Nonparametric Mean Shift Functional Detection in the Functional Space for Task and Resting-state fMRI

Jian Cheng, Feng Shi, Kun Wang, Ming Song, Jiefeng Jiang, Lijuan Xu, Tianzi Jiang

Nonparametric Mean Shift Functional Detection in the Functional Space for Task and Resting-state fMRI

Jian Cheng[1,2], Feng Shi[3], Kun Wang[1], Ming Song[1], Jiefeng Jiang[1], Lijuan Xu[1], Tianzi Jiang[1]
[1] LIAMA Research Center for Computational Medicine, Institute of Automation, Chinese Academy of Science
[2] Odyssée Project Team, INRIA Sophia Antipolis – Méditerranée, France
[3] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
jiancheng@nlpr.ia.ac.cn

Abstract. In functional Magnetic Resonance Imaging (fMRI) data analysis, normalization of time series is an important and sometimes necessary preprocessing step in many widely used methods. The space of normalized time series with n time points is the unit sphere $\boldsymbol{S}^{n-2}$, named the functional space. Riemannian framework on the sphere, including the geodesic, the exponential map, and the logarithmic map, has been well studied in Riemannian geometry. In this paper, by introducing the Riemannian framework in the functional space, we propose a novel nonparametric robust method, namely Mean Shift Functional Detection (MSFD), to explore the functional space. The first merit of the MSFD is that it does not need many assumptions on data which are assumed in many existing method, e.g. linear addition (GLM, PCA, ICA), uncorrelation (PCA), independence (ICA), the number and the shape of clusters (FCM). Second, MSFD takes into account the spatial information and can be seen as a multivariate extension of the functional connectivity analysis method. It is robust and works well for activation detection in task study even with a biased activation reference. It is also able to find the functional networks in resting-state study without a user-selected "seed" region. Third, it can enhance the boundary between different functional networks. Experiments were conducted on synthetic and real data to compare the performance of the proposed method with GLM and ICA. The experimental results validated the accuracy and robustness of MSFD, not only for activation detection in task study but also for functional network exploration in resting-state study.

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) has become a powerful technique to study spatial-temporal neural activity in human brain. A large number of methods have been proposed for fMRI data analysis, including both task and resting state analysis. In general, the existing methods could be categorized into two families: model-based methods and model-free (or data-driven) methods.

Model-based methods often parametrically fit a prior model to the data by statistical techniques, such as correlation analysis, variation analysis, t-test, linear regression, and so on. Among them, General Linear Model (GLM) [1] is a typical one, in which false positive ratio could be limited through selecting an appropriate P value

from hypothesis test. However, it has several critical limitations: (1) a linear addition model is assumed; (2) the residuals are assumed to follow a normal distribution, and so are the parameters; (3) the choice of the variables in the design matrix is sometimes arbitrary; (4) linear convolution is assumed and every voxel shares the same Hemodynamic Response Function (HRF). Another popular one is the functional connectivity method [2], which needs a user-selected 'seed' region of interest. It takes the mean time course in the seed region as the reference signal in the GLM. However, in most studies it is unclear how to select an appropriate seed region. Most model-based methods, including the GLM and the functional connectivity analysis, are univariate methods which assume every time series is independent of each other.

In model-free methods, effects or components of interest in the data are found based on some specific criterions. For example, Principal Component Analysis (PCA) [3] assumes that the structures of interest in the data are uncorrelated both in the temporal and spatial domains. Independent Component Analysis (ICA), including the spatial ICA [4, 5] and the temporal ICA [6], assumes that the data is a linear addition mixture of some independent sources. Another kind of model-free methods is clustering, including Fuzzy C-Means (FCM) [7], Gaussian mixture model [8], etc. These methods are based on the general physiological fact that the activities within a specific functional system have a certain similarity, which is the base of functional networks. Clustering methods partition the brain into some clusters based on some similarity measures. Model-free methods are multivariate and could gather more information and have less prior assumption than model-based methods. Moreover, model-free methods could be used for both activation detection in task study and functional networks analysis in resting-state study, which is difficult for model-based methods. However, there are still some intrinsic assumptions in model-free methods, such as linear addition, uncorrelation and independence which cannot be fully satisfied in the brain. How to determine an appropriate similarity measure and the number of clusters is also an open problem for clustering methods.

In both model-based and model-free methods, normalization of time series is an important and sometimes necessary preprocessing step. Normalization means that the time course is subtracted by its mean (centering) and then divided by its L2 norm (scaling). The space of normalized time series with n time points is a unit sphere S^{n-2}, named as the functional space by Friston [3]. In fact, the results of most methods, for which the normalization may not be necessary, do not change under normalization transform, which could be called as "normalization invariant". That means most of the current fMRI data analysis methods work in the functional space. In this paper, by introducing the Riemannian framework in the functional space, we propose a novel robust nonparametric Mean Shift Functional Detection method (MSFD) to explore the functional space.

2 Riemannian Framework in the Functional Space

In fMRI study, each normalized time course could be considered as a point scattered in the functional space $\boldsymbol{S}^{n-2}$, and the point is called as the shape of the time course. Every point $\boldsymbol{x}$ in the sphere is a representative element of an equivalence class in $\boldsymbol{R}^n$,

i.e. $\{\tilde{\boldsymbol{x}} = a\boldsymbol{x} + b \mid \forall a > 0, \forall b\}$. The distribution of these points embodies the functional topography [3]. The closer two points are, the higher their correlation (i.e. functional connectivity) is. In fact, the correlation is the cosine of the angle subtended at the origin (see Fig. 1). Considering in GLM, $\tilde{\boldsymbol{y}} = \beta\tilde{\boldsymbol{x}} + \varepsilon$, we have $\beta = r = cos\theta$. Model-based methods use the predefined shape of the HRF (a point on $\boldsymbol{S}^{n-2}$) to find a group of points that have close shape. Comparatively, model-free methods analyze the distribution and the structure of these points. PCA and ICA methods try to find some mapping directions to properly interpret the distribution. Clustering methods try to partition these points into some functional homogeneous groups based on some certain similarity measures. In summary, the most important issue in fMRI data analysis is how to describe and interpret the distribution and how to measure the degree of the "close", which needs a rigorous mathematical formulation.

Considering the sphere is a simple manifold that has been well studied in Riemannian geometry, we thus adopt the geodesic distance to describe how close two points are. Moreover, tangent space theory could be used for algorithms device [9]. In particular, assuming that $\boldsymbol{x}$ and $\boldsymbol{y}$ are two points, the geodesic is the part of the great circle between them. Formulae (1) (2) (3) give the geodesic distance, the exponential map and the logarithmic map. A pictorial representation is shown in Fig.1.

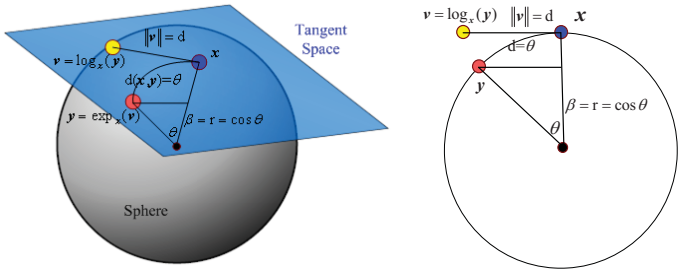

Fig. 1 Pictorial representation of the functional space. d: the geodesic distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. $\boldsymbol{v}$: a vector in the tangent space of $\boldsymbol{x}$. r: correlation coefficient. $\|\cdot\|$: L2-norm.

Geodesic distance: $d(\boldsymbol{x},\boldsymbol{y}) = \theta = arccos\,(r),\ r = cos(\theta) = \sum_i \boldsymbol{x}_i\boldsymbol{y}_i$ **(1)**

Exponential map: $\boldsymbol{y} = exp_{\boldsymbol{x}}(\boldsymbol{v}) = \boldsymbol{x}\,cos\,\theta + \boldsymbol{v}\,sin\,\theta\,/\|\boldsymbol{v}\|$, where $\theta = \|\boldsymbol{v}\|$ **(2)**

Logarithm map: $\boldsymbol{v} = log_{\boldsymbol{x}}\,\boldsymbol{y} = \theta(\boldsymbol{y} - \boldsymbol{x}\,cos\,\theta)/\|\boldsymbol{y} - \boldsymbol{x}\,cos\,\theta\|$ **(3)**

3 Mean Shift Functional Detection (MSFD)

As we discussed above, we formulate fMRI data processing as the problem of analyzing the structure of points in the functional space. A commonly accepted physiological fact is that time courses in the same functional area have covariant fluctuation, which forms the base of functional networks. In other words, they have close shapes (i.e. small geodesic distances). The functional space $\boldsymbol{S}^{n-2}$ is a very high dimensional unit sphere. Points in the functional space are sparsely scattered and relatively assemble in groups. If we consider these points as the observations from a

certain distribution, i.e. the probability distribution of the time course distributed in the whole brain, obviously hills appear in the area with many points, and valleys appear in the area with few points in the distribution on sphere. The idea of MSFD is to find the optimal representative time courses (peaks) and partition these points based on these peaks of different hills. Time courses in the same hill have close shapes (i.e. small geodesic distances). The spatial location distribution of these time courses within the same hill could be considered as a functional network (see Fig. 2).

Mean Shift (MS) is a powerful nonparametric clustering method for feature space analysis in computer vision and machine learning [10]. MS could iteratively move points towards the high density regions in the feature space and cluster data into some groups. Compared with other clustering methods, it is independent of the initialization, and can cluster data into arbitrary number of arbitrary shaped groups [10,12]. Traditional MS method is devised for low dimension Euclidean space (e.g. gray or color natural image). Two modified versions, respectively for low dimension analytic manifold [11] and high dimension Euclidean space [12], have been proposed recently. Considering the requirements in fMRI, we integrate these two versions of MS into the functional space, i.e. a high dimension manifold. If $\{\boldsymbol{x}_i, i = 1, \cdots, N\}$ are N points in the functional space, then the adaptive kernel density estimation (KDE) at $\boldsymbol{x}$ with the kernel profile $k(\cdot)$ could be expressed in (4). Define the mean shift vector $m(\boldsymbol{x})$ in the tangent space as (5). It could be proved that $m(\boldsymbol{x})$ is collinear with the gradient of the probability density function (PDF) at $\boldsymbol{x}$ [11]. $k(\cdot)$ is the kernel profile, and $\mathrm{g}(\mathbf{x}) = -k'(\boldsymbol{x})$. $k(\cdot)$ is chosen as the truncated Gaussian profile. So $g = k$. $h_{\boldsymbol{x}}$ is the kernel bandwidth at $\boldsymbol{x}$, which is adaptively determined as a half of the distance between $\boldsymbol{x}$ and its k^{th} nearest neighbor (KNN method). It has been proposed in high dimension Euclidean space that KNN adaptive MS is robust [12]. In this paper, the results of both synthetic and real experiments change little, even if k ranges from 500 to 1500. Therefore, we choose k=500 for computational efficiency.

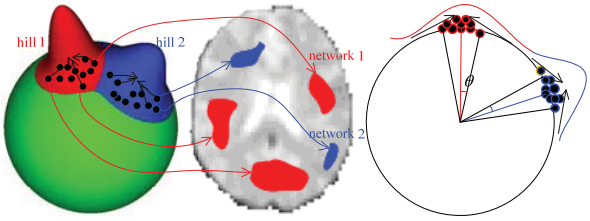

Fig. 2 A distribution of two hills in the functional space. Red hill and blue hill are partitioned by a valley. The black arrows are the movement directions of points. The red and blue arrows denote the locations of time courses in brain. They belong to two networks.

Adaptive KDE: $$\hat{f}(\boldsymbol{x}) = \frac{C_k}{n}\sum_{i=1}^{n} k\left(\frac{d^2(\boldsymbol{x},\boldsymbol{x}_i)}{h_{\boldsymbol{x}}^2}\right) \quad (4)$$

Mean Shift Vector: $$m(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} log_{\boldsymbol{x}}(\boldsymbol{x}_i) g\left(\frac{d^2(\boldsymbol{x},\boldsymbol{x}_i)}{h_{\boldsymbol{x}}}\right)}{\sum_{i=1}^{n} g\left(\frac{d^2(\boldsymbol{x},\boldsymbol{x}_i)}{h_{\boldsymbol{x}}}\right)} \quad (5)$$

MSFD could be used not only for activation detection in task study, but also for functional network analysis in resting-state study. Both of them are based on Mean Shift Iteration (MSI) that is given below. Two images can be obtained after MSI. One is a 4D functional image, which contains the time courses after MSI. The other one is

a 3D ***dist*** image for statistical post-processing, which denotes how far a point is away from itself in the original position in the functional space. After the MSI, these new $\boldsymbol{x}_i$ gather in some peaks of hills that are the optimal representative time courses. Thus if two points are belonged to two groups, the distance between them will be enlarged after MSI, which will enhance the boundary between different networks (Fig.2) [13].

Algorithm: Mean Shift iteration (MSI)

Input: time courses, $\boldsymbol{x}_i,\ i = 1, \cdots, N$

normalize $\boldsymbol{x}_i,\ i = 1, \cdots, N$

for $i \leftarrow 1, \cdots, N$

$\boldsymbol{x} \leftarrow \boldsymbol{x}_i\ \ dist(i) = 0$

Repeat :

Determine $h_{\boldsymbol{x}}$ (KNN)

$$m(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} log_{\boldsymbol{x}}(\boldsymbol{x}_i) g\left(\frac{d^2(\boldsymbol{x}, \boldsymbol{x}_i)}{h_{\boldsymbol{x}}}\right)}{\sum_{i=1}^{n} g\left(\frac{d^2(\boldsymbol{x}, \boldsymbol{x}_i)}{h_{\boldsymbol{x}}}\right)},$$

$$\boldsymbol{x} \leftarrow exp_{\boldsymbol{x}}\big(m(\boldsymbol{x})\big), \quad dist(i) \leftarrow dist(i) + \|m(\boldsymbol{x})\|$$

Until $\|m(\boldsymbol{x})\| < \varepsilon$

Output: save every new $\boldsymbol{x}_i$ and $dist(i),\ i = 1, \cdots, N$

MSFD for functional network analysis in resting-state study. We can estimate these representative time courses through the following process. *First*, many clusters could be generated by grouping together all the new x_i that are closer than a given small threshold $d_{th} = 0.05$ (correlation coefficient=0.9988). *Second*, many cluster centers are calculated by averaging new time courses in the same cluster. *Third*, these centers are sorted in descending order according to the number of voxels in the cluster. More voxels in the cluster, more believable the representative time course is. For every representative course $\boldsymbol{r}_k$ and every point $\boldsymbol{x}_i$ in the whole brain, a distance index d_{ki} could be calculated, $d_{ki} = d_{center} + dist$, where d_{center} is the distance between the new $\boldsymbol{x}_i$ and $\boldsymbol{r}_k$, Compared with GLM, an analogous T_{ki} value is used to measure how close a point $\boldsymbol{x}_i$ is to the $\boldsymbol{r}_k$.

$$T_{ki} = \sqrt{M-2}\cos(d_{ki}) / \sqrt{1-\cos^2 d_{ki}}, \quad \text{where } M \text{ is length of time course} \quad (\mathbf{6})$$

Therefore, for $\boldsymbol{r}_k$ there is a T map $\{T_{ki}, i = 1, \cdots, N\}$. If $k = 0$, MSI dose not start and $dist = 0$, $d_{ki} = d_{center}$. Every point is the representative time courses itself since there is just one point in every cluster, i.e. $\boldsymbol{r}_i = \boldsymbol{x}_i$. Then for every voxel $\boldsymbol{x}_i$, we have a *T* map. In that case, MSFD is the same as the functional connectivity method.

MSFD for activation detection in task study. Because the aim in task study is to find the activity region instead of all activity networks in the brain, it can be seen as a special case of whole brain network analysis. We just need to find the right hill (i.e. the activation hill) that the given point (the expected hemodynamic response) belongs to. So the computational cost is much lower than in resting-state study. In practical, the given hemodynamic response (HR) is usually biased by many factors, such as the noise, the assumption of different regions share the same HR, etc. A novel idea to avoid these limitations and make full use of the given biased reference is to consider it as a real time course in brain, and update it through MSI. We consider the corrected expected HR as the real activation reference for this subject. Obviously, different

subjects will have different references. However, these references are close if all subjects do the same task, which could be seen from the small ***dist*** values of all references. The corrected reference is considered as the representative time course of the activation hill. MSI is needed only for points which are in a certain cone (e.g. correlation coefficient > 0.05). For the points after MSI, $d_{ms} = d_{center} + dist$, where d_{center} is the distance between new $\boldsymbol{x}_i$ and the corrected reference. Analogously, T values (formula 6) could be calculated for these points. Thus we have an activity significant map, i.e. the T map, for the reference. It should be noted that if $k = 0$, MSI does not start and the reference does not update. The T map is just the same as the T map used in the GLM with the only one repressor, i.e. the reference.

4 Experiments on Synthetic and Real Data

Two synthetic and two real experiments are given to compare the proposed method with widely used GLM and ICA. For synthetic experiments, we utilize receiver operating characteristic (ROC) analysis [14] to evaluate MSFD, GLM (SPM2, http://www.fil.ion.ucl.ac.uk/spm/) and ICA (GIFT, http://icatb. sourceforge.net/).

In the first synthetic experiment, we will test the activity detection performance of three algorithms with the accurate reference. We use a resting-state data as null data [15]. Then we preprocess the null data by SPM2 including slice timing, realigning, EPI template normalization of $3 \times 3 \times 3$ voxel size and Gaussian smoothing of 4mm FWHM. Last, synthetic data is generated by replacing resting time courses in given position (two $10 \times 10 \times 10$ boxes, Fig.3(a)) with simulated response which is generated by mixing reference signal with white noise under a given contrast-to-noise ratio (CNR=0.4) [7]. Reference signal is simulated by a boxcar (24s rest and 26s stimulation, totals 7 cycles) convolved with the HRF that is a combination of two γ-functions in SPM2. With the accurate estimation of reference, GLM and ICA both achieve good performance (perfect classification), which is obvious because the activation signal generation satisfies the assumptions of GLM (linear addition, white noise) and ICA (spatially independent). Our MSFD is also a perfect classification even without these assumptions.

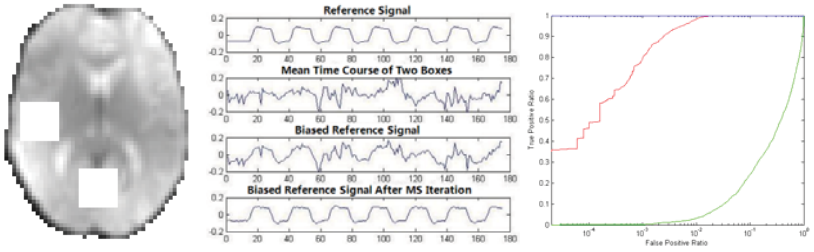

Fig. 3 ROC analysis of synthetic data. From left to right: (a) Illustration of block position. (b) the real reference, the mean time course of boxes, the biased reference and the corrected reference; (c) ROC curves of three methods with the biased reference.

The second synthetic experiment is to evaluate the robustness of methods by using a biased reference signal since the real activation reference cannot be accurately

estimated in real experiment. The average time course of the original resting data in the two boxes was added to the reference with CNR=1.0 to simulate the biased reference. Correlation coefficient between the real reference and the biased reference was 0.723. Fig. 3(b) shows that after MSI, the corrected reference is very close to the real reference (correlation coefficient is 0.96), which makes the activity detection result unchanged and validates the robustness of MSFD. The ROC curves of the three methods with the biased reference are given in Fig.3(c). It reveals that the performance of GLM and ICA relies on the reference estimation. A wrong component was chosen in ICA based on the biased reference, which got the lowest performance, although these components were decomposed independently of the reference.

In the first real experiment, the real fMRI data is the auditory bi-syllabic data from the SPM public dataset (http://www.fil.ion.ucl.ac.uk/spm/data/auditory.html). The preprocessing procedures are similar to that in synthetic data. Fig. 4 shows one slice of the results. These three methods all found the activated auditory regions. From the color bars of Fig. 4, we can find that MSFD have the biggest contrast, because it can enhance the boundary between different networks.

In the second real experiment, the real resting-state data that is the null data in the synthetic experiments is used to explore the functional networks. We can get the representative time courses after MSFD. And for each of them, we have a *T* map. Here we just show some of these *T* maps for the default-mode network, hippocampus, sensorimotor and visual area (Fig.5). The slice number is shown in each axial image.

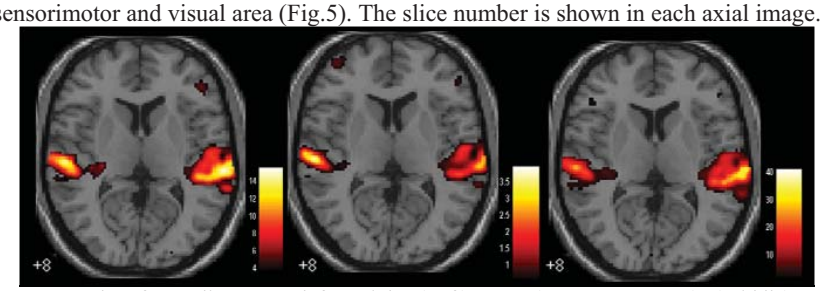

Fig. 4 Results of one slice. From left to right: (Left) GLM (SPM2, $p<0.0001$); (Middle) ICA (GIFT, $z> 0.6$); (Right) MSFD ($T>2$).

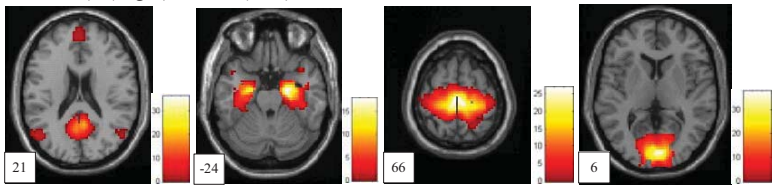

Fig. 5 From left to right: default-mode network, hippocampus, sensorimotor, visual area.

5 Conclusions

In this paper, we propose a nonparametric Mean Shift Functional Detection (MSFD) method based on the Riemannian framework in the functional space. Compared with other classic methods, GLM and ICA, MSFD does not assume the linear addition and

the independence. It is based on the intrinsic manifold formulation, does not depend on the initialization and does not assume the number and the shape of clusters. MSFD considers the local spatial information, and is a multivariate extension of the functional connectivity method. For every optimal representative time course, MSFD can get a *T* map, which is analogous with the *T* map in GLM. It can enhance the boundary between different functional networks, and robustly detect the optimal representative time courses of the data. It has a good performance validated by the experiments both on synthetic and real data. Therefore, it is a data-driven, robust method based on manifold formulation, which is appropriate for both the activation detection in task study and the functional network analysis in resting-state study.

References

1. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping **2** (1995) 189-210
2. Biswal, B., Zerrin Yetkin, F., Haughton, V.M., Hyde, J.S.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine **34** (1995) 537-541
3. Friston, K.J.: Eigenimages and multivariate analyses. In: Frackowiak, R.S.J. (ed.): Human Brain Function. 1st edn (1997)
4. McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J.: Analysis of fMRI data by blind separation into independent spatial components. Human Brain Mapping **6** (1998) 160-188
5. Beckmann, C.F., Smith, S.M.: Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE transactions on medical imaging **23** (2004) 137-152
6. Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J.: Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Human Brain Mapping **13** (2001) 43-53
7. Fadili, M.J., Ruan, S., Bloyet, D., Mazoyer, B.: A multistep Unsupervised Fuzzy Clustering Analysis of fMRI time series. Human Brain Mapping **10** (2000) 160-178
8. Golland, P., Golland, Y., Malach, R.: Detection of spatial activation patterns as unsupervised segmentation of fMRI data. MICCAI 2007, Vol. 4791. Springer, Heidelberg (2007) 110-118
9. Buss, S.R., Fillmore, J.P.: Spherical averages and applications to spherical splines and interpolation. ACM Transactions on Graphics **20** (2001) 95-126
10. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24** (2002) 603-619
11. Subbarao, R., Meer, P.: Nonlinear Mean Shift for Clustering over Analytic Manifolds. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, New York, **1** (2006) 1168–1175
12. Georgescu, B., Shimshoni, I., Meer, P.: Mean shift based clustering in high dimensions: a texture classification example. IEEE International Conference on Computer Vision**, Vol.1** (2003) 456-463
13. Cohen, A.L., Fair, D.A., Dosenbach, N.U., Miezin, F.M., Dierker, D., Van Essen, D.C., Schlaggar, B.L., Petersen, S.E.: Defining functional areas in individual human brains using resting functional connectivity MRI. Neuroimage **41** (2008) 45-57
14. Skudlarski, P., Constable, R.T., Gore, J.C.: ROC Analysis of Statistical Methods Used in Functional MRI: Individual Subjects. Neuroimage **9** (1999) 311-329
15. Lu, Y., Jiang, T., Zang, Y.: Region growing method for the analysis of functional MRI data. Neuroimage **20** (2003) 455-465