Magnetic Resonance Image Selection for Multi-Atlas Segmentation Using Mixture Models

Mauricio Orbes-Arteaga[1], David Cárdenas-Peña[1(✉)], Mauricio A. Álvarez[2], Alvaro A. Orozco[2], and Germán Castellanos-Dominguez[1]

[1] Signal Processing and Recognition Group,
Universidad Nacional de Colombia, Manizales, Colombia
dcardenasp@unal.edu.co
[2] Grupo de Investigación en Automática,
Universidad Tecnológica de Pereira, Pereira, Colombia

Abstract. In this paper, magnetic resonance image similarity metrics based on generative model induced spaces are introduced. Particularly, three generative-based similarities are proposed. Metrics are tested in an atlas selection task for multi-atlas-based image segmentation of basal ganglia structure, and compared with the mean square metric, as it is assessed on the high dimensional image domain. Attained results show that our proposal provides a suitable atlas selection and improves the segmentation of the structures of interest.

Keywords: Generative embedding · Fisher score · Magnetic resonance imaging · Multi-atlas segmentation · Template selection

1 Introduction

Brain magnetic resonance images (MRI) play an important role in the diagnosis and treatment of medical diseases. Applications such as disease progression, brain mapping, and surgery planning require of accurate brain structure segmentation [1]. However, such task is difficult to perform due to the presence of artifacts and low contrast between the tissues, mainly inside the subcortical region.

Atlas-based techniques are commonly used for dealing with the above constraints, as they allow to include shape and intensity distribution of any structure as *a priori* knowledge (atlas). To this end, the atlases are usually non-linearly mapped to the target image space and finally combined into a single labeling image using a procedure, known as atlas voting or label fusion. However, as brain shapes are not unimodal distributed, anatomically non relevant atlases can bias the achieved segmentation. Moreover, the computational cost linearly increases with the number of atlases to be registered. To overcome these issues, multi-atlas approaches have been proposed to properly select and combine the independent

A. Pardo and J. Kittler (Eds.): CIARP 2015, LNCS 9423, pp. 391–399, 2015.
DOI: 10.1007/978-3-319-25751-8_47

contributions of an atlas [2]. In this sense, proper selection of atlases allows to improve the resulting segmentation, while keeping the number of atlases as low as possible. Usually, the selection criterion is based on image similarity metrics [3], as mutual information or mean squares, being computed in a common image space [2] or in the target image space [4]. However, a large similarity value does not necessarily imply high quality propagated labels [5]. Moreover, as the metrics are computed in the original image space, the intrinsic image morphological properties may not be highlighted.

Hence, new methodologies have been introduced to find low dimensional spaces to map the images and assess the similarities. Techniques as manifold learning [6] and locality preserving projections [7] have been used in this regard. On the other hand, generative embeddings have proved to be efficient for representation and discrimination of high dimensional data structures [8]. These approaches take advantage of the low dimensional space induced by the generative model parameters or scores.

In this work, we propose to use the induced metrics from generative models as image similarity function in the atlas selection in a multi-atlas segmentation scheme. Our proposal uses generative models to represent MRI, so they are mapped into a more compact and discriminative space highlighting anatomical differences. In order to compute the image similarities, three probability based approaches are considered: Likelihood-based, Parameter-based, and Fisher score-based. Additionally, as the intensity probability distribution of the image is unknown, Gaussian and Student's t mixture models are used to estimate it. Obtained results show that the similarities in the new representation space achieve a more suitable selection of atlases improving the segmentation accuracy compared with metrics computed in the original image space, such as the means square.

2 Materials and Methods

2.1 Multi-atlas Based Segmentation

The input MRI space is described as follows: Let $\mathscr{X}=\{\boldsymbol{X}^n,\boldsymbol{L}^n{:}n=1,\ldots,N\}$ be a labeled MRI dataset holding N image-segmentation pairs, where $\boldsymbol{X}^n=\{x_r^n\in\mathbb{R}{:}\ r\in\Omega\}$ is the n-th MR image, the value r indexes all spatial elements (*spel*), and $\boldsymbol{L}^n=\{l_r^n\in\{1,C\}{:}r\in\Omega\}$ is the provided image segmentation into $C\in\mathbb{N}$ classes. In the case of 3D-volume analysis, both, $\boldsymbol{X}^n$ and $\boldsymbol{L}^n$, have dimension $\Omega=\mathbb{R}^{T_a\times T_s\times T_c}$, with $\{T_a,T_s,T_c\}$ being the Axial, Sagittal, and Coronal sizes, respectively. Thus, the segmentation of each target image is accomplished by combining the subset holding the most similar labeled atlases of $\mathscr{X}$, which are selected by a given similarity criterion.

In addition, the majority vote strategy is used to carry out MRI segmentation. This straightforward procedure assigns the most agreed label among the selected atlases to each spel of the target image. So, let $\boldsymbol{X}^q$ be a target image and $\mathscr{X}_q=\{\boldsymbol{X}^t,\boldsymbol{L}^t{:}t=1,\ldots,T\}$ be a subset of $\mathscr{X}_q\subset\mathscr{X}$ that holds $T\leq N$ selected atlases, which are ranked by the similarity measure $\kappa\{\cdot,\cdot\}$, so that

$\kappa\{\boldsymbol{X}^q, \boldsymbol{X}^t\} > \kappa\{\boldsymbol{X}^q, \boldsymbol{X}^{t+1}\}$. Also, let $\hat{\boldsymbol{L}}^t$ be the provided segmentation of the t-th atlas after carrying out the deformable registration of the target image so that the matrix $\hat{\boldsymbol{L}}^t$ assigns the label $\hat{l}_r^t$ to each spel r. Afterward, labeling from all atlases is gathered into a single estimated segmentation matrix $\hat{\boldsymbol{L}}^q$ with elements:

$$\hat{l}_r^q = \underset{c\in\{1,C\}}{\arg\max} \sum\nolimits_{t\in T} \delta(\hat{l}_r^t - c),\ \hat{l}_r^q \in \{1, C\}$$

where $\delta(\cdot)$ is the delta Dirac function. So, the image similarity function is crucial for selecting the closest templates to target images.

2.2 Generative Mixture Models for Extracting Image Features

Provided a set of parameters Θ, the intrinsic MRI features are proposed to be described by a generative model maximizing the conditional probability, $P(\boldsymbol{X}|\Theta)$. For fitting parameters Θ to a given image $\boldsymbol{X}=\{x_r\in\mathbb{R}:r\in\Omega\}$, this task is equivalent to the minimization of the negative log-likelihood cost function, $J=-\log P(\boldsymbol{X}|\Theta)$, that under the assumption of independent and identically distributed spel intensities is written as:

$$\Theta^* = \underset{\Theta}{\arg\min}\{-\sum\nolimits_{r\in\Omega} \log P(x_r|\Theta)\} \tag{1}$$

where $P(x_r|\Theta)$ is the probability that a pixel has intensity x_r, given the model parameters. Since the MR images may include several structures with different intensity ranges, we hypothesize that image features are better described by mixture models with $\boldsymbol{K}$ components and parameters $\Theta=\{\theta_k:k=1,\ldots,K\}$. Thus, the conditional probability is written as: $P(x_r|\Theta)=\sum_{k\in K}\omega_k P_k(x_r|\theta_k)$, subject to: $\sum_{k\in K}\omega_k=1$, where the mixture weight, $\omega_k\in\mathbb{R}^+$, stands for the prior probability of each spel to belong to the k-th component. $P_k(x_r|\theta_k)$ is the class conditional probability for the k-th component. We will discus the use of the following functions:

- *Gaussian Distribution*: $P_k(x_r|\theta_k)=\frac{1}{\sigma_k\sqrt{2\pi}}\exp\{-(x_r-\mu_k)^2/2\sigma_k^2\}$, where in the parameter set $\theta_k=\{\omega_k, \mu_k, \sigma_k\}$, $\mu_k\in\mathbb{R}$ is the mean and $\sigma_k\in\mathbb{R}^+$ is standard deviation.
- *Students't Distribution* ($\Gamma(\cdot)$ notates the Gamma function):

$$P_k(x_r|\theta_k)=\frac{\Gamma((\nu_k+1)/2)}{\Gamma(\nu_k/2)\sqrt{\pi\nu_k}\sigma_k}\left(1+\frac{1}{\nu_k}\left(\frac{x_r-\mu_k}{\sigma_k}\right)^2\right)^{-(\nu_k+1)/2}$$

  Therefore $\theta_k=\{\omega_k, \mu_k, \sigma_k, \nu_k\}$, with $\nu_k\in\mathbb{R}^+$ as the degrees of freedom.

2.3 Generative-Model Based Measures of Pair-Wise Image Similarity

- *Likelihood-based similarity:* Due to the $\log P(\boldsymbol{X}^m|\Theta^n)$ is the probability that the image $\boldsymbol{X}^m$ is generated by the model parameters Θ^n, the following log-likelihood measure of pairwise similarity is defined [8]:

$$\kappa\{\boldsymbol{X}^n, \boldsymbol{X}^m\} = \log P(\boldsymbol{X}^m|\Theta^n) \tag{2}$$

– *Parameter-based similarity:* For better handling of size variant images, each image is represented by a concatenated parameter vector Θ that results from the model optimization, making the image characterization be only dependent on the size of the model parameters instead of the whole image domain:

$$\kappa\{\boldsymbol{X}^n, \boldsymbol{X}^m\} = f(\Theta^n, \Theta^m) \tag{3}$$

where $f(\Theta^n, \Theta^m)$ is a similarity function between the vectors Θ^n and Θ^m.

– *Fisher-score-based similarity:* For describing the direction in which the model parameters should be modified to better fit the data, the gradient of the log-likelihood in Eq. (1), $\nabla_\Theta \log P(\boldsymbol{X}|\Theta)$ is used (termed the Fisher score [8]) as follows:

$$\kappa\{\boldsymbol{X}^n, \boldsymbol{X}^m\} = f(\nabla_{\Theta^n} J(X^n|\Theta^n), \nabla_{\Theta^m} J(X^m|\Theta^m)) \tag{4}$$

where $\nabla_\Theta J(X|\Theta) = \{\partial J/\partial \mu_k, \partial J/\partial \sigma_k\}_{k=1}^K$ for the case of the Gaussian mixtures, and $\nabla_\Theta J(X|\Theta) = \{\partial J/\partial \mu_k, \partial J/\partial \sigma_k, \partial J/\partial \nu_k\}_{k=1}^K$ for Student's t mixtures.

3 Experimental Set-Up

To evaluate the performance of the proposed measures of similarity between MRIs, a multi-atlas segmentation scheme is considered so that the atlases are ranked according to the degree of similarity with a target image. Also, the corresponding label images are combined through a majority voting scheme for estimating the final segmentation. Thus, the evaluation process have the following stages: *i)* Image preprocessing, *ii)* Generative model optimization, and *iii)* Similarity metric evaluation for the Atlas voting.

3.1 MRI Database and Image Preprocessing

The MRI collection used is a subset of the Open Access Series of Imaging Studies (OASIS) database that was proposed for the MICCAI 2012 *Multi-atlas labeling and Statical Fusion Challenge*. The dataset holds T1-Weighted structural MRI scans from 35 subjects (13 males and 22 females) aging from 18 to 90 years old. Each 256×256×287 MRI volume has a voxel size of 1×1×*1mm*. All images were expertly labeled for 26 structures. Due to our research interest in Parkinson surgery, only the following structures are considered: hypothalamus, amygdala, putamen, caudate nucleus, thalamus, and pallidum. Fig. 1 shows a sample image subject and its segmentation provided.

To measure image similarities within a single common image space, input MRI set is spatially normalized into the Talairach space using a rigid alignment to the MNI305 atlas. For the label propagation, every atlas image is also spatially mapped into the target image spatial coordinates with an elastic deformation (ANTS toolbox[1]).

[1] http://picsl.upenn.edu/software/ants

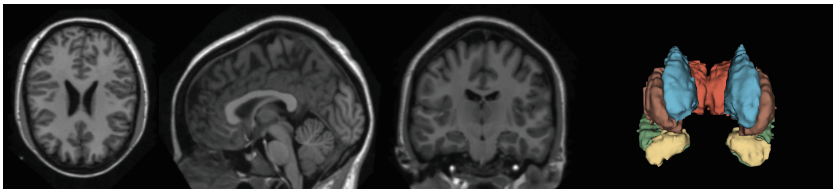
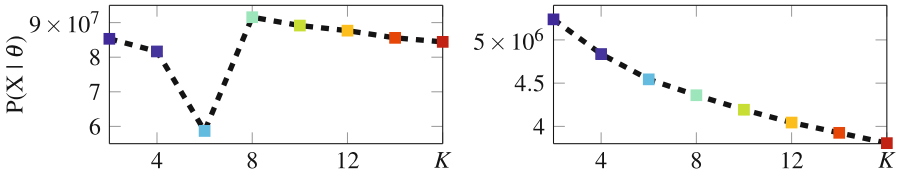


Fig. 1. Left to Right: Axial, Sagittal, Coronal views, and ground-truth segmented structures.

3.2 Generative Model Optimization

We use the Expectation-Maximization (EM) algorithm to find the parameter set of the generative mixture models representing an image (see Eq. (1)). Aiming each mixture to represent the same regions in all images, the EM at each of them is initialized as follows: i) k-means algorithm is performed over a subset of randomly taken spels from the input dataset. ii) Those resulting centroids are used further as seeds for the EM.

The ability of each distribution function considered for describing the input MRI set is analyzed by incrementing the number of mixtures $K{=}2,\ldots,16$ on the model as seen in Fig. 2 showing the average log-likelihood of the Gaussian mixture model. Fig. 2a relates the case when the whole image is fit, and Fig. 2b – when fitting only the region of interest (ROI) corresponding to the basal ganglia location. As a result, the former GMM fitting becomes more complex due to the larger amount of structures of the entire image. It is worth noting, for the ganglia region, that the larger is the number of mixtures, the better the fitting in the generative process. Nevertheless, the model can be over-fitted for a considerable number of mixtures.



(a) Whole image modeling (b) ROI modeling

Fig. 2. Gaussian Mixture Model fitting for several number of mixtures.

For the case of the Student's t mixture, we evaluate the image fitting by varying the degrees of freedom, $\nu{=}1,\ldots,20$. As seen on the top file in Fig. 3, the results achieved of the ROI modeling show that lower values of ν allow improving the generative model performance. The whole image modeling seems to have the same behavior but with worse consistency. Additionally, the bottom file in Fig. 3 show the degrees of freedom obtained from EM for a given number of mixtures. Thus, ν tends to be more stable when the number of components is increased. However, it is known that the Student's t resembles the Gaussian distribution for large values of ν.

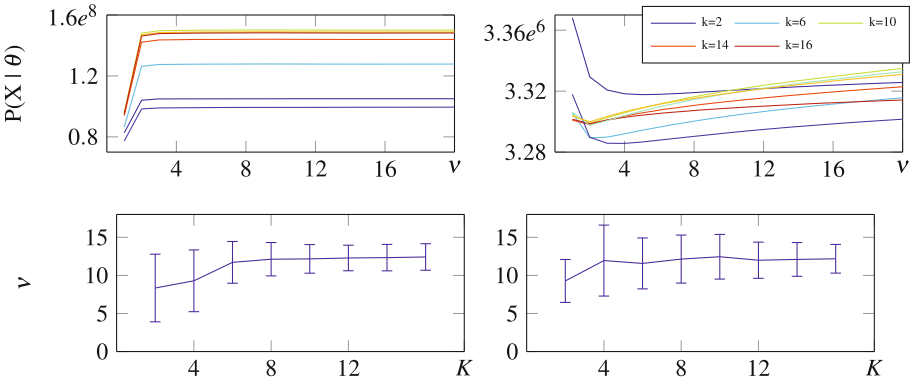


Fig. 3. Student's t parameter tuning. **Top:** Log-likelihood versus the degrees of freedom. **Bottom:** Degrees of freedom versus number of mixtures (subject mean and standard deviation depicted). **Left:** Whole image modeling. **Right:** ROI modeling.

3.3 Validation of Image Similarity in Multi-atlas-Based Segmentation Tasks

We assess the performance of the proposed similarity approaches within an atlas selection task, where all structures are segmented using the atlas-voting label propagation approach in the target image space. Specifically for the (see Eq. (3)) parameter-based and (Eq. (4)) Fisher score-based measures, we make use of the Gaussian kernel, $f(\Theta^n, \Theta^m)$=exp($\|\Theta^n - \Theta^m\|_2^2/2\sigma_f^2$), as the similarity function between feature vectors, where the scale parameter $\sigma_f \in \mathbb{R}^+$ is tuned using the maximum dispersion criterion [9]. For the sake of comparison, we also assess as a similarity metric the voxel-wise Mean Squares (MS) in the image domain space.

Fig. 4 shows the results for multi-atlas segmentation using Gaussian distributions. As seen on the top file in Fig. 4 modeling the whole image, the accuracy achieved by all image similarities is not affected by the number of components used to model the input MRI space. However, the Fisher score-based measure outperforms the others. In turn, the bottom file in the figure Fig. 4 display the ROI-based modeling performance that improves the one achieved by the whole image modeling. Particularly, in Fig. 4 the bottom left figure shows that the larger the number of mixtures the lower the segmentation performance due to the model over-fitting at each image. Once again, the Fisher score-based selection outperforms the other strategies; this result may be explained since the derivatives take into account the degree of agreement between models.

Likewise, we estimate the accuracy using the Student's t distribution as seen in the top file in Fig. 5, showing a similar performance to the Gaussian distribution for the whole image modeling. For the ROI modeling, the parameter-based and Fisher score-based selection methods achieve the highest accuracy at a larger number of mixtures.

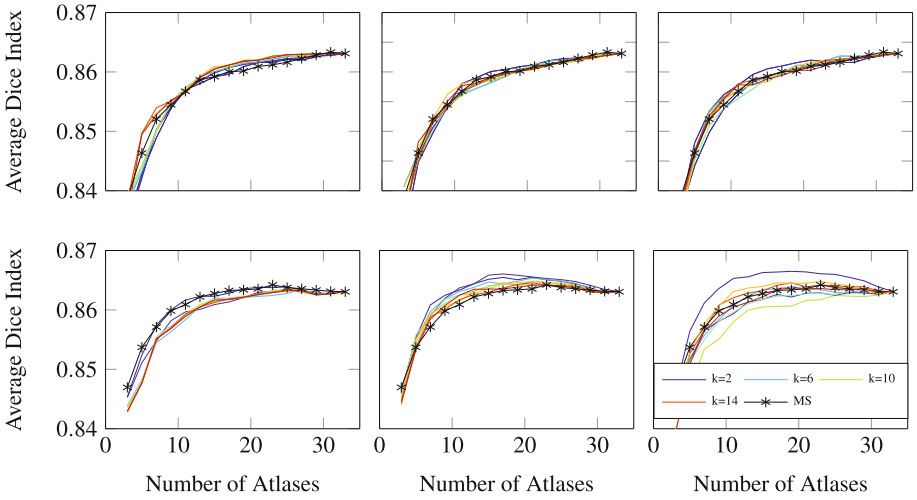


Fig. 4. Dice Index versus the number of atlases for all considered image similarities using Gaussian mixtures. **Top:** Whole image modeling. **Bottom:** ROI image modeling. **Left to Right:** Likelihood-based, Parameter-based and Fisher score-based atlas selection.

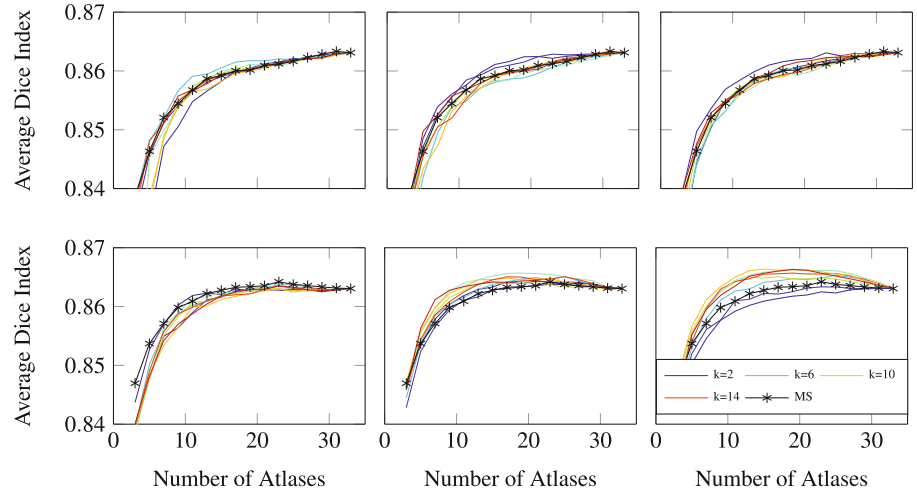


Fig. 5. Dice Index versus the number of atlases for all considered image similarities using Student's t mixtures. **Top:** Whole image modeling. **Bottom:** ROI image modeling. **Left to Right:** Likelihood-based, Parameter-based and Fisher score-based atlas selection.

4 Discussion

We introduce a new strategy for measuring MRI similarities supporting a multi-atlas segmentation scheme. The proposal allows computing pairwise similarities in a low dimensional space being induced by a generative model. As a

result, a new space becomes more discriminative and compact than the original spel-wise image representation. For computing the pair-wise image affinity, three approaches are proposed, namely, Likelihood-based, Parameter-based, and Fisher score-based similarity. Moreover, two different distributions are asssumed as components of the mixture model: Gaussian and Student's t distributions.

The training approach is validated on selecting the most similar atlases to a target, and then combining them into a single segmentation. For testing, two strategies of feature extraction are considered: whole image and the ROI. Obtained results for the basal ganglia location show that the proposed approach outperforms the segmentation achieved by the baseline spel-wise MS metric.

Regarding the mixture base distributions, it is clear that for the Gaussian distribution the larger the number of mixtures, the better the fitting for the generative process. However, for modeling the whole image, the number of required components tends to be large, due to the whole image holding significantly more structures than the ROI. Contrarily, for the Student's t distribution, a better fitting is achieved as the number of components decrease. Taking into account that the degrees of freedom parameter, ν, allows to differentiate a Student's t from Gaussian like shapes, the tuning of such parameter is more complex than the location parameter and the scale parameter, at each mixture. This is mainly because the log-likelihood cost function is not convex. Hence, the parameter tuning may lead to suboptimal values, which are different to the ones obtained by exhaustive search, as seen in the Fig. 3. In this sense, we conclude that the Gaussian distribution is more appropriate for modeling the images.

For the sake of evaluation, all considered image similarities are used as a selection criterion in the multi-atlas segmentation scheme. According to results in the right column in Figs. 4 and 5, the likelihood-based similarity approach achieves the worst accuracy, this is because the similarity measure becomes highly sensitive to poorly estimated or improper models. In the parameter-based approach, each image is characterized by the vector of estimated parameters, and the similarity is measured by comparing vector pairs using a Gaussian kernel function. Obtained results in the middle columns in Figs. 4 and 5 show that improvement in accuracy is achieved with respect to likelihood-based similarity. Therefore, the information captured by the vector of parameters is more discriminative than the obtained by assessing the likelihood over the images, specially in high dimensional images. Also, in order to capture the influence of the parameters on the generative process, the gradient of the log-likelihood cost function with respect to the parameters is used as feature extraction. As a result this measure selects a more appropriate subset of atlases than the former introduced measures, achieving a higher accuracy respect to them, as seen in Figs. 4 and 5. Finally, aiming to compare the proposals against the conventional similarities, the well know Mean squares is used. In this case, the new similarities outperform the MS baseline in the atlas selection task, with the advantage of being only dependent on the number of parameters in the model, which is considerably smaller than the number of spels an image.

As a future work, other image generative models are to be tested (e.g. Markov Random Fields), where the spatial information is also taken into consideration, providing more robust estimation to the artifacts present on MR images. Other methods for model comparison, such as dissimilarities and kernel methods, will also be included.

Acknowledgments. This work was supported by Programa Nacional de Formación de Investigadores "Generación del Bicentenario", 2011/2012, the research projects 111065740687, 111056934461 and 20101008258 funded by COLCIENCIAS.

References

1. Despotovic, I., Vansteenkiste, E., Philips, W.: Brain volume segmentation in newborn infants using multi-modal mri with a low inter-slice resolution. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2010, pp. 5038–5041 (2010)
2. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage **46**(3), 726–738 (2009)
3. Cárdenas-Peña, D., Orbes-Arteaga, M., Castellanos-Dominguez, G.: Supervised brain tissue segmentation using a spatially enhanced similarity metric. In: Vicente, J.M.F., Álvarez-Sánchez, J.R., de la Paz López, F., Toledo-Moreo, F.J., Adeli, H. (eds.) Artificial Computation in Biology and Medicine. LNCS, vol. 9107, pp. 398–407. Springer, Heidelberg (2015)
4. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solorzano, C.: Combination strategies in multi-atlas image segmentation: application to brain mr data. IEEE Transactions on Medical Imaging **28**(8), 1266–1277 (2009)
5. Langerak, T.R., Berendsen, F.F., Van der Heide, U.A., Kotte, A.N.T.J., Pluim, J.P.W.: Multiatlas-based segmentation with preregistration atlas selection. Medical Physics **40**(9), 091701 (2013)
6. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D.: Leap: learning embeddings for atlas propagation. NeuroImage **49**(2), 1316–1325 (2010)
7. Cao, Y., Yuan, Y., Li, X., Turkbey, B., Choyke, P.L., Yan, P.: Segmenting images by combining selected atlases on manifold. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 272–279. Springer, Heidelberg (2011)
8. Bicego, M., Murino, V., Figueiredo, M.A.T.: Similarity-based classification of sequences using hidden markov models. Pattern Recognition **37**, 2281–2291 (2004)
9. Álvarez Meza, A.M., Cárdenas-Peña, D., Castellanos-Dominguez, G. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications SE - 41