

# **A Portable Modular System for Automatic Acquisition of 3D Objects**

N. Alberto Borghese<sup>1</sup> and Stefano Ferrari<sup>2</sup>

<sup>1</sup> Laboratory of Human Motion Analysis and Virtual Reality (MAVR),

Istituto Neuroscienze e Bioimmagini – CNR. Milano, LITA. via f.lli Cervi, 93 – 20090 Segrate

(Milano) - Italy.

<http://www.inb.mi.cnr.it/Borghese/Borghese.html>

<sup>2</sup> Center for Bioengineering, Fondazione ProJuventute, Department of Electronics and Information,

Politecnico di Milano, P.zza Leonardo da Vinci, 32, 20133 Milano, Italy.

**Keywords:** Scanner 3D, Neural Networks, Radial Basis Functions, Modular Architectures, Range Data, Real-time, Multi-resolution.

## **Abstract**

*A portable and flexible system for 3D scanning is here presented. It is constituted of four main modules. The first module is devoted to the acquisition of a set of 3D points over the surface through laser scanning (digitization). The second module reconstructs a continuous 3D surface, filtering the measurement noise. Whenever required (e.g. in CAD applications), a third module converts the surface into a 3D mesh which can then be colored by projecting over it a bitmap of the surface, obtained from a snapshot to obtain a highly realistic textured 3D model. This instrument improves upon the commercially available scanners in two main aspects. The digitizer proves to be highly flexible and accurate, and it can easily accommodate objects of different dimension. The construction of the surface and the filtering of the digitization noise are performed in a single step through a fully adaptive algorithm which produces a multi-scale surface and can be parallelized to work in real time. Results on the reproduction of human faces are reported and discussed.*

## **1. Introduction**

Virtual 3D models are required by an increasing number of applications ranging from basic image processing to video conferencing, constructive and plastic surgery, 3D fax, reversed engineering and 3D CAD (virtual prototyping). A host of devices (3D scanners), have come to the market in the last few years to provide these models. Essentially a 3D scanner captures the 3D visible surface of an object as a mesh suitable to be processed by CAD and graphical systems [1]. Although ultrasound [2] or mechanical (e.g. Microscribe<sup>TM</sup>) devices are available, optical technology is preferred because it allows a high resolution and it does not require any contact with the surface. The gold-standard is represented by the Cyberware<sup>TM</sup> scanners which are suitable to most applications. However, apart

from their very high cost, they have two drawbacks: a complex structure, inside which the object is placed, has to be set up, and only objects within limited size ranges can be digitized.

In this paper, a modular 3D scanner, which is devoted to opposite criteria is described (Fig. 1). The first module is devoted to the acquisition of a set of 3D points over the surface through laser scanning (digitization). The second module reconstructs a 3D continuous surface which filters out the noise introduced by the digitization process. At this stage, a 3D model is already available. For some applications in CAD/CAM and computer graphics, the representation of the surface in the form of a mesh is more suitable; this is done in a third module, which converts the surface into a mesh. In a fourth module, a bitmap of the surface, obtained from a snapshot, is aligned and projected over the 3D mesh to obtain a highly realistic textured 3D model. The aim of this paper is to present the overall structure, focusing on the accuracy of the digitization process and on the reconstruction of the 3D surface. The overall processing allows achieving a high accuracy and can be implemented onto PC boards to achieve real-time 3D modeling. The system has been widely tested in the reconstruction of 3D objects and of human faces in particular.

INSERT FIGURE 1 HERE

## **2. Data acquisition**

The first step towards creating a 3D model is the digitization of a set of 3D points on the object's surface, which is carried out here by the Autoscan system introduced in [3]. This is constituted of a commercial laser pointer of 5mW of power, a set of video cameras, which provide each an image of  $256 \times 256$  pixels with a frame rate of 100 frames/second, a real-time image processor and a host computer. The processor computes a cross-correlation between a  $6 \times 6$  mask, template of the laser spot, and each image of the stream coming from the video cameras. This allows achieving a high

SNR (a laser spot can be detected also in outdoor conditions) and sub-pixel accuracy (by averaging the over-threshold pixels of the laser spot, weighted with their cross-correlation value), which has been experimentally determined in 0.1 pixel [4]. Digital cameras can therefore be used as a valid alternative to analog sensors for spot detection (cf. [5]). The pattern matching procedure allows also to automatically discard those spots which are poorly seen: in this case, cross-correlation is over-threshold only for one row and/or column of the image, losing sub-pixel accuracy in the computation of the spot center. The cross-correlation is computed in real-time through a custom board which contains essentially delay lines, adders and multipliers [6]; it can be replaced by less expensive general-purpose hardware (e.g. on FPGA board), without degrading the performances. The 3D position of the laser spot is computed with standard photogrammetric procedures from the spot position on two cameras. For this purpose, the cameras are calibrated by surveying a rigid bar with two spherical markers on its extremities, the same size of the laser spot, inside the volume where the object is scanned [7]. This makes the system extremely simple to set-up. We explicitly remark that there are no constraints on the set-up and on the size of the object to be scanned: this depends only on the lens used and the relative position of the cameras. Flexibility, scalability and portability therefore characterize the acquisition module.

Surface digitization is carried out by moving the laser pointer manually (Fig. 1a). To help in directing the laser beam, a real-time feedback is provided on the host PC monitor. This scanning procedure offers the great advantage to increase the number of measured points in those regions which are richer of spatial details (where the surface has the highest spatial frequency content), achieving a denser sampling there. The digitization accuracy has been experimentally assessed similarly to [5], and the results are reported here (Fig. 2). The position of 5,000 laser spots projected over a plane positioned in the center of the working volume is analyzed. For a medium-size volume ( $0.8\text{m} \times 0.6\text{m} \times 1\text{m}$ ), the accuracy is of  $0.27\text{mm} \pm 0.0063\text{mm}$  (rms error), averaged over 100 different experiments, with the two cameras at an angle close to ninety degrees ( $86.9$

degrees). This figure scales linearly with the dimension of the working volume, and it is consistent with the accuracy reported in the motion analysis field [7]. The distribution of the measurement error is zero-mean and normally distributed (cf. Fig. 2b) as it is the outcome of several error sources (quantization error in the video cameras, numerical approximation in the cross-correlators, non-uniform spot appearance, etc.).

At the end of the digitization process, a set of  $N$  3D points is obtained  $\{P_j = (x_j, y_j, z_j)\}$  (cf. Fig. 3a). At the same time, through a standard CCD camera a color snapshot of the surface is acquired (cf. Fig. 3b), which is converted into a bitmap of  $576 \times 768$  (PAL standard), with 24bits/pixel. The sustained peak acquisition rate is of 100 spots/second. This rate decreases when the spot is not visible to the cameras. In real acquisitions, an average rate of 90–95spots/second is obtained. By using arrays of laser pointers in place of a single pointer, this rate could be increased.

INSERT FIGURE 2 HERE

### 3. Surface reconstruction through the HRBF model

The conversion of the cloud of 3D points into a 3D geometrical model is the critical operation of any 3D scanner. Due to measurement noise a direct tessellation of the data points obtained by simply connecting them, would produce an undesirable wobbling surface (Fig. 3c). Such a surface is useless to graphics, CAD or any other application, and the need of some sort of filtering is evident.

When the shape of the object surface is known, model fitting can be a valuable solution. In this approach a parameterized surface model is progressively adapted to the range data by minimizing their distance from the surface (e.g. [8] for human faces). This approach is improved when features can be extracted from the range data [9] (cf. also [10, 11]). However these approaches are heavily based on predefined models and lack in generality. Generic parametric shapes (semi-parametric fitting), which are warped to fit the data topology, offer a more general solution [12]. The simplest of

these is the 2D lattice, proposed in the seventies in the connectionist domain, as a tool to represent multivariate data distributions. This model, called Self-Organizing Maps (SOM) [13], has been more recently applied to reconstruct 3D surfaces from range data [14]. Its limitation lies in the requirement that the surface topology is homomorphic to the lattice structure (usually a plane). When this is not the case, twisting of the plane and poor approximation arise. Superquadrics have been introduced as more powerful semi-parametric geometrical models [9]. Introducing a B-spline surface on the top of superquadric [9] or meshes [15] can further refine the obtained surface. This approach, termed “surface on surface”, gives rise to a weak hierarchical structure: the superquadric is adapted first to capture the overall shape of the data, and the spline patch is used to deform the superquadrics surface to better fit the data locally, by reproducing the surface details. To reconstruct complex surfaces, more than one superquadric can be employed [16].

All these approaches are iterative and cannot operate in real-time on the data points. A novel different approach is presented here aimed to produce 3D models in real-time. The reconstruction is achieved with a sparse approximation with adaptive variable resolution spatial filtering, where the correlation between neighbor data points is exploited to eliminate noise introduced by the sampling process. It is based on the Hierarchical Radial Basis Function Network (HRBF) model, proposed originally in the connectionist domain as a tool for multi-scale signal processing [17]; it is extended here to surface approximation from scattered data. In the following, the overview of the HRBF model is presented first along with the robust determination of its parameters. The hierarchical, multi-scale, structure is then introduced to make the approach fully adaptive both in resolution and scale.

INSERT FIGURE 3 HERE

### **3.1 Construction of a 3D surface through approximation**

We assume that the surface can be represented as a function  $S(P): R^2 \rightarrow R$ . This assumption is motivated by the shape of the human face which can be described as a height field (2½D). However this does not lack in generality as full 3D surfaces can be recovered by zippering multiple 2½D models [18]. Under this hypothesis, it is more convenient to reframe the data points collected with the scanning procedure as:  $\{(x_j, y_j, z_j) \mid j = 1, \dots, N\}$  as the set  $\{(P_j, S(P_j)) \mid P_j = (x_j, y_j) \in R^2; S(P_j) = z_j \in R; S: R^2 \rightarrow R\}$ .

Our final goal is the reconstruction of the true surface up to the measurement error. That is, the distance between the points reconstructed by the HRBF model, and those sampled on the surface, should be distributed according to the measurement error. The reconstructed surface  $S_r(P)$  should satisfy the following conditions:

$$\frac{1}{N} \sum_{j=1}^N (S_r(P_j) - S(P_j)) = E(v) = 0 \quad (1a)$$

as the measurement noise is zero mean (Section 2.), and:

$$\frac{1}{N} \sum_{j=1}^N (S_r(P_j) - S(P_j))^2 = \sigma^2_n \quad (1b)$$

which can be evaluated, for a given set-up, through the experiment reported in Section 2. As the measurement error is uniform, (1) should be satisfied not only on the entire input domain, but also locally in sub-regions of it (cf. Section 3.4).

### 3.2 The HRBF network

The HRBF model combines the output of many simple units to achieve the reconstruction of a complex surface. In particular, the HRBF network is composed of radially symmetric Gaussian units:

$$g(P; \mathbf{m} | \mathbf{s}) = \frac{1}{\sqrt{p\mathbf{s}}} e^{-\frac{(P-\mathbf{m})^2}{\mathbf{s}^2}} \quad (2)$$

where  $P, \mathbf{m} \in R^2$  and  $\mathbf{s} \in R$ . In the HRBF model, the units are organized in layers (subnetworks), where the  $k^{\text{th}}$  layer is composed of equally spaced Gaussians, which have the same standard deviation,  $\mathbf{s}_k$  (cf. Figs. 4 and 6). Therefore the surface,  $S_r(P)$ , is constructed by adding the contribution of a few grids of Gaussian functions, where, as shown in Fig. 5, each grid operates at a certain scale (or cut-off frequency).

Given a set,  $\mathbf{a}$ , of parameters which characterize the HRBF network, the actual shape of the reconstructed surface (i.e. the output of the net),  $S(P|\mathbf{a})$ , is:

$$S(P|\mathbf{a}) = \sum_{l=1}^L \sum_{k=1}^{M_l} w_{kl} g(P; P_{kl} | \mathbf{s}_l) \quad (3)$$

where  $L$  is the number of grids and  $M_l$  is the number of Gaussian units in the  $l^{\text{th}}$  grid.  $w_{kl}$  is the weight associated to the  $k^{\text{th}}$  Gaussian in the  $l^{\text{th}}$  grid, and  $P_{kl}$  is its position in the grid.  $\mathbf{s}_l$  is the standard deviation of all the Gaussians in the  $l^{\text{th}}$  grid which determines the filtering scale of that grid. The parameters  $\mathbf{a} = \{L, M, \{w_{kl}\}, \{P_{kl}\}, \{\mathbf{s}_l\}\}$  determine the actual shape of  $S(P|\mathbf{a})$ . Although these parameters could be determined by global optimization, e.g. [19], the time required to obtain a reasonable solution has suggested exploring alternative solution schemes. HRBF networks, in particular, offer a very fast solution as the determination of the parameters in (3) does not require any iteration, and it is performed with only local operations on the data points. This makes this approach particularly suitable to 3D scanning application.

### 3.3 Determination of the parameters in each layer

Each grid,  $l$ , of the HRBF model realizes a low-pass filter, which is able to reconstruct the surface up to a certain scale, determined by  $\mathbf{s}_l$ . It can be shown that  $\mathbf{s}_l$  and the spacing between two consecutive Gaussians on the same grid,  $DP_l$ , are related with:

$$\mathbf{s}_l = 1.465 DP_l \quad (4)$$



This relationship is obtained by accepting a maximum attenuation in the pass band of -3dB and a minimum attenuation in the stop band of -40dB [20]<sup>1</sup>. Different attenuation values lead to a different proportionality constant between  $\mathbf{DP}_l$  and  $\mathbf{s}_l$ . To apply the Gaussian filter, the data should be equally sampled in correspondence to the grid crossings  $\{k\}$ . That is, the set  $\{z_{k,l}\} = \{S(P_{kl})\}$  should be available. Unfortunately this is not the case here where the data are unevenly sampled. However, since many points are usually digitized (surface oversampling), a reliable estimate of  $S(P_{kl})$  can be obtained through the following weighted mean estimate (cf. also [21]):

$$\tilde{S}(P_{kl}) = \frac{\sum_{P_r \in A(P_{kl})} S(P_r) e^{-\frac{\|P_{kl} - P_r\|^2}{\mathbf{s}_l^2}}}{\sum_{P_r \in A(P_{kl})} e^{-\frac{\|P_{kl} - P_r\|^2}{\mathbf{s}_l^2}}} \quad (5)$$

where  $A(P_{kl})$  is the Receptive field associated to the  $k^{\text{th}}$  Gaussian in the  $l^{\text{th}}$  grid. It is set, somewhat arbitrarily, as the square region centered  $P_{kl}$  of side  $\mathbf{DP}_l$ . The estimate in (5) is carried out locally on the input space, and it can be parallelized to achieve quasi-real-time processing.

The grid filter can be now written as:

$$S_r(P) = \sum_{k=1}^{M_l} \tilde{S}(P_{kl}) g(P; P_{kl} | \mathbf{s}_l) \Delta P_l^2 \quad (6)$$

Comparing (6) with (3) it can be demonstrated that the parameters  $\{w_{kl}\}$  can be obtained simply as  $\tilde{z}(P_{kl}) \mathbf{DP}_l^2$  [20].

### 3.4 The hierarchical multi-scale structure

If only one grid was adopted, a serious drawback is introduced: the Gaussian scale should be small enough to resolve the finest details. This requires very dense packing of the Gaussians also in

---

<sup>1</sup> These values produce a cut-off frequency of  $v_{\text{cut-off}} = 0.1874/\sigma_l$  and a transition band of  $[0.1874/\sigma_l, 0.7327/\sigma_l]$ .

those regions where the details can be resolved at a coarser scale, causing a waste of resources and overfitting in those space regions. Moreover, as the sampled point local density reflects the local richness of details, those regions which lack high-frequency details will contain points spaced apart far, and receptive fields  $A(P_{kl})$  may come out empty there. A better solution would be to adaptively allocate the Gaussian units, with an adequate scale and, consequently, adequate receptive fields, in the different space regions. The novel idea is to analyze the local residual and to stack non-complete grids over a first grid at a coarse scale, obtaining what is termed a sparse approximation (cf. Fig. 6). The first grid will output a rough estimate of the surface,  $a_1(P)$  (Fig. 5a) as:

$$a_1(P) = \sum_{k=1}^{M_l} \tilde{S}(P_{kl}) g(P; P_{kl} | \mathbf{s}_l) \Delta P_l^2 \quad (7)$$

The residual  $\{r_1(P_j)\}$  is computed for each sampled data point,  $(P_j, S(P_j))$ , as:

$$r_1(P_j) = S(P_j) - a_1(P_j) \quad (8)$$

This residual will be the input to a second grid which features a smaller scale than the first one; somewhat arbitrarily we have chosen  $\mathbf{s}_2 = \mathbf{s}_1/2$  which is the same choice made in wavelet decomposition [22]. This second grid does not need to reconstruct the original surface but only the residual one. Its output,  $a_2(P)$ , will be a reconstruction of the residual  $r_1(\cdot)$  at the scale  $\mathbf{s}_2$ , and it will provide a second residual:

$$r_2(P_j) = r_1(P_j) - a_2(P_j) = S(P_j) - a_1(P_j) - a_2(P_j) \quad (9)$$

This grid will not be full, but Gaussians will be inserted only when a poor approximation is given.

This is evaluated through the residual itself: a Gaussian is inserted in the grid crossing  $P_{kl}$  only if:

$$\frac{\sum_{P_r \in A(P_{kl})} r_1^2(P_r)}{N_k} < \mathbf{s}_n^2 \quad (10)$$

where  $\sigma_n^2$  is the noise variance and  $N_k$  is the number of sampled points which belong to  $A(P_{kl})$ . A mathematical proof of HRBF convergence can be found in [23] and goes beyond the scope of this paper. Grids are created one after the other, until the condition in (10) is not true over the entire input domain (Figs. 5a–d).

INSERT FIGURE 4 HERE

#### **4. Conversion into a mesh and color application**

When required, the reconstructed continuous surface can be sampled as densely as desired (at 1mm in Fig. 7a) to obtain a mesh of triangular elements. The sampling could be optimized according to topological or geometrical criteria to obtain a lighter mesh [24]. When the color appearance is important (e.g. in computer graphics applications), a bitmap obtained from the snapshot of the surface is applied. This is achieved by using standard texture mapping procedures after having aligned it to the mesh through a semi-automatic graphical tool [25, 26]. The final result is a 3D very realistic reconstruction of human face (Figs 2b and 7b). A detailed description of this procedure goes beyond the scope of this paper.

#### **5. Results**

The system has been intensively used in scanning human body parts and faces in particular. This is a particularly difficult task because the face's spatial frequency content is highly variable. Three to four layers are usually sufficient. A typical result is reported in Figs 5 and 7 for Stefano's face. In the set-up used the angle between the cameras was of  $\approx 50$  degrees with an estimated rms measurement error of 0.65mm. The surface is reconstructed starting from  $N = 17,080$  data points collected in three minutes. As can be seen in Fig. 6, in the higher layers the units are inserted in clusters, only in

those regions where the face has the highest frequency content: the HRBF network automatically adapts its local cut-off frequency and resolution to the local frequency of the data. Only 10,355 Gaussian units have been used versus the 21,750 units, which were required by a complete Gaussian filter at the smallest scale. The quantitative results are reported in Table I, where the residuals and the parameters of each layer are shown. As it can be seen, the mean error on the reconstructed surface (bias) is already very close to zero in the second layer, while the standard deviation decreases down to 0.72mm in the fourth layer. The distribution of the residuals becomes progressively Gaussian shaped and approaches that of the digitization error as can be seen in Figs. 5.

INSERT FIGURE 5 HERE

INSERT FIGURE 6 HERE

INSERT FIGURE 7 HERE

## **6. Discussion**

The 3D scanner presented here is based on a digitization module characterized by scalability, flexibility and accuracy. In fact, acquisition of objects of complex topology can be easily accomplished by zippering multiple acquisitions carried out with a single pair of cameras placed in different positions and orientations [27] or increasing the number of video cameras used simultaneously. The use of pattern recognition to compute the spot's centroid gives a spot accuracy higher than that of commercial 3D scanners [1]. This figure decreases when the angle between the pair of video cameras, used to reconstruct the 3D position of the spot, differs from ninety degrees with approximately a quadratic curve [28, 29]. In real situations, where the angle between the cameras has to be negotiated with the amplitude of the common field of view of the two cameras, an

angle between 40 and 60 degrees is usually adopted, with an accuracy decrease between 1.5 and 3 times. When higher accuracy is required, smaller working volumes can be set-up or the angle between the cameras increased.

Given the relatively long acquisition time, such a high accuracy can be reached only when still objects are scanned. Therefore, when scanning human faces, maximum comfort of the subject has to be provided to avoid non-intentional movements, which do decrease the accuracy. These small movements could be eliminated by using at least three circular markers solid with the head, of the same dimension of the laser spot; they would be recognized by the pattern matching processor, and used to compensate head motion. Given the high accuracy achieved, this procedure has not been adopted, but it could suggest a possible implementation in which moving objects can be scanned.

This high accuracy is preserved in the reconstruction of the 3D surface, which is achieved by a novel hierarchical model: the HRBF network. The requirement of only operations carried out locally on the data points with no iterations is the main characteristic, which makes it suitable to real-time implementations. Few algorithms aimed to real-time, can be found in the literature. The closest approach is the hierarchical reconstruction based on B-splines [21] where the reconstructed surface is the sum of a set of intermediate surfaces generated through a hierarchy of control grids. The main difference is that, in the HRBF model, the grids in the superior layers are not complete, but Gaussian units are inserted in clusters where the residual is over threshold, with a large saving in resources. The same adaptive unit allocation schema offers to HRBF an advantage versus Multi-Resolution Analysis (MRA) carried out through wavelet decomposition [22]. The MRA works in a direction opposite to HRBF as it produces a hierarchy of approximating surfaces from fine to coarse. Each surface is the sum of a very low-frequency surface and a set of surfaces called details, where each of them is obtained as a linear combination of basis functions. Small coefficients can be set to zero, giving a sparse approximation of the surface. The MRA structure closely resembles the reconstruction in the HRBF model (cf. Fig. 4), although the coefficients are computed differently. In fact, the  $k^{\text{th}}$  detail

surface in the HRBF is computed from the difference between the surface approximated at the lower level and the surface samples. This allows recovering errors in the computation, e.g. due to finite precision, or the effect of zeroing the coefficients in the previous layers. In MRA instead, the detail is computed directly by projecting the residual at the higher layer onto an adequate basis. This procedure, although faster as it does not require computing the residual for each data point, does not allow correcting for any error in the computation.

Another advantage of the HRBF model is its minimalism, as it requires a single basis, the Gaussian, versus the four bases required by MRA, at least in the most common biorthogonal setting [30].

Finally, the scale of each layer in HRBF does not have to be necessarily halved as in MRA, but can be adapted to the data frequency content. Moreover, if a norm different from  $L^2$  were required, it can be introduced by simply changing the norm in which the error and the residuals are evaluated. For example, if outliers are often collected, a more suitable norm to measure the residual would be  $L^1$  [20].

The limitation of HRBF networks, and of basis function approximation in general, lies in the use of filtering, which makes the reconstruction of structured objects, with sharp edges and corners, difficult. For these, procedures which reconstruct the surface through an intermediate stage based on extraction of objects' geometrical primitives can be more powerful [31, 32].

## **7. Conclusion**

Overall, the 3D scanner system presented here is able to furnish a detailed 3D reconstruction of surfaces by using simple hardware components in conjunction with adaptive data processing. The operations involved both in spot detection and mesh construction are performed locally on the data set. They can be easily parallelized and implemented on general purpose, low-cost, processing

boards (e.g. FPGA) to obtain a low cost system which shall be seen as a standard measurement device of next generation graphical workstations.

## References

- [1] M. Petrov, A. Talapov, T. Robertson, A. Lebedev, A. Zhilayev and L. Polonskiy, "Optical 3D Digitisers: Bringing Life to the Virtual World," *IEEE Computer Graphics & Applications*, vol. 18, no. 3, pp. 28–37, 1998.
- [2] H. Urban, "Ultrasonic Imaging for industrial scene analysis," *Sensor Devices and System for Robotics*, NATO ASI series, Computer and System Sciences, Alicia Casals (ed.), vol. 52, pp. 187–194, Springer Verlag, 1989.
- [3] N.A. Borghese, G. Ferrigno, G. Baroni, R. Savarè, S. Ferrari and A. Pedotti, "AUTOSCAN: A flexible and portable scanner of 3D surfaces," *IEEE Computer Graphics & Applications*, vol. 18, no. 3, pp. 2–5, May/June 1998.
- [4] G. Ferrigno, and A. Pedotti. "Modularly expansible system for real-time processing of a TV display, useful in particular for the acquisition of coordinates of known shapes objects," *U.S. patent* 4,706,296, 1990.
- [5] E. Gagnon and J.F. Rivest, "Laser Range Imaging Using the Self-Mixing Effect in a Laser Diode," *IEEE Trans. Instrum. Meas.*, vol. 48, no. 3, pp. 693–699, 1999.
- [6] N.A. Borghese, M. Di Rienzo, G. Ferrigno and A. Pedotti, "Elite: a goal-oriented vision system for moving objects detection," *Robotica*, vol. 9, pp 275–282, 1990.
- [7] N.A. Borghese and P. Cerveri, "Calibrating a video camera pair with a rigid bar," *Pattern Recognition*, vol. 33, no. 1, pp. 81–95, 2000.
- [8] K. Waters and D. Terzopoulos, "Modeling and Animating Faces using Scanned Data", *J. Visualization and Computer Animation*, vol. 2, pp. 123–128, 1991.
- [9] D. Terzopoulos, and D. Metaxas, "Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics," *IEEE Trans. on PAMI*, vol. 13, no. 7, pp. 703–714 , July 1991.
- [10] Y. Lee, D. Terzopoulos, and K. Waters, "Realistic Modeling for Facial Animation," *Proc. SIGGRAPH 95*, Los Angeles. CA, pp. 55–62, August, 1995.
- [11] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, D.H. Salesin, "Synthesizing Realistic Facial Expressions from Photographs," In *Computer Graphics Proceedings*, Annual Conference Series, 1998, ACM SIGGRAPH, pp. 151–160.



- [12] R. Mencl and H. Müller, "Interpolation and Approximation of Surfaces from Three-Dimensional Scattered Data Points," *Eurographics '98 State of the Art Reports*, Lisbon, Portugal, pp. 51–67, 1998.
- [13] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag Berlin, Heidelberg, New York, 1995.
- [14] A. Baader and G. Hirzinger, "Three-Dimensional Surface Reconstruction based on Self-Organizing Feature Map," *Proc. 6<sup>th</sup> Int. Conf. Advan. Robotics*, pp. 273–278, Tokio, 1993.
- [15] V.K. Krishnamurthy and M. Levoy, "Fitting Smooth Surfaces to Dense Polygonal Meshes," In *Computer Graphics Proceedings*, Annual Conference Series, 1995, ACM SIGGRAPH, pp. 151–160.
- [16] D. De Carlo, and D. Metaxas, "Shape Evolution With Structural and Topological Changes Using Blending," *IEEE Trans. on PAMI*, vol. 20, no. 11, pp. 312–324, 1998.
- [17] N.A. Borghese, S. Ferrari, "Hierarchical RBF networks and local parameters estimate," *Neurocomputing*, vol. 19, pp. 259–283, 1998.
- [18] C. Dorai, G. Wang and A.K. Jain, "Registration and Integration of Multiple Object Views for 3D Model Construction", *IEEE Trans. PAMI*, vol. 20, no. 1, pp. 83–89, Jan 1998.
- [19] T. Poggio and F. Girosi, "Regularization algorithms for learning that are equivalent to multilayer networks," *Science*, vol. 247, pp. 978–981, 1990.
- [20] N.A. Borghese and S. Ferrari, "Automatic acquisition of 3D objects through portable modular system" *Proceeding IMTC*, vol. 3, pp. 1823–1827, 1999.
- [21] S.Lee, G. Wolberg, and S.Y. Shin, "Scattered Data Interpolation with Multilevel B-Splines," *IEEE Trans. on Visualization and Computer Graphics*, vol. 33, no. 3, pp. 228–244, July–September 1997.
- [22] S.G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. PAMI*, vol. 11, no. 7, pp. 674–493, 1989.
- [23] N.A. Borghese, M. Maggioni and S. Ferrari, "Multi-Scale Approximation with Hierarchical Radial Basis Functions Networks," *Tech. Report no. 99-67*, Dept. Electronics and Information, Politecnico of Milano, 1999.
- [24] P.S. Heckbert and M. Garland, "Survey of Polygonal Surface Simplification Algorithms". *Tech. Rep., Carnegie Mellon University*, 1997.
- [25] F. Weinhaus and V. Devarajan, "Texture Mapping 3D Models of Real-World Scenes," *ACM Computing Surveys*, vol. 29, no. 4, pp. 325–365, 1997.

- [26] P. Rigioli, "Sistema per acquisizione di modelli 3D di volti umani e loro animazione ad alto grado di realismo mediante motion-capture," *Laurea Thesis, Dept. Electronics and Information, Politecnico of Milano*, 2000.
- [27] F. Bernardini, J. Mittleman, C. Silva and G. Taubin, "The Ball-Pivoting Algorithm for Surface Reconstruction," *IEEE Trans. Computer Graphics*, 5(4), pp. 349–362, 1999.
- [28] N.A. Borghese, G. Ferrigno, "An algorithm for 3D Automatic Movement Analysis by means of standard Tvcameras," *IEEE Trans. on Biom. Eng.*, vol. 37, pp. 1221–1225, 1990.
- [29] J. Oliensis, Rigorous Bounds for Two-Frame Structure from Motion. *Proc. ECCV '96, LNCS-Series*, Springer-Verlag, Berlin, vol. 1065, pp. 184–195, 1996.
- [30] A. Cohen, "Biorthogonal Wavelets," in *Wavelets: A Tutorial in Theory and Applications*, C.K. Chu Ed., New York, Academic, 1992.
- [31] H. Hoppe, "Surface reconstruction from unorganized points," PhD Thesis, Dept. of Computer Science and Engineering, Univ. of Washington, June 1994.
- [32] P.E. Debevec, C.J. Taylor, J. Malik, "Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach," In *Computer Graphics Proceedings*, Annual Conference Series, 1995, ACM SIGGRAPH, pp. 151–160.
- [33] R.A. Dwyer, "A Faster Divide-and-Conquer Algorithm for Constructing the Delaunay Triangulation," *Int. J. of Computer and Information Science*, vol. 9, no. 3, pp. 219–242, 1997.

## **Figures & Table Legend**

**Figure 1.** The System: scanning a Human face. The beam of a commercial laser pointer is directed over the subject's face. Scanning is performed moving the laser manually with the aim of collecting more data points in the most difficult regions: lips, eyes.... At the same time a commercial CCD video camera acquires a snapshot of the face (a). A schematic diagram of the processing blocks is reported in (b).

**Figure 2.** The distance of 5,000 spots with respect to the best fitting plane is reported in panel (a). Panel (b) shows how the distribution of the error can be considered Gaussian; its standard deviation is assumed as a measure of the digitization rms error.

**Figure 3.** Ensemble of the  $N = 17,080$  3D points (a) and bitmap (b) of Stefano's face. A direct tessellation of the data points (e.g. through Delaunay triangulation) [33] would produce a jerky surface which is plotted with texture (c).

**Figure 4.** HRBF network structure.

**Figure 5.** The multi-scale reconstruction of Stefano's face is reported here. It is obtained adding the contributions of all the four layers (a); of the first three layers (b); of the first two layers (c). In (d) the output of the first layer is plotted. The residual distribution is plotted besides each layer in the form of a histogram whose area is normalized to one.

**Figure 6.** Position of the Gaussian units in the four layers for Stefano reconstruction. Cluster of Gaussians are inserted in the higher layers determined an increase in the local frequency of the reconstruction.

**Figure 7.** The tessellation resulting from sampling the HRBF surface at 1mm step (a) and the final result where texture has been applied (b).

**Table I.** Performance indexes and parameters of each layer of the HRBF network.

**Table I**

#grid	Error mean [mm]	Error std [mm]	#gauss	$n_{\text{cut-off}}$ [Hz]	Grid dimension
1	1.89	3.66	282/315	0.015	15x21
2	0.07	1.45	1124/1333	0.03	31x43
3	-0.005	0.96	3020/5394	0.06	62x87
4	-0.002	0.72	5929/21750	0.12	125x174