# A Continuous STAPLE for Scalar, Vector and Tensor Images: An Application to DTI Analysis

Olivier Commowick\*, Simon K. Warfield\*, Senior Member IEEE

\* Computational Radiology Laboratory, Department of Radiology, Children's Hospital, 300 Longwood Avenue, Boston, MA, 02115, USA *E-mail: {Olivier.Commowick, Simon.Warfield}@childrens.harvard.edu* 

Abstract—The comparison of images of a patient to a reference standard may enable the identification of structural brain changes. These comparisons may involve the use of vector or tensor images (i.e. 3D images for which each voxel can be represented as an  $\mathbb{R}^N$  vector) such as Diffusion Tensor Images (DTI) or transformations. The recent introduction of the Log-Euclidean framework for diffeomorphisms and tensors has greatly simplified the use of these images by allowing all the computations to be performed on a vector-space. However, many sources can result in a bias in the images, including disease or imaging artifacts.

In order to estimate and compensate for these sources of variability, we developed a new algorithm, called continuous STAPLE, that estimates the reference standard underlying a set of vector images. This method, based on an Expectation-Maximization method similar in principle to the validation method STAPLE, also estimates for each image a set of parameters characterizing their bias and variance with respect to the reference standard.

We demonstrate how to use these parameters for the detection of atypical images or outliers in the population under study. We identified significant differences between the tensors of diffusion images of multiple sclerosis patients and those of control subjects in the vicinity of lesions.

*Index Terms*—Atlas, STAPLE, ground truth, Expectation-Maximization, DTI, transformations, Kullback-Leibler divergence, validation.

# I. INTRODUCTION

Databases of images are widely used to detect abnormalities in patient images or specific characteristics of a population, either by computing average tensor derived parameters on ROIs delineated on each patient [1], or based on their statistical difference with respect to a reference frame [2], [3], [4], [5]. Directly processing vector images (i.e. 3D images for which each voxel is a vector in  $\mathbb{R}^N$  such as diffusion tensor images (DTI), transformations or Jacobian matrix data) however requires complex computation schemes to remain on the corresponding manifold (see [6] for tensors). Recently, the Log-Euclidean framework was introduced both for tensors [7] and diffeomorphisms [8], greatly simplifying the computations by allowing all of them to be performed on a vector-space. We are therefore particularly interested in the study of these vector images.

Early studies explored ways to summarize the information of each voxel vector into a scalar. This scalar information may be, for instance, Fractional Anisotropy (FA) [9] or Mean Diffusivity (MD) when processing tensor images. The Jacobian matrix determinant is also often used to gather the information when studying transformations [10], [11]. However, the use of the full tensor rather than scalar parameters such as the MD or the FA has shown to be more effective in many cases [12], [5], [3]. The full Jacobian matrix was also used in the computation of statistics on deformations [13]. Using the Log-Euclidean framework, Commowick et al. [12] showed the superiority of using the whole Jacobian tensor as opposed to the determinant to constrain non linear registration. Many studies have then used the complete information from either the DTI or transformations to compute their statistics. For example, Fillard et al. [14] studied the brain variability and Lepore et al. [5] presented statistical tests on the Jacobian tensors of deformations in the frame of an HIV/AIDS study. Finally, a growing literature emphasizes the use of the full tensor for voxel-wise comparison of patient groups with respect to controls. Whitcher et al. [3] introduced recently multivariate hypothesis testing for statistical group comparison using the Log-Euclidean framework. Verma et al. [15] and Khurd et al. [16] further underlined the value, when computing such comparisons, to take into account the specific subspace formed by the observed tensors at each voxel. They therefore proposed to capture the manifold of variation of the considered tensor data, using either Isomap manifold learning [15] or kernelbased approaches [16].

1

Several factors can introduce bias and variance in the analysis of these images. First, the study generally requires non linear registration either to register the images on a reference frame or to build the reference frame from a dataset of control subjects. However, differences in acquisition protocols or anatomies may lead to registration discrepancies, therefore resulting in a bias on the transformed DTI or the Jacobian tensors. Other sources of bias and variance are the artifacts due to the acquisition itself, for example movement artifacts or DTI distortion. Applications of population studies involving vector images are becoming increasingly common. Disease state may be recognizable as an atypical increase in bias or variance. Comparisons of individuals to population reference standard databases would be facilitated by the ability to identify and compensate for such sources of bias and variance.

Copyright (c) 2008 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The simultaneous estimation of a reference standard and performance parameters was proposed in the context of segmentation validation. STAPLE [17] uses an EM algorithm [18], [19] to estimate iteratively the reference standard underlying the expert segmentations and parameters for each of the experts. Therefore, experts who do not concur with the prior knowledge or the general consensus (for example images not well registered in an atlas construction process) have a smaller influence. Moreover, the parameters obtained for each expert can be used to characterize his agreement (sensitivity and specificity) with respect to the consensus. This method was extended [20] in order to estimate the reference standard together with bias and variance parameters for expert segmentations. However, this method was developed to use a set of scalar images. It is therefore not adapted to population studies involving images with higher dimensional voxels such as DTI, Jacobian tensor images or diffeomorphisms. We therefore present in this article a new method, called continuous STAPLE, that estimates the reference standard and image parameters underlying a set of vector images using a similar EM approach. We also propose a parameter comparison framework in order to compute hypothesis tests to detect statistically significant differences in the parameters. We present below an application of this algorithm to the study of DTI images of patients with multiple sclerosis and their comparison to control subjects.

# II. METHOD

# A. Description of Continuous STAPLE

We observed a set of I vector images  $S_i$ , i.e. each voxel j of  $S_i(s_{ij})$  is a vector of  $\mathbb{R}^N$ . All these images are assumed to be an independent evaluation of a true vector image: the reference standard. Our goal is then to estimate both this unknown reference standard as well as performance parameters for each image i in order to compare them. If the reference standard was known, then estimating the performance parameters for each image would be straightforward. However, as this reference standard is unknown, we will use an Expectation-Maximization approach [18], [19] in order to estimate the true vector image and the performance parameters of the images. The EM algorithm proceeds iteratively, alternating two steps:

- E-Step: Estimate the expected value of the complete data log-likelihood and the reference standard T from the known image parameters and their decisions  $s_{ij}$ . See Section II-B.
- M-Step: Estimate the performance parameters of each image, knowing the current estimate of the reference standard. See Section II-C.

As mentioned previously, we assume there are I images, each estimating an image consisting of voxels whose dimension is N. For example, if N = 1, this type of image can be the segmentations obtained by distance transforms or level-set approach. In higher dimensions, these could be transformation vectors (N = 3) [8] or tensors in the Log-Euclidean framework (N = 6) [7]. We associate to each image i an  $\mathbb{R}^N$  vector  $\beta_i$  (bias parameter) as well as an  $N \times N$  covariance matrix  $\Lambda_i$ . In other words, we associate each image *i* with a Gaussian probability density function describing  $P(s_{ij}|T_j, \beta_i, \Lambda_i)$ . In our case, we choose to use a multivariate normal distribution:

$$P(s_{ij}|T_j,\beta_i,\Lambda_i) = \phi_{T_j+\beta_i,\Lambda_i}(s_{ij}) = \phi_{\mu_{ij},\Lambda_i}(T_j) \quad (1)$$

where  $\mu_{ij} = s_{ij} - \beta_i$  and  $\phi_{\mu,\Lambda}(x)$  is the probability density function of the normal distribution  $N(\mu, \Lambda)$ . The parameters of the images are grouped in a variable  $\theta = \{\beta_1, \Lambda_1, ..., \beta_I, \Lambda_I\}$ for notation simplicity. As mentioned earlier, as the reference standard is unknown, an Expectation-Maximization algorithm is used to alternately estimate the parameters  $\theta$  and the hidden reference standard. We will now describe in detail the Expectation and Maximization steps of our method.

# B. E-Step: Estimation of the Reference Standard from the Images

The goal of the Expectation step is to compute the expected value of the complete data log-likelihood  $Q(\theta|\theta^{(k)})$  knowing the image parameters at the preceding iteration  $\theta^{(k)}$ . Evaluating this expression requires the knowledge of the posterior probability of the true score  $T: P(T|S, \theta^{(k)})$ . In our case, the computation of this posterior probability is sufficient to perform the Maximization step and it is therefore not necessary to explicitly form  $Q(\theta|\theta^{(k)})$ .

It is assumed that there is no spatial dependency between voxels, i.e.  $P(T|S, \theta^{(k)}) = \prod_j P(T_j|S, \theta^{(k)})$ . Using Bayes' theorem and noting the true score is independent of the rater bias and covariances,  $P(T_j|S, \theta^{(k)})$  can then be expressed as follows:

$$P(T_j|S,\theta^{(k)}) = P(S|T_j,\theta^{(k)}) \frac{P(T_j)}{P(S|\theta^{(k)})}$$
(2)

where  $P(T_j)$  is the prior distribution of  $T_j$ . Since the marginal distribution of  $T_j$ ,  $P(T_j|S, \theta^{(k)})$  integrates to 1, it can be rewritten as:

$$P(T_j|S,\theta^{(k)}) = \frac{P(T_j)P(S|T_j,\theta^{(k)})}{\int_{T_i} P(T_j)P(S|T_j,\theta^{(k)})}$$
(3)

A large variety of distributions can be used for the prior distribution  $P(T_j)$ . In our case, we have chosen to use a multivariate uniform distribution, with a scale parameter  $h = \prod_{a=1}^{N} h_a$ ,

$$P(T_j) = U_h(T_j) = \frac{1}{h} \tag{4}$$

Equation (3) therefore simplifies to:

$$P(T_j|S,\theta^{(k)}) = \frac{1}{Z_j} P(S|T_j,\theta^{(k)})$$
(5)

where  $Z_j = \int_{T_j} \prod_i \phi_{\mu_{ij}^{(k)}, \Lambda_i^{(k)}}(T_j)$  is a normalizing constant. We also know that, in the absence of spatial correlations between voxels,  $P(S|T_j, \theta^{(k)})$  is a product of multivariate Gaussians  $N(\mu_{ij}^{(k)}, \Lambda_i^{(k)})$ , defined in Equation (1), where  $\mu_{ij}^{(k)} = s_{ij} - \beta_i^{(k)}$ . We therefore obtain the posterior probability of  $T_j$ :

$$P(T_j|\theta^{(k)}) = \frac{1}{Z_j} \prod_i \phi_{\mu_{ij}^{(k)}, \Lambda_i^{(k)}}(T_j)$$
  
=  $\frac{1}{Z_j} \left( \prod_i (2\pi)^{-\frac{N}{2}} |\Lambda_i^{(k)}|^{-\frac{1}{2}} \right) \exp(w_j)$  (6)

This probability is then a product over the images i of multivariate Gaussians and,

$$w_j = -\frac{1}{2} \sum_{i} (\mu_{ij}^{(k)} - T_j)^T (\Lambda_i^{(k)})^{-1} (\mu_{ij}^{(k)} - T_j)$$
(7)

Then, completing the square over this expression yields a multivariate Gaussian probability density function, i.e.  $T_j \sim N(\mu_i^{(k+1)}, \Lambda^{(k+1)})$  with the following parameters:

$$(\Lambda^{(k+1)})^{-1} = \sum_{i} (\Lambda_i^{(k)})^{-1}$$
(8)

$$\mu_j^{(k+1)} = \Lambda^{(k+1)} \left( \sum_i (\Lambda_i^{(k)})^{-1} \mu_{ij}^{(k)} \right)$$
(9)

The new estimate of the covariance of the distribution  $(\Lambda^{(k+1)})$  can therefore be seen as an harmonic mean of the individual covariance matrices  $\Lambda_i^{(k)}$ . The new estimate of the reference value at a voxel j  $(\mu_j^{(k+1)})$  can also be seen as an inverse rater covariance weighted sum of the bias corrected rater values.

## C. M-Step: Optimization of the Image Parameters

Given the new estimate of the posterior probability  $P(T|S, \theta^{(k)})$ , the expected value of the complete data loglikelihood can be computed. The parameter estimates can then be computed in the M-step by maximizing  $Q(\theta|\theta^{(k)})$  with respect to the parameters  $\theta$ , i.e.

$$\theta^{(k+1)} = \arg\max_{\theta} Q(\theta|\theta^{(k)})$$
  
=  $\arg\max_{\theta} \sum_{i} \sum_{j} \mathbb{E} \left( \log(P(s_{ij}|T_j, \theta)) \right)$  (10)

Similar to [20],  $\mathbb{E}(\log(P(s_{ij}|T_j, \theta)))$  can then be simplified:

$$\mathbb{E} \left( \log(P(s_{ij}|T_j, \theta)) \right) = \frac{1}{2} \log(|\Lambda_i^{-1}|) \\
- \frac{1}{2} (s_{ij} - \beta_i)^T \Lambda_i^{-1} (s_{ij} - \beta_i) \\
+ (s_{ij} - \beta_i)^T \Lambda_i^{-1} \mu_j^{(k+1)} - \frac{1}{2} (tr(\Lambda_i^{-1} \Lambda^{(k+1)}) \\
+ (\mu_j^{(k+1)})^T \Lambda_i^{-1} \mu_j^{(k+1)})$$
(11)

On computing the derivative of Equation (10) with respect to the parameters  $\theta$ , and then solving the system for a maximum, the following expression for the bias parameters  $\beta_i^{(k+1)}$  is found:

$$\beta_i = \frac{1}{J} \sum_j (s_{ij} - \mu_j^{(k+1)})$$
(12)

where J corresponds to the number of voxels. Then, knowing  $\beta_i^{(k+1)}$ , assuming that  $\Lambda_i^{(k+1)}$  is invertible and using the expression from [21] that  $\frac{\partial |\Lambda_i^{-1}|}{\partial (\Lambda_i^{-1})_{ml}} = |\Lambda_i^{-1}| (\Lambda_i)_{ml}, \Lambda_i^{(k+1)}$  is given by the following:

$$\Lambda_i^{(k+1)} = (\Lambda^{(k+1)})^\top + \frac{1}{J} \sum_j \gamma_{ij} \gamma_{ij}^\top$$
(13)

where  $\gamma_{ij} = \mu_j^{(k+1)} + \beta_i^{(k+1)} - s_{ij}$ . As the first estimates for the  $\Lambda_i^{(k)}$  are given by the user and symmetric, an interesting result seen in these last equations is that the  $\Lambda_i^{(k)}$  are guaranteed to remain symmetric over the successive E-steps and M-steps iterations. When the dimension N of the input images is 1, i.e. scalar images are used, these expressions of the E and M-steps also simplify to the expressions given in [20].

#### D. Framework for the Comparison of Image Parameters

We have presented above a method to compute both a reference standard from a dataset of vector images and parameters describing the characteristics of each image with respect to the underlying reference. These parameters can now be used for the comparison of the images ; for example, to detect differences between them. In this section, we introduce a comparison framework for the image parameters derived from the continuous STAPLE algorithm.

1) Overview: We assume here that all the input images are in the same reference frame. This can be achieved, for example, by their registration on an atlas or the construction of an atlas from a population. Our comparison framework is then divided in the following steps:

- computation of bias and covariance parameters using continuous STAPLE,
- comparison of the distributions from the bias and covariance parameters using Kullback-Leibler Divergence.

We assume that a set of images  $\tilde{S}_i$  is available, all of them aligned in the same reference frame. We then can use our method to compute the reference standard underlying these images together with the bias  $\beta_i$  and covariance parameters  $\Lambda_i$  of each image *i*. Depending on the specific application, however, some pre-processing is required before employing our method. Tensors do not belong to a vector space but to a Riemannian manifold [6], requiring the use of an appropriate computation framework so that the computed tensors remain on the tensor manifold. The Log-Euclidean framework for tensors [7] has been proposed to this end, showing that performing all the Euclidean calculations on the logarithm of the tensors allows to remain on the tensor manifold. The same type of preprocessing is also required when manipulating diffeomorphisms, this time using the Log-Euclidean framework on diffeomorphisms [8].

2) Kullback-Leibler Divergence Statistics: Knowing the parameters for each image, we outline here a method to compare the Gaussian probability density functions defined by them. Each set of parameters,  $\beta_i$  and  $\Lambda_i$ , describes a multivariate Gaussian probability density function as expressed in Eq. (1). One possibility would be to directly compare the parameters

 $\beta_i$  and  $\Lambda_i$  but this would not reflect the different meaning of mean and covariance parameters.

We therefore utilize a framework based on the Kullback-Leibler Divergence (KLD) [22] to compare the probability distributions directly. First we compute the mean parameters  $\bar{\beta}$ and  $\bar{\Lambda}$ . We then form the distribution of the mean parameters  $N(\bar{\beta}, \bar{\Lambda})$ . Hence, we obtain the following expressions:

$$\bar{\beta} = \frac{1}{I} \sum_{i=1}^{I} \beta_i$$
(14)
$$\bar{\Lambda} = \frac{1}{I} \sum_{i=1}^{I} \left( \Lambda_i + (\bar{\beta} - \beta_i)(\bar{\beta} - \beta_i)^T \right)$$
(15)

where I is the number of images used in STAPLE. We can now compare the probability density functions described by each set of image parameters  $(\beta_i, \Lambda_i)$  with respect to the mean normal density. This is done using the KLD, which is expressed as follows when comparing two multivariate Gaussian distributions:

$$D(N_0 || N_1) = \frac{1}{2} \left( \log(\frac{|\Lambda_1|}{|\Lambda_0|}) + tr(\Lambda_1^{-1}\Lambda_0) + (\mu_1 - \mu_0)^T \Lambda_1^{-1}(\mu_1 - \mu_0) - N \right)$$
(16)

We then obtain a set of  $KL_i = D\left(N(\beta_i, \Lambda_i) || N(\bar{\beta}, \bar{\Lambda})\right)$ from the images. This allows us to look for significant differences using an outlier-identification method. Since the sample distribution of  $KL_i$  is approximately normalwhen the number of images is sufficiently large, we compute its mean  $\mu_{KL}$  and variance  $\Lambda_{KL}^2$ . Then, each Kullback-Leibler divergence value is compared to this univariate normal distribution. For each image *i*, we test the parameters  $KL_i$  to see if they are likely to be a sample from the normal distribution  $N(\mu_{KL}, \Lambda_{KL}^2)$ . This is done by computing the area outside the bell of the Gaussian:  $p_i = P(|X| > \frac{|KL_i - \mu_{KL}|}{\Lambda_{KL}})$ , where X follows a normal distribution N(0, 1). For a univariate normal distribution, this expression simplifies to the following:

$$p_i = 1 - \operatorname{erf}\left(\frac{1}{\sqrt{2}} \frac{|KL_i - \mu_{KL}|}{\Lambda_{KL}}\right) \tag{17}$$

A small  $p_i$  will therefore indicate a small probability for the parameters extracted from image  $\tilde{S}_i$  to agree with the other images. This  $p_i$  can also be viewed as the p-value of a z-test where the sample size of the tested population is 1. These  $p_i$  consequently allow outliers or atypical images to be identified.

# **III. RESULTS**

We present two applications of our work. First, we have created a database of simulated noisy DTI images, together with a group of outliers. We used this simulated database in order to evaluate the results of our method and compare them with the average DTI image obtained by the classical Log-Euclidean mean. We present here the evaluation, on clinical images, of local tensor differences in multiple sclerosis (MS) patients with respect to a database of control subjects.

## A. Simulated Experiments

1) Simulation Database: To evaluate our method with respect to a known ground truth and therefore be able to quantitatively evaluate our results, we created a database of control simulated DTI images. This database is illustrated in Fig. 1. These illustrations, as well as the others in this article, were obtained using the visualization interface MedINRIA [23]. The database is composed of 20 2D images  $(40 \times 40)$ and is divided into two groups. Each group was generated from the known ground truth (image (a) in Fig. 1), by adding a multivariate Gaussian noise with different parameters. The first group (example in image (b)) was generated by using a mean of 0.2 and a variance of 0.2 for each component of the log-tensors, while the parameters were a mean of -0.2and variance of 0.2 for the second group (example in image (c)). The additive noise was applied on the Log-Euclidean representation of the tensors as this allows us to stay on the tensor manifold.



Fig. 1. Simulated Database of Tensor Images. Illustration of the images used in our simulation: (a): Ground truth ; (b), (c): examples of noisy images from the first (bias: 0.2, variance 0.2) and second group (bias: -0.2, variance 0.2) of database samples ; (d): Outlier reference image ; (e), (f) : examples of noisy outliers (bias: 0.2, variance 0.2 ; and bias: -0.2, variance 0.2).

We then added to this database a group of 4 outliers created

by adding a multivariate Gaussian noise to a reference outlier image (illustrated in image (d)), in which each tensor has been rotated by an angle of  $-\pi/4$  about the z-axis (going through the image on Fig. 1). To further test the capabilities of our outlier detection method, we used two different sets of noise parameters (bias of 0.2 and variance of 0.2 for images 21 and 22; bias of -0.2 and variance of 0.2 for images 23 and 24) to generate the outlier images illustrated in images (e) and (f).

2) Evaluation of the Results: Using this database of simulated DTI tensors, we compared the results that were attained using a classical Log-Euclidean average [7] and the "ground truth" obtained using our continuous STAPLE. To do so, we supposed that the outlier images were not known and added them to the database as images 21 to 24. We then computed the Log-Euclidean average and the STAPLE ground truth estimation from these 24 patients. The results are illustrated and compared with the reference image in Fig. 2.



Fig. 2. **Comparison of STAPLE and Log-Euclidean Mean Results**. Results obtained from the whole simulated database (20 control images and 4 outliers) (a): Reference noiseless image (zoom in image (d)); (b): Log-Euclidean mean result (zoom in image (e)); (c): STAPLE ground truth (zoom in image (f)).

In this figure, we can see that the Log-Euclidean average (images (b) and (e)) is affected both by the noise in the database and by the presence of the 4 outliers. As a result, the mean tensors are rotated clockwise and sometimes appear swollen when compared to the reference tensors. The use of our method, however, provides a much more accurate result, with the tensors in images (c) and (f) being more similar to the reference. These qualitative results were confirmed quantitatively by computing the average Log-Euclidean distance between the computed averages and the reference image. The distance was 0.11 for the STAPLE average and 0.221 for the Log-Euclidean mean: the average error with our new method is better by a factor of 2.02.

We then evaluated the parameters obtained by our method and used our framework to try to detect the outlier images among the database. The parameters obtained for the two groups of control images were very close to those used to generate the images. For example, the mean of the bias parameters obtained for the first group is  $0.1922 \pm 0.0207$ , which includes the reference value of 0.2. For the second group, the mean bias found is  $-0.2079 \pm 0.0214$ , which again includes the reference value of -0.2.

Finally, to evaluate our Kullback-Leibler comparison framework, we computed the Gaussian scores for each image of the database with respect to the mean distribution of errors. We report the results in Table I. This table shows the capabilities of our framework to detect the outliers among a database. The Gaussian scores for the control images are large, ranging from 0.591 to 0.777, while the scores of the outliers (images 21 to 24) are much smaller, below the limit of 0.05 and ranging from 0.018 to 0.04, therefore identifying them as outliers. Moreover, our framework identified both groups of outliers, even though their noise parameters were different.

Subj. #	Score	Subj. #	Score	Subj. #	Score
1	0.619	9	0.694	17	0.777
2	0.614	10	0.663	18	0.634
3	0.714	11	0.654	19	0.624
4	0.601	12	0.732	20	0.697
5	0.607	13	0.652	21	0.018
6	0.644	14	0.768	22	0.028
7	0.648	15	0.716	23	0.040
8	0.591	16	0.621	24	0.035

#### TABLE I

PARAMETER EVALUATION RESULTS ON THE SIMULATED DATABASE. GAUSSIAN SCORES BETWEEN EACH IMAGE AND THE MEAN PARAMETERS OF THE DATABASE USING THE KLD-BASED EVALUATION FRAMEWORK (SEE EQ. (17)).

# B. Evaluation of Tensor Differences in Multiple Sclerosis Patients

We now present the evaluation of tensor differences in multiple sclerosis (MS) patients with respect to control subjects. As mentioned previously, our method requires a reference frame on which the images are registered. The DTI analysis therefore require two steps:

• construction of a mean DTI atlas from the control subjects that will be used as the reference frame for the study (see Section III-B2),

• registration of the MS patients onto the atlas and use of our framework to analyze the regions of lesions (see Section III-B3).

1) DTI Databases: We have used two databases of images in our experiments. The first database was composed of 3D T1 images (size  $256 \times 256 \times 175$  with a resolution of  $1 \times 1 \times 1.3 \text{ mm}^3$ , illustrated on Fig. 3) and associated DTI of the brain for 20 normal adult subjects (with an original size of  $256 \times 256 \times 60$  and resolution of  $1 \times 1 \times 2 \text{ mm}^3$ ).



Fig. 3. **Examples of Subjects from the Normal Database**. Axial slices showing four subjects T1 images of the normal subjects database registered using a global affine transformation on an atlas.

The second dataset was composed of 8 patients (illustrated on Fig. 4) exhibiting multiple sclerosis lesions in the brain. Both DTI and T1 images were acquired with the same size and resolution characteristics as for the first database. The FLAIR images were also added in this database and were used together with the T1 images to manually delineate the lesions in each 3D patient image.

Some preprocessing was performed on both databases. First, the skull of each brain was removed automatically for both databases using the BET algorithm [24]. We are only interested in the study of the brain and removing the skull allows for a better registration of the structures close to it.

Then, the DTI of each patient was registered on the corresponding T1 image. This was done in two steps: first a global affine transformation was found between the T1 image and the mean diffusivity image of the DTI using [25]. Then, a second step optimized a B-Splines transformation using FFD registration [26]. Finally, the tensors were moved and reoriented according to the transformations. It is usually assumed that the structure of the tensor should not change with the transformation [27]. Therefore, only a local rotation R should be applied to the tensor D:  $D' = RDR^T$ . For the non linear transformation, this local rotation was computed



Fig. 4. **Examples of the Multiple Sclerosis Database**. Axial slices of four patients T1 images, registered using a global affine transformation on an atlas.

using the polar decomposition of the Jacobian matrix  $J = \nabla T$  [27]. The polar decomposition theorem states that any nonsingular square matrix can be decomposed into a rigid rotation component R and a deformation component U: J = UR.

2) Mean DTI Atlas Construction: As mentioned in Section II-D, we needed to transfer all the DTI onto the same reference frame in order to use continuous STAPLE on these databases. Using any subject as a reference would introduce a bias due to the specific anatomy of the reference image. We therefore chose to build an unbiased atlas from the dataset of controls, and use it as the reference for our future evaluations. The construction of the DTI atlas is divided into two main steps:

- construction of an average T1 image from the T1 images,
- application of the obtained transformations to the DTI.

We have chosen to compute the atlas using a block-matching based non linear registration of the anatomical T1 images [28] as it was readily available in our laboratory. However, the proposed algorithm is independent of the choice of the registration algorithm. Therefore, recent tensor based registration algorithms taking into account the tensors specific information [29], [30], [31], [32] could also be used. This would result in a simplification of the atlas construction, requiring only to utilize the atlas construction framework proposed by [33] on the DTI images.

The construction of the average T1 image was done using an unbiased atlas construction method proposed by Guimond et al. [33]. This algorithm basically iterates over the following steps: at each iteration *i*, all the images  $I_k$  are non linearly registered on a reference image  $R_i$ , therefore finding transformations  $T_k$ , and bringing them on  $R_i$ . Then, a mean image  $M_i$  can be built from these registered images simply by averaging the intensities of the images. At the same time, the non linear transformations  $T_k$  are averaged and inverted using the Log-Euclidean framework for diffeomorphisms [8] to get a transformation  $\overline{T}^{-1}$ , i.e.  $\overline{T}^{-1} = \exp(-\frac{1}{N}\sum_k \log(T_k))$ . This inverse average transformation is then applied to  $M_i$  and used as the new reference image  $R_{i+1}$  for the next iteration:  $R_{i+1} = M_i \circ \overline{T}^{-1}$ .

This iterative process is complete when the variation between successive mean images is sufficiently small, i.e. when the difference between the mean transformations  $\overline{T}$  at two successive iterations are small. At the end of the average T1 image construction, a mean image  $\tilde{M}$  is obtained as well as transformations  $\tilde{T}_k$  deforming each image  $I_k$  on  $\tilde{M}$ . These obtained transformations are then applied to the tensor images (in the same way as shown in Section III-B1), therefore obtaining a set of DTI  $\tilde{DT}_k$  in the same reference frame.



Fig. 5. **Mean Tensor Image associated to the Mean Atlas.** Axial slices showing (a) the average T1 image and (b) the mean tensors obtained using our method superimposed on the atlas.

Our continuous STAPLE was used on the log-tensors of the registered DTI. Afterwards, an exponentiation step provides the reference standard tensors. As an illustration of this process, we show in Fig. 5 the average T1 image obtained from the database of 20 subjects, as well as an axial slice of the mean tensors obtained using continuous STAPLE.

We also performed the KLD evaluation proposed in Section II-D on the 20 sets of parameters obtained from continuous STAPLE. The obtained Gaussian scores (see Eq. (17)), presented in Table II, showed that 19 of the 20 normal subjects had similar results and were not significantly different from the mean reference standard population, with their probabilities ranging from 0.259 to 0.984. However, subject 16, image (d) on Fig. 3, had a probability of  $2 \times 10^{-4}$ , pointing out a significant difference in this subject compared to the rest of the population. This difference is due to the specific anatomy of the subject. This subject's ventricles are indeed much larger than any of the others in the database, resulting in registration discrepancies and in significant differences in the transformed tensors.

3) DTI Analysis of Multiple Sclerosis Patients: Once we created this atlas, the second step of our DTI analysis was to use it as a reference frame for the evaluation of the tensors inside the lesions of MS patients. To this end, we have followed, for each of the patients taken separately in the MS database, the following process:

• non linear registration moving the patient T1 image onto the atlas,

Subj. #	Score	Subj. #	Score	Subj. #	Score
1	0.294	8	0.561	15	0.934
2	0.282	9	0.667	16	$2 \times 10^{-4}$
3	0.867	10	0.807	17	0.652
4	0.633	11	0.907	18	0.259
5	0.657	12	0.903	19	0.953
6	0.792	13	0.984	20	0.640
7	0.858	14	0.970		

#### TABLE II

PARAMETER EVALUATION RESULTS ON THE NORMAL SUBJECTS DATABASE. GAUSSIAN SCORES BETWEEN EACH SUBJECT AND THE MEAN PARAMETERS OF THE ATLAS USING THE KLD-BASED EVALUATION FRAMEWORK (SEE EQ. (17)) SHOWING AN OUTLIER SUBJECT DUE TO REGISTRATION DISCREPANCIES.

- application of the transformation to the patient tensors,
- use of continuous STAPLE (over a predefined region of interest) on the patient DTI pooled with the atlas DTIs,
- patient results computation using the KLD evaluation framework.

The patient was registered in two steps on the atlas. First, a global affine transformation was computed. Then, a non linear registration was performed to align locally the anatomies of the atlas and of the patient. As intensities in voxels with lesions are different from those of the atlas, the registration was performed in a slightly different way as it was done for the atlas construction, following a method similar to [34]. No correspondences were computed inside a dilated mask of the patient's lesions and the transformation was therefore interpolated from neighboring regions. This method allows to take into account abnormalities in lesions signal so that they did not generate registration artifacts in the evaluated data.

The computed non linear transformation was then applied to the patient's DTI. We used continuous STAPLE over the set of DTI from the normal subjects and added the transformed DTI of the patient among the images. The algorithm was only run on a subregion of the images, i.e. on the 3D masks of the patient's lesions. We therefore obtained parameters for all of the subjects and the patient, and used them in our image comparison framework to test for a significant difference between the normal parameters and those of the patient.

We show the results of this evaluation, repeated for the 8 patients of the database, in Table III. The obtained probabilities of the patient parameters belonging to the mean Gaussian in our experiments were very small for all the patients, ranging from  $9.4 \times 10^{-89}$  to  $1.6 \times 10^{-4}$ . This shows, therefore, that there exists a significant difference between the MS patients' tensors and the normal subjects' tensors in the lesions regions.

#### **IV. CONCLUSION**

We have presented a new method, called continuous STA-PLE, to compute a reference standard from a set of vector or tensor images (i.e. where each voxel of the image can be represented as an  $\mathbb{R}^N$  vector) and parameters describing the bias and covariance of each image with respect to this reference. This algorithm was associated to a framework

Patient #	Score	Patient #	Score
1	$1.6 \times 10^{-4}$	5	$2.3 \times 10^{-87}$
2	$9.4 \times 10^{-89}$	6	$1.1 \times 10^{-12}$
3	$3.8 \times 10^{-9}$	7	$1.4 \times 10^{-83}$
4	$1.6 \times 10^{-38}$	8	$3.1 \times 10^{-16}$

#### TABLE III

PARAMETER EVALUATION RESULTS ON THE MS PATIENTS DATABASE. GAUSSIAN SCORES BETWEEN THE MS PATIENTS AND THE REFERENCE STANDARD TENSORS ON THE MS LESION REGIONS. ALL VALUES SHOW A STRONG DISAGREEMENT BETWEEN THE MS LESIONS TENSORS AND THE CONTROL PATIENTS TENSORS.

for the comparison of these parameters, which was utilized for the evaluation of local differences of tensors in DTI of multiple sclerosis patients, showing significant differences in the vicinity of lesions.

This DTI analysis will be performed in the future on other regions of the patients, such as the normal appearing white matter along the cortical spinal tract, to quantify the brain structure differences due to the disease. This algorithm could also be associated with the local comparison between the controls and the patients of FA, MD, eigenvalues or other characteristic values extracted from the DTI. This would allow to combine this improved detection framework based on the full tensor information to existing measurements to get an improved characterization of the various effects of a disease on the local white matter structure.

The comparison framework presented here is dependent on non linear registration: various algorithms may be used to create the atlas and register the patient on it. These algorithms can be based on anatomical images or DTI. DTI registration may even be used in conjunction with anatomical registration to compute the reference atlas. Registration errors may appear when the anatomy of a patient differs substantially from the mean anatomy, or due to the high variability of brain sulci from one subject to another. The evaluation of the robustness of the obtained results with respect to various registration algorithms would therefore be of interest to determine if significant differences may arise due to those methods. Interestingly, the framework presented here could also be used for the comparison of different registration methods and therefore to identify the good registration methods for a particular application.

Many other applications exist for this algorithm. As it can use general vector images as an input, it could be used on non linear transformations, again using the Log-Euclidean framework [8] to ensure a proper vector space. Another potential application would be the study of the Jacobians of transformations for the detection of abnormal anatomy. The comparison of these results with other hypothesis tests such as those proposed in Lepore et al. [5] would then be very interesting.

To be able to perform those studies, this comparison framework may be extended to perform group studies on a given region of interest (for example using the Cramers' test [3] with the proposed Kullback-Leibler divergence as a distance between the STAPLE parameters). Moreover, the generalization of this framework to the computation of voxelwise statistics, for example using blocks around each voxel, would further improve the power of this method allowing for the automatic detection of the regions where statistically significant differences exist. This would take into account the different local variabilities and therefore reduce the number of false negatives, particularly when the region of interest is large (e.g. when the study is performed on the whole white matter). This extended framework could then be an invaluable tool for the comparison of populations, the characterization of pathologies and of their evolution in response to treatment.

#### **ACKNOWLEDGMENTS**

This investigation was supported in part by a research grant from CIMIT; by NMSS grants RG 4032A1/1 and RG 3478A2/2; by NIH grants R03 CA126466, R01 RR021885, R01 GM074068, R01 EB008015 and P30 HD018655; and by NSF grant ITR 0426558.

#### REFERENCES

- M. Filippi, M. Cercignani, M. Inglese, and M. H. G. Comi, "Diffusion tensor magnetic resonance imaging in multiple sclerosis," *Neurology*, vol. 56, pp. 304–311, 2001.
- [2] A. Schwartzman, R. Dougherty, and J. Taylor, "Cross-subject comparison of principal diffusion direction maps," *Magnetic Resonance in Medicine*, vol. 53, no. 6, pp. 1423–1431, 2005.
- [3] B. Whitcher, J. J. Wisco, N. Hadjikhani, and D. S. Tuch, "Statistical group comparison of diffusion tensors via multivariate hypothesis testing," *Magnetic Resonance in Medicine*, no. 57, pp. 1065–1074, 2007.
- [4] J.-M. Peyrat, M. Sermesant, X. Pennec, H. Delingette, C. Xu, E. R. McVeigh, and N. Ayache, "A computational framework for the statistical analysis of cardiac diffusion tensors: Application to a small database of canine hearts," *IEEE Transactions on Medical Imaging*, vol. 26, no. 11, pp. 1500–1514, November 2007.
- [5] N. Lepore, C. A. Brun, Y.-Y. Chou, M.-C. Chiang, R. Dutton, K. Hayashi, E. Luders, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, and P. M. Thompson, "Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors," *IEEE Transactions on Medical Imaging*, vol. 27, no. 1, pp. 129–141, Jan. 2008.
- [6] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, January 2006.
- [7] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic Resonance in Medicine*, vol. 56, no. 2, pp. 411–421, Aug. 2006.
- [8] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache, "A Log-Euclidean framework for statistics on diffeomorphisms," in *Proc. of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), Part I*, ser. LNCS, no. 4190, Oct. 2006, pp. 924–931.
- [9] C.-F. Westin, S. Maier, H. Mamata, A. Nabavi, F. Jolesz, and R. Kikinis, "Processing and visualization for diffusion tensor MRI," *Medical Image Analysis*, vol. 6, no. 93-108, 2002.
- [10] D. Rey, G. Subsol, H. Delingette, and N. Ayache, "Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis," *Medical Image Analysis*, vol. 6, no. 2, pp. 163–179, Jun. 2002.
- [11] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration," *IEEE Trans Med Imaging*, vol. 22, no. 8, pp. 1014–25, Aug. 2003.
- [12] O. Commowick, R. Stefanescu, P. Fillard, V. Arsigny, N. Ayache, X. Pennec, and G. Malandain, "Incorporating statistical measures of anatomical variability in atlas-to-subject registration for conformal brain radiotherapy," in *MICCAI 2005, Part II*, ser. LNCS, J. Duncan and G. Gerig, Eds., vol. 3750. Palm Springs, CA, USA: Springer Verlag, Oct. 2005, pp. 927–934.
- [13] R. Woods, "Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation," *NeuroImage*, vol. 18, no. 3, pp. 769–788, 2003.

- [14] P. Fillard, V. Arsigny, X. Pennec, K. M. Hayashi, P. M. Thompson, and N. Ayache, "Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines," *NeuroImage*, 2006, in press.
- [15] R. Verma, P. Khurd, and C. Davatzikos, "On analyzing diffusion tensor images by identifying manifold structure using isomaps," *IEEE Transactions on Medical Imaging*, vol. 26, no. 6, pp. 772–778, Jun. 2007.
- [16] P. Khurd, R. Verma, and C. Davatzikos, "Kernel-based manifold learning for statistical analysis of diffusion tensor images," in *Information Processing in Medical Imaging*, 20th International Conference, IPMI 2007, ser. Lecture Notes in Computer Science, vol. 4584, 2007, pp. 581–593.
- [17] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, Jul. 2004.
- [18] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39 (Series B), 1977.
- [19] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. John Wiley and Sons, 1997.
- [20] S. K. Warfield, K. H. Zou, and W. M. Wells, "Validation of image segmentation by estimating rater bias and variance," in *Proceedings* of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), Part II, ser. LNCS, no. 4191, Oct. 2006, pp. 839–847.
- [21] D. B. Madan, "Differentiating a determinant," *The American Statistician*, vol. 36, no. 3, pp. 178–179, Aug. 1982.
- [22] S. Kullback, *Information theory and statistics*. John Wiley and Sons, 1968.
- [23] N. Toussaint, J. Souplet, and P. Fillard, "Medinria: Medical image navigation and research tool by inria," in *Proc. of MICCAI'07 Workshop* on Interaction in medical image analysis and visualization, Brisbane, Australia, 2007.
- [24] S. M. Smith, "Fast robust automated brain extraction," Human Brain Mapping, vol. 17, no. 3, pp. 143–155, Nov. 2002.
- [25] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multimodal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, pp. 35–52, 1996.
- [26] D. Rueckert, L. L. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [27] J. Ruiz-Alzola, C.-F. Westin, S. K. Warfield, C. Alberola, S. Maier, and R. Kikinis, "Nonrigid registration of 3D tensor medical data," *Medical Image Analysis*, vol. 6, no. 2, pp. 143–161, Jun. 2002.
- [28] O. Commowick and G. Malandain, "Evaluation of atlas construction strategies in the context of radiotherapy planning," in *Proceedings of the SA2PM Workshop (From Statistical Atlases to Personalized Models)*, Oct. 2006, Held in conjunction with MICCAI 2006.
- [29] H. Zhang, B. Avants, P. Yushkevich, J. Woo, S. Wang, L. McCluskey, L. Elman, E. Melhem, and J. Gee, "High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences in amyotrophic lateral sclerosis," *IEEE Transactions* on Medical Imaging - Special Issue on Computational Diffusion MRI, vol. 26, no. 11, pp. 1585–1597, Nov. 2007.
- [30] M.-C. Chiang, A. Leow, A. Klunder, R. Dutton, M. Barysheva, S. Rose, K. McMahon, A. de Zubicaray, G.I.and Toga, and P. Thompson, "Fluid registration of diffusion tensor images using information theory," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 442–456, Apr. 2008.
- [31] J. Yang, D. Shen, C. Misra, X. Wu, S. Resnick, C. Davatzikos, and R. Verma, "Spatial normalization of diffusion tensor images based on anisotropic segmentation," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 6914, Apr. 2008.
- [32] B. T. T. Yeo, T. Vercauteren, P. Fillard, X. Pennec, P. Golland, N. Ayache, and O. Clatz, "DTI registration with exact finite-strain differential," in *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*, May 2008.
- [33] A. Guimond, J. Meunier, and J.-P. Thirion, "Average brain models: A convergence study," *Computer Vision and Image Understanding*, vol. 77, no. 2, pp. 192–210, 2000.
- [34] R. Stefanescu, O. Commowick, G. Malandain, P.-Y. Bondiau, N. Ayache, and X. Pennec, "Non-rigid atlas to subject registration with pathologies for conformal brain radiotherapy," in *Proc. of the 7th Int. Conf on Medical Image Computing and Computer-Assisted Intervention - MICCAI* 2004, ser. LNCS, vol. 3216, Saint-Malo, France, September 2004, pp. 704–711.