GEODESIC NEIGHBORHOODS FOR PIECEWISE AFFINE INTERPOLATION OF SPARSE DATA

Gabriele Facciolo, Vicent Caselles

Universitat Pompeu Fabra - Dept. Tecnologies de la Informació, Barcelona, Spain
{gabriele.facciolo,vicent.caselles}@upf.edu

ABSTRACT

We propose an interpolation method for sparse data that incorporates the geometric information of a reference image. The idea consists in defining for each sample a geodesic neighborhood and then fit a model (affine for instance) to interpolate at the current point.

In the field of remote sensing for urban areas, two widely used techniques are laser range scanning (LIDAR) and stereo photogrammetry. Both techniques have a common drawback, for a variety of reasons the information they provide is sparse or incomplete. But in both cases it is fair to assume that a high resolution image of the scene is available, and we propose in this paper a diffusion algorithm that takes into account the geometry of the image u to refine the range data. This allows us to interpolate the data set while respecting the edges of u. The core of the algorithm is a fast method for computing *geodesic distances* between image points, which has been successfully applied to colorization by Yatziv et al. and supervised segmentation by Bai et al.

The geodesic distance is used to find the set of points that are used to interpolate a piecewise affine model in the current sample. This first interpolation result is refined by merging the obtained affine patches using a greedy Mumford-Shah like algorithm. The output is a piecewise affine interplation of the data set that respects both the given data and the radiometric information provided by u.

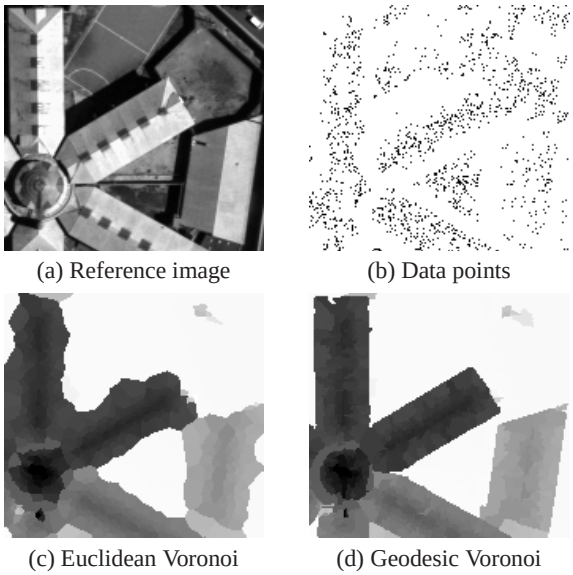
***Index Terms*—** Interpolation, Remote sensing, Image processing, Diffusion processes

1. INTRODUCTION

We consider the problem of interpolating a set of range measurements of a scene using the additional knowledge of the radiometric information of the same scene given by the image u.

This scenario is common in the case of LIDAR measurements, since a digital image has a higher density, and its acquisition is faster, when compared to the range data. We will take advantage of the information provided by a image to interpolate the sparser range measurements. The same applies to the case of stereo reconstruction of urban *Digital Elevation Models (DEM)*, since in this case the height information can be accurately determined only at a few locations in the image [1], and therefore using the reference image to interpolate them will provide a denser height map.

As in most works we adopt the Lambertian hypothesis. That is, a uniform surface (in the model) with a constant angle will be seen with a constant intensity in the image. This assumption allows us to extrapolate information across uniform regions of the image. Clearly not all uniform regions will have some data sample in it, especially if we consider textured surfaces. In those cases we extrapolate from the "nearest" sample, where nearest refers to the geodesic distance that takes into account the radiometric and edge information of the image. We have to acknowledge that areas of constant intensity in the image do not necessarily correspond to the same surface in the model, but this situation is unlikely since in most cases a discontinuity in the model corresponds to a change of intensity in the image.



(a) Reference image (b) Data points

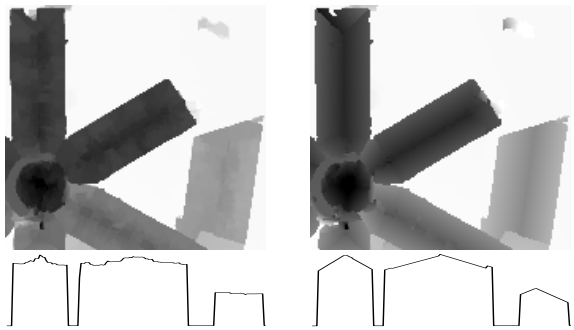
(c) Euclidean Voronoi (d) Geodesic Voronoi

Fig. 1. Piecewise constant geodesic Voronoi interpolation of height information (darker is higher). In (a) is shown the reference image used to compute the geodesic distances, (b) shows the position of the samples with known depth (5% of the total pixels). Interpolating the samples of (b) with Euclidean Voronoi cells produces (c), while (d) corresponds to the geodesic Voronoi interpolation.

1.1. The basic idea

Let us denote by $u(x) : \Omega \to \mathbb{R}^+$ the monochromatic image of a scene defined on $\Omega \subset \mathbb{R}^2$, and let $G(\lambda) : \Lambda \to \mathbb{R}$, $\lambda \in \Lambda \subset \Omega$, be a given depth function which we assume to be known only in Λ. As in [2, 3] the idea is to use the geodesic distance to incorporate the radiometric information provided by the image u in the interpolation of the sparse data G to produce a dense depth map $H(x) : \Omega \to \mathbb{R}$ that fits the values of G. This distance measures the minimum variation of u between two points. Therefore, the distance of two points along the same isophote is 0, while the distance of two points at both sides of a discontinuity of u will be proportional to the "jump" of u.

With the geodesic distance we define the geodesic Voronoi diagram for the sites in Λ, and then interpolate in each cell using an affine model. The geodesic Voronoi cell represents an advantage when compared to the Euclidean nearest sample interpolation, since



(a) Linear blend (b) Affine adjust

Fig. 2. Interpolation of the samples in Fig 1(b) using the 5 nearest (geodesic) samples. Left: Yatziv's [2] weighted mean blending. Right: interpolation by affine plane estimation using the same 5 samples. In the second line we compare a profile from each images.

it permits to recover sharp and meaningful boundaries of the model (see Fig 1). Moreover, since the distances can be efficiently computed with a modified pixel queue algorithm, this method provides a fast alternative to iterative anisotropic diffusion algorithms.

In [2] the authors use the K nearest geodesic samples to blend the chrominance information provided by color scribbles, while here we re-interpret those *geodesic neighborhoods* as samples belonging to the same surface as the current point. The linear blending used in [2] cannot recover kinks in the model (see Fig 2(a)). Because, as pointed out in [4], they consider just the distance and not the spatial distribution of the nearby samples. It becomes clear that, in order to recover kinks and discontinuities, an affine model must be adjusted to each geodesic neighborhood (Fig 2(b)).

But this method alone is not robust in the presence of noise or outliers in the measurements. Solving this issue involves the robust selection of the neighborhoods and the merging of regions with compatible affine models. It is possible to propose plenty of merging strategies to address this issue [5], which will eventually lead us to a higher level representation of the scene. To verify this conjecture and to improve the estimation in case of noisy samples, we will merge adjacent geodesic Voronoi cells using a greedy criterium and an error based stopping condition.

The idea of using the information provided by the image u to guide the interpolation is not new. In [6] the authors solve an anisotropic minimal surface problem to interpolate sparse disparity data. But minimal surfaces are unable to resolve discontinuities and TV is unable to recover kinks in the model, moreover, this type of schemes are likely to be slow due to its iterative nature. In [5] the authors incorporate u through an initial segmentation, adjusting the data points of each region with a robust estimation and use a region merging strategy to merge similar regions. Ideally, with a "perfect" initial segmentation, this method gives the best possible result, but since each initial region must already contain the points needed for the estimation, a bad initialization could be catastrophic. In our method there is no initial segmentation, the regions are induced by Voronoi neighborhood configurations, and the affine planes are estimated using the geodesic neighborhood relations.

Let us give the plan of the paper. In the next Section we will define the geodesic distance, the geodesic neighborhoods and in Section 3 we use them to fit a piecewise affine model. In Section 4 we present a basic region merging algorithm to increase the robustness of the estimation. Section 5 is devoted to the discussion of the results, the limitations of the presented method and possible solutions. In Section 6 we give some conclusions and future work.

2. GEODESIC DISTANCES AND NEIGHBORHOODS

Let s and t be two points in Ω and let $C(p) : [0,1] \to \Omega$ be a curve in Ω. We write $C_{s,t}$ to refer to a curve connecting s and t such that $C_{s,t}(0) = s$ and $C_{s,t}(1) = t$. Then we define the geodesic distance between s and t as:

$$d(s,t) = \min_{C_{s,t}} \int_0^1 |\nabla u \cdot \dot{C}_{s,t}(p)| + \varepsilon |\dot{C}_{s,t}(p)| dp. \qquad (1)$$

This distance is a regularized anisotropic distance. When $\varepsilon \ll 1$ it behaves as the geodesic distance $\min_{C_{s,t}} \int_0^1 |\nabla u \cdot \dot{C}_{s,t}| dp$ (at least for short curves). This distance respects the contrasted boundaries of the image as long as the Euclidean length of the shortest curve is less than ε^{-1}, therefore we select ε inversely proportional to the size of the domain and the range of u.

Finally, we define the *geodesic neighborhood* $GN_K(p)$ as the set formed by the K-nearest (in the geodesic sense) samples of Λ to the point p.

Remark 1 The geodesic distance (1) embodies the edge information contained in the image u, however under some circumstances the shortest curve may cross an edge (in the discrete case). Edges (or other information) can be introduced as a hard constraints in the metric in order to penalize curves crossing them.

2.1. Implementation details

To compute the geodesic neighborhoods we adopt an algorithm similar to the one proposed in [2]. Indeed we run Dijkstra's algorithm over a 4-connected pixel lattice, once for each data sample and store all the geodesic distances to data points. The overall performance of this method is improved by early stopping each Dijkstra execution, since we only require the K nearest neighbors. Let us briefly summarize our algorithm.

The data structures

For each pixel in Ω we define:

- a status *label*: needed for the Dijkstra algorithm (ACCEPTED, FAR, TRIAL),
- *distance*: stores the current minimum distance,
- *nearest*: refers to the nearest pixel in $\Omega \setminus \Lambda$,
- the *K Nearest List*: stores the K Nearest distances to samples obtained to the moment. This list only keeps the K smallest values.
- *maxDistance*: the value of the maximum distance stored in the list, but if the list is not full then *maxDistance* is infinity.

The geodesic neighborhood algorithm

- Define a priority queue of pixels, sorted by the *fitness* field.
- For each pixel $p \in \Omega \setminus \Lambda$
  1 Clear the priority queue and label all the pixels of Ω as FAR.
  2 Add p to the priority queue, with distance 0.
  3 While the priority queue is not empty:
  3.1 Extract the top of the queue : q.
  3.2 Update $q.label$ as ACCEPTED.
  3.3 Add the $q.distance$ to the *K Nearest List* of q.
  3.4 For each non ACCEPTED neighbor of q : r
  3.4.1 Compute the distance from p to r trough q: $dNew = d(p,q) + d(q,r)$.

3.4.2 If $r.label$ is TRIAL and, $dNew < r.distance$ then update $r.distance = dNew$.

3.4.3 If $r.label$ is FAR and, $p.distance < r.maxDistance$ then update $r.label$ as TRIAL, set $r.distance = dNew$ and add r to the priority queue.

3. ROBUST AFFINE PLANE INTERPOLATION

For each data sample $p \in \Lambda$ we have defined a geodesic Voronoi cell and a geodesic neighborhood. Fig 1 shows that the geodesic Voronoi diagram successfully accounts for discontinuities in the image. We propose to use the geodesic neighborhood $GN_K(p)$ to fit an affine plane trough its points and extend it to the whole geodesic Voronoi cell. This will give a piecewise affine model H.

To interpolate H at $p = (p_x, p_y) \in \Omega$, we determine its nearest (with geodesic distance) data sample $p^* \in \Lambda$, and its geodesic neighborhood $GN_K(p^*)$. Then compute

$$H(p) = w_p^T v_{p^*}, \tag{2}$$

where $w_p = (p_x, p_y, 1)^T$ are the homogeneous (spatial) coordinates of p, and $v_{p^*} \in \mathbb{R}^3$ contains the affine parameters determined by the least squares regression

$$v_{p^*} = \arg\min_{v \in \mathbb{R}^3} err(v, GN_K(p^*)), \tag{3}$$

$$err(v, GN_K(p^*)) = \sum_{q \in GN_K(p^*)} |w_q^T v - G(q)|^2. \tag{4}$$

Observe that the samples of $GN_K(p^*)$ used in (3) are (except for p^* itself) all outside the geodesic Voronoi cell of p^*.

The size K of the neighborhood is a critical parameter here. If it is too small, then the estimation of the plane will be poor. If it is too big, then we will not be able to recover small planar surfaces (with only a few samples on them). Moreover if the geodesic neighborhood contains outlier samples (that do not belong to the same object as the center p) then the result will be biased.

To remove the outliers we use a *RANdom SAmple Consensus* (RANSAC) [7], which we modified to assure that the consensus set always contains the 3-nearest geodesic neighbors of p. This choice of the 3 neighbors is arbitrary, however this parameter depends on the density of the samples. Indeed with this selection we are supposing that each planar region contains at least 3 samples in it.

4. CONSTRAINED REGION MERGING

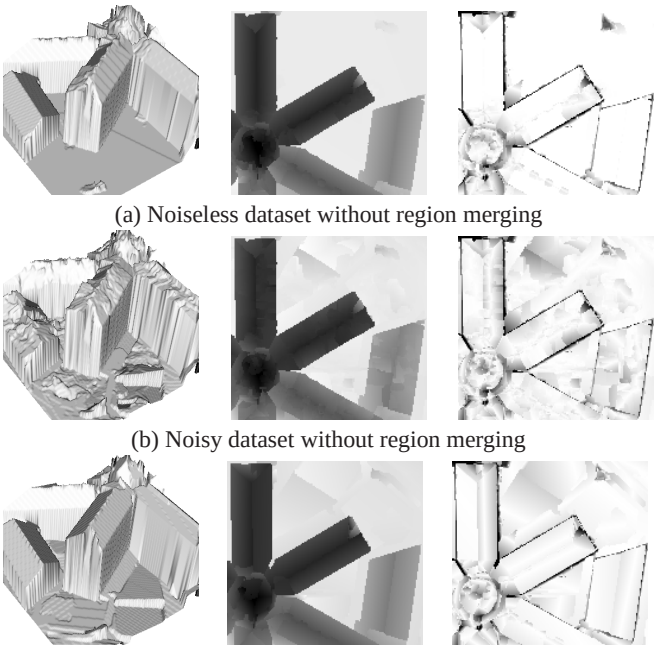
Fitting each individual Voronoi region with an independent plane represents an inconvenient when the height information is perturbed by noise (or small measurement errors). Even for adjacent regions, that share some samples, the different fittings may not coincide resulting in an irregular model as shown in Fig 3(b).

To improve the plane estimation we will merge the models of adjacent geodesic Voronoi regions obtained in the previous section. For that we minimize a simplified Mumford-Shah functional

$$E(B, f) = \sum_{R \in \mathcal{P}(\Omega)} RErr(R, f) + \lambda \int_B g(s) ds, \quad \lambda \geq 0. \tag{5}$$

with $RErr(X, f) = \sum_{x \in X \cap \Lambda} |f_X(x) - H(x)|^2$. Where H denotes the output of the affine plane interpolation and f_X is the affine model for the region X obtained by

$$f_X(x) = w_x^T \left(\arg\min_v \sum_{p \in X \cap \Lambda} err(v, GN_K(p)) \right). \tag{6}$$



(a) Noiseless dataset without region merging

(b) Noisy dataset without region merging

(c) Noisy dataset after region merging

Fig. 3. Left: interpolated map. Middle: interpolated map as gray level. Right: error map. In (a) we display the interpolation of the noiseless data without region merging. In (b) the same for the case of noisy data. In (c) we display the region merging applied to (b).

The boundary length term $g(s)$ is a function that is big at poorly contrasted boundaries and very small at the well contrasted ones. In practice, we take $g(x) = \exp\left(-\frac{|\nabla u(x)|^2}{\sigma^2}\right)$, with $\sigma^2 = \mathrm{Var}(u)$, so that the length of a curve along a contrasted boundary is almost zero. As in [8] a greedy minimization of (5) will merge two regions A and B when

$$\lambda > \frac{RErr(A \cup B, f) - RErr(A, f) - RErr(B, f)}{\ell_g(\partial(A, B))}, \tag{7}$$

where $\ell_g(\partial(A, B))$ is the weighted length of the common boundary between A and B.

In the context of urban landscape interpolation there are evident boundaries that should not be violated, for instance, we will include restrictions along the segments detected by the LSD algorithm [9] (Fig 4(a)), which reflects the geometry of the urban landscape. If the edge crosses a line segment then its length will be set to zero, and this will avoid the merge of these two regions. Moreover, we will keep track of all the edges in this situation and forbid any merge that removes them from the segmentation.

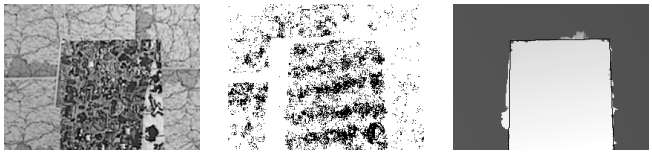
Lastly, the selection of λ is critical, however we will select it indirectly by controlling the merging error $RErr(A \cup B, f)$ at each step and stopping the algorithm if it exceeds a threshold.

5. EXPERIMENTS

Let us test the proposed method in two cases: when the depth measures are exact, and the more realistic case when they are noisy. For that we consider two datasets associated to the scene displayed in Fig 1, the first dataset was extracted from the ground truth depth information, which was available. Indeed this image is part of a stereo

(a) LSD segments (b) Before merging (c) After merging

Fig. 4. Region merging. The region merging algorithm do not merge regions across detected segments (a), and stops if the error is too big. (c) shows the regions obtained after merging the regions of (b).



(a) Reference image (b) Known samples (c) Interpolated map

Fig. 5. Middelburry *MAP* set. The complex non contrasted texture of (a), complicates the obtention of the edges of the model (c).

data set kindly provided by CNES from which we only use the reference image and the disparity. In the second dataset the depth information was computed with the subpixel stereo algorithm described in [1]. For this last dataset the variance of the error is estimated to be 0.013, and we use it to select the parameters for RANSAC and the region merging stopping condition.

Both datasets where decimated to 5% of the image pixels (240×240 pixels), and the depth information along the detected segments was erased. Such a low density is unlikely in a real application, however this may prove the efficacy of the method. For these images the computation of the 25 geodesic nearest neighborhoods takes 14 seconds on a 1.6 Ghz Pentium M CPU with non optimized C code. Without tuning the RANSAC parameters the processing time for this stage is less than 5 seconds and the region merging takes 2 seconds.

In Table 1 we summarize the performance (in terms of Mean Square Error) for each step of the proposed method. First, let us consider the noiseless case. In Fig 3(a) we display the result of the affine interpolation without merging. The error map of shown in the rightmost column of Fig 3(a) (darker means higher error with respect to the ground truth information) shows that most of the errors come from small planes, non-planar surfaces, and discontinuities. Note that, despite of the low sample density, kinks are recovered with great precision and planes are well recovered.

In the noiseless case we can skip the merging step since all the height information is consistent across the neighborhoods. However in the noisy case this is not possible since the result of the neighborhood interpolation is very irregular (see Fig 3(b)). The result shown in Fig 3(c) is obtained after merging the region models of the Fig 3(b). As mentioned in Section 4 the algorithm avoids merging regions that are separated by LSD segments (shown in Fig 4(a)), and stops when the merged error variance exceeds 0.013. It is interesting to see that the resulting partition (see Fig 4(c)) also resembles a segmentation of the model.

Let us point out some limitations of this method. Since we do not construct a geometric description of the scene, there is no distinction between kinks and discontinuities in the model (note that in Fig 3(c) most roof planes do not coincide at the top). Lastly, in presence of poorly contrasted edges between strongly textured regions the method tend to produce artifacts as seen in Fig 5. More results

Algorithm	Noiseless Disp.	Noisy Disp. [1]
Linear Blend [2]	0.0902	0.0983
Affine (no RANSAC)	0.0810	0.0829
Affine	0.0687	0.0708
Affine+Merge	0.0649	0.0637

Table 1. Mean Square Errors (MSE) of the results obtained using the ground truth information. All the experiments where performed with the 25 geodesic nearest neighbors. The first column corresponds to the interpolation of noiseless data (the values of samples are exact), and in the second the samples are computed with a stereo correlation algorithm [1].

can be found at `http://gpi.upf.edu/static/geoint`.

6. CONCLUSIONS AND FUTURE WORK

We have proposed a method for interpolating range data that incorporates the information provided by an image of the scene. We have seen that the geodesic neighborhoods given by the metric (1) constitute a fast and robust tool for modelling the local range information. This method can be easily adapted to other (non-affine) scene models [4]. In future work we will study the robust selection of the neighborhood specially since RANSAC discards all geodesic distance information. We will also study the merging process.

Acknowledgements: G.F. acknowledges L. Podesta and J. Martinez for their useful comments. V.C. acknowledges partial support by PNPGC project, reference MTM2006-14836 and by "ICREA Acadèmia" for excellence in research funded by the Generalitat de Catalunya.

7. REFERENCES

[1] N. Sabater, A. Almansa, and J-M. Morel, "Rejecting wrong matches in stereovision," CMLA Preprint, 2008.

[2] L. Yatziv and G. Sapiro, "Fast image and video colorization using chrominance blending," *IEEE Trans. on Image Proc.*, vol. 15, no. 5, 2006.

[3] X. Bai and G. Sapiro, "A geodesic framework for fast interactive image and video segmentation and matting," in *IEEE 11th Int. Conf. on Computer Vision*, 2007.

[4] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM national conference*, New York, NY, USA, 1968, ACM.

[5] L. Igual, J. Preciozzi, L. Garrido, A. Almansa, V. Caselles, and B. Rougé, "Automatic low baseline stereo in urban areas," *Inverse Problems and Imaging*, vol. 1, no. 2, 2007.

[6] G. Facciolo, F. Lecumberry, A. Almansa, A. Pardo, V. Caselles, and B. Rougé, "Constrained anisotropic diffusion and some applications," in *Proceedings of British Machine Vision Conference*, Edinburgh, U.K., 2006.

[7] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol.24, no.6, 1981.

[8] J-M. Morel and S. Solimini, *Variational methods in image segmentation*, Birkhauser Boston Inc., Cambridge, USA, 1995.

[9] R. Grompone, J. Jakubowicz, J-M. Morel, and G. Randall, "Lsd: a line segment detector on digital images," *to appear in IEEE Trans. Pattern Anal. Mach. Intell.*, 2007.