Curve Reconstruction from Noisy Data

Emilio Vital Brazil　　Ives Macêdo　　Luiz Henrique de Figueiredo

IMPA – Instituto Nacional de Matemática Pura e Aplicada, Brazil
email: {emilio,ijamj,lhf}@visgraf.impa.br

Introduction

The problem of shape reconstruction from noisy dot patterns is a classical problem in low level computer vision, pattern recognition, and cluster analysis. We extend [1], which uses results from graph theory and differential topology to construct correct polygonal approximations of differentiable arcs from a dense set of sample points on them (i.e., exact, noiseless data, as depicted in figure 1(a)). Following directions pointed out by [1] our extension is an approach for a more general class of curves, including loops, multiple connected components, and represented by noisy sample sets, figure 1(b). We use probability theory results to estimate parameters that allow automatic reconstruction based on a single user-selected probability.

Method

Similar to the approach in [1], our method consists of four steps: computation of an *Euclidian minimum spanning tree* ($EMST$) over the sample set $\mathcal{S}$, *bridge deletion*, *topological noise filtering*, and *loop closure*. In the bridge and loop identification steps, we use thresholds to delete bridges and to close loops. The computation of these thresholds is based on some statistics defined on $EMST(\mathcal{S})$.

Euclidean MST computation. The $EMST(\mathcal{S})$ of a finite set of points $\mathcal{S} \subset \mathbb{R}^2$ is a *minimum spanning tree* of the complete weighted graph $\mathcal{K}(\mathcal{S})$ whose vertices are the elements of $\mathcal{S}$ and the edges' weights are the *Euclidean distances* of the adjacent vertices. We avoid building $\mathcal{K}(\mathcal{S})$, by exploiting the fact $EMST(\mathcal{S}) \subset Del(\mathcal{S})$ and precompute a *Delaunay triangulation* of $\mathcal{S}$.

Threshold estimation. The estimation of thresholds used to identify bridges and loops exploits the following result from *probability theory* [2, section V.7]:

> **One-Tailed Chebyshev's Inequality:** *"Let X be a random variable with expected value μ and finite variance σ^2, then $P(X - \mu \geq \kappa\sigma) \leq \frac{1}{1+\kappa^2}, \forall \kappa > 0$".*

This gives us the opportunity of fixing an *upper bound* $p \in (0, 1]$ for the probability that a random variable tresspasses a certain amount $\epsilon = \mu + \kappa\sigma$. By setting $p = \frac{1}{1+\kappa^2}$, we have $\kappa = \sqrt{\frac{1-p}{p}}$, by the *one-tailed Chebyshev's inequality*, $p \geq P(X - \mu \geq \kappa\sigma) = P(X \geq \epsilon)$.

Bridge deletion. To detect connected components we delete every edge of $EMST(\mathcal{S})$ whose weight is greater than the threshold given by $\kappa_{bridge} = \sqrt{\frac{1-p_{bridge}}{p_{bridge}}}$, obtaining a minimum spanning forest $MSF(\mathcal{S})$, figure 2(b). Analyzing our threshold estimation method we observe that, if the probability p_{bridge} is too high, edges other than bridges may be deleted, if it is too low ($p_{bridge} \approx 0$) then bridges can change the topology of the curve, as depicted in figure 4.

Topological noise filtering. To filter topological noise (i.e., vertices with valence greater than 2), we use the following graph properties: the *Topological length* of a path is the number of its edges; the *Euclidean length* of a path is the sum of weights (Euclidean lengths) of its edges; A *diameter path* is a path with maximal length (in the topological or Euclidean sense). To filter out the noise from $MSF(\mathcal{S})$, a diameter path $\mathcal{P}_i(\mathcal{S})$ is computed for each connected component $\mathcal{C}_i(MSF(\mathcal{S}))$ and the edges not in the diameter path are deleted (figure 3(a)).

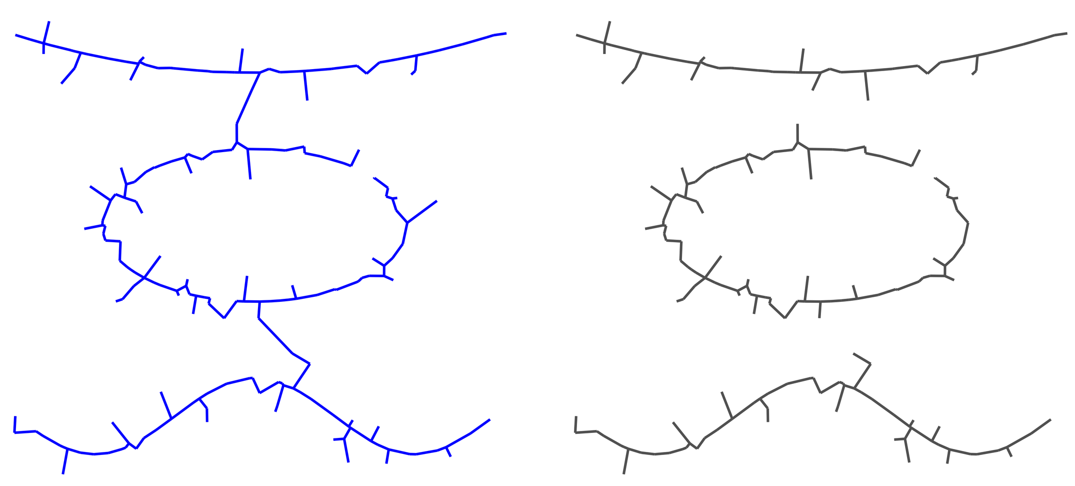
Loop closure. To identify the loops that must be closed in each $\mathcal{P}_i(\mathcal{S})$, we check whether the distance d_i between its extremes is lower than the threshold $\epsilon_{loop} = \mu_i + \kappa_{loop}\sigma_i$, where $\kappa_{loop} = \sqrt{\frac{1-p_{loop}}{p_{loop}}}$, ensuring that the probability of an open arc be erroneously closed in the *loop closure* step is lower than $p_{loop} \in (0, 1]$ (figure 3(b)). Observe that, if the probability p_{loop} is too high, some loops may fail to be closed, and, if it is too low, some edges which do not belong to the curve may be added (figure 5).
The whole pipeline of our method is depicted in figures 2, 3 figures 6–11.

Results

We have performed a number of different experiments to evaluate our method and the effects of parameters. Although we have experimented with both topological and Euclidean lenghts of graph paths, no relevant qualitative difference has been noticed. We have added to the method in [1] a more intuitive set of parameters and have shown that the modified, automatic method, constructs a good polygonal approximation for low-noise sets even if you have just a rough knowledge of the sampled curve topology, expressed by the probabilities p_{bridge} and p_{loop}. As future work, we intend to use wavelets to filter geometric noise from the reconstructed polygonal approximation [3] and experiment with 3D noisy sampled curves, since our method is easily applicable to the three-dimensional setting.

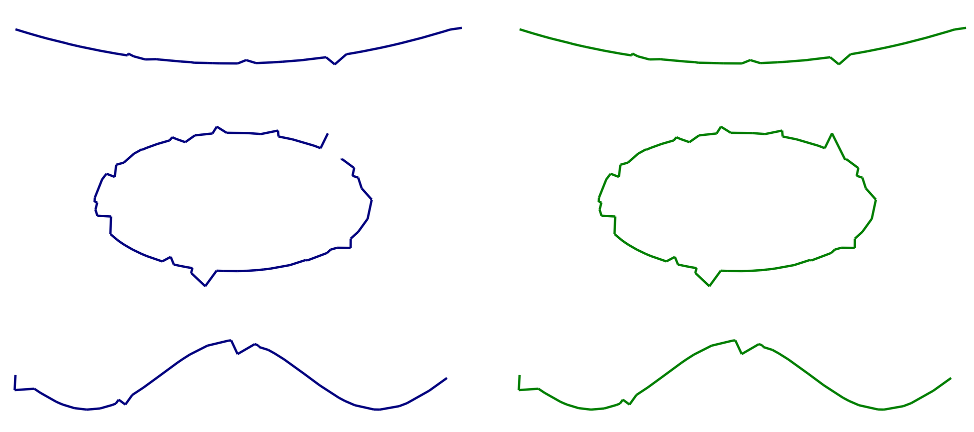
(a)　(b)

Figure 1: densely sampled curves: (a) exact, on curve, sample set; (b) general curve noisily sampled



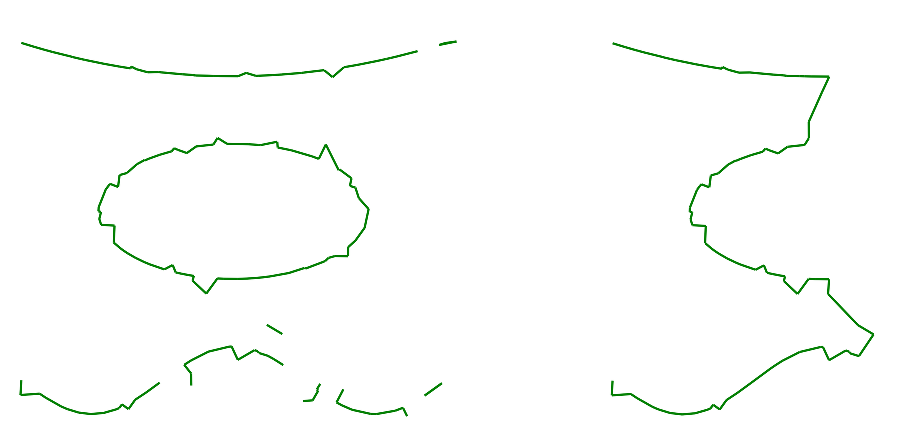
(a)　(b)

Figure 2: (a) euclidean MST; (b) bridge deletion



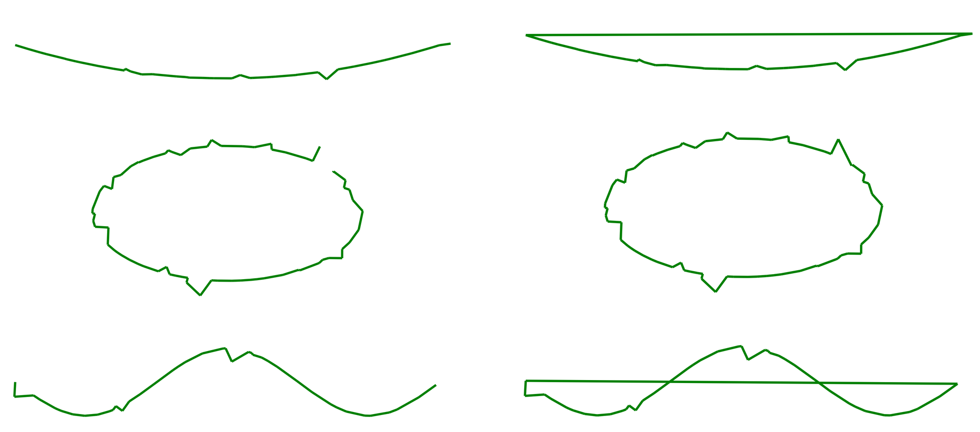
(a)　(b)

Figure 3: (a) topological noise filtering; (b) loop closure



(a)　(b)

Figure 4: (a) high p_{bridge} ($= 0.5$) creates many connected components; (b) low p_{bridge} ($= 0.01$) doesn't identify the bridges



(a)　(b)

Figure 5: (a) high p_{loop} ($= 0.2$) doesn't find the loop; (b) low p_{loop} ($= 0.0001$) closes "obviously" open arcs

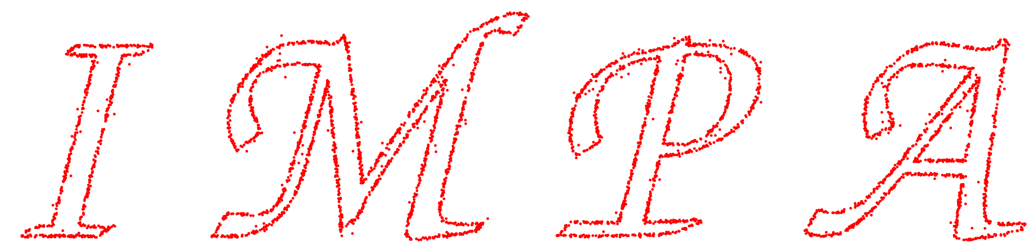


Figure 6: point cloud

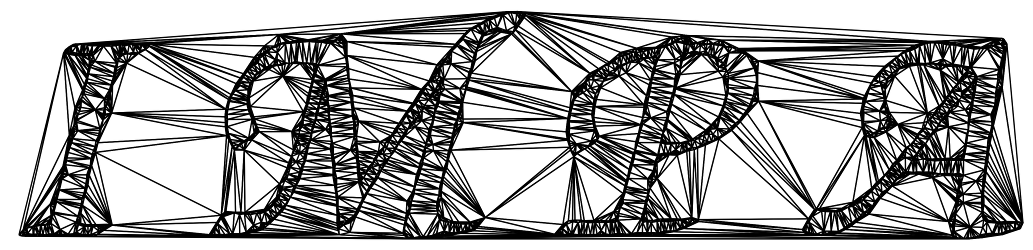


Figure 7: delaunay triangulation

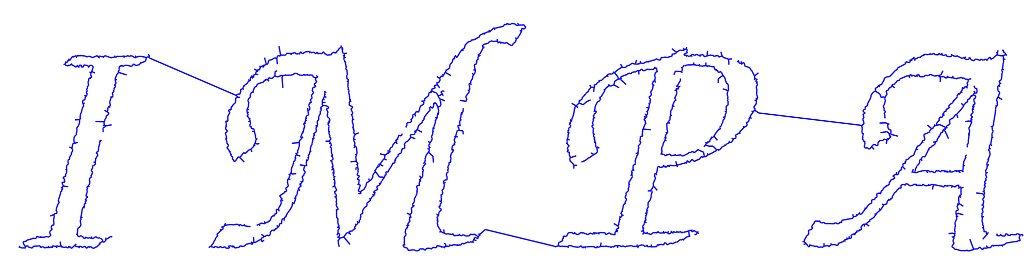


Figure 8: euclidean MST

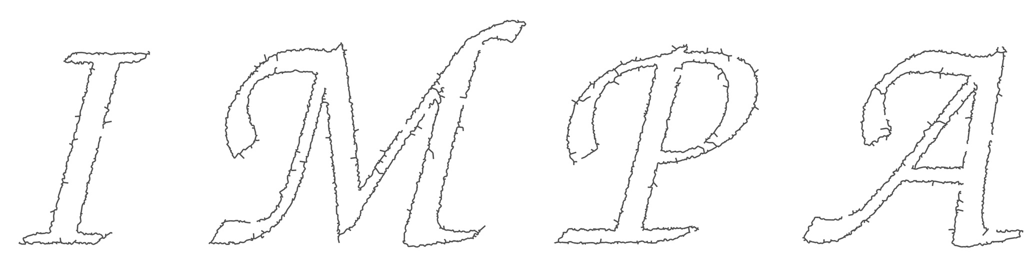


Figure 9: bridge deletion

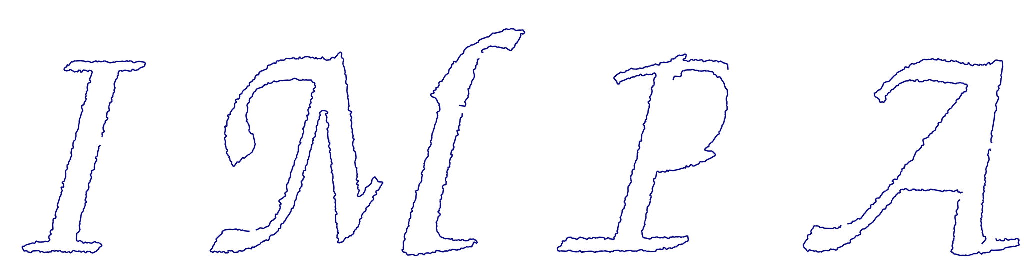


Figure 10: topological noise filtering

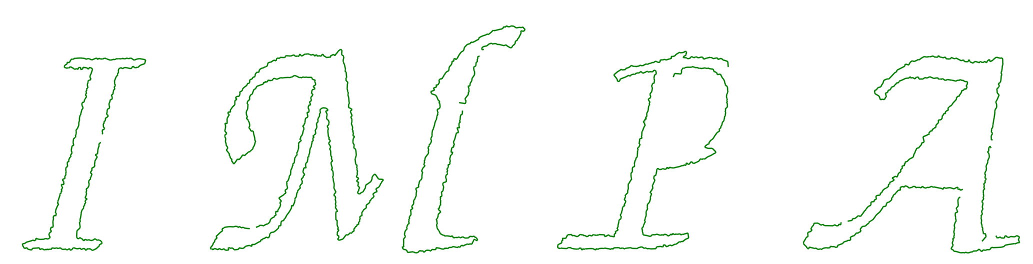


Figure 11: loop closure

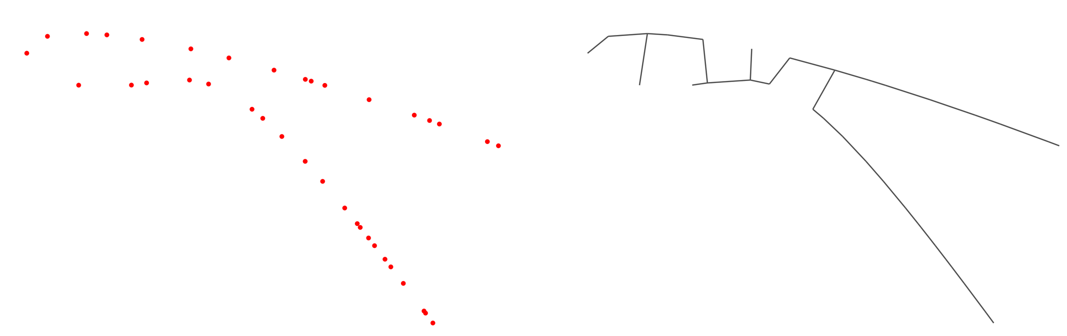


Figure 12: point cloud and MST

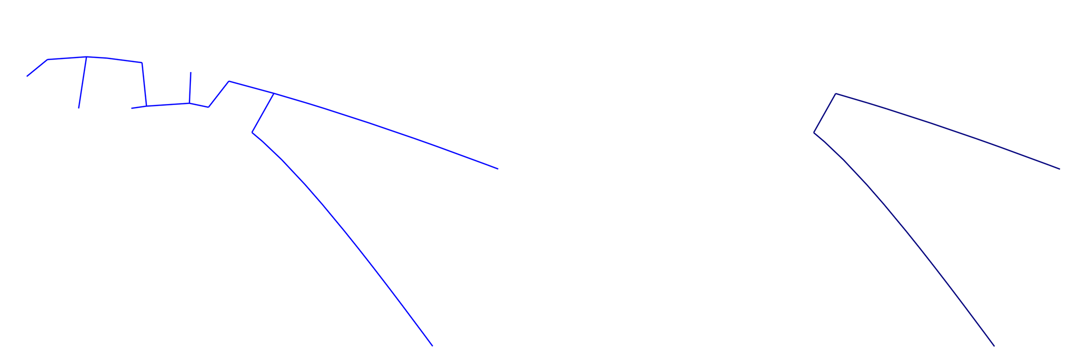


Figure 13: bridge deletion and topological noise filtering

References

[1] Luiz Henrique de Figueiredo and Jonas Gomes. Computational morphology of curves. *The Visual Computer*, 11(2):105–112, 1995.

[2] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[3] Adam Finkelstein and David H. Salesin. Multiresolution curves. In *Proceedings of SIGGRAPH 94*, Computer Graphics Proceedings, Annual Conference Series, pages 261–268, July 1994.