Medical Image Analysis 48 (2018) 58–74

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Superpixel and multi-atlas based fusion entropic model for the segmentation of X-ray images

D.C.T. Nguyen[a,b,*], S. Benameur[b], M. Mignotte[a], F. Lavoie[b,c]

[a] Département d'Informatique & Recherche Opérationnelle (DIRO), Faculté des Arts et des Sciences, Université de Montréal, Montréal, Québec, Canada
[b] Eiffel Medtech Inc., Montréal, Québec, Canada
[c] Orthopedic Surgery Department, Centre Hospitalier de l'Université de Montréal (CHUM), Montréal, Québec, Canada

ARTICLE INFO

Article history:
Received 12 October 2017
Revised 9 May 2018
Accepted 11 May 2018
Available online 18 May 2018

Keywords:
Bone structures of the lower limb
Bone segmentation
Consensus segmentation
Multi-atlas segmentation
Superpixel map
Variation of information based fusion step
X-ray images

ABSTRACT

X-ray image segmentation is an important and crucial step for three-dimensional (3D) bone reconstruction whose final goal remains to increase effectiveness of computer-aided diagnosis, surgery and treatment plannings. However, this segmentation task is rather challenging, particularly when dealing with complicated human structures in the lower limb such as the patella, talus and pelvis. In this work, we present a multi-atlas fusion framework for the automatic segmentation of these complex bone regions from a single X-ray view. The first originality of the proposed approach lies in the use of a (training) dataset of co-registered/pre-segmented X-ray images of these aforementioned bone regions (or multi-atlas) to estimate a collection of superpixels allowing us to take into account all the nonlinear and local variability of bone regions existing in the training dataset and also to simplify the superpixel map pruning process related to our strategy. The second originality is to introduce a novel label propagation step based on the entropy concept for refining the resulting segmentation map into the most likely internal regions to the final consensus segmentation. In this framework, a leave-one-out cross-validation process was performed on 31 manually segmented radiographic image dataset for each bone structure in order to rigorously evaluate the efficiency of the proposed method. The proposed method resulted in more accurate segmentations compared to the probabilistic patch-based label fusion model (PB) and the classical patch-based majority voting fusion scheme (MV) using different registration strategies. Comparison with manual (gold standard) segmentations revealed that the good classification accuracy of our unsupervised segmentation scheme is, respectively, 93.79% for the patella, 88.3% for the talus and 85.02% for the pelvis; a score that falls within the range of accuracy levels of manual segmentations (due to the intra inter/observer variability).

1. Introduction

X-ray images are used by physicians all over the world for the preliminary diagnosis of several bone diseases, to plan surgical intervention, and for pre and post-operative treatments. In this context, accurate extraction of bone contours or regions from these 2D X-ray images, is often the preliminary, and also crucial, step for three-dimensional (3D) bone reconstruction which can then be a great help in determining the extent of a fracture or for improving the diagnosis, follow-up, and treatment of major bone diseases, such as osteoporosis and osteoarthritis.

X-ray segmentation of bone structures of the lower limb, especially of the pelvic region, remains a challenging problem due both to intrinsic and extrinsic difficulties. Intrinsic difficulties refer to the intrinsic properties of the X-ray imaging systems that result in imaging noise. The major source of noise that degrades image quality is from the X-ray source and is caused by radiation scattering and source leakage (quantum noise). Extrinsic difficulties are closely related to the patients. Indeed, bone structures in radiographic images often overlap with other bones or dense soft tissues, making it difficult to accurately delimit the boundaries of different bone regions. This is particularly true when neighboring structures have similar absorption rates. Also, we must include the effects of the density variation of the bone structure being imaged, the inter-subject variability in bone morphology, the variability in the 3D imaging poses, and the motion unsharpness (caused

* Corresponding author at: Département d'Informatique & Recherche Opérationnelle (DIRO), Faculté des Arts et des Sciences, Université de Montréal, Montréal, Québec, Canada.
E-mail addresses: dac.cong.tai.nguyen@umontreal.ca (D.C.T. Nguyen), sbenameur@eiffelmedtech.com (S. Benameur), mignotte@iro.umontreal.ca (M. Mignotte), flavoie@eiffelmedtech.com (F. Lavoie).

https://doi.org/10.1016/j.media.2018.05.006
1361-8415/

by movement of the patient). These problems are all the more difficult when dealing with large and complex bone shapes such as the human pelvis or small bone structures such as the talus or patella. In these cases, the segmentation of complex bones is usually still done manually or semi-automatically, which is time consuming and tedious.

In order to propose a fully automatic segmentation method of bone structures in X-ray images, some *a priori* anatomical knowledge of the region to be extracted is necessary and must be integrated in the segmentation model. This can be done in different ways, for example by modeling the grey level distributions associated with each tissue or region class to be segmented (Mignotte et al., 2000), in the form of deformable shape templates constrained by a family of parametric or non-parametric curves (Mignotte et al., 2001; Mignotte and Meunier, 2001), or by one or several prototype templates together with the set of admissible deformations (Destrempes et al., 2007).

Recently, an efficient example-based segmentation strategy has been proposed, which turns out to be an interesting alternative in order to incorporate prior information into the segmentation model. This alternative uses a training dataset of X-ray images with the corresponding pre-segmentations called a multi-atlas (Artaechevarria et al., 2009; Aljabar et al., 2009; Dowling et al., 2011). In this strategy, the unsupervised segmentation process is divided into two stages; a first registration procedure is conducted between the target image to be segmented and each image and/or segmentation of the multi-atlas. Next, from the set of co-registered segmentations (and possibly its corresponding X-ray image), a second *fusion* or *label propagation* step is finally achieved to estimate both the segmentation into regions and also to infer the most likely internal region labels to the final consensus segmentation result. Regarding the second stage of any multi-atlas based segmentation scheme, also called *label propagation* or *decision fusion* step, the commonly used techniques, already proposed in this context, are generally based on different types of voting rules, such as the majority voting (Rohlfing et al., 2004a), (possibly locally) weighted voting (Artaechevarria et al., 2009) in which weights can be possibly estimated by the Expectation-Maximization (EM) algorithm (Rohlfing et al., 2004b), or shape-based methods (Rohlfing and Maurer, 2007) to name a few.

The limits of these techniques are mainly due to the registration errors. Regions from atlases may be associated to the wrong regions in the target image. To address this limitation, a first strategy consists in using a combination of a linear rigid registration, followed by a nonlinear or local registration. The first step provides an initial rigid alignment, while the second step rather considers the specific non-rigid and/or nonlinear deformations of the target (for example, to take into account the inter-individual variability of the imaging pose and bone morphology). In this context, the nonlinear registration can be achieved in many ways. For example, by considering small perturbations (as parameters of the nonlinear transform) on landmark points of a preliminary rigid registration, or by considering nonlinear deformations of a grid of Bezier or BSpline points, using deformation field (optical flows), demons, finite element models, to name a few (Morin et al., 2012). In addition to the fact that this nonlinear registration is often slow and suboptimal due to non-convexity of the energy function that is optimized, let us recall, above all, that the nonlinear deformation required for this registration is generally not learned from the multi-atlas dataset. A second strategy focuses on patch-based label fusion methods to compensate registration error by searching for correspondences between the target image and atlases. In this spirit, Coupé et al. (2010) use expert segmentation priors to achieve the nonlocal patch-based label fusion. Rousseau et al. (2011) propose a patch-based image labeling method, relying on image intensity similarities between the input image and an anatomy textbook. Wang et al. (2011) propose a regression-based approach for label fusion. Patch-based approach in Wu et al. (2012) uses $l_{2,1}$-norm regularization to enforce joint sparsity during the label fusion. Wang et al. (2013) combine (possibly weighted) majority voting and a patch-based approach to achieve unsupervised segmentation. Bai et al. (2013) formulate patch-based label fusion in a probabilistic Bayesian framework. Shi et al. (2014) combine spectral matching with multi-atlas PatchMatch (MAPM) to introduce multi-atlas spectral Patch-Match (MASP) segmentation. Finally, Wu et al. (2015) propose a hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition. Multi-scale feature representation increases the accuracy of the patch-based similarity measurement and label-specific atlas patches makes the label fusion process more specific and flexible. The interest of these methods is that they don't need explicit registration between the individual putative segmentations of the atlas and the subject image to be segmented. In fact, the main assumption of all these patch-based approaches is that, if patches of the input subject image are locally similar to the patches of atlases, they should have a similar label. In terms of voting rule, the atlases whose reference images are more similar to the target image should contribute more to the segmentation. This strategy borrows the main idea of the non-local means denoising algorithm (Buades et al., 2008) which assumes that image information is redundant.

In our approach, we prefer not to rely too much on patch-based similarity between two X-ray images mainly because grey level information of these patches are not always as informative as might be desired. Indeed, quantum noise can be different for different X-ray imaging systems and numerous artifacts exist, such as bone and soft-tissue artifacts and inter-individual variability of bone densities and structures. In fact, the most informative and reliable visual cues in an X-ray image of the pelvic region remain the boundary contours between the different pelvic bone structures, and particularly the external bone contour of the pelvis; our proposed method will fully rely on this information.

In our model, this boundary information between bone regions existing in the pre-segmented images of the multi-atlas dataset allows us to estimate a collection of superpixels which has the interesting property to capture, easily and non-parametrically, all the nonlinear and local variability of Regions Of Interest (ROIs) present in the multi-atlas dataset. The interest is then twofold; this allows to efficiently estimate the target-specific nonlinear deformations required by the registration process while simplifying the optimization problem involved in the pruning process (*via* a superpixel map based refining process aiming at reducing the remaining registration errors). Then, the *label propagation* step is formulated into an optimization problem, also called the median partition problem, aiming to find the best segmentation which maximizes an entropy-based similarity criterion between the compromise solution and the set of co-registered segmentations, over the space of possible segmentations. This fusion procedure has already turned out to be very efficient for combining a set of weak segmentation results to get a final improved segmentation result (Mignotte, 2014a).

Let us mention that the concept of superpixels was initially proposed in multi-atlas segmentation by Wang and Yushkevich (2013) and Yu et al. (2016) but with a very different approach compared to us, both in terms of data and algorithm used to generate the superpixel map but also in terms of their utility in the underlying registration/segmentation model. Indeed, in the above mentioned works, the superpixel representation is generated, in a classical way, through low-level unsupervised image segmentation of an input Magnetic Resonance (MR) image (from the Simple Linear Iterative Clustering proposed in Achanta et al., 2012 or the Graph-Based Image Segmentation technique described in Felzenszwalb and Huttenlocher, 2004). Nevertheless, unlike MR

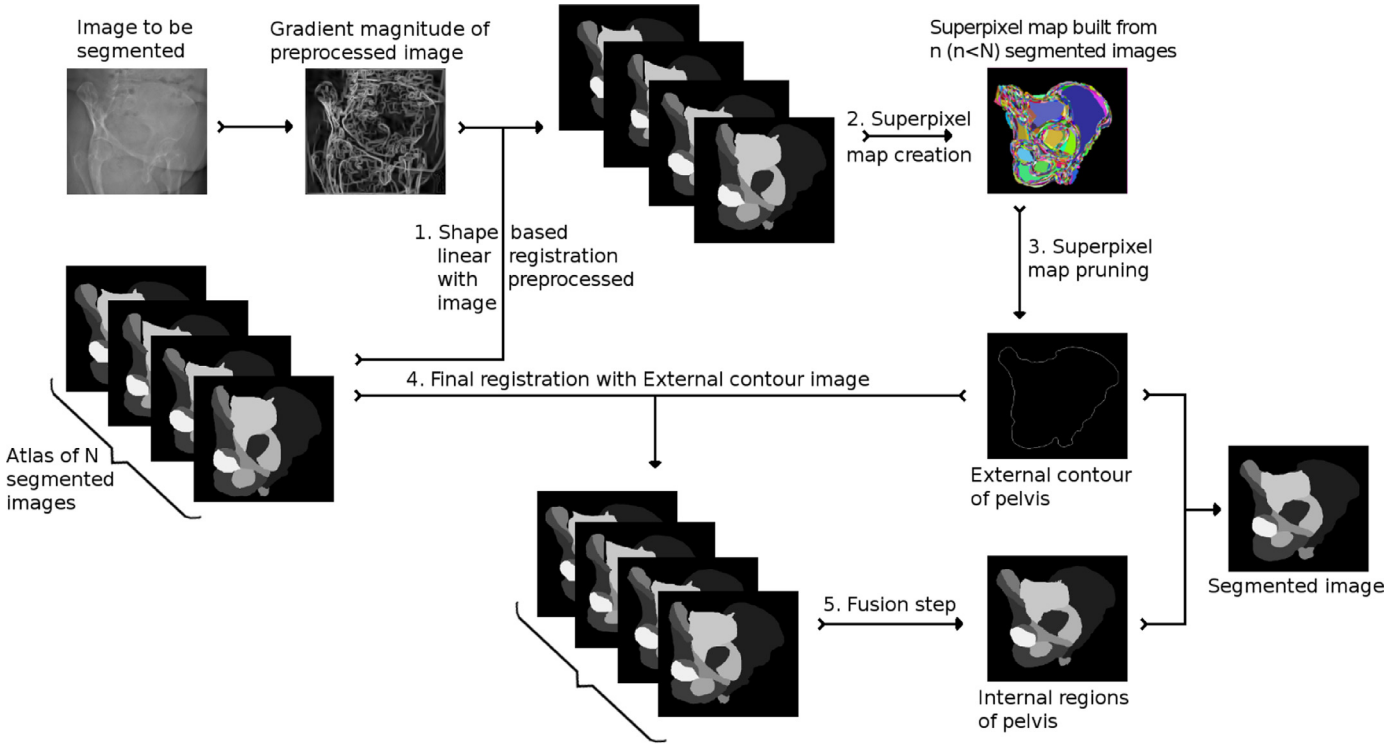


Fig. 1. The pipeline of the proposed approach.

images, which have a better quality, X-Ray images contain quantum noise, numerous artifacts (overlap and blurring) that can degrade any over-segmentation or superpixel map estimation procedure. In addition, the input image is over-segmented into small homogeneous regions or superpixels (sharing similar features) but which are not necessarily of interest in terms of bone or anatomical structures. Unlike this, our application generates the superpixel map by taking the intersection operation on the warped training label images which were manually segmented into different ROIs by experts. That is why our superpixel map takes into account the local and nonlinear deformations of ROIs (i.e., bone regions), existing in the training atlases, without too much error. Besides, our superpixel map is not used as an additional matching constraint for establishing the voxel-wise label transfer and fusion (as proposed in the two above mentioned references) but in order to refine the segmentation process.

As mentioned in our preliminary work (Nguyen et al., 2015) and to our knowledge, there is no reported work that proposes and uses, for the *label propagation* step, the result of a median partition-based optimization problem or that exploits a superpixel strategy for taking into account the nonlinear variability of the ROIs in label multi-atlas population. The remainder of this paper is divided into the following sections. In Section 2, we introduce details about the dataset that will be used in our model. Section 3 describes the DCT-based pre-processing step made on each X-ray image to be segmented in order to enhance the contour of the different anatomical structures to be segmented. The proposed model with its five steps, namely (1) shape-based linear registration and multi-atlas selection, (2) superpixel map creation, (3) superpixel map pruning, (4) final selection and filtering of the multi-atlas dataset, and (5) entropy-based fusion procedure are presented in Section 4. Finally, we show experimental results in Section 6.2 and conclude in Section 7. The pipeline of superpixel and multi-atlas segmentation of bone structures for the pelvic structure is given in Fig. 1.

2. Dataset and multi-atlas creation

For our experiments (approved by the local ethics board of our University), we used 31 anonymized 45-degree oblique X-ray images of the pelvis, talus, and patella that were manually segmented into different ROIs by experts well trained in lower extremity and medical image segmentation. Images of the pelvis were partitioned into 14 different ROIs including the whole pelvis with sacrum, and both hemipelvis. Images of the talus were manually segmented into 4 different ROIs, including the body, dome and head of talus. Finally, images of the patella had 2 different ROIs (see Fig. 2).

3. X-ray image pre-processing

In our application, a pre-processing step is required in order to enhance the visibility of the bone contours which, in fact, constitute the most important and reliable low-level visual cue in each radiographic image. To this end, a histogram equalization technique and a DCT-based denoising step (Mignotte et al., 2008; Yu and Sapiro, 2011) was used.

The DCT-based denoising procedure used in this application is an iterative strategy of denoising. Each iteration progressively denoises the image and consists in repeating a translation invariant DCT-based frequential filtering using a simple adaptive hard thresholding rule within a sliding subimage. Algorithmically and, more precisely, the iterative denoising scheme thus simply consists of alternating, until a maximal number of iterations is reached, two steps (which are also summarized in pseudo-code in Algorithm 1):

(1) The first step is a frequential filtering using the DCT transform of each 8×8 subimage extracted from the current denoised version of the degraded image (initially, this current image estimate is y, the degraded image itself). It is well known by the compression community that a DCT filtering (i.e., a threshold on these DCT coefficients) on each 8×8 blocks extracted from a (noisy or not) image can denoise, but also can cause blocky artifacts on the processed image after reconstruction (i.e., inverse DCT). In order

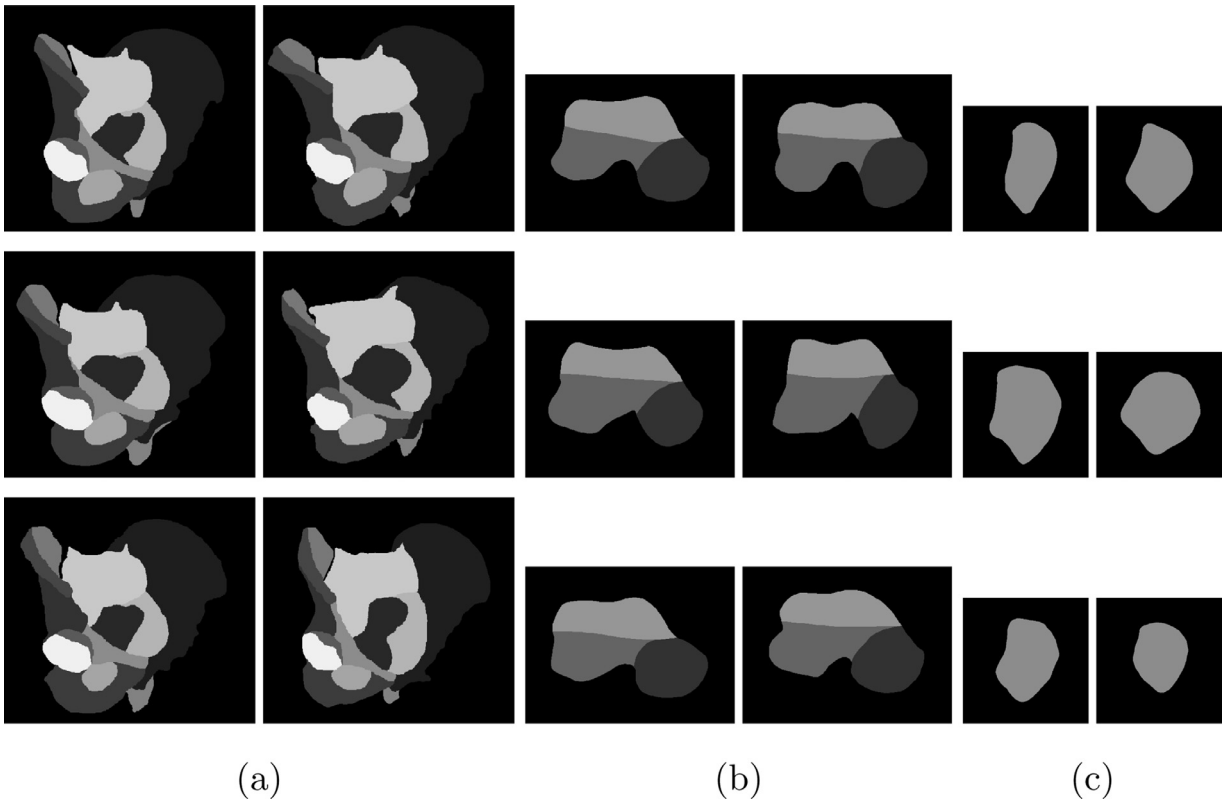


Fig. 2. Six examples of 45-degree oblique manual segmentations of the human pelvis (a), talus (b), and patella (c) used in our multi-atlas.

DCT-Based Denoising Algorithm

y Input image to be denoised (Input)
$\hat{x}$ Denoised estimated image (Output)
T Threshold

1. Initialization: $\hat{x}^{[1]} \leftarrow y$

2. for *iteration* $n = 1$ *to* 5 **do**

I. Frequential Filtering

for *All (8 horizontal and 8 vertical) possible shifts of* $\hat{x}^{[n]}$ **do**

▷ **for** *All* [8×8] *blocks extracted from* $\hat{x}^{[n]}$ **do**

• DCT Transform

• Threshold the obtained DCT coefficients w with the hard thresholding rule

$$\lambda_T^{\text{hard}} = \begin{cases} 0 & \text{if } |w| \leq T \\ w & \text{otherwise} \end{cases}$$

• Inverse DCT Transform of these thresholded coefficients

▷ Unshift the filtered image and store it

II. Averaging

$\hat{x}^{[n+1]} \leftarrow$ Averaging of these 64 denoised images

$\hat{x} \leftarrow \hat{x}^{[n+1]}$

Algorithm 1. DCT-Based Denoising Algorithm.

to reduce these artifacts, a standard way is to use the DCT of all circularly translated version of the image, herein assumed to be toroidal. This thus implies to compute a set of $N_S = 8 \times 8 = 64$ (8 horizontal shifts and 8 vertical shifts) transformed images (which are then averaged in the second step). For the filtering operation in the DCT domain, we used the simple hard thresholding rule classically used in the wavelet-based denoising approaches. After an inverse DCT transform, we have to un-shift the filtered image and store the result for the second step.

(2) The second operation consists in doing an averaging of these 64 denoised image estimates in order to compute the current denoised estimate and to make translation-invariant the proposed denoising procedure.

We returned to the first step for another denoising operation by considering that $y = \hat{x}$ and so on, until the stability of the solution, typically at the end of 5 full iterations of our application. For the implementation, we have used the fast 8×8, 16×16 FFT2D DCT package implemented in C code by T. Ooura (functions DDCT8x8S and DDCT16x16S tested in program SHRTDCT.C) and available on-line at http address given in Ooura (2001).

4. Proposed multi-atlas segmentation model

The proposed multi-atlas segmentation process is mainly divided into two stages. The first part allows us to estimate the external contour of the bone structure to be segmented (steps 4.1, 4.2 and 4.3) and the second part (steps 4.4 and 4.5) is dedicated to the refinement of the external contour of the bone structure and, above all, to estimate the region labels within the bone structure to be semantically segmented.

4.1. Shape-based linear registration and multi-atlas selection step

Since the most significant and informative visual information of a bone structure in an X-ray image (especially in the lower limbs) remains its external bone contour, each segmented image of the multi-atlas was first independently and rigid-registered, through affine transformations, to the target image using a contour-based registration technique. This rigid registration was done to produce

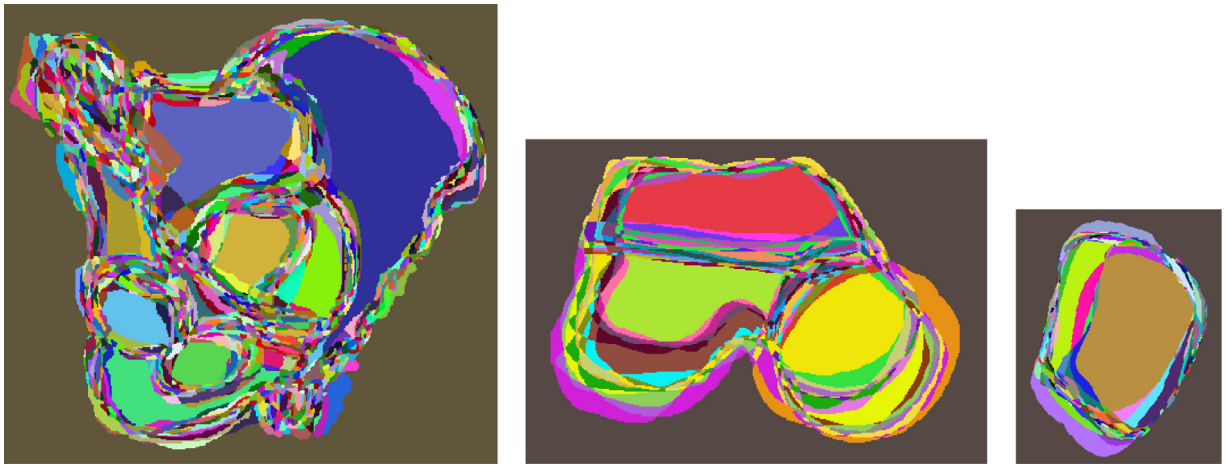


Fig. 3. From left to right; superpixel map obtained for a given X-ray 45 degree image for respectively the pelvis, talus, and patella to be segmented.

an initial alignment for each multi-atlas segmented image that maximized a similarity measure between an edge potential field estimated from the pre-processed target X-ray image and the external contour of the corresponding bone region in the segmented image. The edge potential was simply generated by the calculation of the gradient magnitude convolved with a Gaussian kernel (Gaussian blur) with a variance controlling the degree of smoothness of this contour field (Benameur et al., 2003) and making nearly convex the energy function used in the optimization-based registration technique (see Fig. 7). The rigid registration that maximized the summation of this potential map over all the pixels located on the boundary of the external contour of the bone structure to be segmented (given by the manual segmentation) thus allowed us to get a rough alignment of each atlas segmented image to the target image.

Furthermore, our contour-based similarity registration score was also used to rank the set of rigid-registered manual segmentations of our multi-atlas by decreasing order of similarity. The first half of this rigid-registered segmentations was then used for estimating the adaptive superpixel representation associated to each input image to be segmented.

4.2. Superpixel map creation

A set of superpixels is thus estimated from the first half of the optimally rigidly registered segmented images belonging to the multi-atlas by simply taking the intersection of all the segments (or regions) existing in this pre-selected (rigid-registered) manual segmentations (see Fig. 3). Conceptually, each generated superpixel corresponds to a group of connected pixels belonging to the same region label and the set of superpixels provides a (superpixel) map which is assumed to divide the input image to be segmented into atomic regions sharing the same label (thus reducing the number of entities to be labeled or/and to guide and adding constraints to the segmentation process). Fig. 4 illustrates different superpixel maps of patella with different number of input atlases. Each superpixel (i.e., each color) represents a local and nonlinear deformations of ROIs of the patella.

Ren and Malik first used the superpixels (Ren and Malik, 2003), as a pre-processing step, for the segmentation of natural images in order to simplify and better spatially constrain the segmentation task. This concept has also been proposed in Jodoin et al. (2007) however, with the different purpose of combining a segmentation map (into spatial regions) and a coarsely estimated and to-be-refined higher level vision application (e.g., occlusion map, motion segmentation/estimation field, etc.). In multi-atlas segmentation, the concept of superpixels have also been used, but especially for MR images in Wang and Yushkevich (2013); Yu et al. (2016) and in order to oversegment the target images and all atlases into small regions, sharing similar features but which are not necessarily of interest in terms of bone structures. In our application, the superpixel maps were generated directly from the well segmented training atlases and were thus adaptive for each radiographic image to be segmented. They will allow us both to carry out an adaptive local non-rigid registration taking into account the local and nonlinear deformations of ROIs specific to the target but also to guide and to simplify the superpixel map pruning step by better spatially constraining it.

4.3. Superpixel map pruning

At this stage, an adaptive superpixel map, which non-parametrically incorporated all the nonlinear variability existing in a selected subset of the multi-atlas dataset (inter-patient variability of the imaging pose and bone morphology), was generated using each radiographic target image to be segmented. Since the superpixel map was already linearly registered to the target image (because this intersection map was built from a subset of selected and linearly rigidly registered segmentations), a pruning step, using an iterative pruning algorithm, was easily achieved by simply finding the set of connected superpixels which maximized the contour-based similarity between the outer contour of this superpixel map and the edge map of the radiographic image[1]. More precisely, this iterative pruning procedure can be described in the following manner; Each superpixel connected with the background label (i.e., which is contiguous with the outer bone region to be segmented), in a lexicographic order, was classified as belonging to the considered bone region if the outer contour-based similarity metric increased and until convergence or a maximum number of iterations was reached. At convergence, this iterative procedure (See Algorithm 2) allowed us to estimate a precise closed (and continuous) outer contour of the bone structure.

4.4. Final selection/filtering of the multi-atlas dataset

At this stage, we had to refine the previously closed external contour of the bone structure and also to infer both the internal

[1] Experiments have shown that slightly better results are given if the superpixel map is scaled by a factor slightly greater than one in order to ensure that the external bone contour of the X-ray target image is fully contained in the superpixel map. To this end, after trial and error, a scale factor of 1.02 allows us to give the best segmentation results.

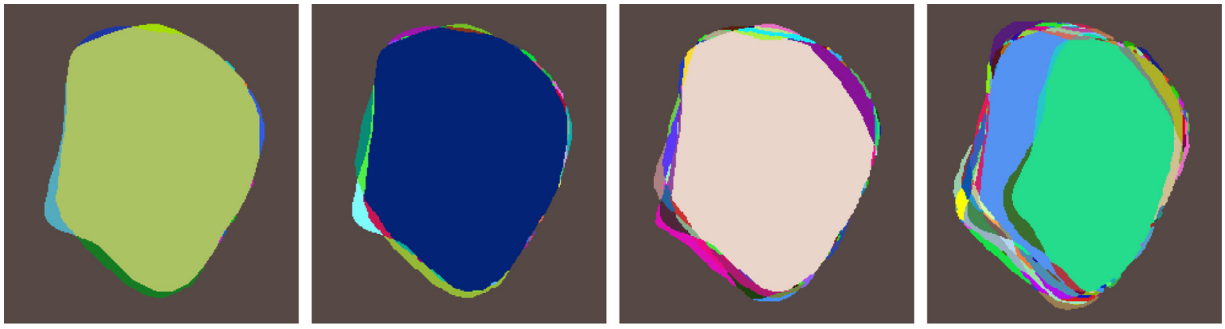


Fig. 4. Examples of superpixel maps of the patella created from (a) 2 atlases, (b) 3 atlases, (c) 5 atlases and (d) 10 atlases. Each superpixel (i.e. each color) represents a nonlinear deformation between each bone regions.

Iterative Pruning Algorithm

S_{SupPix}	Superpixel map (Input)
I_{Grad}	Gradient potential field of the pre-processed X-ray image to be segmented (Input)
$\hat{S}_{\text{ExtCont}}$	External contour of the bone (Output)
$N_{\max}$	Maximal number of iterations
c_i	Thresholds ($i \in \{1,2,3\}$)
E_{ExtCont}	Energy of the external contour

1. Initialization

$\hat{S}^{[0]}_{\text{SPixPruned}} \leftarrow S_{\text{SupPix}}$ & $n \leftarrow 0$ & $E^{[0]}_{\text{ExtCont}} \leftarrow 0$

• Estimation of the biggest (usually central) superpixel *bsp* (which should never be pruned)

2. Iterative Pruning

do

  for *each superpixel sp (sp* $\neq$ *bsp) of* $\hat{S}^{[n]}_{SPixPruned}$*, both in the lexicographic order and belonging to the external contour* **do**

    • Compute energy E_{sp} of *sp* (using I_{Grad})

    if $(Esp < c_2)$ **then**

      Delete *sp* from $\hat{S}^{[n]}_{\text{SPixPruned}}$

    else

      • Compute E_1, E_2: the energy of the external contour of $\hat{S}^{[n]}_{\text{SPixPruned}}$ with & without *sp* (using I_{Grad})

      • Compute E_3, E_4: the energy of the external contour of *sp* and the energy of the superpixel which replaces *sp* (in the external contour of $\hat{S}^{[n]}_{\text{SPixPruned}}$) (using I_{Grad})

      if $(E_2 \geq c_3 \cdot E_1)$ *and* $(E_4 \geq c_3 \cdot E_3)$ **then**

        Delete *sp* from $\hat{S}^{[n]}_{\text{SPixPruned}}$

  • Compute energy $E^{[n]}_{\text{ExtCont}}$ (using I_{Grad})

  $n \leftarrow n+1$

while $\left(\hat{S}^{[n]}_{SPixPruned} \neq \hat{S}^{[n-1]}_{SPixPruned}\right)$ *and* $\left(n < N_{max}\right)$ *and* $\left(E^{[n]}_{ExtCont} \geq c_1 \cdot E^{[n-1]}_{ExtCont}\right)$;

$\hat{S}_{\text{ExtCont}} \leftarrow$ External contour of $\hat{S}^{[n]}_{\text{SPixPruned}}$

Algorithm 2. Iterative Pruning Algorithm.

VoI-Based Mean Segmentation Estimation

$\{S_k\}_{k \leq L}$	Set of L segmentations to be fused (Input)
$\hat{S}_{\overline{\text{VoI}}}$	Estimated fused result (Output)
$T_{\max}$	Maximal number of iterations
$\overline{\text{VoI}}$	Mean VoI (See Eq. (2))

1. Initialization:

$$\hat{S}^{[0]}_{\overline{\text{VoI}}} = \arg\min_{S \in \{S_k\}_{k \leq L}} \overline{\text{VoI}}\left(S, \{S_k\}_{k \leq L}\right) \quad \& \quad n \leftarrow 0$$

2. Steepest Local Energy Descent

while $n < T_{max}$ **do**

  for *each pixel of* $\hat{S}^{[n]}_{\overline{VoI}}$ *with label* l_s *at site* s **do**

    • Let $\mathcal{E}$ the set of labels $\neq l_s$ contained in the local (squared) fixed-size ($N_w = 7$) neighborhood of s

    • Draw a new label x according to the uniform distribution in the set $\mathcal{E}$

    • **if** $x = \emptyset$ **then**

      continue;

    • **if** *pixel s with label x is not 4-connected with the x-th region in* $\hat{S}^{[n]}_{\overline{VoI}}$ **then**

      continue;

    • Compute $\Delta\overline{\text{VoI}}\left(\hat{S}^{[n]}_{\overline{\text{VoI}}}, \{S_k\}_{k \leq L}\right)_{s: l_s \rightarrow x}$ (See Equation (3))

    • **if** $\Delta\overline{VoI}\left(\hat{S}^{[n]}_{\overline{VoI}}, \{S_k\}_{k \leq L}\right)_{s: l_s \rightarrow x} > 0$ **then**

      replace label l_s by label x at site s

  $n \leftarrow n+1$

$\hat{S}_{\overline{\text{VoI}}} \leftarrow \hat{S}^{[n+1]}_{\overline{\text{VoI}}}$

Algorithm 3. VoI-Based Segmentation Fusion Algorithm.

regions (within the bone structure) along with their most likely semantic labels.

To this end, since the previously estimated external contour was closed, a final linear registration, between it and each manual segmentation of the multi-atlas, was achieved with a region-based similarity metric (namely the F-measure[2]) and also allowed us to identify and select the first half of the manual segmentations of our multi-atlas dataset, which were best registered.

[2] In this binary region-based classification step, the F-measure (also called F-score) is a region-based measure of accuracy. The accuracy of this classification is expressed by means of recall (R), precision (P) and their harmonic mean, the F-score: $F = 2PR/(P+R)$ with $P = TP/(TP+FP)$ and $R = TP/(TP+FN)$, where *TP, FP* and *FN* indicate, respectively, the number of correct detections, false alarms and missed detections at pixel-level. The F-score can be also interpreted as a weighted average of the precision and recall, where an F-score reaches its best value at 1 and worst at 0.

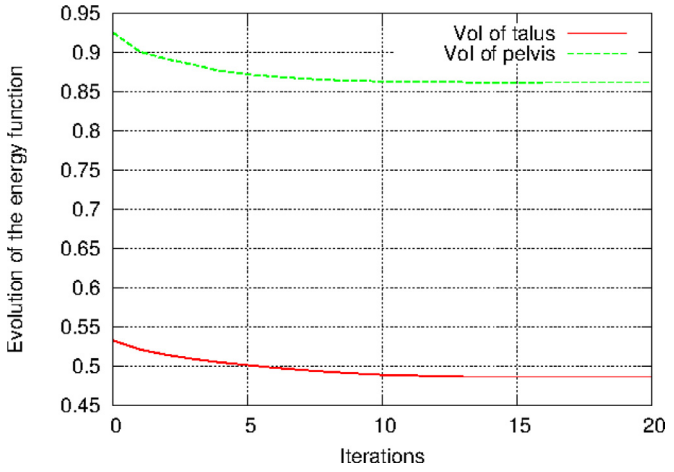


Fig. 5. Evolution of the consensus energy function of the VoI-based fusion procedure along the number of iterations for two different bone structures.

Finally, these above-mentioned pre-selected rigid-registered manual segmentations were combined in the variation of information (VoI) sense (Mignotte, 2014a), to yield the final semantic segmentation.

4.5. VoI-based fusion procedure

The fusion procedure, used in our application, is based on the variation of information (VoI) distance (Meilă, 2005; 2007) which has the crucial advantage of being a true metric on the space of segmentations, i.e., a metric which is positive, symmetric and obeying the triangle inequality (Meilă, 2007). This metric is commonly used for comparing the similarity of two segmentation results (or clusterings) or also as an efficient quality score that measures the agreement of the segmentation result with a given ground truth. To this end, it was used in image segmentation (Mignotte, 2011a; 2011b; 2014b) in order to objectively benchmark the efficiency of an unsupervised segmentation algorithm giving an automatic machine segmentation of an image to a ground truth segmentation (e.g., a manually hand-segmented image given by an expert).

Let $S^{\mathrm{A}} = \{C_1^{\mathrm{A}}, C_2^{\mathrm{A}}, \ldots, C_{R^{\mathrm{A}}}^{\mathrm{A}}\}$ and $S^{\mathrm{B}} = \{C_1^{\mathrm{B}}, C_2^{\mathrm{B}}, \ldots, C_{R^{\mathrm{B}}}^{\mathrm{B}}\}$ be respectively the first and second segmentation (or the machine segmentation result to be evaluated and the ground truth segmentation of the same size) between which the VoI distance has to be estimated. S^{A} is composed of n pixels and R^{A} regions C_i (also called cluster, segments or superpixels in the following). Let n_i^{A} be the number of pixels in the ith region (C_i) of the segmentation S^{A}, n_j^{B} the number of pixels in the jth region (C_j) of the segmentation S^{B} and finally n_{ij} the number of pixels which are together in the ith region of the segmentation S^{A} and in the jth region of the segmentation S^{B}. $P(i) = n_i^A/n$ represents the probability that a pixel belongs to the ith region of S^{A} (respectively $P(j) = n_j/n$ being the probability that a pixel belongs to the jth region of S^{B}). Let finally $P(i, j) = n_{ij}/n$ represents the joint probability that a pixel belongs to C_i^{A} and to C_j^{B}. The VoI between S^{A} and S^{B} is always in the range $[0, \ldots 1]$ and defined as:

$$\mathrm{VoI}(S^{\mathrm{A}}, S^{\mathrm{B}}) = H(S^{\mathrm{A}}) + H(S^{\mathrm{B}}) - 2 \cdot I(S^{\mathrm{A}}, S^{\mathrm{B}})$$

$$\text{with:} \quad H(S^{\mathrm{A}}) = -\sum_{i=1}^{R^{\mathrm{A}}} P(i) \log P(i) = -\sum_{i=1}^{R^{\mathrm{A}}} \frac{n_i^{\mathrm{A}}}{n} \log \frac{n_i^{\mathrm{A}}}{n}$$

Segmentation algorithm of X-ray images

I	X-ray image to be segmented (Input)
$\{\mathcal{S}_k\}_{k \le N}$	DataSet (multi-atlas) of N (*a priori*) segmentations of the bone structure (Input)
I_{Seg}	Segmented X-ray image (Output)
T_{DCT}	Threshold of the denoising filter
σ_b	Std of the contour-based potential field blur
ρ	Selection % of the best registrations
$\{c_i\}_{i=1,2,3}$	Thresholds of the pruning algorithm

Initialization: X-Ray Image Pre-processing

. ▷ **From** I **do**

- Histogram equalization
- DCT-based denoising algorithm (T_{DCT})
- Gradient potential field (gradient magnitude + Gaussian blur with standard deviation σ_b)

. ▷ **Result:** I_{Grad}

1. Bone Structure External Contour Estimation

▷ A. Shape-based Linear Registration and Multi-Atlas Selection

$\{\mathcal{S}_k^{\bullet}\} \leftarrow$ Rigid registration of each segmentation of $\{\mathcal{S}_k\}_{k \le N}$ and selection of the ρ% best registrations (in term of our contour-based potential using I_{Grad})

▷ B. Superpixel Map Creation

$S_{\mathrm{SupPix}} \leftarrow$ intersection of all the regions existing in $\{\mathcal{S}_k^{\bullet}\}$

▷ C. Superpixel Map Pruning

$\hat{S}_{\mathrm{ExtCont}} \leftarrow$ Pruning Algorithm $\left(S_{\mathrm{SupPix}}, I_{\mathrm{Grad}}, \{c_i\}_{i=1,2,3}\right)$

2. External Contour Refinement and Internal Segmentation Estimation

▷ D. Final Selection of the Multi-Atlas

$\{\mathcal{S}_k^{\circ}\} \leftarrow$ Rigid registration of each segmentation of $\{\mathcal{S}_k\}_{k \le N}$ in the (region-based) F-score sense between $\hat{S}_{\mathrm{ExtCont}}$ and $\mathcal{S}_k$ & selection of the ρ% best registrations

▷ E. VoI-Based Mean Segmentation Estimation

$I_{\mathrm{Seg}} \leftarrow$ Voi-Based Mean Seg. Algorithm $\left(\{\mathcal{S}_k^{\circ}\}\right)$

Algorithm 4. The proposed X-ray segmentation algorithm.

$$I(S^{\mathrm{A}}, S^{\mathrm{B}}) = \sum_{i}^{R^{\mathrm{A}}} \sum_{j}^{R^{\mathrm{B}}} P(i, j) \log \frac{P(i, j)}{P(i)\,P(j)} \qquad (1)$$

where $H(S^{\mathrm{A}})$ and $H(S^{\mathrm{B}})$ denote respectively the classical entropy associated with the segmentation S^{A} and S^{B} and $I(S^{\mathrm{A}}, S^{\mathrm{B}})$ the mutual information between these two segmentations.

Let $\{S_k\}_{k \le L}$ be a finite ensemble of L segmentations $\{S_k\}_{k \le L} = \{S_1, S_2, \ldots, S_L\}$ to be fused or combined in the VoI metric sense in order to obtain $\hat{S}_{\overline{\mathrm{VoI}}}$, *i.e.*, (equivalently) the average (mean) segmentation in the VoI distance sense also called the best compromise segmentation solution (resulting in a consensus in terms of contour accuracy, detail level, clutters, etc. exhibited by each segmentations ($\in \{S_k\}_{k \le L}$)). $\hat{S}_{\overline{\mathrm{VoI}}}$ can be regarded as the aver-

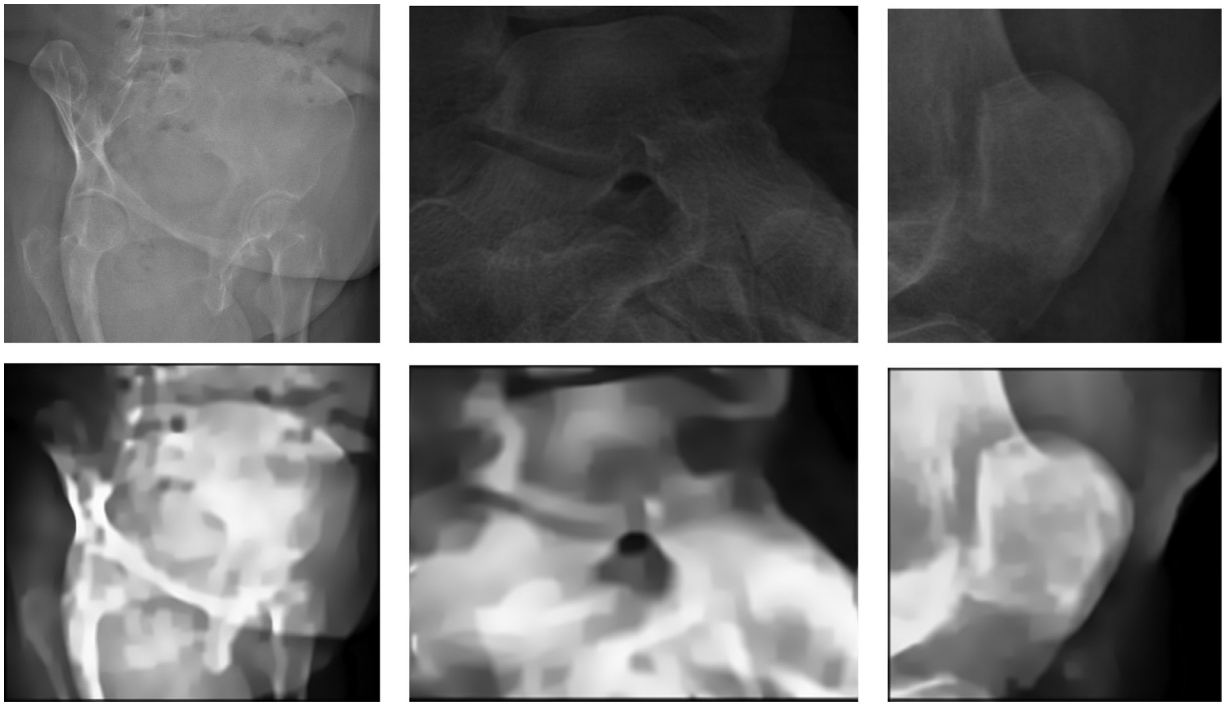


Fig. 6. Original oblique X-ray radiographic images of respectively the pelvis, the talus and the patella before and after the denoising step (see Section 3).

age pairwise agreement between putative individual segmentations ($\in \{S_k\}_{k\leq L}$) or the solution of the following optimization (or so-called *median partition* Vega-Pons and Ruiz-Shulcloper, 2011) problem:

$$\hat{S}_{\overline{\mathrm{VoI}}} = \arg\min_{S\in\mathcal{S}_n} \underbrace{\frac{1}{L}\sum_{k=1}^{L} \mathrm{VoI}\,(S, S_k)}_{\overline{\mathrm{VoI}}\,(S,\{S_k\}_{k\leq L})} \tag{2}$$

with $\mathcal{S}_n$ is the set of all possible segmentations using n pixels. This consensus segmentation solution can be expressed as the result of a minimization problem on the consensus function $\overline{\mathrm{VoI}}$ and can be efficiently solved with a iterative local gradient descent algorithm (Mignotte, 2014a) in which a new label x is assigned to pixel s (initially with label l_s), if this pixel is connected to the xth region and if there is a local decrease in the energy $\overline{\mathrm{VoI}}(.)_{s:l_s\to x}$ with:

$$\begin{aligned}
&\Delta\overline{\mathrm{VoI}}\left(\hat{S}^{[p]}_{\overline{\mathrm{VoI}}}, \{S_k\}_{k\leq L}\right)_{s:m\to x} \\
&= L\cdot\left\{-\frac{n_m}{n}\log\left(\frac{n_m}{n}\right) - \frac{n_x}{n}\log\left(\frac{n_x}{n}\right)\right. \\
&\quad + \frac{(n_m-1)}{n}\log\left(\frac{n_m-1}{n}\right) + \frac{(n_x+1)}{n}\log\left(\frac{n_x+1}{n}\right)\Big\} \\
&\quad -2\cdot\sum_{l=1}^{L}\left\{\frac{n_{m,\mathcal{L}_s^l}}{n}\log\left(\frac{n_{m\mathcal{L}_s^l}}{n}\cdot\frac{n}{n_m}\cdot\frac{n}{n_{\mathcal{L}_s^l}}\right)\right. \\
&\quad + \frac{n_{x\mathcal{L}_s^l}}{n}\log\left(\frac{n_{x\mathcal{L}_s^l}}{n}\cdot\frac{n}{n_x}\cdot\frac{n}{n_{\mathcal{L}_s^l}}\right) \\
&\quad - \frac{(n_{m\mathcal{L}_s^l}-1)}{n}\log\left(\frac{(n_{m\mathcal{L}_s^l}-1)}{n}\cdot\frac{n}{(n_m-1)}\cdot\frac{n}{n_{\mathcal{L}_s^l}}\right) \\
&\quad \left.\left. - \frac{(n_{x\mathcal{L}_s^l}+1)}{n}\log\left(\frac{(n_{x\mathcal{L}_s^l}+1)}{n}\cdot\frac{n}{(n_x+1)}\cdot\frac{n}{n_{\mathcal{L}_s^l}}\right)\right\}\right\}
\end{aligned} \tag{3}$$

where $\mathcal{L}_s^l$ denotes the label at site s of the lth segmentations ($l\leq L$) of the segmentation ensemble $\{S_k\}_{k\leq L}$ and we recall that $n_{m\mathcal{L}_s^l}$ designates the number of pixels which are together in the mth region of the segmentation S and in the $\mathcal{L}_s^l$th region of the segmentation $S_l\in\{S_k\}_{k\leq L}$ (see Algorithm 1). As initialization of this steepest gradient descent, we can start from the segmentation result (among the L segmentation results to be averaged), ensuring the minimal consensus energy in the $\overline{\mathrm{VoI}}$ sense (Mignotte, 2014a).

In the *label propagation* step, the VoI-based fusion procedure allowed us both to refine the external contour of the bone structure to be segmented, but also, and especially, to infer the internal region labels from the selected atlases (previously estimated in Section 4.4) to the final segmentation map. Fig. 5 shows two examples of the evolution of the consensus energy function of the VoI-based fusion procedure along the number of iterations for two different bone structures.

5. Time complexity

The purpose of this section is to compute an estimate for the time complexity required to segment an X-ray image by our strategy. We performed a complexity analysis on each step of our approach.

First, the time complexity of our X-ray image pre-processing step is linear with respect to the number of pixels with $\mathcal{O}(p(n_{dct}b+g))$ where p is the number of pixels in the image, n_{dct} is the number of DCT iterations, b is the extracted block size and g is the Gaussian kernel size.

Second, the time complexity of bone structure external contour estimation step is $\mathcal{O}(p(N_{dataset}ts+n_{pruning}P))$. More precisely, the time complexity of shape-based registration is $\mathcal{O}(N_{dataset}tsp)$, where $N_{dataset}$ is the number of images in training dataset, t is the number of translations, and s is the number of scaling factors. This complexity grows linearly with $N_{dataset}$ and p, while t and s can be fixed. The time complexity of superpixel map creation is

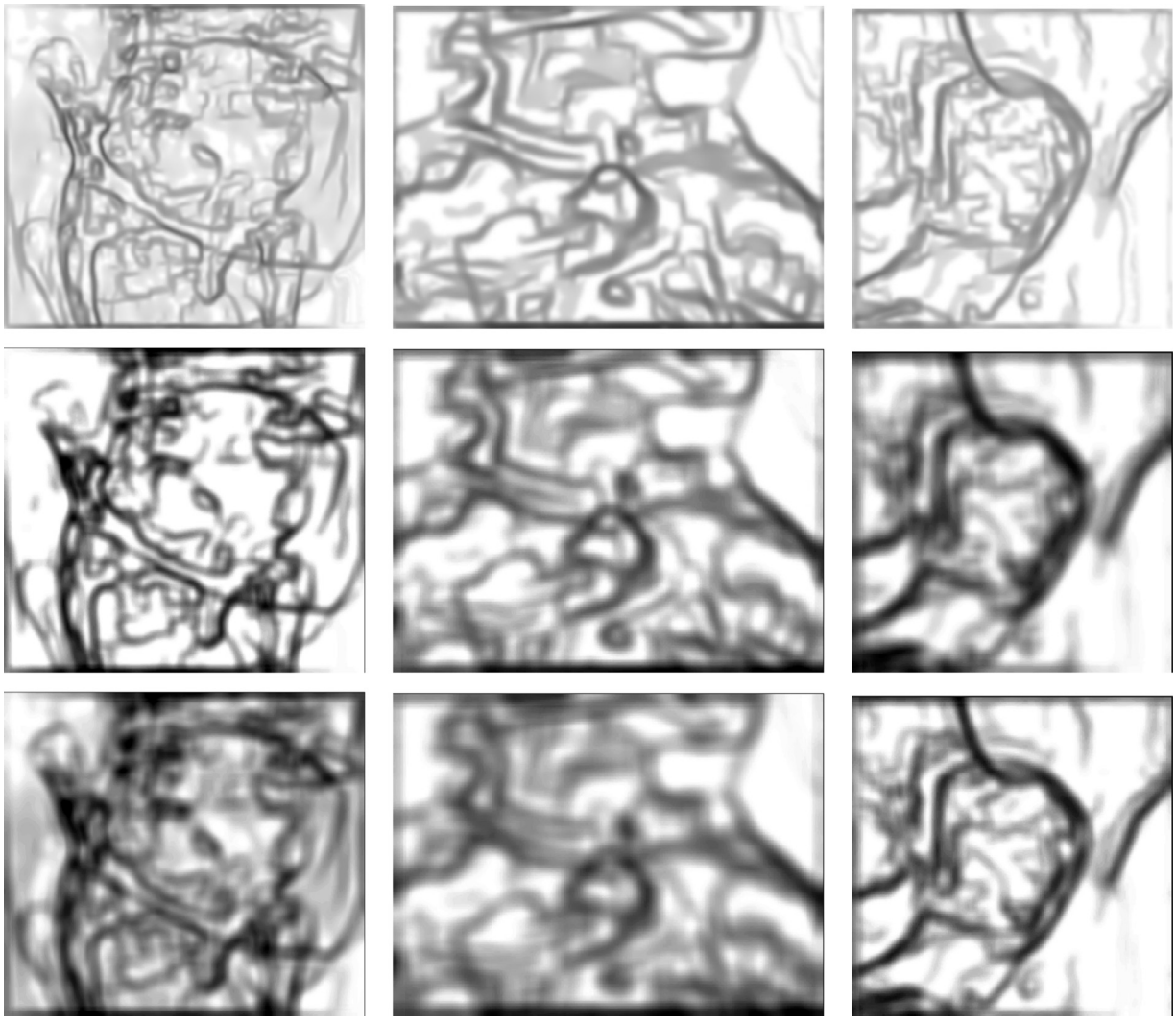


Fig. 7. From left to right; gradient magnitude of three original oblique pre-processing X-ray radiographic images of the pelvis, talus, and the patella, respectively. From top to bottom; potential field with Gaussian filter with an increasing standard deviation σ_b.

$\mathcal{O}(N_{map}p)$, where N_{map} ($N_{map} < N_{dataset}$) is the number of images used for creating the superpixel map. Also, the time complexity of the superpixel map pruning algorithm is $\mathcal{O}(n_{pruning}Pp)$, where $n_{pruning}$ is the number of iterations required and P is the number of superpixels.

Third, the time complexity of external contour refinement and internal segmentation estimation step is $\mathcal{O}(p(N_{dataset}ts + n_{voi}N_{fusion}))$. More precisely, the time complexity of final selection of the multi-atlas (i.e. rigid registration) step, which is similar to the shape-based registration, is $\mathcal{O}(N_{dataset}tsp)$. Also the time complexity of the VoI based fusion algorithm is $\mathcal{O}(n_{voi}N_{fusion}p)$, where n_{voi} is the number of iterations required, and N_{fusion} ($N_{fusion} < N_{dataset}$) is the number of images used for fusion step.

Finally, the overall complexity of our approach remains linear with respect to the number p of pixels in the image with $\mathcal{O}(p(n_{dct}b + g + N_{dataset}ts + n_{pruning}P + n_{voi}N_{fusion}))$.

6. Experiments

6.1. Data

For the experiments, we validated our multi-atlas based segmentation approach on 31 45-degree oblique X-ray radiographic images and on three different human bone structures (pelvis, talus, and patella) acquired using a X-ray imaging system. Every pelvis, talus and patella image was cropped for the best view, resulting in image sizes of 1500×1350, 400×300 and 350×350 pixels respectively.

6.2. Experimental results

The overall algorithm is summarized in pseudo-code in Algorithm 4. The internal parameters of the proposed method are respectively the threshold T_{DCT} of the DCT-denoising filter for the pre-processing step, the standard deviation σ_b of the potential field used in the shape-based linear registration, and local non rigid contour based reconstruction (step 1.A and 1.C). The selection parameter ρ of the multi-atlas selection (step 1.A and 2.D), and finally the three thresholds of the pruning algorithm c_1, c_2, c_3. We have set these parameters, chosen after some trial and error experiments, to the following respective values $T_{\mathrm{DCT}} = 40$, $\sigma_b = 2$, $\rho = 50\%$ and $c_1 = 0.95, c_2 = 3.5, c_3 = 0.99$.

The performance of segmentation algorithms is usually evaluated based both on the quality or accuracy of the segmentation result and computational requirement. The quality or accuracy of the segmentation algorithm is evaluated by comparing the output

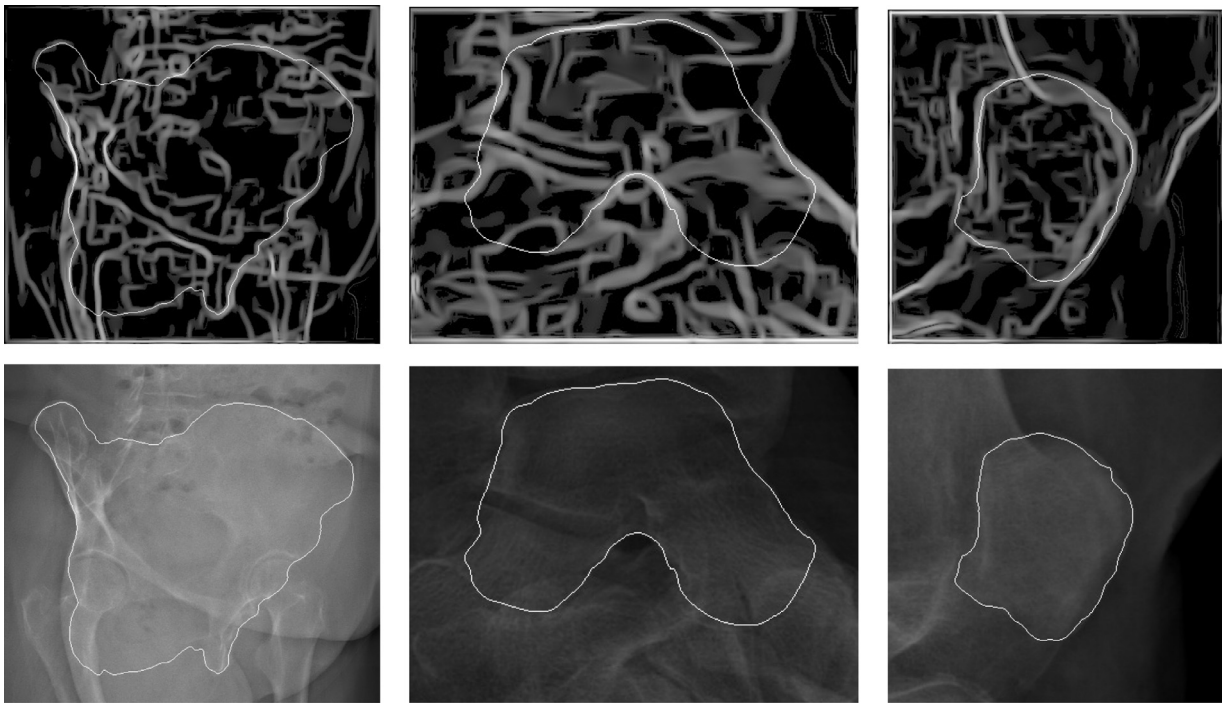


Fig. 8. Respectively the pelvic, talar and patellar external contours on the 45 degree oblique X-ray radiographic images (gradient and original).

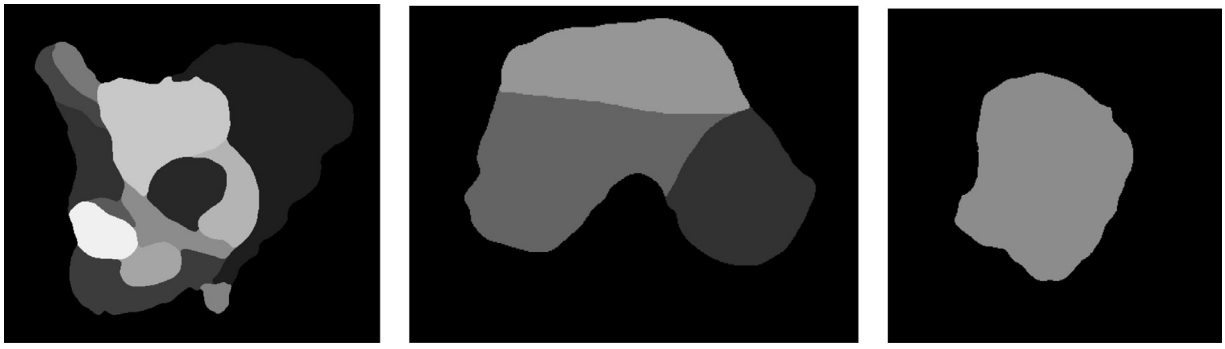


Fig. 9. Resulting fusion image estimated after the VoI-based *label propagation* step, with the estimated internal region labels of the pelvis, talus, and patella.

of the method with a ground truth, usually obtained from manual segmentation done by experienced radiologists. However, the ground truth always suffers from inter and intra-observer variations. To solve this problem, we have performed a leave-one-out procedure, *i.e.,* we removed each existing manual segmentation from the multi-atlas data set while other manual segmentations remained. Each X-ray image associated to the previously removed segmentation map was then segmented by our strategy and compared, in terms of classification error rate (i.e., similarity index), with its manual segmentation.

Fig. 6 shows the original oblique X-ray radiographic image of pelvis, talus and patella respectively before and after the denoising step (without the computation of the magnitude gradient) (see Section 3). Fig. 7 shows an example of gradient magnitude and potential field with Gaussian filter with an increasing standard deviation σ_b of three original oblique pre-processing X-ray radiographic images of respectively the pelvis, talus, and patella. Fig. 8 shows the external pelvic, talar, and patellar contours respectively on the corresponding 45-degree oblique X-ray radiographic (original and gradient) images and Fig. 9 shows the resulting segmentations estimated after the VoI-based *label propagation* step, with the estimated internal region labels of the pelvic, talar and patellar structures. Figs. 10, 11, and 12 show respectively, some segmentation results from our automatic segmentation approach compared to a manual gold standard segmentation.

The accuracy of our automatic segmentation approach was, with our leave-one-out procedure, respectively 85.02% for the pelvis (e.g. in Fig. 10), 88.3% for the talus (e.g. in Fig. 11), and 93.79% for the patella (e.g. in Fig. 12). These scores take into account the accuracy of the labeling of the different regions within the bone structure to be segmented. The complete segmentation of one image took, on average, approximately 161 s for the pelvis, 62 seconders for the talus, and 80 s for the patella on a 64-bit desktop PC (2,50 GHz Core i5-3210M CPU and a graphic card with Intel, 8 GB RAM). Multi-core parallel computing with OpenMP was used during the rigid registration step for each individual registration to speed up the overall procedure, Table 1 shows the running time of the different steps of our segmentation approach.

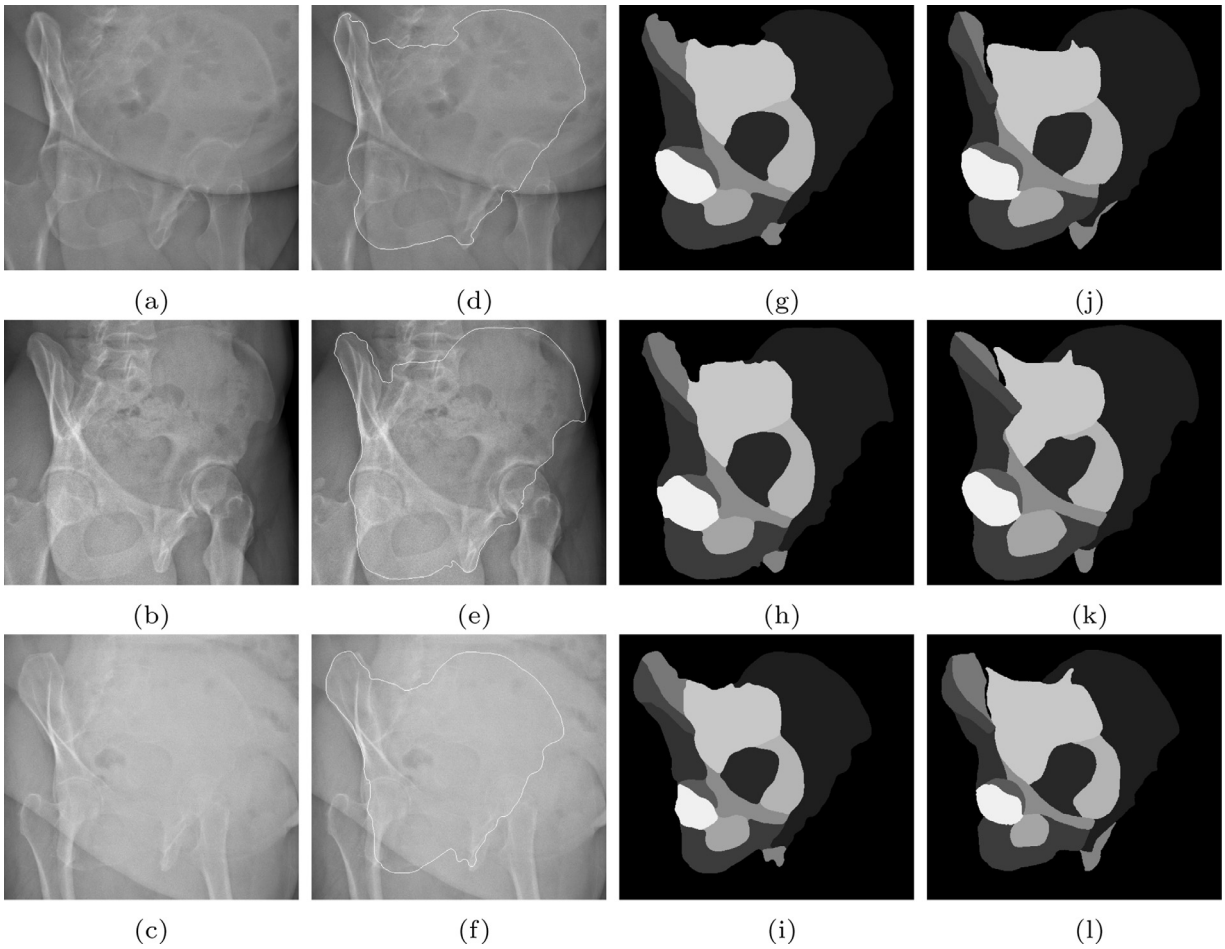


Fig. 10. Comparison of segmentation results from our approach and a manual segmentation of the pelvis. (a), (b), and (c) original oblique X-ray radiographic images; (d), (e), and (f) external pelvic contours on the corresponding X-ray images; (g), (h), and (i) resulting images; (j), (k), and (l) manual segmentations.

Table 1
Computational time of different steps of the algorithm for the pelvis.

Step	Time (s) and %	
Pre-processing	1	(0.5%)
Shape-based linear registration	4	(2.5%)
Superpixel map estimation	2	(1%)
Iterative pruning algorithm	15	(9%)
External contour-based registration	44	(27%)
VoI-based mean segmentation	95	(60%)
Total computational time	161	(100%)

In spite of the numerous artifacts existing in X-ray images, such as the overlapping edges and concealed boundaries, the proposed segmentation method can cope with these artifacts and achieve good segmentation performance in an unsupervised way.

6.3. Comparison with other methods

We compared our method to the probabilistic patch-based label fusion model and the classical patch-based majority voting fusion scheme[3] using different registration strategies. Briefly, as in Bai et al. (2013), atlases and corresponding label maps were first warped to the target image using image registration. Then, in order to determine the label at a given voxel, each atlas patch in the target image was compared to a number of patches (in a local search region) in a possibly warped atlases and was then weighted according to its similarity and distance to the target patch. Finally, labels from all the atlas patches were combined by a weighted majority voting based strategy, to give the label estimate in the target image. The resulting segmentation was then incorporated back into the image registration process to refine the registration results. We also compared our VoI based fusion step to Shape-based averaging (SBA)[4] fusion step presented in Rohlfing and Maurer (2007). These two fusions of segmentations are based on a different similarity metric between two segmentation maps (i.e., the entropy concept

[3] Patch-based majority voting is a powerful tool and also the simplest label fusion method exploited in multi-atlas based segmentation. It is based on the assumption that, if patches of the input subject image are locally similar to the patches of atlases, they should have a similar label. To this end, the more similar patches are selected and the label in the target image can then be determined by choosing the majority vote, i.e., the label which is the most frequent at this voxel (Huo et al., 2015).

[4] SBA is based on the distances from each pixel to the contours of each label, and thus, it implicitly includes the neighborhood information of pixels (Rohlfing and Maurer, 2007).

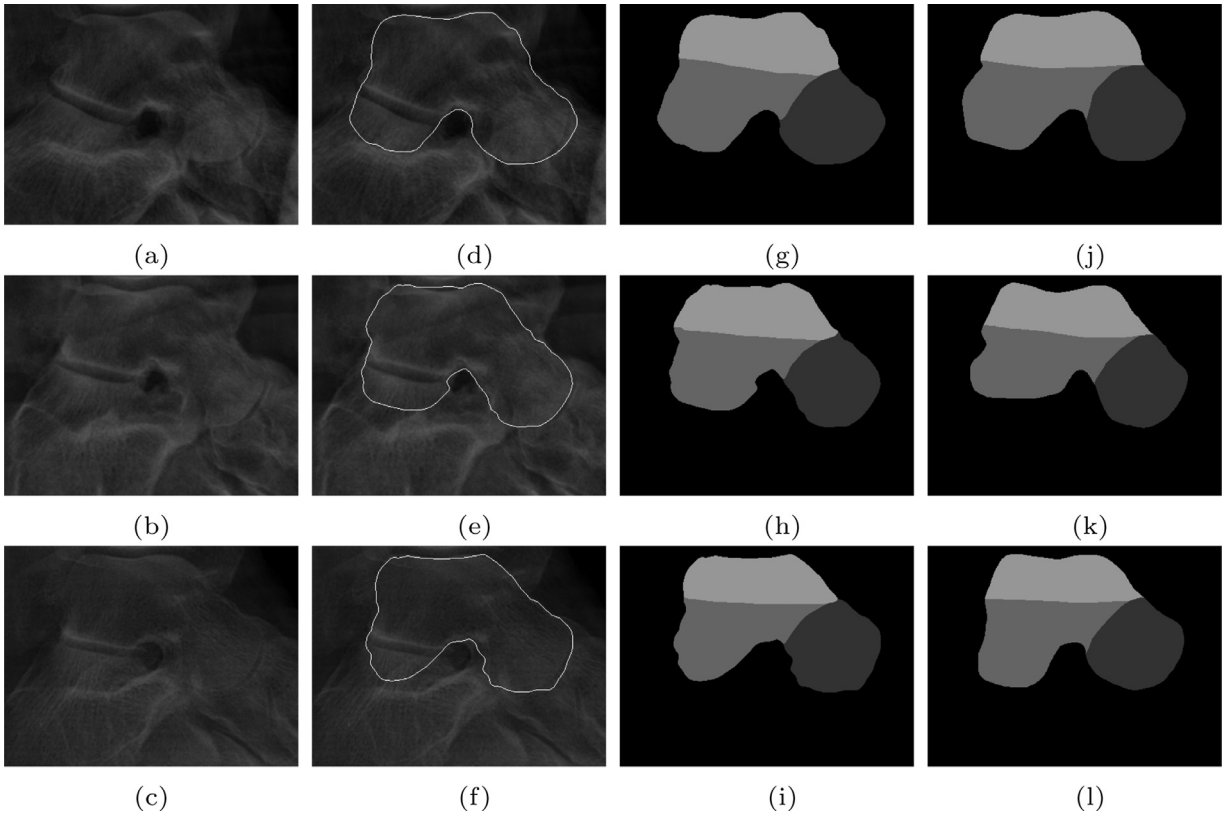


Fig. 11. Comparison of segmentation results from our approach and a manual segmentation of the talus. (a), (b), and (c) original oblique X-ray radiographic images; (d), (e), and (f) external talar contours on the corresponding X-ray images; (g), (h), and (i) resulting images; (j), (k), and (l) manual segmentations.

Table 2
Accuracy of our method versus the MV and the PB fusion model with different registration and fusion methods for multi-atlas segmentation (*:Patella bone has only 1 ROI).

		Patella (%)	Talus (%)	Pelvis (%)
Our method	VoI based fusion	93.79*	88.3	85.02
	SBA fusion		88.38	85.01
Majority voting	Affine registration	89.99	85.36	67.77
	Shape-based registration	90.26	85.26	76.91
Probabilistic patch-based	Affine registration	90.01	85.41	68.03
	Shape-based registration	90.27	85.44	77.59

for the VoI and a distance, related to the Chamfer distance for the SBA).

This comparison was evaluated on 31 radiographic image datasets for each bone structure and performed on above-mentioned desktop PC. In this comparison, we used an affine registration (AR) and our shape-based registration (SBR) to align the atlas with the target, then the label fusion process (MV, PB, SBA and our method VoI) was applied to obtain the final result. The segmentation quality was estimated with the similarity index by comparing the manual segmentations (considered as a gold standard) with the combinations of different registration and label fusion strategies (MV-AR, MV-SBR, PB-AR, PB-SBR, our method with VoI, and our method with SBA). For two binary segmentations A and B, the similarity index is computed as the ratio between $2|A \bigcap B|$ et $|A| + |B|$. Figs. 13, 14, and 15 show respectively some segmentation results from these 6 abovementioned segmentation approaches compared to a manual gold standard segmentation. The Tables 2 and 3 show the comparison of the similarity index (i.e., accuracy) and the computational time of these three segmentation methods respectively. Fig. 16 shows the accuracy of each atlas in dataset for leave-one-out validation.

7. Discussion and conclusion

In our experiments, we observed that the VoI based fusion and SBA fusion have almost the same accuracy, probably due to the small number of ROIs. The experiments also showed that our shape-based registration gave a better result than image registration in the probabilistic patch-based label fusion model and the classical majority voting fusion scheme. Our method is slightly better than the other methods (MV and PB) with two different registration strategies in case of simple bones (patella, talus), and much better in the case of a complex structure (pelvis). Note that probabilistic patch-based label fusion methods are used for MRI, and our method is proposed for X-ray radiographic imaging. As already mentioned in the introduction, image registration and patch-based

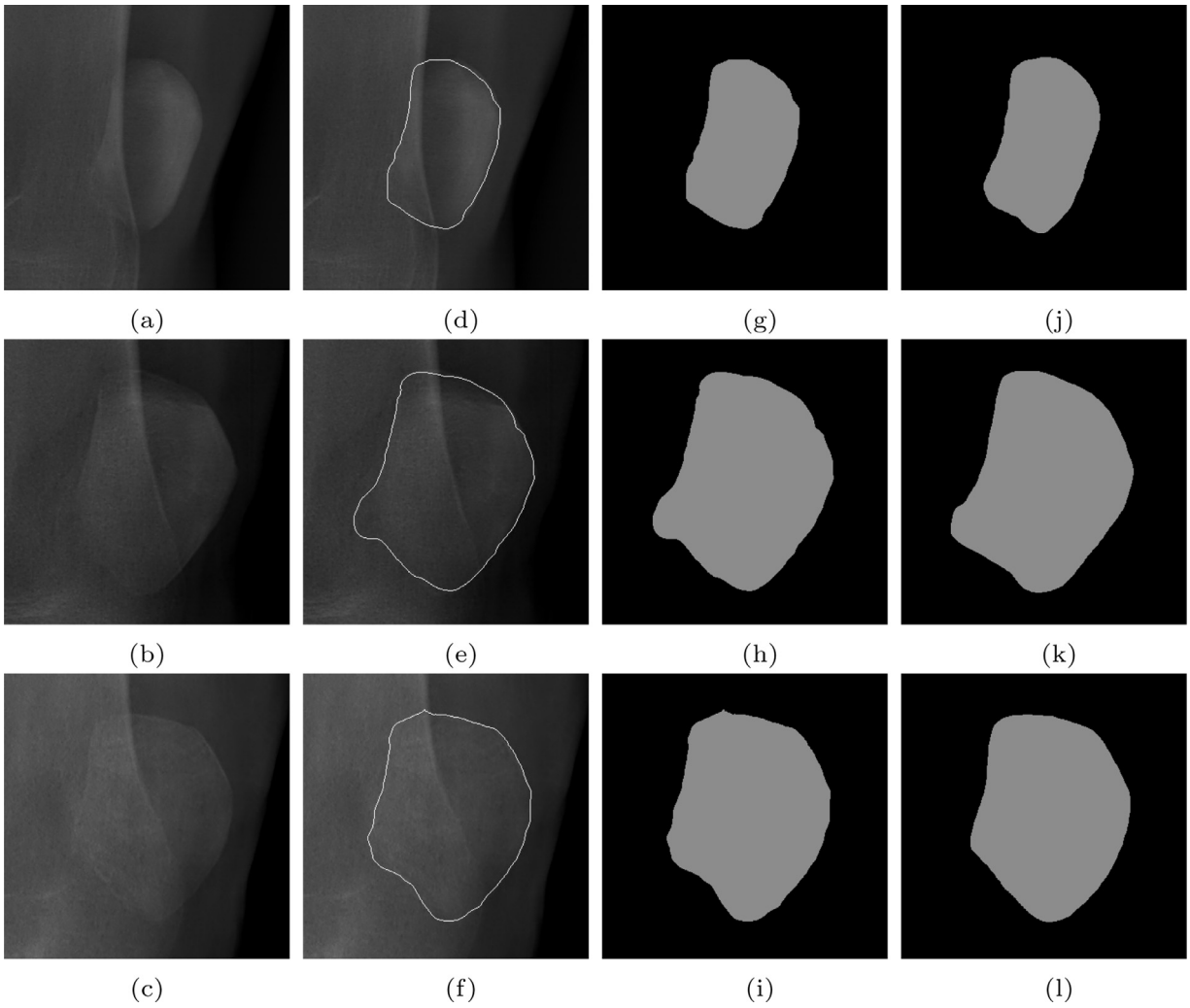


Fig. 12. Comparison of segmentation results from our approach and a manual segmentation of the patella. (a), (b), and (c) original oblique X-ray radiographic images; (d), (e), and (f) external patellar contours on the corresponding X-ray images; (g), (h), and (i) resulting images; (j), (k), and (l) manual segmentations.

Table 3
Computational time (sec.) of our method versus the MV and the PB fusion model with different registration and fusion methods for multi-atlas segmentation (*:Patella bone has only 1 ROI).

		Patella	Talus	Pelvis
Our method	VoI based fusion	80*	62	161
	SBA fusion		52	129
Majority voting	Affine registration	176	165	212
	Shape-based registration	56	6	9
Probabilistic patch-based	Affine registration	181	170	235
	Shape-based registration	58	12	18

methods probably turn out not to be well suited for X-ray images since quantum noise can be different for different X-ray imaging systems and also because numerous artifacts exist in these images, mainly due to the fact that neighboring tissues inside human body may have similar X-ray absorption rates and this phenomenon can considerably bias the estimations of the state-of-the-art patch-based methods.

The strength of our approach lies in the combination of multiple weak, but complementary visual cues, such as contour-based and region-based similarity measures all along the different steps of our registration/segmentation process. In addition, our superpixel-based framework that takes into account all the nonlinear and local variability of bone regions existing in the training dataset along with our label propagation fusion step, in the variation of information sense, combined with our outlier removal step, allows us to efficiently constrain the space of candidate segmentations, to finally propose a reliable segmentation map even in presence of low quality input X-ray images. This also

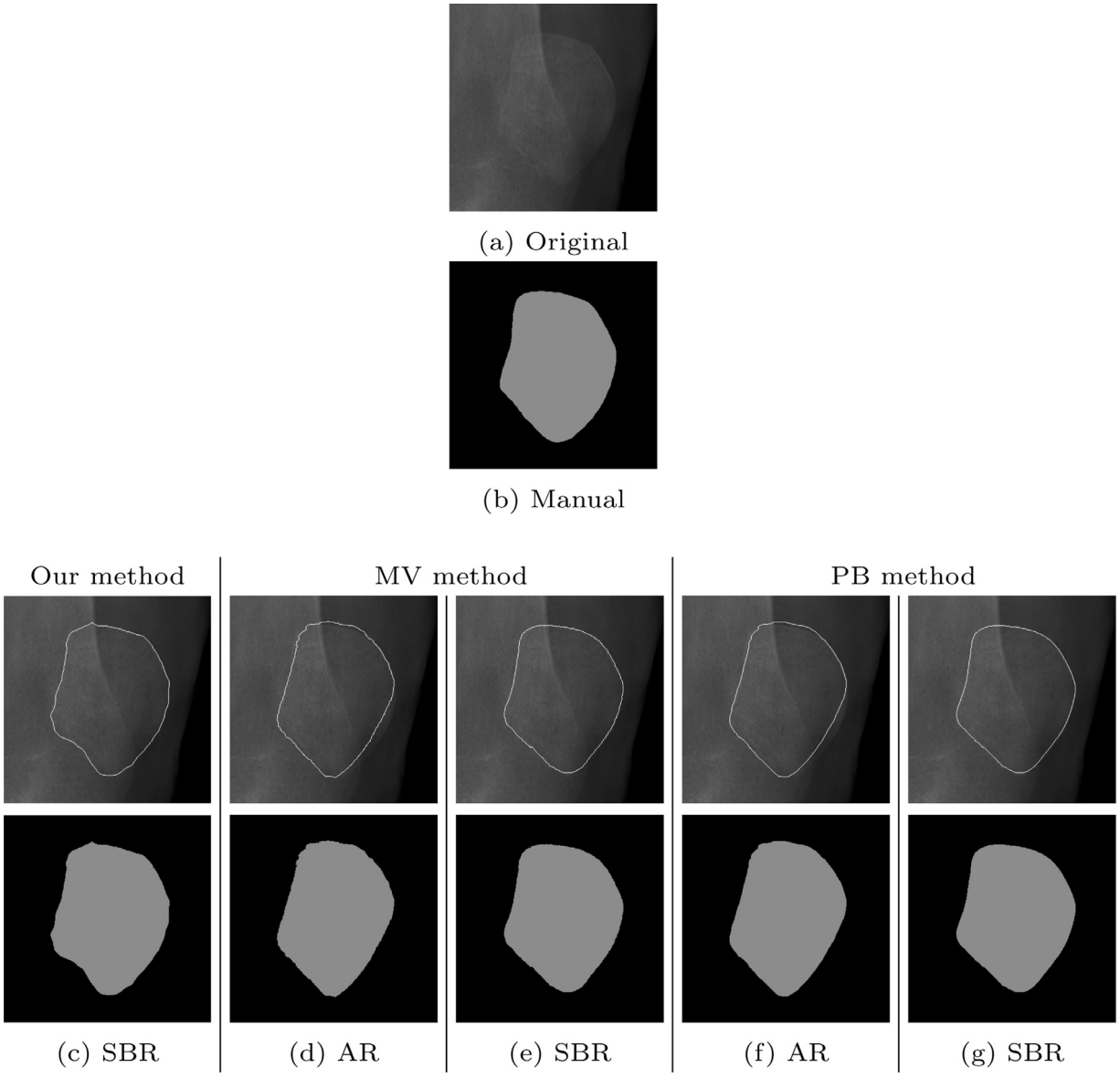


Fig. 13. Comparison of segmentation results of the patella. (a) original oblique X-ray radiographic image, (b) manual segmentation, third row: external patellar contours on the corresponding X-ray image, fourth row: resulting images.

makes our method original, robust and particularly appropriate for the segmentation of complex structures in X-ray imaging. To the best of our knowledge, no study has been done on the X-ray oblique image segmentation of the pelvis, talus, and patella. Unlike most existing multi-atlas-based segmentation approaches, the proposed method makes the following main contributions. First, the multi-atlas-based training dataset allows us to estimate, non-parametrically, a collection of superpixels which capture all the nonlinear and local variability of ROIs present in the multi-atlas dataset. This allows us both to increase the accuracy of the superpixel map pruning step while simplifying the optimization problem involved in this crucial step. Second, the contour-based similarity measure used in our first registration step and combined with the region-based similarity measure exploited in the final selection/filtering of the multi-atlas dataset allowed us to efficiently and adaptively single out the outlier segmentations, existing in the multi-atlas dataset, for each specific segmentation case. Third, the label propagation fusion step was herein performed in the variation of information sense and allowed us to refine the external contour of the bone structure to be segmented, but also, and especially, to infer the internal region labels from the preselected segmentations of the atlas to the final segmentation result. In addition, the proposed approach is robust since it mainly relies on the external bone contour of the bone structure to be segmented, which is the most informative and reliable visual cues in an X-ray image, especially in the case of the lower limb bone structures (such as the pelvis, talus, and patella). Also, the average classification rate of our method was within the range of accuracy typically obtained in case of manual segmentations (w.r.t. intra inter/observer variability). The proposed segmentation approach has potential clinical applications, one of them being three-dimensional reconstruction from biplanar images to assess skeletal morphology in the context of various pathological processes like osteoarthritis.

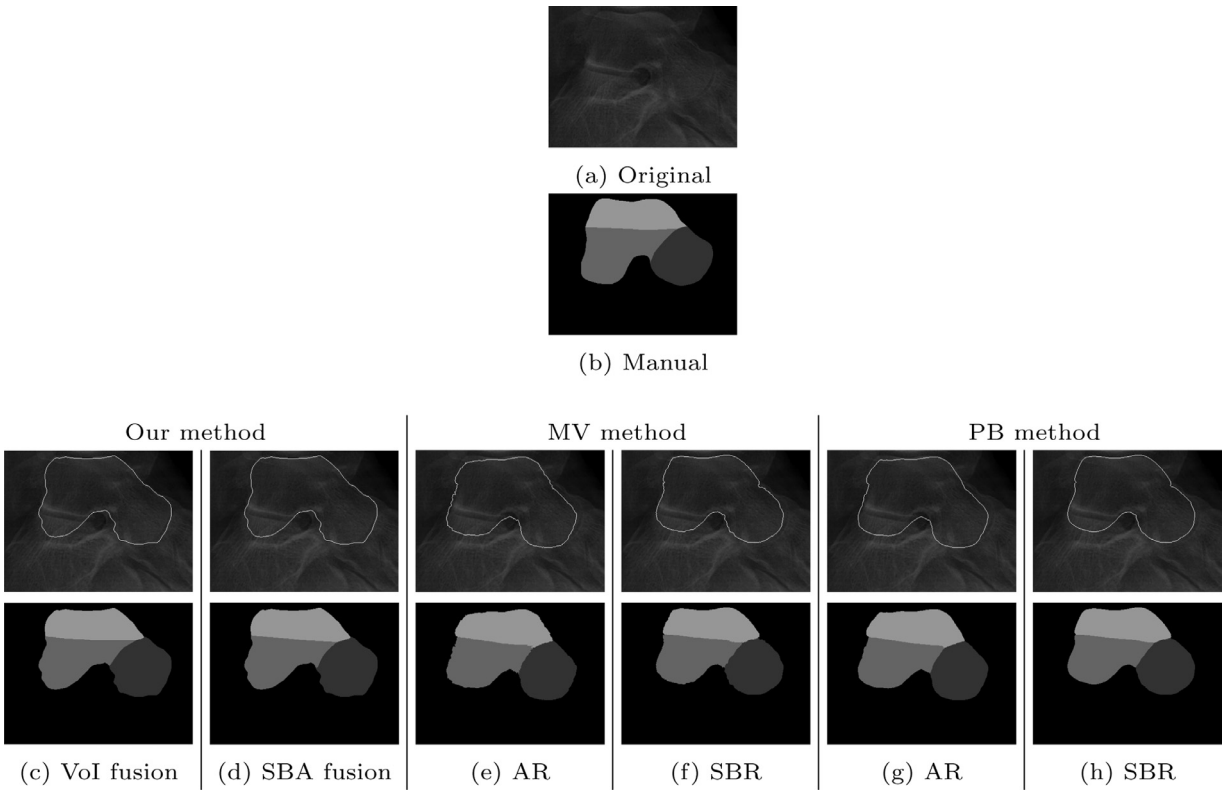


Fig. 14. Comparison of segmentation results of the talus. (a) original oblique X-ray radiographic image, (b) manual segmentation, third row: external talar contours on the corresponding X-ray image, fourth row: resulting images.

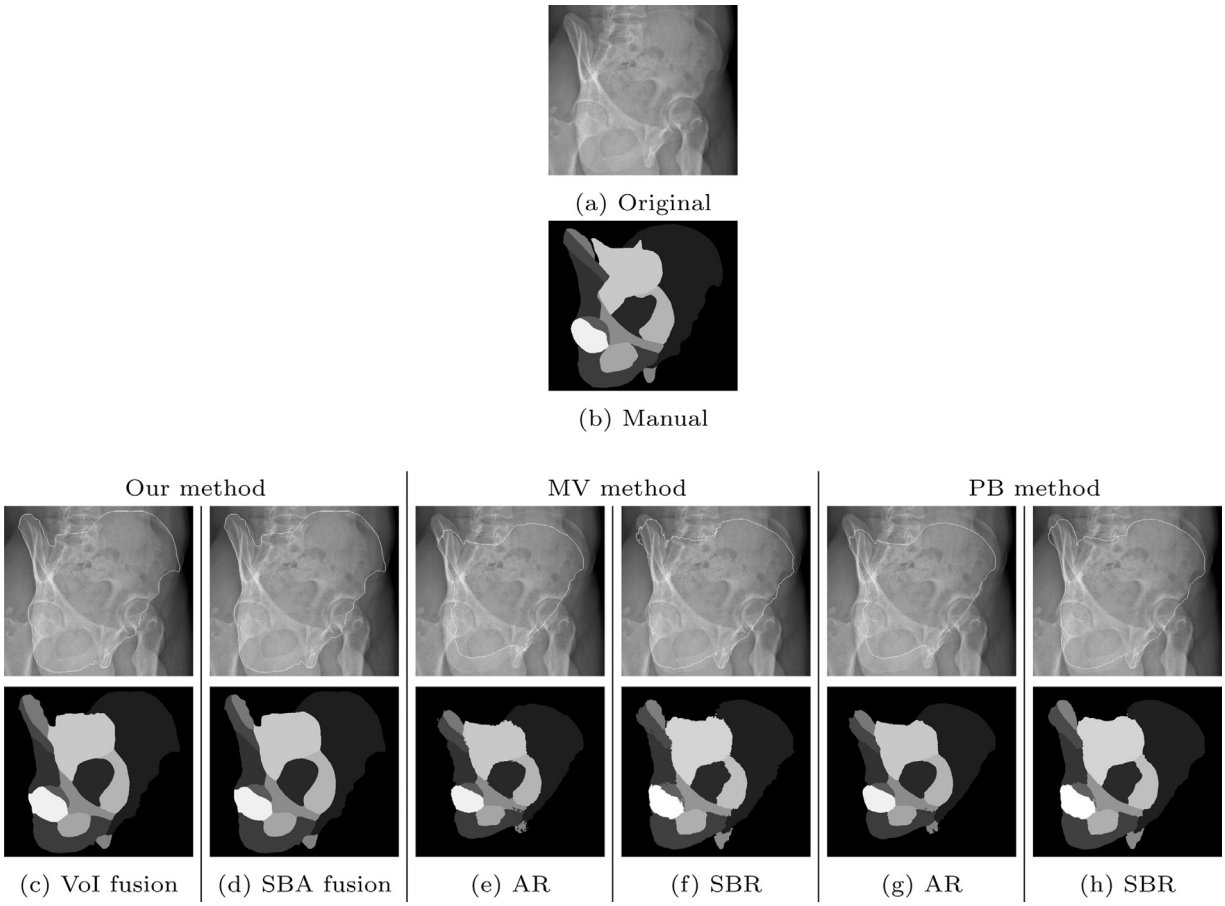
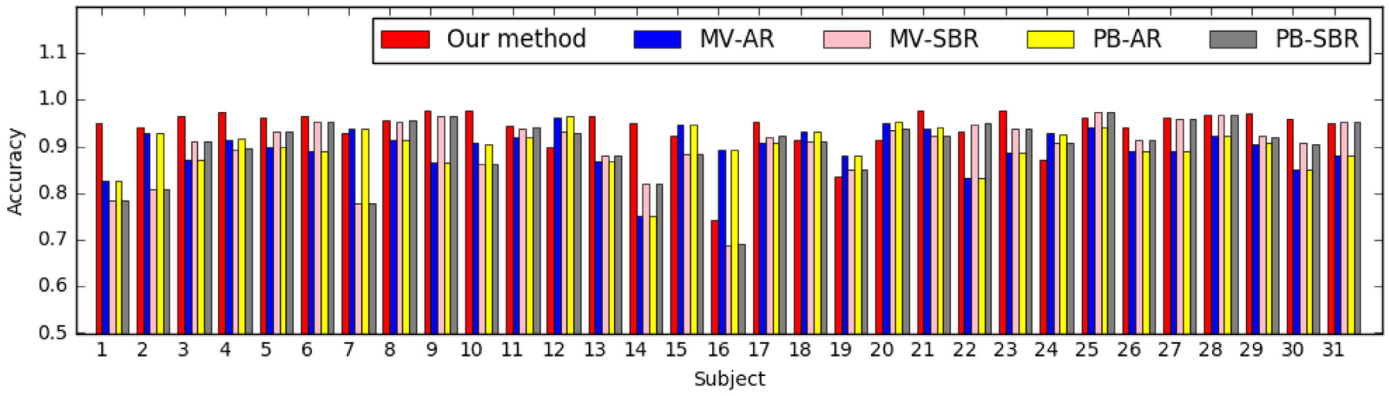
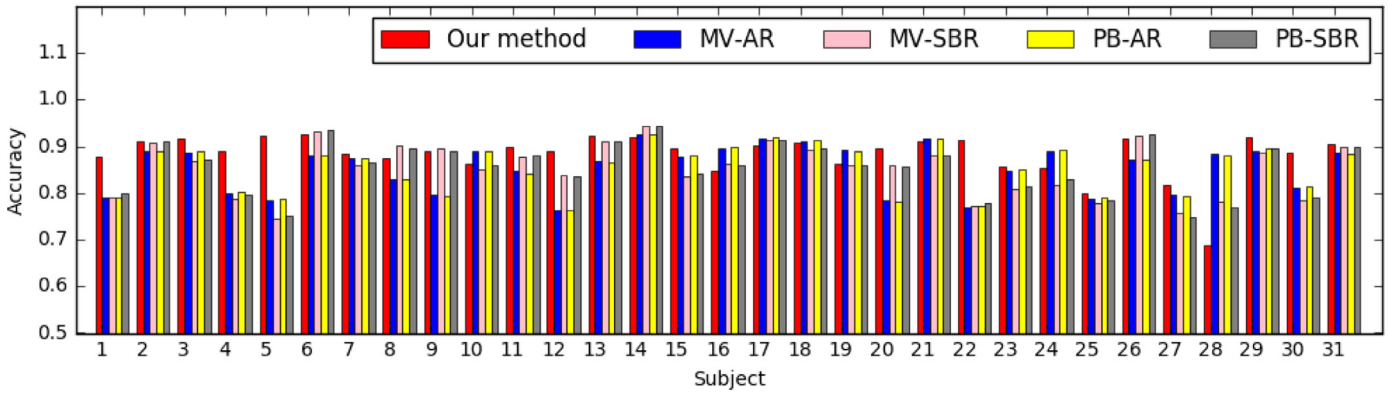


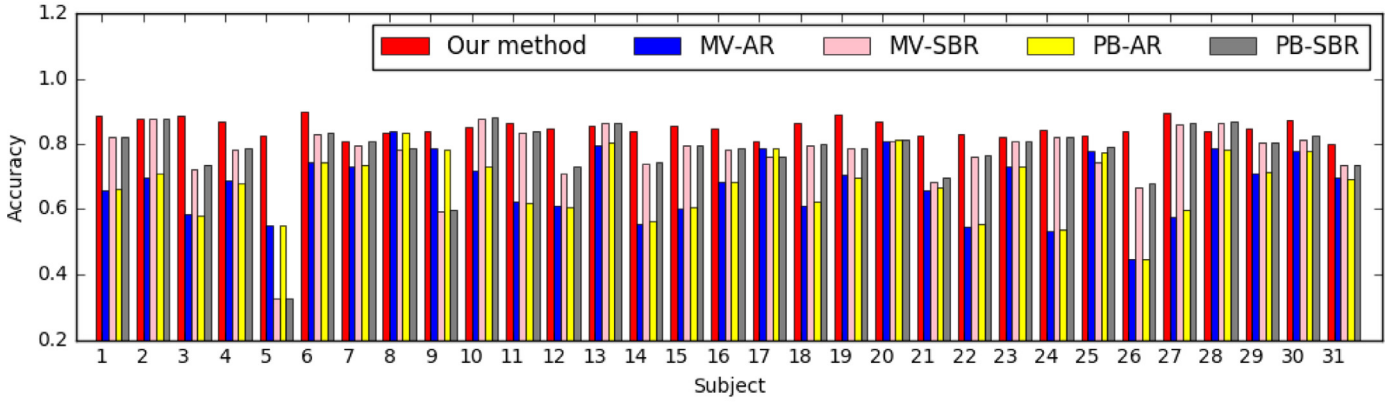
Fig. 15. Comparison of segmentation results of the pelvis. (a) original oblique X-ray radiographic image, (b) manual segmentation, third row: external pelvic contours on the corresponding X-ray image, fourth row: resulting images.



(a) Patella



(b) Talus



(c) Pelvis

Fig. 16. Accuracy of each subject in dataset. Each bin represents the similarity index of the resulting image comparing to the corresponding manual segmentation.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34 (11), 2274–2282.

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

Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solórzano, C., 2009. Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans. Med. Imaging 28 (8), 1266–1277.

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 seg-

mentation with registration refinement: application to cardiac mr images. IEEE Trans. Med. Imaging 32 (7), 1302–1315.
Benameur, S., Mignotte, M., Parent, S., Labelle, H., Skalli, W., de Guise, J., 2003. 3D/2d registration and segmentation of scoliotic vertebrae using statistical models. Comput. Med. Imaging Graph. 27 (5), 321–337.
Buades, A., Coll, B., Morel, J.-M., 2008. Nonlocal image and movie denoising. Int. J. Comput. Vis. 76 (2), 123–139.
Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L., 2010. Nonlocal patch-based label fusion for hippocampus segmentation. In: Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI 2010, 6363. Springer, pp. 129–136.
Destrempes, F., Mignotte, M., Angers, J.-F., 2007. Localization of shapes using statistical models and stochastic optimization. IEEE Trans. Pattern Anal. Mach. Intell. 29 (9), 1603–1615.
Dowling, J.A., Fripp, J., Chandra, S., Pluim, J.P.W., Lambert, J., Parker, J., Denham, J., Greer, P.B., Salvado, O., 2011. Fast automatic multi-atlas segmentation of the prostate from 3D MR images. In: Proceedings of the International Conference on Prostate Cancer Imaging: Image Analysis and Image-guided Interventions. Springer-Verlag, Berlin, Heidelberg, pp. 10–21.
Felzenszwalb, P.F., Huttenlocher, D.P., 2004. Efficient graph-based image segmentation. Int. J. Comput. Vis. 59 (2), 167–181.
Huo, J., Wang, G., Wu, Q.J., Thangarajah, A., 2015. Label fusion for multi-atlas segmentation based on majority voting. In: Proceedings of the International Conference Image Analysis and Recognition. Springer, pp. 100–106.
Jodoin, P.-M., Mignotte, M., Rosenberger, C., 2007. Segmentation framework based on label field fusion. IEEE Trans. Image Process. 16 (10), 2535–2550.
Meilă, M., 2005. Comparing clusterings: an axiomatic view. In: Proceedings of the 22nd International Conference on Machine Learning. ACM, Bonn, Germany, pp. 577–584.
Meilă, M., 2007. Comparing clusterings-an information based distance. J. Multivar. Anal. 98 (5), 873–895.
Mignotte, M., 2011a. A de-texturing and spatially constrained k-means approach for image segmentation. Pattern Recognit. Lett. 32 (2), 359–367.
Mignotte, M., 2011b. MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation. IEEE Trans. Neural Netw. 22 (3), 447–460.
Mignotte, M., 2014a. A label field fusion model with a variation of information estimator for image segmentation. Inf. Fusion 20, 7–20.
Mignotte, M., 2014b. A non-stationary MRf model for image segmentation from a soft boundary map. Pattern Anal. Appl. 17 (1), 129–139.
Mignotte, M., Collet, C., Perez, P., Bouthemy, P., 2000. Sonar image segmentation using an unsupervised hierarchical MRF model. IEEE Trans. Image Process. 9 (7), 1216–1231.
Mignotte, M., Meunier, J., 2001. A multiscale optimization approach for the dynamic contour-based boundary detection issue. Comput. Med. Imaging Graph 25 (3), 265–275.
Mignotte, M., Meunier, J., Soucy, J.-P., 2008. DCT-Based complexity regularization for EM tomographic reconstruction. IEEE Trans. Biomed. Eng. 55 (2), 801–805.
Mignotte, M., Meunier, J., Tardif, J.-C., 2001. Endocardial boundary e timation and tracking in echocardiographic images using deformable template and Markov random fields. Pattern Anal. Appl. 4 (4), 256–271.
Morin, J.-P., Desrosiers, C., Duong, L., 2012. A random walk approach for multiatlas-based segmentation. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR),. IEEE, pp. 3636–3639.
Nguyen, D.C.T., Benameur, S., Mignotte, M., Lavoie, F., 2015. Superpixel and entropy-based multi-atlas fusion framework for the segmentation of x-ray images. In: Proceedings of the 18th International Conference on Image Analysis and Processing, ICIAP'15, Lecture Notes in Computer Science, 9280, pp. 151–161. Genova, Italy
Ooura, T., 2001. General Purpose FFT (fast fourier/cosine/sine transform) Package. http://momonga.t.u-tokyo.ac.jp/~ooura/fft.html.
Ren, X., Malik, J., 2003. Learning a classification model for segmentation. In: Proceedings of the 9th IEEE International Conference on Computer Vision, 1. IEEE, p. 10.
Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R., 2004a. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 21 (4), 1428–1442.
Rohlfing, T., Maurer, C.R., 2007. Shape-based averaging. IEEE Trans. Image Process. 16 (1), 153–161.
Rohlfing, T., Russakoff, D.B., Maurer, C.R., 2004b. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Trans. Med. Imaging 23 (8), 983–994.
Rousseau, F., Habas, P.A., Studholme, C., 2011. A supervised patch-based approach for human brain labeling. IEEE Trans. Med. Imaging 30 (10), 1852–1862.
Shi, W., Lombaert, H., Bai, W., Ledig, C., Zhuang, X., Marvao, A., Dawes, T., ORegan, D., Rueckert, D., 2014. Multi-atlas spectral patchmatch: application to cardiac image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI, 8673. Springer, pp. 348–355.
Vega-Pons, S., Ruiz-Shulcloper, J., 2011. A survey of clustering ensemble algorithms. Int. J. Pattern Recognit. Artif. Intell. 25 (03), 337–372.
Wang, H., Suh, J.W., Das, S., Pluta, J., Altinay, M., Yushkevich, P., 2011. Regression-based label fusion for multi-atlas segmentation. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR),. IEEE, pp. 1113–1120.
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 (3), 611–623.
Wang, H., Yushkevich, P.A., 2013. Multi-atlas segmentation without registration: a supervoxel-based approach. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 535–542.
Wu, G., Kim, M., Sanroma, G., Wang, Q., Munsell, B.C., Shen, D., Initiative, A.D.N., et al., 2015. Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition. Neuroimage 106, 34–46.
Wu, G., Wang, Q., Zhang, D., Shen, D., 2012. Robust patch-based multi-atlas labeling by joint sparsity regularization. In: Proceedings of the MICCAI Workshop on Sparsity Techniques in Medical Imaging (STMI).
Yu, G., Sapiro, G., 2011. DCT Image denoising: a simple and effective image denoising algorithm. Image Process. Line 1, 292–296.
Yu, N., Wang, H., Yushkevich, P.A., 2016. Supervoxel-based hierarchical Markov random field framework for multi-atlas segmentation. In: Proceedings of the International Workshop on Patch-based Techniques in Medical Imaging. Springer, pp. 100–108.