Comparison of Methods for Hyperspherical Data Averaging and Parameter Estimation

Kai Rothaus, Xiaoyi Jiang and Martin Lambers
Department of Computer Science, University of Münster, Germany
{rothaus,xjiang,marlam}@math.uni-muenster.de

Abstract

Averaging is an important concept which has found numerous applications in general and in pattern recognition and computer vision in particular. In this paper we consider averaging directional vectors of arbitrary dimensions. Given a set of vectors, we intend to compute an average vector which optimally represents the input vectors according to some formal criterion. Several optimisation criteria are formulated. In particular, we present a class of robust estimators of up to 50% outlier tolerance. Furthermore, we propose a technique to estimate another distribution parameter. Experimental results on spherical data are presented to demonstrate the usefulness of the proposed methods.

1 Introduction

The general concept of averaging has turned out to be useful in numerous contexts of science and engineering. In pattern recognition and computer vision, interesting applications have been demonstrated for graphs [6], 2D shapes [12], brain models [5], and anatomical structures [15].

In this work, we treat data samples on the surface of the unit hypersphere HS^{D-1} embedded in the vector space $\mathbb{R}^D$. In the following this hyperspherical surface is the underlying data space. We use the Cartesian vector representation for the data points: $\vec{v} = \left(v^1, \ldots, v^D\right) \in \mathbb{R}^D$ with $\|\vec{v}\| = 1$.

The fundamental difficulty in dealing with such vectors is their cyclic nature. Looking at the 2D case as an example, although a vector can be specified by one angle, the average of a vector set cannot be computed by the arithmetic mean of the angles. The cyclic property is responsible for the substantially higher complexity of computing average vectors.

In this work we consider averaging vectors (of arbitrary dimensions), i.e. finding a vector which optimally represents a set of vectors according to some formal criterion. The solution of this problem may be applied to smooth vector fields from, say, optical flow and shape from shading computation. Additionally, the results of several such methods may be combined this way to achieve improved performance. Another application lies in handling 3D rotations. Since a rotation can be represented as a quaternion, averaging vectors immediately leads to averaging rotations [4].

There is very few work on this topic. Gramkow [4] and Moakher [10] consider averaging 3D rotations. While Gramkov starts with the quaternion representation, Moakher's work is based on the matrix representation of rotations. Both approaches are presenting rotation-specific solutions which are not general enough for arbitrary vectors.

We present different methods of computing a representative vector $\vec{v}_0 \in HS^{D-1}$ on a finite observation set $S = (\vec{v}_1, \ldots, \vec{v}_n)$. Thereby, the special geometry of the hypersphere surface is focused, whereas the cyclicity is the most important geometrical feature. The distance of any fixed point to a point moving on a hyperspherical line is a periodical function (Fig. 1).

Micó et al. [9] propose an efficient, randomised algorithm to compute the set median in general nonmetric spaces. This restriction of the search space is justified in cases of a concentrated observation set. We search for a representative vector in the whole domain but not only in S.

In statistical research areas the definition of a median is based on a depth function which has to be maximised [13]. A depth function is viewed as a measure how central a point is lying in a data distribution or an empirical observation set. Adapted from the simple case of the L_1-median on the line, various depth functions extend the L_1-median to higher dimensional vector spaces and manifolds. Liu et al. [7] treat the unit sphere and adapt commonly used depth functions (Tukey's depth, simplicial depth) as well as a reformulation of the spherical median (see LS in Sec. 2.1 and Fisher [1]).

The concept of generalised median is based on a distance function d. The problem is to find the median $\vec{v}_0$ such that the mean distance to the observation S is minimised, i.e.

$$\vec{v}_0 = \arg \min_{\vec{v} \in HS^{D-1}} \frac{1}{n} \cdot \sum_{i=1}^{n} d\left(\vec{v}, \vec{v}_i\right).$$

In this work, four different generalised median solutions (Bary, Eigen, LS and LSQ) for computing the vector representative on a hyperspherical surface are taken into consid-

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00

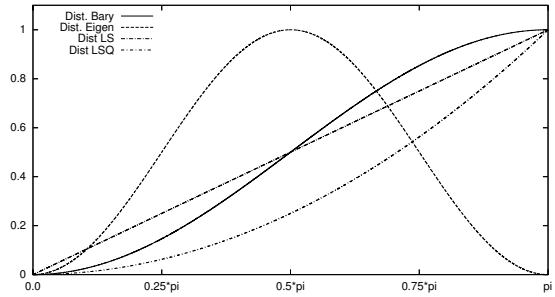

Figure 1. Distance functions on HS^{D-1}

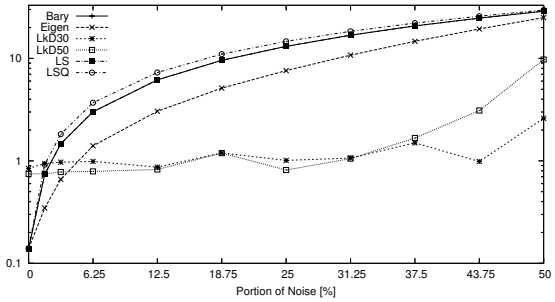

Figure 2. μ-Error depending on α

eration (Sec. 2.1). Beside these generalised median methods we present another class of estimators, which is a generalisation of the least median error approach (Sec. 2.2).

A good method for computing a representative vector should be robust against outliers (noise) and their distribution and should provide a good estimation for further data distribution parameters. We evaluate the performance of the different estimators in the situation of Fisher distributed data and noise samples for computing the representative vector (Sec. 2.3) and the concentration parameter κ, which is the shape parameter of the Fisher distribution (Sec. 3).

For the sake of clarity we first present the methods with unit vectors and then give an extension to non unit length vectors (Sec. 4). This extension could be viewed as weighted generalised median, where the norm of each vector is proportional to its contribution to the error function. Furthermore, we extend the methods for analysing the distributions themselves.

2 Estimation of the representative vector

In this section we present the estimators for the representative vector of a given observation set $S = (\vec{v}_1, \ldots, \vec{v}_n)$. As mentioned before, different distance functions on the data space lead to different generalised median approaches.

2.1 Generalised median approaches

We introduce four approaches of this category. In Figure 1 the underlying distance functions of the presented approaches are shown. Note that the function depends on the arc length distance d_{arc} of the two vectors (x-axis) and that the functions are scaled so that the maximum of each is 1.

Bary: The data space is embedded in the $\mathbb{R}^D$ and the squared euclidian distance is a valid distance function in this vector space. Considering two vectors $\vec{x}$, $\vec{y}$ of unit length, we define

$$d_{\text{Bary}}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|_2^2 = 2 - 2\cos(\vec{x}, \vec{y})$$

where $\cos(\vec{x}, \vec{y})$ defines the cosine of the inner angle between the vectors $\vec{x}$ and $\vec{y}$. Given the observation set S, it is a well-known fact that the barycentre of S

$$\vec{v}_{\text{Bary}} = \left(v^1_{\text{Bary}}, \ldots v^D_{\text{Bary}}\right) \text{ where } v^j_{\text{Bary}} = \frac{1}{n}\sum_{i=1}^{n} v_i^j$$

minimises the error function e_{Bary} introduced by d_{Bary}:

$$e_{\text{Bary}}(\vec{v}_0) = \frac{1}{n}\sum_{i=1}^{n} \|\vec{v}_0 - \vec{v}_i\|_2^2 = \frac{2}{n}\sum_{i=1}^{n} \left(1 - \cos(\vec{v}_0, \vec{v}_i)\right).$$

The problem of this solution is that v_{Bary} does not lie on the hypersphere surface and a projection is needed. However, it can be shown that the normalised vector $\vec{v}_{\text{Bary}}$ minimises the error function subject to HS^{D-1} so that the concept of generalised median is justified on the hypersphere surface. The computation of this estimator is easy to realise in any dimension in $\mathcal{O}(n)$ time.

Eigen: Defining $T = \frac{1}{n}\sum_{i=1}^{n} \vec{v}_i^t \vec{v}_i$, the distance function

$$d_{\text{Eigen}}(\vec{x}, \vec{y}) = \sin^2(\vec{x}, \vec{y}) = 1 - (\vec{x} \cdot \vec{y})^2$$

leads to the error function

$$e_{\text{Eigen}}(\vec{v}_0) = 1 - \sum_{i=1}^{n} \frac{1}{n}(\vec{v}_i \cdot \vec{v}_0)^2 = 1 - \vec{v}_0 T \vec{v}_0^t$$

Obviously, the first principal axis, represented by an eigenvector to the largest eigenvalue of T, minimises e_{Eigen}. Once T is built in $\mathcal{O}(n)$ time, v_0 can be computed very fast. One problem is introduced by the double periodicity of d_{Eigen} (see Fig. 1), so that each vector is treated like its antipole. This property should be taken into account particularly since both v_0 and its antipole minimise e_{Eigen}.

LS: Fisher [1] proposes a so-called sample spherical median by using the arc length as underlying distance $d_{\text{LS}} = d_{\text{arc}}$. Since an analytical solution to minimise the corresponding error function does not exist for $D > 2$, we use the scaled Levenberg-Marquart algorithm LMSDER of the GNU Scientific Library.

LSQ: Using the squared arc length as distance $d_{\text{LSQ}} = d^2_{\text{arc}}$ we get another error function, likewise without an analytical optimisation procedure.

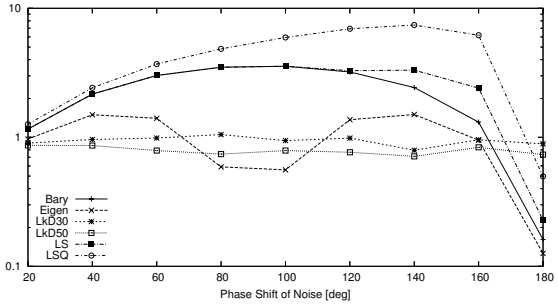

Figure 3. μ-Error depending on δ

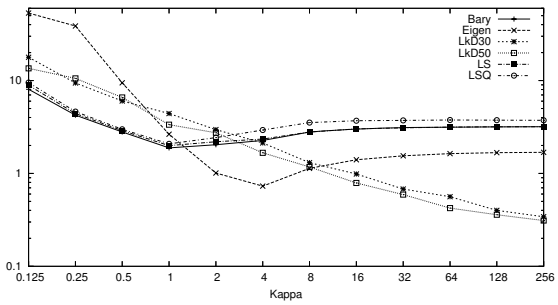

Figure 4. μ-Error depending on κ

2.2 Least Median Error

In this section a novel method of computing a representative vector is proposed. Like the generalised median approaches we have to solve an optimisation task based on an underlying distance function d. In contrast to the generalised median approach, now the median error has to be minimised. Let ψ_i denote the sorted distances $d(\vec{v}_0, \vec{v}_i)$. We estimate the representative vector $v_{\mathrm{L}k\mathrm{D}}$ which minimises ψ_k with $k = n/2$. This method is one of the most popular robust estimators with a breakdown point 0.5 [14].

Obviously, other choices of k are possible: $k = \lceil p \cdot n \rceil$ with $p \in (0, 1)$. In our study we use two different LkD-estimators, namely LkD50 ($p = 50\%$) and LkD30 ($p = 30\%$) in combination with d_{arc}.

Another interpretation of this task is the constrained optimisation problem of minimising $\psi := d(\vec{v}_0, \vec{w})$ under:

$$\sum_{d(\vec{v}_i, \vec{v}_0) \leq \psi} \frac{1}{n} \geq p \quad \Leftrightarrow \quad \sum_{d(\vec{v}_i, \vec{v}_0) > \psi} \frac{1}{n} \leq 1 - p$$

The error function to be minimised is defined as the distance $d(\vec{v}_0, \vec{w})$ of the representative vector v_0 and the k^{th} nearest sample $\vec{w} \in S$. For the optimisation task we use the simplex algorithm of Nelder and Mead [11].

There exists an algorithm to compute an exact solution v_0 of the LkD median. Utilising the fact that there have to be at least D points in S which are the k^{th} nearest neighbours to v_0 simultaneously, we inspect all subsets $S' \subset S$ of size D: Let H be the unique hyperplane, which includes all points of S', we consider the normal vectors $\pm\vec{n}$ as possible solutions for the optimisation task. Afterwards, we check if more than k sample points are lying in the same closed halfspace as $\pm\vec{n}$ bordered by H. A brute-force strategy can be realised to compute v_0 in $\mathcal{O}\left(k\binom{n}{D}\right)$ time. Beside the high complexity for large dimension this solution is numerical instable for concentrated data distributions.

2.3 Results

To define distributions on hyperspherical data usually the polar coordinate representation (in contrast to the Cartesian vector representation) is used:

$$\left(\theta^1, \ldots, \theta^{D-2}, \phi\right) \in [0, \pi)^{D-2} \times [0, 2\pi).$$

For our experiments we use spherical Fisher distributed $F(\mu, \kappa)$ data, with $\mu = (\hat{\theta}, \hat{\phi}) \in HS^2$ in 3D space. The Fisher distribution, which is a generalisation of the von Mises distribution defined on unit circle, plays an important role for the statistical analysis of directions. Important features of the Fisher distribution are the unimodality and the rotational symmetry around the mean direction μ [2]. The probability density function is given by $f(\theta, \phi) =$

$$C_F \exp\left(\kappa\left(\sin\theta \sin\hat{\theta} \cos\left(\phi - \hat{\phi}\right) + \cos\theta \cos\hat{\theta}\right)\right) \sin\theta$$

with $C_F = \frac{\kappa}{4\pi \sinh \kappa}$. The parameter $\kappa > 0$ is the shape parameter of the Fisher distribution. For $\kappa \to 0$ the Fisher distribution converges to the uniform distribution on HS^2, and for $\kappa \to \infty$ to the point distribution at μ.

Furthermore, we add noisy samples, with another underlying Fisher distribution $F(\nu, 0.25\kappa)$ and define $\delta := d_{\mathrm{arc}}(\mu, \nu)$, implying the angle between the mean directions μ and ν of the data and noise distribution. The portion of noisy samples in the data set is denoted by α. We use the method described by Fisher et al. [3] to generate Fisher distributed samples. For each configuration we run 50 test sets and compute the mean errors of the proposed estimators Bary, Eigen, LS, LSQ, LkD30, LkD50.

We have tested the approaches (Sec. 2.1 and 2.2) for estimating the mean direction depending on the parameter α, δ, κ and the sample size n. The standard configuration is set to $\alpha = 6.25\%$, $\delta = 60^{\mathrm{o}}$, $\kappa = 16.0$ and $n = 16384$ $(= 2^{14})$. In Figures 2, 3 and 4 the error (in degree) of estimating the mean direction is plotted against α, δ and κ, respectively.

In the case of undisturbed data ($\alpha = 0\%$) the generalised median methods outperform the LkD approaches, but even for the presence of a slight portion of noise ($\alpha = 1.0/64.0 = 1.5625\%$) the L$k$D approaches become comparable with increasing benefit for largerα (Fig. 2). The GM algorithms Bary, Eigen, LS and LSQ show a similar estimation quality, whereas Eigen performs the best of these. In the situation of a very large portion of noise ($\alpha > 30\%$)

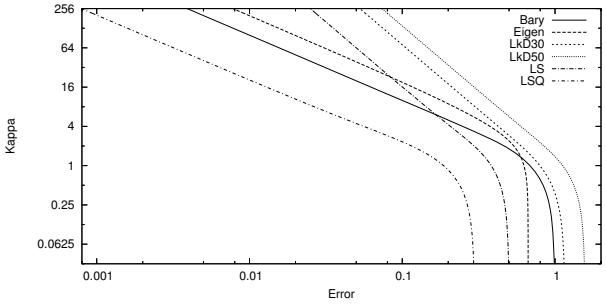

Figure 5. κ depending on the error values

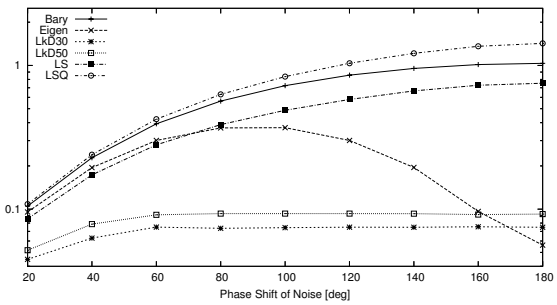

Figure 6. log(κ)-Error depending on δ

the optimisation algorithm for the LkD approaches (Fig. 2) sometimes fails and converges to a local minimum.

The influence of δ for the estimation of μ is negligible for the LkD approaches (Fig. 3). The Eigen estimator performs best for $\delta = 90^{\mathrm{o}}$ since the underlying distance function is low for such noise samples (Fig. 1). Bary, LS and LSQ again produce a much higher mean error than the LkD approaches and Eigen.

This behaviour is also visible by varying κ (Fig. 4). While the error functions of the LkD estimators decrease with increasing κ, those of Bary, LS and LSQ keep stable on a low level. The influence of κ to the quality of the Eigen estimator is remarkable. For medium κ ($\in [2, 8]$) the mean direction is estimated good, but for a different value of κ clearly worse.

For reasons of shortness the influence of the sample size is not presented since it does not discriminate the performance of the estimators. Concluding these results, it should be pointed out that in the situation of no noise, the GM estimators are more accurate than the LkD estimators. The advantage of the latter ones is the robustness against noise and the variation of κ, i.e. different distributions.

3 Estimation of the shape parameter κ

In this section we propose a method for estimating the shape parameter κ of the data distribution. For this purpose the same data model is used as proposed in Section 2.3. The task of estimating further parameters beside the mean is important to analyse the underlying data distribution. The best known example is the estimation of the standard deviation σ of normal distributed data.

We propose a two step strategy, which is applicable to all estimators of this paper. At first the mean direction μ is estimated and afterwards, we compute κ such that the expected error under the distribution $F(\mu, \kappa)$ is equal to the obtained error, which has been minimised.

To estimate κ, we use the functional relationship of κ and the expected error for the different approaches. Mathematical details are omitted due to space limitations. Based upon the obtained errors (which have been used for the optimisation task) we consider the estimation of κ as an inverse problem. In Figure 5 these relationship are shown by plotting κ against the error values. All functions are strictly monotonic decreasing, but for $\kappa \to 0$, the error values converge to those of a uniform distribution, which makes a good estimation numerically difficult and susceptible to noise for small values of κ.

In Figures 6 and 7 the errors of the different estimation strategies are shown. We use the absolute difference between the logarithm of the true and the estimated κ as error. This is due to the fact that the influence of κ is exponential (see the pdf of $F(\mu, \kappa)$). For both tests in Figure 6 and 7 the robust behaviour of the LkD estimators in comparion with the GM approaches is apparent. The error plots for varying the sample size and the portion of noise are omitted due to space restrictions.

4 Extensions

Each proposed method bases upon an optimisation problem. The influence of every data sample to the error function is equal to $1/n$. If the assumption of unit vectors is dropped, the data samples are not lying on a unit hypersphere anymore. For such a situation the methods could be extended by transferring the vector norm to the weight of the corresponding error summand.

Let $r_i = \|\vec{v}_i\|_2$ be the norms of the data samples $\vec{v}_i$. We project each sample to the hypersphere by $\tilde{v}_i := \vec{v}_i / r_i$ and set the weight of $\tilde{v}_i$ to $w_i = r_i / \sum r_i$. Instead of considering the equal weighted summations (Sec. 2.1) we minimise $\sum w_i \cdot d(\tilde{v}_0, \tilde{v}_i)$. In the situation of Section 2.2 we adjust the constraint to

$$\sum_{d(\tilde{v}_i, \tilde{v}_0) > \psi} w_i \ \leq 1 - p.$$

In the case of circular data, i.e. unit vectors in the plane, there are special algorithms for the optimisation problem, which have $\mathcal{O}(n \log n)$ time complexity. These algorithms utilise the unique circular ordering of points on a circle. After sorting the vectors with respect to their angle with the

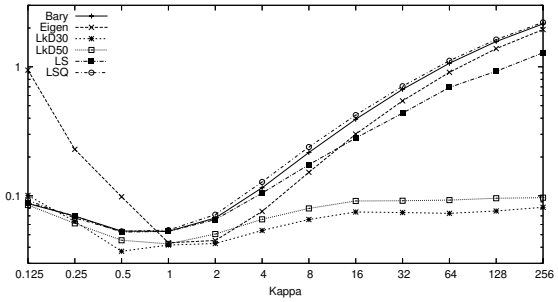

Figure 7. log(κ)-Error depending on κ

x-axis, the minimisation for LS and the LkD methods can be done with a single sweep through the ordered data set.

For all generalised median approaches a statistical analysis of an assumed distribution (with pdf p) is possible. This can be done by minimising the expected error for the underlying distance function d: $HS^{D-1} \times HS^{D-1} \rightarrow \mathbb{R}$:

$$v_0 = \arg \min_{v' \in HS^{D-1}} \int_{HS^{D-1}} d(v', v) \quad p(v)\, \mathrm{d}v.$$

A statistical analysis is also feasible for LkD. In that case the p-quantile $Q_p(v')$ of an assumed distribution model has to be inspected around $v' \in HS^{D-1}$. An optimal solution v_0 is characterised by:

$$v_0 = \arg \min_{v' \in HS^{D-1}} \max_{v \in Q_p(v')} d(v', v)$$

5 Discussion

We have presented different methods to compute representative vectors for hyperspherical data sets. The four generalised median approaches (Bary, Eigen, LS and LSQ) are good in situations of unnoisy data but could fail even for a small portion of noise. Therefore, the Bary and Eigen approach are favourable due to their direct computation. For dimensions $D > 2$, LS and LSQ have to be solved by an iterative optimisation process. The Eigen approach is in some cases more robust for the estimation of the representative vector than the other GM approaches and also gives better estimation for the shape parameter κ of a Fisher distribution. This better robustness of the Eigen approach is explained by the special character of the underlying distance function. The error function shows a maximal value at 90^{o}. Minimising the mean error effects a reduction of vectors orthogonal to the Eigen estimation $\vec{v}_0$ of mean direction. On the other side the vectors near the antipole of $\vec{v}_0$ are neglected. These properties are beneficial in the case of concentrated data distributions for separated noise and data samples. The results show that in other cases (e.g. for widespread data) this benefit is decreased.

The second estimator class presented in this paper is a generalisation of the least median error approach [14]. The benefits of the LkD methods are the scaleability ($k = \lceil p \cdot n \rceil$) and the robustness against noise of up to 50%. Unfortunately, a direct compution is only known for circular data ($D = 2$) and the derivations of the error function cannot be computed directly at all.

We have also tested these methods on circular real data using a Gaussian model [8] and a Von Mises model. It turns out that the LkD methods offer the best estimator also on this application. The task of this application is extracting elongated structures out of images showing heart tissue and estimating the preferred direction of these objects. For this we compute the tangent directions of the midlines at equidistant points as observation set S and estimate the preferred direction with the methods which are proposed in this work. Furthermore, a shape parameter has to be computed. It turns out that the LkD methods outperform the other estimators for all processed images.

References

[1] N. I. Fisher. Spherical medians. *J. R. Statist. Soc. B*, 47:342–348, 1985.

[2] N. I. Fisher. *Statistical analysis of spherical data*. Cambridge University Press, 1987.

[3] N. I. Fisher, A. J. Lee, and M. E. Willcox. Test of discordancy for samples from fisher's distribution on the sphere. *Appl. Stat.* , 30:230–237, 1981.

[4] C. Gramkow. On averaging rotations. *Int. J. of Computer Vision*, 42(1/2):7–16, 2001.

[5] A. Guimond, J. Meunier, and J. P. Thirion. Average brain models: A convergence study. *Computer Vision and Image Understanding*, 77(2):192–210, 2000.

[6] X. Jiang, A. Münger, and H. Bunke. On median graphs: Properties, algorithms, and applications. *IEEE Trans. on PAMI*, 23(10):1144–1151, 2001.

[7] R. Y. Liu and K. Singh. Ordering directional data: concepts of data depth on circles and spheres. *Ann. Stat.* , 20:1468–1484, 1992.

[8] P. P. Lunkenheimer, K. Redmann, N. Kling, K. Rothaus, X. Jiang, C. W. Cryer, F. Wübbeling, P. Niederer, S. Y. Ho, and R. H. Anderson. The three-dimensional architecture of the left ventricular myocardium. *Anat. Rec. (in press)*, 2006.

[9] L. Micó and J. Oncina. An approximate median search algorithm in non-metric spaces. *Pattern Recogn. Lett.* , 22(10):1145–1151, 2001.

[10] M. Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.* , 24(1):1–16, 2002.

[11] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer J.* , 7:308–313, 1965.

[12] T. B. Sebastian, P. N. Klein, and B. B. Kimia. On aligning curves. *IEEE Trans. on PAMI*, 25(1):116–125, 2003.

[13] C. G. Small. A survey of multidimensional medians. *Int. Stat. Review*, 58:263–277, 1990.

[14] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999.

[15] K. Subramanyan and D. Dean. A procedure to average 3d anatomical structures. *Med. Image Anal.* , 4(4):317–334, 2000.