Hierarchical and symmetric infant image registration by robust longitudinal-example-guided correspondence detection

Yao Wu[a)]
School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China and Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Guorong Wu[a)] and Li Wang[a)]
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Brent C. Munsell[a)]
Department of Computer Science, College of Charleston, Charleston, South Carolina 29424

Qian Wang[a)]
Med-X Research Institute, Shanghai Jiao Tong University, Shanghai 200240, China

Weili Lin[a)]
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Qianjin Feng[b)] and Wufan Chen[a)]
School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China

Dinggang Shen[b)]
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599 and Department of Brain and Cognitive Engineering, Korea University, Seoul 136-701, Republic of Korea

(Received 16 January 2015; revised 24 May 2015; accepted for publication 26 May 2015; published 17 June 2015)

Purpose: To investigate anatomical differences across individual subjects, or longitudinal changes in early brain development, it is important to perform accurate image registration. However, due to fast brain development and dynamic tissue appearance changes, it is very difficult to align infant brain images acquired from birth to 1-yr-old.

Methods: To solve this challenging problem, a novel image registration method is proposed to align two infant brain images, regardless of age at acquisition. The main idea is to utilize the growth trajectories, or spatial-temporal correspondences, learned from a set of longitudinal training images, for guiding the registration of two different time-point images with different image appearances. Specifically, in the training stage, an intrinsic growth trajectory is first estimated for each training subject using the longitudinal images. To register two new infant images with potentially a large age gap, the corresponding images patches between each new image and its respective training images with similar age are identified. Finally, the registration between the two new images can be assisted by the learned growth trajectories from one time point to another time point that have been established in the training stage. To further improve registration accuracy, the proposed method is combined with a hierarchical and symmetric registration framework that can iteratively add new key points in both images to steer the estimation of the deformation between the two infant brain images under registration.

Results: To evaluate image registration accuracy, the proposed method is used to align 24 infant subjects at five different time points (2-week-old, 3-month-old, 6-month-old, 9-month-old, and 12-month-old). Compared to the state-of-the-art methods, the proposed method demonstrated superior registration performance.

Conclusions: The proposed method addresses the difficulties in the infant brain registration and produces better results compared to existing state-of-the-art registration methods. [http://dx.doi.org/10.1118/1.4922393]

Key words: infant brain registration, correspondence detection, hierarchical and symmetric registration

1. INTRODUCTION

In the first year of life, the infant brain undergoes rapid growth development. For instance, the size of the neonatal brain is roughly 380 to 420 cm^3, which is approximately 1/4 to 1/3 the size of an adult brain.[1] More specifically, in the first year of life, the total size of an infant brain increases by more than 100%, in contrast to 15% or less in subsequent years.[2,3] As

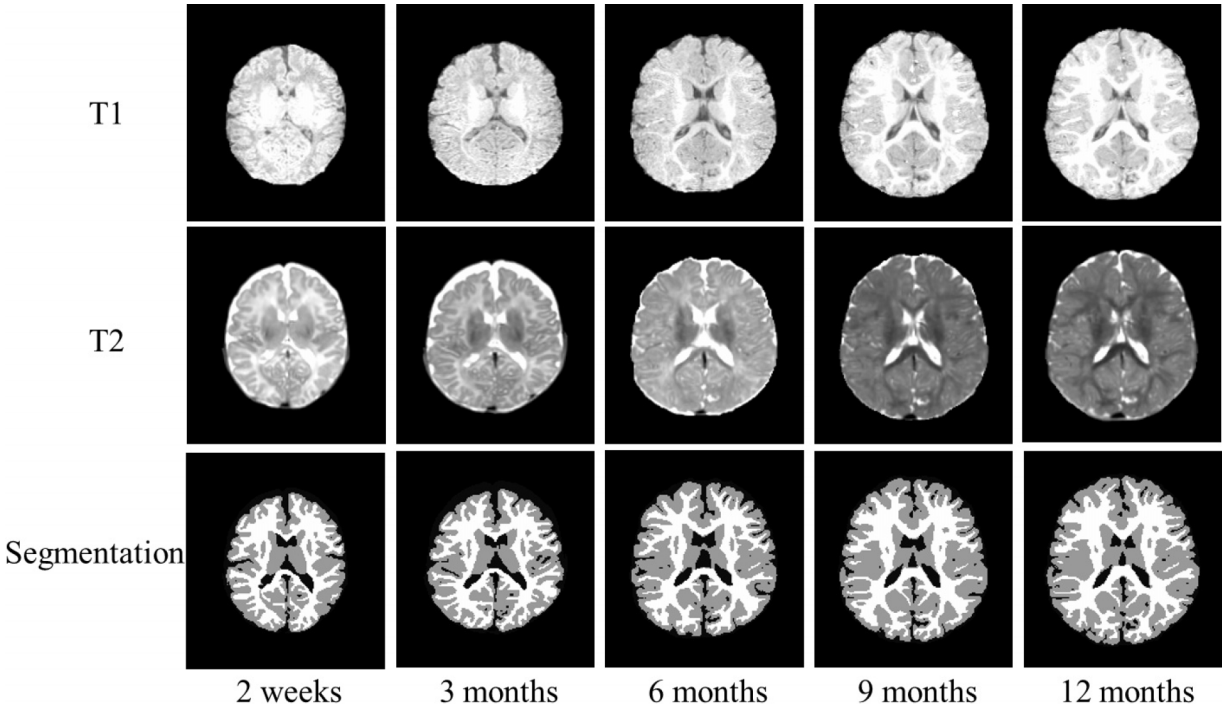

FIG. 1. Dynamic appearance changes and fast brain development in a typical infant brain from 2-week-old to 12-month-old, as shown by T1-weighted MR image (first row) and T2-weighted MR image (second row). The third row shows the automatic segmentation results, which contain three parts, i.e., white matter (white color), gray matter (gray color), and cerebrospinal fluid (black color, except the background).

illustrated in Fig. 1, morphological patterns of key anatomical structures in infant brain images change dramatically in the first year of life,[2] making it very difficult to detect and diagnose developmental disorders, such as autism, or guide disorder interventions.[4]

Magnetic resonance (MR) imaging is widely used in studies because it provides a noninvasive way to discover *morphological patterns* in infant brain images over time. To understand the fast structure changes, the key step is to align individual infant MR brain images to a reference space by using deformable image registration method.[5] In deformable image registration, it is critical to establish accurate anatomical correspondences between two medical images.[6–9] Typically, a patch-based correspondence detection approach is often used, where a patch is a fixed-size symmetric neighborhood of pixel intensity values.[10] If two different patches from two images show similar morphological patterns, the two points (at each patch center) are considered to be well corresponded. Once the imaging data have been aligned, other important applications, such as multiatlas segmentation[11] and longitudinal cortical thickness analysis,[12] may be applied. Unfortunately, the tissue appearance of anatomical structures in infant brain images changes dramatically in the first year of life,[2] as illustrated in Fig. 1, which raises critical issues in measuring the patch-wise similarity in the conventional correspondence detection procedure. Due to the lack of infant registration tool sets, many early brain development studies have to use the registration tool developed only for the adult brains, which may not be well suited for the infant brain images.[13]

In general, deformable image registration tools developed for adults typically do not perform well on infants because of the following reasons: (1) the morphological patterns (*or tissue image appearances*) in the first year of life change very rapidly and may also have unique subject-specific growth patterns;[14,15] (2) due to extremely low tissue contrast[16] in both T1- and T2-weighted images, white matter (WM) and gray matter (GM) tissue regions show very similar intensity levels; and (3) changes in morphological pattern over time are highly nonlinear and can show a considerable amount of spatial variation.[2,15,17,18]

To better understand the complexity of the infant brain image registration problem, examples of longitudinal T1- and T2-weighted images are shown in Fig. 1. Specifically, at 2 weeks of age, the WM tissue appearance (i.e., voxel intensity values) in the T1-weighted MR images are considerably lower than the GM ones, and then, WM and GM tissue appearances become very similar at roughly 6 months. Toward the end of the first year (12 months), the WM and GM tissue appearances in the infant brain start to resemble an adult brain.[19–21] As evident in this example, detecting the corresponded tissue appearances between two infant images with a significant difference in age, i.e. age gap, becomes extremely difficult. Since image registration depends on accurate correspondence, deformable image registration performance can be greatly affected.

Many image registration methods[22–28] are available to align adult brain images that have similar size and appearance. However, few methods can be used for infant brain registration. Xue *et al.*[29] propose a longitudinal registration method for developing neonates; however, this method requires the segmentation of the infant brain images. Csapo *et al.*[8] design a new image similarity metric for longitudinal registration,

which obtains better results than mutual information (MI). However, this method is tested on simulated brains and monkey data, thus its performance needs to be further tested on the real infant brain images. Ha *et al.*[30] use the probabilistic and anatomical structures to match tissue class posteriors and boundaries. However, this method is suitable only for the longitudinal images from the same subject. Serag *et al.*[31] use a 4D population atlas to help the registration of two infant brains. However, a single 4D population atlas is hard to capture the detailed individual development patterns of each subject, which often varies largely across subjects.

To overcome the challenges in infant brain registration, we propose a new deformable image registration method that can automate the alignment of rapidly developing infant brain images within the first year of life. In particular, the novelty of the proposed method is twofold: (1) to overcome dynamic longitudinal changes in the brain tissue appearance, a new patch-based correspondence framework is developed, by using the learned growth trajectory models to guide correspondence detection in the new brain images acquired at different ages; (2) because the learned growth trajectory models may not be able to estimate the complete deformation between two new infant images with very large age differences (e.g., 2-weeks of age and 1-yr of age), our infant image registration method is further combined with a hierarchical and symmetric deformation approach to iteratively reduce the magnitude of the deformation field needed for aligning the two images. Particularly, hierarchical approach can avoid local minima and capture more global deformations, and symmetric approach is unbiased toward the choice of the template (fixed) image and further generates inverse consistent deformations.

Compared to the state-of-the-art deformable image registration methods, our method improves the registration performance across different or same infant subjects, for both small and large age differences. Specifically, through visual inspection and quantitative measurements, the proposed method outperforms several state-of-the-art deformable image registration methods, including the SyN registration method in ANTs package (http://sourceforge.net/projects/advants/) using mutual information[32,33] and cross correlation (CC)[34,35] as similarity measures. We also compare with 3D-HAMMER[36] that uses the segmented images obtained with iBEAT software (http://www.nitrc.org/projects/ibeat/).[37]

The remaining sections of this paper are organized as follows. In Sec. 2, we present the technical details of the proposed infant brain registration method. In Sec. 3, we outline the performed experiments and their results. Finally, in Sec. 4, conclusion is provided.

2. METHOD

Our goal is to register the infant image F_i with the infant image M_j. Their scanning ages, i and j, can be as young as 2-week or as old as 1-yr. Since the intensity contrast between WM and GM changes dynamically in the first year of life, it is difficult to *directly* register two infant images that might belong to different brain development stages. Also, the direct modeling of the complex appearance changes in the infant brains is challenging. Here, we take advantage of the growth trajectories that can be learned from the training subjects with complete longitudinal brain scans (Sec. 2.B), to bridge the appearance gap between F_i and M_j. The entire deformation pathway $\psi_{F\to M}$, for registering M_j and F_i, consists of three segments (Sec. 2.C): (1) the correspondences from F_i to the training images in the i time-domain, (2) the growth trajectories of the training images from i to j time-domains, and (3) the correspondence from the training images in the j time-domain to image M_j. To increase the accuracy of the proposed registration method, we further present a hierarchical and symmetric registration framework in Sec. 2.D. Finally, in Sec. 2.E, a step-by-step procedure is used to summarize the proposed image registration method.

The current paper involves a lot of notations, which are summarized in Table I.

2.A. Preprocessing

To facilitate the subsequent image registration, we preprocess all the images in our dataset. Specifically, the preprocessing includes bias correction[38] and skull-stripping.[39] Note that the brain size increases significantly in the first year of life. Thus, we need to remove such global variations (i.e., translation, rotation, and scaling) via affine registration. Particularly, the affine registration of training images (with complete longitudinal data) consists of two parts. First, within each training subject, we register all other time-point images to the 12-month-old image. Then, across individual subjects, we randomly select the 12-month scan of a certain subject as the common space, to which all other subjects are further registered based on their 12-month scans. Considering the drastic intensity appearance changes over time, mutual information is used as the similarity metric by FSL's linear registration tool (FLIRT) in FSL package (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/).[40]

2.B. Estimation of the growth trajectory

All training subjects have longitudinal data denoted by, $\mathbf{I} = \{I_t^s | s = 1,\ldots,N; t = 1,\ldots T_s\}$, where N denotes the number of training subjects and T_s denotes the number of time points in the image sequence of each training subject. We require the segmentation results of each longitudinal image sequence to be consistent along all time points. To this end, we apply a 4D segmentation method that integrates complementary multimodality information (T1, T2, and FA) and enforces both the cortical thickness and the longitudinal consistency.[11] Particularly, 4D segmentation and 4D-HAMMER registration steps are iteratively applied to refine the 4D segmentation results.[11] Typical segmentation results are shown in the last row of Fig. 1, where WM is denoted by white, GM by gray, and cerebrospinal fluid (CSF) by black.

After obtaining tissue segmentation results for each training subject, we use 4D-HAMMER registration method[41] to establish temporal correspondences in each longitudinal image sequence, since 4D-HAMMER is equipped with

TABLE I. Descriptions of notations used in this paper.

Notation	Description
F_i	A new infant brain image at i time point
F_i^k	Image F_i in the kth iteration
M_j	A new infant brain image at j time point
M_j^k	Image M_j in the kth iteration
$\psi_{F\to M}$	The deformation field from F_i to M_j
I_i^s	A training image at i time point of subject s
Φ	A set of growth trajectories in the training subjects
$\varphi_{i\to j}^s$	The growth trajectory from i time point to j time point of training subject s
$Q(x,i)$	An image patch centered at point x in image F_i
$\vec{b}$	A column vector of patch $Q(x,i)$
$P_s(v,i)$	An image patch centered at point v in the training image I_i^s
$\vec{a}_p$	A column vector of patch $P_s(v,i)$
$n(x)$	A search neighborhood centered at point x in training images at i time point (black box in Fig. 2)
$D(x,i)$	A dictionary of point x containing patches at i time point of all training subjects
A	A matrix containing training samples in dictionary $D(x,i)$
η	The total number of training samples in A
$\vec{w}$	The weight vector of matrix A in representation of vector $\vec{b}$
w_i^p	The pth element in weight vector $\vec{w}$
λ_1	The weight of L_1 term in Eqs. (1) and (2)
λ_2	The weight of L_2 term in Eqs. (1) and (2)
P^*	A set of training correspondences at i time point with weight $w_i^p > 0$
p	A sample in P^* (black dot in Fig. 2)
U	A set of training points at j time point (blue triangles in Fig. 2)
u_v^s	A training point in U which is transferred from point v in the training image I_i^s
$P_s(u_v^s,j)$	An image patch centered at point u_v^s and the j time point of training image I_j^s
$\vec{c}$	A column vector of patch $P_s(u_v^s,j)$
$D'(u_v^s,j)$	A dictionary of point u_v^s which contains training patches in image M_j
A'	A matrix contains training samples in dictionary $D'(u_v^s,j)$
$\vec{w}$	The weight vector of matrix A' in representation of vector $\vec{c}$
$w_j^{p,q}$	The qth element in weight vector $\vec{w}$
$f(x)$	The correspondence of key point x from F_i to M_j
$z^{p,q}$	The coordinate vector of the center of the qth image patch in dictionary D'
ψ_i	The deformation field derived from F_i
ψ_j	The deformation field derived from M_j
ϕ_i^k	The incremental deformation fields obtained by TPS in the kth iteration from F_i
ϕ_j^k	The incremental deformation fields obtained by TPS in the kth iteration from M_j

longitudinal constraint in registration to make the estimated temporal correspondence consistent from one time point to another time point. Given the temporal deformations by 4D-HAMMER, we can form any temporal growth trajectories $\varphi_{i\to j}^s$ from time point i to time point j in each training subject, as shown by the purple dashed curves in Fig. 2. Since the difference of image appearances between the two new infant images (F_i and M_j) could be very large, we propose to establish correspondences between F_i and M_j through the learned growth trajectories $\Phi = \left\{\varphi_{i\to j}^s | s = 1,\ldots,N\right\}$, as detailed below.

2.C. Robust correspondence detection

Since the image points located at distinctive regions can identify the correspondences more accurately than image points in the uniform regions, we determine the correspondences only at key points and let these key points steer the entire deformation estimation.

2.C.1. Key point selection

We use the importance sampling strategy[42] to select key points. Specifically, we first smooth and normalize the gradient magnitude values to be 1 over the two to-be-registered images F_i and M_j, respectively. Then, we use the obtained gradient values to represent the importance (or probability) for each voxel to be selected as the key point. Based on the probability map, a set of key points can be sampled via Monte Carlo simulation.[43] During the hierarchical registration (Sec. 2.D), we relax the selection criterion, which is defined based on the probability map, and add more key points to both images in each iteration. Most of the key points locate at distinctive regions, such as the boundaries of WM, GM, and ventricle. There are still some key points located at the uniform regions to cover the whole brain and guide the deformation of the entire brain. In general, the key points distribute more in the gradient-rich regions but less in the uniform regions. Although more advanced method could be used here for key point selection, we use a simple gradient-guided strategy since

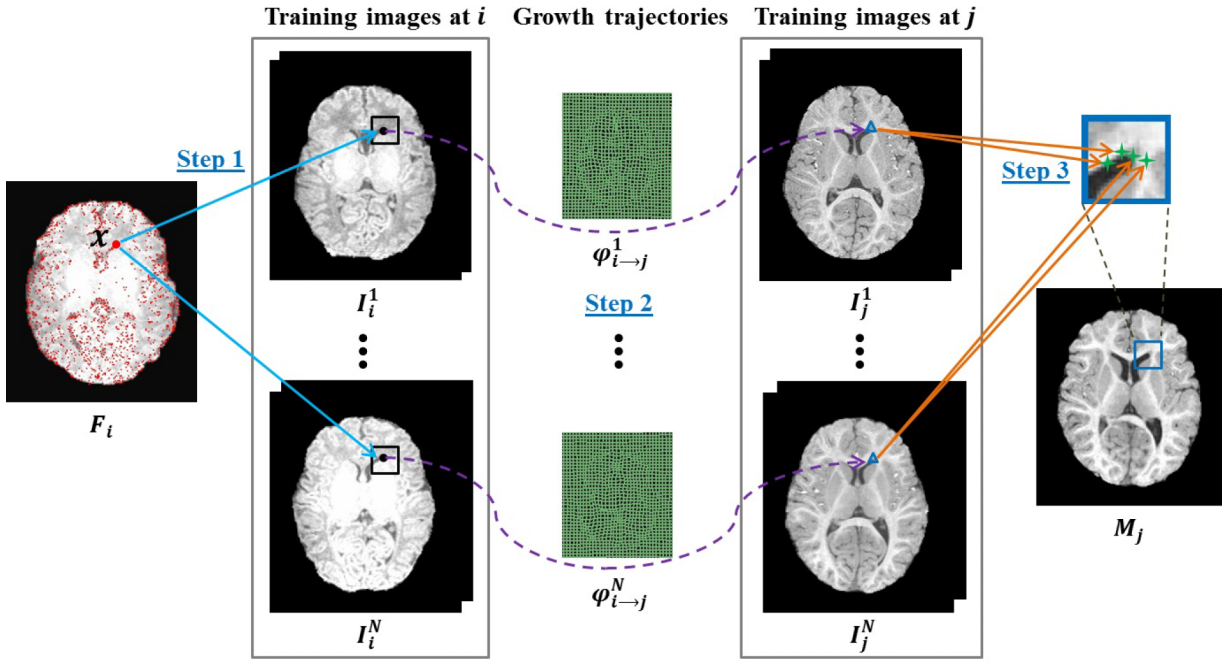

FIG. 2. Overview of our correspondence detection procedure for a particular key point x in image F_i to the candidate points in image M_j, with helps from the images with similar ages in the training set (middle). Arrows denote the possible correspondences, and the two deformed meshes (middle) represent the temporal growth trajectories $\varphi_{i\to j}^s$ from time point i to time point j in each training subject (top and bottom).

it is computationally fast. The typical key points selected are colored in red and overlaid on image F_i in Fig. 2 (left).

As we will explain in Sec. 2.D, we simultaneously detect correspondences on both F_i and M_j in the sense of symmetric deformation. However, for clear illustration, we only introduce the robust correspondence detection (in three steps) associated with the key point x in image F_i, as detailed below.

Step 1: Detect correspondence with respect to the training images of similar scanning age. First, we extract the reference patch $Q(x,i) \subset F_i$ centered at the key point x. Next, we collect a set of patches $\{P_s(v,i)|P_s(v,i) \subset I_i^s\}$ across all training images $\{I_i^s\}$ at the same time point i. The center point v of $P_s(v,i)$ is located within the search neighborhood $n(x)$ centered at x (i.e., black box in Fig. 2), to generate an overcomplete dictionary $D(x,i) = \{P_s(v,i)|P_s(v,i) \subset I_i^s, s = 1,\ldots,N, v \in n(x)\}$. Since F_i and I_i^s are acquired at the same time point i, the appearances of their local patches are often very similar. Thus, it is relatively easy to represent the reference patch $Q(x,i)$ with patches in the dictionary $D(x,i)$, acquired from the training images with similar scanning age.

For clarity, we vectorize the reference patch $Q(x,i)$ into a column vector $\vec{b} \in R^m$, where m denotes the total number of points in the image patch. Also, we arrange each patch $P_s(v,i)$ into a column vector $\vec{a}_p \in R^m$ and further assemble them into the matrix $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_\eta] \in R^{m\times n}$, where $\eta = N \cdot |n(x)|$ denotes the total number of patches in D. Note that matrix A is used to mathematically represent dictionary D. Inspired by the power of sparse representation,[44,45] we seek for a sparse coefficient vector $\vec{w}_i \in R^\eta$ to represent the target patch $\vec{b}$ based on the matrix/dictionary A, i.e., $\vec{b} \leftarrow A\vec{w}_i$. Each element of $\vec{w}_i$ indicates the contribution of a particular patch $\vec{a}_p$ in representing the target patch $\vec{b}$. The estimation of $\vec{w}_i$ falls into the elastic net problem,[46]

$$\vec{w}_i = \arg\min_{\vec{w}_i} \left\|\vec{b} - A\vec{w}_i\right\|_2^2 + \lambda_1\left\|\vec{w}_i\right\|_1 + \lambda_2\left\|\vec{w}_i\right\|_2^2, \quad \text{s.t.}\, \vec{w}_i > 0. \tag{1}$$

The first term is the data fitting term. λ_1 and λ_2 control the strength of sparsity (the second term) and ridge regularization (the third term), respectively. Specifically, the second term is an L_1 LASSO (least absolute shrinkage and selection operator) constraint,[47] which controls the sparsity of the dictionary coefficient vector $\vec{w}_i$; the last term is a L_2 ridge regression, which ensures the similarity of coefficients for similar patches.[48] The weighting vector $\vec{w}_i$ reflects the representation of patch $\vec{b}$ by using the image patches in the dictionary A.[49] Intuitively, larger element w_i^p in the coefficient vector $\vec{w}_i$ implies higher confidence in identifying the center point of the image patch $\vec{a}_p$ as the correspondence of key point x (see blue arrows in Fig. 2).

Step 2: Identify correspondences from one time-domain to another time-domain. After sparse patch matching for the reference point x as described above (Step 1), we can obtain a small set of possible correspondences $P^* = \{p = (v,s)|w_i^p > 0\}$, which excludes all $\vec{a}_p$'s with $w_i^p = 0$. Black dots in Fig. 2 denote the established correspondences P^* in the i time-domain. The correspondences in P^* might be located at different training images. Then, we can propagate each obtained correspondence $p \in P^*$ from the i time-domain (belonging to F_i) to the j time-domain (belonging to M_j) by following the learned growth trajectory $\varphi_{i\to j}^s$ (the purple dashed curves in Fig. 2). Thus, we can identify the correspondences of the key point x in the j time-domain, denoted by $U = \{u_v^s = \varphi_{i\to j}^s(v)|(v,s) \in P^*\}$ (see blue triangles in Fig. 2).

Step 3: Establish the final correspondence. The goal of this step is to continue the detection of the correspondence for each u_v^s (Step 2) toward image M_j. Since each u_v^s belongs to the same time-domain as M_j, we can follow the similar procedure in Step 1 to determine the correspondence. Specifically, image patch $P_s(u_v^s,j)$ now becomes the reference patch here, which is centered at u_v^s and takes the image content from the training image I_j^s at j time-domain. The dictionary $D'(u_v^s,j)=\{P(o,j)|P(o,j)\subset M_j, o\in n(u_v^s)\}$ consists of the patches $\{P(o,j)\}$ in image M_j, where the center point o of patch $P(o,j)$ is within the search neighborhood $n(u_v^s)$ (as designated by the blue box in Fig. 2). It is worth noting that the dictionary size is only $\theta=|n(u_v^s)|$ since we confine the search to be in M_j only. To find the correspondence for each u_v^s toward image M_j, we vectorize the reference patch $P_s(u_v^s,j)$ into a column vector $\vec{c}\in R^m$. We further arrange each patch in D' into a column vector $\vec{a}_{p,q}\in R^m$ and assemble them into the matrix $A'=[\vec{a}_{p,1},\vec{a}_{p,2},\ldots,\vec{a}_{p,\theta}]\in R^{m\times\theta}$, where $\theta=|n(u_v^s)|$ denotes the total number of patches in D'. Here, we solve another elastic net problem which is similar with Eq. (1),

$$\vec{w}_j^p=\arg\min_{\vec{w}_j^p}\left\|\vec{c}-A'\vec{w}_j^p\right\|_2^2+\lambda_1\left\|\vec{w}_j^p\right\|_1+\lambda_2\left\|\vec{w}_j^p\right\|_2^2,$$
$$\text{s.t.}\,\vec{w}_j^p>0. \tag{2}$$

After solving Eq. (2), we can obtain a sparse coefficient vector $\vec{w}_j^p\in R^\theta$, where element $w_j^{p,q}$ is the weight for the sample $\vec{a}_{p,q}$ in A'. After we sequentially perform Steps 1–3, we can calculate the final correspondence [i.e., $f(x)$] of the key point x from F_i to M_j by using a soft correspondence strategy,[50]

$$f(x)=\frac{\sum_{p\in P^*}w_i^p\left(\sum_{q=1}^{\theta}w_j^{p,q}z^{p,q}\right)}{\sum_{p\in P^*}w_i^p\left(\sum_{q=1}^{\theta}w_j^{p,q}\right)}, \tag{3}$$

where $z^{p,q}$ represents the coordinate vector of the center of the qth image patch in dictionary D'. In Fig. 2, $z^{p,q}$ is displayed by green cross.

2.C.2. Dense deformation interpolation

Given the correspondences determined on all key points, we can use thin-plate splines (TPSs)[51] to interpolate the dense deformation field by considering the key points as control points.

Recall that we also detect the key points in image M_j. Thus, it is straightforward to apply the same correspondence detection procedure to determine correspondences for the key points y in image M_j, which leads to the estimation of the symmetric deformation (along with the correspondences already estimated from F_i to M_j) as described in Sec. 2.D.

2.D. Hierarchical and symmetric deformation estimation

2.D.1. Hierarchical deformation

In the beginning of image registration, we only select a small number of key points in the infant images to steer the entire deformation estimation. That is, we only detect correspondence for each selected key point by the correspondence detection procedure in Sec. 2.C. Other nonkey points just follow the deformations of the nearby key points. In this way, our registration method can alleviate the ambiguity in correspondence detection by letting the key points with reliable correspondences to steer the entire image deformation. Moreover, since only a small number of key points are selected, our registration can be very efficient. After two images have been approximately aligned, we can gradually increase the number of key points by relaxing the selection criterion in importance sampling (as we can see for the gradually increased number of red (blue) points in image F_i^k (M_j^k) in Fig. 3, where

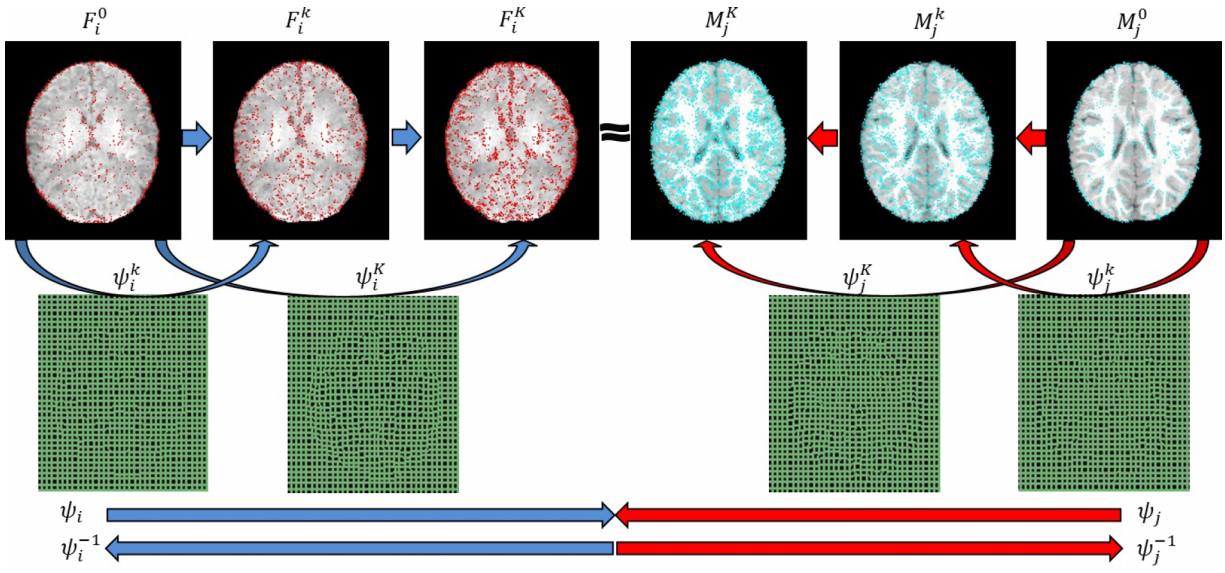

FIG. 3. The iterative symmetric registration framework. Both images gradually deform from their own domains, until they meet with each other.

superscript k denotes the number of iteration). Meanwhile, we can gradually reduce the search neighborhood in correspondence detection, since only local refinement is needed in the latter stages as registration progresses. In the beginning of iteration, the number of key points is set to approximately 20 000, and then, this number is gradually increased to approximately 60 000 in the end of registration.

2.D.2. Symmetric registration

Instead of directly estimating the deformation pathway from F_i to M_j, symmetric deformation estimation can significantly improve registration accuracy. Specifically, we can simultaneously deform F_i and M_j toward each other until the deformed F_i and deformed M_j become similar in the common space. As shown in Fig. 3, we can obtain two deformation pathways, i.e., (1) ψ_i derived from F_i to the common space and (2) ψ_j derived from M_j to the common space, as denoted by the blue and red arrows in Fig. 3, respectively. Deformation pathways ψ_i and ψ_j are calculated by composing the deformation fields in the previous iteration and the incremental deformation fields obtained in the current iteration (as detailed in the following paragraph). After obtaining ψ_i and ψ_j, we use the following two steps to calculate the deformation field from F_i to M_j: (1) inverse deformation ψ_j to $(\psi_j)^{-1}$; (2) compose deformation ψ_i and the inversed deformation $(\psi_j)^{-1}$, i.e., $\psi_{F\to M}=\psi_i\circ(\psi_j)^{-1}$, where "$\circ$" denotes the deformation composition[52] and $\psi_{F\to M}$ denotes the deformation field from F_i to M_j.

To make the deformation field invertible, we adapt the registration procedure to a space of diffeomorphic transformation by following the approach in diffeomorphic demons.[53] Since the deformation pathways are iteratively refined, we use k $(k=0,\ldots,K)$ to denote the round of iteration. In the beginning of registration $(k=0)$, $F_i^0=F_i$ and $M_j^0=M_j$, along with their respective identity deformation pathways ψ_i^0 and ψ_j^0. During the registration $(k>0)$, the incremental deformation fields obtained by TPS in the kth iteration (i.e., ϕ_i^k and ϕ_j^k) are mapped to the space of diffeomorphism through the exponentials, i.e., $\exp(\phi_i^k)$ and $\exp(\phi_j^k)$.[53] The deformation pathways from both ends in the kth iteration can be obtained by $\psi_i^k=\psi_i^{k-1}\circ\exp(\phi_i^k)$ and $\psi_j^k=\psi_j^{k-1}\circ\exp(\phi_j^k)$, respectively. Based on these two deformation pathways (i.e., ψ_i^k and ψ_j^k), F_i gradually deforms to $F_i^k=F_i(\psi_i^k)$ and M_j deforms to $M_j^k=M_j(\psi_j^k)$. Meanwhile, more and more key points are selected to refine the deformation pathways ψ_i^k and ψ_j^k. The iteration will stop until the deformed F_i^K and the deformed M_j^K become very similar in the end of registration.

2.E. Summary

Given the growth trajectories Φ learned from the training subjects, the registration of image F_i at the time point i and image M_j at the time point j is summarized as follows:

1. Affine register both M_j and F_i to the common space where all training images were previously aligned.
2. Set $k=0$.
3. Detect key points in F_i^k and M_j^k, respectively.
4. For each key point x in F_i, compute the correspondence $f(x)$ from F_i^k to M_j^k, according to the correspondence detection procedure in Sec. 2.C.
5. For each key point in M_j, compute the correspondence from M_j^k to F_i^k, according to the correspondence detection procedure in Sec. 2.C.
6. Use TPS to interpolate the incremental deformation fields ϕ_i^{k+1} and ϕ_j^{k+1}, respectively.
7. Use the scaling and squaring method[53] to calculate the exponentials, i.e., $\exp(\phi_i^{k+1})$ and $\exp(\phi_j^{k+1})$.
8. Calculate the deformation pathways from both ends by $\psi_i^{k+1}=\psi_i^k\circ\exp(\phi_i^{k+1})$ and $\psi_j^{k+1}=\psi_j^k\circ\exp(\phi_j^{k+1})$, respectively.
9. Calculate the inverse deformation pathways from both ends by $(\psi_i^{k+1})^{-1}=\exp(-\phi_i^{k+1})\circ(\psi_i^k)^{-1}$ and $(\psi_j^{k+1})^{-1}=\exp(-\phi_j^{k+1})\circ(\psi_j^k)^{-1}$, respectively.
10. Deform F_i to F_i^{k+1} based on ψ_i^{k+1} and also deform M_j to M_j^{k+1} based on ψ_j^{k+1}.
11. $k\leftarrow k+1$.
12. If $k<K$, go to (3) and relax the key point selection criterion to get more key points.
13. Otherwise, the final deformation pathway $\psi_{F\to M}$ is computed as $\psi_{F\to M}=\psi_i^K\circ(\psi_j^K)^{-1}$.

3. EXPERIMENTAL RESULTS

In this section, we applied our novel infant registration method to 24 infant subjects. Each subject has T1- and T2-weighted MR images at 2 weeks, 3, 6, 9, and 12 months. The T1-weighted images were acquired on a Siemens head-only 3T MR scanner and have 144 sagittal slices with resolution $1\times1\times1$ mm^3. The T2-weighted images have 64 axial slices with resolution $1.25\times1.25\times1.95$ mm^3. For each subject, the T2 image is linearly aligned to the T1 image of the same time point by FLIRT in FSL package (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/)[40] and then further isotropically resampled to $1\times1\times1$ mm^3.

We use two sets of segmentation data to evaluate the performance of all comparison methods. (1) We use the manually edited results of automated tissue segmentations[11] (on WM and GM) by a clinical expert as ground truth to evaluate the registration accuracy, by measuring the Dice ratio of tissue overlap between the aligned images. During manual correction, we follow a protocol (http://neobrains12.isi.uu.nl/example.php) for manual editing of automated tissue segmentations. Note here that we use T2 images to segment 2-week and 3-month-old infant images while we use T1 images to segment 6-, 9-, and 12-month-old infant images, considering the strong tissue contrast of MR images in the respective time points.[11] (2) We have also included the pure manual segmentations of hippocampus from the longitudinal images of ten subjects in our dataset and then use them as ground truth. In addition, normalized mutual information and cross correlation of the registered images are calculated to further evaluate the registration performance.

We evaluate intra-/intersubject registration performance in the leave-one/two-subject(s)-out manner. In each leave-one-subject-out cross validation, we apply our registration method

to register any two images in the same longitudinal image sequence, while the other 23 are considered to be training images. Furthermore, in order to evaluate the registration performance of infant images from different subjects and at different time points, we leave all images of two subjects out of the training image set and then regard the remaining 22 infant subjects as the training subjects.

In both *intrasubject* longitudinal and *intersubject* cross-sectional image registration scenarios as described above, we show the registration results of aligning 2-week, 3-, 6-, and 9-month images to 12-month images. Considering the tissue contrast, we use T2 images at 2 weeks and 3 months, while T1 images at 6, 9, and 12 months. Since the growth trajectory model is built upon the presegmented images and also the correspondence detection procedure between two new images under registration is always performed at the same age, using different modality MR images at different ages does not affect our infant registration method. Actually, this is also the advantage of the proposed method.

For comparison, we use the state-of-the-art 3D-HAMMER registration method (based on the segmented images) and SyN in ANTs package (based on the original intensity images) as the counterpart registration methods. Specifically, for any two testing images to be registered with HAMMER, we use 3D tissue segmentation method in iBEAT software[37] to segment them, instead of using longitudinal data for assisting the segmentation since we assume no complete longitudinal images are available for the testing samples. Furthermore, we use mutual information and cross correlation as the two respective image similarity metrics for SyN based registration, thus resulting in two intensity-based registration versions, termed as MI-method and CC-method, respectively. In our experiment, we performed an intrasubject registration to determine the parameters, where the testing subject was randomly chosen from the 24 subjects and the remaining 23 subjects were used as the training dataset. Specifically, we set the patch size to {3, 5, 7, 9, 11}, the size of search neighborhood to {3, 5, 7, 9, 11}, the weight of L_1 term λ_1 [Eqs. (1) and (2)] to {0.01, 0.05, 0.1, 0.15, 0.2}, and the weight of L_2 term λ_2 to {0.001, 0.005, 0.01, 0.015, 0.02}. We calculated the combined WM and GM Dice ratios with respect to different combinations of these values to set the parameters in our study. Finally, the patch size was set to $5\times5\times5$ mm^3, while the size of search neighborhood was initially $7\times7\times7$ mm^3 and then gradually reduced to $3\times3\times3$ mm^3. In solving the elastic net problem in Eqs. (1) and (2), we set $\lambda_1 = 0.2$ and $\lambda_2 = 0.01$. It is worth noting that we fixed all these parameters in the following experiments.

In Subsections 3.A–3.E, we first evaluate the advantage of using elastic net, against the use of only the sparsity constraint, during correspondence detection in Sec. 3.A. Next, we examine the registration performance gained from hierarchical and symmetric registration framework in Sec. 3.B. From Secs. 3.C and 3.D, we demonstrate the overall performance for *intrasubject* longitudinal image registration and the *intersubject* image registration, respectively, by also comparing with 3D-HAMMER, MI-method, and CC-method. We finally present the discussions in Sec. 3.E.

3.A. Advantage of using elastic net in correspondence detection

Recall that we regard the correspondence detection as sparse representation with elastic net. In order to show its advantage, we compare its registration performance with that obtained by using a typical LASSO-guided correspondence detection procedure. Figure 4 shows the combined WM and GM Dice ratios in each of 24 leave-one-subject-out cases (by registering 2-week-old image to 12-month-old image) using elastic net (red curve) and using only L_1 (LASSO) constraint (blue curve), respectively. The results shown in Fig. 4 suggest that the proposed method outperforms the counterpart method using only L_1 (LASSO) constraint. Specifically, Dice ratio of WM is $77.2\% \pm 1.5\%$ by using our proposed (elastic-net) method and $75.6\% \pm 1.8\%$ by the counterpart method using only L_1 (LASSO) constraint, thus achieving 1.6% improvement by our method (paired t-test; $p < 0.001$). Similarly, we also achieve 1.7% improvement for the Dice ratio of GM by our method ($77.6\% \pm 1.4\%$) over the counterpart method ($75.9\% \pm 1.7\%$) (paired t-test; $p < 0.001$).

3.B. Advantages of hierarchical and symmetric registration framework

In order to evaluate the impact of hierarchical deformation mechanism and symmetric registration component one by one, we compare our full registration method (Method 4) with three slightly different versions: (1) our method without these two components (Method 1), (2) our method with hierarchical deformation mechanism only (Method 2), and (3) our method with symmetric registration only (Method 3). Like Sec. 3.A, we evaluate the registration performance in 24 *intrasubject* leave-one-subject-out registration cases. Note that besides registering 2-week-old image to the 12-month-old image of same subject as described above, we also register 3-month-old, 6-month-old, and 9-month-old images to their own 12-month-old image. The mean and standard deviation of combined WM and GM Dice ratios are displayed in Table II,

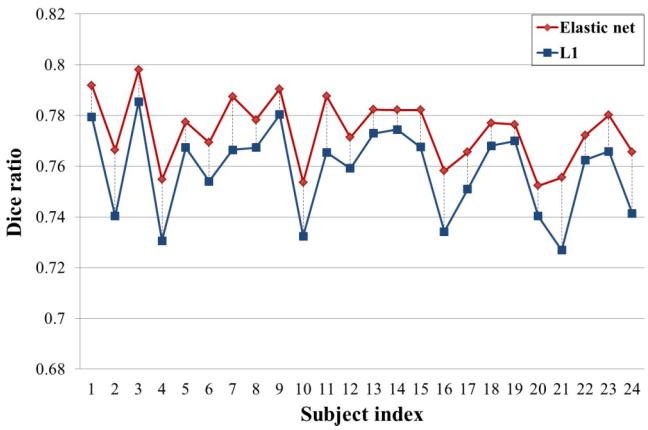

FIG. 4. The combined WM and GM Dice ratios obtained by registering 2-week-old image to the 12-month-old image of the same subject in 24 leave-one-subject-out cases. The registration performances with the correspondence detection by using elastic net and using only L_1 (LASSO) constraint are displayed by red and blue curves, respectively.

TABLE II. The mean and standard deviation of combined WM and GM Dice ratios by four registration methods. Method 1 is our method without hierarchical deformation mechanism and symmetric registration components; Method 2 is our method with hierarchical deformation mechanism only; Method 3 is our method with symmetric registration only; Method 4 is our full registration method. The best result for each column is shown in bold.

	2-week to 12-month	3-month to 12-month	6- month to 12-month	9-month to 12-month
Method 1	0.755 ± 0.020[a]	0.756 ± 0.025[a]	0.775 ± 0.023[a]	0.819 ± 0.022[a]
Method 2	0.763 ± 0.018[a]	0.764 ± 0.022[a]	0.788 ± 0.019[a]	0.831 ± 0.020[a]
Method 3	0.768 ± 0.017[c]	0.771 ± 0.019[b]	0.798 ± 0.020[a]	0.842 ± 0.018[a]
Method 4	**0.774** ± 0.015	**0.778** ± 0.018	**0.818** ± 0.016	**0.858** ± 0.015

[a] $p < 0.001$.
[b] $p < 0.01$.
[c] $p < 0.05$ for paired t-test.

where each row in the table denotes for one of the four above registration methods. We conducted paired t-test between each of Methods 1–3 and our full registration method (Method 4) to evaluate the performances of hierarchical deformation mechanism and symmetric registration. The significance levels are shown in Table II, where a denotes $p < 0.001$, b denotes $p < 0.05$, and c denotes $p < 0.01$.

Our dataset contains the manual segmentations of hippocampus in a set of longitudinal images from ten subjects. For these ten subjects, we deform the hippocampus segmentation images from the 2-week-old, 3-month-old, 6-month-old, and 9-month-old to the 12-month-old domain of the same subject based on the deformation fields obtained by Methods 1–4. We calculate the Dice ratio of the hippocampus between the two registered images to further evaluate the performances of hierarchical deformation mechanism and symmetric registration. The mean and standard deviation of hippocampus Dice ratios obtained by different methods are shown in Table III. We also conducted paired t-test between each of Methods 1–3 and our full registration method (Method 4), and the significance levels are shown in Table III. We can see that both hierarchical deformation mechanism and symmetric registration contribute to improve the performance of infant image registration, and our full registration method combining these two components obtains the best results.

3.C. Evaluation on *intrasubject* image registration

In this experiment, we evaluate the performance of the proposed infant registration method by registering two new infant images from the same subject, but scanned at two different time points. Here, we register 2-week-old, 3-month-old, 6-month-old, and 9-month-old images to the 12-month-old image. For visual inspection, Fig. 5 shows registration results by five different methods on a typical infant subject. In Fig. 5, the 12-month-old image is shown in Fig. 5(a), and the 2-week-old, 3-month-old, 6-month-old, and 9-month-old images are shown in the first row from columns (b) to (e), respectively. From the second row to the last row, we show the registration results by (1) affine registration, (2) MI-method, (3) CC-method, (4) 3D-HAMMER, and (5) our infant registration method. As indicated by the arrows and dashed boxes in Fig. 5, our method achieves better registration results, in terms of the structure similarity with those in the 12-month-old image. Next, we provide the quantitative measurements to show the improved registration performance by our method.

3.C.1. Tissue overlap ratio

Here, we use tissue overlap ratio to evaluate the accuracy of our registration method. Given the estimated deformation fields, we can deform the tissue maps from the 2-week-old, 3-month-old, 6-month-old, and 9-month-old to the 12-month-old domain, respectively. After that, we can quantitatively measure the registration accuracy by computing the Dice ratio of each brain tissue between the two registered images. Table IV shows the mean and standard deviation of combined WM and GM Dice ratios in registering different time-point images to their own 12-month-old images by using FLIRT, MI-method, CC-method, 3D-HAMMER, and our

TABLE III. The mean and standard deviation of hippocampus Dice ratios on ten subjects by registering images at different time points of the same subject. Method 1 is our method without hierarchical deformation mechanism and symmetric registration components; Method 2 is our method with hierarchical deformation mechanism only; Method 3 is our method with symmetric registration only; Method 4 is our full registration method. The best result for each column is shown in bold.

	2-week to 12-month	3-month to 12-month	6-month to 12-month	9-month to 12-month
Method 1	0.593 ± 0.069[a]	0.615 ± 0.063[a]	0.673 ± 0.062[a]	0.735 ± 0.057[a]
Method 2	0.602 ± 0.067[a]	0.623 ± 0.057[a]	0.685 ± 0.057[a]	0.746 ± 0.053[a]
Method 3	0.605 ± 0.064[b]	0.627 ± 0.059[b]	0.687 ± 0.060[a]	0.751 ± 0.051[b]
Method 4	**0.614** ± 0.063	**0.635** ± 0.055	**0.702** ± 0.059	**0.762** ± 0.049

[a] $p < 0.001$.
[b] $p < 0.01$ for paired t-test.

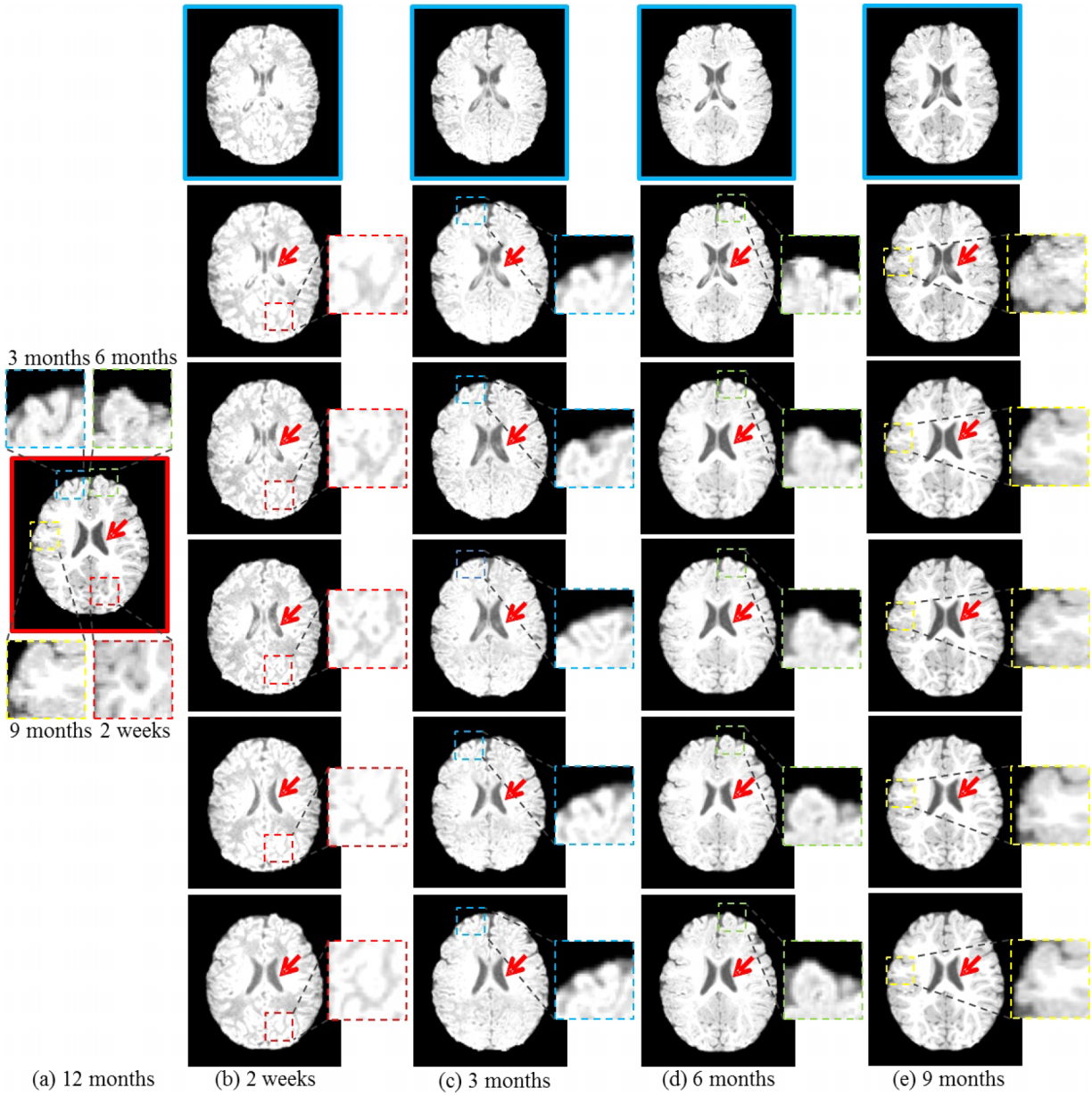

FIG. 5. Intrasubject registration results on images at different time points. Column (a) shows the 12-month-old image. In the first row, from columns (b) to (e) are the 2-week-old, 3-month-old, 6-month-old, and 9-month-old images, respectively. Registration results by affine registration, MI-method, CC-method, 3D-HAMMER, and our infant registration method are shown in the 2–6 rows, respectively. Registration improvement by our method can be viewed by arrows and dashed boxes.

method, respectively. Particularly, the combined WM and GM Dice ratios in registering 2-week-old image to 12-month-old image are 62.8% ± 3.8% by FLIRT, 70.9% ± 2.8% by MI-method, 71.4% ± 2.5% by CC-method, 76.4% ± 2.7% by 3D-HAMMER, and 77.4% ± 1.5% by our method, where our method achieves around 6.5%, 6.0%, and 1.0% improvements than MI-method, CC-method, and 3D-HAMMER, respectively. For the registration results from 3- and 6-month-old images to the 12-month-old image, our method achieves 8.4% and 1.9% improvements than MI-method, 9.1% and 2.1% improvements than CC-method, and 2.2% and 1.2% improvements than 3D-HAMMER. These improvements are statistically significant (paired t-test), and the significance levels are shown in Table IV. For the registration from 9-month-old image to the 12-month-old image, there is no statistical significant difference between the registration results of MI-method, CC-method, 3D-HAMMER, and our method (paired t-test; $p > 0.05$).

3.C.2. Hippocampus overlap ratio

We also provide manual segmentations of hippocampus of ten subjects. For these ten subjects, we deform the hippocampus segmentations from the 2-week-old, 3-month-old, 6-month-old, and 9-month-old to the 12-month-old domain of the same subject based on the deformation fields obtained by FLIRT, MI-method, CC-method, 3D-HAMMER, and the proposed method, respectively. We further compute the Dice

TABLE IV. Intrasubject registration results. The mean and standard deviation of combined WM and GM Dice ratios, obtained for registering 2-week-old, 3-month-old, 6-month-old, and 9-month-old images to the 12-month-old image by five different methods over 24 leave-one-subject-out registration cases. The best result for each column is shown in bold.

	2-week to 12-month	3-month to 12-month	6-month to 12-month	9-month to 12-month
FLIRT	0.628 ± 0.038[a]	0.622 ± 0.035[a]	0.753 ± 0.026[a]	0.790 ± 0.028[a]
MI-method	0.709 ± 0.028[a]	0.694 ± 0.030[a]	0.799 ± 0.029[a]	0.850 ± 0.018
CC-method	0.714 ± 0.025[a]	0.687 ± 0.031[a]	0.797 ± 0.028[a]	0.851 ± 0.019
3D-HAMMER	0.764 ± 0.027[c]	0.756 ± 0.028[a]	0.806 ± 0.024[b]	0.848 ± 0.021
Our method	**0.774** ± 0.015	**0.778** ± 0.018	**0.818** ± 0.016	**0.858** ± 0.015

[a] $p < 0.001$.
[b] $p < 0.01$.
[c] $p < 0.05$ for paired t-test.

ratio of the hippocampus between the two registered images to evaluate the accuracy of the proposed method. The mean and standard deviation of Dice ratios for all compared methods are displayed in Table V. For the registration results from 2-week-old, 3-month-old, 6-month-old, and 9-month-old images to 12-month-old image, our method achieves 3.3%, 4.2%, 2.6%, and 1.1% improvements than MI-method, 3.5%, 5.0%, 2.1%, and 1.4% improvements than CC-method, and 1.3%, 2.2%, 1.9%, and 2.6% improvements than 3D-HAMMER. All improvements are statistically significant (paired t-test), and the significance levels are shown in Table V.

3.C.3. Normalized mutual information of registered images

To further evaluate the performances of all comparison methods in 24 intrasubject leave-one-subject-out registration cases, we calculate the normalized mutual information of the registered images. Figure 6 shows the whisker plots of the normalized mutual information of registered images between different time points by using five registration methods. In Fig. 6, we use black, blue, green, cyan, and red to denote the plots for the results by FLIRT, MI-method, CC-method, 3D-HAMMER, and the proposed method, respectively. Each box in Fig. 6 denotes the distribution of the normalized mutual information of 24 leave-one-subject-out cases, with the top, middle, and bottom lines corresponding to the 25th percentile, median value, and 75th percentile, respectively. Here, we use an asterisk above the box to denote the statistical significant difference between the results produced by the compared method and our method (paired t-test; $p < 0.05$).

3.C.4. Cross correlation of registered images

This section shows the cross correlation of the registered images in 24 leave-one-subject-out registration cases. Figure 7 shows the whisker plots of the cross correlation of registered images between different time points by using FLIRT (black), MI-method (blue), CC-method (green), 3D-HAMMER (cyan), and the proposed method (red), where each box denotes the distribution of the cross correlation of registered images in 24 leave-one-subject-out cases. For the registration results from the 2-week-old image to the 12-month-old image, our method achieves significant improvements compared to FLIRT, MI-method, and CC-method (paired t-test; $p < 0.05$). For the registration from the 3- and 6-month-old images to the 12-month-old image, our method achieves significant improvements compared to other four methods (paired t-test; $p < 0.05$). For the registration from the 9-month-old image to the 12-month-old image, there is no statistical significant difference between the registration results of four deformable registration methods (i.e., MI-method, CC-method, 3D-HAMMER, and the proposed method) (paired t-test; $p > 0.05$). For clarity, we use an asterisk above the box to denote the statistical significant difference between the results produced by the compared method and our method.

3.D. Evaluation on intersubject image registration

The most difficult task for infant image registration is to align two infant images from different subjects with large age gap. In this section, we provide both visual and quantitative evaluations of registration performance as follows. First, we

TABLE V. Intrasubject registration results. The mean and standard deviation of hippocampus Dice ratios on ten subjects by registering images at different time points. The best result for each column is shown in bold.

	2-week to 12-month	3-month to 12-month	6-month to 12-month	9-month to 12-month
FLIRT	0.543 ± 0.083[a]	0.557 ± 0.085[a]	0.615 ± 0.073[a]	0.705 ± 0.061[a]
MI-method	0.581 ± 0.072[a]	0.593 ± 0.070[a]	0.676 ± 0.061[a]	0.751 ± 0.053[c]
CC-method	0.579 ± 0.066[a]	0.585 ± 0.065[a]	0.681 ± 0.064[a]	0.748 ± 0.045[b]
3D-HAMMER	0.601 ± 0.061[b]	0.613 ± 0.062[a]	0.683 ± 0.065[a]	0.736 ± 0.051[a]
Our method	**0.614** ± 0.063	**0.635** ± 0.055	**0.702** ± 0.059	**0.762** ± 0.049

[a] $p < 0.001$.
[b] $p < 0.01$.
[c] $p < 0.05$ for paired t-test.

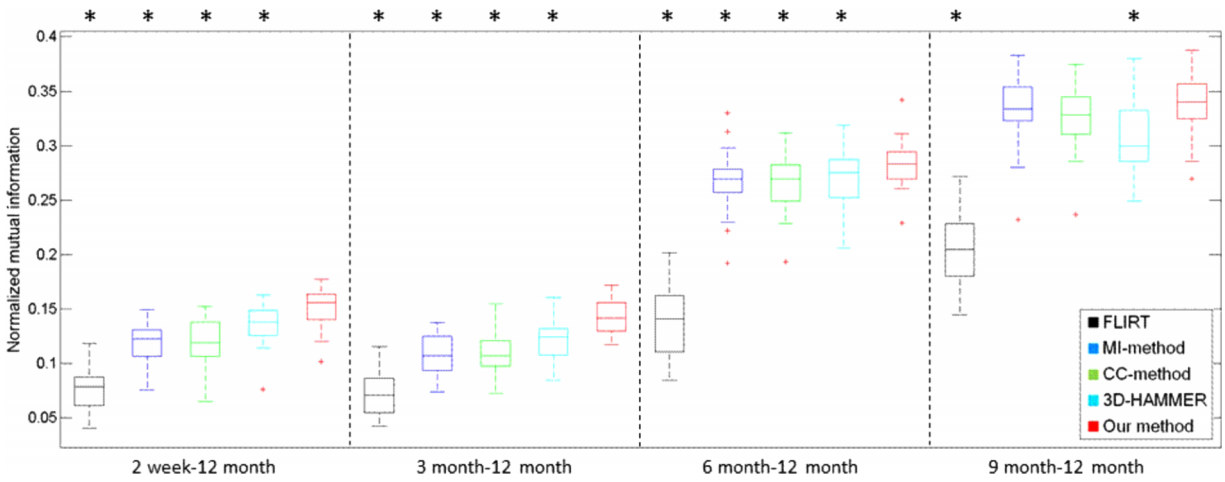

FIG. 6. Intrasubject registration results. Whisker plots of the normalized mutual information of registered images by using five methods: FLIRT (black), MI-method (blue), CC-method (green), 3D-HAMMER (cyan), and our method (red). From left to right, each column denotes the results of registering 2-week-old, 3-month-old, 6-month-old, and 9-month-old images to their respective 12-month-old images. An asterisk above the box denotes the statistical significant difference between the results produced by the compared method and our method.

choose a 12-month-old image from one of 24 subjects. Then, we separately register the infant images of other subjects (at 2-week-old, 3-month-old, 6-month-old, and 9-month-old) to the selected 12-month-old image. Our method uses the growth trajectories of the remaining 22 subjects for correspondence detection. The two subjects with their images used for testing are excluded from the training set. We repeat the above procedure for 24 times, with each time leaving two subjects out.

Figure 8 shows the typical registration results by FLIRT (second row), MI-method (third row), CC-method (fourth row), 3D-HAMMER (fifth row), and our method (sixth row), with the selected 12-month-old image shown in Fig. 8(a). From left to right in columns (b)–(e), the 2-week-old, 3-month-old, 6-month-old, and 9-month-old images are displayed in the first row, respectively. Through visual inspection, our method achieves the best registration results, especially for the locations pointed by the arrows and dashed boxes.

3.D.1. Tissue overlap ratio

We further examine the tissue overlap ratio of different subjects scanned at different time points. In Table VI, we show the mean and standard deviation of combined WM and GM Dice ratios in registering 2-week-old, 3-month-old, 6-month-old, and 9-month-old images to a 12-month-old image of another subject by FLIRT, MI-method, CC-method, 3D-HAMMER, and our infant registration method, respectively. Specially, for registering the 2-week-old, 3-month-old, and 6-month-old images to the 12-month-old image of another subject, our method significantly improves the combined WM and GM Dice ratios for 5.9%, 6.3%, and 2.0% over MI-method, 3.0%, 4.7%, and

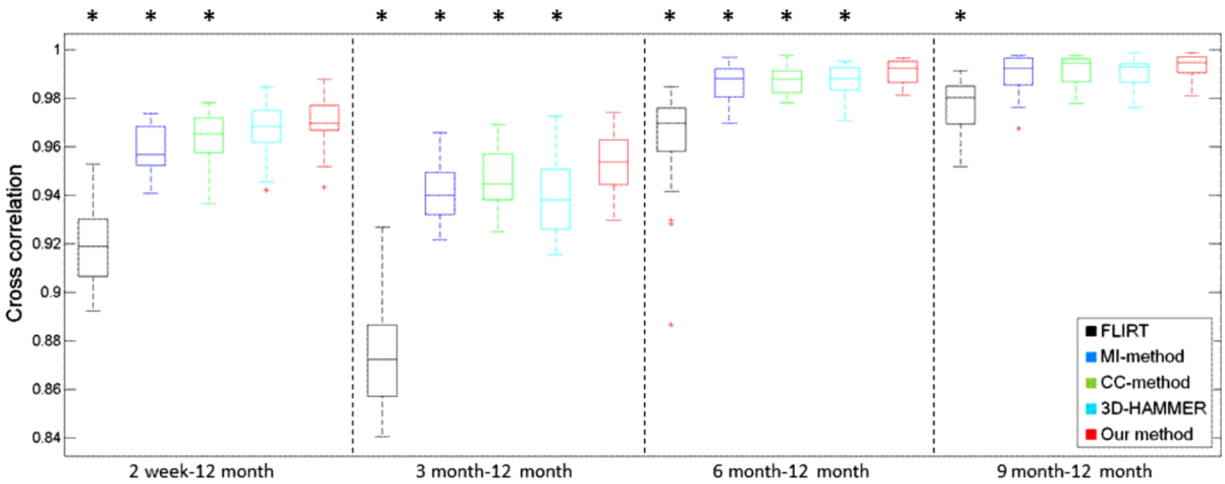

FIG. 7. Intrasubject registration results. Whisker plots of the cross correlation of registered images by using five methods: FLIRT (black), MI-method (blue), CC-method (green), 3D-HAMMER (cyan), and our method (red). From left to right, each column denotes the results of registering 2-week-old, 3-month-old, 6-month-old, and 9-month-old images to their respective 12-month-old images. An asterisk above the box denotes the statistical significant difference between the results produced by the compared method and our method.

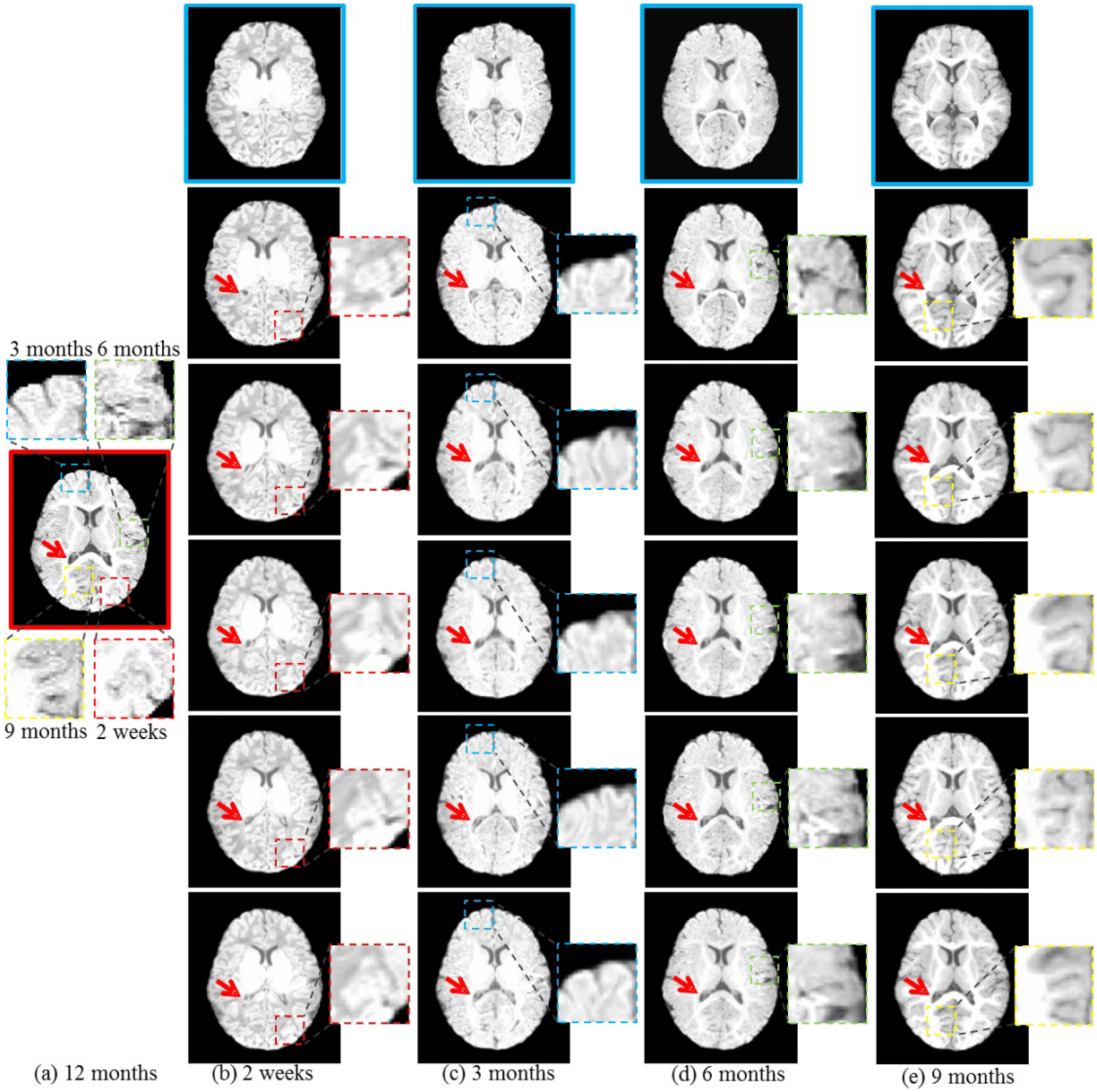

FIG. 8. Intersubject registration results of aligning 2-week-old, 3-month-old, 6-month-old, and 9-month-old images (shown in columns (b)–(e) of the first row) to a 12-month-old image in (a). The registration results by FLIRT, MI-method, CC-method, 3D-HAMMER, and our method are shown in the rows 2–6, with the obvious improvements by our method indicated by arrows and dashed boxes.

1.1% over CC-method, and 1.1%, 2.3%, and 0.7% over 3D-HAMMER, respectively. The statistical significance levels are also shown in Table VI. For registering the 9-month-old image to the fixed 12-month-old image of another subject, we found no statistically significant differences between our method and other three deformable registration methods (i.e., MI-method, CC-method, and 3D-HAMMER) (paired t-test; $p > 0.05$).

3.D.2. Hippocampus overlap ratio

This section provides the evaluation results of hippocampus Dice ratios for the intersubject registration. Table VII shows the registration results from the 2-week-old, 3-month-old, 6-month-old, and 9-month-old to the 12-month-old domain by FLIRT, MI-method, CC-method, 3D-HAMMER, and our method. We can see that in most of the cases (i.e., for registering the 2-week-old, 3-month-old, and 6-month-old to the 12-month-old domain), the proposed method achieves significantly higher hippocampus Dice ratio than any other compared registration method, which indicates the effectiveness of the proposed method for infant brain registration.

3.E. Discussion

Based on the above results on intrasubject and intersubject registrations, it can be observed that the intrasubject registration results are significantly better than the intersubject registration results, since the intersubject anatomical variation is larger than the intrasubject anatomical variations. Also, it is apparent that the registration of 2-week-old image to the

TABLE VI. Intersubject registration results. The mean and standard deviation of combined WM and GM Dice ratios, obtained by FLIRT, MI-method, CC-method, 3D-HAMMER, and our method in registering the 2-week-old, 3-month-old, 6-month-old, and 9-month-old images to a 12-month-old image of another subject in the leave-two-subjects-out cross validation. The best result for each column is shown in bold.

	2-week to 12-month	3-month to 12-month	6-month to 12-month	9-month to 12-month
FLIRT	0.573 ± 0.032[a]	0.570 ± 0.036[a]	0.606 ± 0.034[a]	0.618 ± 0.030[a]
MI-method	0.614 ± 0.025[a]	0.602 ± 0.022[a]	0.675 ± 0.030[a]	0.717 ± 0.022
CC-method	0.643 ± 0.030[a]	0.618 ± 0.025[a]	0.684 ± 0.024[b]	**0.722** ± 0.020
3D-HAMMER	0.662 ± 0.023[b]	0.642 ± 0.026[a]	0.688 ± 0.025[c]	0.715 ± 0.023
Our method	**0.673** ± 0.016	**0.665** ± 0.017	**0.695** ± 0.015	0.718 ± 0.018

[a] $p < 0.001$.
[b] $p < 0.01$.
[c] $p < 0.05$ for paired t-test.

12-month-old image is more challenging than the registration of 9-month-old image to the 12-month-old image. This is because of rapid appearance changes of infant brain during the first year of life. As the infant brain is becoming more and more similar to the 12-month template image, the registration gets easier as the age gap between the two images under registration becomes smaller, which is reflected by the results using affine registration (i.e., FLIRT).

From Fig. 1, we can see that for images obtained before 6 months, the tissue contrast of T2 images is stronger than that of T1 images. However, for the images obtained later than 6 months, the tissue contrast of T1 images is stronger than that of T2 images. Therefore, we use T2 images at 2 weeks and 3 months, but we use T1 images at 6, 9, and 12 months, for all the comparison methods. The proposed method can handle multimodality image registration. To test whether it is better to use T2 images for early scans in the comparison methods, we conducted experiments to register T1 images at 2 weeks and 3 months with the 12-month T1 image by using SyN (i.e., MI-method and CC-method). For the intrasubject registration from the 2-week T1 image to the 12-month T1 image, the mean and standard deviation of combined WM and GM Dice ratios by using MI-method and CC-method are 0.701 ± 0.026 and 0.704 ± 0.029, respectively. Using T2 images at 2 weeks improves 0.8% and 1.0% mean Dice ratios, compared to the case of using T1 images. For the intrasubject registration from the 3-month T1 image to the 12-month T1 image, the combined WM and GM Dice ratios by using MI-method and CC-method are 0.687 ± 0.032 and 0.676 ± 0.034, respectively. Using T2 images at 3 months improves 0.7% and 1.1% mean Dice ratios, compared to the case of using T1 images. Besides, we also calculated the normalized mutual information and the cross correlation of the registered images to further evaluate the registration performance, where using T2 images at early scans also obtains higher normalized mutual information and cross-correlation values than using T1 images. For example, when using CC-method in the intrasubject registration, the mean normalized mutual information of 24 subjects obtained by using T1 and T2 at 2 weeks is 0.093 and 0.121, respectively; the mean cross-correlation values of 24 subjects obtained by using T1 and T2 at 2 weeks are 0.951 and 0.963, respectively. Similarly, the proposed method also performs better when using T2 images at 2 weeks and 3 months than using T1 images. Therefore, in our experiment, we use T2 at 2 weeks and 3 months, while T1 at 6, 9, and 12 months. Considering that T1 images and T2 images belong to two different modalities, we use MI and CC in SyN, since MI and CC are regularly used as the similarity measures in multimodality image registration.

It is worth noting the slightly worse registration result by 3D-HAMMER, compared to our proposed infant registration method. This is because 3D-HAMMER was performed on the infant images that were separately segmented at different time points with iBEAT, without using any longitudinal information. As a result, the segmentation results are not accurate, compared to those segmented with guidance of longitudinal information in iBEAT. On the other hand, when given the 4D segmentation results by iBEAT, 3D-HAMMER can obtain the respective combined WM and GM Dice ratios of 78.5% ± 1.7% and 70.1% ± 1.5% for registering the 2-week-old image to the 12-month-old image of same subject (intrasubject registration case) and different subjects (intersubject registration case), respectively, while our method only can achieve

TABLE VII. Intersubject registration results. The mean and standard deviation of hippocampus Dice ratios on ten subjects by registering images at different time points. The best result for each column is shown in bold.

	2-week to 12-month	3-month to 12-month	6-month to 12-month	9-month to 12-month
FLIRT	0.413 ± 0.092[a]	0.404 ± 0.097[a]	0.458 ± 0.076[a]	0.509 ± 0.075[a]
MI-method	0.437 ± 0.073[a]	0.465 ± 0.072[a]	0.569 ± 0.058[a]	**0.631** ± 0.061
CC-method	0.442 ± 0.081[a]	0.471 ± 0.078[a]	0.573 ± 0.053[b]	0.627 ± 0.053
3D-HAMMER	0.461 ± 0.071[b]	0.512 ± 0.069[a]	0.568 ± 0.062[a]	0.621 ± 0.050
Our method	**0.478** ± 0.063	**0.535** ± 0.067	**0.585** ± 0.051	0.625 ± 0.055

[a] $p < 0.001$.
[b] $p < 0.01$ for paired t-test.

77.4% ± 1.5% and 67.3% ± 1.6% in intra- and intersubject registration applications. Although 3D-HAMMER can obtain better registration result, it highly depends on the quality of presegmented images, which are dependent of the availability of longitudinal data. Our method is more flexible and can be applied for registering images with no complete longitudinal images.

In our experiment, we did not compare with 4D-HAMMER registration method. This is because 4D-HAMMER is performed on the 4D segmentation maps of the longitudinal images of each subject. Generally, it is hard to obtain the longitudinal image sequence of a subject and also hard to get accurate 4D segmentations of the longitudinal images. The current study aims to register any two intensity images of the same subject or different subjects. 4D-HAMMER is equipped with longitudinal constraint in registration to make the estimated temporal correspondence consistent from one time point to another time point, which obtains satisfactory registration results. Here, we show the Dice ratio calculated on WM and GM by using 4D-HAMMER. The mean and standard deviation of combined WM and GM Dice ratios obtained for registering 2-week-old, 3-month-old, 6-month-old, and 9-month-old images to the 12-month-old image over 24 intrasubject registration cases are 0.834 ± 0.012, 0.860 ± 0.008, 0.882 ± 0.013, and 0.895 ± 0.011, respectively. In our study, we use 4D-HAMMER, combined with a 4D segmentation method, to establish the growth trajectories in our training dataset for guiding the registration of two new intensity images.

To evaluate the registration performance, we use Dice ratio to calculate the tissue (i.e., WM and GM) and hippocampus overlaps. On the other hand, it is worth noting that the validation on tissue types, even ignoring segmentation errors, cannot accurately indicate the registration accuracy.[54] Since the registration performance validation is difficult, no metric alone is sufficient to evaluate the effectiveness of the registration methods.[31] Besides using the Dice overlap measure, we also calculate the normalized mutual information and cross correlation of the registered images to further evaluate the registration performance.

All experiments were performed on a computer cluster with 3.10 GHz Intel processors, 12 M L3 cache, and 48 GB memory. The average computation times of all comparison methods, i.e., FLIRT, MI-method, CC-method, 3D-HAMMER, and the proposed method, are 4.1, 12.5, 14.6, 28.7, and 26.3 min, respectively. Future work includes using CUDA (compute unified device architecture) to significantly reduce the computation time of the proposed method.

4. CONCLUSION

We have presented a novel method for the registration of infant brain MR images. In the proposed method, two new strategies were particularly developed to deal with the dynamic appearance changes in the first year images. The proposed method has been evaluated on 24 infant subjects, each with longitudinal T1- and T2-weighted MR images at five different time points (2-week-old, 3-month-old, 6-month-old, 9-month-old, and 12-month-old). It achieved better performance compared to several state-of-the-art image registration methods. Future work includes more extensive evaluation of our current method and its potential applications in other studies.

ACKNOWLEDGMENTS

This research was supported by the grants from the National Natural Science Funds of China (NSFC) (Nos. 81101109, 31371009, 61471187, and 61473190), the National High Technology Research and Development Program of China (863 Program) (No. 2012AA02A616), the National Science and Technology Pillar Program of China (No. 2012BAI14B02), the Program of Pearl River Young Talents of Science and Technology in Guangzhou (Nos. 2013J2200065 and 2012J2200041), Guangdong Natural Science Foundation (2014A030313316), and National Institutes of Health (NIH) (Nos. MH100217, AG042599, MH070890, EB006733, EB008374, EB009634, NS055754, MH064065, and HD053000).

[a]Electronic addresses: yaowu1588@gmail.com, grwu@med.unc.edu, li_wang@med.unc.edu, bmunsell@cofc.edu, wang.qian@sjtu.edu.cn, weili_lin@med.unc.edu, and wufanchen@gmail.com
[b]Authors to whom correspondence should be addressed. Electronic addresses: qianjinfeng08@gmail.com and dgshen@med.unc.edu

[1]M. S. Choe, S. Ortiz-Mantilla, N. Makris, M. Gregas, J. Bacic, D. Haehn, D. Kennedy, R. Pienaar, V. S. Caviness, Jr., A. A. Benasich, and P. E. Grant, "Regional infant brain development: An MRI-based morphometric analysis in 3 to 13 month olds," Cereb. Cortex **23**, 2100–2117 (2013).
[2]R. Knickmeyer, S. Gouttard, C. Kang, D. Evans, K. Wilber, J. Smith, R. Hamer, W. Lin, G. Gerig, and J. Gilmore, "A structural MRI study of human brain development from birth to 2 years," J. Neurosci. **28**, 12176–12182 (2008).
[3]B. Casey, N. Tottenham, C. Liston, and S. Durston, "Imaging the developing brain: What have we learned about cognitive development?," Trends Cognit. Sci. **9**, 104–110 (2005).
[4]J. Giedd and J. Rapoport, "Structural MRI of pediatric brain development: What have we learned and where are we going?," Neuron **67**, 728–734 (2010).
[5]F. Shi, P.-T. Yap, G. Wu, H. Jia, J. H. Gilmore, W. Lin, and D. Shen, "Infant brain atlases from neonates to 1- and 2-year-olds," PLoS One **6**, e18746 (2010).
[6]Z. Xue, D. Shen, and C. Davatzikos, "Determining correspondence in 3-D MR brain images using attribute vectors as morphological signatures of voxels," IEEE Trans. Med. Imaging **23**, 1276–1291 (2004).
[7]Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," Med. Image Anal. **15**, 622–639 (2011).
[8]I. Csapo, B. Davis, Y. Shi, M. Sanchez, M. Styner, and M. Niethammer, "Longitudinal image registration with temporally-dependent image similarity measure," IEEE Trans. Med. Imaging **32**, 1939–1951 (2013).
[9]D. Shen, W. H. Wong, and H. H. S. Ip, "Affineinvariant image retrieval by correspondence matching of shapes," Image Vis. Comput. **17**, 489–499 (1999).
[10]M. Kim, G. Wu, and D. Shen, "Sparse patch-guided deformation estimation for improved image registration," in *Machine Learning in Medical Imaging* (Springer, New York, NY, 2012), pp. 54–62.
[11]L. Wang, F. Shi, P.-T. Yap, J. H. Gilmore, W. Lin, and D. Shen, "4D multi-modality tissue segmentation of serial infant images," PLoS One **7**, e44596 (2012).
[12]J. Nie, G. Li, and D. Shen, "Development of cortical anatomical properties from early childhood to early adulthood," NeuroImage **76**, 216–224 (2013).
[13]P. Aljabar, K. Bhatia, M. Murgasova, J. Hajnal, J. Boardman, L. Srinivasan, M. Rutherford, L. Dyet, A. Edwards, and D. Rueckert, "Assessment of brain growth in early childhood using deformation-based morphometry," NeuroImage **39**, 348–358 (2008).

[14]Y. Fan, F. Shi, J. K. Smith, W. Lin, J. H. Gilmore, and D. Shen, "Brain anatomical networks in early human brain development," NeuroImage **54**, 1862–1871 (2011).
[15]T. Paus, D. Collins, A. Evans, G. Leonard, B. Pike, and A. Zijdenbos, "Maturation of white matter in the human brain: A review of magnetic resonance studies," Brain Res. Bull. **54**, 255–266 (2001).
[16]N. I. Weisenfeld and S. K. Warfield, "Automatic segmentation of newborn brain MRI," NeuroImage **47**, 564–572 (2009).
[17]S. C. Deoni, E. Mercure, A. Blasi, D. Gasston, A. Thomson, M. Johnson, S. C. Williams, and D. G. Murphy, "Mapping infant brain myelination with magnetic resonance imaging," J. Neurosci. **31**, 784–791 (2011).
[18]R. Verma, S. Mori, D. Shen, P. Yarowsky, J. Zhang, and C. Davatzikos, "Spatiotemporal maturation patterns of murine brain quantified by diffusion tensor MRI and deformation-based morphometry," Proc. Natl. Acad. Sci. U. S. A. **102**, 6978–6983 (2005).
[19]F. Shi, P.-T. Yap, J. H. Gilmore, W. Lin, and D. Shen, "Spatial-temporal constraint for segmentation of serial infant brain MR images," in *Medical Imaging and Augmented Reality* (Springer, New York, NY, 2010), pp. 42–50.
[20]L. Wang, F. Shi, Y. Gao, G. Li, J. H. Gilmore, W. Lin, and D. Shen, "Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain MR image segmentation," NeuroImage **89**, 152–164 (2014).
[21]R. B. Dietrich, W. Bradley, E. Zaragoza 4th, R. Otto, R. Taira, G. Wilson, and H. Kangarloo, "MR evaluation of early myelination patterns in normal and developmentally delayed infants," Am. J. Roentgenol. **150**, 889–896 (1988).
[22]A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," IEEE Trans. Med. Imaging **32**, 1153–1190 (2013).
[23]G. Wu, H. Jia, W. Qian, and D. Shen, "SharpMean: Groupwise registration guided by sharp mean image and tree-based registration," NeuroImage **56**, 1968–1981 (2011).
[24]A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M. C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," NeuroImage **46**, 786–802 (2009).
[25]J. Yang, D. Shen, C. Davatzikos, and R. Verma, "Diffusion tensor image registration using tensor geometry and orientation features," in *Lecture Notes in Computer Science* (Springer, Berlin, Heidelberg, 2008), Vol. 11, pp. 905–913.
[26]Z. Xue, D. Shen, and C. Davatzikos, "Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping," Med. Image Anal. **10**, 740–751 (2006).
[27]S. Tang, Y. Fan, G. Wu, M. Kim, and D. Shen, "RABBIT: Rapid alignment of brains by building intermediate templates," NeuroImage **47**, 1277–1287 (2009).
[28]P. T. Yap, G. Wu, H. Zhu, W. Lin, and D. Shen, "TIMER: Tensor image morphing for elastic registration," NeuroImage **47**, 549–563 (2009).
[29]H. Xue, L. Srinivasan, S. Jiang, M. Rutherford, A. D. Edwards, D. Rueckert, and J. V. Hajnal, "Longitudinal cortical registration for developing neonates," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007* (Springer, Berlin, Heidelberg, 2007), Vol. 10, pp. 127–135.
[30]L. Ha, M. Prastawa, G. Gerig, J. H. Gilmore, C. T. Silva, and S. Joshi, "Efficient probabilistic and geometric anatomical mapping using particle mesh approximation on GPUs," Int. J. Biomed. Imaging 1–16 (2011).
[31]A. Serag, P. Aljabar, S. Counsell, J. Boardman, J. V. Hajnal, and D. Rueckert, *Presented at the 9th IEEE International Symposium on Biomedical Imaging (ISBI)* (unpublished).
[32]W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," Med. Image Anal. **1**, 35–51 (1996).
[33]A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, *Presented at the Information Processing in Medical Imaging* (unpublished).
[34]W. K. Pratt, "Correlation techniques of image registration," IEEE Trans. Aerosp. Electron. Syst. **AES-10**, 353–358 (1974).
[35]A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Medical Image Computing and Computer-Assisted Interventation—MICCAI'98* (Springer, New York, NY, 1998), pp. 1115–1124.
[36]D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," IEEE Trans. Med. Imaging **21**, 1421–1439 (2002).
[37]Y. Dai, F. Shi, L. Wang, G. Wu, and D. Shen, "iBEAT: A toolbox for infant brain magnetic resonance image processing," Neuroinformatics **11**, 211–225 (2013).
[38]J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," IEEE Trans. Med. Imaging **17**, 87–97 (1998).
[39]F. Shi, L. Wang, J. H. Gilmore, W. Lin, and D. Shen, "Learning-based meta-algorithm for MRI brain extraction," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011* (Springer, Berlin, Heidelberg, 2011), Vol. 14, pp. 313–321.
[40]M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," Med. Image Anal. **5**, 143–156 (2001).
[41]D. Shen and C. Davatzikos, "Measuring temporal morphological changes robustly in brain MR images via 4-dimensional template warping," NeuroImage **21**, 1508–1517 (2004).
[42]G. Wu, M. Kim, Q. Wang, and D. Shen, "S-HAMMER: Hierarchical attribute-guided, symmetric diffeomorphic registration for MR brain images," Hum. Brain Mapp. **35**, 1044–1060 (2013).
[43]J. M. Hammersley and D. C. Handscomb, "Monte Carlo methods," Monographs on Applied Probability & Statistics **30**(2–3), 347–385 (1964).
[44]J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," Proc. IEEE **98**, 1031–1044 (2010).
[45]W. E. Vinje and J. L. Gallant, "Sparse coding and decorrelation in primary visual cortex during natural vision," Science **287**, 1273–1276 (2000).
[46]H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," J. R. Stat. Soc.: Ser. B (Stat. Methodol.) **67**, 301–320 (2005).
[47]R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Stat. Soc.: Ser. B (Stat. Methodol.) **58**(1), 267–288 (1996).
[48]L. Wang, F. Shi, G. Li, Y. Gao, W. Lin, J. H. Gilmore, and D. Shen, "Segmentation of neonatal brain MR images using patch-driven level sets," Neuroimage **84**, 141–158 (2014).
[49]H. Cheng, Z. Liu, and J. Yang, *Presented at the IEEE 12th International Conference on Computer Vision* (unpublished).
[50]G. Wu, P. T. Yap, M. Kim, and D. Shen, "TPS-HAMMER: Improving HAMMER registration algorithm by soft correspondence matching and thin-plate splines based deformation interpolation," NeuroImage **49**, 2225–2233 (2010).
[51]F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 567–585 (1989).
[52]J. Ashburner, "A fast diffeomorphic image registration algorithm," NeuroImage **38**, 95–113 (2007).
[53]T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," NeuroImage **45**, S61–S72 (2009).
[54]T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable," IEEE Trans. Med. Imaging **31**, 153–163 (2012).