Region reconstruction from noisy samples

Emilio Ashton Vital Brazil Luiz Henrique de Figueiredo (orientador)
IMPA, Rio de Janeiro, Brazil

Abstract

We describe a heuristic method for reconstructing a region in the plane from a noisy sample of points. The method uses radial basis functions with Gaussian kernels to compute a fuzzy membership function which provides an implicit approximation for the region.

1. Introduction

We consider the problem of reconstructing a region in the plane from a noisy sample of points in it. Figure 1 shows the setting: Λ is a region of $\mathbf{R}^2$ and points are sampled in or near Λ. Note how the points are well distributed in the interior of Λ and that there are sample points outside Λ; these are the effect of noise in the sampling. Note also that the boundary of Λ is not sampled at all, except by accident.

The classical geometrical solutions for shape reconstruction from points, such as α-shapes [4] and β-skeletons [6], work well in the absence of noise but are too sensitive to the presence of noise, because they use *all* sample points in the reconstruction graph. We seek a method that can automatically identify points that are definitely in the interior of the region (these are trustworthy) and points that are near the boundary (these are less trustworthy because of noise).

To handle noise and to quantify the trustworthiness of each point, we approach the region reconstruction problem as a function reconstruction problem:

> Given a sample S of points well distributed in or near an unknown region $\Lambda \subseteq \mathbf{R}^2$, find an approximation $\widehat{\chi}$ for the characteristic function χ_Λ that is consistent with S.

A sample is *well distributed* when the number of sample points per area unit does not vary much in or near Λ. To model noise in the sample, we shall assume that the sampling has been done according to the density implied by an unknown *fuzzy membership function* $\widetilde{\chi}$ for Λ, that is, a function $\widetilde{\chi}\colon \mathbf{R}^2 \to [0,1]$ that satisfies

$$\widetilde{\chi}(x) = 1 \Rightarrow x \in \Lambda \quad \text{and} \quad \widetilde{\chi}(x) = 0 \Rightarrow x \notin \Lambda.$$

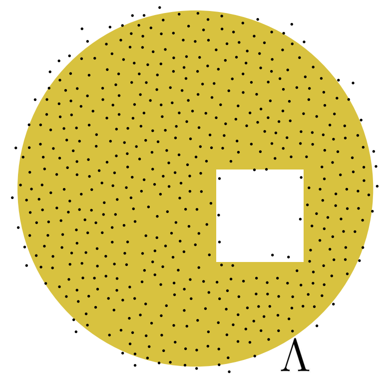


Figure 1. A noisy sample of a planar region.

In contrast, the characteristic function $\chi = \chi_\Lambda$ satisfies

$$\chi(x) = 1 \Leftrightarrow x \in \Lambda \quad \text{and} \quad \chi(x) = 0 \Leftrightarrow x \notin \Lambda.$$

Thus, a fuzzy membership function only provides partial information about Λ, but it never lies: when $\widetilde{\chi}$ coincides with χ, the membership information provided by $\widetilde{\chi}$ is correct. The more $\widetilde{\chi}$ differs from χ, the noisier the sample is. Note that the noise is concentrated near the boundary of Λ.

We want to approximate $\widetilde{\chi}$ with another fuzzy membership function $\widehat{\chi}$, and from $\widehat{\chi}$ obtain an approximation $\widehat{\Lambda}$ for Λ. Recall that both $\widetilde{\chi}$ and Λ are unknown. The only information comes from the sample points S; the values of $\widetilde{\chi}$ in S are not provided. Thus, we are not dealing with a function interpolation problem, one that could be approached by giving a constant value to every sample point. As mentioned above, the challenge is to identify the interior points.

Our solution to this region reconstruction problem computes an approximation $\widehat{\chi}$ using radial basis functions [1]. From $\widehat{\chi}$ we compute an implicit approximation $\widehat{\Lambda}$ for Λ as

$$\widehat{\Lambda} = \{x \in \mathbf{R}^2 : \widehat{\chi}(x) \geq \delta\},$$

for a suitable value of $\delta \in [0,1]$. An example of the method in action is shown in Figure 2.

To the best of our knowledge, there has been no research focused on the region reconstruction problem. Virtually all previous work has focused on the reconstruction of curves

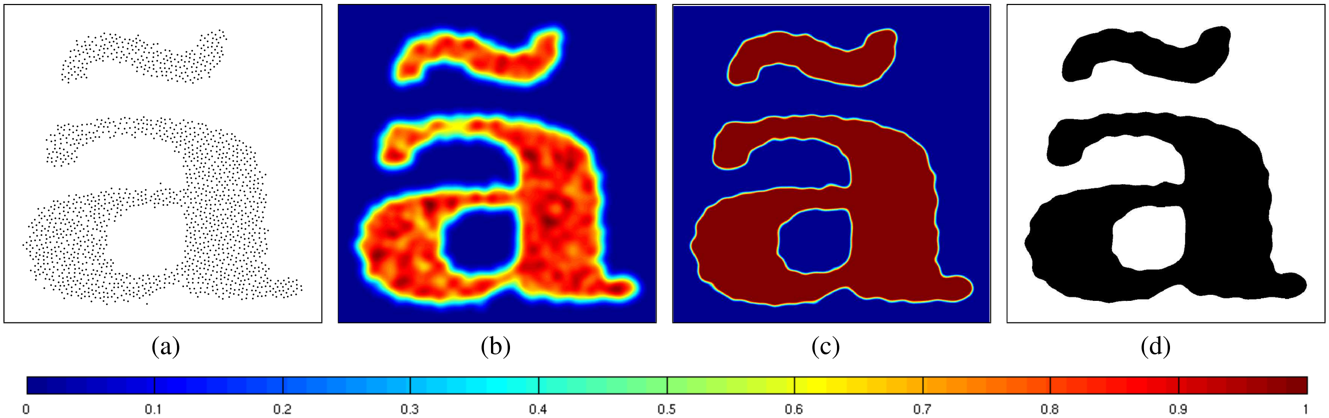


Figure 2. Overview of our reconstruction method: (a) sample points S, (b) prereconstruction function Φ, (c) fuzzy membership function $\widehat{\chi}$, (d) reconstructed region $\widehat{\Lambda}$.

and surfaces [3, 7] because of its practical importance. In these cases, the points are sampled on or near the boundary of the object, not in its interior. Moreover, most approaches that use radial basis functions attack the problem as a function interpolation problem [2]. In our case, as mentioned above, only the points are given, not the values of $\widetilde{\chi}$.

2. Approximation by spatial coherence

Here is an overview of our approximation method, which is illustrated in Figure 2.

We start by using radial basis functions [1] to combine the local influence of all sample points into a single *prereconstruction function* $\Phi\colon \mathbf{R}^2 \to \mathbf{R}$ given by

$$\Phi(x) = \sum_{\xi \in S} K_\xi(x),$$

where K_ξ is a radial basis function, or *kernel*, centered at the sample point $\xi \in S$. We shall discuss the selection of the kernel later. Having selected a suitable kernel, we define our approximation function $\widehat{\chi}$ as the following normalization of Φ:

$$\widehat{\chi}(x) = \begin{cases} 0, & \Phi(x) \le A \\ \dfrac{\Phi(x) - A}{B - A}, & A \le \Phi(x) \le B \\ 1, & \Phi(x) \ge B \end{cases}$$

where A and B will be explained later. This normalization maps the interval $[A, B]$ linearly onto $[0, 1]$, cutting values below A or above B.

Finally, as mentioned in Section 1, the region Λ is approximated implicitly by

$$\widehat{\Lambda} = \{x \in \mathbf{R}^2 : \widehat{\chi}(x) \ge \delta\},$$

for some threshold $\delta \in (0, 1]$, typically 0.5.

2.1. Choosing the kernel

Since we assume that the sample is well distributed in and near Λ, we consider only isotropic kernels, that is, functions whose value at a point $x \in \mathbf{R}^2$ depend only on the distance from x to the sample point ξ:

$$K_\xi(x) = \psi\left(\frac{|x - \xi|}{r}\right),$$

where $\psi\colon \mathbf{R}^+ \to \mathbf{R}$ is a *basis function* and r is the *radius of influence*, which we take the same for all sample points ξ.

We tested several candidates for ψ. Some had compact support and satisfied $\psi(u) = 0$ for $u > 1$. (In terms of K_ξ, this means that ξ does not influence points x that are farther than r from ξ.) We tested the following candidates for ψ:

- $\psi(u) = 1$ (constant)
- $\psi(u) = 1 - u$ (linear)
- $\psi(u) = 1 - 2u^k + u^{k+1}$ (polynomial)
- $\psi(u) = \dfrac{1}{\varepsilon^{-1}u^k + 1}$ (rational)
- $\psi(u) = e^{-\log(\varepsilon^{-1})u^k}$ (compact exponential)

where $\varepsilon > 0$ and $k \in \mathbf{N}$ are parameters.

We also tested candidates without compact support:

- $\psi(u) = e^{-u^k}$ (exponential)
- $\psi(u) = e^{-u^2/2}$ (Gaussian)

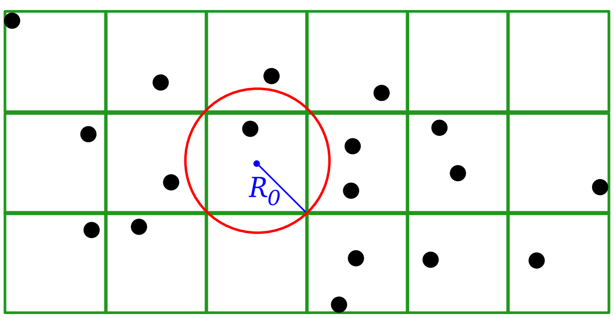
We tried several variations of the parameters involved, but the functions with compact support did not give good results. We chose the Gaussian basis function for the rest of the research because it was less sensitive to variations in the spatial uniformity of the sample. The other functions generated false holes in the reconstruction wherever there were voids in the sample (Figure 9). The only parameter left to choose to define the kernel is then the radius of influence r.

2.2. Choosing the radius

The radius of influence is the main parameter in our approach. We want to choose a radius that is adapted to the sample: small for dense samples, larger for sparse samples.

Since we assume that the sample well distributed, the number of sample points that influence the value of Φ at a point x near Λ should be approximately the same for all points x. The radius of influence measures not only the density of the sample but also its spatial structure, if any. For instance, if the sample is taken on a rectangular grid, we expect that about 4 sample points will influence any given point; if the sample is taken via a Poisson process, we expect that about 6 sample points will influence any given point.

We choose the radius automatically (but empirically) as follows. If the sample was uniformly distributed in a rectangle Ω containing Λ, and if we laid a regular grid of square cells so that each cell contained just one sample point, then the diameter of the cells would be $\sqrt{2 \cdot \text{area}(\Omega)/N}$, and so the radius of the disk circumscribing a cell would be $R_0 = \sqrt{\text{area}(\Omega)/2N}$, as shown below:



However, the sample is not uniformly distributed in Ω, because it is concentrated near Λ. So, we lay a regular grid of square cells of side $2R_0$ in Ω and look at the sample points that land in each cell. For each sample point ξ, we compute the radius $R(\xi)$ of the smallest ball centered at ξ containing least n sample points, for a small fixed $n \geq 2$, which will depend on the spatial structure of the sample. If a cell C contains more than n sample points, we take $R(C)$ to be the average of all $R(\xi)$ for $\xi \in C$. If C contains less than n sample points, we take $R(C) = R_0$. We take the average of all $R(C)$ as our first estimate $\widehat{r}$ of the influence radius r.

The final value of r is computed as follows: We need to decide which n to use. We start with $n = 2$ and compute the relative standard deviation $\sigma' = \sigma/\widehat{r}$ of the $R(\xi)$. If $\sigma' \leq 0.01$, we infer that the sample has strong spatial structure and we take $r = \widehat{r}/2$. If $\sigma' \leq 0.25$, we infer that the sample has some spatial structure and we take $r = \widehat{r}$. If $\sigma' > 0.25$, we increase n and repeat the previous analysis until we reach $\sigma' \leq 0.25$ or $n = 12$. If we reach $n = 12$ without reaching $\sigma' \leq 0.25$, then we pronounce the sample to be not well distributed and abort the reconstruction. (This never happened in our tests.)

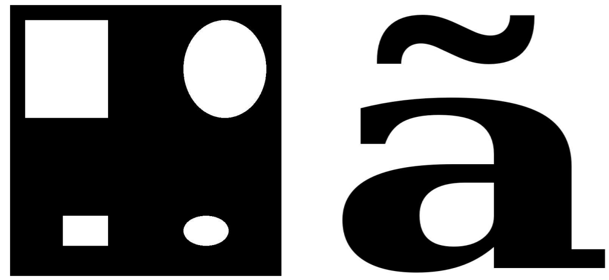


Figure 3. Regions used in tests: 'Q' and 'a'.

2.3. Choosing the normalization parameters

It remains to tell how we choose the parameters A and B for normalizing the pre-reconstruction function Φ. Again, this is done empirically. We set $A = 0.7c$, $B = 0.5c$, where

$$c = \frac{2\max\{\Phi(x) : x \in \Omega\}}{\log(N)}$$

3. Results

We tested and evaluated our reconstruction method for several regions and sampling conditions. Here we report on the results for the two synthetic regions Λ shown in Figure 3. The region on the left was chosen to try to assess the effect of smoothing on the edges. The region on the right was chosen to try to assess how well the reconstruction handled topological features, such as multiple connected components and holes.

We selected a rectangular region Ω containing Λ. As mentioned in Section 1, the sampling was done in Ω according to the density implied by a fuzzy membership function $\widetilde{\chi}$ for Λ, using a simple rejection method [8]. To simulate noise, we used a convolution of χ_Λ with a Gaussian low-pass filter with σ^2 equal to 2% and 4% of the area of Ω, as shown in Figure 4.

To test the behavior of samples with spatial structure, we chose four sampling schemes: points in a regular grid, points in a perturbed regular grid, points with a Poisson disk distribution, and points uniformly distributed in Ω with no spatial structure, as shown in Figure 5. The actual samples used in the tests are shown in Figure 10.

We used two error measures to quantify the quality of the reconstruction in Ω:

$$\mathscr{E}_{\widehat{\chi}} = \frac{1000}{\text{area}\,(\Omega)} \int_\Omega \left(\widetilde{\chi}(x) - \widehat{\chi}(x)\right)^2 dx$$

$$\mathscr{E}_{\widehat{\Lambda}} = \frac{1000}{\text{area}\,(\Omega)} \int_\Omega \left(\chi_\Lambda(x) - \chi_{\widehat{\Lambda}}(x)\right)^2 dx$$

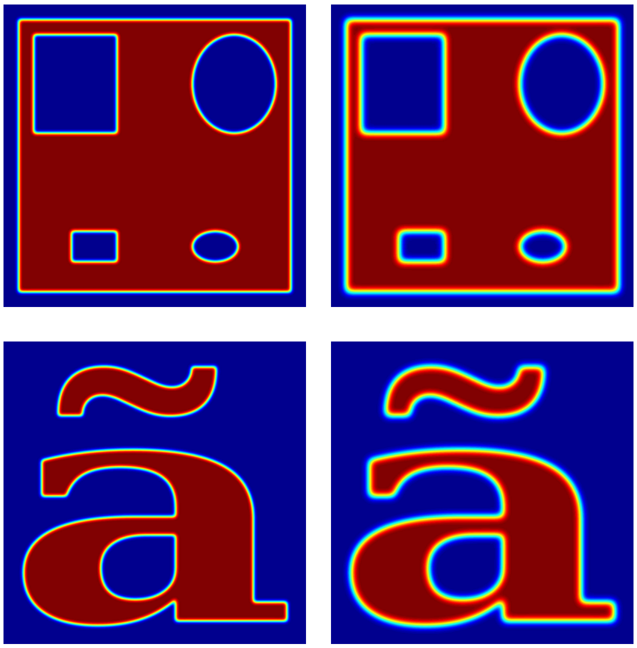


Figure 4. Fuzzy membership functions.

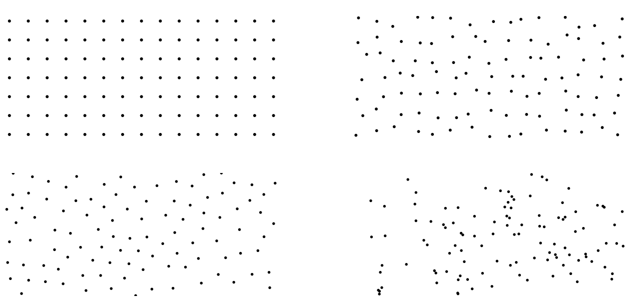


Figure 5. Sampling schemes: regular grid, perturbed regular grid, Poisson disk distribution, no spatial structure.

500	RG		PRG		PD		U	
χ_Q	**082**	081	088	076	092	**076**	194	172
$\widetilde{\chi}_Q$ 2%	082	081	**079**	**079**	080	079	178	172
$\widetilde{\chi}_Q$ 4%	100	099	**082**	**085**	084	089	167	171
χ_a	**061**	**058**	077	062	086	066	175	162
$\widetilde{\chi}_a$ 2%	073	072	**065**	**063**	074	069	165	162
$\widetilde{\chi}_a$ 4%	077	079	**068**	**070**	075	077	158	164
1000	**RG**		**PRG**		**PD**		**U**	
χ_Q	**050**	**047**	056	053	065	054	137	118
$\widetilde{\chi}_Q$ 2%	060	056	**056**	**056**	059	059	123	119
$\widetilde{\chi}_Q$ 4%	091	082	**073**	**068**	073	070	120	123
χ_a	**041**	**040**	052	043	055	043	116	102
$\widetilde{\chi}_a$ 2%	055	054	**047**	**047**	048	047	106	104
$\widetilde{\chi}_a$ 4%	071	067	**058**	**055**	058	056	102	107
4000	**RG**		**PRG**		**PD**		**U**	
χ_Q	**027**	**027**	036	034	040	038	060	047
$\widetilde{\chi}_Q$ 2%	047	040	**044**	**040**	045	042	052	050
$\widetilde{\chi}_Q$ 4%	088	078	071	056	070	057	**065**	**056**
χ_a	**028**	**027**	032	030	032	029	050	041
$\widetilde{\chi}_a$ 2%	049	046	**038**	**035**	038	036	043	042
$\widetilde{\chi}_a$ 4%	072	063	061	049	057	**046**	**052**	048
10000	**RG**		**PRG**		**PD**		**U**	
χ_Q	**022**	**021**	025	023	028	027	037	030
$\widetilde{\chi}_Q$ 2%	051	044	042	036	042	036	**036**	**033**
$\widetilde{\chi}_Q$ 4%	084	072	072	055	070	053	**057**	**040**
χ_a	**018**	**018**	021	021	023	022	031	026
$\widetilde{\chi}_a$ 2%	046	043	036	031	035	030	**031**	**024**
$\widetilde{\chi}_a$ 4%	069	058	061	047	058	044	**047**	**035**

Table 1. Average reconstruction error. From top to bottom: 500, 1000, 4000, 10000 sample points. From left to right: regular grid; perturbed regular grid; Poisson disk distribution; no spatial structure. Best results shown in bold.

Table 1 shows the average reconstruction errors. Each block gives $\mathscr{E}_{\widehat{\chi}} \mid \mathscr{E}_{\widehat{\Lambda}}$ for the test regions shown in Figures 3 and 4. The best results are shown in bold. The numbers suggest that the two error measures are essentially the same. We also performed convergence tests and they show that the reconstruction errors decrease fairly fast as the size of the sample increases.

The reconstructions obtained with our method for the samples shown in Figure 10 are shown in Figure 11 for noiseless samples and in Figure 12 for noisy samples; we show the reconstructed region $\widehat{\Lambda}$ and the reconstruction error $|\chi_\Lambda - \chi_{\widehat{\Lambda}}|$. For noiseless samples (Figure 11), the reconstruction was better for spatially structured samples (RG, PRG, PD), even for small samples. For noisy samples (Figure 12), spatially structured samples gave better results when the sample was small, but unstructured samples gave better results when the sample was large.

On the other hand, the numerical differences in reconstruction performance shown in Table 1, which are given in thousandths of the area of Ω, do not always translate into clear visual differences in Figures 11 and 12. This suggests that our reconstruction method is robust and works for different kinds of samples. Of course, the larger the sample, the better the reconstruction.

Figure 6 shows a reconstruction of a Miró print that tries to recreate the imaginary region sampled by the artist. Figure 7 shows the effect of smoothing on the edges. As expected, the smoothing decreases with the sample size.

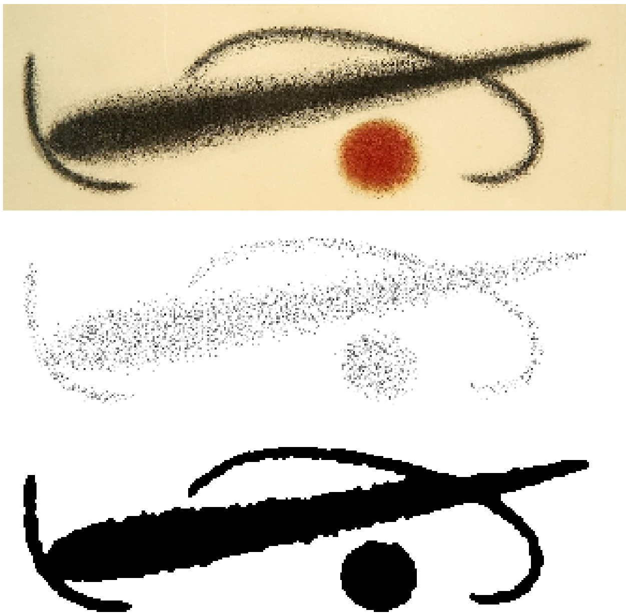


Figure 6. Reconstructing a Miró print. From top to bottom: original print, sample points, reconstructed region.

4. Conclusion

Although it is based on heuristics and empirical choice of parameters, the region reconstruction method that we have proposed is simple to understand and implement. The implicit representation provided for the reconstructed region $\widehat{\Lambda}$ can be exploited for geometric processing, such as area computation and boundary evaluation.

As shown by our tests, the method can reliably identify the sample points that are clearly inside Λ and the sample points that are probably near the boundary of Λ. Figure 8 shows the fuzzy membership function $\widehat{\chi}$ for the reconstructions shown in Figures 11 and 12. The empirically chosen normalization parameters A and B allow us to identify the interior points quite well; boundary points are limited to a narrow band. The method faithfully reproduced all topological features and successfully reconstructed almost all of the interior of Λ, as quantified by our error measures (the typical error was around 5%). On the other hand, the boundary of $\widehat{\Lambda}$ is not as smooth as we tend to expect. (This expectation is probably due to the strong intuitive meaning of the test shapes.)

At least two lines of research are natural from this point. One is to use principal components analysis (PCA) [5] to generate anisotropic kernels and try to improve the smoothness of the reconstructed boundaries. Another is that an analysis of the variation of the radius $R(\xi)$ can be used to determine the spatial structure of the sample.

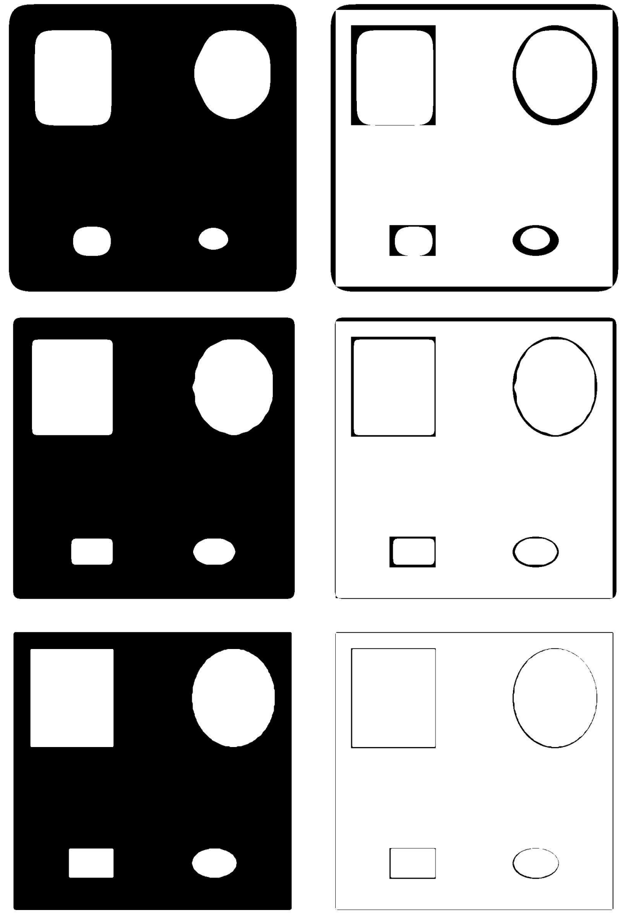


Figure 7. Smoothing effect on boundaries. Left: reconstructed region; right: reconstruction error. From top to bottom: 1000, 10000, and 100000 points.

Acknowledgements. This work is part of the first author's M.Sc. work at IMPA [9]. The authors are partially supported by CNPq. This work was done in the Visgraf laboratory at IMPA, which is sponsored by CNPq, FAPERJ, FINEP, and IBM Brasil.

References

[1] M. D. Buhmann. *Radial Basis Functions*. Cambridge University Press, 2003.

[2] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of SIGGRAPH '01*, pages 67–76. ACM Press, 2001.

[3] T. K. Dey. *Curve and Surface Reconstruction: Algorithms with Mathematical Analysis.* Cambridge University Press, 2006.

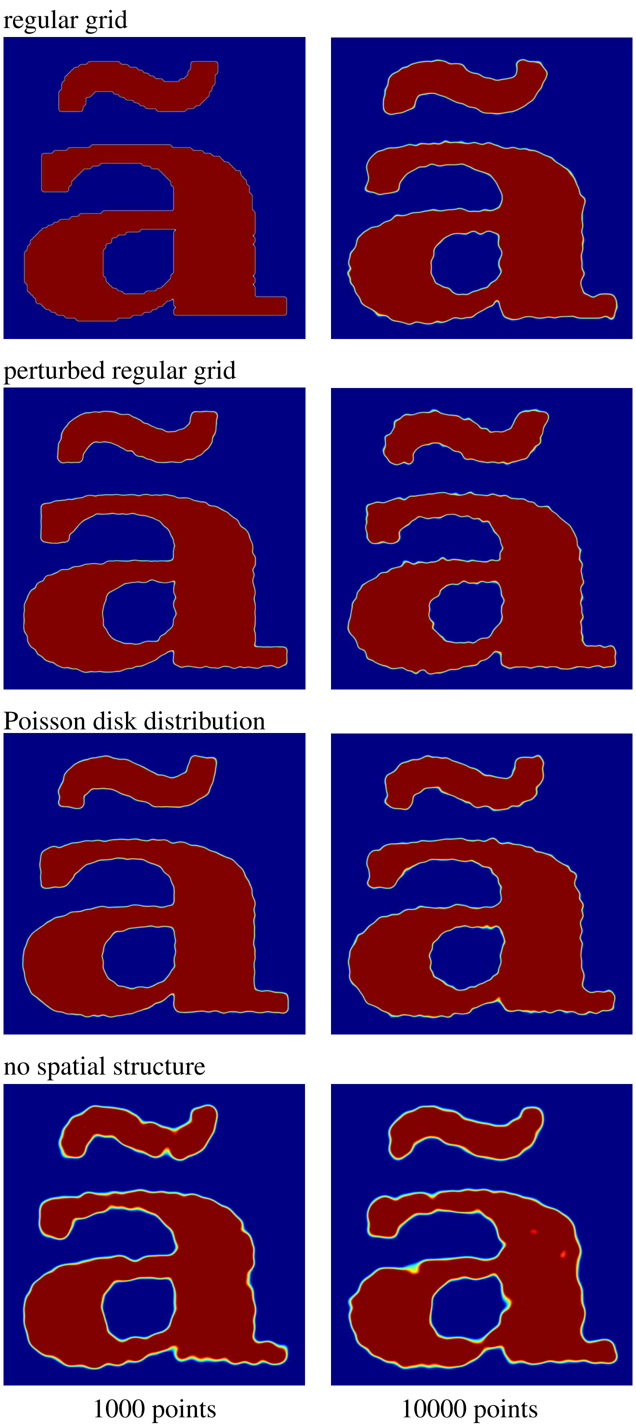


Figure 8. Fuzzy membership functions for 10000 sample points. Left column: noiseless, right column: 2% noise. From top to bottom: regular grid; perturbed regular grid; Poisson disk distribution; no spatial structure.

[4] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–558, 1983.

[5] I. T. Jolliffe. *Principal component analysis*. Springer Series in Statistics. Springer-Verlag, New York, 1986.

[6] D. G. Kirkpatrick and J. D. Radke. A framework for computational morphology. In G. T. Toussaint, editor, *Computational Geometry*, pages 217–248. North-Holland, 1985.

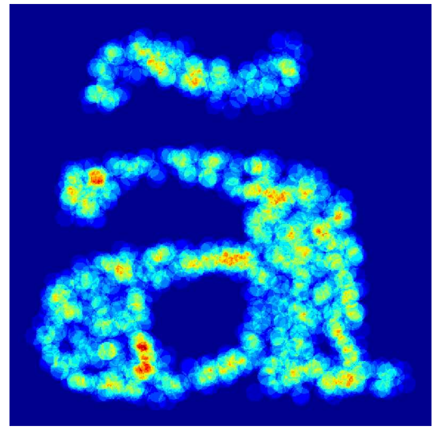
[7] Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, and H.-P. Seidel. Multi-level partition of unity implicits. In *Proceedings of SIGGRAPH '03*, pages 463–470. ACM Press, 2003.

[8] S. M. Ross. *Introduction to Probability Models*. Academic Press, third edition, 1985.

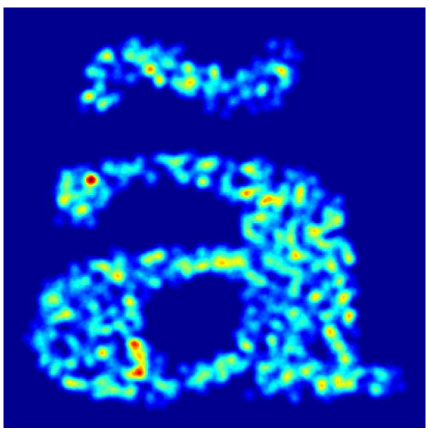
[9] E. A. Vital Brazil. Reconstrução de regiões a partir de amostras com ruído. Master's thesis, IMPA, March 2007.

(a) fuzzy membership function

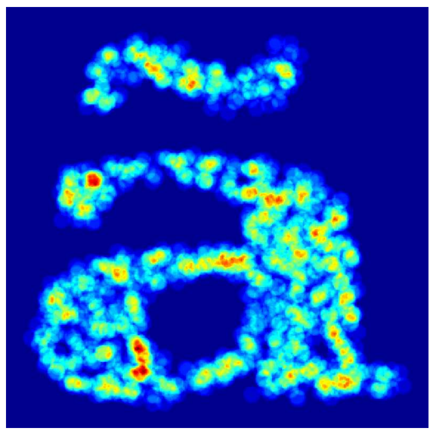
(b) 10000 sample points



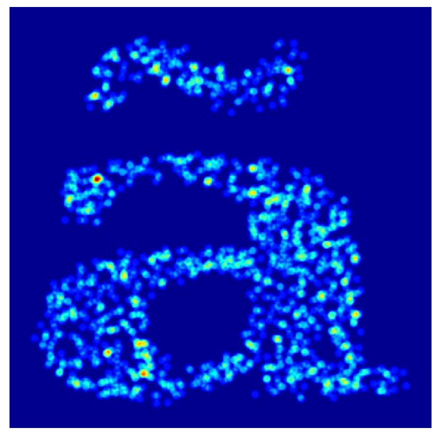
(c) constant kernel



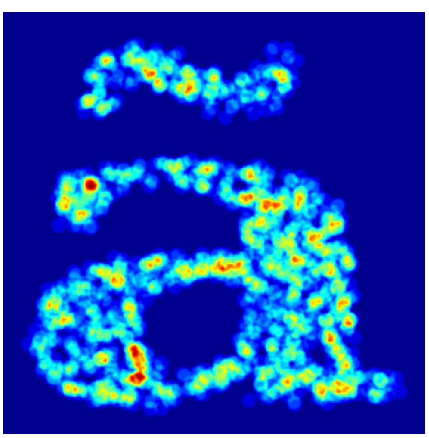
(d) linear kernel



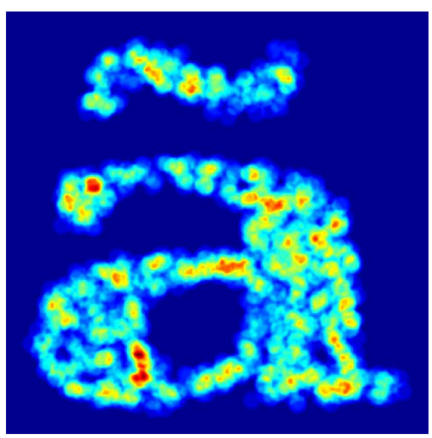
(e) polynomial kernel



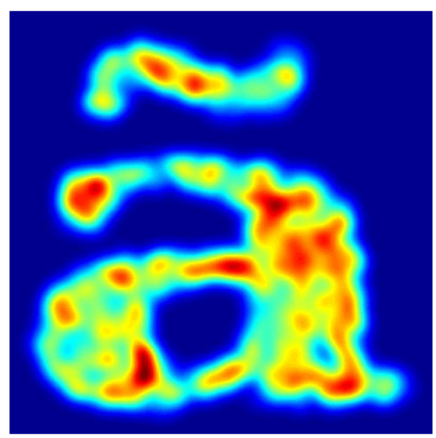
(f) rational kernel



(g) compact exponential kernel



(h) exponential kernel



(i) Gaussian kernel

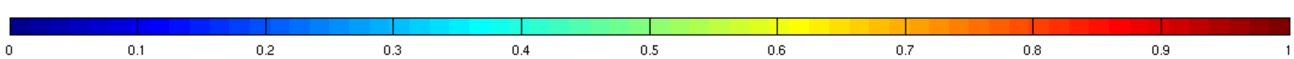


Figure 9. Pre-reconstruction for several kernels. False holes appear in the dark blues areas in the interior of the region, for all kernels except the Gaussian kernel.

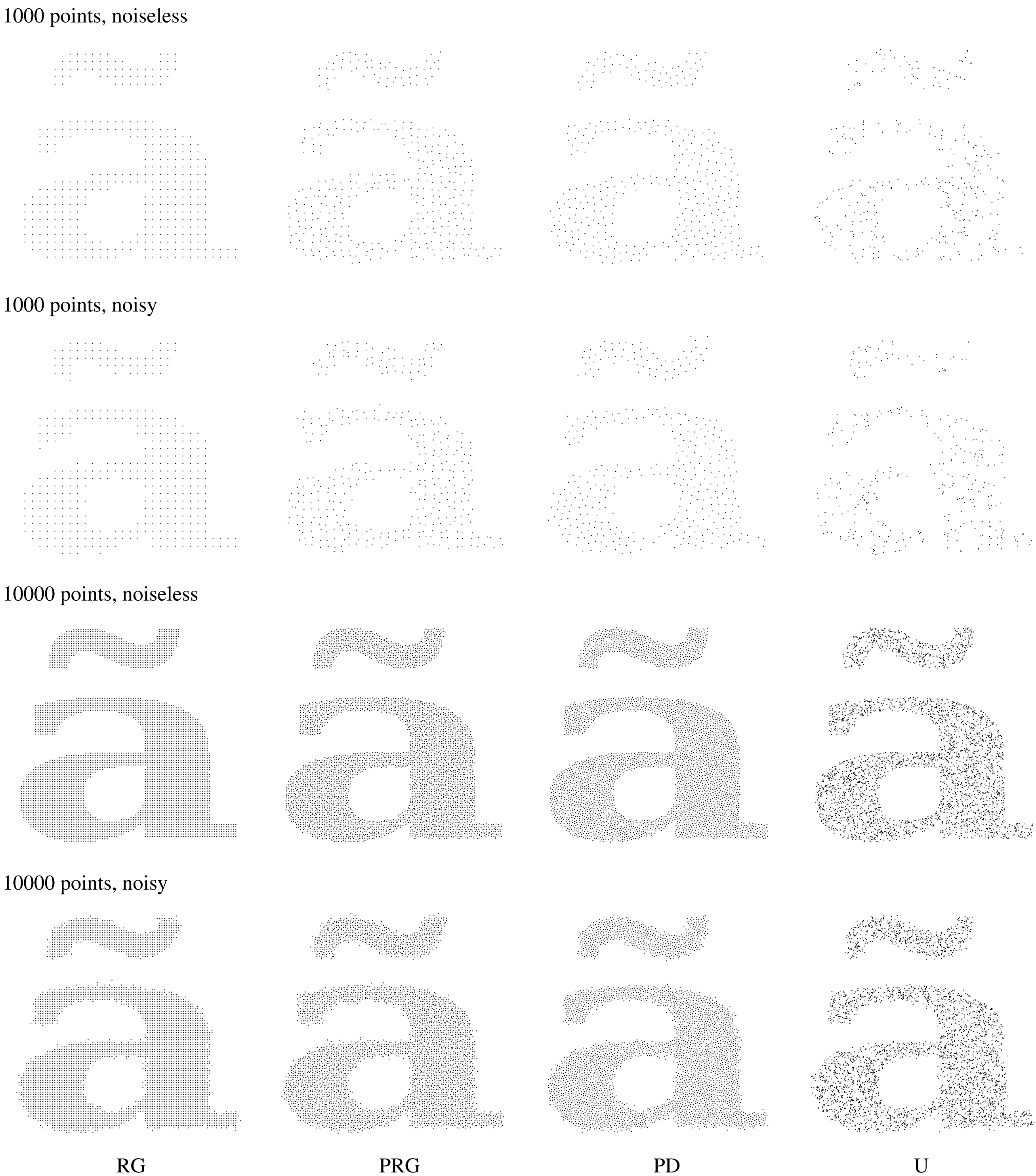


Figure 10. Samples used in tests. From top to bottom: 1000 sample points, noiseless and noisy; 10000 sample points, noiseless and noisy. From left to right: regular grid; perturbed regular grid; Poisson disk distribution; no spatial structure.

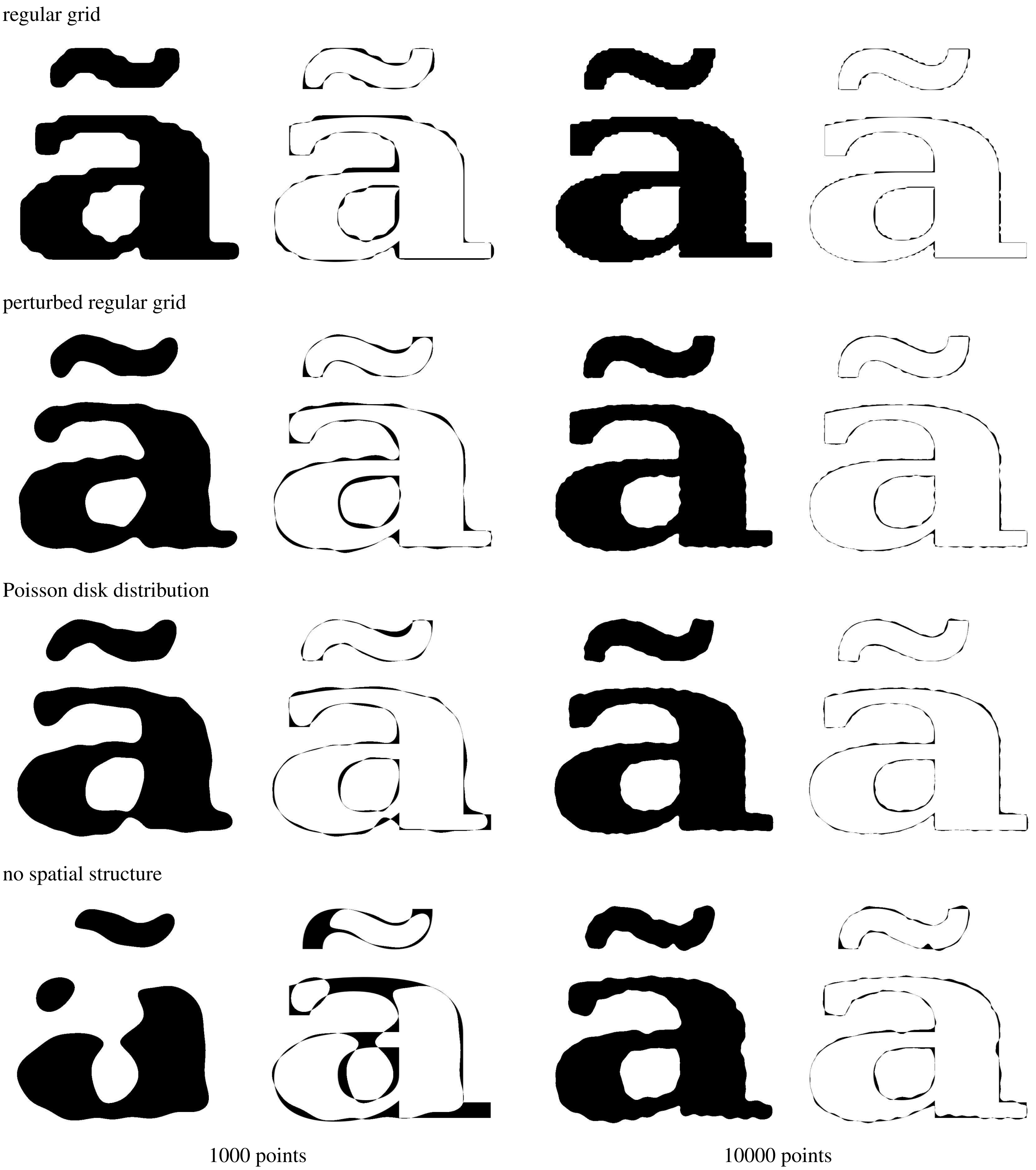


Figure 11. Reconstructed regions and reconstruction errors for noiseless samples. Left two columns: 1000 sample points; right two columns: 10000 sample points. From top to bottom: regular grid; perturbed regular grid; Poisson disk distribution; no spatial structure.

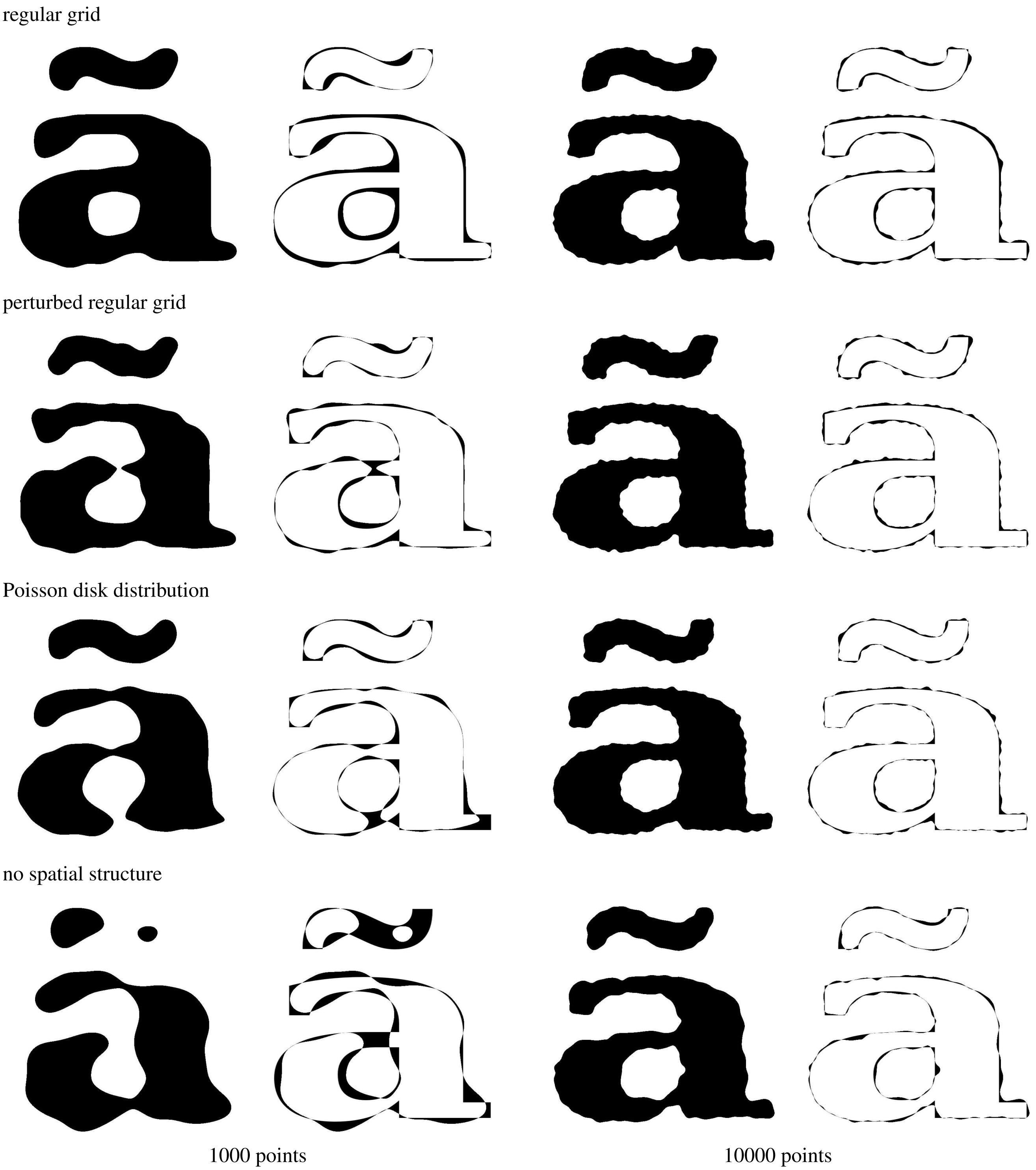


Figure 12. Reconstructed regions and reconstruction errors for samples with 2% noise. Left two columns: 1000 sample points; right two columns: 10000 sample points. From top to bottom: regular grid; perturbed regular grid; Poisson disk distribution; no spatial structure.