Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference On Medical Imaging Understanding and Analysis 2016, MIUA 2016, 6-8 July 2016, Loughborough, UK

Cluster analysis of Diffusion Tensor fields with application to the segmentation of the Corpus Callosum

MIUA 2016 Submission # 33,
Safa Elsheikh[a], Andrew Fish[a], Roma Chakrabarti[a], Diwei Zhou[b,*]

[a]*School of Computing, Engineering and Mathematics, University of Brighton*
[b]*Department of Mathematical Sciences, Loughborough University*

Abstract

Accurate segmentation of the Corpus Callosum (CC) is an important aspect of clinical medicine and is used in the diagnosis of various brain disorders. In this paper, we propose an automated method for two and three dimensional segmentation of the CC using diffusion tensor imaging. It has been demonstrated that Hartigan's K-means is more efficient than the traditional Lloyd algorithm for clustering. We adapt Hartigan's K-means to be applicable for use with the metrics that have a f-mean (e.g. Cholesky, root Euclidean and log Euclidean). Then we applied the adapted Hartigan's K-means, using Euclidean, Cholesky, root Euclidean and log Euclidean metrics along with Procrustes and Riemannian metrics (which need numerical solutions for mean computation), to diffusion tensor images of the brain to provide a segmentation of the CC. The log Euclidean and Riemannian metrics provide more accurate segmentations of the CC than the other metrics as they present the least variation of the shape and size of the tensors in the CC for 2D segmentation. They also yield a full shape of the splenium for the 3D segmentation.

Peer-review under responsibility of the Organizing Committee of MIUA 2016.

Keywords: Corpus Callosum; Hartigan's K-means; f-mean; Cholesky; Procrustes; Riemannian

1. Introduction

The corpus callosum (CC) is a great fiber bundle in the white matter of the brain. It connects the two hemispheres of the brain. An accurate segmentation of the CC is needed in clinical applications (e.g. surgical planning and disease prognosis). There is a large quantity of research that has used Magnetic Resonance Imaging (MRI) to segment the CC. Diffusion Tensor Imaging (DTI) is an advanced technique of MRI which is able to measure the diffusion of the water inside the brain tissues. DTI has also been used for the segmentation of the CC (e.g. [1–4]). Lenglet et al. [1] proved that the segmentation of the CC using the Riemannian metric is superior to the segmentation using Euclidean metric. Their method was based on *surface evolution* which seeks the optimal partition through a Bayesian formulation. Lee et al. [2] proposed an automatic $2D$ segmentation of the CC using the color coded map of the diffusion tensors. Goh and Vidal [3] proposed the Locally Linear algorithm for diffusion tensor clustering (LLDTC) to segment the fiber bundles (e.g. Cingulum and the CC) using Riemannian and log-Euclidean metrics. In fact, LLDTC is a generalization of a locally linear embedding method which is a dimensionality reduction method based on embedding

* Corresponding author. Tel.:01509223482.
E-mail address: D.Zhou2@lboro.ac.uk

1877-0509
Peer-review under responsibility of the Organizing Committee of MIUA 2016.

high dimensional data on Euclidean space into low dimensional space. The LLDTC uses the K-means algorithm to cluster the embedded points. Nazem-Zadeh et al. [4] proposed a three dimensional segmentation of the CC using DTI. They used the diffusivity pattern of the CC as prior information. A similarity measure, based on a *speed function*, has been proposed to segment the CC and its subdivisions.

Non-Euclidean means are alternatives to the Euclidean mean for the space of covariance matrices. In the context of tensor interpolation, the linear interpolation of tensors (using the Euclidean mean) produces tensors with parabolic, non monotonic, determinants and hence large sized tensors in the interpolated path [5]. As opposed to the Euclidean mean, the log Euclidean [5] and Riemannian [6] means yield monotonic determinants of the interpolated tensors. Carmichael et al. [7] stated that using the log Euclidean and Riemannian means are more effective in smoothing the tensor field in anisotropic regions than the Euclidean mean. Cholesky, Procrustes size and shape and root Euclidean means are other non-Euclidean alternatives for interpolating and smoothing diffusion tensors [8]. For more details about non-Euclidean methods see [8,9].

In K-means cluster analysis, Hartigan's method [10] (also known as the Exchange method [11]) is more efficient than the Lloyd algorithm [12], as it results with tighter clusters than the Voronoi diagram [13] and hence it converges to a solution with smaller within cluster sum of squares (WCSS). The Lloyd method considers how close an object is to the centroid of the cluster and hence each object is assigned to its closest centroid. The result of Hartigan's method guarantees that no movement of an object to any cluster will reduce the WCSS.

In this article, in Section 2, we present the objective function for the Lloyd algorithm and Hartigan's method. In Section 3, we show that Hartigan's method can be generalized to the metrics which have the f-mean property. In Section 4 we provide a 2D and 3D segmentation of the CC. We show that the log Euclidean and Riemannian metrics outperform the other metrics in the segmentation of the CC as they provide the least size and shape variations of the tensors in the CC for 2D segmentation, and they segment the splenium (posterior end) as part of the CC for 3D segmentation.

2. Hartigan's method

We begin by introducing some basic notations that are used in this section. Suppose $\mathbf{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_M\}$ is a set of covariance matrices. Suppose $C = \{C_1, \ldots, C_K\}$ where C_k, for $k \in \{1, \ldots, K\}$, is the set of the indices of the covariance matrices inside the cluster k (i.e. C is a set of sets of the indices arising from a partition of A into K clusters). Let $\bar{\mathbf{A}}_p$ be the centroid of the cluster p and $m(p)$ be the number of covariance matrices in the cluster p.

Both Hartigan's method and Lloyd's algorithm aim to find a partition of a data set that minimizes the WCSS, but they use different methods to reduce the WCSS. The objective of Lloyd's algorithm is to find a partition C that minimizes the WCSS, i.e. to $\arg\min_{C} \sum_{k=1}^{K} \sum_{i \in C_k} d^2(\mathbf{A}_i, \bar{\mathbf{A}}_k)$, where $d(\mathbf{A}_i, \bar{\mathbf{A}}_k)$ is the distance between $\mathbf{A}_i$ and $\bar{\mathbf{A}}_k$, where $\bar{\mathbf{A}}_k$ is the mean of the covariance matrices in the cluster k. Hartigan's method iterates over all the covariance matrices in $\mathbf{A}$ to draw each covariance matrix out of its cluster, updates the centroid of the clusters, and assigns the covariance matrix to another cluster if doing so will reduce the WCSS. Suppose C_p and $C_j \in C$ and let $i \in C_p$. The method tries to move $\mathbf{A}_i$ into cluster j where $j \in \{1, \ldots, k\}$ and $j \neq p$, then calculates the new WCSS. Suppose the partitioning before the movement was $C = \left\{C_1, \ldots, C_p, \ldots, C_j, \ldots, C_K\right\}$. Then the partitioning after the movement is $C = \left\{C_1, \ldots, C_p \setminus \{i\}, \ldots, C_j \cup \{i\}, \ldots, C_K\right\}$. After the movement, we need to calculate the new centroids and the new WCSS. Spath [11] showed that, using the Euclidean distance, the new centroids and the new WCSS can be calculated in a simple and quick way as a function of the previous centroids and the previous WCSS respectively. Let $\bar{\mathbf{A}}_{np}$ be the new centroid of the cluster p after the movement of $\mathbf{A}_i$ from the cluster p. Let $C_{np} = C_p \setminus \{i\}$ and $C_{nj} = C_j \cup \{i\}$. Then WCSS of the clusters p before the movement is $WCSS_p = \sum_{t \in C_p} \left\|\mathbf{A}_t - \bar{\mathbf{A}}_p\right\|^2$. The new WCSS of the cluster p and j when $\mathbf{A}_i$ is moved from the cluster p to j are $WCSS_{np} = WCSS_p - \frac{m(p)}{m(p)-1} \left\|\mathbf{A}_i - \bar{\mathbf{A}}_p\right\|^2$ and $WCSS_{nj} = WCSS_j + \frac{m(j)}{m(j)+1} \left\|\mathbf{A}_i - \bar{\mathbf{A}}_j\right\|^2$ (see [11]). As the movement is between the cluster p and the cluster j, the WCSS of other clusters will not be affected. The new total of the WCSS over all clusters is $WCSS_{new} = WCSS_{pre} + \frac{m(j)}{m(j)+1} \left\|\mathbf{A}_i - \bar{\mathbf{A}}_j\right\|^2 - \frac{m(p)}{m(p)-1} \left\|\mathbf{A}_i - \bar{\mathbf{A}}_p\right\|^2$ where $WCSS_{pre}$ is the total of WCSS, over all clusters, before the movement. Let $G_j = \frac{m(j)}{m(j)+1} \left\|\mathbf{A}_i - \bar{\mathbf{A}}_j\right\|^2 - \frac{m(p)}{m(p)-1} \left\|\mathbf{A}_i - \bar{\mathbf{A}}_p\right\|^2$, then for the movement to be successful $\exists j \in \{1, \ldots, k\}$ such that $G_j < 0$. In this case the movement of $\mathbf{A}_i$ will be to the cluster j with smallest G_j as we want the largest reduction in the $WCSS_{new}$. Therefore, the objective function for Hartigan's method is $\min_{j \neq p, G_j < 0} G_j$.

3. Hartigan's method with f-mean

The f-mean (also known as the quasi-arithmetic mean and the Kolmogorov mean) is a generalization of the arithmetic and the geometric means. Suppose f is a continuous and injective function, then $f : \mathbb{I} \subset \mathbb{C} \longrightarrow \mathbb{C}$ where $\mathbb{C}$ denotes the class of covariance matrices. The f-mean for the covariance matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ is defined as $\bar{\mathbf{A}} = f^{-1}(\frac{1}{n}\sum_{i=1}^{n} f(\mathbf{A}_i))$. The f-mean is a special case of the Fréchet[14] mean obtained by using the metric $\|f(\mathbf{A}_1) - f(\mathbf{A}_2)\|$ as the distance between $\mathbf{A}_1$ and $\mathbf{A}_2$. The Fréchet mean, and hence the f-mean, is unique for Euclidean and non-Euclidean spaces[8]. Suppose the spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^\intercal$ where $\mathbf{V}$ is an orthogonal matrix, $\mathbf{V} \in O(n)$, and Λ is a diagonal matrix of the eigenvalues of $\mathbf{A}$. Suppose $\mathbb{I} = \mathbb{C}$, then the Euclidean, Cholesky and root Euclidean means are examples of f-mean with $f(\mathbf{A}) = \mathbf{A}$, $f(\mathbf{A}) = \text{chol}(\mathbf{A})$ and $f(\mathbf{A}) = \mathbf{A}^{1/2} = \mathbf{V}\Lambda^{1/2}\mathbf{V}^\intercal$ respectively. Suppose $\mathbb{I} = \mathbb{C}^+$ is the class of positive-definite covariance matrices, then the log Euclidean mean is an example of f-mean with $f(\mathbf{A}) = \log(\mathbf{A}) = \mathbf{V}\log(\Lambda)\mathbf{V}^\intercal$. Here, we develop a formula for Hartigan's objective function for the metrics which have the f-mean property. Given the notations in Section 2, the details are below.

Theorem 1. *In Hartigan's method, suppose the covariance matrix $\boldsymbol{A}_i$ is moved from the cluster p to the cluster j. Suppose the distance between two covariance matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ is $\|f(\boldsymbol{A}_1) - f(\boldsymbol{A}_2)\|$, where $\|.\|$ is the Frobenius norm and $\bar{\boldsymbol{A}}_j = f^{-1}(\frac{1}{m(j)}\sum_{t\in C_j} f(\boldsymbol{A}_t))$ is the f-mean of the covariance matrices in the cluster j, then the function G_j for Hartigan's method is $G_j = \frac{m(j)}{m(j)+1}\left\|f(\boldsymbol{A}_i) - f(\bar{\boldsymbol{A}}_j)\right\|^2 - \frac{m(p)}{m(p)-1}\left\|f(\boldsymbol{A}_i) - f(\bar{\boldsymbol{A}}_p)\right\|^2$.*

Proof. For arbitrary $k \in C_{nj}$:

$$f(\mathbf{A}_k) - f(\bar{\mathbf{A}}_{nj}) = \left[f(\mathbf{A}_k) - f(\bar{\mathbf{A}}_j)\right] + \left[f(\bar{\mathbf{A}}_j) - f(\bar{\mathbf{A}}_{nj})\right] \tag{1}$$

$$\text{and } f(\bar{\mathbf{A}}_{nj}) = \frac{1}{m(j)+1}\sum_{k\in C_{nj}} f(\mathbf{A}_k) = \frac{1}{m(j)+1}\left[\sum_{t\in C_j} f(\mathbf{A}_t) + f(\mathbf{A}_i)\right] = \frac{1}{m(j)+1}\left[m(j)f(\bar{\mathbf{A}}_j) + f(\mathbf{A}_i)\right]. \tag{2}$$

The centroid of the new cluster nj can be calculated by taking the inverse function of $f(\bar{\mathbf{A}}_{nj})$ in Equation 2. Furthermore, using Equation 2:

$$f(\bar{\mathbf{A}}_j) - f(\bar{\mathbf{A}}_{nj}) = f(\bar{\mathbf{A}}_j) - \frac{1}{m(j)+1}\left[m(j)f(\bar{\mathbf{A}}_j) + f(\mathbf{A}_i)\right] = \frac{1}{m(j)+1}\left[f(\bar{\mathbf{A}}_j) - f(\mathbf{A}_i)\right]. \tag{3}$$

Substitute Equation 3 into Equation 1 as follows:

$$f(\mathbf{A}_k) - f(\bar{\mathbf{A}}_{nj}) = \left[f(\mathbf{A}_k) - f(\bar{\mathbf{A}}_j)\right] + \frac{1}{m(j)+1}\left[f(\bar{\mathbf{A}}_j) - f(\mathbf{A}_i)\right]. \tag{4}$$

Equation 4 implies $\sum_{k\in C_{nj}}\left\|f(\mathbf{A}_k) - f(\bar{\mathbf{A}}_{nj})\right\|^2 = \sum_{t\in C_j}\left\|f(\mathbf{A}_t) - f(\bar{\mathbf{A}}_j)\right\|^2 + \frac{m(j)}{m(j)+1}\left\|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_j)\right\|^2$. Therefore $WCSS_{nj} = WCSS_j + \frac{m(j)}{m(j)+1}\left\|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_j)\right\|^2$. Similarly, the new centroid of the cluster p can be calculated by taking the inverse of $f(\bar{\mathbf{A}}_{np})$ where

$$f(\bar{\mathbf{A}}_{np}) = \frac{1}{m(p)-1}\sum_{k\in C_{np}} f(\mathbf{A}_k) = \frac{1}{m(p)-1}\left[m(p)f(\bar{\mathbf{A}}_p) - f(\mathbf{A}_i)\right]. \tag{5}$$

In addition, the WCSS of the cluster p, after the movement, is $WCSS_{np} = WCSS_p - \frac{m(p)}{m(p)-1}\left\|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_p)\right\|^2$. Then the change amount in the WCSS, where $j \neq p$, is $G_j = \frac{m(j)}{m(j)+1}\left\|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_j)\right\|^2 - \frac{m(p)}{m(p)-1}\left\|f(\mathbf{A}_i) - f(\bar{\mathbf{A}}_p)\right\|^2$ □

For large data sets, Hartigan's method runs very slowly as it depends on moving objects one at a time. To improve efficiency, the centroids resulting from Lloyd's algorithm can be used as an initialization to Hartigan's method.

Hartigan's method cannot be generalized for the metrics which need numerical solutions for mean computations (e.g. the Procrustes and Riemannian metrics). That is because Hartigan's method is based on the relation between the old centroid of a cluster and the new centroid of that cluster after a movement (as in Equation 2 and 5). In such metrics

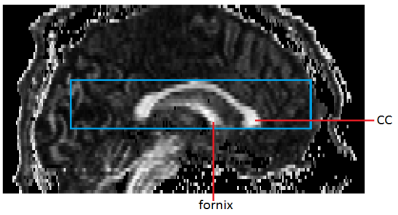
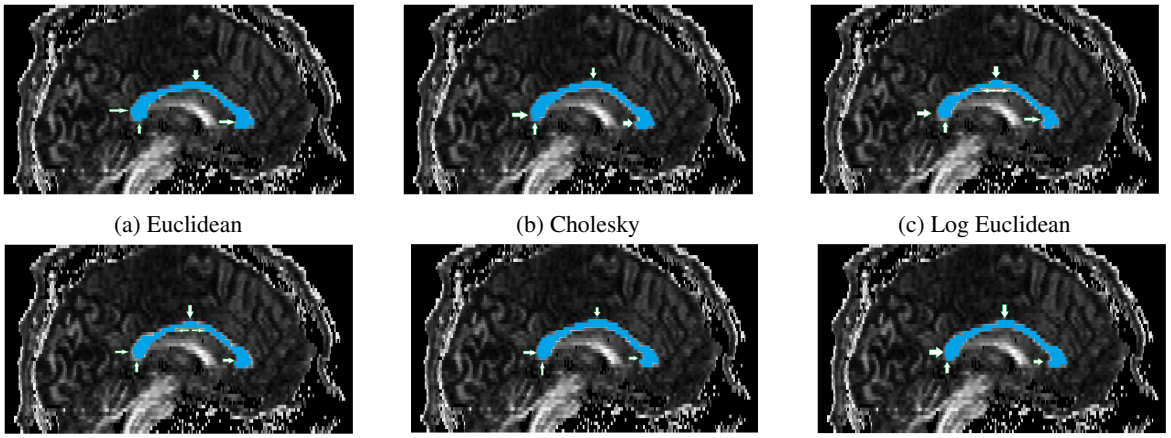


Fig. 1: The Fractional Anisotropy map of the sagittal view of the brain. The ROI is indicated by the blue rectangle.



(a) Euclidean (b) Cholesky (c) Log Euclidean

(d) Riemannian (e) Root Euclidean (f) Procrustes

Fig. 2: Segmentation of the CC (the blue color) from the sagittal view. Different metrics yield different segmentations (see the arrows).

which need numerical solutions for mean computations, it is not possible to guess the new centroid explicitly, in terms of the old centroid. Hence, for those metrics we add a condition to run Hartigan's method only if that produces a smaller WCSS.

4. Application

In this section, 2D and 3D segmentation of the CC are provided. To reduce clustering time, the algorithm used is Lloyd's algorithm followed by Hartigan's method, which we refer to them together briefly as "the k-means algorithm". The Academic Radiology Department of Queens Medical Center in Nottingham provided us with a set of MR images of a healthy human brain. The MR images were acquired using a $3T$ Phillips Achieva scanner with a receive-only eight elements head coil. The diffusion tensor MR images were acquired using a spin echo, EPI, sequence with diffusion weighting gradients applied with a weighting factor of $b = 1000\ s/mm^2$. Then, 52 interleaved contiguous transaxial slices were acquired throughout the subject's head in a matrix of 112×112 (interpolated to 224×224) with an acquisition voxel size of $1 \times 1 \times 2\ mm^3$. For each slice, the acquisition was repeated to acquire diffusion weighted images in 32 non-collinear directions, as well as one acquisition with no diffusion weighting ($b = 0$). The acquisition was repeated twice and the data averaged to improve the signal to noise ratio.

4.1. Two dimensional segmentation

The region of interest (ROI) from a sagittal view of the brain is selected (the blue rectangle in the Fractional Anisotropy (FA) map in Fig. 1). The diffusion tensors in the ROI are clustered into five groups (see below for illustration) using the k-means algorithm with the Euclidean and non-Euclidean metrics (see Fig. 2). The CC is the bright (white) region in the FA map. Note that the small bright shape near the CC is not a part of the CC and it is called the fornix. Although, the fornix has high FA values just like the CC, our clustering method distinguishes between them and clusters them in two different clusters. The segmentation results are different from each other (see the arrows in Fig. 2).

The Silhouette method[15] is used to validate the cluster size and choose the appropriate number of clusters for the segmentation of the CC. Suppose we have a set of n objects clustered into K clusters. The object $i \in \{1, \ldots, n\}$, is a member of the cluster P. We need to check whether P is the appropriate cluster for i (i.e. i is nearer to P than other

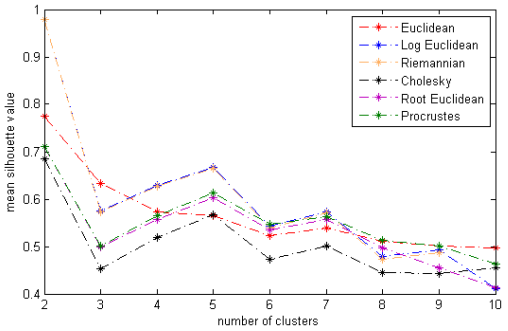


Fig. 3: Silhouette curves show that the highest mean of silhouette values is at 2 clusters, followed by 5 clusters for non-Euclidean metrics.

Table 1: The SEMean of FA, determinant (det) and ϕ for the CC region. The accuracy (AC) and specificity (SP) measures are also calculated.

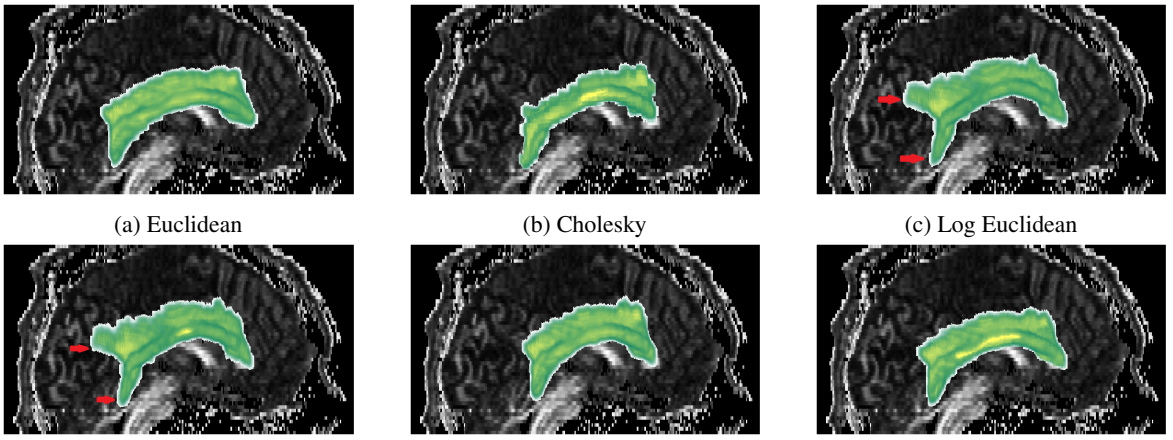
Metric	$FA \times 10^{-2}$	det$\times 10^{-12}$	ϕ	AC%	SP%
Euclidean	0.920	20.778	0.3288	95.77	95.41
log Euclidean	**0.681**	**8.3014**	0.3463	97.96	98.07
Riemannian	**0.690**	**8.3011**	0.3494	97.91	98.01
Cholesky	0.817	15.745	0.3284	96.48	96.30
root Euclidean	0.816	15.217	**0.3225**	96.54	96.35
Procrustes	0.808	15.112	**0.3233**	96.59	96.41

clusters). Suppose $a(i)$ is the average of the distances between the object i and other objects in the same cluster, $b(i)$ is the lowest average of the distances between i and any other clusters (different from P). The Silhouette value for i, denoted $s(i)$, is given by $s(i) = [b(i) - a(i)]/\max\{a(i), b(i)\}$. To evaluate the number of clusters, the K with the highest overall average of the silhouette values, $avg(s(i))$, is the appropriate cluster size. In particular, we are interested in the appropriate cluster size where CC is visible. We used the Euclidean and non-Euclidean distances and means to calculate $s(i)$ then $avg(s(i))$ (Fig. 3). For all non-Euclidean metrics, the highest $avg(s(i))$ is obtained at 2 then 5 clusters (see Fig. 3). As the CC is not really distinguishable from other brain regions at 2 clusters, then 5 is the most appropriate cluster size for the segmentation of the CC. For the Euclidean metric, 2 clusters has a higher $avg(s(i))$ than 3 clusters. However, the CC shape is still not distinguishable either for 2 or 3 clusters. The silhouette value indicates that 4 clusters are slightly more appropriate than 5 clusters in the Euclidean metric. When comparing the performance of the Euclidean metric with 4 clusters and non-Euclidean metrics with 5 clusters, the CC shape using the Euclidean metric is more swollen at 4 clusters (contains 256 tensors) and has a higher variance. On the other hand, the number of tensors in the CC using 5 clusters are in the range $[183, 218]$ for non-Euclidean metrics.

To evaluate the segmentation results, the standard error of the mean (SEMean) is used. The shape of a tensor $\mathbf{A}$ can be measured using $FA = \sqrt{[(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2]\,/\,\left[2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2\right]}$, where λ_j, $j \in \{1, 2, 3\}$ are the eigenvalues of $\mathbf{A}$. In addition, the determinant of $\mathbf{A}$, $|\mathbf{A}|$, can be used to measure the size of $\mathbf{A}$. Suppose ϕ is the angle between each principal eigenvector, $\mathbf{v}_i$ and the principal eigenvector of the mean of the tensors in the CC, $\mathbf{vm}$. Then ϕ can be calculated as $\phi = \arccos\left([\mathbf{v}_i \cdot \mathbf{vm}]\,[\|\mathbf{v}_i\| * \|\mathbf{vm}\|]\right)$. Suppose var is the variance of the FA, $|\mathbf{A}|$ or ϕ, then SEMean $= \sqrt{var/n}$. To evaluate the shape, size and orientation of the tensors in the CC, we use the SEMean of FA, $|\mathbf{A}|$ and ϕ respectively (Table 1). The metric that yields the least SEMean of at least two of the measures is the accurate metric for the segmentation of the CC. The bold and red entries in Table 1 indicate the least and the highest SEMean of each measure. It is clear, from Table 1, that log Euclidean and Riemannian are the most accurate metrics for estimating the shape and size of the CC. The Euclidean metric yields the highest variation of the shape and size of the tensors in the CC and hence other methods are preferable for the segmentation. Regarding the Orientation, the log Euclidean and Riemannian metrics produce the highest SEMean(ϕ). Although the root Euclidean and Procrustes provide the smallest SEMean(ϕ), they provide higher variation of the size and shape of the tensors in the CC. The segmentation results are compared with the manual segmentation drawn by an expert and the accuracy (AC) and specificity (SP) measures [16] are calculated (see Table 1). The log Euclidean and Riemannian yield the highest accuracy and specificity of the segmentation of the CC.

4.2. Three dimensional segmentation

The ROI from the middle of the brain is used as input to the K-means algorithm. The CC is not distinguishable from other brain regions at 2 and 3 clusters, for any of the Euclidean and non-Euclidean metrics. At 4 and 5 clusters, the CC is clearly distinguishable from other brain regions. However, there are disjoint components within the CC cluster, appearing outside the main CC region. For each metric, these disjoint components are always smaller at 5 clusters than at 4 clusters. At cluster size 6, the CC is divided into two different clusters for each metric. That indicates the most suitable cluster size for segmentation of the CC is 5. The segmentation after removing the small objects around the CC is shown in Fig. 4. The splenium of the CC is entirely visible using the log Euclidean and Riemannian metrics (see the red arrows in Fig. 4) while other metrics fail to provide a full shape of the splenium.



(a) Euclidean (b) Cholesky (c) Log Euclidean

(d) Riemannian (e) Root Euclidean (f) Procrustes

Fig. 4: The three dimensional segmentation of the CC, superimposed over a 2D background, after removing the small objects around the CC.

5. Discussion

For 2D segmentation, the log Euclidean and Riemannian metrics yield accurate segmentation of the CC regarding the size and shape of the tensors in the CC. They also produce a full shape of the splenium of the CC for the 3D segmentation. This confirms the results obtained by Lenglet et al. [1]. Therefore, the segmentation of the CC using the log Euclidean and Riemannian metrics are preferable than the other metrics (Euclidean, Cholesky, root Euclidean and Procrustes). While the Riemannian mean needs a numerical method for the computation, the log Euclidean mean is easy and quick to compute without resorting to a numerical method and hence the segmentation using the log Euclidean method is recommended as it is more efficient. To obtain a smoother shape of the CC without holes, we may wish to carry out basic tensor data processing (e.g. smoothing and interpolation) before the segmentation. The data set used in this paper is from one healthy brain so using more data sets will be considered in future work.

References

1. Lenglet, C., Rousson, M., Deriche, R.. A statistical framework for DTI segmentation. In: *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*. IEEE; 2006, p. 794–797.
2. Lee, S.P., Cheng, J.Z., Chen, C.M., Tseng, W.Y.I.. An automatic segmentation approach for boundary delineation of corpus callosum based on cell competition. In: *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE; 2008, p. 5514–5517.
3. Goh, A., Vidal, R.. Segmenting fiber bundles in diffusion tensor images. In: *Computer Vision–ECCV 2008*. Springer; 2008, p. 238–250.
4. Nazem-Zadeh, M.R., Saksena, S., Babajani-Fermi, A., Jiang, Q., Soltanian-Zadeh, H., Rosenblum, M., et al. Segmentation of corpus callosum using diffusion tensor imaging: validation in patients with glioblastoma. *BMC medical imaging* 2012;**12**(1):10.
5. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J Matrix Anal Appl* 2007;**29**:328–347.
6. Pennec, X., Fillard, P., Ayache, N.. A Riemannian framework for tensor computing. *International Journal of Computer Vision (Springer)* 2006;**66**(1):41–66.
7. Carmichael, O., Chen, J., Paul, D., Peng, J.. Diffusion tensor smoothing through weighted Karcher means. *Electronic Journal of Statistics* 2013;**7**:1913–1956.
8. Dryden, I., Kolydenko, A., Zhou, D.. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics* 2009;**3**:1102–1123.
9. Zhou, D., Dryden, I.L., Koloydenko, A.A., Audenaert, K.M., Bai, L.. Regularisation, interpolation and visualisation of diffusion tensor images using non-Euclidean statistics. *Journal of Applied Statistics* 2016;**43**(5):943–978.
10. Hartigan, J.A.. *Clustering algorithms*. John Wiley & Sons, Inc.; 1975.
11. Spath, H.. *The cluster dissection and analysis theory FORTRAN programs examples*. Prentice-Hall, Inc.; 1985.
12. Lloyd, S.P.. Least squares quantization in PCM. *Information Theory, IEEE Transactions on* 1982;**28**(2):129–137.
13. Telgarsky, M., Vattani, A.. Hartigan's method: k-means clustering without voronoi. In: *International Conference on Artificial Intelligence and Statistics*. 2010, p. 820–827.
14. Fréchet, M.. Les éléments aléatoires de nature quelconque dans un espace distancié. In: *Annales de l'institut Henri Poincaré*; vol. 10. Presses universitaires de France; 1948, p. 215–310.
15. Rousseeuw, P.J.. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 1987;**20**:53–65.
16. Xu, T., Mandal, M., Long, R., Cheng, I., Basu, A.. An edge-region force guided active shape approach for automatic lung field detection in chest radiographs. *Computerized Medical Imaging and Graphics* 2012;**36**(6):452–463.