2013 IEEE 10th International Symposium on Biomedical Imaging:
From Nano to Macro
San Francisco, CA, USA, April 7-11, 2013

PATCH BASED INTENSITY NORMALIZATION OF BRAIN MR IMAGES

Snehashis Roy, Aaron Carass, and Jerry L. Prince

Image Analysis and Communications Laboratory, Electrical and Computer Engineering,
The Johns Hopkins University
{snehashisr,aaron_carass,prince}@jhu.edu

ABSTRACT

Magnetic resonance (MR) imaging (MRI) is widely used to study the structure of human brains. Unlike computed tomography (CT), MR image intensities do not have a tissue specific interpretation. Thus images of the same subject obtained with either the same imaging sequence on different scanners or with differing parameters have widely varying intensity scales. This inconsistency introduces errors in segmentation, and other image processing tasks, thus necessitating image intensity standardization. Compared to previous intensity normalization methods using histogram transformations–which try to find a global one-to-one intensity mapping based on histograms–we propose a patch based generative model for intensity normalization between images acquired under different scanners or different pulse sequence parameters. Our method outperforms histogram based methods when normalizing phantoms simulated with various parameters. Additionally, experiments on real data, acquired under a variety of scanners and acquisition parameters, have more consistent segmentations after our normalization.

***Index Terms*—** MRI, intensity normalization, intensity standardization, brain, segmentation.

1. INTRODUCTION

MRI is a non-invasive modality that has been widely used to gain understanding of the human brain. MR image processing techniques, such as segmentation, are needed to understand normal aging as well as the progression of diseases. For almost all MR image processing algorithms, image intensities are used as a primary feature. Unlike CT, MR intensities do not possess any tissue specific meaning. Thus a subject scanned in different scanners or with different pulse sequence parameters, will have variable image contrast between tissues (see Fig. 1 for example). It has been shown that the variability in tissue contrast gives rise to inconsistencies in segmentation [1].

This work was supported by the NIH/NIBIB under grant 1R21EB012765. We would like to thank Dr. Sarah Ying for providing the MR data. Some data used for this study were downloaded from the Biomedical Informatics Research Network (BIRN) Data Repository, supported by grants to the BIRN Coordinating Center (U24-RR019701), Function BIRN (U24-RR021992), Morphometry BIRN (U24-RR021382), and Mouse BIRN (U24-RR021760) Testbeds, funded by the National Center for Research Resources at the National Institutes of Health, U.S.A.

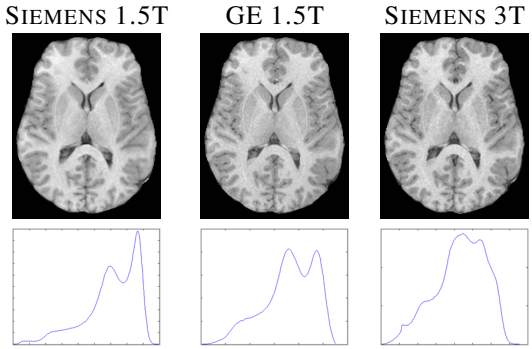

Fig. 1. Top row shows a subject's SPGR (spoiled gradient recalled) from three different scanners [2], while the bottom row shows their histograms–scale is arbitrary.

Several intensity standardization algorithms have been proposed to bring MR intensities to a common scale. To normalize a subject T_1-weighted scan to another reference T_1-w scan, histogram matching algorithms try to match the subject histogram to the reference histogram by matching the intensities of manually or automatically chosen landmarks [3] or by minimizing some information-theoretic criteria [4] between the histograms. Another method uses intrinsic MR properties, such as longitudinal and transverse relaxation times (T_1 and T_2) or proton density (P_D), to normalize the subject intensities to a reference volume, acquired with a different pulse sequence [5]. A segmentation based technique has also been proposed to account for the scanner variations, by matching intensity distributions of similar tissues between two images [6]. Landmark based methods usually suffer from the fact that reliable landmarks are sometimes difficult to find as they are either manually chosen or based on a segmentation itself. While, using tissue properties can be difficult as one needs precise knowledge of the imaging equations and pulse sequence parameters, which are often not known or difficult to estimate accurately.

We have previously reported a patch-based intensity normalization method called MIMECS [7] to overcome these problems. In this paper, a generative model based intensity normalization technique is developed independent of but having strong parallels with the idea of coherent point drift [8]. We still use patches instead of intensities of a single voxel as features, because a patch around a voxel contains neighbor-

hood information of that center voxel. We propose a patch matching technique that takes patches from the subject image and finds its best matching patches in a target image. We benefit from not having to choose landmarks and no pulse sequence parameters are required.

2. METHOD

We want to normalize the intensities of a normal subject $\mathcal{S}$ to a normal atlas $\mathcal{A}$. We assume that $\mathcal{S}$ and $\mathcal{A}$ are acquired with similar imaging sequences, e.g. either T_1-w SPGR or MPRAGE (magnetization prepared rapid gradient echo), but differ in either or both the scanner or the imaging parameters (e.g., repetition time, echo time, or flip angle). At each voxel of an image, 3D patches are considered and stacked into 1D vectors of size $d \times 1$. There are N subject patches $\mathbf{x}_i$ and M atlas patches $\mathbf{y}_k$, where N and M are the number of non-zero image voxels in $\mathcal{S}$ and $\mathcal{A}$, respectively. Define $\mathbf{X} = \{\mathbf{x}_i\}$ and $\mathbf{Y} = \{\mathbf{y}_j\}$ to be the collection of subject and atlas patches. To match the subject intensities to the atlas, we match $\mathbf{X}$ to $\mathbf{Y}$. First, we make sure that the peak white matter (WM) intensities of both $\mathcal{S}$ and $\mathcal{A}$ are the same (which provides a rough normalization of the two data sets). We assume that each subject patch $\mathbf{x}_i$ is a realization of a Gaussian random vector whose mean is one of the atlas patches—i.e., $\mathbf{x}_i \sim \mathcal{N}(\mathbf{y}_j, \Sigma_j)$ for some j. Then both the best matching atlas patch for each subject patch and the covariances of all atlas patches are defined by maximum likelihood and found using the expectation maximization algorithm (EM) [9], as described in detail below. The normalized image is produced by replacing the center pixel of each observed patch by the corresponding value of the matching atlas patch.

To motivate the key assumption of the Gaussian perturbation, we first observe that the image intensities are related to the intrinsic MR tissue properties, T_1, T_2, $T_2^\star$, and P_D, by nonlinear imaging equations. However, we can use a linear approximation to write the intensities of a patch as $\mathbf{x}_i = H_\mathcal{S}\mathbf{\Theta}_\mathcal{S}(i)$ and $\mathbf{y}_j = H_\mathcal{A}\mathbf{\Theta}_\mathcal{A}(j)$, where $H_\mathcal{S}$ and $H_\mathcal{A}$ are linear approximations of the imaging equations and $\mathbf{\Theta}_\mathcal{S}(i)$ and $\mathbf{\Theta}_\mathcal{A}(j)$ are the underlying MR tissue properties of the patches. If patches $\mathbf{x}_i$ and $\mathbf{y}_j$ are from the same tissue, we can assume that their MR tissue properties follow a Gaussian distribution, $\mathbf{\Theta}_\mathcal{A}(j), \mathbf{\Theta}_\mathcal{S}(i) \sim \mathcal{N}(\boldsymbol{\mu}, \Omega)$. Then, $\mathbf{x}_i \sim \mathcal{N}(H_\mathcal{S}\boldsymbol{\mu}, H_\mathcal{S}\Omega H_\mathcal{S}^T)$ and $\mathbf{y}_j \sim \mathcal{N}(H_\mathcal{A}\boldsymbol{\mu}, H_\mathcal{A}\Omega H_\mathcal{A}^T)$. Similarly, $\mathbf{x}_i - \mathbf{y}_j \sim \mathcal{N}((H_\mathcal{S} - H_\mathcal{A})\boldsymbol{\mu}, (H_\mathcal{S} + H_\mathcal{A})\Omega(H_\mathcal{S} + H_\mathcal{A})^T)$. We note that it is not feasible to estimate $H_\mathcal{A}$ or $H_\mathcal{S}$ because neither the imaging equations nor the pulse sequence parameters are always exactly known. But we assume $\mathcal{S}$ and $\mathcal{A}$ are acquired with "similar" pulse sequences, so we can thus make the assumption that $H_\mathcal{A} \approx H_\mathcal{S}$, implying $\mathbf{x}_i - \mathbf{y}_j \sim \mathcal{N}(\mathbf{0}, \Omega')$. Here, Ω' is a tissue-specific constant dependent on the nature of $\mathbf{\Theta}_\mathcal{A}(j)$ and $\mathbf{\Theta}_\mathcal{S}(i)$, or in essence, the classification of $\mathbf{x}_i$ and $\mathbf{y}_j$. However, it is very difficult to estimate intrinsic tissue MR properties (μ and Ω') since the exact nature of the imaging equations ($H_\mathcal{A}$ or $H_\mathcal{S}$) is usually not known. We avoid the classification and estimation problem of tissue MR properties by assuming that Ω' is a characteristic of $\mathbf{y}_j$, thereby creating an M-class Gaussian mixture model with means as the M atlas patches. We note that the same analysis can be extended even if a patch contains more than one tissue class. This idea of an M-class problem was explored previously for a registration algorithm [8].

To find the correspondence between patches, let z_{ij} be an indicator function having the value one when the ith subject patch $\mathbf{x}_i$ originates from a Gaussian distribution having its mean as the jth atlas patch $\mathbf{y}_j$ with co-variance matrix Σ_j, and is zero otherwise. The probability of observing $\mathbf{x}_i$ is written as

$$P(\mathbf{x}_i|z_{ij} = 1, \Sigma_j, \mathbf{y}_j) = \frac{1}{\sqrt{2\pi}|\Sigma_j|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_i - \mathbf{y}_j)^T\Sigma_j(\mathbf{x}_i - \mathbf{y}_j)\right\}. \quad (1)$$

For computational simplicity, we make the assumption that Σ_j is a diagonal matrix–i.e. $\Sigma_j = \sigma_j^2 I$, where I is a $d \times d$ identity matrix. Assuming the i.i.d. nature of the patches and a uniform prior probability $P(z_{ij} = 1|\Sigma_j, \mathbf{y}_j) = \frac{1}{M}$ and , the joint probability distribution of all the subject patches is given by

$$P(\mathbf{X}, \mathbf{Z}|\mathbf{Y}, \Theta) = C\prod_{j=1}^{M}\prod_{i=1}^{N}\left(\frac{1}{\sigma_j}\exp\left\{-\frac{||\mathbf{x}_i - \mathbf{y}_j||^2}{2\sigma_j^2}\right\}\right)^{z_{ij}} \quad (2)$$

where $\mathbf{Z} = \{z_{ij}\}$, $\Theta = \{\sigma_j : j = 1, \ldots, M\}$, and C is a normalizing constant. An estimate of σ_j's and posterior probabilities $P(z_{ij} = 1|\mathbf{X}, \mathbf{Y}, \Theta)$ can be obtained by maximizing the joint probability using EM.

EM is a two-step iterative process, that estimates the true z_{ij}'s based on the current estimates of the parameters Θ, and then updates Θ based on the estimates of z_{ij}. This is described as:

1. **E**-step: to find new update $\Theta^{(m+1)}$ at the mth iteration, compute the expectation $Q(\Theta^{(m+1)}|\Theta^{(m)}) = E[\log P(\mathbf{Z}|\mathbf{X}, \mathbf{Y}, \Theta^{(m+1)})|\mathbf{X}, \mathbf{Y}, \Theta^{(m)}]$.
2. **M**-step: find new estimates $\Theta^{(m+1)}$ based on the previous estimates by solving, $\Theta^{(m+1)} = \arg\max_{\Theta^{(m+1)}} Q(\Theta^{(m+1)}|\Theta^{(m)})$.

Assuming $w_{ij} = E(z_{ij}|\mathbf{x}_i, \mathbf{y}_j, \sigma_j)$, the **E**-step gives

$$\begin{aligned} w_{ij} &= P(z_{ij} = 1|\mathbf{x}_i, \mathbf{y}_j, \sigma_j), \\ &= \frac{P(\mathbf{x}_i|z_{ij} = 1, \mathbf{y}_j, \sigma_j)P(z_{ij} = 1|\mathbf{y}_j, \sigma_j)}{\sum_k P(\mathbf{x}_i|z_{ik} = 1, \mathbf{y}_k, \sigma_k)P(z_{ij}|\mathbf{y}_j, \sigma_j)}, \\ &= \frac{\frac{1}{\sigma_j}\phi\left(\frac{\mathbf{x}_i - \mathbf{y}_j}{\sigma_j}\right)}{\sum_{k=1}^{M}\frac{1}{\sigma_k}\phi\left(\frac{\mathbf{x}_i - \mathbf{y}_k}{\sigma_k}\right)}, \ \phi(\mathbf{t}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}||\mathbf{t}||^2}. \end{aligned} \quad (3)$$

The **M**-step gives the update equations for σ_js as

$$\sigma_j^2 = \frac{\sum_{i=1}^{N} w_{ij}||\mathbf{x}_i - \mathbf{y}_j||^2}{\sum_{i=1}^{N} w_{ij}}. \quad (4)$$

The algorithm is deemed to have converged when the maximum difference between the posteriors in successive itera-

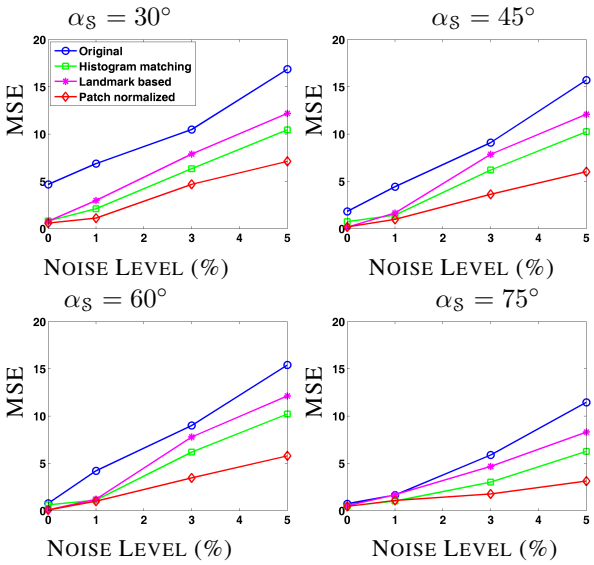

Fig. 2. Mean squared errors (MSE) between atlas ($\alpha_{\mathcal{A}} = 90°$) and subject images ($\alpha_{\mathcal{S}} = 30°, 45°, 60°, 75°$) are shown for various noise levels.

tions is below a threshold. Then the central voxel of the subject patch $\mathbf{x}_i$ is replaced with the central voxel of $\mathbf{y}_t$, where $t = \arg\max_j(w_{ij})$.

For all of the experiments, we use $3 \times 3 \times 3$ patches, thus $d = 27$. For a $256 \times 256 \times 199$ MR brain image of resolution 1mm^3, M and N are typically $\approx 10^7$. It is therefore memory and time intensive to compute w_{ij}'s for every $\mathbf{x}_i$'s. We observed experimentally that w_{ij} is very close to zero if $||\mathbf{x}_i - \mathbf{y}_j||$ is large. Following the assumption that $\mathbf{\Theta}_{\mathcal{S}}(i)$ and $\mathbf{\Theta}_{\mathcal{A}}(j)$ should be the same tissue, we assume $w_{ij} = 0$ if $||\mathbf{x}_i - \mathbf{y}_j|| > \delta$, which is similar to a non-local criteria [10]. However, in our computation, for each i, we choose its closest D atlas patches to have non-zero w_{ij}, their indices denoted by Ψ_i, a set of cardinality D. Thus if any atlas index j is not in Ψ_i, we fix $w_{ij} = 0$. Eqn. 4 is modified as follows

$$\sigma_j^2 = \frac{\sum_{i;j\in\Psi_i} w_{ij}||\mathbf{x}_i - \mathbf{y}_j||^2}{\sum_{i;j\in\Psi_i} w_{ij}}. \quad (5)$$

3. RESULTS

3.1. Phantom validation

We first validate the algorithm on the Brainweb phantom data [11]. We use flip angle α as the varying imaging parameter of SPGR scans of a normal phantom. Phantoms with $\alpha_{\mathcal{S}} = 30°, 45°, 60°, 75°$ with noise levels $n = 0, 1, 3, 5\%$ are used to normalize to a phantom with $\alpha_{\mathcal{A}} = 90°$ and 0% noise. Our method is compared with histogram matching and a landmark based method [3] where the landmarks are found using a Gaussian mixture model algorithm. Fig. 2 shows the mean squared errors (MSE) between the atlas and the subjects before and after normalization with these three methods. Clearly the patch based method outperforms the other two for all values of $\alpha_{\mathcal{S}}$, and $n = 1, 3, 5$. For 0% noise, all three

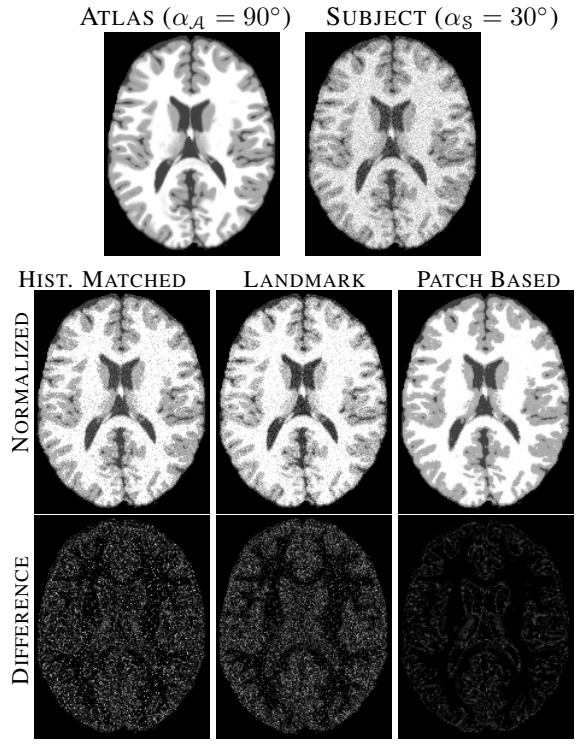

Fig. 3. Top row shows the atlas ($\alpha_{\mathcal{A}} = 90°$) and the subject ($\alpha_{\mathcal{S}} = 30°$). Middle row shows the histogram matched, landmark based normalized and patch based normalized images. Bottom row shows their differences with the atlas.

methods perform similarly, because the lack of any partial volume or noise makes the choice of landmarks accurate and histogram matching perfect. At higher noise levels, histogram matching becomes dependent on the number of bins and the estimation of landmarks becomes less robust. The atlas and a subject with $\alpha_{\mathcal{S}} = 30°$ at 5% noise are shown in Fig. 3 the top row, the normalized images and their differences from the atlas are shown in the middle and bottom row, respectively. Clearly, both histogram matching and landmark based normalization leaves a bias in intensities in large regions, e.g., gray matter (GM), while our patch-based method fails only in recovering sharp edges such as at GM to WM transitions.

3.2. Experiments on real data

We next used five healthy subjects from the morphometry BIRN [2] calibration study to show the effects of normalization. Each subject has 14 SPGR scans, 7 each with $\alpha = 20°$ and $30°$ from 4 different scanners. Each scan of a subject was first rigidly registered to its GE 1.5T $\alpha = 30°$ scan and then skull-stripped [12] using the same mask generated from the GE 1.5T $\alpha = 30°$ scan. Then each image is corrected for any intensity inhomogeneity using N3 [13]. For every subject, we normalize every $\alpha = 20°$ scan to the corresponding $\alpha = 30°$ scan. Fig. 4 shows one set of $\alpha = 20°$ and $\alpha = 30°$ scans of a subject along with the normalized image in the top row.

Our method does not seek to match histograms, however it is useful to compare them after normalization. Fig. 4 bottom

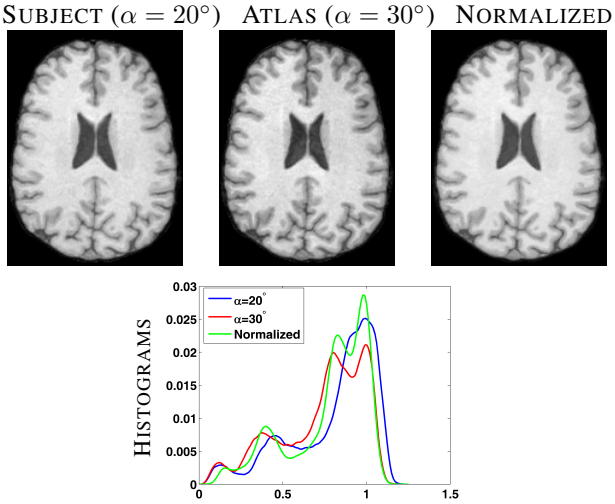

Fig. 4. Top row shows the subject ($\alpha = 20°$), atlas ($\alpha = 30°$) and the corresponding normalized image. Bottom row shows a plot of the histograms of the above three images.

row, shows that the histogram of the normalized image more strongly resembles the atlas than the original subject image.

As described earlier, segmentations can become inconsistent with the variation of scanners and imaging parameters [1, 5, 6]. We demonstrate that the patch based normalization achieves more consistent segmentation between images acquired with different scanners and pulse sequence parameters. We compare our method with histogram matching and a landmark based intensity transformation [3] where three landmarks on the subject and the atlas are automatically chosen based on a 3-class Gaussian mixture model segmentation. After normalization, we use fuzzy C-means [14] to generate segmentations. We report Dice similarity coefficients in Table 1 between segmentations of the three primary tissues, cerebro-spinal fluid (CSF), GM, and WM, and their weighted average. The values are averaged over five subjects, each with seven scans. Patch based normalization produces significantly higher Dice coefficients for GM and WM compared to the other two, while the CSF Dice segmentations are comparable. Histogram matching often decreases the similarity with the original, as it is heavily dependent on the number of bins. Landmark based matching can also deteriorate the segmentation consistency, when a suitable landmark is not found. This is often the case if the image histogram is not multi-modal (e.g., the blue line in Fig. 4). Both Figs. 3 and 4 reveal a smoothing effect on the reconstructed images. Though not unexpected in patch-based methods, this effect needs further investigation.

4. DISCUSSION AND CONCLUSION

We have described a patch based MR intensity normalization framework that can normalize scans having similar acquisition protocol but acquired on different scanners or with different acquisition parameters. We validated our algorithm on phantoms and real scans and shown it gives rise to more consistent segmentations after normalization.

Table 1. Average Dice coefficients of hard segmentations are obtained from 35 scans, before and after normalization, comparing our method with histogram matching and a landmark based matching [3].

	CSF	GM	WM	Mean
Original	0.882	0.788	0.927	0.884
Hist. Match	0.881	0.828	0.917	0.891
Landmark based	0.876	0.776	0.926	0.882
Patch based	0.874	0.845⋆	0.943⋆	0.902

⋆ Statistically significantly larger than the other three (p-value < 0.05).

5. REFERENCES

[1] K. A. Clark, R. P. Woods, D. A. Rottenber, A. W. Toga, and J. C. Mazziotta, "Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images," *NeuroImage*, vol. 29, pp. 185–202, 2006.

[2] L. Friedman *et. al.*, "Test-Retest and Between-Site Reliability in a Multicenter fMRI Study," *Human Brain Mapping*, vol. 29, no. 8, pp. 958–972, August 2008.

[3] L. G. Nyul and J. K. Udupa, "On Standardizing the MR Image Intensity Scale," *Mag. Res. in Medicine*, vol. 42, no. 6, pp. 1072–1081, 1999.

[4] F. Jager and J. Hornegger, "Nonrigid Registration of Joint Histograms for Intensity Standardization in Magnetic Resonance Imaging," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 137–150, 2009.

[5] B. Fischl *et. al.*, "Sequence-independent segmentation of magnetic resonance images," *NeuroImage*, vol. 23, pp. 69–84, 2004.

[6] X. Han and B. Fischl, "Atlas Renormalization for Improved Brain MR Image Segmentation Across Scanner Platforms," *IEEE Trans. Med. Imag.*, vol. 26, no. 4, pp. 479–486, 2007.

[7] S. Roy, A. Carass, and J. L. Prince, "A Compressed Sensing Approach For MR Tissue Contrast Synthesis," in *22nd Conf. on Info. Proc. in Medical Imag. (IPMI 2011)*, 2011, pp. 371–383.

[8] A. Myronenko and X. Song, "Point-Set Registration: Coherent Point Drift," *IEEE Trans. Patt. Anal. and Machine Intell.*, vol. 32, no. 12, pp. 2262–2275, 2010.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Stat. Soc. Series B (Method.)*, vol. 39, no. 1, pp. 1–38, 1977.

[10] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *IEEE Intl. Conf. on Comp. Vision*, 2009, pp. 2272–2279.

[11] C.A. Cocosco, V. Kollokian, R.K.S. Kwan, and A.C. Evans, "BrainWeb: Online Interface to a 3D MRI Simulated Brain Database," *NeuroImage*, vol. 5, no. 4, pp. S425, 1997.

[12] A. Carass, J. Cuzzocreo, M. B. Wheeler, P. L. Bazin, S. M. Resnick, and J. L. Prince, "Simple paradigm for extra-cerebral tissue removal: Algorithm and analysis," *NeuroImage*, vol. 56, no. 4, pp. 1982–1992, 2011.

[13] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A non-parametric method for automatic correction of intensity non-uniformity in mri data," *IEEE Trans. Med. Imag.*, vol. 17, no. 17, pp. 87–97, 1998.

[14] J. C. Bezdek, "A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 20, no. 1, pp. 1–8, 1980.