MULTI-RESOLUTION 3D RECONSTRUCTION THROUGH TEXTURE MATCHING

Federico Pedersini, Augusto Sarti, Stefano Tubaro
Dip. di Elettronica e Informazione (DEI), Politecnico di Milano
Piazza L. Da Vinci 32, 20133 Milano, Italy
E-mail: pedersin/sarti/tubaro@elet.polimi.it

ABSTRACT

In this paper we present a general and robust approach to the problem of close-range partial 3D reconstruction of objects from multi-resolution texture matching. The method is based on the progressive refininement of a parametric surface, which is described using an increasing number of radial functions.

1 Introduction

Typical stereometric methods for 3D data extraction from multiple views are based on the detection, matching and backprojection onto the object space of viewer-invariant features such as corner points and sharp edges. Such methods, unfortunately, are unable to produce dense clouds of 3D data, therefore it is usually quite difficult to interpolate them into a global surface that resembles that of the imaged object. Another approach to stereometric reconstruction that produces dense depth maps is stereopsis, which consists of the matching of the luminance profiles of small image areas of the available views [1]. The 3D coordinates of the surface patch that originated the corresponding luminance profiles are determined through geometric triangulation, while the matching process is performed by maximizing a similarity function between the luminance profiles. A generalized version of this approach has been proposed in the literature [2, 3], which is able to perform area matching while accounting for geometric and radiometric distortions of the luminance profiles. The object, in fact, is modeled as a bundle of tangent planes, whose position and orientation in the 3D space is determined in such a way to maximize the similarity (correlation) between the luminance profiles projected onto them from the available views. Such solutions, however, need an initial approximation of the object surface to begin with, in order to prevent the algorithm from encountering relative minima.

The 3D modeling approach that we present in this able, on the contrary, is able to effectively and efficiently perform an accurate area matching from scratch (modeling bootstrap), with virtually no outliers. In order to do so, we adopt a multi-resolution strategy for shaping the surface. At each resolution level, we determine the surface shape that maximizes the correlation between the original image and the luminance profile of the other views, as transferred through the 3D surface model. The object surface is modeled as a hierarchical radial basis function (RBF) network [4], i.e. as an array of gaussian functions scattered on regular hexagonal grids of progressively increasing density (see Fig. 1).

2 Luminance transfer

Our shape estimation mechanism is based on the comparison between the original luminance profile on an image and the luminance profiles that are "transferred" from the other images. Luminance transferral consists of a back-projection onto the surface model, followed by re-projection onto the destination image plane.

Let us consider a smooth surface modeled as an unknown function (*depth-map*) of the form $d(x, y)$, where (x, y) are the image coordinates of a point as seen on one of its views (reference view), and d is the "depth" (distance between this point and optical center of the reference viewpoint). Estimating the surface shape means determining the coordinates $(x, y, d(x, y))$ of a generic point of the imaged surface. If what we have is a pair of images, then the mapping between the projections s_1 and s_2 of the same 3D point onto the two image planes will be of the form $s_2 = K(s_1)$, where the mapping K depends on the shape of the surface and on the projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ associated to the two views [3]. When the surface is planar, the mapping becomes a linear projective transformation [3]. If $I_1(s_1)$ and $I_2(s_2)$ are corresponding luminances and the surface reflectivity is perfectly Lambertian, then we have

$$I_1(s_1) = I_2(s_2) \ . \tag{1}$$

Based on this equality we can estimate the surface shape through the minimization of a cost function of the form

$$C(d) = \iint \left[I_1(s_1) - I_2(K(s_1))\right]^2 ds_1 \ , \tag{2}$$

where $I_2(K(s_1))$ depends on the depth function d (surface shape) and on the projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$.

The cost function, of course, can only be computed for all the points that are visible on both images.

In several applications the depth function d is expressed in parametric form, therefore it depends on the reference image coordinates (x, y) and on a set of parameters $(\alpha_1, \ldots, \alpha_N)$. The cost function, in this case, depends on such parameters and the best estimate of the surface is given by

$$(\widehat{\alpha}_1, \ldots, \widehat{\alpha}_N) = \arg\min_{(\alpha_1,\ldots,\alpha_N)} \{C(\alpha_1, \ldots, \alpha_N)\}$$

3 RBF and HRBF surface models

Among the possible parametric representations of smooth surfaces, an interesting choice is to model $d(x, y)$ as a Radial Basis Function (RBF) [4], i.e. as a grid of radially symmetric Gaussian functions

$$d(s) = \sum_{m=1}^{M} w_m G(s; s_m, \sigma_m) = \sum_{m=1}^{M} \frac{w_m}{\pi\sigma_m^2} e^{-\frac{|s-s_m|^2}{\sigma_m^2}} \quad ,$$

where M is the total number of Gaussians, while w_m, s_m and σ_m represent their weight, location and standard deviation, respectively. Usually the Gaussians are placed on a regular square grid, while their number is chosen in such a way as to cover the whole area of interest. The surface parameters are thus given by the sole weights. The RBF representation can be constructed in a multi-resolution fashion by organizing the surface in layers of Gaussians, laying on regular grids of increasing resolution (at constant std deviation). In this case the RBF is said to be hierarchical (HRBF) and its representation is given by

$$d(s) = \sum_{l=1}^{L} \sum_{k_l=1}^{K_l} w_{lk} G(s; s_{lk}, \sigma_l) \quad ,$$

where L is the number of RBF layers and K_l is the number of Gaussians of the l-th layer. Such Gaussians have all the same standard deviation σ_l. In order to obtain a good surface representation the value of σ_l needs to be linked to the grid density. A good choice is $\sigma_l = 1.465\Delta_l$, where Δ_l is the grid stepsize [4].

RBFs and HRBFs can also be used as surface interpolators. Given the magnitude of a (continuous and smooth) depth function $d(s)$ on an arbitrary set of points, it is possible to estimate its value at an unknown point $\overline{s}$ as follows

$$d(\overline{s}) = \frac{\sum_{s_k \in A(\overline{s})} d(s_k) \exp\left(-\frac{|\overline{s}-s_k|^2}{\sigma_e^2}\right)}{\sum_{s_k \in A(\overline{s})} \exp\left(-\frac{|\overline{s}-s_k|^2}{\sigma_e^2}\right)} \quad ,$$

where, in practice, the summation is usually extended to only a neighborhood of $\overline{s}$.

We can thus estimate the value of $d(s)$ by using a RBF network

$$\widehat{d}(s) = \sum_{k=1}^{K} d(s_k) \frac{\Delta^2}{\pi\sigma^2} \exp\left(-\frac{|\overline{s} - s_k|^2}{\sigma^2}\right) \quad ,$$

Δ being the grid step of the RBF network. In order to speed up the interpolation process, we can use a multiresolution approach based on a HRBF network

$$\widehat{d}(s) = \sum_{l=1}^{L} \sum_{k=1}^{K} r_{lk}(s_k) \frac{\Delta_l^2}{\pi\sigma_l^2} \exp\left(-\frac{|\overline{s} - s_{lk}|^2}{\sigma_l^2}\right) \quad ,$$

where the residuals r_{lk} are defined as $r_l(s_k) = \widehat{d}(s_k) - d_{l-1}(s_k)$, $l \neq 1$, and $r_1(s_k) = d(s_k)$. Therefore the surface is built, level by level, increasing its local detail.

4 3D reconstruction

The approach to 3D modeling that we propose in this paper is based on the minimization of the mean square error betwen original and transferred luminance profiles, while modeling the surface with an HRBF network. The values of the parameters that describe this surface are estimated by a comparison of the luminance profiles of the three available images.

Although the Gaussian functions of an HRBF network are usually arranged on a regular square grid, we adopted a set of hexagonal grids of increasing density (see Fig. 1), which ensures a slightly better packing of the Gaussian functions with respect to an equivalent square grid. Furthermore, this choice ensures that, especially at low resolutions, some Gaussian functions will be placed at the center of the region of interest.

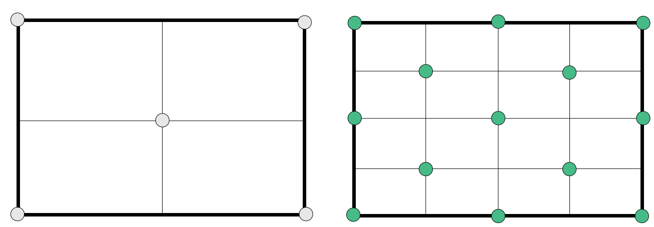

Figure 1: The resolution increases through a densification fo the hexagonal reference grid.

The adopted HRBF network is organized in 8 layers. The resolution doubles (along an axis) from one layer to the next. There is also a layer 0, made of a single Gaussian radial function with $\sigma \to \infty$.

The function $d(x, y)$ is defined as a "projective" depth map referred to the image coordinates x and y (the change of angular density of the grid is assumed as negligible). This choice tends to make the reconstruction consistent with what is viewed.

4.1 Global parameter estimation

In order to model the surface at the lowest levels of resolution (low-density RBF) we perform a global optimization of all surface parameters, using a modified version of eq. (2) as a cost function. The most relevant changes concern the fact that all luminance transfers between views are simultaneously considered; and the fact that the luminance transfer incorporates a gain factor (to be estimated in the optimization process) in order to model the electro-optical differences between cameras. In order to avoid problems of luminace offsets between views, the average luminance is previously subtracted from each image. As the resulting cost function is highly nonlinear, the global optimization process is based on a simplex algorithm.

4.2 Local parameter estimation

The global minimization process described in the previous subsection works well as long the number of parameters of the minimization problems is modest. Unfortunately, as the resolution of the model increases, the number of Gaussians to be adjusted soon becomes unmanageable. In fact, the RBF layers are initially made of a reasonable number of radial functions: 5 for level 1, 13 for level 2, 41 for level 3, and 145 for level 4. Beyond level 4, however, the number of Gaussians becomes unreasonably high, therefore the proposed algorithm is forced to switch to a "local mode", in which only a small image region is considered at each time.

On the reference image, square patches that include five Gaussians each are selected. For each patch, the five Gaussian weights are estimated through optimization. What we obtain is a number of individually estimated patches that locally improve the resolution of the previous HRBF layer. Although the analysis windows relative to each patch overlap with each other, the small surface patches, as we may expect, do not match at the borders. As a consequence, only the coordinates of the central point of each window can be used. The global shape of the new HRBF layer is then built through interpolation based, once again, on a HRBF approach.

It is important to notice that, as the resolution increases, the generation of outliers becomes more frequent because of smaller analysis region; small luminance gradient; specularities on the surface; occlusions, etc. Such problems, however, are usually characterized by high values of the cost function and/or by large depth corrections on the previous layer. It is thus possible to eliminate such outliers through a proper thresholding process. In fact, the lack of incremental depth information on some regions of the network layer does not constitute a problem as the correction is incremental and the interpolation is done at the end of the layer's construction process.

The local optimization is also characterized by a further improvement of the cost function, which now uses the original luminance profiles (instead of the deviations from their averages), and incorporates an unknown luminance transfer offset (to be estimated through optimization) in order to model reflections that are modestly non-lambertian.

The organization of the algorithm, when working in local mode, is shown in Fig. 2.

5 Conclusions

In this paper we proposed a general and robust method for the close-range 3D reconstruction of surfaces through multi-resolution area matching. The method is based on the progressive refinement of a parametric surface, which is described by increasing number of radial functions organized in an HRBF network. This solution enables the 3D surface reconstruciton without any initial model.

The algorithm has been tested on different real image triplets obtaining significant results (see Figs. 3, 4 and 5).

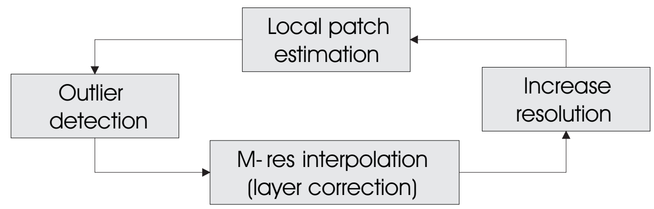

Figure 2: Global scheme of the reconstruction algorithm when working in "local mode"

References

[1] Y. Otha, T. Kanade, "*Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming*", IEEE Tr. On PAMI, Vol. 7, N. 2, pp. 139-154, 1985.

[2] F. Pedersini, A. Sarti, S. Tubaro, "*Robust Area Matching*". ICIP '97, Oct. 26-29, 1997, Santa Barbara, CA, USA.

[3] P. Pigazzini, F. Pedersini, A. Sarti, S. Tubaro, "*3D Area Matching with Arbitrary Multiview Geometry*". Signal Processing: Image Communications, special issue on *3D Video Technology*. Vol 14, Nos. 1-2, 1998, pp. 71-94.

[4] N. A. Borghese, S. Ferrari, "Hierarchical RBF networks and local parameters estimate", Neurocomputing, Vol. 19, 1998, pp. 259-283.

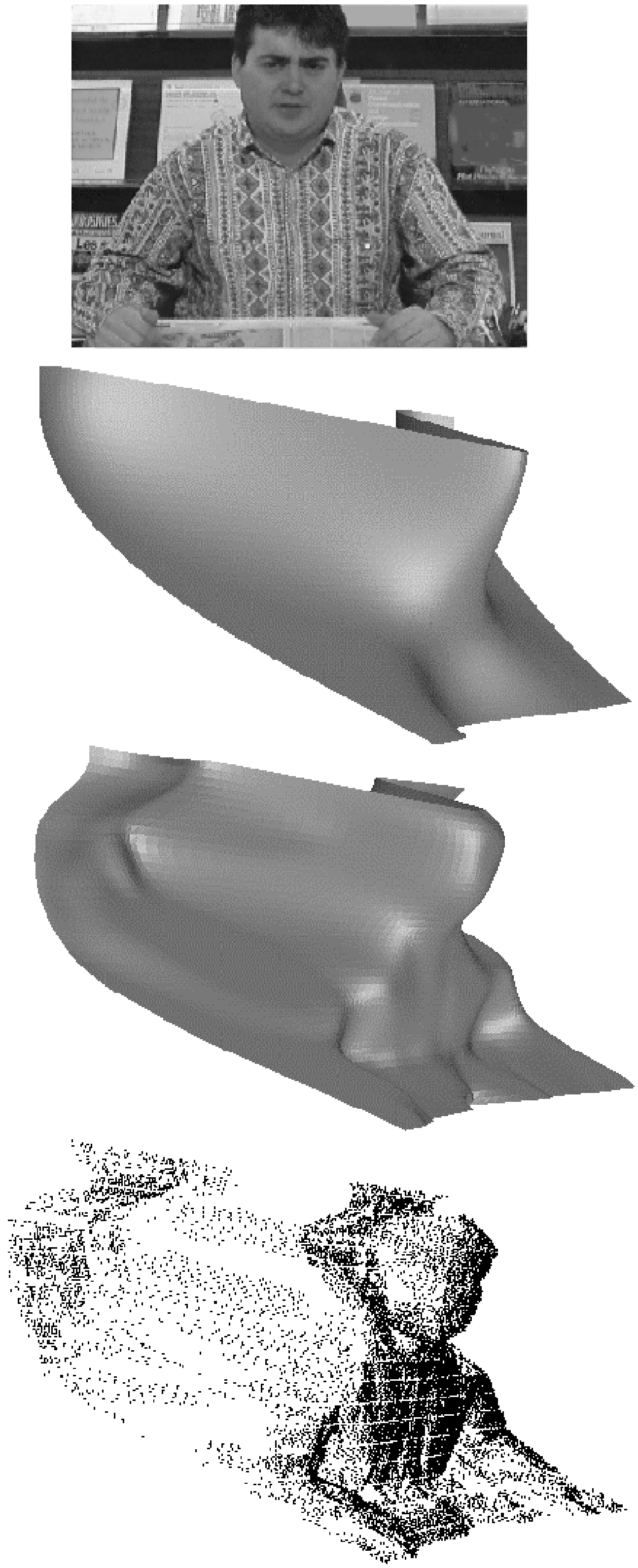

Figure 3: Example of 3D reconstruction of a teleconferencing scene from three calibrated views. From top to bottom: one of the original views; progressive refinement of the reconstructed surface, final cloud of points after outlier thresholding.

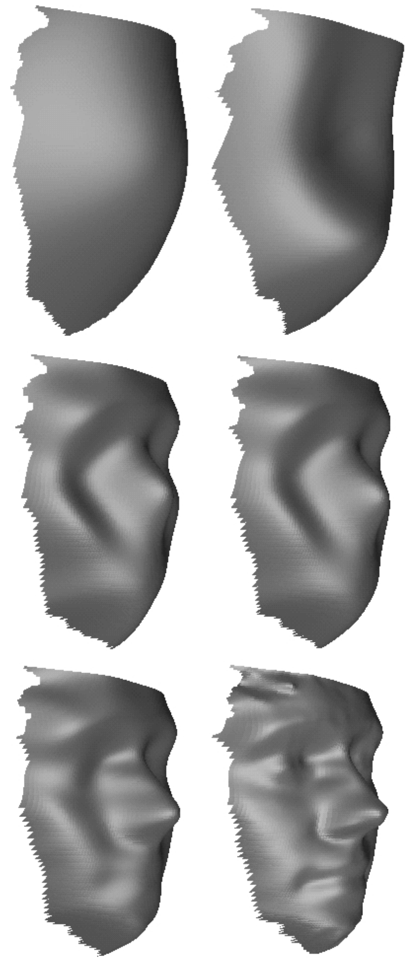

Figure 4: Progressive surface reconstruction of a face using multi-resolution area matching.

Figure 5: Example of 3D reconstruction of a face. From top to bottom: final reconstruction with texture mapping.