Longitudinal Intensity Normalization of Magnetic Resonance Images using Patches

Snehashis Roy, **Aaron Carass**, and **Jerry L. Prince**
Image Analysis and Communications Laboratory, The Johns Hopkins University

Abstract

This paper presents a patch based method to normalize temporal intensities from longitudinal brain magnetic resonance (MR) images. Longitudinal intensity normalization is relevant for subsequent processing, such as segmentation, so that rates of change of tissue volumes, cortical thickness, or shapes of brain structures becomes stable and smooth over time. Instead of using intensities at each voxel, we use patches as image features as a patch encodes neighborhood information of the center voxel. Once all the time-points of a longitudinal dataset are registered, the longitudinal intensity change at each patch is assumed to follow an auto-regressive (AR(1)) process. An estimate of the normalized intensities of a patch at every time-point are generated from a hidden Markov model, where the hidden states are the unobserved normalized patches and the outputs are the observed patches. A validation study on a phantom dataset shows good segmentation overlap with the truth, and an experiment with real data shows more stable rates of change for tissue volumes with the temporal normalization than without.

Keywords

Intensity normalization; intensity standardization; MRI; patch; brain

1. INTRODUCTION

MR is a widely used noninvasive modality to image the structure of human brains. Associated image processing techniques, such as segmentation, are needed to understand normal aging[1,2] as well as the progression of diseases. Analysis of 4D temporal data—multiple 3D images at different time-points–is relevant in this scenario to estimate the rates of change of image statistics (i.e., tissue shapes, volumes, and cortical thickness). Unlike computed tomography, MR image intensities do not possess any tissue specific meaning, which introduces inconsistencies in MR segmentations.[3] Thus image intensities need to be on a standardized scale to achieve temporally consistent results. However, 3D segmentation algorithms, performed independently on a longitudinal dataset, seldom achieve the desired longitudinal stability, giving rise to the need for a 4D normalization technique.

Several 3D intensity normalization techniques have been proposed in the literature to bring the varying MR intensity ranges to a common scale. Most of these methods are based on deforming the intensity histogram to match a template histogram from an atlas[4] or using spatial landmarks[5,6], or some information-theoretic criteria[7,8] to match histograms. They work well in normalizing one 3D volume to another, however they do not produce the desired stability on a 4D dataset. A state-of-the-art 4D image segmentation technique, CLASSIC[9], was proposed to account for the desired longitudinal smoothness of the 4D

Further author information: Send correspondence to Jerry L. Prince prince@jhu.edu.

segmentations. A 4D segmentation algorithm was also proposed to obtain consistent segmentations and cortical thicknesses of infant brains[10], where the segmentations of the later years are used as a prior to the first year and a novel 4D penalty is introduced on the cortical thickness measurements[11]. Several probabilistic atlas based methods[12,13] have also been proposed where subject specific atlases are created from either manual segmentation or an average template to segment both cortical and sub-cortical structures. Unfortunately these algorithms are closely tied to a particular segmentation algorithm and can not easily be extended to other segmentation methods. In this paper, we propose a 4D image intensity normalization method that can be used as a pre-processing step to any segmentation algorithm. Instead of using intensities at each voxel, our method uses patches as features, because a 3D patch around a voxel encodes the neighborhood information at that voxel. We assume that the longitudinal change of intensity of a patch follows an AR(1) process and then estimate the normalized intensity of that patch using a hidden Markov model (HMM)[14]. The hidden states of the HMM are the unobserved normalized patches, while the corresponding outputs are the observed image patches. The motivation for such a model is shown in Figs. 1(a)–(c) where three time-points are shown for a longitudinal dataset[2]. The images are scaled so that their white matter (WM) peaks are unity and Figs. 1(b) and (c) are rigidly registered to Fig. 1(a). The scaled intensities of a selected voxel, indicated as the center of the blue square in Fig. 1(a), are plotted as a blue line in Fig. 1(d). An AR(1) fit of the intensities is also shown in Fig. 1(d) as a red line, with $R^2 = 0.91$. Thus for normal aging or gradual progression of diseases (e.g., Alzheimer's Disease), we assume that the intensities of the registered voxels smoothly change. In the following section, we describe the algorithm in detail.

2. METHOD

We assume that there are T time-points in our 4D dataset. The 3D volume of each time-point is denoted as S_t, $t = 1, \ldots, T$, and are all registered to S_1. The S_t's have also been linearly scaled such that their WM peaks are at unity. Each 3D volume S_t is made up of small $p \times q \times r$ 3D patches, which we stack as 1D vectors $y_i^{(t)} \in \mathbb{R}^d$, where i denotes the spatial location of the center voxel of the patch and d is the dimension, $d = pqr$. In our HMM model, shown in Fig. 2, each of the $y_i^{(t)}$,s has a hidden state, denoted as $x_i^{(t)}$, which is the corresponding unobserved normalized patch. Since the images are registered, we assume the hidden states $x_i^{(1 \cdots T)} \equiv \left\{ x_i^{(1)}, \ldots, x_i^{(T)} \right\}$ follow a Markov process and the transition of the normalized patches from $(t-1)^{th}$ time-point $x_i^{(t-1)}$ to the tth time-point $x_i^{(t)}$ is an AR(1) process

$$x_i^{(t)} = M_i x_i^{(t-1)} + \epsilon_i^{(t)} \quad \text{for} \quad t \geq 2, \quad \text{with} \quad \epsilon_i^{(t)} \sim \mathcal{N}\left(0, \sigma_{\epsilon,i}^2 I\right), \quad (1)$$

where M is a spatially varying matrix denoting the parameter of the AR(1) process. $\epsilon_i^{(t)}$ denotes the time-varying noise, and is assumed to be a zero mean uncorrelated Gaussian with diagonal covariance, I being the identity matrix.

The parameter of the AR(1) process, M_i, controls how much the previous state $x_i^{(t-1)}$ contributes to the next state $x_i^{(t)}$. For computational simplicity, we assume that M_i is a diagonal matrix, $M_i = \text{diag}\left\{ m_i^{(1)}, \ldots, m_i^{(d)} \right\}$, If $m_i^{(\ell)}$ is close to zero, then the process looks like white noise. As $m_i^{(\ell)}$ approaches 1, the tth state gets more contribution from the $(t-1)^{th}$ state, and the result is a smoothing effect of the intensities. One example is shown in Fig. 1(d), where the 1D intensities are best fitted with parameter $m = 0.89$. If $m_i^{(\ell)} = 1$, the

intensities only vary according to the noise limit $\sigma_{\varepsilon,i}$, which is expected for a voxel having the same tissue class for all the time-points, e.g., a deep WM patch. However, without any prior spatial knowledge on the values of $m^{(\ell)}$, we assume a Gaussian distribution, $m_i^{(\ell)} \sim \mathcal{N}\left(1, \sigma_{m,i}^2\right)$. This is based on the assumption that most often intensities remain similar within a noise range, e.g. inside deep WM, but occasionally, as shown in a voxel near ventricles (Fig. 1(a)) or near edges, they decrease in intensity over time, indicating a change of tissue type at that location.

The observed patches are the outputs of the HMM and they are assumed to be obtained from the hidden states as

$$P\left(\mathrm{y}_i^{(t)}|\mathrm{x}_i^{(t)}\right) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left\{-\frac{1}{2}\left(\frac{\mathrm{y}_i^{(t)} - \mathrm{x}_i^{(t)}}{\sigma_i}\right)^2\right\}. \quad (2)$$

Here, we have also assumed the errors in obtaining $\mathrm{y}_i^{(t)}$ from $\mathrm{x}_i^{(t)}$ to be uncorrelated zero mean Gaussian with variance σ_i^2. Using the conditional independence of $\mathrm{x}^{(1\cdots T)}$ and $\mathrm{y}^{(1\cdots T)}$, as often assumed in HMM, we can write

$$\begin{aligned} P\left(\mathrm{x}_i^{(t)}|\mathrm{x}_i^{(1)}, \ldots, \mathrm{x}_i^{(t-1)}\right) &= P\left(\mathrm{x}_i^{(t)}|\mathrm{x}_i^{(t-1)}\right), \\ P\left(\mathrm{y}_i^{(t)}|\mathrm{x}_i^{(1\cdots T)}\right) &= P\left(\mathrm{y}_i^{(t)}|\mathrm{x}_i^{(t)}\right). \end{aligned} \quad (3)$$

The collection of all unknown parameters are denoted as $\Theta = \left\{\sigma_i, \sigma_{m,i}, \sigma_{\epsilon,i}, m_i^{(\ell)}\right\}$. Now the parameter set Θ and the hidden states $\mathrm{x}_i^{(1\cdots T)}$ are estimated by a maximum-a-posteriori (MAP) estimation criteria by maximizing the posterior probability of $\mathrm{x}_i^{(1\cdots T)}$ as

$$\begin{aligned} \left(\mathrm{x}_i^{(1\cdots T)}, \Theta\right) &= \underset{\mathrm{x}_i^{(1\cdots T)}, \Theta}{\operatorname{argmax}} P\left(\mathrm{x}_i^{(1\cdots T)}, \Theta|\mathrm{y}_i^{(1\cdots T)}\right), \\ &= \underset{\mathrm{x}_i^{(1\cdots T)}, \Theta}{\operatorname{argmax}} P\left(\mathrm{y}_i^{(1\cdots T)}|\mathrm{x}_i^{(1\cdots T)}, \Theta\right) P\left(\mathrm{x}_i^{(1\cdots T)}|\Theta\right) P\left(\Theta\right). \end{aligned} \quad (4)$$

We simplify the conditional probability of the observed data $P\left(\mathrm{y}_i^{(1\cdots T)}|\mathrm{x}_i^{(1\cdots T)}, \Theta\right)$ using Eqn. 3 as

$$\begin{aligned} P\left(\mathrm{y}_i^{(1\cdots T)}|\mathrm{x}_i^{(1\cdots T)}, \Theta\right) &= P\left(\mathrm{y}_i^{(T)}|\mathrm{y}_i^{(1\cdots T-1)}, \mathrm{x}_i^{(1\cdots T)}, \Theta\right) \ldots P\left(\mathrm{y}_i^{(1)}|\mathrm{x}_i^{(1\cdots T)}, \Theta\right), \\ &= P\left(\mathrm{y}_i^{(T)}|\mathrm{x}_i^{(T)}, \Theta\right) P\left(\mathrm{y}_i^{(T-1)}|\mathrm{x}_i^{(T-1)}, \Theta\right) \ldots P\left(\mathrm{y}_i^{(1)}|\mathrm{x}_i^{(1)}, \Theta\right), \\ &= \prod_{t=1}^{T} P\left(\mathrm{y}_i^{(t)}|\mathrm{x}_i^{(t)}, \Theta\right). \end{aligned} \quad (5)$$

The conditional independence of $\mathrm{x}_i^{(t)}$'s, as given in Eqn. 3, are used to simplify $P\left(\mathrm{x}_i^{(1\cdots T)}|\Theta\right)$

$$\begin{aligned} P\left(\mathrm{x}_i^{(1\cdots T)}|\Theta\right) &= P\left(\mathrm{x}_i^{(T)}|\mathrm{x}_i^{(1\cdots T-1)}, \Theta\right) P\left(\mathrm{x}_i^{(T-1)}|\mathrm{x}_i^{(1\cdots T-2)}, \Theta\right) \ldots P\left(\mathrm{x}_i^{(2)}|\mathrm{x}_i^{(1)}, \Theta\right) \\ &= P\left(\mathrm{x}_i^{(T)}|\mathrm{x}_i^{(T-1)}, \Theta\right) P\left(\mathrm{x}_i^{(T-1)}|\mathrm{x}_i^{(T-2)}, \Theta\right) \ldots P\left(\mathrm{x}_i^{(2)}|\mathrm{x}_i^{(1)}, \Theta\right) = \prod_{t=2}^{T} P\left(\mathrm{x}_i^{(t)}|\mathrm{x}_i^{(t-1)}, \Theta\right) \end{aligned} \quad (6)$$

Eqn. 4 can now be expanded using Eqns. 5 and 6 and also using the prior distribution ($P(\Theta)$) of $m_i^{(\ell)}$, as assumed earlier. Assuming i.i.d. nature of patches and discarding some normalization constants, the posterior probability becomes

$$P\left(\mathrm{x}_i^{(1\cdots T)},\Theta|\mathrm{y}_i^{(1\cdots T)}\right) \propto \frac{1}{\sigma_i^T \sigma_{\epsilon,i}^{T-1} \sigma_{m,i}} \exp\left\{-\sum_{t=1}^{T}\frac{||\mathrm{y}_i^{(t)}-\mathrm{x}_i^{(t)}||^2}{2\sigma_i^2} - \sum_{t=2}^{T}\frac{||\mathrm{x}_i^{(t)}-M_i\mathrm{x}_i^{(t-1)}||^2}{2\sigma_{\epsilon,i}^2} - \sum_{\ell=1}^{d}\frac{\left(m_i^{(\ell)}-1\right)^2}{2\sigma_{m,i}^2}\right\}. \quad (7)$$

Maximizing the posterior with respect to Θ yields the following update equations

$$m_i^{(\ell)} = \frac{\frac{1}{\sigma_{\epsilon,i}^2}\Sigma_{t=2}^{T}\mathrm{x}_i^{(t)}(\ell)\mathrm{x}_i^{(t-1)}(\ell)+\frac{1}{\sigma_{m,i}^2}}{\frac{1}{\sigma_{\epsilon,i}^2}\Sigma_{t=2}^{T}\mathrm{x}_i^{(t-1)}(\ell)\mathrm{x}_i^{(t-1)}(\ell)+\frac{1}{\sigma_{m,i}^2}},$$

$$\sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}||\mathrm{y}_i^{(t)}-\mathrm{x}_i^{(t)}||^2, \quad \sigma_{m,i}^2 = \sum_{\ell=1}^{d}\left(m_i^{(\ell)}-1\right)^2,$$

$$\sigma_{\epsilon,i}^2 = \frac{1}{T-1}\sum_{t=2}^{T}||\mathrm{x}_i^{(t)}-M_i\mathrm{x}_i^{(t-1)}||^2,$$

$$\mathrm{x}_i^{(t)} = \left(\left(\frac{1}{\sigma_i^2}+\frac{1}{\sigma_{\epsilon,i}^2}\right)I+\frac{1}{\sigma_{\epsilon,i}^2}M_i^T M_i\right)^{-1}\left(\frac{\mathrm{y}_i^{(t)}}{\sigma_i^2}+\frac{M_i\mathrm{x}_i^{(t+1)}}{\sigma_{\epsilon,i}^2}+\frac{M_i\mathrm{x}_i^{(t-1)}}{\sigma_{\epsilon,i}^2}\right), \quad t=2,\ldots,T-1, \quad (8)$$

$$\mathrm{x}_i^{(1)} = \left(\frac{I}{\sigma_i^2}+\frac{1}{\sigma_{\epsilon,i}^2}M_i^T M_i\right)^{-1}\left(\frac{\mathrm{y}_i^{(1)}}{\sigma_i^2}+\frac{M_i\mathrm{x}_i^{(2)}}{\sigma_{\epsilon,i}^2}\right),$$

$$\mathrm{x}_i^{(T)} = \left(\frac{1}{\sigma_i^2}+\frac{1}{\sigma_{\epsilon,i}^2}\right)^{-1}\left(\frac{\mathrm{y}_i^{(T)}}{\sigma_i^2}+\frac{M_i\mathrm{x}_i^{(T-1)}}{\sigma_{\epsilon,i}^2}\right).$$

We use a coordinate-descent type optimization to solve Eqn. 8 until the difference in the posterior (Eqn. 7) between successive iterations is sufficiently small. After Eqn. 8 reaches convergence, the center voxel of $\mathrm{y}_i^{(t)}$ replaced with the center voxel of $\mathrm{x}_i^{(t)}$.

3. RESULTS

We use a tissue atrophy simulation method[15] on a real T_1-w image to simulate temporal data. Ventricles of the subject image are deformed using six different atrophy radii to simulate the normal aging at six different time-points. Then the images are normalized using this method and the hard segmentations[16] are compared with the ground truth, obtained from the simulation. We calculated Dice coefficients between the hard segmentations of un-normalized and normalized images with the truth. A paired t-test shows that the Dice coefficients are not statistically different (p-value = 0.014), indicating that normalization does not deteriorate the stability of the segmentation when the segmentation is very close to the truth.

We used seven subjects from the BLSA dataset[2] to show the segmentation stability. Each subject has 8–11 time-points. Each subject is scanned in the same GE 1.5T scanner using an SPGR imaging sequence, giving $256 \times 256 \times 199$ images with $0.9375 \times 0.9375 \times 1.5\text{mm}^3$ resolution. For one subject, four time points and the corresponding normalized images are shown in Fig. 3. We compare our method with histogram matching and a landmark based histogram transformation method[5]. For these two algorithms, we used each subject's first time-point as the reference histogram and matched the consecutive time-points to the first one using automatically chosen landmarks.

To show the improvement in segmentation stability, we segment each image with a fuzzy c-means based algorithm[16]. The relative volumes of cerebro-spinal fluid (CSF), gray matter (GM), and WM are obtained from the hard segmentations of the original, histogram

matched, landmark based normalized and our 4D normalized volumes, plotted for one subject in Fig. 3. Clearly, 4D normalized volumes are more stable over time. Since these are normal subjects, the gradual increase in WM volume and the decrease in GM volume are expected and reported previously[1]. Although the original unnormalized volumes as well as histogram matching produces a similar trend, our normalization introduces stability in the segmentation. This is also reflected quantitatively in Table 1, where the average coefficient of determination R^2 from linear fitting from the volumes are shown. Larger R^2 for normalized volumes indicate smoother rates of change. R^2 values of the fits for longitudinal normalization are significantly larger (p-value < 0.05) than the original and two other normalization algorithms, except for the CSF on the landmark based method. This can be attributed to the high variability of CSF segmentation in the SPGR images. Average coefficient of variations (CV) are also shown in Table 2, where smaller CV indicates more stability in segmentation. Our method outperforms the others with significantly smaller CV for all the tissues.

4. CONCLUSION

We have described a generative model to normalize intensities of a 4D longitudinal normal dataset that can be used as a pre-processing step to any segmentation algorithm. Instead of using intensities, we use patches as features to model an HMM to obtain an estimate of the normalized intensities via a MAP estimation framework. Although this method currently assumes that the time-points are equally spaced, future work would overcome this limitation.

Acknowledgments

REFERENCES

[1]. Resnick SM, Pham DL, Kraut MA, Zonderman AB, Davatzikos C. Longitudinal magnetic resonance imaging studies of older adults: A shrinking brain. Journal of Neuroscience. 2003; 23(8):3295–3301. [PubMed: 12716936]

[2]. Resnick SM, Goldszal AF, Davatzikos C, Golski S, Kraut MA, Metter EJ, Bryan RN, Zonderman AB. One-year age changes in MRI brain volumes in older adults. Cerebral Cortex. 2000; 10(5): 464–472. [PubMed: 10847596]

[3]. Clark KA, Woods RP, Rottenber DA, Toga AW, Mazziotta JC. Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. NeuroImage. 2006; 29(1):185–202. [PubMed: 16139526]

[4]. Weisenfeld NI, Warfield SK. Normalization of Joint Image-Intensity Statistics in MRI Using the Kullback-Leibler Divergence. Intl. Symp. on Biomed. Imaging (ISBI). Apr.2004 1:101–104.

[5]. Nyúl LG, Udupa JK, g X. New Variants of a Method of MRI Scale Standardization. IEEE Trans. on Med. Imag. 2000; 19(2):143–150.

[6]. Nyúl LG, Udupa JK. On Standardizing the MR Image Intensity Scale. Mag. Res. in Medicine. 1999; 42(6):1072–1081.

[7]. Jäger F, Hornegger J. Nonrigid Registration of Joint Histograms for Intensity Standardization in Magnetic Resonance Imaging. IEEE Trans. on Med. Imag. 2009; 28(1):137–150.

[8]. Jäger, F.; Nyúl, L.; Frericks, B.; Wacker, F.; Hornegger, J. Informatik Aktuell. Springer Berlin Heidelberg; 2007. Whole Body MRI Intensity Standardization; p. 459-463.Bildverarbeitung für die Medizin 2008

[9]. Xue Z, Shen D, Davatzikos C. CLASSIC: Consistent Longitudinal Alignment and Segmentation for Serial Image Computing. NeuroImage. 2006; 30(2):388–399. [PubMed: 16275137]

[10]. Wang L, Shi F, Yap P-T, Gilmore JH, Lin W, Shen D. 4D Multi-Modality Tissue Segmentation of Serial Infant Images. PLoS One. 2012; 7(9):e44596. [PubMed: 23049751]

[11]. Shi F, Fan Y, Tang S, Gilmore JH, Lin W, Shen D. Neonatal brain image segmentation in longitudinal MRI studies. NeuroImage. 2010; 49(1):391–400. [PubMed: 19660558]

[12]. Wang, B.; Prastawa, M.; Awate, SP.; Irimia, A.; Chambers, MC.; Vespa, PM.; Van Horn, JD.; Gerig, G. Intl. Symp. on Biomed. Imaging (ISBI). 2012. Segmentation of serial MRI of TBI patients using personalized atlas construction and topological change estimation; p. 1152-1155.

[13]. Riklin-Raviv T, Leemput KV, Menze BH, Wells WM III, Golland P. Segmentation of Image Ensembles via Latent Atlases. Med. Image Analysis. 2010; 14(5):654–665.

[14]. Baum LE, Petrie T, Soules G, Weiss N. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. Annals of Math. Statistics. 1970; 41(1): 164–171.

[15]. Karacali B, Davatzikos C. Simulation of tissue atrophy using a topology preserving transformation model. IEEE Trans. Med. Imag. 2006; 25(5):649–652.

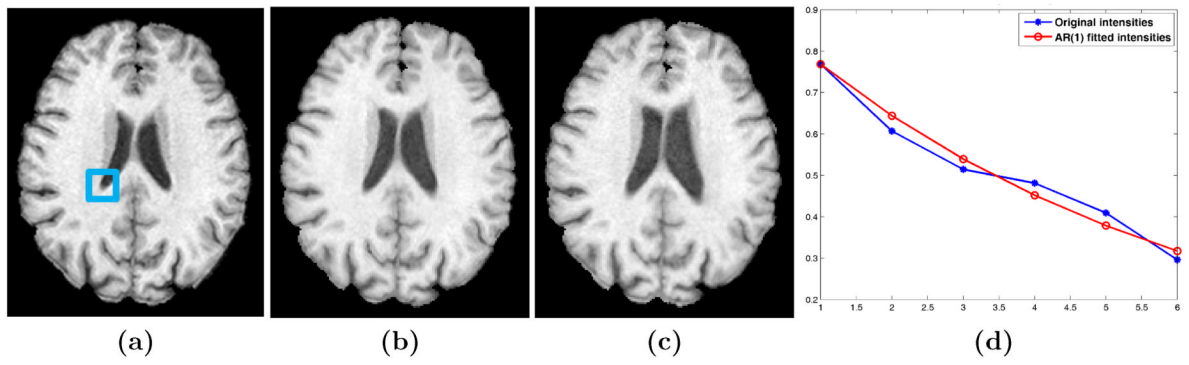[16]. Pham DL. Spatial Models for Fuzzy Clustering. Comp. Vision & Image Understand. 2001; 84(2): 285–297.

Figure 1.
(a)-**(c)** shows the first, fourth, and sixth year of a longitudinal dataset[2]. Original intensities of a voxel, the blue box in (a), are plotted in **(d)** as a blue line, while an AR(1) fit of the intensities is shown as a red line.
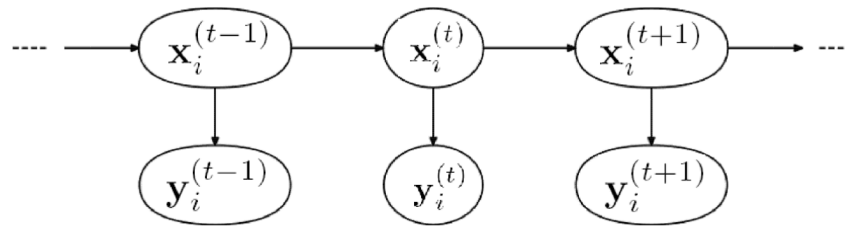
Figure 2.

A hidden Markov model with normalized patches $x_i^{(t)}$ as hidden states, while the observed patches $y_i^{(t)}$'s are the output. The transition from the $(t-1)^{th}$ time-point to the tth time-point is an AR(1) process.
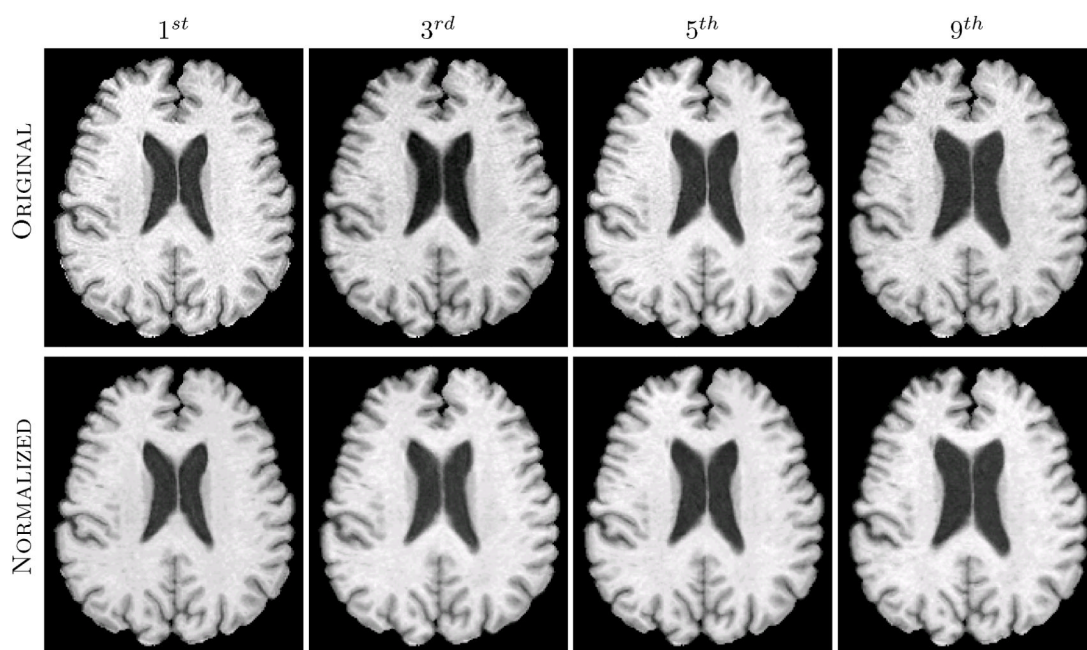
Figure 3.
Four time-points (1st, 3rd, 5th and 9th) of a subject and the corresponding normalized images are shown.
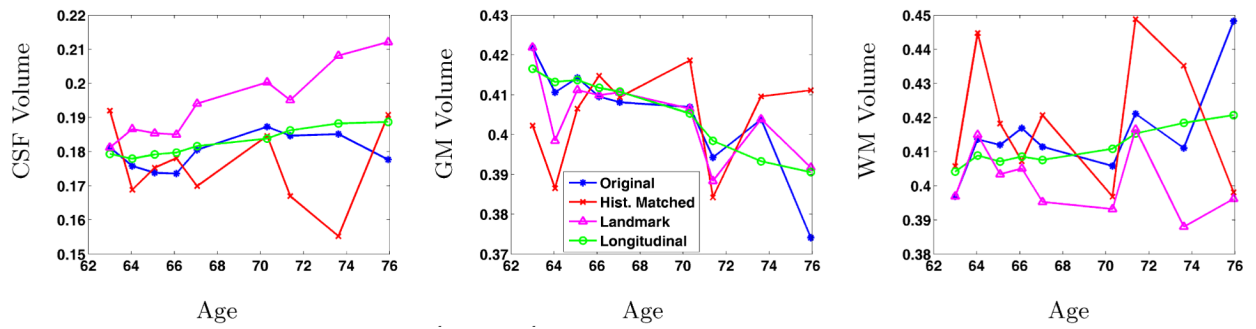
Figure 4.
(a) Three time-points (1st, 4th and 9th) of a subject and the corresponding normalized images are shown. (b) After segmentation,[16] CSF, GM and WM relative volumes are plotted w.r.t. age.

Table 1

The table shows the (mean ± std) R^2 of linear fits of the tissue volumes obtained from hard segmentations of the original image and also the hard segmentations based on three normalization approaches: histogram matched; landmark based[5]; and our approach (4D Longitudinal). The R^2 values are based on an average of seven subjects.

	Original	Hist. Matched	Landmark Based	Longitudinal
CSF	0.362 ± 0.361	0.045 ± 0.027	0.710 ± 0.280	0.724 ± 0.380[†]
GM	0.819 ± 0.067	0.068 ± 0.079	0.630 ± 0.183	0.930 ± 0.037[*]
WM	0.578 ± 0.180	0.075 ± 0.117	0.379 ± 0.248	0.903 ± 0.019[*]

[*] Indicates statistically significantly larger than the other three (p-value < 0.05).

[†] Indicates statistically significantly larger than histogram matching and original (p-value < 0.05).

Table 2

The table shows the (mean ± std) coefficient of variations of the tissue volumes obtained from hard segmentations of the original image and also the hard segmentations based on three normalization approaches: histogram matched; landmark based[5]; and our approach (4D Longitudinal). The values are based on an average of seven subjects.

	Original	Hist. Matched	Landmark Based	Longitudinal
CSF	0.028 ± 0.012	0.068 ± 0.019	0.060 ± 0.020	$0.023 \pm 0.011^{*}$
GM	0.034 ± 0.013	0.030 ± 0.016	0.026 ± 0.014	$0.024 \pm 0.010^{\dagger}$
WM	0.034 ± 0.011	0.047 ± 0.019	0.024 ± 0.017	$0.014 \pm 0.013^{*}$

[*] Indicates statistically significantly larger than the other three (p-value < 0.05).

[†] Indicates statistically significantly larger than histogram matching and original (p-value < 0.05).