NONPARAMETRIC LEAST SQUARES REGRESSION
FOR IMAGE RECONSTRUCTION ON THE SPHERE

Tamara Tošić, Ivana Tošić and Pascal Frossard

Signal Processing Laboratory (LTS4)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne 1015 - Switzerland
email:{tamara.tosic, ivana.tosic, pascal.frossard}@epfl.ch

ABSTRACT

This paper addresses the problem of interpolating signals defined on a 2-d sphere from non-uniform samples. We present an interpolation method based on locally weighted linear and nonlinear regression, which takes into account the differences in importance of neighboring samples for signal reconstruction. We show that for optimal kernel function variance, the proposed method performs interpolation more accurately than the nearest neighbor method, especially in noisy conditions. Moreover, this method does not have memory limitations which set the upper bound on the possible interpolation points number, like in the method presented in [1].

Index Terms— Spherical Function, Interpolation, Signal Reconstruction, Nonparametric Model

1. INTRODUCTION

High dimensional signals with complex structures are usually hard to sample uniformly. For example, spherical images obtained from omnidirectional cameras are generally sampled non-uniformly due to the geometry of acquisition system. On the other hand, methods for compression or analysis of spherical signals are usually designed for regularly sampled data. Therefore, it is necessary to introduce a signal interpolation step between the data acquisition step and the compression or analysis blocks.

Interpolation values of the function to be reconstructed can be obtained by explicit modeling, or by using nonparametric modeling procedure. Classical parametric image processing methods rely heavily on the model of the signal. On the contrary, nonparametric techniques permit to estimate the signal or its significant components directly from the data [2]. The method is widely used in the field of pattern recognition and it can also be applied for the interpolation. For example, [3] proposes the use of Parzen windows or k-Nearest Neighbor estimation method for the estimation of the signal density.

Regression, and in particular linear regression modeling, is one of the most studied and widely used statistical tools. It has emerged lately as a useful tool in the domain of image processing. However, there are some cases when linear modeling can not be applied because of the intrinsic nonlinearity of the data. In such cases, nonparametric regression becomes more appropriate.

We address the problem of the reconstruction of a signal non-uniformly sampled on a 2-d sphere. In [1], a method is proposed for the reconstruction of a spherical signal from nonuniform samples using the Fourier transform on a 2-d sphere. We have observed

that smoother reconstruction results can be obtained in the case when weights are used to compensate for the different density of the samples caused by the geometry of the sphere. This observation has motivated us to develop interpolation methods that adapt to the importance of the different samples.

We therefore propose to extend data adapted kernel regression methods [4] to the interpolation of signals on the sphere. We propose a generic reconstruction scheme, that provides better performance than classical interpolation strategies when the sampling density becomes small, especially in a noisy environment.

This paper is organized as follows. Section 2 overviews the nonparametric kernel regression method for signal interpolation. The interpolation method is then formulated for the case of signals living on the 2-d unit sphere in Section 3. Section 4 shows experimental results. We conclude the paper with section 5.

2. INTERPOLATION BY KERNEL REGRESSION

2.1. Signal interpolation

In many signal processing applications we deal with a problem of interpolating signal values at predetermined positions from a set of signal samples acquired at different, non-uniformly spaced positions. Signal values at specific positions are usually required by the processing or display techniques. For example, filtering of spherical images is usually performed on an equiangular grid or a healpix grid[1]. However, spherical signals, such as omnidirectional images, star-shaped 3D point clouds, astrophysics data, etc., are typically sampled in a non-uniform manner. In such cases it is necessary to estimate the values of the function for desired reconstruction points. To achieve stable reconstruction one needs to carefully choose the signal samples that will be taken into account for the interpolation at a particular point. Global interpolation methods use all available samples to interpolate a signal value at a certain point, which makes them unstable in the presence of undersampled parts. Local methods, such as the one presented in this work, use only a set of *significant samples* for each interpolation point, which makes them more robust to undersampled data and less computationally expensive.

2.2. Nonparametric Kernel-Regression Method

We briefly overview the data interpolation method presented in [5] and adapted to image interpolation in [4]. For simplicity, we present the 1D case, but the method is also extendable to higher dimensions. Given the data y_i measured at points x_i, where $i = 1, \ldots, q$ and q

[1]http://healpix.jpl.nasa.gov

represents the number of samples, the goal is to estimate the value of the function (signal) z at a point x_i in a presence of noise ϵ_i:

$$y_i = z(x_i) + \epsilon_i. \tag{1}$$

The influence of noise can be interpreted as sampling the signal at a point $x_i + \Delta x_i$ instead of a point x_i, or as recording the signal value $z(x_i) + \Delta z(x_i)$ instead of the value $z(x_i)$. In both cases, analytic underlying function value has to be be estimated.

In the local interpolation framework, the signal value in some specific point is estimated from its values in neighboring points. Therefore, the authors in [5] propose to represent the unknown signal z with its Taylor expansion around the neighboring sample $z(x_0)$:

$$z(x_i) = \sum_{n=0}^{\infty} \frac{z^{(n)}(x_0)}{n!}(x_i - x_0)^n = \sum_{n=1}^{\infty} \frac{\beta_n}{n!}(x_i - x_0)^n,$$

where $\beta_n = z^{(n)}(x_0) = \partial^n z(x_0)/\partial x_i^n$ represent the Taylor expansion coefficients in the neighboring point x_0, and signal z is assumed to be infinitely differentiable in the neighborhood of x_0. The signal $z(x_i)$ is then approximated by taking the first k Taylor coefficients, and further estimated by finding the values $\beta_n, n = 0, ..., k$ that give the best fitting to the known data. Minimizing the Mean Square Error (MSE) to all samples can be one approach, but its main drawback is that all the data values, no matter if they are in the neighborhood of the interpolating point or not, will have equal impact on the final β_n values estimation. Authors in [5] propose to weight the samples according to their 'importance' (distance from the interpolation point) by a data-adaptive kernel K_{ad}, and they formulate a following minimization problem to estimate β_n for $k = 2$:

$$\min_{\{\beta_n\}_{n=0}^2} \sum_{i=1}^{q} \|y_i - \beta_0 - \beta_1(x_i - x_0) - \beta_2(x_i - x_0)^2\|^2 \cdot K_{ad}. \tag{2}$$

The kernel function can represent the sample influence based only on the distance for the interpolation point ($K_{ad} = K_L$), on the difference between the estimated and the current data values ($K_{ad} = K_R$), or both ($K_{ad} = K_{LR}$).

Dual principle in the probabilistic domain is well known as the Parzen window [3] and it has been broadly used in the pattern recognition. While in deterministic case the kernel function estimates the sample significance for the interpolation, Parzen window is used to estimate the probability density function of the data drawn from an unknown function. The kernel density estimator, similar to the weight estimator, can be any function $f(x) \geq 0$ with the property $\int f(x)dx = 1$, which ensures that the resulting probability distribution is nonnegative and integrates to one.

3. INTERPOLATION ON THE SPHERE

Specific geometry of the signal on the sphere requires formulating the minimization problem in Eq. (2) for the spherical coordinates. Therefore, we adapt the Taylor expansion to the geometry of the sphere and construct the spherical kernel functions accordingly.

3.1. Signal on the sphere and operator definitions

Points on the 2-d sphere are defined with the azimuth angle $\phi \in [0, 2\pi)$, the zenith angle $\theta \in [0, \pi]$ and radius $r > 0$. The same point representation in Cartesian coordinates (x, y, z) is related to the spherical coordinates by:

$$x = r\cos\phi\sin\theta, \ y = r\sin\phi\sin\theta, \ z = r\cos\theta.$$

The del operator on the sphere is given with: $\vec{\nabla} = \hat{r}\frac{\partial}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial}{\partial \theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial}{\partial \phi}$. The Laplacian is a scalar operation that is defined as: $\nabla^2 z = \vec{\nabla} \cdot (\vec{\nabla} z)$. A simple calculation transforms the Laplacian operator for a spherical case to:

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial \phi^2} + \frac{\cos\theta}{r^2\sin\theta}\frac{\partial}{\partial \theta} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}. \tag{3}$$

For simplicity, we will consider the unit sphere ($r = 1$). The Laplacian in that case is:

$$\nabla^2 = \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial \phi^2} + \frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial \theta} + \frac{\partial^2}{\partial \theta^2}. \tag{4}$$

3.2. Taylor expansion

In general, a differentiable N-dimensional real function can be represented through the Taylor series function as a sum of terms calculated from the values of its derivatives at a single point. For practical reasons, we assume that the function is well approximated if only first k coefficients have a nonzero values. Therefore, the assumption that a function on the sphere is k-differentiable function is sufficient to use this representation. Similarly as in [5], we assume that good approximation of the function is achieved for k=2.

For a 2-d unit sphere ($r = 1$), the Taylor function simplifies to:

$$z(\theta_0, \phi_0) = z(\theta_i, \phi_i) +$$
$$\frac{\partial z(\theta_0, \phi_0)}{\partial \theta_i}(\theta_i - \theta_0) + \frac{1}{\sin\theta_i}\frac{\partial z(\theta_0, \phi_0)}{\partial \phi_i}(\phi_i - \phi_0) +$$
$$\frac{1}{2!}\Big[\frac{1}{\sin^2\theta_i}\frac{\partial^2 z(\theta_0, \phi_0)}{\partial \phi_i^2}(\phi_i - \phi_0)^2 + \frac{\cos\theta_i}{\sin\theta_i}\frac{\partial z(\theta_0, \phi_0)}{\partial \theta_i}(\theta_i - \theta_0) +$$
$$\frac{\partial^2 z(\theta_0, \phi_0)}{\partial \theta_i^2}(\theta_i - \theta_0)^2\Big].$$

This equation can be also rewritten as:

$$z(\theta_0, \phi_0) = \beta_0 + \beta_{11}(\theta_i - \theta_0)(1 + \frac{1}{2}arctg(\theta_i)) + \beta_{12}\frac{\phi_i - \phi_0}{sin\theta_i}$$
$$+\beta_{21} \cdot \frac{1}{2}(\frac{\phi_i - \phi_0}{\sin\theta_i})^2 + \beta_{23}\frac{1}{2}(\theta_i - \theta_0)^2, \tag{5}$$

where each of the coefficients β_i represents the derivation of z at point (θ_0, ϕ_0): $\beta_0 = z(\theta_i, \phi_i)$, $\beta_{11} = \frac{\partial z(\theta_0, \phi_0)}{\partial \theta_i}$, $\beta_{12} = \frac{\partial z(\theta_0, \phi_0)}{\partial \phi_i}$, $\beta_{21} = \frac{\partial^2 z(\theta_0, \phi_0)}{\partial \phi_i^2}$, $\beta_{23} = \frac{\partial^2 z(\theta_0, \phi_0)}{\partial \theta_i^2}$.

3.3. Estimation of Taylor coefficients β_n

We want to estimate the value of the function z with help of the coefficients β_n in the Taylor expansion. The coefficients can be computed from the value of z at neighboring samples. This can be done by minimizing the following expression:

$$\min_b\{\sum_{i=1}^{q} \|y_i - z(\theta_0, \phi_0)\|^2 \cdot K_{ad}(\theta_i - \theta_0, \phi_i - \phi_0)\},$$

where $b = [\beta_0\ \beta_{11}\ \beta_{12}\ \beta_{21}\ \beta_{23}]$, and the function $z(\theta_0, \phi_0)$ depends on b as given by Eq. (5).

The Kernel function K_{ad} is a function on the 2-d sphere, which takes into account a distance between the interpolation and data points. A spherical kernel is constructed by choosing a 2D planar Kernel function and projecting it onto the sphere by inverse

stereographic projection. The stereographic projection is a method used to map a 2-d sphere to the plane tangential to the South pole. Each point (θ, φ) on the sphere is uniquely mapped to a point (x, y) on the plane with the following mapping: $x = 2tg\frac{\theta}{2}\cos\phi$, $y = 2tg\frac{\theta}{2}\sin\phi$. Inverse stereographic projection maps the point on the plane tangential to the South pole back to the sphere. Positioning the Kernel at a particular point on the sphere is easily performed with a single rotation in the SO(3) group.

The problem in Eq. (6) is now rewritten in matrix representation. Given q samples, the measurement data $M_i(y_i, \theta_i, \phi_i)$, $i = 1, \cdots, q$, can be represented with the matrices $\mathbf{y} = [y_1, y_2, \cdots, y_q]^T$, $\theta = [\theta_1, \theta_2, \cdots, \theta_q]$ and $\phi = [\phi_1, \phi_2, \cdots, \phi_q]$, which, respectively, represent sample values and spherical coordinates of samples positions. Matrices $\mathbf{X}_{\theta_0, \phi_0}$ and $\mathbf{W}_{\theta_0, \phi_0}$ defined in Eq. (6) and Eq. (7) respectively, represent the coefficients specific to the spherical problem, and values of the Kernel function at sample points:

$$\mathbf{X}_{\theta_0,\phi_0} = [x_1 \ x_2 \ x_3 \ x_4 \ x_5] \tag{6}$$
$$\mathbf{W}_{\theta_0,\phi_0} = diag[K_{ad}(\theta_1,\phi_1)K_{ad}(\theta_2,\phi_2)\cdots K_{ad}(\theta_q,\phi_q)] \tag{7}$$

where

$$x_1 = [1 \ 1 \cdots 1]^T$$
$$x_2 = [(\theta_1 - \theta_0)[1 + arctg\theta_1] \cdots (\theta_q - \theta_0)[1 + arctg\theta_q]]^T$$
$$x_3 = [\frac{\phi_1 - \phi_0}{sin\theta_1} \ \frac{\phi_2 - \phi_0}{sin\theta_2} \cdots \frac{\phi_q - \phi_0}{sin\theta_q}]^T$$
$$x_4 = [\frac{1}{2}(\frac{\phi_1 - \phi_0}{sin\theta_1})^2 \ \frac{1}{2}(\frac{\phi_2 - \phi_0}{sin\theta_2})^2 \cdots \frac{1}{2}(\frac{\phi_q - \phi_0}{sin\theta_q})^2]^T$$
$$x_5 = [\frac{1}{2}(\theta_1 - \theta_0)^2 \ \frac{1}{2}(\theta_2 - \theta_0)^2 \cdots \frac{1}{2}(\theta_q - \theta_0)^2]^T$$

represent vectors of a dimension $q \times 1$.

The weight function given in Eq. (7) represents the contribution of every data point to the estimated value of the current interpolation point. Logical choice of such a function is a probability density estimator, which is dependent on the total number of data q, the surface of the region containing the interpolation point and the number of sample points in this surface.

The optimization problem in Eq. (6) can be efficiently solved using the weighted least squares method, giving the estimate of the unknown coefficient vector $\mathbf{b} = b^T$ as:

$$\hat{\mathbf{b}} = \min_{\mathbf{b}} \|\mathbf{y} - \mathbf{X}_{\theta_0,\phi_0}\mathbf{b}\|^2_{W_{\theta_0,\phi_0}} \tag{8}$$
$$= \min_{\mathbf{b}}(\mathbf{y} - \mathbf{X}_{\theta_0,\phi_0}\mathbf{b})^T \cdot \mathbf{W}_{\theta_0,\phi_0} \cdot (\mathbf{y} - \mathbf{X}_{\theta_0,\phi_0}\mathbf{b}). \tag{9}$$

To check the convexity of the objective function:

$$f = \sum_{i=1}^{q}\{\|y_i - z(\theta_0, \phi_0)\|^2 \cdot K_{ad}(\theta_i - \theta_0, \phi_i - \phi_0)\}$$

we examine its first and second order derivatives. It can be shown that a global minimum exists (the function f is convex) in case where the weight matrix has positive values. In the current work, we use a Kernel which gives a dependency between the data points distance only. It is thus easy to show that the first order derivative of a function f with respect to the coefficients β_k for a second order Taylor series estimation leads to the following result:

$$\begin{bmatrix} \sum_i y_i \\ \sum_i y_i w_i a \\ \sum_i y_i w_i b \\ \sum_i y_i w_i c \\ \sum_i y_i w_i d \end{bmatrix} = \begin{bmatrix} w_i & w_i a & w_i b & w_i c & w_i d \\ w_i a & w_i a^2 & w_i ab & w_i ac & w_i ad \\ w_i b & w_i ab & w_i b^2 & w_i bc & w_i bd \\ w_i c & w_i ac & w_i bc & w_i c^2 & w_i cd \\ w_i d & w_i ad & w_i bd & w_i cd & w_i d^2 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{23} \end{bmatrix} \tag{10}$$

where w_i represents the K_{ad} value for i-th sample, and the coefficients are given with:

$$a = (\theta_i - \theta_0)[1 + \frac{1}{2}arctg\theta_i], b = \frac{\phi_i - \phi_0}{\sin\theta_i},$$
$$c = \frac{1}{2}(\frac{\phi_i - \phi_0}{\sin\theta_i})^2, d = \frac{1}{2}(\theta_i - \theta_0)^2.$$

A more compact version of this equation is $Y = M \cdot N$, where the solution involves calculation of the inversion matrix: $N = M^{-1} \cdot Y$. Further on, a second order derivative of the function f with respect to β_k coefficients gives the following result:

$$\frac{\partial^2 f}{\partial\beta_0^2} = 2w_i, \frac{\partial^2 f}{\partial\beta_{11}^2} = 2w_i a^2, \frac{\partial^2 f}{\partial\beta_{12}^2} = 2w_i b^2,$$
$$\frac{\partial^2 f}{\partial\beta_{22}^2} = 2w_i c^2, \frac{\partial^2 f}{\partial\beta_{23}^2} = 2w_i d^2.$$

If there exists a measurement point whose coordinates are the same as for the interpolation point, the function interpolation at this point is not calculated. Because of that, the coefficients a, b, c and d have nonzero values. Square of a nonzero value is always positive, which implies that a global minimum solution exists if the Kernel function values are all positive.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Setup

In this work, we use the 2-d Gaussian as a kernel function. It permits the weighting of samples based on the geodesic distance between the interpolation and the sampling point, which is not possible with a simple 2-d step function. Therefore, for each interpolation point, the set of neighboring points is selected as the points whose geodesic distance to the interpolation point is smaller than the kernel variance.

In order to project the Kernel function on the sphere, we use the inverse stereographic projection, to ensure the axi-symmetric selection of points. First, the interpolating point (θ_0, ϕ_0) is rotated to coincide with the South pole of the 2-d sphere. Further on, all the samples are rotated for the same angles. Gaussian 2-d kernel function on a plane tangent to the South pole is then projected to the sphere, thus resulting in an axisymmetric spherical Gaussian Kernel with a center at the interpolation point. Therefore, neighboring samples are weighted according to their distance from the interpolation point.

We examine the performance of the kernel regression method for zero, first and second order Taylor expansions. We chose to perform our experiments with omnidirectional images whose value points lie on the uniform grid. The resolution of the omnidirectional image used in experiments is 1024×1024, out of which 1024×326 pixels have non-zero values (recall that the omnidirectional image has blocks of zero values close to the poles). The unfolded non-zero region of the test image used in experiments is given on the Fig.1. We then remove points from the images, and we try to interpolate the value of the spherical function at the corresponding positions from the remaining samples. The number of interpolation points is respectively $16k$ and $33k$. To achieve statistically significant results we observed 10 randomly chosen interpolation sets per dataset.

For the comparison of the reconstruction values accuracy, we calculate MSE (mean of the spherical l_2 norm of the error) between the interpolated and the original image values. The performance of our method is compared with the nearest neighbor (NN) method, where five nearest neighbors are used for interpolation. The selected
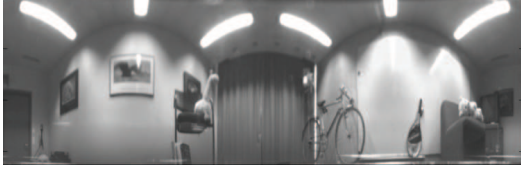
Fig. 1. Omnidirectional image

number of neighbors was shown to give the best performance of NN interpolation.

4.2. Experimental results

We first examine the performance of our method in the noiseless (clean) case. Results given in the Table 1 show that the proposed interpolation method gives better results than the NN method in terms of MSE for an optimal value of the kernel parameter. The second order Taylor expansion coefficients do not provide improvement over the first order ones.

We illustrate also the behavior of the reconstruction accuracy for different values of the kernel variance. We see that the selection of the kernel variance is important for this method, as it drives the number and the distance of the neighboring samples that participate in the interpolation of the missing point.

MSE values ($\cdot 10^{-4}$)						
	Kernel variance (σ)					
method	0.02	0.01	0.009	0.008	0.007	0.0069
zero	13	3.859	3.491	2.662	2.516	2.516
first	12	3.584	3.228	2.515	2.362	2.358
second	12	3.584	3.228	2.515	2.362	2.358
NN	2.435	2.435	2.435	2.435	2.435	2.435

Table 1. MSE values: $16k$ interpolation points, clean conditions.

Then, we perform interpolation in the presence of additive Gaussian noise and observe the influence of noise on the MSE values for high SNR values. The results are given in the Table 2. It can be observed that all proposed methods stay resilient to the noise in this case and that second and first order Taylor expansion methods give lower MSE then the NN method and zero order case. Note that these results are given for one randomly chosen set of interpolation points from the image.

MSE values ($\cdot 10^{-4}$)						
	SNR(dB)					
method	100	90	80	70	60	50
zero	2.516	2.516	2.515	2.516	2.517	2.530
first	2.358	2.358	2.358	2.358	2.360	2.372
second	2.358	2.358	2.358	2.358	2.360	2.372
NN	2.435	2.435	2.435	2.435	2.437	2.457

Table 2. MSE: noisy case, $16k$ sample points, kernel variance $\sigma = 0.0069$

The results on the Fig. 2 show the evaluation of the PSNR when the SNR values are lower (i.e. when the noise is more important). In order to get statistically significant results, the following result represent the mean value from ten independent experiments (for each

dataset we performed ten random data removal procedures). We see that the proposed Kernel method performs the interpolation more accurately than the NN method for all Taylor expansion orders and that the improvement in performance with respect to NN method increases when the noise energy is higher.
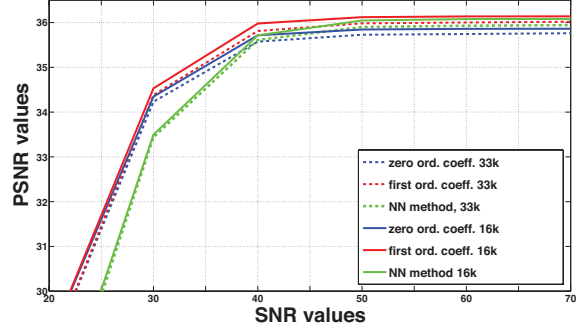

Fig. 2. PSNR value results for the $16k$ and $33k$ interpolation points in presence of additive Gaussian noise

5. CONCLUSIONS

This paper proposes a new method for interpolation of the points of a function on the sphere from non-uniformly sampled points. Our method achieves better performance in terms of the MSE than the baseline NN method, especially in the presence of noise. However, we note that the performance of the proposed method depends on the kernel variance, which is currently the topic of more detailed investigations. The Taylor series coefficients are calculated separately from a small set of neighboring points for every interpolation point and the method can use sampled points located arbitrarily. Therefore, there are no memory restrictions that would introduce the upper bound to the total number of interpolation points. This is one of the advantages of this solution compared to the reconstruction method that calculates Fourier coefficients on the 2-d sphere [1].

6. REFERENCES

[1] I. Tošić and P. Frossard, "FST-based Reconstruction of 3d-models from Non-Uniformly Sampled Datasets on the Sphere," *Proc. of Picture Coding Simposium*, 2006.

[2] C. Loader, *Local Regression and Likelihood (Statistics and Computing)*, chapter 2, Springer, 1999.

[3] R. Duda, P. Hart, and D. Stork, *Pattern classification*, chapter 4, A Willey-Interscience Publication, 2001.

[4] H. Takeda, S. Farsiu, and P. Milanfar, "Robust kernel regression for restoration and reconstruction of images from sparse noisy data," *Proc. of the International Conference on Image Processing*, 2006.

[5] A. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*, Clarendon Press - Oxford Science Publication, 1997.