Published in final edited form as:

Patch Based Tech Med Imaging (2015). 2015 October ; 9467: 28–36. doi:10.1007/978-3-319-28194-0_4.

Isointense Infant Brain Segmentation by Stacked Kernel Canonical Correlation Analysis

Li Wang^a, Feng Shi^a, Yaozong Gao^{a,b}, Gang Li^a, Weili Lin^c, and Dinggang Shen^a

^aIDEA Lab, Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

^bDepartment of Computer Science, University of North Carolina at Chapel Hill, NC, USA

^cMRI Lab, Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

Abstract

Segmentation of isointense infant brain (at ~6-months-old) MR images is challenging due to the ongoing maturation and myelination process in the first year of life. In particular, signal contrast between white and gray matters inverses around 6 months of age, where brain tissues appear isointense and hence exhibit extremely low tissue contrast, thus posing significant challenges for automated segmentation. In this paper, we propose a novel segmentation method to address the above-mentioned challenges based on stacked kernel canonical correlation analysis (KCCA). Our main idea is to utilize the 12-month-old brain image with high tissue contrast to guide the segmentation of 6-month-old brain images with extremely low contrast. Specifically, we use KCCA to learn the common feature representations for both 6-month-old and the subsequent 12-month-old brain images of same subjects to make their features comparable in the common space. Note that the longitudinal 12-month-old brain images are not required in the testing stage, and they are required only in the KCCA based training stage to provide a set of longitudinal 6- and 12-month-old image pairs for training. Moreover, for optimizing the common feature representations, we propose a stacked KCCA mapping, instead of using only the conventional one-step of KCCA mapping. In this way, we can better use the 12-month-old brain images as multiple atlases to guide the segmentation of isointense brain images. Specifically, sparse patch-based multi-atlas labeling is used to propagate tissue labels in the (12-month-old) atlases and segment isointense brain images by measuring patch similarity between testing and atlas images with their learned common features. The proposed method was evaluated on 20 isointense brain images via leave-one-out cross-validation, showing much better performance than the state-of-the-art methods.

1 Introduction

The first year of life is the most dynamic phase of postnatal human brain development, with the rapid tissue growth and development of a wide range of cognitive and motor functions. To accurately measure the early brain development, one critical step is to segment the infant brain image into different regions of interest (ROIs), e.g., white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). It is well-known that the segmentation of infant brain MRI is considerably more difficult than that of the adult brain MRI [1], due to reduced tissue contrast [2], increased noise, severe partial volume effect [3], and ongoing white matter

myelination [2, 4]. In fact, as shown in Fig. 1, there are three distinct phases in the first-year brain MR images, with each phase having quite different white-gray matter contrast patterns (in chronological order) [5]: (1) the infantile phase (< 5 months), in which GM shows higher signal intensity than WM in the T1 MR images, as shown in Fig. 1(a); (2) the isointense phase (6–8 months), in which the signal intensity of WM is increasing during the development due to the myelination and maturation process; in this phase, GM has the lowest signal difference from WM, resulting in the lowest tissue contrast and hence significant difficulty for tissue segmentation; (3) the early adult-like phase (> 12 months), where GM intensity is much lower than that of WM in the T1 MR images, and this pattern is similar with that of the adult MR images.

Although many methods have been proposed for infant brain image segmentation, most of them focused either on segmentation of the infantile-phase images (< 5 months) or early adult-like images (> 12 months) [3, 4, 6, 7], which demonstrate a relatively good contrast between white matter and gray matter. **There is limited work that could effectively address the difficulties in segmentation of the isointense infant images.** Shi *et al.* [8] first proposed a 4D joint registration and segmentation framework for the segmentation of infant MR images in the first year of life. In this method, longitudinal images in both infantile and early adult-like phases were used to guide the segmentation of images in the isointense phase. The major limitation of these methods is that they fully depend on the availability of longitudinal images. On the other side, Wang *et al.* [9] proposed a sparse patch-based labeling method, which employs a number of isointense atlases with T1, T2 and FA (calculated from diffusion tensor imaging) for guiding the segmentation of isointense infant images. However, requirement of using three modality images (T1, T2 and FA) limits its application for subjects with no longitudinal images. *Moreover, the reliability of manual segmentation for the isointense atlases is also limited due to the extremely low tissue contrast in the isointense phase.*

In this paper, we propose a novel segmentation method to address the above-mentioned difficulties, by using stacked kernel canonical correlation analysis (KCCA) to learn common features for different time-point brain images. Our main motivation is to use the brain images with high tissue contrast (e.g., 12-month brain images) to guide the segmentation of 6-month-old brain images with low tissue contrast. Since the intensity appearances between them are dramatically different, it is not appropriate to directly utilize intensity appearance as features. To this end, we propose to employ KCCA to learn the common feature representations for the 6-month-old and the follow-up 12-month-old images of same subjects in the training stage. Note that the follow-up 12-month-old images are required only in the KCCA based training stage, but not required in the testing stage. Specifically, we consider image patches extracted from those pairs of training images as raw features, and then project them to a common space by KCCA. Also, instead of using the conventional one step of KCCA mapping, which is often not sufficient to capture those highly non-linear correlations between raw features, we iteratively project the previously mapped features and thus form a stacked KCCA mapping to seek an optimal common space for feature representation. After that, we employ a sparse patch-based multi-atlas labeling [9] to segment isointense brain images by propagating tissue labels in the training 12-month-old

images (used as atlases), where patch similarity between testing (6-month-old) and atlas (12-month-old) images is measured by the learned common features. In experiments, we have comprehensively validated our method on 20 isointense brain images, obtaining much better results than the counterpart state-of-the-art methods.

2 Method

Our dataset consists of 20 infant subjects scanned at 6 months and 12 months of age. T1-weighted images were acquired for 144 sagittal slices with resolution of $1 \times 1 \times 1 \text{ mm}^3$ on a Siemens head-only 3T scanners. During the scan, infants were asleep, unsedated, fitted with ear protection, and their heads were secured in a vacuum-fixation device. For image preprocessing, each 6-month-old image was non-rigidly registered to their corresponding 12-month-old image. Afterwards, standard image preprocessing steps were performed before segmentation, including skull stripping, intensity inhomogeneity correction [10], histogram matching, and removal of the cerebellum and brain stem by using in-house tools. To generate ground-truth segmentations of 12-month-old images, we first generated an initial reasonable segmentation by FreeSurfer [11], and then performed manual editing by an experienced rater to correct segmentation errors.

2.1 Stacked Kernel Canonical Correlation Analysis

Given a set of aligned 6-month-old and follow-up 12-month-old images, we randomly selected N pairs of image patches from 6-month-old and corresponding 12-month-old images, which are denoted as $\{x_n\}_{n=1}^N \in \mathbb{R}^D$ and $\{y_n\}_{n=1}^N \in \mathbb{R}^D$, respectively. Here, x_n and y_n are D -dimensional column vectors. For simplicity, let $\mathbf{X} = [x_1, \dots, x_n, \dots, x_N] \in \mathbb{R}^{D \times N}$, $\mathbf{Y} = [y_1, \dots, y_n, \dots, y_N] \in \mathbb{R}^{D \times N}$. The aim of canonical correlation analysis (CCA) is to find a pair of linear projections ($w_x \in \mathbb{R}^D$ and $w_y \in \mathbb{R}^D$) to maximize the correlation between $w_x^T \mathbf{X}$ and $w_y^T \mathbf{Y}$ [12]. However, CCA is limited to linear transformation, which is often not sufficient to capture the highly non-linear relationship between $\mathbf{X}$ and $\mathbf{Y}$. KCCA offers an alternative solution by first projecting the data (i.e., $\mathbf{X}$) into a high dimensional feature space $\mathcal{H}_x$ via a mapping ϕ_x such that data in the feature space become $\Phi_X = [\phi_x(x_1) \cdots \phi_x(x_N)] \in \mathbb{R}^{\mathcal{N}_x \times N}$, where $\mathcal{N}_x$ is the dimension of feature space $\mathcal{H}_x$ that can be very high or even infinite. The mapping ϕ_x from input data to the feature space $\mathcal{H}_x$ is performed implicitly by considering a positive definite kernel function k_x satisfying $k_x(x_1, x_2) = \langle \phi_x(x_1), \phi_x(x_2) \rangle$, where $\langle \cdot, \cdot \rangle$ is an inner product in $\mathcal{H}_x$, rather than by giving the coordinates of ϕ_x explicitly.

Here, we choose k_x as a Gaussian kernel: $k_x(x_1, x_2) = \exp(-\|x_1 - x_2\|^2 / (2\sigma_x^2))$. Similarly, we can project $\mathbf{Y}$ into a feature space $\mathcal{H}_y$ associated with a Gaussian kernel k_y through function ϕ_y such that $\Phi_Y = [\phi_y(y_1) \cdots \phi_y(y_N)] \in \mathbb{R}^{\mathcal{N}_y \times N}$. Thus, after mapping $\mathbf{X}$ to Φ_X and $\mathbf{Y}$ to Φ_Y , we then can apply the ordinary linear CCA to the data pair (Φ_X, Φ_Y) . Let $K_X = \langle \Phi_X, \Phi_X \rangle \in \mathbb{R}^{N \times N}$ and $K_Y = \langle \Phi_Y, \Phi_Y \rangle \in \mathbb{R}^{N \times N}$ be *kernel matrices*. KCCA seeks linear transformation in the Reproducing Kernel Hilbert Space by expressing the projections as linear combinations of the training data, i.e.,

$$w_x = \Phi_X \alpha = \sum_{n=1}^N \alpha_n \phi_x(x_n), w_y = \Phi_Y \beta = \sum_{n=1}^N \beta_n \phi_y(y_n), \text{ where } \alpha, \beta \in \mathbb{R}^N \text{ are called } \textit{dual}$$

vectors. The first pair of *dual vectors* can be determined by solving the following optimization problem:

$$\{\alpha_1, \beta_1\} = \arg \max_{\alpha \in \mathbb{R}^N, \beta \in \mathbb{R}^N} \frac{\alpha^T K_X K_Y \beta}{\sqrt{(\alpha^T K_X K_X \alpha) \cdot (\beta^T K_Y K_Y \beta)}} \quad (1)$$

which is often regularized as the following optimization problem to overcome the risk of over-fitting problem [12, 13],

$$\{\alpha_1, \beta_1\} = \arg \max_{\alpha \in \mathbb{R}^N, \beta \in \mathbb{R}^N} \frac{\alpha^T K_X K_Y \beta}{\sqrt{(\alpha^T K_X K_X \alpha + r_x \alpha^T K_X \alpha) \cdot (\beta^T K_Y K_Y \beta + r_y \beta^T K_Y \beta)}} \quad (2)$$

where r_x and r_y are the positive regularization constants. The subsequent *dual vectors* $\{\alpha_j, \beta_j\}$ are found iteratively by solving (2) with the constraints that $\{\alpha_j, \beta_j\}$ should be orthogonal to all previous *dual vectors*. Specifically, let $A = [\alpha_1 \cdots \alpha_l]$ and $B = [\beta_1 \cdots \beta_l]$ be the sets of *dual vectors* for X and Y , respectively, where $l = \min(\text{rank}(K_X), \text{rank}(K_Y))$. Therefore, the projections of the training data X and new testing X^{test} from their respective original spaces to the common space can be given by $\hat{X} = A^T K_X$ and $\hat{X}^{\text{test}} = A^T K_X^{\text{test}}$, where $[K_X^{\text{test}}]_{mn} = k_x(x_m, x_n^{\text{test}})$. Similarly, the projections of the training data Y is computed as $\hat{Y} = B^T K_Y$.

Algorithm 1

Stacked Kernel CCA Mapping

Input: Training data $X \in \mathbb{R}^{D \times N}$, $Y \in \mathbb{R}^{D \times N}$

Repeat

Construct and center kernel matrices K_X, K_Y

Compute *dual vectors* A, B in Eq. (2)

Set $\hat{X} \leftarrow A^T K_X, \hat{Y} \leftarrow B^T K_Y$

Until reaching the maximum number of iterations

Output: Stacked KCCA projected $\hat{X}, \hat{Y}$

Due to the highly non-linear relationship between native image features X and Y , one step of KCCA is often not sufficient since the error between projections of X and Y , defined as $\|A^T K_X - B^T K_Y\|^2$, may still be large [13]. To reduce this error and seek the optimal common feature space, we can further apply KCCA on the previously projected $\hat{X}$ and $\hat{Y}$. By iteratively finding projecting functions via the KCCA, we can then obtain a sequence of projections for a stacked mapping. The stacked KCCA mapping algorithm is given in Algorithm 1. The advantage of stacked mapping is demonstrated in Fig. 2, where we use the Euclidian distance between projected features as similarity measure. Specifically, Fig. 2(a) and 2(b) show a testing 6-month-old image and an atlas of 12-month-old image. The similarity map is computed by comparing the respective projected features of the reference

point (indicated by red cross) w.r.t. all other points in the atlas. Fig. 2(c) and 2(d) show the respective similarity maps by one step KCCA and the two steps KCCA. It can be clearly seen that the learned common features by two steps KCCA are more distinctive than those by one step KCCA mapping.

2.2 Sparse patch-based labeling

Given a new 6-month-old testing image, we use the 12-month-old training images (of other subjects) with tissue labels as multiple atlases to guide the segmentation. Let C be the total number of atlases. Before labeling, all the atlases are first affine aligned to the testing image space. Based on the learned projections for the common feature representation, we can project each testing image patch and each atlas image patch into the common space. Since the projected features are comparable in the common space, we can perform labeling based on sparse patch-based representation. Specifically, for each point p in the testing 6-month-old image, its patch can be rearranged as a column vector x_p^{test} and its projected features as $\hat{x}_p^{\text{test}} = A^T K_{x_p}^{\text{test}}$, where $K_{x_p}^{\text{test}} \in \mathbb{R}^N$ is a N -dimensional column vector with n -th entry as $k_x(x_n, x_p^{\text{test}})$. Similarly, for each point q in the search neighborhood $\mathcal{N}(p)$ of the c -th aligned atlases, its patch y_q^c is projected as $\hat{y}_q^c = B^T K_{y_q}^c$, where $K_{y_q}^c \in \mathbb{R}^N$ is a N -dimensional column vector with n -th entry as $k_y(y_n, y_q^c)$. Thus, a dictionary can be defined as $D_p \stackrel{\text{def}}{=} \{\hat{y}_q^c, c=1, \dots, C, \forall q \in \mathcal{N}(p)\}$. Then, the label fusion can be formulated as a sparse representation problem:

$$\min_{\theta \geq 0} \frac{1}{2} \|\hat{x}_p^{\text{test}} - D_p \theta\|_2^2 + \lambda_1 \|\theta\|_1 + \frac{\lambda_2}{2} \|\theta\|_2^2 \quad (3)$$

In the above formulation, the first term is the data fitting term, and the second term is the ℓ_1 regularization term for enforcing the sparse constraint on the reconstruction coefficient θ , and the last term is the ℓ_2 smoothness term for enforcing similar coefficients for similar elements in D_p . The parameters λ_1 and λ_2 are the positive constants. Eq. (3) is a convex combination of ℓ_1 lasso [14] and ℓ_2 ridge penalty, which encourages a grouping effect while keeping a similar sparsity of representation [15]. Each element of the sparse coefficient vector θ , i.e., θ_q^c reflects the similarity between the testing patch $\hat{x}_p^{\text{test}}$ and each atlas patch $\hat{y}_q^c$ in the patch dictionary. Based on the assumption that similar patches should share similar tissue labels, we use the estimated sparse coefficients θ to compute the probability of the voxel p belonging to the j -th tissue, $j \in \{WM, GM, CSF\}$, i.e., $P_p(j) = \sum_c \sum_q \theta_q^c$, if $L_q^c = j$, where L_q^c is the tissue label (WM, GM, or CSF) for voxel q in the c -th atlas.

3 Experimental results and analysis

We validate the proposed method on 20 infant subjects via a leave-one-out strategy, where each subject contains 6-month-old and 12-month-old MR T1 images. Specifically, we learn common feature representations based on 19 subjects with aligned 6-month-old and 12-

month-old images via stacked KCCA and then use 19 12-month-old images with tissue labels as multiple atlases to guide the segmentation of 6-month-old image of the remaining testing subject. In our implementation, we randomly selected 10,000 pair of patches with size of $11 \times 11 \times 11$ as the original features from each pair of 6-month-old and corresponding 12-month-old images of 19 training subjects. We set the regularization parameters $r_x=r_y=0.5$, and Gaussian kernel parameters $\sigma_x=\sigma_y=2.5$ for each KCCA mapping. In our study, we found two-step KCCA mapping is sufficient.

We first perform qualitative comparisons with (1) majority voting (MV), (2) conventional patch-based method (CPM) [16], and (3) patch-based labeling based on sparse representation (SR) [9], as shown in Fig. 3(a). Specifically, MV assigns the tissue label by obtaining the most votes (from the aligned atlases) to each voxel, with result shown in Fig. 3(b). CPM employs the Sum of the Squared Difference (SSD) to measure the similarity between the original image patches of the testing image and the atlas images, while SR employs the estimated sparse coefficients as similarities. Since both CPM and SR directly employ image patches as features, these two methods are not able to produce reasonable segmentation results, due to large appearance difference between 6-month-old testing image and 12-month-old atlases. For fair comparison with these two methods, we use 19 6-month-old training images (with tissue labels) as multiple atlases to guide the segmentation in these two methods. Results of these two comparison methods are shown in Fig. 3(c–d). It can be seen that these two comparison methods still cannot achieve satisfactory results due to the extremely low tissue contrast in the 6-month-old testing image. Fig. 3(e) shows segmentation results obtained by the proposed method with one step of KCCA mapping. It can be seen that the estimated probabilities are more accurate than those by all comparison methods. After a stacked KCCA mapping, the correlation between 6-month-old and 12-month-old images are further increased, thus the estimated probabilities are further improved, as shown in Fig. 3(f). Fig. 3(g) shows the ground-truth. To better compare the results of these different methods, the zoomed probabilities are also presented in the bottom row. Fig. 4 shows the WM/GM surfaces obtained by different methods. The zoomed views are also provided in the bottom row for better comparison. It can be observed that the result with stacked KCCA is more consistent with the ground truth.

We then quantitatively evaluate the performance of different methods by employing Dice ratio. The average Dice ratio of each method on segmentation of isointense brain images, tested with 20 leave-one-out validations, is reported in Table 1. Our proposed method with stacked KCCA achieves significantly better results than other methods ($p\text{-value} < 0.05$). Besides using Dice ratio, we also measure the mean surface distance error between the automatically estimated WM/GM (GM/CSF) surfaces and the ground-truth surfaces, as shown in Table 1, further demonstrating the advantage of our method.

4 Discussion and Conclusion

In this paper, we have proposed a novel method for segmentation of isointense infant brain images. Our main motivation is to employ the 12-month-old brain images with high image contrast as multiple atlases to guide the segmentation of 6-month-old brain images with low image contrast. We particularly employ a stacked KCCA mapping to seek the common

feature representations so that the extracted features from 6-month-old and 12-month-old images are comparable. The proposed method was evaluated on 20 6-month-old brain images using leave-one-out cross-validation, showing much better results than the state-of-the-art methods. It is worth noting that the common feature representations learned by stacked KCCA can also be applied to other applications, such as registration of 6-month-old and 12-month-old brain images. In the future, we will extend the stacked KCCA mapping from two time-points, i.e., 6-month-old and 12-month-old, to multiple time-points in the first year of life [12].

References

1. Išgum I, Benders MJNL, Avants B, et al. Evaluation of automatic neonatal brain segmentation algorithms: The NeoBrainS12 challenge. *Med Image Anal.* 2015; 20:135–151. [PubMed: 25487610]
2. Weisenfeld NI, Warfield SK. Automatic segmentation of newborn brain MRI. *Neuroimage.* 2009; 47:564–572. [PubMed: 19409502]
3. Xue H, Srinivasan L, Jiang S, et al. Automatic segmentation and reconstruction of the cortex from neonatal MRI. *Neuroimage.* 2007; 38:461–477. [PubMed: 17888685]
4. Gui L, Lisowski R, Faundez T, et al. Morphology-driven automatic segmentation of MR images of the neonatal brain. *Med Image Anal.* 2012; 16:1565–1579. [PubMed: 22921305]
5. Paus T, Collins DL, Evans AC, et al. Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain Research Bulletin.* 2001; 54:255–266. [PubMed: 11287130]
6. Prastawa M, Gilmore JH, Lin W, et al. Automatic segmentation of MR images of the developing newborn brain. *Med Image Anal.* 2005; 9:457–466. [PubMed: 16019252]
7. Warfield SK, Kaus M, Jolesz FA, et al. Adaptive, template moderated, spatially varying statistical classification. *Med Image Anal.* 2000; 4:43–55. [PubMed: 10972320]
8. Shi F, Yap P-T, Gilmore JH, et al. Spatial-temporal constraint for segmentation of serial infant brain MR images. *MIAR.* 2010
9. Wang L, Shi F, Li G, et al. Integration of Sparse Multi-modality Representation and Geometrical Constraint for Isointense Infant Brain Segmentation. *MICCAI.* 2013:703–710. [PubMed: 24505729]
10. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging.* 1998; 17:87–97. [PubMed: 9617910]
11. Dale AM, Fischl B, Sereno MI. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage.* 1999; 9:179–194. [PubMed: 9931268]
12. Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical Correlation Analysis: An Overview with Application to Learning Methods. *Neuroal Computation.* 2004; 16:2639–2664.
13. Hardoon D, Shawe-Taylor J. Convergence analysis of kernel Canonical Correlation Analysis: theory and practice. *Machine Learning.* 2009; 74:23–38.
14. Tibshirani RJ. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B.* 1996; 58:267–288.
15. Zou H, Hastie T. Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B.* 2005; 67:301–320.
16. Coupé P, Manjón J, Fonov V, et al. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage.* 2011; 54:940–954. [PubMed: 20851199]

Fig. 1. Illustration of three distinct phases in infant brain development, with each phase having quite different WM/GM contrast patterns in the T1 MR images: (a) infantile, (b) isointense, and (c) early adult-like phases. The tissue contrast in the isointense phase (b) is the lowest, resulting in significant difficulty for tissue segmentation.

Fig. 2.

Similarity maps between (a) a point (indicated by red cross) in the testing (6-month-old) image and (b) all points in the atlas (12-month-old) image, by (c) one-step KCCA mapping and (d) two-step KCCA mapping, respectively. Dotted lines are placed in the same location in different images.

Fig. 3. Segmentation results on a typical isointense brain image (a). Results by comparison methods: (b) Majority voting (MV), (c) conventional patch-based method (CPM) [16], and (d) patch-based sparse representation (SR) [9]. Results by proposed methods: (e) with *one step KCCA*, and (f) *stacked KCCA*. (g) shows the ground truth. In each column, we show the estimated probability maps of WM, GM and CSF, along with the zoomed WM probability map, by different method.

Fig. 4.

WM/GM surfaces obtained by (a) majority voting (MV), (b) conventional patch-based method (CPM), (c) patch-based sparse representation (SR), (d) the proposed method with *one step KCCA*, (e) *stacked KCCA*, and (f) ground-truth segmentation.

Segmentation performances of five different methods on isointense brain images, tested by 20 leave-one-out validations.

Table 1

Method		MV	CPM [16]	SR [9]	One KCCA	Stacked KCCA
Dice ratio (%)	WM	73.95±3.41	73.80±2.33	73.65±1.45	80.25±1.22	82.34±0.68
	GM	72.96±2.48	74.47±1.98	75.25±1.38	78.95±1.12	81.57±1.00
	CSF	70.32±0.95	73.94±1.34	75.26±1.01	76.25±0.77	78.80±0.59
Mean surface distance (in mm)	WM/GM	3.06±0.58	2.98±0.66	2.75±0.55	2.33±0.33	1.90±0.34
	GM/CSF	5.84±0.71	5.47±0.88	5.25±0.57	3.12±0.41	2.65±0.29