Multiscale Approximation With Hierarchical Radial Basis Functions Networks

Stefano Ferrari, Mauro Maggioni, and N. Alberto Borghese

***Abstract*—An approximating neural model, called hierarchical radial basis function (HRBF) network, is presented here. This is a self-organizing (by growing) multiscale version of a radial basis function (RBF) network. It is constituted of hierarchical layers, each containing a Gaussian grid at a decreasing scale. The grids are not completely filled, but units are inserted only where the local error is over threshold. This guarantees a uniform residual error and the allocation of more units with smaller scales where the data contain higher frequencies. Only local operations, which do not require any iteration on the data, are required; this allows to construct the network in quasireal time. Through harmonic analysis, it is demonstrated that, although a HRBF cannot be reduced to a traditional wavelet-based multiresolution analysis (MRA), it does employ Riesz bases and enjoys asymptotic approximation properties for a very large class of functions. HRBF networks have been extensively applied to the reconstruction of three-dimensional (3-D)models from noisy range data. The results illustrate their power in denoising the original data, obtaining an effective multiscale reconstruction of better quality than that obtained by MRA.**

***Index Terms*—Multiscale reconstruction, range data, radial basis function (RBF) networks, wavelet decomposition.**

I. Introduction

MULTISCALE techniques are particularly suitable to deal with nonstationary data, as the majority of real-world interesting data are. Their appealing derives from their ability to efficiently locate short-term events characterized by high-frequency content, and events which have a larger span, but vary slower. Multiresolution analysis (MRA) carried out through wavelet decomposition [1], [2] is the most general tool for multiscale data processing, thanks to its fast machinery. Data processing comprehends a large set of problems, among which hypersurface approximation from sparse noisy data has been widely studied in the connectionist domain [3]. We propose here, a new approach to this problem, alternative to MRA, where multiscale approximation is achieved through a neural-network model called hierarchical radial basis function (HRBF) network. This is a particular multiscale version of radial basis function (RBF) networks [3], which self-organizes to allocate more units where the data contain higher frequencies, guaranteeing a uniform residual error. The constructive approach in HRBF networks is based only on local operations, inspired by linear filtering theory and multiscale analysis, which do not require any iteration on the data. This allows to construct the network in quasi-real time. Thanks to their properties, HRBFs have been recently introduced in different applications [4]–[6]; here, results on their application to three-dimensional (3-D) scanners [7] are reported and discussed.

This paper is organized as follows. In Section II, the HRBF network is formally introduced and it is compared with MRA in Section III. The mathematical proof of convergence is worked out in this section along with remarks on the degree of redundancy of the bases and the convergence speed of the approximation. In Section IV, the reconstruction of a human face from range data obtained with HRBF and MRA is reported. The results are discussed in Section V and in Section VI few concluding remarks are reported.

II. HRBFN

HRBF networks operate a mapping s: $\mathbb{R}^D \rightarrow \mathbb{R}$, as the sum of J hierarchical approximations, $\{a_j(\cdot)\}_{j=1,\ldots,J}$

$$s(\boldsymbol{x}) = \sum_{j=1}^{J} a_j(\boldsymbol{x}) \tag{1}$$

where $a_j(\cdot)$ is a linear combination of Gaussian units with $g(\cdot) = 1/\sqrt{(\pi\sigma^2)^D}\,\exp(-(\|\cdot\|^2/\sigma^2))$

$$a_j(\boldsymbol{x}) = \sum_{c=1}^{M_j} w_{c,j}\, g(\boldsymbol{x} - \boldsymbol{x}_{c,j}; \sigma_{c,j}). \tag{2}$$

Every $a_j(\cdot)$ can be regarded as the output of a RBF network itself [3], which is composed of Gaussians units having the same scale parameter σ: $\sigma_{c,j} = \sigma_j$. Hence, an HRBF network can be described as a pool of J hierarchical subnets, called *layers*, which operate in parallel during the reconstruction [cf. Fig. 1(d)].

The actual shape of $s(\cdot)$ depends on the number of Gaussian units in each layer, M_j, their variance $\{\sigma_j \,|\, \sigma_j \in \mathbb{R}\}$ and their position $\{\boldsymbol{x}_{c,j} \,|\, \boldsymbol{x}_{c,j} \in \mathbb{R}^D\}$, which constitute the *structural parameters* of the network, and the $\{w_{c,j} | w_{c,j} \in \mathbb{R}\}$ which are the *synaptic weights*. In this paper, a new fully constructive approach to compute the parameters is presented. Starting from the input data $\{(\boldsymbol{x}_j, s_m(\boldsymbol{x}_j))$: $\boldsymbol{x}_j \in \mathbb{R}^D$, $s_m(\boldsymbol{x}_j) \in \mathbb{R}\}$, the parameters are computed layer-by-layer. For each layer [cf. Fig. 1(b)], the structural parameters are defined first, the weights are then estimated node-by-node through local operations.

A. Setting the Parameters of a Single Layer

Let us consider a single layer and assume for now that both the data and the Gaussians are equally spaced on a regular grid, A, with Δx the spacing of two adjacent Gaussians. In this case, the RBF can be regarded as the Gaussian interpolation of the

Manuscript received August 27, 1999; revised June 27, 2001 and August 21, 2002. This work was supported in part by the Italian CNR-MIUR by Grant 449/97.

S. Ferrari is with the Department of Information Technology, University of Milan, 20135 Milan, Italy.

M. Maggioni is with the Department of Mathematics, University of Milan, 20135 Milan, Italy.

N. A. Borghese is with the Department of Computer Science, University of Milan, 20135 Milan, Italy (e-mail: borghese@dsi.unimi.it).

Digital Object Identifier 10.1109/TNN.2003.811355

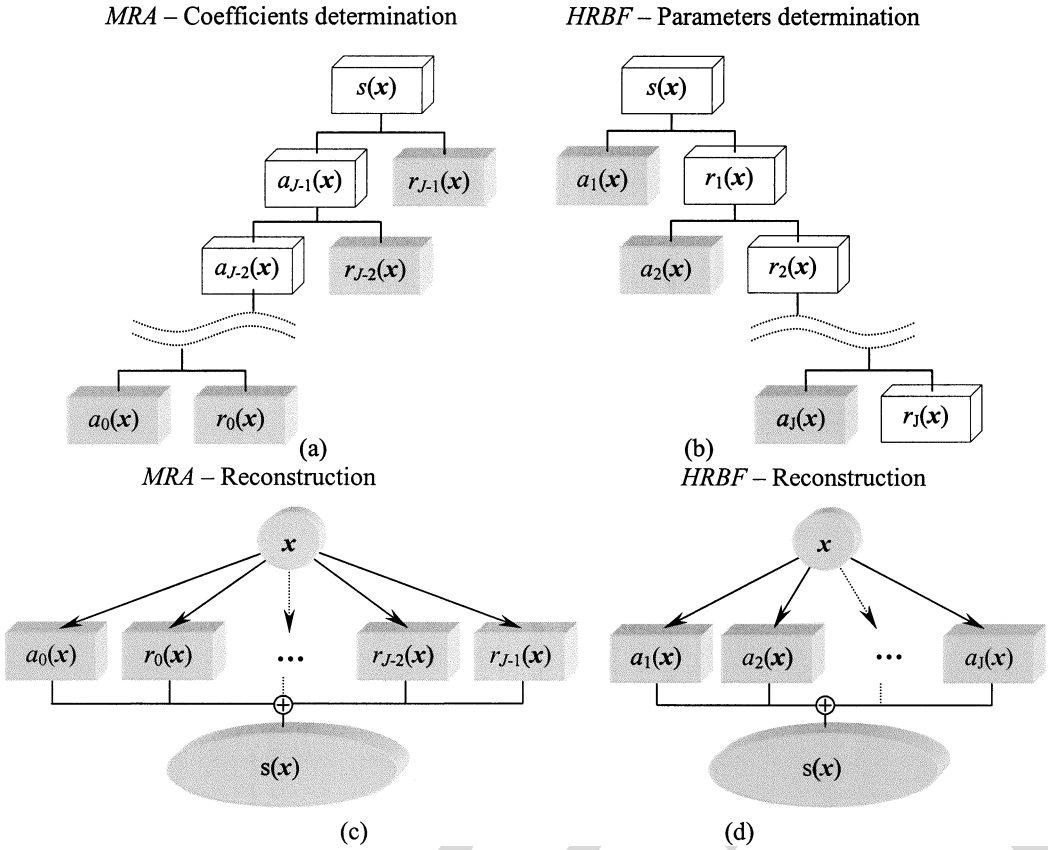

Fig. 1. In the constructive phase (a) MRA proceeds from fine to coarse, while (b) HRBF proceed from coarse to fine. In the reconstruction phase, (c) MRA and (d) HRBF add J and $J-1$ details to an approximation at a large scale, respectively. The difference is in the nature of the details. In MRA they are obtained through projection of the input data over an adequate base. In HRBF the detail of layer j is obtained by projecting the residual of the previous layer (obtained by difference) over the Gaussian basis associated to layer j.

discrete convolution on the grid A of the weights sequence with the discretized Gaussian kernel of scale σ [8]. It follows that the RBF surface sampled on A can be written as

$$\begin{aligned} s(\boldsymbol{x}_p) &= \sum_c w_c\, g(\boldsymbol{x}_p - \boldsymbol{x}_c;\sigma) \\ &= w(\cdot) * g(\boldsymbol{x}_p - \cdot;\sigma), \qquad \boldsymbol{x}_p, \boldsymbol{x}_c \in A \end{aligned} \tag{3}$$

where $*$ denotes the discrete convolution product [it is implicitly assumed $w(\cdot) \equiv 0$ outside the convex hull of A]. This has a strong relationship with linear filtering theory [8]–[10]. In this field, the Gaussian is regarded as a low-pass filter whose bandwidth is inversely proportional to σ. The surface $s(\boldsymbol{x})$ can be reconstructed as

$$s(\boldsymbol{x}) = s(\cdot) * g(\boldsymbol{x} - \cdot;\sigma) = \sum_{c=1}^{M} s(\boldsymbol{x}_c)\, g(\boldsymbol{x} - \boldsymbol{x}_c;\sigma) \tag{4}$$

where $s(\boldsymbol{x}_c)$ is the surface height measured at the grid crossings and $\|g(\cdot)\|_{\mathbb{L}^1} = 1$. Comparing (3) and (4), it can be noticed [8] that the weights can be obtained from the surface height measured at the grid crossings.[1]

[1]More precisely $w_c = s(\boldsymbol{x}_c)\Delta x^D$ due to discrete sampling of $s(\cdot)$ [11]. In the following Δx^D will be incorporated into $g(\cdot)$: $g(\cdot) = (1/\sqrt{(\pi\sigma^2)^D})\exp(-(\|\cdot\|^2/\sigma^2))\Delta x^D$.

Linear filtering theory suggests also a criterion to set a proper value for σ and Δx as a function of the spatial frequency content [8]

$$\sigma_{\min} = \frac{1.465}{\nu_S} = 1.465\,\Delta x \le \sigma \le \frac{0.1874}{\nu_{\max}} = \sigma_{\max} \tag{5}$$

where $\nu_{\max}$ is the maximum frequency content of the reconstructed surface and ν_S is the spatial sampling frequency ($\nu_S = 1/\Delta x$). Finer details can therefore be reconstructed at the price of having more Gaussian units, more densely packed. From the value of σ determined by (5), Δx can be derived (or viceversa). In turns, from Δx, the number of Gaussians, M, follows. At this stage all the parameters of one layer have been defined.

In real applications, the surface height at the grid crossings, $s(\boldsymbol{x}_c)$, is not available [cf. Fig. 2(a)]. To preserve the schema in (4), $s(\boldsymbol{x}_c)$ can be estimated through a local weighted mean [8], [12] as

$$\widehat{s}(\boldsymbol{x}_c) = \frac{\sum_{\boldsymbol{x}_m \in F(\boldsymbol{x}_c)} s(\boldsymbol{x}_m)\, g(\boldsymbol{x}_m - \boldsymbol{x}_c;\sigma)}{\sum_{\boldsymbol{x}_m \in F(\boldsymbol{x}_c)} g(\boldsymbol{x}_m - \boldsymbol{x}_c;\sigma)} \tag{6}$$

where $F(\boldsymbol{x}) \subset \mathbb{R}^D$ selects a region around the point $\boldsymbol{x} \in \mathbb{R}^D$. As there is no advantage in considering the data points very distant from $\boldsymbol{x}_c$, only those in the neighborhood of $\boldsymbol{x}_c$ are considered. Hence, for each Gaussian, the receptive field, $F(\boldsymbol{x}_c) \subset \mathbb{R}^D$, has

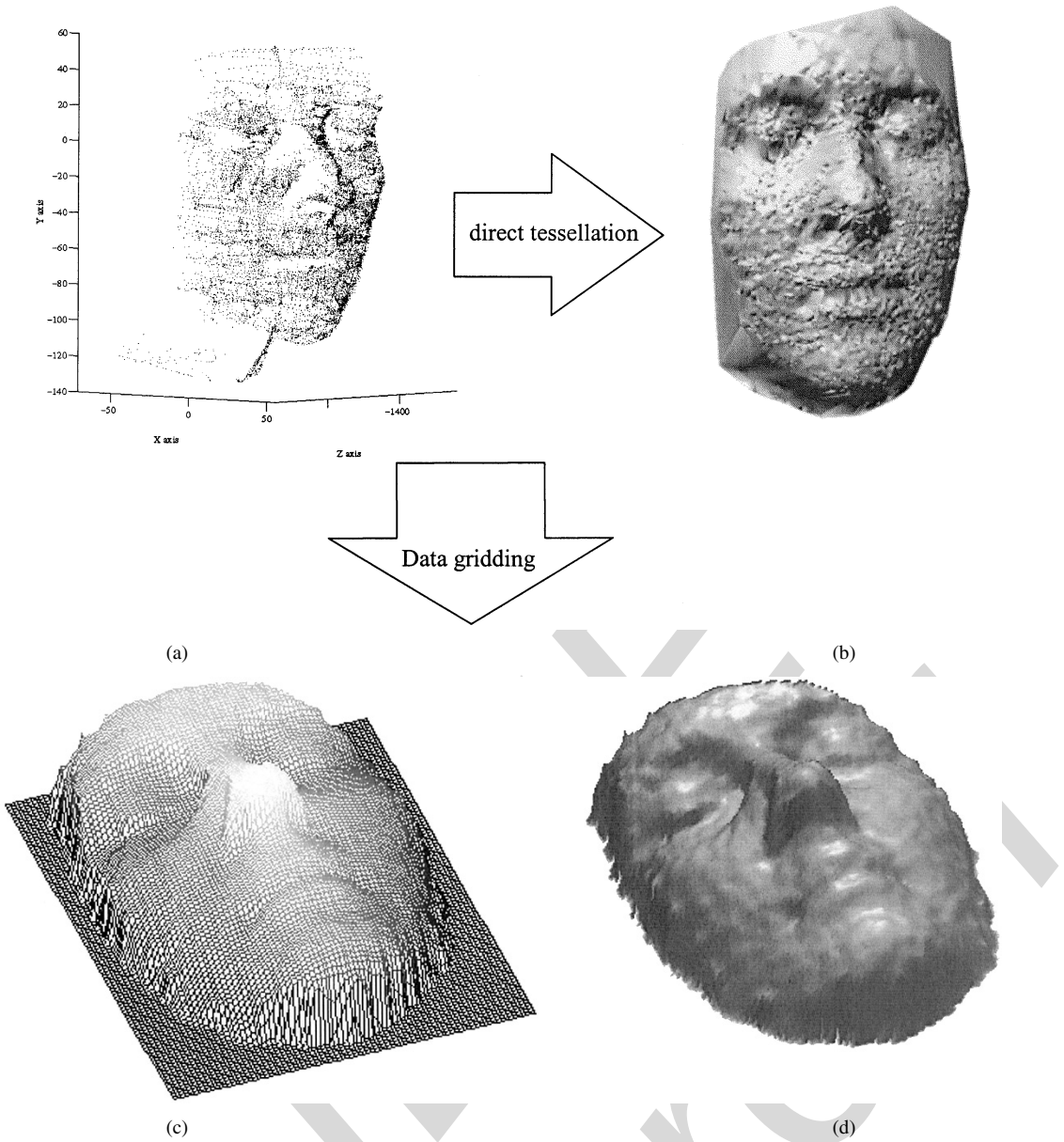

Fig. 2. Typical ensamble or range data. (a) A set of $N = 12\,641$ 3-D data points were digitized by the Autoscan system on a human face [17]. (b) The direct tessellation of the data points produces a poor model: the need of some sort of filtering is evident. (c) The mesh obtained by gridding the input data is plotted along with (d) its Gouraud shaded version. This operation is not adequate to obtain a high-quality mesh and a better reconstruction schema is required.

been assumed here as the D-dimensional hypersphere centered in $\boldsymbol{x}_c$ with radius $2\Delta x$.

Summarizing, the surface $s(\boldsymbol{x})$ can be reconstructed at the scale σ as [cf. (4)]

$$s(\boldsymbol{x}) = \sum_{c=1}^{M} w_c\, g(\boldsymbol{x} - \boldsymbol{x}_c;\, \sigma) = \sum_{c=1}^{M} \hat{s}(\boldsymbol{x}_c)\, g(\boldsymbol{x} - \boldsymbol{x}_c;\, \sigma). \quad (7)$$

B. Constructing the Hierarchical Architecture

If a single layer were used, its scale should be small enough to reconstruct the finest surface details. Unfortunately, a small grid side may cause: 1) waste of resources in low frequency content regions; 2) overfitting; and 3) risk of unreliable estimate of $\hat{s}(x_c)$ due to too few points in $F(x_c)$. To overcome these drawbacks, HRBF networks stack grids of Gaussians with decreasing scale. Units are inserted only in those crossings where the reconstruction error is greater than the measurement error. This allows to achieve a uniform error over all the input domain with a reduced number of units. In details: the first layer reconstructs the surface with few Gaussians featuring a large scale, σ_1, which catches only the outline of the surface. For each measured point $\boldsymbol{x}_m$ a residual $r_1(\boldsymbol{x}_m)$ is computed

$$r_1(\boldsymbol{x}_m) = s(\boldsymbol{x}_m) - a_1(\boldsymbol{x}_m). \quad (8)$$

The successive layers are devoted to reconstruct the details. The jth layer receives as input the residual $r_{j-1}(\boldsymbol{x}_m)$ of the previous layer [Fig. 1(b)]

$$r_{j-1}(\boldsymbol{x}_m) = s(\boldsymbol{x}_m) - \sum_{k=1}^{j-1} a_k(\boldsymbol{x}_m) = r_{j-2}(\boldsymbol{x}_m) - a_{j-1}(\boldsymbol{x}_m). \quad (9)$$

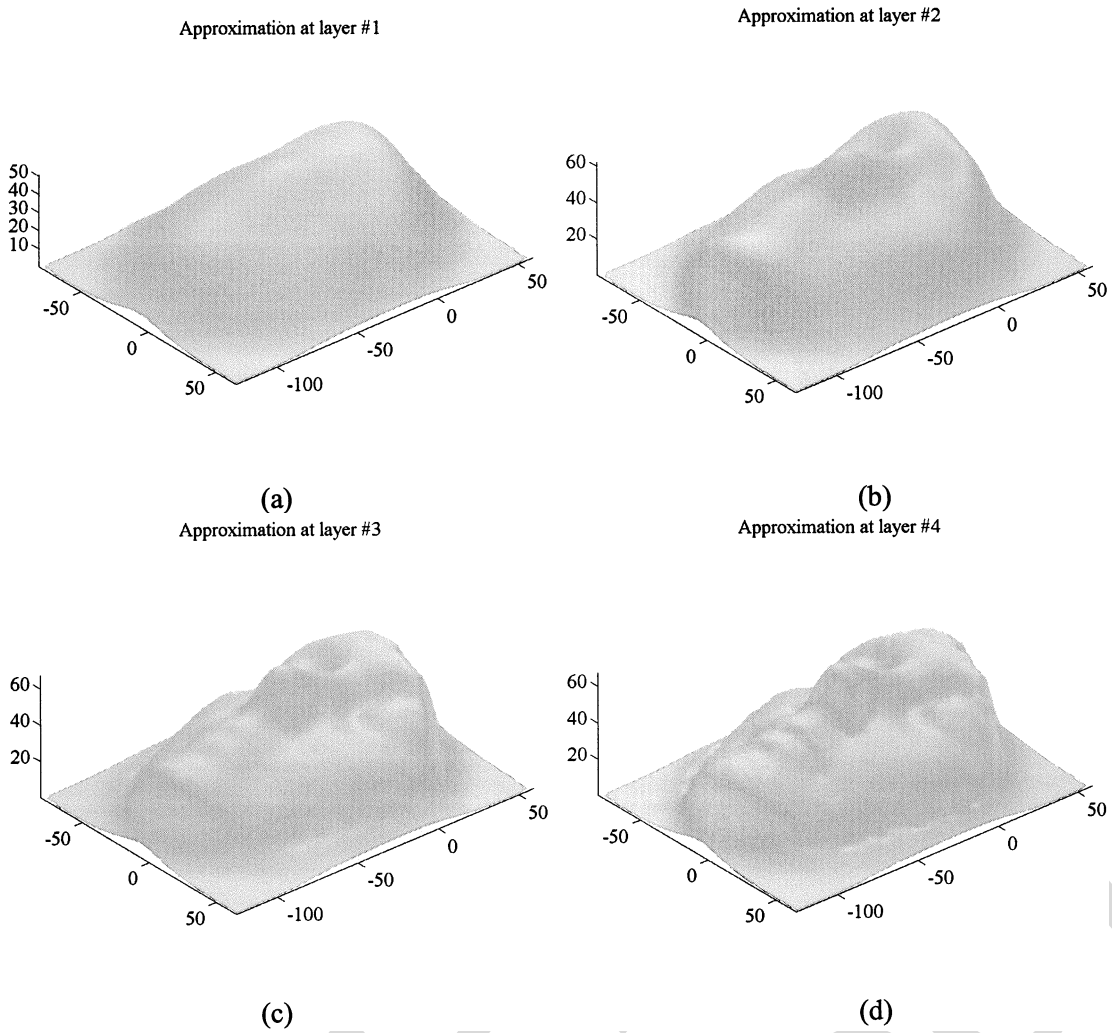

Fig. 3. The reconstruction of the face from the range data in Fig. 2(a) through an HRBF network is reported (cf. Table I). (a) First layer alone. (b) Two layers. (c) Three layers. (d) Four layers.

The scale, somehow arbitrarily, halved at each layer. In each layer the residuals are analyzed node-by-node. For each node, the local reconstruction error $E_j(\boldsymbol{x}_{c,j})$ is computed using all the points inside the receptive field of that node $F(\boldsymbol{x}_{c,j})$

$$E_j(\boldsymbol{x}_{c,j}) = \frac{\sum\limits_{\boldsymbol{x}_m \in F(\boldsymbol{x}_{c,j})} |r_{j-1}(\boldsymbol{x}_m)|}{N_{c,j}} \tag{10}$$

where $N_{c,j}$ is the number of data points which fall inside $F(\boldsymbol{x}_{c,j})$. An integral norm, like $\mathbb{L}^k_{loc}$ norm in error evaluation, allows minimizing the impact of the outliers. $\mathbb{L}^1_{loc}$ has been adopted here, but any other integral norm is suitable. If $E_j(\boldsymbol{x}_{c,j})$ is below the measurement error, no Gaussian is inserted in $\boldsymbol{x}_{c,j}$. Otherwise, its associated weight $w_{c,j}$ is computed through (6), where $r_{j-1}(\boldsymbol{x}_m)$ is substituted to $s(\boldsymbol{x}_m)$ [cf. Fig. 1(b)]

$$\hat{r}_j(\boldsymbol{x}_c) = \frac{\sum\limits_{\boldsymbol{x}_m \in F(\boldsymbol{x}_{c,j})} r_{j-1}(\boldsymbol{x}_m) g(\boldsymbol{x}_m - \boldsymbol{x}_{c,j}; \sigma_j)}{\sum\limits_{\boldsymbol{x}_m \in F(\boldsymbol{x}_{c,j})} g(\boldsymbol{x}_m - \boldsymbol{x}_{c,j}; \sigma_j)}. \tag{11}$$

This procedure is iterated, creating grids at smaller scales and estimating their associated coefficients until $E_j(\cdot)$ goes uniformly under the measurement error (cf. Fig. 3). We explicitly remark that the number of layers J is not given *a priori*, but it can be determined run time by $E_J(\cdot)$ going under threshold over all the input domain.

In the reconstruction phase [Fig. 1(d)], the different layers operate in parallel. The approximating surface, up to the scale σ_k $1 \le k \le J$ is obtained as the sum of the output of the first k layers. The first layer outputs a surface draft, which is corrected with the details output by the higher layers

$$s(\cdot) = \sum_{j=1}^{J} a_j(\cdot) = \sum_{c=1}^{M_1} \hat{s}(\boldsymbol{x}_{c,1})\, g(\cdot - \boldsymbol{x}_{c,1}; \sigma_1) + \sum_{j=2}^{k} \sum_{c=1}^{M_j} \hat{r}_{j-1}(\boldsymbol{x}_{c,j})\, g(\cdot - \boldsymbol{x}_{c,j}; \sigma_j). \tag{12}$$

III. Comparison With Multiresolution Analysis

It is intuitively clear that the HRBF multilayer approach is actually a multiscale approach: at each layer, finer details of

the surface are reproduced by smaller scale Gaussian functions. From this point of view, HRBF can be conceptually compared with MRA [13].

In Section III-A, the basic definitions and properties of MRA [1], [2] are reported. Then, it is shown that the Guassian basis is a Riesz basis, and from this the uniqueness and the stability of the coefficients in (4) are derived (Section III-B). In Section III-C, it is shown that using a set of RBF grids a traditional MRA cannot be obtained. Nevertheless, in Section III-D it is demonstrated that HRBF multiscale reconstruction does enjoy approximation capability for a very large class of functions. An estimate of the convergence speed of a HRBF network, is given in Section III-E.

A. MRA Decomposition

MRA is a particular function multiscale decomposition, which employs Riesz bases.

Definition 1: A collection $\{V_j\}_{j\in\mathbb{Z}}$ of subspaces of $\mathbb{L}^2$ is a Multi-Resolution Analysis (MRA) if: 1) $V_j \subset V_{j+1}$ for all $j \in \mathbb{Z}$; 2) $\bigcap_j V_j = \{\mathbf{0}\}$ and $\bigcup_j V_j$ is dense in $\mathbb{L}^2$; 3) $\forall j$, $s(\cdot) \in V_j$ if and only if $s(2\cdot) \in V_{j+1}$; and 4) there exists a *scaling function* $\varphi(\cdot) \in V_0$ such that $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}}$ is a Riesz basis for V_0. In particular, $\{\varphi(\cdot - k)\}$ is also a Schauder basis for V_0.

Moreover, for every j, a Riesz basis for V_j, can be constructed by dilating and scaling the basis function of V_0. Every function $a_j(\cdot) \in V_j$ can be described as a linear combination of dilated versions of $\{\varphi(\cdot - k)\}$ The last assertion guarantees that for every $s(\cdot) \in V_0$ and for any $\epsilon > 0$ there exists and is unique a set of scalars $\{c_0, c_1, c_2, \ldots, c_n\}$ such that [2]

$$\left\| s(\cdot) - \sum_k c_k \varphi(\cdot - k) \right\|_{\mathbb{L}^2} < \epsilon. \tag{13}$$

Moreover, for every j, $\{2^{j/2}\varphi(2^j \cdot - k)\}_k$ is a Riesz basis for V_j.

Moreover, for every j, a Riesz basis for V_j, can be constructed by dilating and scaling the basis function of V_0. Every function $a_j(\cdot) \in V_j$ can be described as a linear combination of dilated versions of $\{\varphi(\cdot - k)\}$

$$a_j(\cdot) = \sum_k c_{k,j} 2^{j/2} \varphi(2^j \cdot - k). \tag{14}$$

Hence, the set $\{2^{j/2}\varphi(2^j \cdot - k)\}_k$ is a Riesz basis for V_j.

One of the most important features of the collection of subspaces $\{V_j\}$ of a MRA is the *approximation property*. This is the ability of a MRA to approximate, to any prescribed degree of accuracy, any function (in $\mathbb{L}^2$), by projecting it on some subspace V_j, with j large enough. This property derives from $\bigcup_j V_j$ being dense in $\mathbb{L}^2$.

In practice, the MRA decomposition of a function $s(\cdot)$ is carried out from the finest to the coarsest level [cf. Fig. 1(a)]. The first level, J, is $s(\cdot)$ itself [the function is assumed to belong to V_J, cf. Fig. 4(a)]. $s(\cdot)$ is then projected onto V_{J-1} [spanned by the scaling functions, $\{2^{(J-1)}\varphi(2^{J-1} \cdot - k)\}_k$, which feature half the scale of those in V_J], and on its complement in V_J, W_{J-1} (spanned by the wavelet functions): $V_J = W_{J-1} \oplus V_{J-1}$. The projections allow to compute the coefficients $\{c_{k,j}\}$ and $\{d_{k,j}\}$ used to construct, respectively, the approximation of $s(\cdot)$ in V_{J-1}: $s_{J-1}(\cdot)$ [cf. (14)], and the detail

$$r_j(\cdot) = \sum_k d_{k,j} 2^{j/2} \psi(2^j \cdot - k) \tag{15}$$

which lies in W_{J-1}. This schema is iterated on the approximation V_{J-1}, until the largest desired scale is achieved. The function is reconstructed as the sum of an approximation at a very large scale plus $J-1$ details [cf. Fig. 1(c)]. When orthonormal bases are used, each W_j is orthogonal to V_j and all the W_js are orthogonal: $\oplus_j W_j = \mathbb{L}^2$. In this case, fast algorithms are available to compute the coefficients of the decomposition [1], [2], based on projecting the input on the basis. The projections can be interpreted as the convolution with a bank of two mirror filters, and the MRA as the iterative application of the filter bank to the original function [14].

The fact that $\{\varphi(\cdot - k)\}$ is a Riesz basis guarantees also that there exist two constants A and B, independent from $s(\cdot)$, such that

$$A \sum_k |\langle s(\cdot), \varphi(\cdot - k)\rangle|^2 \le \|s(\cdot)\|^2_{\mathbb{L}^2} \le B \sum_k |\langle s(\cdot), \varphi(\cdot - k)\rangle|^2. \tag{16}$$

This inequality is responsible for the numerical stability of the decomposition/reconstruction (i.e., how errors in the data and in the calculation affect the approximation at a given scale [14]). In this respect, it is not important how large/small A and B are, but how near to one is their ratio. This ratio is a measure of the *redundancy* of the basis, in the sense of correlation between its elements: $A/B = 1$ implies no correlation (orthogonality), $B/A \gg 1$ implies large correlation. For an orthonormal basis $A = B = 1$ holds.

B. RBF Grids Are Riesz Bases

In the HRBF framework, the Gaussian $g(\cdot, \sigma_j)$ can be assimilated to a scaling function, and the HRBF *approximation space* V_j is the subspace generated by the translates of the Gaussian with scale σ_j

$$V_j = \langle \{g(\cdot - k\Delta x_j; \sigma_j)\}_k \rangle . \tag{17}$$

In the next theorem, it is shown that each layer, j, of the HRBF network constitutes a Riesz basis for V_j: this implies that the decomposition of $s(\cdot)$ on V_j through the jth grid of the network is unique and stable.

Theorem 1: The set of functions $\{g(\cdot - k\Delta x_j;\ \sigma_j)\}$ is a Riesz basis for V_j.

Proof: Let us introduce the function

$$\beta_j(\nu) = \frac{1}{\Delta x_j} \sum_{k\in\mathbb{Z}} \left| G\left(\xi + \frac{2k\pi}{\Delta x_j}; \sigma_j\right)\right|^2 \tag{18}$$

where $G(\cdot)$ is the Fourier transform of $g(\cdot)$. To be a Riesz basis two constants A, $B > 0$ such that [2]

$$A \le \beta_j(\nu) \le B$$

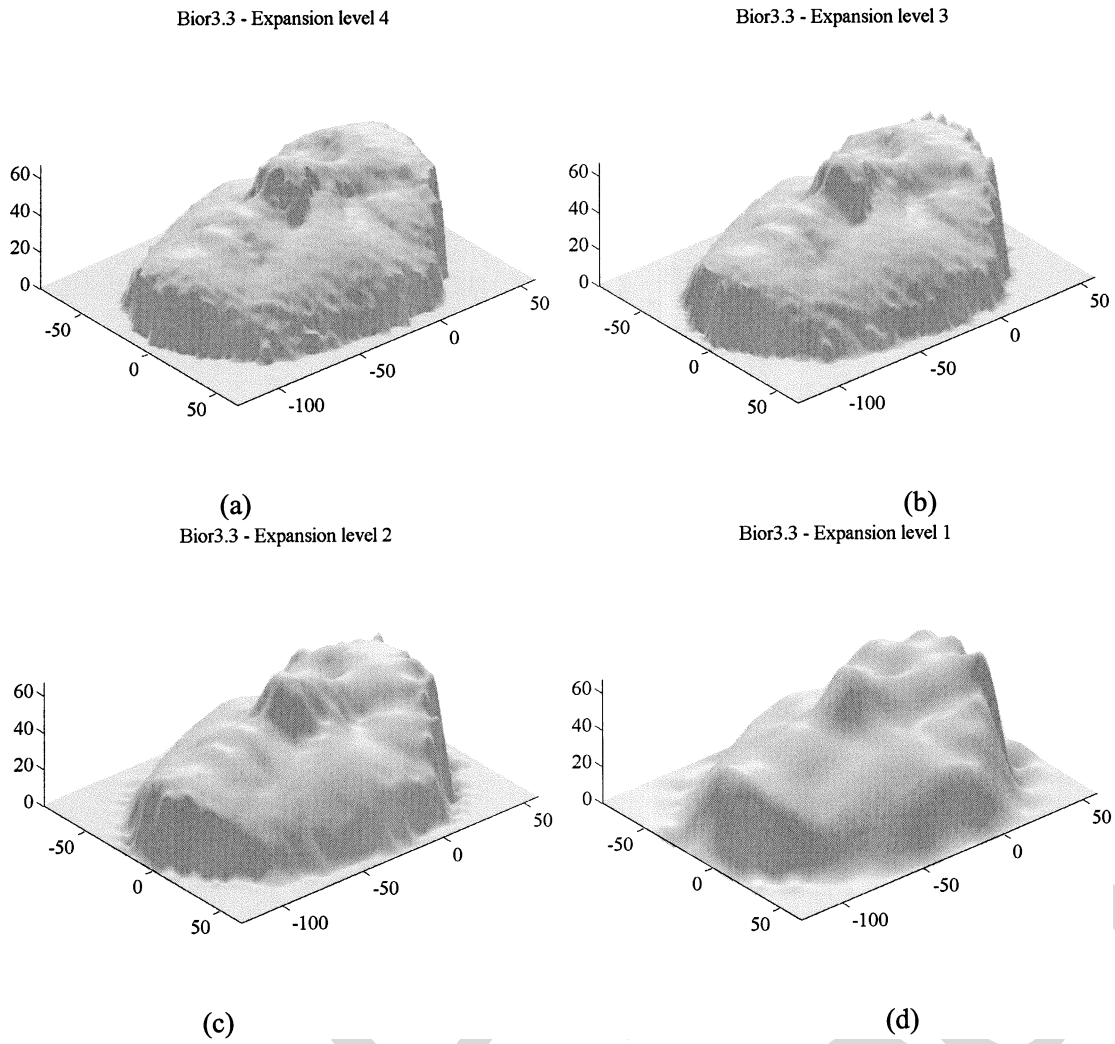

Fig. 4. The reconstruction of the same face as in Fig. 2(a) using MRA implemented by the Matlab package. Biorthogonal Wavelets 3.3, among the most commonly used, are adopted.

for all ν, should exist. In our case, the existence of such constants follows from the periodicity of $\beta_j(\cdot)$, its continuity [a consequence of the uniform convergence of the series in (18)], and its being positive on the compact interval of periodicity. ■

The set of the coefficients $\{w_c\}$ in (7) is, therefore, unique. We can now estimate the stability of the decomposition by determining the coefficients A and B. It can be shown that A and B do not depend on j, but only on the parameters Δx_j and σ_j. Let us consider the case $j = 0$. We observe that the series which defines $\beta_0(\cdot)$ not only converges uniformly, but so do also all its term-by-term derivatives. This implies that $\beta_0(\cdot)$ is a C^∞, $2\pi/\Delta x_0$ periodic function, which is also even with respect to zero and $\pi/\Delta x_0$. Its maximum is clearly attained at 0 , while the minimum is attained at $\pi/\Delta x_0$, since this point is flat by symmetry and it is the only flat point internal to the interval of periodicity (this can be seen by taking the derivative of the series). All of this implies that

$$\frac{A}{B} \approx \frac{\sum_k e^{-2k^2\pi^2\left(\frac{\sigma_0}{\Delta x_0}\right)^2}}{\sum_k e^{-(2k+1)^2\left(\frac{\sigma_0}{\Delta x_0}\right)^2\frac{\pi^2}{2}}}. \tag{19}$$

It follows from this relation that the ratio A/B depends only on $\sigma_0/\Delta x_0$ which, in fact, controls the amount of overlap of the Gaussians in each layer and, therefore, how much correlation is there between adjacent Gaussians. The ratio A/B is the same in the higher layers as both σ_0 and Δx_0 are halved at each step [cf. (5)]. It follows that the Gaussian bases are uniformly (in j) stable. The degree of stability is determined by A and B which can be estimated numerically or analytically from the series in (18). For $\sigma_0 = 1.465\Delta x_0$ [cf. (5)], $B/A \approx 1.6 \cdot 10^4$, which indicates a considerable redundancy.

C. HRBF Does Not Perform an MRA

We show here that, although each grid of RBF is a Riesz basis, stacking them does not produce a traditional MRA: the HRBF approximating subspaces $\{V_j\}$ are not nested ($V_j \nsubseteq V_{j+1}$).

Theorem 2: The subspaces $\{V_j\}$ of an HRBF do not satisfy the inclusion relation: $V_j \subset V_{j+1}$.

Proof: Suppose, by contradiction, that $V_j \subset V_{j+1}$. Suppose also, without loss of generality, $\Delta x = 1$. Since

$\{g(\cdot - k;\, \sigma_j)\}_k$ and $\{g(\cdot - k;\, \sigma_{j+1})\}_k$ are Riesz bases for V_j and V_{j+1}, respectively, we should be able to write

$$g(\cdot;\, \sigma_j) = \sum_k c_k g(\cdot - k;\, \sigma_{j+1})$$

for some sequence $\{c_k\}_{k\in\mathbb{Z}}$ in $l^2(\mathbb{Z})$. Passing to the Fourier transform, there should exist a 2π periodic, square integrable function, $m(\cdot)$, such that [2] $G(2\xi) = m(\xi)\, G(\xi)$, which is absurd, as $G(2\xi)$ and $G(\xi)$ are Gaussian functions.

D. Approximation Capabilities of HRBFs

Although HRBF cannot perform a traditional MRA, we demonstrate here that the multiscale reconstruction produced by HRBF does enjoy similar approximation capabilities. We start from the continuous case[2] and we analyze the jth layer, whose output can be written as [8], [10]

$$a_j(\boldsymbol{x}) = s(\boldsymbol{x}) * g(\boldsymbol{x};\, \sigma_j). \tag{20}$$

Theorem 3: The family of subspaces $\{V_j\}$ associated to the HRBF has the *approximation property*, i.e., for any function $s(\cdot) \in \mathbb{L}^2$ and for any $\epsilon > 0$ we can choose a j (large enough) for which [cf. (13)] $\|s(\boldsymbol{x}) - a_j(\boldsymbol{x})\|_{\mathbb{L}^2} < \epsilon$.

Proof: The claim follows immediately from the result in harmonic analysis known as *summability kernels* or *approximate identities* (cf. [15]). For sake of completeness we give here a simple proof, which takes advantage of the structure of our particular spaces V_j and introduces estimates which will be useful in the Appendix. Suppose $s(\cdot)$ is $\mathcal{C}^\infty$, and compactly supported. We have the following pointwise estimate of the error:

$$\begin{aligned}
|s(\boldsymbol{x}) - a_j(\boldsymbol{x})| &= |s(\boldsymbol{x}) - s(\boldsymbol{x}) * g(\boldsymbol{x};\, \sigma_j)| \\
&= |s(\boldsymbol{x}) - \int_{\mathbb{R}^D} s(\boldsymbol{x} - \boldsymbol{t}) g(\boldsymbol{t};\, \sigma_j)\, \mathrm{d}\boldsymbol{t}| \\
&\le \int_{\mathbb{R}^D} |s(\boldsymbol{x}) - s(\boldsymbol{x} - \boldsymbol{t})| g(\boldsymbol{t};\, \sigma_j)\, \mathrm{d}\boldsymbol{t} \\
&\le \max_{\|\boldsymbol{t}\| \le \delta} |s(\boldsymbol{x}) - s(\boldsymbol{x} - \boldsymbol{t})| \\
&\quad + \|s(\cdot)\|_\infty \int_{\|\boldsymbol{t}\| \ge \delta} g(\boldsymbol{t};\, \sigma_j)\, \mathrm{d}\boldsymbol{t}
\end{aligned} \tag{21}$$

which holds for every δ; $\|s(\cdot)\|_\infty = \max_{\boldsymbol{x}\in T} s(\boldsymbol{x})$, T being the support of $s(\cdot)$. Now we can choose δ small enough to make the first term as small as we want. Fixed that δ, we can take j large enough (σ_j small enough) so that the last integral is smaller than any prescribed positive real number. This implies that we can approximate, to any assigned degree of accuracy, any compactly supported $s(\cdot) \in \mathcal{C}^\infty$. ∎

The approximation is not only pointwise, but is actually uniform (since the right member of the last inequality does not depend on $\boldsymbol{x}$), and hence is valid also in any $\mathbb{L}^p$, for $1 \le p \le +\infty$. Moreover, it can be shown that this result holds also in a rather wide class of Banach spaces (e.g., Sobolev $\mathbb{L}^p_{loc}$ [15]). This result is in accordance with [16] where Bochner and Chandrasekharan theorem was used. In the discrete case, we are approximating the convolution integral in (20) [cf. (3)]. This approximation is uniform and gets better and better as $\Delta x_j \to 0$, which is exactly what happens when the number of layers, j, increases. This shows that the *approximation property* also holds for the discrete case.

Let us remark that in the last theorem a single layer j was considered. This does not automatically imply that the HRBF approximation converges. In fact, in the HRBF decomposition, the function to be approximated changes at every layer, since it is the residual of the previous layer to be projected on the higher layers. However, in the Appendix, we provide the mathematical proof that the HRBF decomposition converges for all the functions, $s(\cdot)$, which have compact support, belong to $\mathcal{C}^\infty$ and for which there exists a value M such that, for every n, $\|s^{(n)}(\cdot)\|_\infty < M$ (equilimited derivatives). Most real surfaces belong to this class of functions.

E. Approximation Spaces of HRBF

In biorthogonal MRAs, which are the most common, the angle between two consecutive approximation spaces is zero while the details spaces are orthogonal. As a consequence, a large difference between two approximations is obtained [2] (cf. Fig. 4 and Table II). In HRBF, the detail spaces are not defined; the approximation is obtained by projecting on the RBF grids at decreasing scale [Fig. 1(d) and Table I]. To get an idea of HRBF speed of convergence, the angle α_j between two consecutive approximation spaces V_j and V_{j+1} is estimated

$$\cos\,\alpha_j = \sup_{f(\cdot)\in V_j, h(\cdot)\in V_{j+1}} \frac{\langle f(\cdot),\, h(\cdot)\rangle}{\|f(\cdot)\|_2 \|h(\cdot)\|_2} = \cos\alpha_{j-1}. \tag{22}$$

The angle α_j is constant for all the layers of a HRBF; to estimate its value, $f(\cdot) = g(\cdot;\sigma_0)$ and $h(\cdot) = g(\cdot;\sigma_1)$ can be assumed, obtaining $\alpha_j \le 0.15\,\pi$. This is quite small, which means that, on one side, little more is added layer-by-layer, and, on the other side, the convergence is slower. This property is useful in all the cases when a given function should be approximated up to a certain degree of accuracy. For example, when a function has to be approximated from noisy data, the final residual should match the measurement error (cf. Section IV) or when the level of detail of the reconstruction should be controlled [cf. Results Section and Fig. 1(b)].

IV. Application to 3-D Reconstruction From Range Data

HRBF networks have been widely applied to reconstruct 3-D surfaces from range data. A typical ensemble of range data points obtained through the Autoscan system [17] from a human face is plotted in Fig. 2(a). The root-mean-square (rms) measurement error was 0.8 mm. Connecting the points to form a triangular mesh produces an undesirable wavy mesh [cf. Fig. 2(b)]. Traditional linear filtering cannot be applied to clean the surface as data are not equally spaced; morevoer, the highly variable spatial frequency content of a face makes the adaptive approach of HRBF particularly suitable.

The face obtained through a HRBF network is shown in Fig. 3. Four layers have been used, featuring scales ranging from 16 to 2 mm. As it can be appreciated, HRBF is able to give an effective progressive multiscale reconstruction. The

[2]This can be obtained from (4) for $\Delta x \to 0$.

TABLE I
QUANTITATIVE EVALUATION OF THE SURFACE RECONSTRUCTION THROUGH HRBF WITH $\sigma = [16, 8, 4, 2, 1]$ mm

#Layers	Mean error [mm]	Error std [mm]	# of gaussians	σ [mm]
1	2.33	4.38	141/270	16
2	0.03	2.06	573/980	8
3	0.04	1.02	1,778/3,795	4
4	0.06	0.77	3,078/14,933	2
5	0.05	0.68	3,719/58,752	1

TABLE II
QUANTITATIVE EVALUATION OF THE SURFACE RECONSTRUCTION THROUGH MRA WITH BIORTHOGONAL BASES 3.3

#Levels	Mean error [mm]	Error std [mm]	# of Bases	Δx [mm]
1	$-4.37 \cdot 10^{-5}$	0.088	2,489/12,528	1.365
2	$5.21 \cdot 10^{-4}$	0.44	1,921/3,744	2.73
3	0.019	0.76	705/1,311	5.46
4	0.15	1.47	283/585	10.92

quality of the network output increases with the number of layers, adding details mainly in the most difficult regions like the nose, the eyes, and the lips (cf. Fig. 3). These details are obtained by means of Gaussian clusters at smaller scales in the higher layers. These clusters are created by the network itself at configuration time by inserting a Gaussian only where the local reconstruction error (10) is larger than measurement error. The final reconstruction is obtained using only 5570 units distributed over the four layers, while a RBF network with a single grid, would have required 14 933 units (109×137) (cf. Table I). If an additional layer is introduced (fifth layer with $\sigma_5 = 1$ mm), the quality of the reconstruction is degraded. Error falls to 0.68 mm, well below the measurement error, clearly indicating overfitting.

MRA cannot be applied directly to the original data of Fig. 2(a), as these are not equally spaced on a grid. For the sake of comparison, these data have been transformed into gridded data as follows. A grid, 109×137 with spacing $\Delta x = 1.365$ mm, is built (this is equal to a RBF grid featuring $\sigma = 2$ mm). The surface height in the grid crossings, $s(\boldsymbol{x}_c)$, is estimated with the local weighted mean (6). The resulting gridded data are reported in Fig. 2(c) and (d). Notice that gridding, by itself, does not avoid the need of filtering. From the gridded data, a multiscale surface is obtained applying a four levels MRA, with biorthogonal 3.3 bases. This adopts short filters (eight samples long), and the wavelet used in the reconstruction phase closely resembles a quadratic spline [14]. The Matlab^{TM} wavelet package was used. The surface is reconstructed using the scaling function at the coarser scale, $\varphi_0(\cdot)$, and the wavelets of all the four levels [Fig. 1(c)]. To compare with HRBF, only 5570 coefficients have been used: all the coefficients of $\varphi_0(\cdot)$ and the 5398 largest wavelets coefficients [14] (cf. Table II). All the others have been discarded.

As it can be seen from Fig. 4, a smooth reconstruction is produced by MRA only at the first two levels while at the last two, the surface is rather wavy. On the other side, the first two levels are not able to sufficiently reconstruct the face details (lips and eyes). Similar results were obtained if the discarded wavelet coefficients were forced to belong only to the first two levels (which contain the highest frequencies), or if they were not discarded at all. The same applies if shrinkage [18] were applied to the MRA procedure. Different bases offered by the Matlab package have been tested with similar results. On the other side, the reconstruction of the surface through HRBF from the gridded data used to test MRA is equivalent to that shown in Fig. 3. The greater ability of HRBF to produce a reconstruction of good quality is evident.

Other wavelet based techniques, which produce an extension of a traditional MRA, have been proposed to reconstruct 3-D models from sparse noisy data points [19]. However, to be applied, they require a preprocessing stage to dispose the points onto a two-dimensional (2-D) manifold mesh with semiregular connectivity.

V. DISCUSSION

A. Structural Aspects

Although wavelet decomposition and HRBF networks both offer a multiscale approximation, their operation is very different. While in MRA, approximations at different scales are created from fine to coarse, HRBF operates the other way round. Coarse-to-fine approach has the advantage that the adequate number of layers has not to be determined *a priori*, but it is the result of the learning process itself: the HRBF network, in fact, grows until the reconstructed surface meets the required quality. This allows saving time and resources as the creation of unuseful dense layers (the highest ones) can be avoided. In MRA, instead, the first level computed is the one which requires the maximum number of coefficients [cf. Fig. 1(a) and Section III-A].

Another main difference is in the speed with which the residuals are caught by the layers at smaller scales, which is related to the angle between the spaces used in the reconstruction (Sections III-A and -E). In MRA, due to the almost orthogonality of the basis, the detail added by each level is maximized. This property is useful in signal analysis and compression, which makes wavelets analysis the *de facto* standard in multidimensional signal compression and has been incorporated in the new JPEG2000 standard. This property is not as valuable when a certain *degree of approximation* is searched. HRBF, instead, adds less detail at each layer, allowing to get closer to a desired degree of approximation.

Reconstruction from noisy data makes this point clear. For the face data of Fig. 2 HRBF proposes two approximations close to the measurement error: the reconstruction error is 1.02 mm with three layers and 0.77 mm with four layers. MRA proposes a third layer with an error of 0.76 mm and a fourth one with 1.47 mm. To this larger numerical difference corresponds a larger difference in the level of the details incorporated in the reconstruction [cf. Fig. 4(c) and (d)]. This difference is larger than that seen in the last two HRBF layers [Fig. 3(c) and (d)]. Notice that although the figure of 0.76 mm is common to the two approaches (four layers HRBF and three levels MRA), it has been obtained very differently: in HRBF the local reconstruction error is almost the same over all the input domain, while in MRA, a slightly smaller error is achieved in most regions, but in

few critical regions (like eyes and nose) it is clearly higher than in HRBF approximation, with an overall lower quality.

We have experimentally verified that, the initial scale, σ_1, is not critical as far as the construction of the network is ended when the residual matches the measurement error. Starting from σ_1 twice the one used before and using five layers (i.e., $\sigma = [32, 16, 8, 4, 2]$ mm) does not produce appreciable differences, nor in the final reconstruction error, nor in the number of units. The same is true if an intermediate value were chosen for σ_1 (e.g., $\sigma_1 = 12$ mm). As the value of σ in each layer is not constrained to be an integer multiple of the previous layer (like in MRA) we can even choose a value of σ on a layer-by-layer basis. This flexibility can be exploited to improve the network performance: the scale parameter of the last layer, σ_L, can be adjusted through single parameter optimization procedure. However, the very little gain in quality, is not worth the additional computational time required by optimization.

B. Coefficients Determination

Once the structural parameters of one HRBF layer have been set, the structure of that layer has been completely defined, and the "synaptic" weights are left to be determined. Equation (7) represents a linear system, which could be solved by matrix manipulation techniques like singular-value decomposition [20], [21]. However, the computation is demanding for the large parameter space, and overparametrization may produce numerical instability and high-frequency oscillations in the reconstructed surface [22]. This is related to the high redundancy of the basis remarked in Theorem 1. The estimate schema adopted in HRBF, instead, computes the weights, unit by unit independently. The local correlation between the data points is therefore exploited to get a better estimate of the weights. The accuracy in the estimate depends on the number of samples which fall around a unit [inside the receptive field of the corresponding Gaussian, (6)]. It is, therefore, critical that data are largely oversampled. This is often the case in neural networks applications, at least in low-dimensionality problems, and, in particular, in reconstructing surfaces from range data.

Not all the HRBF weights and the MRA coefficients are required by the reconstruction. On the contrary, some of them are superflous and may even bring noise into the reconstruction, hence, they have to be discarded. In HRBF a novel elimination strategy is adopted. The weights are discarded (which corresponds to not inserting a Gaussian in the corresponding grid crossing) during the learning phase, on the basis of the local reconstruction error, $E(\cdot)$ (10). There is no analogous mechanism for MRA. In fact, as the MRA coefficients are computed directly projecting the original data on the scaling functions and the wavelets bases, the elimination of one coefficient at a certain level, cannot be taken into account at the higher levels. This drawback does not pertain to HRBF where the weights of each layer are computed from the residuals, which incorporate the effect of having cancelled some of the weights of the previous layers. More elaborated schemas to eliminate wavelet coefficients, based on the statistics of each layer and soft thresholding, have been developed [18]. However, these schemas slow down wavelet decomposition, diminishing the appealing of the technique. Moreover, the threshold parameter cannot be directly related to any measure on the sampled data.

C. General Remarks

As the construction of an HRBF network requires only local operations with no iterations, the resulting configuration algorithm is very fast: less than 1 s are required to build the four-layer 3-D model reported in Fig. 3 on a Pentium III, 800 MHz. Besides a hardware implementation [13], a real-time, biologically inspired, "neural" implementation can be envisaged where the Gaussian functions could be implemented in the dendritic trees [23] or in ensembles of population elements [24], the weights could be implemented in the synaptic connections and the local weighted mean through lateral inhibition mechanisms [3]. The price to be paid to the speed in the configuration algorithm, is the number of units. However, this redundancy does not introduce overfitting because each weight is estimated independently and its associated unit is inserted only if the local reconstruction error is overthreshold (cf. Section III-D). This approach differs from constructive approaches based on orthogonalization (e.g., [20], [21], [25]), whose aim is to minimize the redundancy by choosing a proper set of basis functions. Although these approaches are able to obtain a reconstruction of the same quality with less units, they are based on an iterative machinery which makes them much slower and prone to numerical instability.

The sparse structure of an HRBF network suggests a biological interpretation of the HRBF learning machinery, related to the primate cerebellum. One of the unsolved questions about the cerebellar cortex is the origin of its fracturo-topic structure: the same receptive field is represented in different distinct regions on it [26]. One may hypothesize that the information conveyed by each of these sites is at a different scale, and, in particular, that these sites take shape at different times during learning, representing the information at different level of details. This would agree with the well-known findings in evolution psychology, which describes learning as going through discrete steps [27].

VI. Conclusion

HRBF adopts a novel constructive approach to build an approximating network. The resulting architecture is not claimed to be optimal (nevertheless it may be a good starting point for optimization [28]), but it allows to build an effective network in a very short time. Moreover, as the learning algorithm adopts only local operations, HRBF machinery can be parallelizable. HRBF appears a more suitable tool than MRA for multiscale surface approximation from sparse noisy data as it is clear from the results obtained for the construction of digital models from range data. HRBF networks do not require any *a priori* information and they fully self-organizes (by growing) to represent the desired input/output mapping. This and the units allocation mechanism do not require backpropagation of any error information and suggest new learning schemes in real neural networks.

APPENDIX
CONVERGENCE OF HRBF ARCHITECTURE

We show here that HRBF decomposition converges for all the functions which are compactly supported, belong to $\mathcal{C}^\infty$ and have equilimited derivative. The estimate derived in Theorem 3 (21) holds for any δ and for any $s(\cdot) \in \mathcal{C}^\infty$, compactly supported. If we choose $\delta = \delta_0$ for the level $j = 0$ and $\delta = \delta_j = 2^{-j}\delta_0$ for the other layers, a simple substitution gives

$$\int_{\|\boldsymbol{t}\| \geq \delta_j} g(\boldsymbol{t};\, \sigma_j)\, \mathrm{d}\boldsymbol{t} = \int_{\|\boldsymbol{t}\| \geq \delta_0} g(\boldsymbol{t};\, \sigma_0)\, \mathrm{d}\boldsymbol{t} = K < 1 \qquad (23)$$

where K does not depend on j. Equation (21) can be rewritten as

$$|s(\cdot) - a_j(\cdot)| \leq \|s'(\cdot)\|_\infty 2^{-j}\, \delta_0 + \|s(\cdot)\|_\infty K \qquad (24)$$

which is again a uniform approximation. Let us see what happens at the second HRBF layer. We have to estimate how well we approximate the residual $s(\cdot) - s(\cdot) * g(\cdot;\, \sigma_0)$, by convolving it with $g(\cdot;\, \sigma_0/2)$

$$\begin{aligned}
&\|(s(\cdot) - s(\cdot) * g(\cdot;\, \sigma_0)) - (s(\cdot) - s(\cdot) * g(\cdot;\, \sigma_0)) \\
&\qquad * g(\cdot;\, \sigma_0/2)\|_\infty \\
&\leq \|s'(\cdot) - s'(\cdot) * g(\cdot;\, \sigma_0)\|_\infty \frac{\delta_0}{2} + \|s(\cdot) - s(\cdot) * g(\cdot;\, \sigma_0)\|_\infty K \\
&\leq \|s''(\cdot)\|_\infty \frac{\delta_0^2}{2} + \|s'(\cdot)\|_\infty \frac{3}{2} K \delta_0 + \|s(\cdot)\|_\infty K^2 \qquad (25)
\end{aligned}$$

where we have applied (24) to $s(\cdot) - s(\cdot) * g(\cdot;\, \sigma_0)$. This estimate, carried out in regularity hypotheses for the function $s(\cdot)$, shows that the higher derivatives play an important role in controlling how fast the error decreases in successive approximations. This is expectable since they control how fast the surface varies locally. Let us introduce the quantity $E_j^{(n)}$ equal to the $\mathbb{L}^\infty$ norm of the nth derivative of the error committed at the jth layer of the HRBF decomposition of $s(\cdot)$. Suppose that there exists a value M such that, for every $n \geq 0$, we have $\|s^{(n)}\|_\infty < M$, i.e., the derivatives are equilimited in $\mathbb{L}^\infty$; for $E_j^{(n)}$ the following estimate holds

$$\begin{aligned}
E_0^{(n)} &\leq \left\|s^{(n+1)}\right\|_\infty \delta_0 + \left\|s(\cdot)^{(n)}\right\|_\infty K \leq M(\delta_0 + K) \\
E_1^{(n)} &\leq E_0^{(n+1)} \frac{\delta_0}{2} + E_0^{(n)} K \leq M(\delta_0 + K)\left(\frac{\delta_0}{2} + K\right) \\
&\cdots \\
E_m^{(n)} &\leq E_{m-1}^{(n+1)} \frac{\delta_0}{2} + E_{m-1}^{(n)} K \\
&\leq M(\delta_0 + K)\left(\frac{\delta_0}{2} + K\right) \cdots \left(\frac{\delta_0}{2^m} + K\right). \\
&\cdots \qquad (26)
\end{aligned}$$

Since $K < 1$, for m large enough, we have also $K + (\delta_0/2^m) < 1$ and the products tend to zero. This implies the convergence of the approximation. This is asymptotically geometric and uniform for the function and all its derivatives, in all the $\mathbb{L}^p$ spaces. The uniformity of all estimates above is here used for the sake of simplicity; in fact, one may think of $\|\cdot\|_\infty$ as the supremum taken over a ball or "receptive field" centered around a specific point, and results about local convergency follow immediately.

REFERENCES

[1] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[2] I. Daubechies, *Ten lectures on wavelets*, 1998, vol. 61, Cbms-Nsf Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics.

[3] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural Comput.*, vol. 7, pp. 219–269, 1995.

[4] F. Pedersini, A. Sarti, and S. Tubaro, "Multi-resolution area matching," in *Proc. ICIP 2000 Int. Conf. Image Processing*, vol. 1, Vancouver, BC, Canada, Sept. 2000, pp. 553–556.

[5] N. D'Apuzzo, "Modeling human faces with multi-image photogrammetry," in *Proc. SPIE Three-Dimensional Image Capture and Applications V*, B. Corner, R. Pargas, and J. Nurre, Eds. San Jose, CA: SPIE, 2002, vol. 4661, pp. 191–197.

[6] P. Cerveri, C. Forlani, N. Borghese, and G. Ferrigno, "Distortion correction for X-ray image intensifiers: A comparison between local un-warping polynomials and adaptive neural networks," *Med. Phys.*, 2002.

[7] N. A. Borghese and S. Ferrari, "A portable modular system for automatic acquisition of 3-D objects," *IEEE Trans. Instrum. Meas.*, vol. 49, pp. 1128–1136, Oct. 2000.

[8] N. Borghese and S. Ferrari, "Hierarchical RBF networks in function approximation," *Neurocomput.*, vol. 19, no. 1, pp. 259–283, 1998.

[9] R. Sanner and J. Slotine, "Gaussian networks for direct adaptive control," *IEEE Trans. Neural Networks*, vol. 3, pp. 837–863, Nov. 1992.

[10] M. Orr, "Regularization in the selection of radial basis function centres," *Neural Comput.*, vol. 7, no. 3, pp. 606–623, 1995.

[11] A. Oppenheim and R. Shafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[12] C. Atkenson, A. Moore, and S. Schaal, "Locally weighted learning," *Artif. Intell. Rev.*, 1999.

[13] S. Ferrari, N. Borghese, and V. Piuri, "Multi-resolution models for data processing: An experimental sensitivity analysis," *IEEE Trans. Instrum. Meas.*, vol. 50, pp. 995–1002, Aug. 2001.

[14] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Cambridge, U.K.: Wellesley, Cambridge Press, 1996.

[15] Y. Katznelson, *An Introduction to Harmonic Analysis*. New York: Dover, 1968.

[16] J. Park and I. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Comput.*, vol. 3, pp. 246–257, 1991.

[17] A. Borghese, G. Ferrigno, G. Baroni, R. Savarè, S. Ferrari, and A. Pedotti, "Autoscan: A flexible and portable scanner of 3d surfaces," *IEEE Comput. Graphics Applicat.*, pp. 38–41, 1998.

[18] A. Chambolle, R. DeVore, N. Lee, and B. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Processing*, vol. 7, pp. 319–334, Mar. 1998.

[19] A. Khodakovsky, P. Schröder, and W. Sweldens, "Progressive geometry compression," in *—AUTHOR, IS THIS A PROCEEDINGS OR A BOOK? PLEASE PROVIDE PUBLISHER AND CITY IF BOOK, CONFERENCE SPONSOR IF A CONFERENCE.—Annu. Conf. Series*, 2000, pp. 271–278.

[20] Q. Zhang, "Using wavelet network in nonparametric estimation," *IEEE Trans. Neural Networks*, vol. 8, pp. 227–236, Mar. 1997.

[21] S. Chen, C. Cowan, and P. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, vol. 2, pp. 302–309, Mar. 1991.

[22] S. Marchini and N. Borghese, "Optimal local estimation of RBF parameters," in *Proc. ICANN'94*, vol. 1. New York, 1994, pp. 463–466.

[23] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, 1985.

[24] A. Lukashin, G. Wilcox, and A. Georgopulos, "Overlapping neural networks for multiple motor engrams," in *Proc. Nat. Academy Sci. USA*, vol. 91, 1994, pp. 8651–8654.

[25] J. B. Gomm and D. L. Yi, "Selecting radial basis function network centers with recursive orthogonal least square training," *IEEE Trans. Neural Networks*, vol. 11, pp. 306–314, Mar. 2000.

[26] M. Ito, *The Cerebellum and Neural Control*. New York: Raven, 1984.

[27] J. Piaget, *Origins of Intelligence in Children*. New York: Int. Univ. Press, 1952.

[28] G. P. F. Schwenker and H. A. Kestler, "Three learning phases for radial-basis-function networks," *Neural Networks*, vol. 14, pp. 439–457, 2001.