A STATISTICALLY PROVEN AUTOMATIC CURVATURE BASED CLASSIFICATION PROCEDURE OF LASER POINTS

Fabio Crosilla, Domenico Visintini, Francesco Sepic

Department of Georesources & Territory, University of Udine, via Cotonificio, 114 I-33100 Udine, Italy – (fabio.crosilla)@uniud.it;-(domenico.visintini)@uniud.it;-(francesco.sepic)@e-laser.it

KEY WORDS: Laser scanning, Classification, Feature Recognition, Statistical analysis, Spatial modeling.

ABSTRACT

One of the critical aspects of the curvature based classification of spatial objects from laser point clouds is the correct interpretation of the results. This is due to the fact that measurements are characterized by errors and that simplified analytical models are applied to estimate the differential terms used to compute the object surface curvature values. In particular, the differential terms are the first and second order partial derivatives of a Taylor's expansion used to determine, by the so-called "*Weingarten map*" matrix, the *Gaussian* and the *mean* curvatures. Due to the measurement errors and to the simplified model adopted, a statistical procedure is proposed in this paper. It is based at first on the analysis of variance (ANOVA) carried out to verify the fulfilment of the second order Taylor's expansion applied to locally compute the curvature differential terms. Successively, the variance covariance propagation law is applied to the estimated differential terms in order to calculate the variance covariance matrix of a two rows vector containing the *Gaussian* and the *mean* curvature estimates. An F ratio test is then applied to verify the significance of the *Gaussian* and of the *mean* curvature values. By analysing the test acceptance or rejection for K and H, and their sign, a reliable classification of the whole point cloud into its geometrical basic types is carried out. Some numerical experiments on synthetic and real laser data finally emphasize the capabilities of the method proposed.

1. INTRODUCTION

Dealing with the laserscanning surveying technique, once the automatic points acquisition is carried out, the main methodological challenge is their automatic processing. Within the various computational procedures, a reliable geometrical classification of each laser point is investigated throughout this paper. The work fits in the recent researches conducted by the authors, whose analytical aspects have been mainly presented to the statisticians community (Crosilla et al., 2007) and whose laser scanning applications have been shown instead at various ISPRS events (Crosilla et al., 2004, 2005; Visintini et al., 2006; Beinat et al., 2007).

The procedure of automatic classification proposed by the authors is fundamentally based on the local analysis of the *Gaussian* K and *mean* H curvatures, obtained by applying a non parametric analytical model. In detail (chapter 2), the Z measured coordinates of each point is modelled as a Taylor's expansion of second order terms of X,Y local coordinates. The weighted l. s. estimate of the unknown vector, collecting the differential terms, is obtained by considering a selected number of neighbour points within a bandwidth radius and by applying a function taking into account their distance from the centre.

From the so locally estimated surface differential terms, the corresponding local *Gaussian* K and *mean* H curvature values are obtained, as well as the principal curvatures (chapter 3). As known, such curvatures are invariant to the reference frame.

Since the instrumental noise worsens the data quality and the analytical modeling simplifies the surface true form, the curvature values have to be statistically verified, namely also the variances of the estimated values have to be taken into account, as recommended by Flynn and Jain since 1988 and recently by Hesse and Kutterer (2005), these last specifically for the form recognition of laser scanned objects.

An analysis of variance (ANOVA) is first of all carried out in order to verify the fulfilment of the second order Taylor's expansion model (chapter 4). A Chi-Square ratio test is computed between the l.s. estimated variance factor and the a priori measurement variance. If the null hypothesis is rejected, the second order Taylor's series is not enough extended and a third order series is required. As known from literature (e.g. Cazals and Pouget, 2007), third order series can be applied to detect ridges and crest lines. If the null hypothesis is accepted, the statistical analysis is continued by applying the variance propagation law to compute the variance covariance matrix of the two terms vector containing the *Gaussian* and the *mean* curvature values. A Fisher ratio test is then applied to verify the significance of the obtained curvature values vector. If the null hypothesis is accepted, the surface can be locally accepted as planar. If the null hypothesis is rejected a ratio test for each K and H curvatures is carried out.

By simultaneously analyzing the sign and the values of K and H (chapter 5), a classification of the whole point cloud is indeed achievable, being possible the following surfaces basic types: hyperbolic (if $K < 0$), parabolic ($K = 0$ but $H \neq 0$), planar ($K = H = 0$), and elliptic ($K > 0$).

The subsequent automatic segmentation of each recognized surface in its geometrical elements is carried out by two complementary procedures (chapter 6): the first one by finding the surface analytical functions of each geometrical element and the second one by searching primarily the object edges.

The numerical testing of the proposed procedure has been carried out with satisfactory results for various simulated laser data belonging to the OSU Range Image database (Ohio State

University) and also for real data acquired with a Riegl Z360i laserscanning system on architectural surfaces (chapter 7).

2. ESTIMATION OF LOCAL SURFACE PARAMETERS BY A NON PARAMETRIC REGRESSION MODEL

Dealing with parameters estimation by regression models, the main advantage of a non parametric approach consists in its full generality: in our case, i.e. the local estimation of the bypassing surface through the laser points, it means that neither a priori knowledge of the point geometry nor the fitting analytical function is required. Let us consider the following polynomial model of second order terms (Cazals and Pouget, 2003):

$$Z_j = a_0 + a_1 u + a_2 v + a_3 u^2 + a_4 uv + a_5 v^2 + \varepsilon_j \qquad (1)$$

where the coefficients and the parameters are locally related to a measured value Z_j by a Taylor's expansion of the function $Z = \mu + \varepsilon$ in a neighbour point *i* of *j*, as:

$$a_0 = Z_{0_i}; \quad a_1 = \left(\frac{\partial Z}{\partial X}\right)_{X_i}; \quad a_2 = \left(\frac{\partial Z}{\partial Y}\right)_{Y_i}; \quad a_3 = \frac{1}{2}\left(\frac{\partial^2 Z}{\partial X^2}\right)_{X_i};$$

$$a_4 = \left(\frac{\partial^2 Z}{\partial X \partial Y}\right)_{X_i, Y_i}; \quad a_5 = \frac{1}{2}\left(\frac{\partial^2 Z}{\partial Y^2}\right)_{Y_i}; \quad u = (X_j - X_i); \quad v = (Y_j - Y_i)$$

with X_i, Y_i and X_j, Y_j plane coordinates of points *i* and *j*.
The parameters a_s $(s \neq 0)$ are the first and second order partial derivatives along X,Y directions at the *i*-th point of the best approximating local surface, collected in the [6 x 1] vector $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = [a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5]^T$$

where a_0 is the estimated function value at point *i*.
The weighted least squares estimate of the unknown vector $\boldsymbol{\beta}$ from a selected number of *p* neighbour points results as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{z} \qquad (2)$$

where (for j = 1, …, p):
X is the coefficient matrix, with *p* rows as:

$$\mathbf{X}_j = \begin{bmatrix} 1 & u & v & u^2 & uv & v^2 \end{bmatrix}$$

W is a diagonal weight matrix defined by a symmetric kernel function centred at the *i*-th point, with elements as:

$$w_{ij} = \left[1 - (d_{ij}/b)^3\right]^3 \text{ for } d_{ij}/b < 1 \quad w_{ij} = 0 \text{ for } d_{ij}/b \geq 1$$

where d_{ij} is the distance between the points *i*,*j* and *b* is the half radius (bandwidth) of the window encompassing the *p* closest points to *i*. The value of *b*, rather than the kernel function, is critical for the quality in estimating $\boldsymbol{\beta}$. In fact, the greater is the value of *b*, the smoother the regression function results, while the smaller is the value of *b*, the larger is the variance of the estimated value.

Rewriting model (1) in algebraic form as:

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \qquad (3)$$

and considering the vector $\boldsymbol{\beta}$ estimated by (2), the residual vector $\hat{\mathbf{v}}$ for the *p* points within the bandwidth is given as $\hat{\mathbf{v}} = \mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}$. This last allows computing the least squares variance factor $\hat{\sigma}_0^2$ at point *i* as:

$$\hat{\sigma}_0^2 = \frac{\hat{\mathbf{v}}^T \mathbf{W} \hat{\mathbf{v}}}{p - 6} \qquad (4)$$

For each point *i*, this local value has to be suitably evaluated, as will be better explained in chapter 4, in order to verify by a χ^2 test if it is comparable to the measurement laser noise or if it is sensible also to a systematic effect, due to limitations in the Taylor's expansion order.

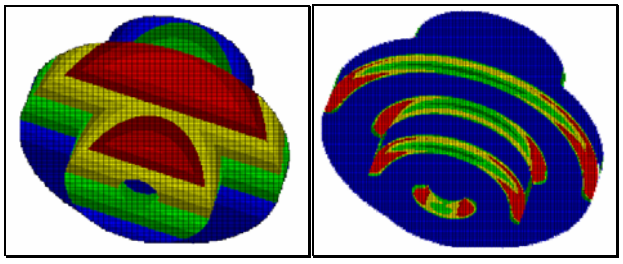


Figure 1: Simulated laser points of the *agpart-2* model (OSU database) coloured by Z_i values (at left) and by $\hat{\sigma}_0^2$ (at right).

Figure 1 reports the simulated scan *agpart-2* as example throughout the paper chapters: it belongs to the OSU Range Image database (Ohio State University). This synthetic object is composed by a cylinder, having a circular cavity in the axis, with a larger coaxial disk: the surfaces are thus cylindrical and planar. The simulated scan is oblique with respect to the object axis, as can be seen in Figure 1 at left, where the points are coloured by the original Z_i values from blue (minimum) to red (maximum). At right, the same points are coloured by the estimated values of $\hat{\sigma}_0^2$ from blue (zero) to red (maximum).

3. COMPUTATION OF LOCAL CURVATURES VALUES

For the local analysis of a surface obtained from a laser point cloud, some fundamental quantities defined in differential geometry are considered. In particular, local *Gaussian, mean* and *principal* curvatures values are taken into account. All these can be obtained from the so-called "*Weingarten map*" matrix **A** of the surface (e.g. Do Carmo, 1976), that is given by:

$$\mathbf{A} = -\begin{bmatrix} e & f \\ f & g \end{bmatrix}\begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1}$$

where E, F, and G are the coefficients of the so-called "*first fundamental form*", computable from a_s $(s \neq 0)$ parameters as:

$$E = 1 + a_1^2; \quad F = a_1 a_2; \quad G = 1 + a_3^2$$

and e, f, and g are the "*second fundamental form*" coefficients:

$$e = 2a_3 \Big/ \sqrt{a_1^2 + 1 + a_2^2} \quad ; \qquad f = a_4 \Big/ \sqrt{a_1^2 + 1 + a_2^2} \quad ;$$
$$g = 2a_5 \Big/ \sqrt{a_1^2 + 1 + a_2^2}$$

The *Gaussian* curvature K corresponds to the determinant of **A**:

$$K = \frac{eg - f^2}{EG - F^2} \tag{5}$$

The *mean* curvature H can be instead obtained from:

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)} \tag{6}$$

The *principal* curvatures k_{max} and k_{min}, corresponding to the eigenvalues of **A**, are given instead from the solution of the system $k^2 - 2Hk + K = 0$, i.e. from $k_{min,max} = H \pm \sqrt{H^2 - K}$.

Substituting the a_s terms into the formulas (5) and (6) (see e.g. Quek et al., 2003), the following expressions for the *Gaussian* K and the *mean* H curvatures can be obtained:

$$K = \frac{a_3 a_4 - a_5^2}{(a_1^2 + 1 + a_2^2)^2} \tag{7}$$

$$H = \frac{a_3(1 + a_2^2) + a_4(1 + a_1^2) - 2a_1 a_2 a_5}{2(\sqrt{a_1^2 + 1 + a_2^2})^3} \tag{8}$$

Summarizing, for each *i*-th point, four local curvature values K, H, k_{max} and k_{min} can be automatically obtained as functions of the vector $\hat{\boldsymbol{\beta}}$ terms. Furthermore, such curvatures are invariant to the adopted reference frame, providing a very much important property in analyzing the surface shape. Figure 2 shows the estimated K and H curvature values for the *agpart-2* scan: while constant values occur in central part of the various unit surfaces, very high curvature variations occur in buffer areas along the edges of the same surfaces.

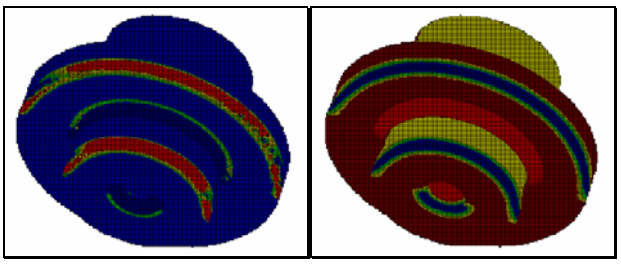


Figure 2: Points of the previous *agpart-2* scan coloured, from blue to red, for K (at left) and for H (at right) values.

4. STATISTICAL ANALYSIS OF THE ESTIMATED CURVATURE VALUES

As mentioned before, for each laser point *i*, the estimated local value of the variance factor, given by formula (4) as:

$$\hat{\sigma}_0^2 = \frac{\hat{\mathbf{v}}^T \mathbf{W} \hat{\mathbf{v}}}{p - 6}$$

is a quality index of the vector $\boldsymbol{\beta}$ estimation process. It is crucial to verify whether, within the bandwidth, the behaviour of the corresponding residuals $\hat{\mathbf{v}} = \mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}$ are due to the noise of the laser measures or rather to limitations in the non parametric Taylor's terms order. For such aim, the application of the following Chi-Square test is proposed, under the null hypothesis H0: $\hat{\sigma}_0^2 = \sigma_{tls}^2$.

$$\frac{\hat{\sigma}_0^2}{\sigma_{tls}^2}(p-1) \le \chi^2_{(p-1)_{1-\alpha}} \tag{9}$$

where:

σ_{tls}^2 is the variance of the terrestrial laser scanning (tls) instrument employed for the data acquisition;

$\chi^2_{(p-1)_{1-\alpha}}$ is the value of the Chi-Square distribution for (p-1) degrees of freedom when α probability for a first kind error is assumed.

The following analysis of the test results can be done, considering that, for most part of the points, the H0 hypothesis is accepted, as can be seen in Figure 3 at left:

H0 is accepted: a good local congruence between laser measures and second order Taylor's model is statistically proved. The values derived from vector $\hat{\boldsymbol{\beta}}$, as the K and H curvatures, are statistically meaningful and thus a curvature based classification can be carried out in such zones.

H0 is rejected: the local congruence between laser measures and the Taylor's model is not statistically fulfilled, i.e. a significant difference between the acquired laser data and the second order polynomial modeling is present. A part the reasons for this discrepancy, the derived curvature values in such zones have to be interpret with particular care.

Usually the values of $\hat{\sigma}_0^2$ significantly differ from σ_{tls}^2 along the edges of the laser scanned objects or along crests. This is explainable as a not sufficient modeling of the Taylor's order terms or as an improper choice of the bandwidth radius.

If the H0 hypothesis is accepted, the next problem is the statistical analysis of the local *Gaussian* and *mean* curvatures. To this purpose, once the least squares solution of the differential terms is obtained by means of (2), the variance covariance matrix of the estimated parameters is also available. The variance-covariance propagation law can be applied to the estimated $\hat{\boldsymbol{\beta}}$ terms to determine the [2 x 2] variance-covariance matrix of the *Gaussian* and *mean* curvature values. For such

end, let rewrite $\hat{\boldsymbol{\beta}} = [\hat{z}_0 \quad \hat{a}_1 \quad \hat{a}_2 \quad \hat{a}_3 \quad \hat{a}_4 \quad \hat{a}_5]^T$ as a partitioned estimated vector $\hat{\boldsymbol{\beta}} = [\hat{z}_0 \quad \hat{\mathbf{a}}]^T$ containing the function value z_0 and the sub vector **a** of the Taylor's expansion differential terms at point *i*. Let $\boldsymbol{\Sigma}_{\beta\beta}$ is the estimated variance-covariance matrix of vector $\hat{\boldsymbol{\beta}}$ terms; it can be yet partitioned as:

$$\boldsymbol{\Sigma}_{\beta\beta} = \begin{bmatrix} \sigma_{z_0}^2 & \boldsymbol{\sigma}_{z_0 a}^T \\ \boldsymbol{\sigma}_{z_0 a} & \boldsymbol{\Sigma}_{aa} \end{bmatrix} \tag{10}$$

where $\boldsymbol{\Sigma}_{aa}$ is the variance-covariance matrix of the sub vector **a** containing the differential terms at point *i*. As known, the variance-covariance matrix $\boldsymbol{\Sigma}_{\beta\beta}$ can be expressed as:

$$\boldsymbol{\Sigma}_{\beta\beta} = \hat{\sigma}_0^2 \mathbf{N}^{-1} = \hat{\sigma}_0^2 \begin{bmatrix} n_{z_0} & \mathbf{n}_{z_0 a}^T \\ \mathbf{n}_{z_0 a} & \mathbf{N}_{aa} \end{bmatrix}^{-1} = \hat{\sigma}_0^2 \mathbf{Q}_{\beta\beta} = \hat{\sigma}_0^2 \begin{bmatrix} q_{z_0} & \mathbf{q}_{z_0 a}^T \\ \mathbf{q}_{z_0 a} & \mathbf{Q}_{aa} \end{bmatrix}$$

where $\mathbf{Q}_{\beta\beta}$ is the covariance matrix of vector $\hat{\boldsymbol{\beta}}$, while $\hat{\sigma}_0^2$ is given from relationship (4).

Of course, the estimated K and H curvature values are not independent, as can be seen observing equations (5) and (6) or (7) and (8). In order to apply a significant test, considering also the correlation between the curvature values K and H, the following [2 x 1] vector is introduced:

$$\boldsymbol{\omega} = [K \quad H]^T \tag{11}$$

Applying the variance-covariance law propagation, the covariance matrix of vector $\boldsymbol{\omega}$ can be obtained as:

$$\mathbf{Q}_{\omega\omega} = \mathbf{F}_{\omega\omega} \mathbf{Q}_{aa} \mathbf{F}_{\omega\omega}^T \tag{12}$$

where:

$$\mathbf{F}_{\omega\omega} = \begin{bmatrix} \frac{\partial K}{\partial a_1} & \frac{\partial K}{\partial a_2} & \frac{\partial K}{\partial a_3} & \frac{\partial K}{\partial a_4} & \frac{\partial K}{\partial a_5} \\ \frac{\partial H}{\partial a_1} & \frac{\partial H}{\partial a_2} & \frac{\partial H}{\partial a_3} & \frac{\partial H}{\partial a_4} & \frac{\partial H}{\partial a_5} \end{bmatrix}$$

For the points where the null hypothesis of the Chi-Square test (9) is fulfilled, in order to verify whether the *Gaussian* and *mean* curvature vector $\boldsymbol{\omega}$ is significantly different from zero, i.e. $E(\boldsymbol{\omega}) \neq 0$, the following F ratio test must be satisfied (Pelzer, 1971) (see Figure 3 at right):

$$\frac{\boldsymbol{\omega}^T \mathbf{Q}_{\omega\omega}^{-1} \boldsymbol{\omega}}{r \, \hat{\sigma}_0^2} > F_{1-\alpha,r,\infty} \tag{13}$$

where:
r = rank ($\mathbf{Q}_{\omega\omega}$) = 2,

$F_{1-\alpha,r,\infty}$ Fisher distribution value for r and ∞ degrees of freedom and α probability for a first kind error.

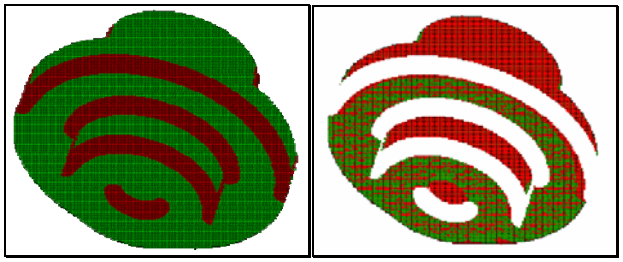


Figure 3: Points coloured by the results of the χ^2 test (at left) and the Pelzer test (at right): green where H0, red where H1.

If $E(\boldsymbol{\omega}) \neq 0$, it is worthwhile to independently test the values of K and H, in order to check if both, or just only one of them, are significantly different from zero. The null hypothesis is independently rejected for K and H, i.e. $E(K) \neq 0$, $E(H) \neq 0$, if:

$$\frac{K^2}{\hat{\sigma}_0^2 \, q_{kk}} > F_{1-\alpha,1,\infty} \qquad \frac{H^2}{\hat{\sigma}_0^2 \, q_{hh}} > F_{1-\alpha,1,\infty}$$

where q_{kk} and q_{hh} are the diagonal terms of matrix $\mathbf{Q}_{\omega\omega}$. Furthermore, these formulas are also useful to determine the minimal values of K and H that can be detected by the test, once a significance level α is fixed:

$$K > \hat{\sigma}_0 \sqrt{F_{1-\alpha,1,\infty} \, q_{kk}} \qquad H > \hat{\sigma}_0 \sqrt{F_{1-\alpha,1,\infty} \, q_{hh}}$$

Of course, K and H tend to diminish, i.e. the test becomes more sensitive, as $\hat{\sigma}_0$, q_{kk} and q_{hh} become smaller; that is if the precision of the laser measurements rises, the curvature values augments and the number of selected points, within a prefixed bandwidth, becomes greater.

This fact makes it possible to suggest, as a new topic of research, the fascinating concept of the *optimal design* of the laser survey, in order to reliably detect real curvature values. For instance, if the geometric characteristics of the surveyed object are approximately known and very rough curvature values K_0 and H_0 can be a priori defined, once the class of the instruments that are going to be used is fixed and the corresponding measurement precision σ_{tls} is known, a simulation procedure can be applied in order to find the minimal number of the bandwidth points that satisfy the following inequalities:

$$\frac{K_0^2}{\sigma_{tls}^2 \, F_{1-\alpha,1,\infty}} > q_{kk} \qquad \frac{H_0^2}{\sigma_{tls}^2 \, F_{1-\alpha,1,\infty}} > q_{hh} \tag{14}$$

The values q_{kk} and q_{hh} can be determined, for selected classes of bandwidth points, once approximate design parameters are fixed. Terms q_{kk} and q_{hh} correspond to the diagonal elements of matrix $\mathbf{Q}_{\omega\omega}$, computed from:

$$\mathbf{Q}_{\omega\omega} = \mathbf{F}_{\omega\omega}\mathbf{Q}_{aa}\mathbf{F}_{\omega\omega}^{T} = \mathbf{F}_{\omega\omega}(\mathbf{N}_{aa} - \frac{1}{n_{z_0}}\mathbf{n}_{z_0 a}\mathbf{n}_{z_0 a}^{T})^{-1}\mathbf{F}_{\omega\omega}^{T}$$

In order to reliably detect particular K and H curvature values, the class of bandwidth points satisfying inequalities (14) will be chosen. The test put in evidence the fact that it is necessary to strongly emphasize the design aspects of the laser survey.

5. CURVATURE BASED SURFACE CLASSIFICATION

By simultaneously analyzing the sign and the values of K and H, the classification of the whole point cloud is finally made possible. As known, each surface can be classified as one of the following types (see Table 4): hyperbolic (if K < 0), parabolic (K = 0 but H ≠ 0), planar (K = H = 0), and elliptic (K > 0).

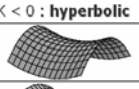
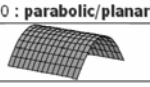
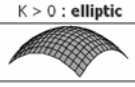
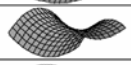
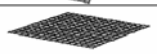
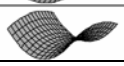
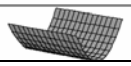
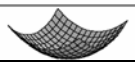
	K < 0 : **hyperbolic**	K = 0 : **parabolic/planar**	K > 0 : **elliptic**
H < 0			
H = 0			not possible
H > 0			

Table 4. Classification of surfaces according to the values of *Gaussian* K and *mean* H curvatures (from Haala et al., 2004).

When the null hypothesis H0: K = 0 is only satisfied, if H > 0 the single curvature surface can be classified as a concave parabolic valley, while if H < 0 as a convex parabolic ridge. Finally whether both null hypotheses are rejected, the surface is classifiable as a concave pit (if K > 0 and H > 0), as a convex peak (K > 0, H < 0), as a saddle valley (K < 0, H > 0), or as a saddle ridge (K < 0, H < 0).

Summarizing, this step allows not only to classify the various volumetric primitives but also to a priori define the polynomial degree of an interpolating parametric model applied for a refined segmentation of the points, as will be explained in 6.1.

6. AUTOMATIC SEGMENTATION OF EACH CLASSIFIED SURFACE

The segmentation of each recognized surface in its geometrical units can be carried out by two complementary procedures: by finding the analytical functions of each surface unit of the object or by directly searching the edges of such units.

6.1 Analytical modeling of the surface units

Within any kind of surface unit, classified as shown before, a region growing method is applied, starting from a random point not yet belonging to any recognized subset. The surrounding points having a distance less than the bandwidth *b* are analysed, by evaluating the values of the estimated height Z_{0_i} and the values of K and H. If the neighbour points present difference values within a threshold, then they are labelled as belonging to the same class and putted into a list. The same algorithm is repeated for each list element, till this is fully completed. Afterwards, the procedure restarts again from a new random point, ending when every point has been analysed. A first raw segmentation of the whole dataset is so carried out: each cluster represents an initial subset to submit to a refining segmentation. For this aim, we now suppose that laser measures can be rightfully represented by the parametric model:

$$\mathbf{z} - \rho\mathbf{W}\mathbf{z} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \qquad (15)$$

where:
z is the vector of laser height/depth values, as for the non parametric model (1);

ρ is a value that measures the mean spatial interaction between neighbouring points;

W is a spatial adjacency (binary) matrix, defined as $w_{ij} = 1$ if the points are neighbours, $w_{ij} = 0$ otherwise;

A is a r column matrix with $\mathbf{A}_i = \begin{bmatrix} 1 & X_i & Y_i & \dots & X_i^s & Y_i^s \end{bmatrix}$ as rows, where X_i and Y_i are X,Y-coordinates of points approximated by a *s* degree orthogonal polynomial;

$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \dots & \theta_{r-1} \end{bmatrix}^T$ is a [r x 1] vector of parameters;

ε is the vector of normally distributed noise errors, with mean 0 and variance σ_ε^2.

Discerning about the differences between the non parametric model (3) and the parametric model (15) applied for processing the same laser points, we can observe that:

the unknown parameters (mainly) involve local differential terms (**β**) of a whatever (and not estimated) function in model (1), while they correspond to the polynomial parameters (**θ**) of the best interpolating global analytical function in model (15);

in both cases, the coefficient matrix involves X and Y (planimetric) coordinates, expressed by relative values with respect to the local reference point for the non parametric case, and by absolute values for the parametric one;

in both cases, the **W** weight matrices consider the distance among the laser points, although with very different geometric and stochastic significance.

To solve equation (15), a Maximum Likelihood (ML) estimation of the unknown parameters is carried out: in particular, the value ρ_{ML} giving the maximum log-likelihood value is assumed as the ML estimation $\hat{\rho}$ of ρ. In this way, the optimal estimation of the SAR unknowns is given by:

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{I} - \hat{\rho}\mathbf{W})\mathbf{z} \qquad (16.1)$$

$$\hat{\sigma}^2 = n^{-1}(\mathbf{z} - \hat{\rho}\mathbf{W}\mathbf{z} - \mathbf{A}\hat{\boldsymbol{\theta}})^T(\mathbf{z} - \hat{\rho}\mathbf{W}\mathbf{z} - \mathbf{A}\hat{\boldsymbol{\theta}}) \qquad (16.2)$$

Within the **z** values, the individual departures from the fitted polynomial trend surface can be estimated by the vector $\mathbf{e} = \sigma^{-1}\boldsymbol{\varepsilon}$ of standardised residuals, computed from (15) as:

$$\mathbf{e} = \hat{\sigma}^{-1}[(\mathbf{I} - \hat{\rho}\mathbf{W})\mathbf{z} - \mathbf{A}\hat{\boldsymbol{\theta}}] \qquad (17)$$

Afterwards, its components are inferentially evaluated to find which measures do not fit the estimated trend surface. The so-called "Forward Search" (FS) (e.g. Cerioli and Riani, 2003) algorithm is applied. It makes possible to execute the robust estimations $\hat{\rho}$ and $\hat{\boldsymbol{\theta}}$ at each step of the search, starting from a partition of the dataset. The basic idea of the FS approach is to repeatedly fit the postulated model to subsets of increasing size, selecting for any new iteration the **z** observations best fitting the previous subset, that is having the minimum absolute value of **e**. Thanks to this growing strategy, the outlier data are potentially included only at the end of the FS process. To understand at which iteration the outlier data enter into the subset, a *F*-Fisher test is continuously applied. If the null hypothesis is rejected, any new point included from now on is an outlier: thus, there is no reason to continue with the FS iterations.

Therefore, after the estimation by formulas (16) of the analytical fitting function of each surface unit, the points segmentation is fulfilled: the edges among the so detected surfaces can be indirectly estimated by means of analytical 3D intersections.

6.2 Edge detection of the geometrical units

To directly detect the edges of each geometric unit, the attention is now focused onto the estimated values of the local *mean* curvature H. The analysis of H values exploits the property that such index is closely related to the first variation (slope) of a surface area that locally well reveals possible edges. Since H is the average of k_{max} and k_{min}, it is numerically slightly less sensitive to the noise with respect to the K curvature, which is instead the product of k_{max} and k_{min}.

Extreme absolute values of H correspond to surface slope discontinuities and then it is sufficient to locally evaluate for each point if the H absolute value is greater than a certain threshold, this last fixed taking into account also the local $\hat{\sigma}_0^2$ value. If this happens, the corresponding buffer volume of points reveals the edges we are looking for.

7. NUMERICAL EXPERIMENTS

The numerical testing of the proposed procedures has been carried out with satisfactory results for the *agpart-2* model and for other synthetic objects of the OSU Range Image database, averagely constituted by about 30.000 simulated points. In particular, the *column1* model is composed by a cylindrical column over a parallelepiped base, upper closed by a circular plane: the scan *column1-5* simulates a pointing-down laser acquisition from a scanning position so that the vertex among three planes of the base occurs at right (see Figures 5 and 6).

Figure 5 at left shows the estimated local values of $\hat{\sigma}_0^2$, coloured again from blue (0 values) to red (maximum values): as expectable, the most part of the plane and cylindrical surfaces have null variance, while along the edges such values dramatically increase. In spite of this, $\hat{\sigma}_0^2$ is not null also along some surface borders, but only because there is a lack of data for the backside surfaces. The result of the Chi-Square test (9) is reported in Figure 5 at right: the red areas, where the test fails, should be not submitted to the succeeding Pelzer test (13).

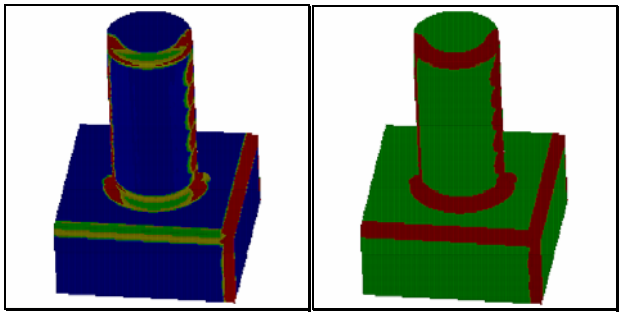


Figure 5: Experiments on the column1-5 model: $\hat{\sigma}_0^2$ values (at left) and χ^2 test results (at right), green where H0 holds.

Figure 6 shows the estimated local curvature values. Since only single curvature surfaces are present, the value of the *Gaussian* K should be always null, i.e. E(K) = 0. This condition is represented by green coloured points in Figure 6 at left. The *mean* H curvature values are instead correctly less than zero for the points belonging to the cylindrical column, as can be seen in Figure 6 at right, where such value is represented by dark yellow colour. Analyzing the border buffer areas, it is also possible to recognize the convex edges in blue (where H << 0) and concave edges in red (where H >> 0).

A last consideration arises from the obtained results: the width of the edge buffer areas meanly corresponds to the bandwidth radius; this might be reduced by choosing a smaller value for it.

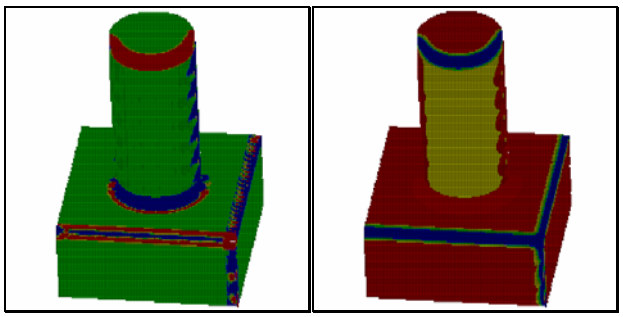


Figure 6: Estimated curvature values for *column1-5* model: *Gaussian* K values (at left) and *mean* H values (at right).

Last but not least, to test the proposed procedure in noisy conditions, some experiments have been performed onto real data acquired with a Riegl Z360i laserscanning system in the Aquileia Basilica (Italy), whose complete surveying is described in Visintini et al. (2006). The obtained results over about 26.000 points relating to a real column with a complex and irregular basement are here briefly reported.

In this case, it is particularly hard to evaluate the correctness of the estimated curvature values, since only the column has a regular surface, while in the basement, the surfaces are only roughly planar, with unequal bricks and ruined stones, as can be well seen in Figure 7 at left. Anyway, the K estimated values are mainly equal to zero, i.e. the yellow points in this false color representation (Figure 7 at centre), with very instable values in correspondence of the edges. In the plotting of the H values (Figure 7 at right), it is possible to well recognize the convex (blue) and the concave (red) edges.

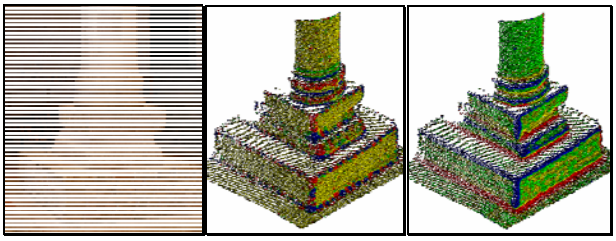


Figure 7: A column of the Aquileia Basilica (Italy): estimated *Gaussian* K (at centre) and *mean* H values (at right).

8. CONCLUSIONS

The paper reports a statistical procedure able to automatically detect reliable *Gaussian* and *mean* curvature values from laser point clouds, computed by applying a local surface non parametric Taylor's expansion. First, the fulfilment of the Taylor's terms is verified by an analysis of variance, comparing the variance factors obtained from the residuals of the estimation process and from the instrumental noise. A second test considers the variance-covariance propagation law applied to the estimated Taylor's terms, in order to compute variances and covariances of the *Gaussian* and *mean* curvature values. If the null hypothesis of the F test applied is rejected, some curvature values are significantly different from zero and their sign analysis allows to correctly classify the surface geometry of each object surface unit. The carried out experiments confirm the capabilities of the proposed procedure.

REFERENCES

Beinat, A., Crosilla, F., Visintini, D., Sepic, F., 2007. Automatic non parametric procedures for terrestrial laser point clouds processing, in: *IAPRS&SIS*, XXXVI, part 3/W49B, Munich, 1-6.

Cazals, F., Pouget, M., 2003. Estimating differential quantities using polynomial fitting of osculating jets, in: *Proceedings of the 1st Symposium on Geometry Processing*, 177-187.

Cazals, F., Pouget, M., 2007. *A generic C++ package for estimating the differential properties on sampled surfaces via polynomial fitting*. Rapport de recherche n. 6093, INRIA, France.

Cerioli, A., Riani, M., 2003. Robust methods for the analysis of spatially autocorrelated data. *Statistical Methods & Applications*, 11, 334-358.

Crosilla, F., Visintini, D., Prearo, G., 2004. A robust method for filtering non-ground measurements from airborne LIDAR data, in: *IAPRS&SIS*, XXXV, B3, Istanbul, 196-201.

Crosilla, F., Visintini, D., Sepic, F., 2005. A segmentation procedure of LIDAR data by applying mixed parametric and nonparametric models, in: *IAPRS&SIS*, XXXVI, 3/W19, Enschede, 132-137.

Crosilla, F., Visintini, D., Sepic, F., 2007. An automatic classification and robust segmentation procedure of spatial objects. *Statistical Methods & Applications*, 15, 329-341.

Do Carmo, M.P., 1976. *Differential geometry of curves and surfaces*. Prentice-Hill International, London.

Flynn, P.J., Jain, A.K., 1988. Surface classification: Hypothesis testing and parameter estimation, in: *Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition*, 261-267.

Haala, N., Reulke, R., Thies, M., Aschoff, T., 2004. Combination of terrestrial laser scanning with high resolution panoramic images for investigations in forest applications and tree species recognition, in: *IAPRS&SIS*, XXXIV, part 5/W16, Dresden.

Hesse, C., Kutterer, H., 2006. Automated form recognition of laser scanned deformable objects, in: Sansò, F., Gil, A.J., (eds.): *Geodetic deformation monitoring: from geophysical to engineering roles*. IAG Symposia, 131, Springer.

Ohio State University: *OSU MSU/WSU Range Image Database*, http://sampl.ece.ohio-state.edu/data/3DDB/RID/index.htm.

Pelzer, H., 1971. *Zur analyse geodätischer deformationsmessungen*. DGK, Reihe C, 164, München.

Quek, F.K.H., Yarger, R.W.I., Kirbas, C., 2003. Surface parametrization in volumetric images for curvature-based feature classification. *IEEE Transactions on Systems, Man, and Cybernetics*, part B: Cybernetics, 33, 758-765.

Visintini, D., Crosilla, F., Sepic, F., 2006. Laser scanning survey of the Aquileia Basilica (Italy) and automatic modeling of the volumetric primitives, in: *IAPRS&SIS*, XXXVI, part 5, Dresden, 6 pag. (on CD)