mathrm{in}}$ except that $\mathbf{M}_c$ in (5) is calculated from the difference of entropy values.

The entropy images are computed as follows (Wachinger and Navab, 2012): For each pixel in a cardiac image, a local patch centered on the pixel is selected. For each of these patches, one
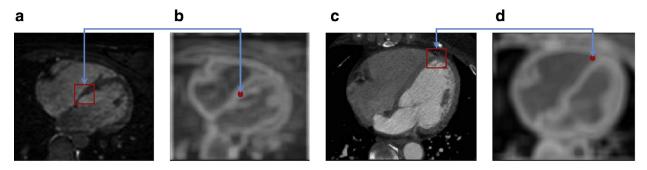
Fig. 2. Examples of entropy image (EI): (a) an axial-view image of a cardiac MRI; (b) the corresponding EI of (a); (c) an axial-view image of a cardiac CT; (d) the corresponding EI of (c). The highlighted square patch and pixel (both in red color) illustrate that the entropy value of the patch is assigned to the pixel at the corresponding location. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

can compute its intensity entropy and store this entropy value in the location of the pixel to create a new image, i.e. the entropy image. In this work, we compute patch entropy using the Parzen-window method (Thevenaz and Unser, 2000). Fig. 2 provides the examples and demonstrates the idea.

- JLF$_{cp}$ computes the expected labeling error of an atlas based on a local similarity measure, i.e. the conditional probability of the target image given the atlas intensity image,

$$P(\mathcal{E}_a(x)) = P(I(x)|\mathcal{A}_a(x)). \quad (6)$$

The labeling error is then embedded into (4) to compute the dependency matrix $\mathbf{M}_c$. The local similarity measure using (6) is derived from the generative model, which is presented in Section 2.1.2.

2.1.2. Estimating local similarity based on the conditional probability of images

Sabuncu et al. (2010) proposed a generative model $m: \Omega \rightarrow \{1, \ldots, N\}$, assuming that the target image is generated from one or more atlas images, and formulated the multi-atlas label fusion via maximum-a-posteriori (MAP) estimation,

$$
\begin{aligned}
L^\star &= \operatorname*{argmax}_L P(L|I, \{\mathcal{A}_a, \mathcal{L}_a\}) \\
&= \operatorname*{argmax}_L P(L, I|\{\mathcal{A}_a, \mathcal{L}_a\}) \\
&= \operatorname*{argmax}_L \sum_m P(m)P(L, I|m, \{\mathcal{A}_a, \mathcal{L}_a\}). \quad (7)
\end{aligned}
$$

Bai et al. (2013) extended the generative model to a patch-based model, assuming that each pixel of the target image is generated from a corresponding pixel in one of the atlas image. They introduced a random vector field to represent this mapping, $m = [m_n, m_p]^t: \Omega \rightarrow \{1, \ldots, N\} \times \{1, \ldots, |\Omega_{\text{Psrc}}|\}$, where Ω_{Psrc} denotes the patch search region for pixel c. Given the mapping field m, one can have two assumptions: First, the target pixels become conditionally independent, providing $P(L, I|m, \{\mathcal{A}_a, \mathcal{L}_a\}) = \prod_{x \in \Omega} P(L(x), I(x)|m(x), \{\mathcal{A}_a, \mathcal{L}_a\})$. Second, the intensity distribution and label distribution of the target image are conditionally independent, leading to $P(L(x), I(x)|m(x), \{\mathcal{A}_a, \mathcal{L}_a\}) = P(I(x)|m(x), \{\mathcal{A}_a\})P(L(x)|m(x), \{\mathcal{L}_a\})$.

In atlas-based segmentation, one can apply a fully nonrigid registration to achieve accurate mapping between the target image and the atlas image. For a MAS method using this registration, the label fusion is performed with the knowledge of the corresponding mapping field, i.e. $P(m)$ is a constant, and $P(L(x)|m(x), \{\mathcal{L}_a\})$ is 1 if $m(x) \in \Omega_{\text{Psrc}}$ and 0 otherwise. As a result, the maximization problem of (7) can be achieved by estimating the label for each pixel independently. The label of pixel c is then computed as follows,

$$
\begin{aligned}
L(c) &= \operatorname*{argmax}_{l \in \{l_1 \ldots l_K\}} \sum_{m_n=1}^{N} \sum_{m_p \in \Omega_{\text{Psrc}}} P(I(c)|m(c), \{\mathcal{A}_a\})\delta(\mathcal{L}_{m_n}(c), l) \\
&= \operatorname*{argmax}_{l \in \{l_1 \ldots l_K\}} \sum_{m_n=1}^{N} \sum_{m_p \in \Omega_{\text{Psrc}}} P(I(c)|\mathcal{A}_{m_n}(m_p))\delta(\mathcal{L}_{m_n}(c), l). \quad (8)
\end{aligned}
$$

Let $a = m_n$, (2) is a specific form of (8) where the search range is constrained to the location of c. The local weight becomes the conditional probability,

$$w_a(c) = P(I(c)|\mathcal{A}_a(c)). \quad (9)$$

This conditional probability represents the local similarity between the two images at location c.

2.2. Hierarchical local atlas ranking based on multi-scale patch

Generally, a patch-based method computes the local similarity between two images within a patch. One can increase the patch size and then compute the similarity within a larger region to involve more global structural information of the images. This method however leads to the following three problems:

First, the conventional patch scheme adopts only one level for patch-based atlas ranking. A multi-level patch strategy for hierarchical local atlas ranking is desired in the MAS label fusion.

Second, the runtime of computing the local patch similarity is proportional to the volume of the patch. For instance, to include the information 10 pixels away from the central pixel, one needs to adopt a patch size of $21 \times 21 \times 21$ pixel, whose computational load is 74 ($21^3/5^3$) times compared with that of a patch with size of $5 \times 5 \times 5$ pixel.

Third, computing the patch similarity based on a large area results in a significantly reduced contribution of the *center pixel* to the computation of the patch similarity. The center pixel refers to the pixel whose label is to be evaluated. This is a dilemma where including global information by increasing the size of a patch is contradictory to the maintaining of the local information around the central pixel.

To address these problems, we propose the MSP and develop the multi-level hierarchical local atlas ranking for the patch-based label fusion. A spatially varying weight (SVW) scheme is also invented to tackle the third issue. The framework is illuminated in Fig. 3.

MSP is based on the scale-space theory, a framework for multi-scale signal representation that is widely developed and applied to feature extraction and image matching (Lindeberg, 1998). Using this theory, one can handle image structures at different scales, by representing an image as a one-parameter family of smoothed images using a kernel function with a scale parameter. Different scales provide different information to the structures of the image.
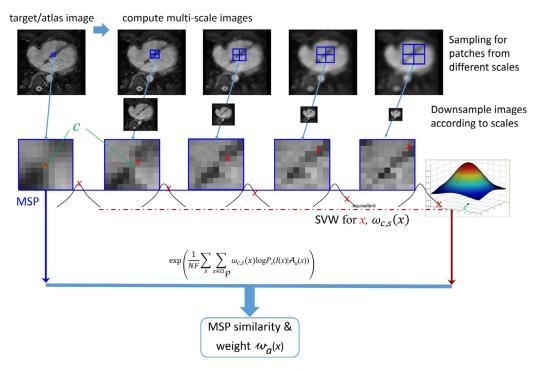
Fig. 3. Diagram illustrating the multi-scale patch (MSP) adopting the spatially varying weight (SVW) scheme, and the multi-level hierarchical atlas ranking. Here, c is the center pixel of the patch whose label is to be evaluated; x, in red color, is the pixel whose contribution to the patch similarity is to be calculated. The weight of x is determined according to its distance to c in the scaled image grid. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

The higher scale image suppresses fine structure but captures more global structural information within a certain vision window, while the lower scale image provides more local details of the structure. Here, we adopt the Gaussian scale space, where the Gaussian kernel function is used and the scale is defined by the standard deviation of the Gaussian (Lindeberg, 2011).

The following steps are employed to generate the MSP with N_s scales and patch volume $\Omega_{\mathcal{P}}$.

(1) We first convolve the original images with N_s scale parameters to generate the multi-scale images (MSIs), $\{I^s\}$ for the target image and $\{\mathcal{A}_a^s\}$ for the atlases. The standard deviations of the Gaussian kernel function for the N_s scales are set as $\sigma_1=0$, $\sigma_2=1$, ..., $\sigma_{N_s}=N_s-1$. Here, $\sigma_1=0$ means no Gaussian convolution is applied for the first level.

(2) For each scale, we generate the patches centered on the pixels to be evaluated. All these patches have the same number of pixels, as Fig. 3 shows. According to the scale-space theory, the higher level scale the patch is generated from, the more global structural information the patch has. This is achieved by using the larger field-of-view and sparser sampling for the patches coming from the image spaces with larger scales. In this work, we set the sample spacings from level-1 to level-N_s scale spaces as $\nu_1 = 1$ pixel, $\nu_2 = 2$ pixels, ..., and $\nu_{N_s} = N_s$ pixels of the original images, as illustrated in Fig. 3, where the blue grids indicate the sample grids of the patches. This scheme is equivalent to the strategy which first down-samples the MSIs according to their scales and then extracts the patches using the same sampling strategy. The reader is referred to Fig. 3 for illustration.

(3) Finally, the MSP is defined using the N_s-level patches from the multiple scale spaces, based on which we compute the local patch similarity.

The conditional probability of images, $P(I(x)|\mathcal{A}_a(x))$ in (9), is used to compute the local similarity. By assuming the conditional independency at each location, the conventional patch similarity can be computed,

$$S(I(\mathcal{P}), \mathcal{A}_a(\mathcal{P})) = \left(\prod_{x \in \Omega_{\mathcal{P}}} P(I(x)|\mathcal{A}_a(x)) \right)^{1/|\Omega_{\mathcal{P}}|}$$

$$= \exp\left(\frac{1}{|\Omega_{\mathcal{P}}|} \sum_{x \in \Omega_{\mathcal{P}}} \log P(i_x|j_x) \right). \quad (10)$$

where $i_x = I(x)$ and $j_x = \mathcal{A}_a(x)$. Here, the patch similarity resembles the geometric mean of the local similarity of all the pixel pairs in the patch.

For the MSP, we similarly compute the geometric mean of the local similarity from each scale,

$$S_{\mathrm{msp}} = \exp\left(\frac{1}{N_F} \sum_{s=1}^{N_s} \sum_{x \in \Omega_{\mathcal{P}}} \omega_{c,s}(x) \log P(i_x|j_x) \right), \quad (11)$$

where $s = 1, \ldots, N_s$ indicates the index of MSIs; $i_x=I^s(x)$ and $j_x=\mathcal{A}_a^s(x)$; N_F is the normalization factor; and $\omega_{c,s}(x)$ is the SVW which is described in detail in Section 2.2.1.

2.2.1. Spatially varying weight

In patch-based label fusion, one can compute the local patch similarity using (10), where each pixel pair is treated equally. As discussed above, this equal-weight-computation can dramatically reduce the contribution of the center pixel whose label is to be decided based on the local patch similarity. For example, for a patch size $3 \times 3 \times 3$ pixel, the contribution of the center pixel to the patch similarity computation is one 27th; and when the patch size increases to $5 \times 5 \times 5$ pixel, the contribution dramatically reduces to be one 125th. Since the patch similarity is computed for evaluating the labeling of the center pixel, a new scheme emphasizing the importance of the center pixel is needed.
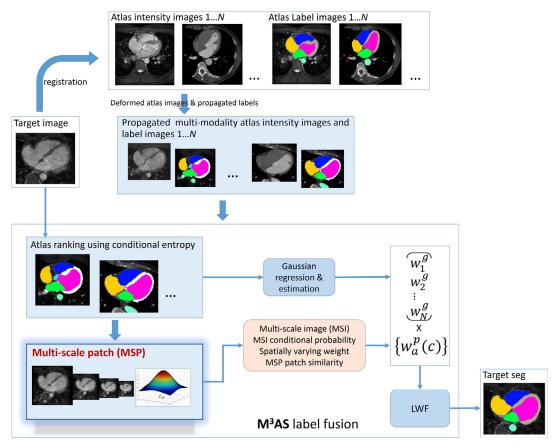
Fig. 4. Diagram demonstrating the new label fusion framework of multi-modality multi-atlas segmentation.

We propose the SVW scheme, where the contribution of a pixel to the computation of local similarity decreases monotonically with respect to the increased distance of the pixel to the center pixel. The local similarity of (10) is then revised as follows,

$$S_{svw} = \exp\left(\frac{1}{N_F} \sum_{x \in \Omega_P} \omega(|x-c|) \log P(i_x|j_x) \right), \tag{12}$$

where c is the center pixel; $\omega(d) = \text{Gaussian}(d, 0, \sigma_{svw})$, assuming the contribution of pixels to the similarity measure be Gaussian distributed and σ_{svw} is the controlling parameter. For MSP in (11), the SVW is defined according to the sample spacing for each level, i.e., $\omega_{c,s}(x) = \omega(\frac{|x-c|}{v_s})$. Here v_s is the patch sample spacing for each scale level.

2.2.2. Estimating conditional probability

The conditional probability can be computed from the quotient of the joint probability and marginal probability,

$$P(i|j) = \frac{P(i,j)}{P(j)} = \frac{P(i,j)}{\sum_i P(i,j)}. \tag{13}$$

Conventionally, the joint intensity distribution of images at scale level s can be estimated from the image pair using the Parzen window estimation (Thevenaz and Unser, 2000),

$$P(i,j) = \frac{1}{|\Omega|} \sum_{y \in \Omega} \beta(i, I^s(y)) \beta(j, \mathcal{A}_a^s(y)), \tag{14}$$

where $\beta(\cdot)$ is the Parzen window estimation function.

However, the sample size from one image pair can be small, and the standard error of the estimated joint distribution using (14) can be large. Consequently, the joint probability of intensity

pair (i, j) can have different values when estimated from different image pairs, namely from different atlases, particularly when multi-modality atlas images are involved.

We therefore propose to estimate the joint intensity distribution between the target image and atlas images using the sample points from all the atlases for each scale space:

$$P(i,j) = \frac{1}{N|\Omega|} \sum_{a=1}^{N} \sum_{y \in \Omega} \beta(i, I^s(y)) \beta(j, \mathcal{A}_a^s(y)). \tag{15}$$

In this work we adopt the cubic spline kernel function for the Parzen window function (Thevenaz and Unser, 2000).

2.3. Multi-modality multi-atlas segmentation (M^3AS)

Fig. 4 provides the diagram of the proposed M^3AS. In this framework, we first register all the atlases to the target image using a comprehensive nonrigid registration algorithm specially designed for whole heart images (Zhuang et al., 2010). Then, we employ the proposed MSP to compute the multi-level patch-based local weights and adopt an atlas ranking scheme to calculate the global weights. Finally we combine the two weighting schemes for LWF.

2.3.1. Atlas ranking for global weight

A number of works have demonstrated the advantages of ranking and selecting a subset of atlases, instead of using all for label fusion (Rohlfing et al., 2004; Heckemann et al., 2006; Aljabar et al., 2009; Bai et al., 2013). This is because some atlases can generate poor segmentation results which affect the consensus in fusing multiple classifiers. One can rank the atlases based on the expected performance and then select a subset with *better* performance. The

conventional schemes compute an intensity-based similarity measure, such as mutual information (MI) or normalized mutual information (NMI), as the ranking criterion. These methods assume that the similarity measure indicate the expected segmentation quality, which may not always hold, particularly when multi-modality atlases are involved.

In this work we employ the conditional entropy (CE) for atlas ranking (Zhuang et al., 2015),

$$w_a^g \propto -\text{CE} = -H(I|\mathcal{L}_a), \tag{16}$$

where w_a^g denotes the global weight and $H(\cdot)$ is the entropy function. CE measures the uncertainty inherent in the intensity distributions of the target image, conditioned on the propagated atlas label image. Compared with the intensity-based similarity measures, it provides a mechanism to compute the relationship between the target image and the atlas segmentation and thus is more effective in representing the segmentation performance of an atlas.

By considering label l as a random variable, one can derive the CE metric in forms of entropy and MI as

$$\text{CE} = \sum_{i \in I} \sum_l p(i, l) \log \frac{p(i, l)}{p(l)} = H(I) - \text{MI}(I, \mathcal{L}_a), \tag{17}$$

where $p(i, l)$ is the joint probability of the intensity value i and label l, and $p(l)$ is the marginal probability of label l. Since the field-of-views and overlap regions between the target image and the atlas segmentation result are determined by the target image, the entropy of the target image, $H(I)$, can be regarded as being constant for all atlases. The global ranking metric is therefore proportional to the minus CE or MI measure,

$$w_a^g \propto R_a, \tag{18}$$

where $R_a = -\text{CE}(I, \mathcal{L}_a)$ or $R_a = \text{MI}(I, \mathcal{L}_a)$.

To compute the global weights, we use the Gaussian kernel function, assuming a truncated normal distribution at μ_g of the atlas ranking values,

$$w_a^g = \text{Gaussian}(\mu_g, \sigma_g, R_a). \tag{19}$$

The mean and standard deviation, μ_g and σ_g, can be estimated by regressing this Gaussian kernel function to the ranking values from all atlases.

2.3.2. Label fusion for M^3AS

The global weight from atlas ranking and the local weight from multi-level MSP are combined to compute $w_a(c)$ for LWF in (2),

$$w_a(c) = w_a^g \times w_a^p(c), \tag{20}$$

where $w_a^p(c) = S_{\text{msp}}$ is computed using (11). Here, we use a product of the global and local weights, assuming that the vote from an atlas should get high weight if and only if both of the global atlas ranking and MSP-based local atlas ranking achieve large values. This is because we have a large number of atlases, 30 subjects in this work, thanks to the usage of multi-modality atlases.

3. Experiments and results

3.1. Materials

Twenty cardiac whole heart volumetric MRI data, provided by the Imaging Division at King's College London, were used. The MRI sequence was the balanced steady state free precession (b-SSFP) for whole heart imaging. A 3D triggering b-SSFP turbo field echo (TFE) sequence, with arrhythmia rejection, was modified to enable the imaging at the end diastolic phase. The sequence was implemented on a 1.5T clinical scanner (Philips Healthcare, Best, The Netherlands) equipped with 32 independent receive channels. A

fat saturation and T2 preparation pulses were used to null fat and to increase the contrast between blood and cardiac muscle, and a navigator beam before data acquisition was implemented to enable the free-breathing scan. All the data were acquired at about $2 \times 2 \times 2$ mm and reconstructed to around $1 \times 1 \times 1$ mm.

Another twenty contrast enhanced cardiac CT data were used. All the data were obtained from two state-of-the-art 64-slice CT scanners (Philips Medical Systems, Netherlands) using a standard coronary CT angiography protocol at two sites affiliated to Shanghai Shuguang Hospital. Images were acquired in the axial view, covering the whole heart from the upper abdominal to the aortic arch. The in-plane resolution was about 0.44×0.44 mm and the average slice thickness was 0.60 mm.

Thirty of the 40 subjects had cardiovascular diseases, including cardiac function insufficiency (NYHA II), cardiac edema, hypertension (III), sick sinus syndrome, arrhythmia, atrial flutter, atrial fibrillation, artery plaque, coronary atherosclerosis, aortic aneurysm (dilated aorta), Tetralogy of Fallot (right ventricle hypertrophy), dilated cardiomyopathy (left ventricle), aortic stenosis, pulmonary artery stenosis. Some patients had a combination of several different types of pathologies. The dataset demonstrates a wide variety of pathologies and heart shapes. Fig. 1 provides an MRI case and a CT case.

3.1.1. Manual segmentation and atlas construction

Seven substructures were of interest in the WHS study, including:

(1) the left ventricular cavity (LV);
(2) the right ventricular cavity (RV);
(3) the left atrial cavity (LA);
(4) the right atrial cavity (RA);
(5) the myocardium of the left ventricle (Myo) where the epicardium (Epi) is assessed in the evaluation of surface delineation;
(6) the ascending aorta trunk from the aortic valve to the superior level of the atria;
(7) the pulmonary artery (PA) trunk from the pulmonary valve to the bifurcation point.

The substructures of interest in each image were manually labeled to generate the atlas label map as well as the gold standard for validation. The descending aorta was also labeled, separately from the ascending aorta, for assisting the atlas-to-target registration (Zhuang et al., 2010), though it was not of interest and its segmentation would not be evaluated in this work. The manual labeling was performed by well-trained students majoring in biomedical engineering or medical physics. Each result was double checked and refined by an observer with expertise in cardiac anatomy. During manual segmentation, the 3D data were re-orientated into the traditional view of the heart, meaning the images were displayed in the short-axis and long-axis views. The observers employed the brush tool in the software ITK-SNAP (Yushkevich et al., 2006) to manually label each substructure slice-by-slice. The manual segmentation took six to ten hours per case. Fig. 1 shows two examples.

3.1.2. Evaluation

Two-fold cross validation strategy was used, by dividing the 20 MRI data into two sets, one be the test set and the other be the atlas set for both single-modality MAS or multi-modality MAS. All of the twenty cardiac CT data were used as atlases to form a thirty-subject, multi-modality atlas set.

For segmentation evaluation, Dice score (DS) (Kittler et al., 1998), Jaccard index (JI) (Jaccard, 1901) and surface-to-surface distance (SD) were used. For WHS evaluation, we adopted the generalized version of DS, JI and SD, which are normalized with
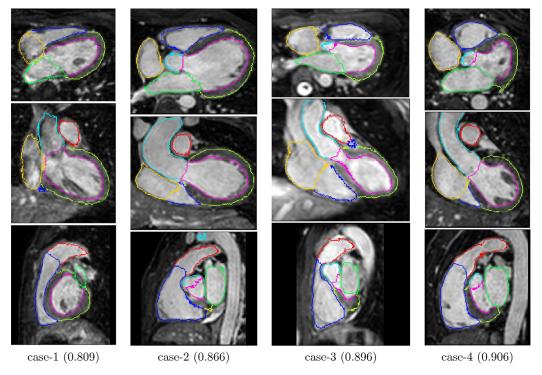
case-1 (0.809)   case-2 (0.866)   case-3 (0.896)   case-4 (0.906)

Fig. 5. Three orthogonal views of four segmentation cases with the WHS Dice scores in brackets, illustrating performance of the proposed segmentation method: case 1 and case 2 are the two worst cases among all the test subjects in terms of WHS Dice scores, and cases 3 and 4 are the two median cases.

Table 1
This table provides the Dice scores of the multi-modality MAS using majority vote label fusion (MVF), the single-modality MAS using the proposed label fusion algorithm (smMAS), the proposed M^3AS, and the inter-observer (Inter-Ob) and intra-observer (Intra-Ob) variations. The Jaccard index (JI) and surface distance (SD) of the proposed M^3AS are also provided.

Dice (%)	LV	Myo	RV	LA	RA	aorta	PA	WHS
MVF	94.1 ± 1.06	80.9 ± 3.87	88.3 ± 5.38	84.2 ± 7.92	84.7 ± 8.78	86.1 ± 7.95	73.6 ± 10.23	86.9 ± 3.99
smMAS	93.9 ± 2.43	84.1 ± 4.77	90.6 ± 4.94	87.5 ± 6.14	87.7 ± 5.39	88.7 ± 4.82	82.0 ± 8.08	89.2 ± 3.75
M^3AS	94.6 ± 1.67	85.2 ± 3.49	91.4 ± 4.48	87.8 ± 6.24	88.5 ± 5.29	89.1 ± 4.99	81.7 ± 8.41	89.9 ± 3.40
Inter-Ob	93.7 ± 1.33	81.1 ± 2.90	90.1 ± 1.96	83.7 ± 4.58	85.8 ± 3.10	87.6 ± 5.24	76.3 ± 14.34	87.8 ± 1.36
Intra-Ob	94.2 ± 0.84	83.9 ± 1.23	91.2 ± 2.59	86.8 ± 3.23	87.2 ± 2.48	91.1 ± 1.65	82.6 ± 3.77	89.5 ± 1.03
M^3AS								
JI (%)	89.8 ± 2.99	74.4 ± 5.28	84.4 ± 7.22	78.8 ± 9.05	79.7 ± 8.16	80.7 ± 7.62	69.8 ± 11.59	81.8 ± 5.49
SD (mm)	0.85 ± 0.72	1.23 ± 1.05	0.99 ± 0.99	1.34 ± 1.28	1.37 ± 1.44	0.85 ± 0.79	0.93 ± 0.85	1.09 ± 1.11

respect to the size of substructures, providing more objective measurements (Crum et al., 2006; Zhuang, 2013). *All the images, including the target image, atlas images and segmentation results, were resampled into images with pixel size* $1 \times 1 \times 1$ *mm.* Hence, a patch or search range with size $a \times b \times c$ pixel is equivalent to the size of $a \times b \times c$ mm.

All the algorithms, including the atlas-to-target registration and label fusion, were implemented on a Lenovo ThinkStation D30 workstation. The implementations were based on single thread for registration and label fusion, but the multiple tasks of atlas-to-target registration were run simultaneously on the workstation which had four Intel Xeon E5-2667 V2 CPUs and 32 cores.

3.2. Performance of the proposed method

Fig. 5 provides the visualization of four segmentation cases by the proposed M^3AS method. This M^3AS method employed the global weighting and the MSP-based local atlas ranking for label fusion. The σ_{svw} in (12) was set to 1 mm and N_s in (11) was set to 5.

Table 1 presents the Dice scores of the single-modality MAS (smMAS), the conventional MAS using MVF (MVF), the inter-

observer and intra-observer variations for comparisons with the proposed M^3AS. M^3AS was better than MVF in all categories with statistical significance using the paired, two-tailed t-test ($p < 0.05$). M^3AS was generally better than smMAS such as in WHS Dice ($p = 0.007$), but the gain was limited due to the fact that the atlas-to-target registration method performed better in single-modality images than in inter-modality images. Note that the gain of M^3AS was more evident in the segmentation of myocardium, which is the particularly challenging and important task in cardiac segmentation.

We selected six subjects for the inter-observer variation study. The WHS Dice scores of these cases by M^3AS were the medians of the twenty cases. For the intra-observer study, we selected four subjects based on the similar criterion. As Table 1 shows, Dice scores of the inter- and intra-observer variations were worse than those of the M^3AS. This was probably because the label fusion algorithm tended to produce a *mean* of the multiple classifiers. The volume overlap between the mean and each of the classifiers can be greater than the overlap between the classifiers themselves.

The Jaccard indices and SD of the proposed M^3AS are provided in Table 1. The average runtime was 2.82 min for the atlas-to-target registration and 12.58 min for the proposed label fusion.
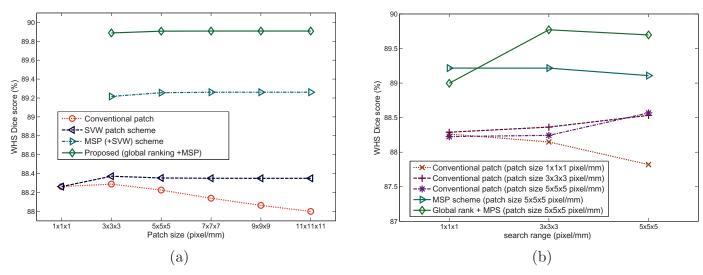
Fig. 6. Patch-based label fusion for multi-modality multi-atlas segmentation: (a): the mean whole heart segmentation Dice scores using the four different patch schemes, with different patch sizes; (b): the Dice scores of the different patch schemes, with different search ranges.
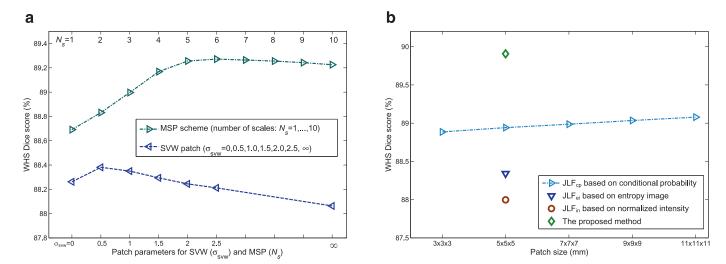

Fig. 7. Study parameterization of the spatially varying weight (SVW) patch scheme and the multi-scale patch (MSP) scheme.

3.3. Study of patch strategies

Fig. 6 (a) presents the mean WHS Dice scores of the four patch schemes, with different patch sizes from $1 \times 1 \times 1$ pixel/mm to $11 \times 11 \times 11$ pixel/mm. The conventional patch scheme achieved the optimal performance with patch size $3 \times 3 \times 3$ pixel. The performance became worse when further increasing the patch size, due to the dramatically reduced contribution of the center pixel whose label is to be evaluated. This problem was solved by the SVW patch scheme, whose performance converged from the patch size $5 \times 5 \times 5$ pixel. The MSP further greatly improved the label fusion performance whose WHS Dice scores also converged from patch size $5 \times 5 \times 5$ pixel, thanks to the usage of SVW. Notice that the MSP strategy could not adopt a patch with size $1 \times 1 \times 1$ pixel. Finally, the proposed label fusion scheme, combining the global atlas ranking and the MSP-based local ranking, generated the best performance among the four patch-based schemes.

Fig. 6 (b) provides the mean WHS Dice scores of the five schemes with different search ranges, from $1 \times 1 \times 1$ pixel to $5 \times 5 \times 5$ pixel. The MSP scheme adopted 5 scales and the proposed method (global rank + MPS) adopted 3 scales. The figure shows that increasing the search range, which is believed to increase the potential of correcting local misalignments, does not guarantee a

significant improvement of the segmentation performance. Particularly, in the conventional patch scheme with patch size $1 \times 1 \times 1$ pixel the mean WHS Dice scores became worse with respect to the increased search range. This was probably due to the fact that the atlas-to-target registration was a fully deformable method and adopted the NMI similarity measure based on global image information to provide a good alignment. By contrast, the patch-match computed patch similarity solely based on the local information of the images. Since the whole heart images had indistinct boundaries between local substructures, e.g. the intensity values among the blood pools of the four chambers and great vessels were similar, the patch-match became more likely to step into local optima, compared to the deformable registration. More importantly, the computation time of the label fusion increased dramatically with respect to the increased size of search radius.

Fig. 7 (a) illustrates the performance of the MSP-based label fusion, with respect to the different values of the key parameters, i.e. σ_{svw} for the SVW patch and N_s (the number of scales) for the MSP scheme. The SVW patch-based label fusion adopted a $9 \times 9 \times 9$ pixel patch. When $\sigma_{\text{svw}}=0$, the method is equivalent to the patch scheme with size $1 \times 1 \times 1$ pixel, and when $\sigma_{\text{svw}}=\infty$, the method is equivalent to the conventional scheme without SVW. One can see that SVW obtained the optimal performance when $\sigma_{\text{svw}} \in [0.5,$

1]. For MSP, the label fusion performance was first improved evidently, and then converged when $N_s \geq 5$.

3.4. Study and comparison with joint label fusion

Fig. 7 (b) presents the mean WHS Dice scores of the three JLF methods (described in Section 2.1.1). At the patch size of $5 \times 5 \times 5$ pixel, JLF_{in} (WHS Dice score $0.880 \pm .0355$) was worse than JLF_{ei} (WHS Dice score $0.883 \pm .0310$) with statistical significance ($p = 0.014$); JLF_{ei} was worse than JLF_{cp} (Dice score $0.889 \pm .0379$) with statistical significance ($p = 0.007$). Using the Bonferroni correction, we obtained the performance ranking of the three JLF methods, from the worst to the best as JLF_{in}, JLF_{ei}, JLF_{cp}, with statistical significance ($p < 0.05$). Compared to the conventional intensity and entropy based measures, the proposed algorithm based on conditional probability of images is a better local similarity measure in multi-modality images.

Fig. 7 (b) also provides the mean WHS Dice score of the proposed method, 0.899 ± 0.0340, which was significantly better than JLF_{cp} ($p = 0.0193$).

4. Discussion and conclusion

This work presents a new whole heart segmentation (WHS) method using multi-modality atlases, i.e., M^3AS, for cardiac MRI. The label fusion algorithm is based on the proposed multi-scale patch (MSP) and a new global atlas ranking scheme. MSP uses the information of images from multi-scale space and thus is able to capture different levels of the structural information of images for hierarchical local atlas ranking. Also, MSP adopts the spatially varying weight (SVW) scheme to leverage the contribution of each pixels to the local similarity computation in the patch-based label fusion. The global atlas ranking estimates the weights using a truncated Gaussian kernel regression, where the atlas ranking value is computed based on the conditional entropy measure. The resulting global atlas weighting scheme avoids the difficulty of manually determining the optimal number of atlases selected for label fusion.

The proposed M^3AS yielded a mean WHS Dice score of 0.899 ± 0.0340, Jaccard index of 0.818 ± 0.0549, surface distance error of 1.09 ± 1.11 mm for the 20 cardiac MRI data, where we adopted a two-fold cross validation strategy and further included 20 cardiac CT data to form a 30-subject atlas database. The average runtime for the label fusion was 12.58 min. The WHS accuracy of M^3AS was significantly better than that of the conventional multi-atlas segmentation (MAS) method ($p < 0.001$) in all the three error metrics.

The patch study showed that SVW improved the conventional patch scheme when the patch size was more than $1 \times 1 \times 1$ pixel and the improvement became increasingly evident with respect to the increased patch size. MSP yielded significantly better WHS Dice scores than the conventional patch scheme, for example 0.893 ± 0.0369 versus 0.882 ± 0.0371 ($p = 0.0016$) using patch size $5 \times 5 \times 5$ pixel and $N_s = 5$. The proposed patch scheme did not benefit evidently from the increased patch size after a patch size of $5 \times 5 \times 5$ pixel. The results of the patch-search study showed that increasing the search range did not necessarily improve the segmentation performance, due to the challenge of indistinct boundaries between local substructures.

The computational load of the MSP-based label fusion is approximately linear to the volume of the adopted patch size, volume of search range, and the number of multi-scale space, namely $\mathbf{O}(\text{MSP})=N_s \times |\Omega_{\mathcal{P}}| \times |\Omega_{\text{search}}| \times \mathbf{O}(\mathcal{P}_0)$. Here, $\mathbf{O}(\mathcal{P}_0)$ indicates the computational complexity of the MSP label fusion which adopts one single scale and no patch strategy. For example, the mean runtime of $\mathcal{P}_0$ was 22.0 s, which included about 20 s for computing the joint histograms between the target image and the 30 atlases

and 1–2 s for loading these images from the hard drive. The runtime of label fusion increased to 2.71 min after using a patch size of $5 \times 5 \times 5$ pixel, then further to 4.66 h with a search range $5 \times 5 \times 5$ pixel, and finally to 11.3 h by using $N_s = 3$ multi-scale space for the MSP. The last scheme achieved an average WHS Dice score of 0.897 ± 0.0352, which was comparable to 0.899 ± 0.0340 of the proposed scheme ($N_s = 5$, patch size $5 \times 5 \times 5$ pixel, no patch-search) which took less than 13 min.

Computation time is a common concern in clinical practice. Parallel computation, such as GPU programming, is an efficient strategy to reduce the runtime of label fusion. This will be our future work. Adopting a scheme with smaller patches and fewer levels of scales, without spoiling the accuracy, is another useful strategy. For example, one can achieve a clinical acceptable WHS Dice score of $.898 \pm .0342$ within three minutes using three scales ($N_s = 3$) and a patch scheme with size of $3 \times 3 \times 3$ pixel. When the number of atlases is significantly greater than the number of cores of the workstation, we recommend to adopt a multi-level atlas ranking scheme, by first using a fast atlas ranking method to select a subset of good atlases to build a subject-specific atlas database (Zhuang et al., 2015).

In the literature, Peters et al. (2009) evaluated their method on a set of 42 volumetric cardiac MRI images using a 4-fold cross validation strategy. The reported runtime for the tool was 10–20 s for one case and the mean distance between the automatically fitted mesh model and the gold standard mesh derived from manual adaptation of the model, was about 0.60–0.83 mm for different substructures. Zhuang et al. (2010) employed one atlas to segment 37 subjects involving 9 different pathologies. The mean Dice scores were about 0.77 to 0.92 and the computation time was about two to four hours. Kirisli et al. (2010) performed a MAS validation study using the leave-one-out cross validation scheme on 8 subjects. In this work, the authors did not report the performance of myocardium segmentation, which is the particularly challenging and important task of cardiac segmentation. The Dice scores of other substructures were about 0.89 to 0.95 and the mean runtime was 20 min.

A number of works employed the MAS and patch-based label fusion for other segmentation tasks of cardiac images. Bai et al. (2013) developed a probabilistic patch-based label fusion algorithm for myocardium segmentation. The authors proposed to iteratively refine the registration after each label fusion step, to further improve the performance of segmentation. In their work, the patch similarity was computed based on intensity difference between the target and atlas images. Hence, the image intensity of cardiac MRI needed to be normalized before performing the label fusion. Their MAS method was also applied to the right ventricle segmentation reported in (Petitjean et al., 2015). Zhuang et al. (2015) proposed an atlas ranking and selection criterion based on conditional entropy for MAS-based WHS of cardiac CT. The work demonstrated the importance of atlas ranking measures. However, similar to most of the existing label fusion algorithms, the number of atlases selected for label fusion was manually determined. Wang et al. (2014) proposed a learning-based algorithm to train a local classifier based on the random forest method for each patch of the target image. The local classifiers were then used to predict labels of the patches. The authors applied this method to the MAS of cardiac ultrasound images. Bai et al. (2015) explored patch-based label fusion with augmented features. The proposed fusion strategy was also a learning-based algorithm, i.e. using the support vector machine (SVM) approach to train local classifiers. The authors evaluated their proposed method using a set of 83 cardiac MRI data, and compared the performance with the conventional MVF and the state-of-art JLF. Their proposed method and the JLF generated similar accuracy for myocardium segmentation, and both of them performed evidently better than the MVF. Notice that due to the

difference in the test data sets, evaluation metrics and implementation, an objective inter-work comparison can be difficult.

In conclusion, the proposed M^{3}AS is able to take advantage of atlases from different modalities and outperforms the single-modality based MAS. The WHS method is fully automatic and robust. This method can generate accurate results, comparable to manual segmentation from experts, for geometrical modeling of the heart using cardiac MRI. Hence, it can be useful in clinical care of cardiovascular diseases.

Acknowledgment

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, 726–738.

Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solorzano, C., 2009. Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans. Med. Imaging 28, 1266–1277.

Awate, S., Whitaker, R., 2014. Multiatlas segmentation as nonparametric regression. IEEE Trans. Med. Imaging 33, 1803–1817.

Bai, W., Shi, W., Ledig, C., Rueckert, D., 2015. Multi-atlas segmentation with augmented features for cardiac MR images. Med. Image Anal. 19, 98–109.

Bai, W., Shi, W., O'Regan, D.P., Tong, T., Wang, H., Jamil-Copley, S., Peters, N.S., Rueckert, D., 2013. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. IEEE Trans. Med. Imaging 32, 1302–1315.

Coupe, P., Manjon, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L., 2010. Nonlocal patch-based label fusion for hippocampus segmentation. In: Medical Image Computing and Computer-Assisted Intervention, pp. 129–136.

Crum, W.R., Camara, O., Hill, D.L.G., 2006. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans. Med. Imaging 25, 1451–1461.

Ecabert, O., Peters, J., Walker, M., Ivanc, T., Lorenz, C., Berg, J., Lessick, J., Vembar, M., Weese, J., 2011. Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. Med. Image Anal. 15, 863–876.

Eskildsen, S.F., Coupe, P., Fonov, V., Manjon, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Qstergaard, L.R., Collins, D.L., 2012. BEaST: brain extraction based on nonlocal segmentation technique. Neuroimage 59, 2362–2373.

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

Holden, M., Griffin, L.D., Saeed, N., Hill, D.L.G., 2004. Multi-channel mutual information using scale space. Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 797–804.

Hu, S., Coupe, P., Pruessner, J.C., Collins, D.L., 2014. Nonlocal regularization for active appearance model: application to medial temporal lobe segmentation. Hum. Brain Mapp. 35, 377–395.

Iglesias, J.E., Sabuncu, M.R., Leemput, K.V., 2013. A unified framework for cross-modality multi-atlas segmentation of brain MRI. Med. Image Anal. 17, 1181–1191.

Isgum, 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 Trans. Med. Imaging 28, 1000–1010.

Jaccard, P., 1901. Étude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37, 547–579.

Kirisli, H., Schaap, M., Klein, S., Papadopoulou, S.L., Bonardi, M., Chen, C.H., Weustink, A.C., Mollet, N.R., Vonken, E.J., van der Geest, R.J., van Walsum, T., Niessen, W.J., Kirisli, H.a., 2010. Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: a large-scale, multicenter, and multivendor study. Med. Phys. 37, 6279–6291.

Kittler, J., Hatef, M., Duin, R., Matas, J., 1998. On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. 20, 226–239.

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 Trans. Med. Imaging 29, 2000–C2008.

Leutenegger, S., Chli, M., Siegwart, R.Y., 2011. BRISK: binary robust invariant scalable keypoints. In: Proceedings of International Conference on Computer Vision, pp. 2548–2555.

Lindeberg, T., 1998. Feature detection with automatic scale selection. Int. J. Comput. Vis. 30, 77–116.

Lindeberg, T., 2011. Generalized gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. J. Math. Imaging Vis. 40, 36–81.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

World Health Organization, 2011. Cardiovascular Diseases (CVDs) Fact Sheet No. 317.

Peters, J., Ecabert, O., Meyer, C., Kneser, R., Weese, J., 2009. Optimizing boundary detection via simulated search with applications to multi-modal heart segmentation. Med. Image Anal. 14, 70–84.

Petitjean, C., Zuluaga, M.A., Bai, W., Dacher, J.N., Grosgeorge, D., Caudron, J., Ruan, S., Ayed, I.B., Cardoso, M.J., Chen, H.C., Jimenez-Carretero, D., Ledesma–Carbayo, M.J., Davatzikos, C., Doshi, J., Erus, G., Maier, O.M., Nambakhsh, C.M., Ou, Y., Ourselin, S., Peng, C.W., Peters, N.S., Peters, T.M., Rajchl, M., Rueckert, D., Santos, A., Shi, W., Wang, C.W., Wang, H., Yuan, J., 2015. Right ventricle segmentation from cardiac MRI: a collation study. Med. Image Anal. 19, 187–202.

van Rikxoort, E., Isgum, 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. Med. Image Anal. 14, 39–49.

Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R., 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21, 1428–1442.

Rohlfing, T., Maurer, J.C.R., 2005. Multi-classifier framework for atlas-based image segmentation. Pattern Recognit. Lett. 26, 2070–2079.

Sabuncu, M., Yeo, B., Leemput, K.V., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. IEEE Trans. Med. Imaging 29, 1714–1729.

Shi, F., Yap, P.T., Fan, Y., Gilmore, J.H., Lin, W., Dinggang, 2010. Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. NeuroImage 51, 684–693.

Thevenaz, P., Unser, M., 2000. Optimization of mutual information for multiresolution image registration. IEEE Trans. Image Process. 9, 2083–2099.

Tong, T., Wolz, R., Gao, Q., Guerrero, R., Hajnal, J.V., Rueckert, D., 2014. Multiple instance learning for classification of dementia in brain MRI. Med. Image Anal. 18, 808–818.

Tong, T., Wolz, R., Wang, Z., Gao, Q., Misawa, K., Fujiwara, M., Mori, K., Hajnal, J.V., Rueckert, D., 2015. Discriminative dictionary learning for abdominal multi-organ segmentation. Med. Image Anal. 23, 92–104.

Vasan, R.S., Benjamin, E.J., Levy, D., 1996. Congestive heart failure with normal left ventricular systolic function clinical approaches to the diagnosis and treatment of diastolic heart failure. Arch. Intern. Med. 156, 146–157.

Wachinger, C., Navab, N., 2012. Entropy and Laplacian images: structural representations for multi-modal registration. Med. Image Anal. 16, 1–17.

Wang, H., Cao, Y., Syeda-mahmood, T., 2014. Multi-atlas segmentation with learning-based label fusion. In: Machine Learning in Medical Imaging, pp. 256–263.

Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A., 2013. Multi-atlas segmentation with joint label fusion. IEEE Trans. Pattern Anal. Mach. Intell. 35, 611–623.

Wang, L., Gao, Y., Shi, F., Li, G., Gilmore, J.H., Lin, W., Shen, D., 2015. Links: learning-based multi-source integration framework for segmentation of infant brain images. NeuroImage 108, 160–172.

Wolz, R., Chu, C., Misawa, K., Mori, K., Rueckert, D., 2013. Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE Trans. Med. Imaging 32, 1723–1730.

Wu, G., Kim, M., Sanroma, G., Wang, Q., Munsell, B.C., Shen, D., 2015. Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition. NeuroImage 106, 34–46.

Wu, Y., Liu, G., Huang, M., Guo, J., Jiang, J., Yang, W., Chen, W., Feng, Q., 2014. Prostate segmentation based on variant scale patch and local independent projection. IEEE Trans. Med. Imaging 33, 1290–1303.

Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage 31, 1116–1128.

Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S., 2011. Deformable segmentation via sparse shape representation. In: Medical Image Computing and Computer-Assisted Intervention, pp. 451–458.

Zhang, S., Zhan, Y., Metaxas, D.N., 2012. Deformable segmentation via sparse representation and dictionary learning. Med. Image Anal. 16, 1385–1396.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D., 2008. Four-chamber heart modeling and automatic segmentation for 3-d cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imaging 27, 1668–1681.

Zhuang, X., 2013. Challenges and methodologies of fully automatic whole heart segmentation: a review. J. Healthc. Eng. 4, 371–407.

Zhuang, X., Bai, W., Song, J., Zhan, S., Qian, X., Shi, W., Lian, Y., Rueckert, D., 2015. Multi-atlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection. Med. Phys. 42, 3822–3833.

Zhuang, X., Rhode, K., Razavi, R., Hawkes, D.J., Ourselin, S., 2010. A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Trans. Med. Imaging 29, 1612–1625.