A two-step method for texture/geometry inpainting

Simon Masnou

Laboratoire Jacques-Louis Lions

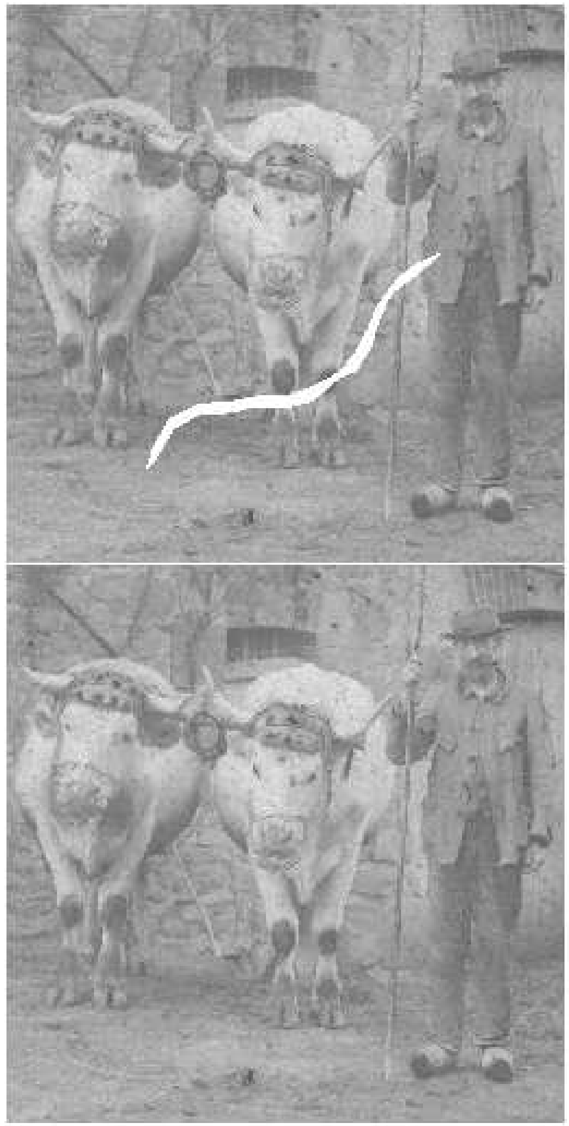
Université Paris 6

in collaboration with
- Frédéric Cao (Irisa, Rennes)
- Yann Gousseau (ENST, Paris)
- Patrick Pérez (Irisa, Rennes)

Inpainting (or disocclusion) refers to the problem of recovering missing areas in a digital image

Many applications:

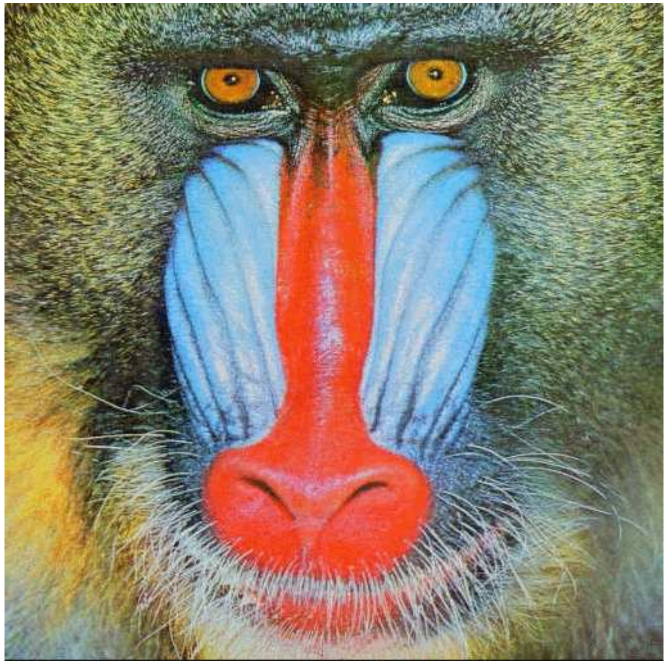
- restoration of old photographs;
- digital photo edition;
- digital video post-processing;
- special effects.



Problem : how to interpolate an image in the most realistic way ?

Main difficulty

How to deal simultaneously with geometry and texture ?



- geometry, i.e. shapes and objects

  (eyes, nose, etc.)

- texture, i.e. a local, stationary

  and (possibly) oscillatory signal (fur)

In the last ten years, many methods have been developed for interpolating either

- the geometry (Masnou-Morel '98, Chan-Shen '01, Caselles et al '01, Bertalmio et al '01, Bornemann et al '05, Tschumperlé-Deriche '04, etc.);
- or the texture – in the context of texture synthesis (Efros-Leung '99, Wei-Levoy '00, Efros-Freeman '01, Igehy-Pereira '97, Ashikmin et al '01, de Bonet '97, Harrison '01, Gousseau '00, etc.).

Only recently, different approaches have been proposed for the simultaneous interpolation of both geometry and texture (Bertalmio et al '03, Bornard et al '02, Criminisi-Pérez-Toyama '04, Starck-Fadili '04, etc.)

Example 1: Bertalmio, Vese, Sapiro, Osher : prior separation of the image into geometry and texture (cf talks by Y. Meyer, L. Blanc-Ferraud and J.-F. Aujol). Both components are interpolated independently and the final image is recovered by addition. The results are not fully satisfactory.

Example 2: J.-L. Starck- M.J. Fadili (cf talks): much more elegant approach. The interpolated image admits an "optimal" representation both in a basis that only "sees" the geometry and in a basis that only "sees" the texture. Much more convincing results.

From texture synthesis to inpainting

Most classical methods assume that the image is a realization of a *Markov Random Field*; the synthesis is possible after the distribution law has been learnt on the whole image. This gives terribly slow algorithms that actually do not perform very well.

A new class of very heuristic methods (Efros-Leung '99, Wei-Levoy '00, Liang et al '01, Criminisi-Pérez-Toyama '04) has been proposed in recent years. They are based on a simple "copy-and-paste" procedure.

Principle

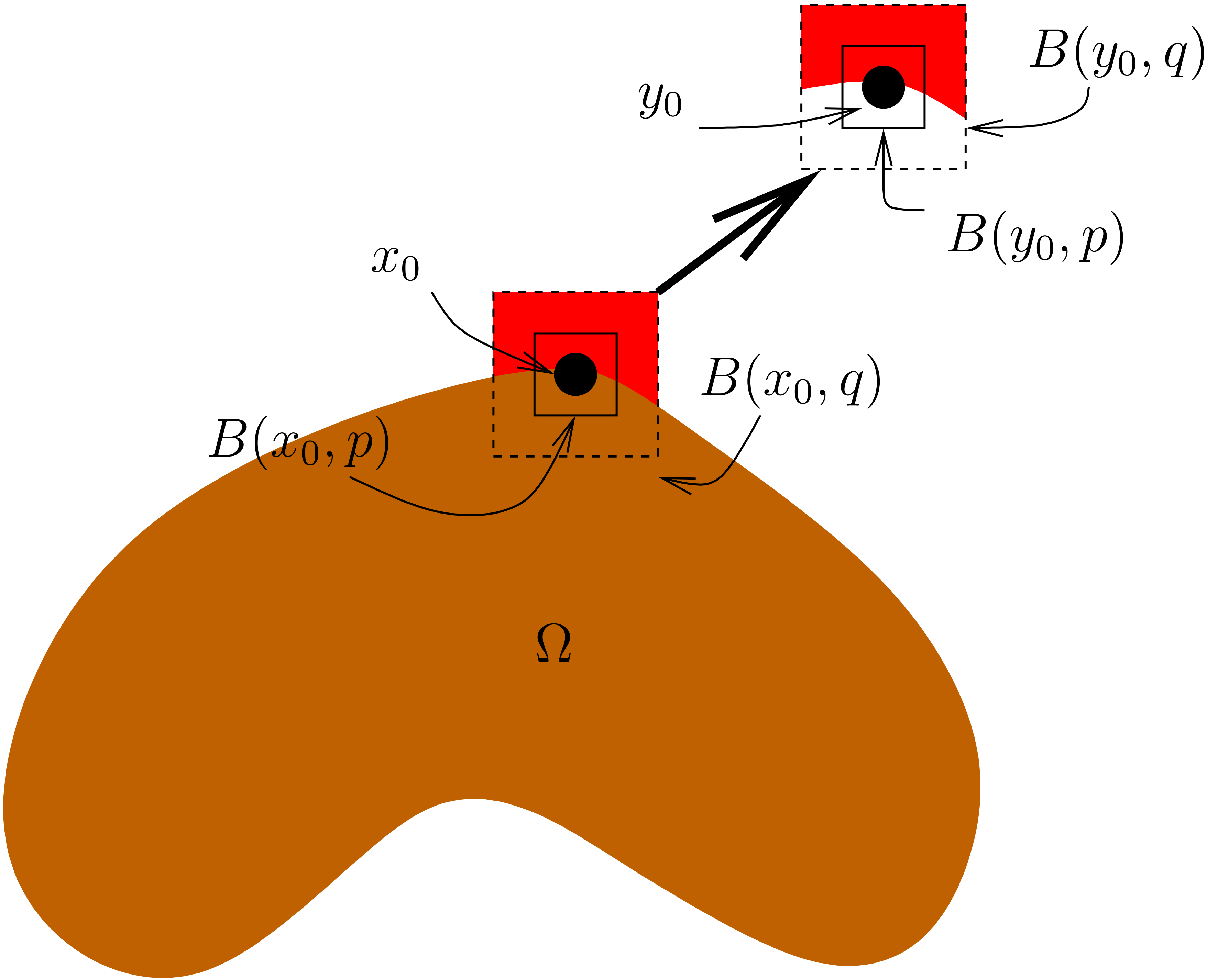
Let Ω denote the hole and $0 < p < q$ two parameters.

1) take $x_0 \in \partial\Omega$ maximizing the number of neighbors in Ω^c.

2) find y_0 in Ω^c that minimizes the L^2 norm between the patches $B(x_0, p) \setminus \Omega$ and $B(y_0, p) \setminus (\Omega + y_0 - x_0)$.

3) for each $x \in B(x_0, p) \cap \Omega$ let $I(x) = I(x + y_0 - x_0)$.

4) Replace Ω with $\Omega \setminus B(x_0, p)$ and iterate.

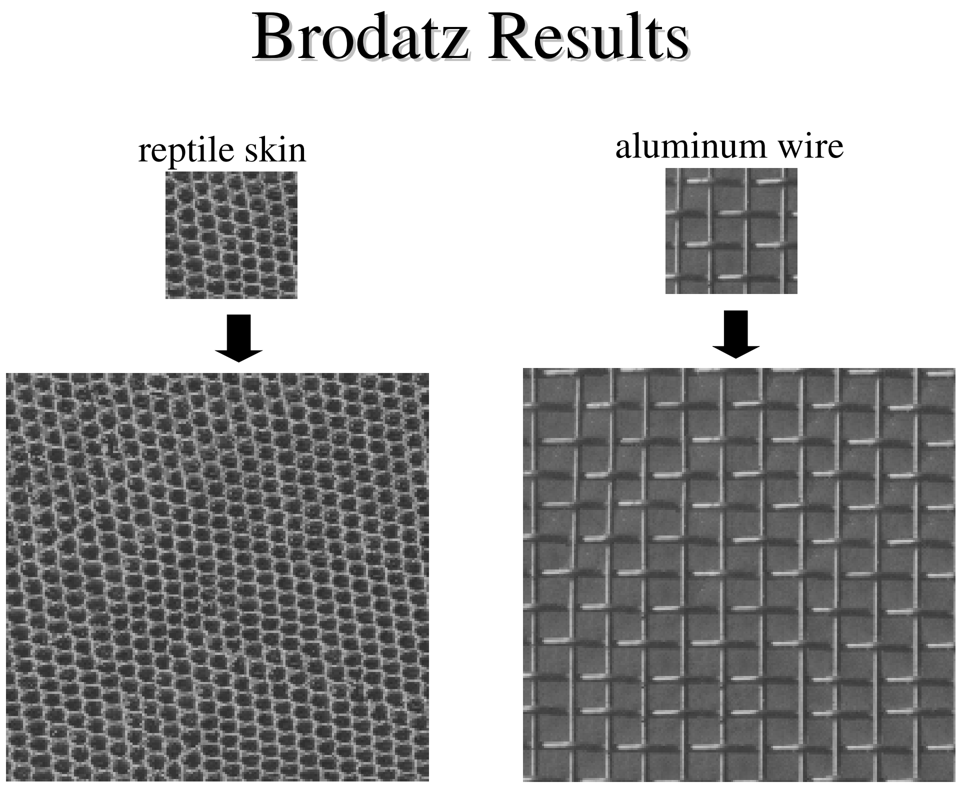


$B(y_0, q)$
y_0
$B(y_0, p)$
x_0
$B(x_0, q)$
$B(x_0, p)$
Ω

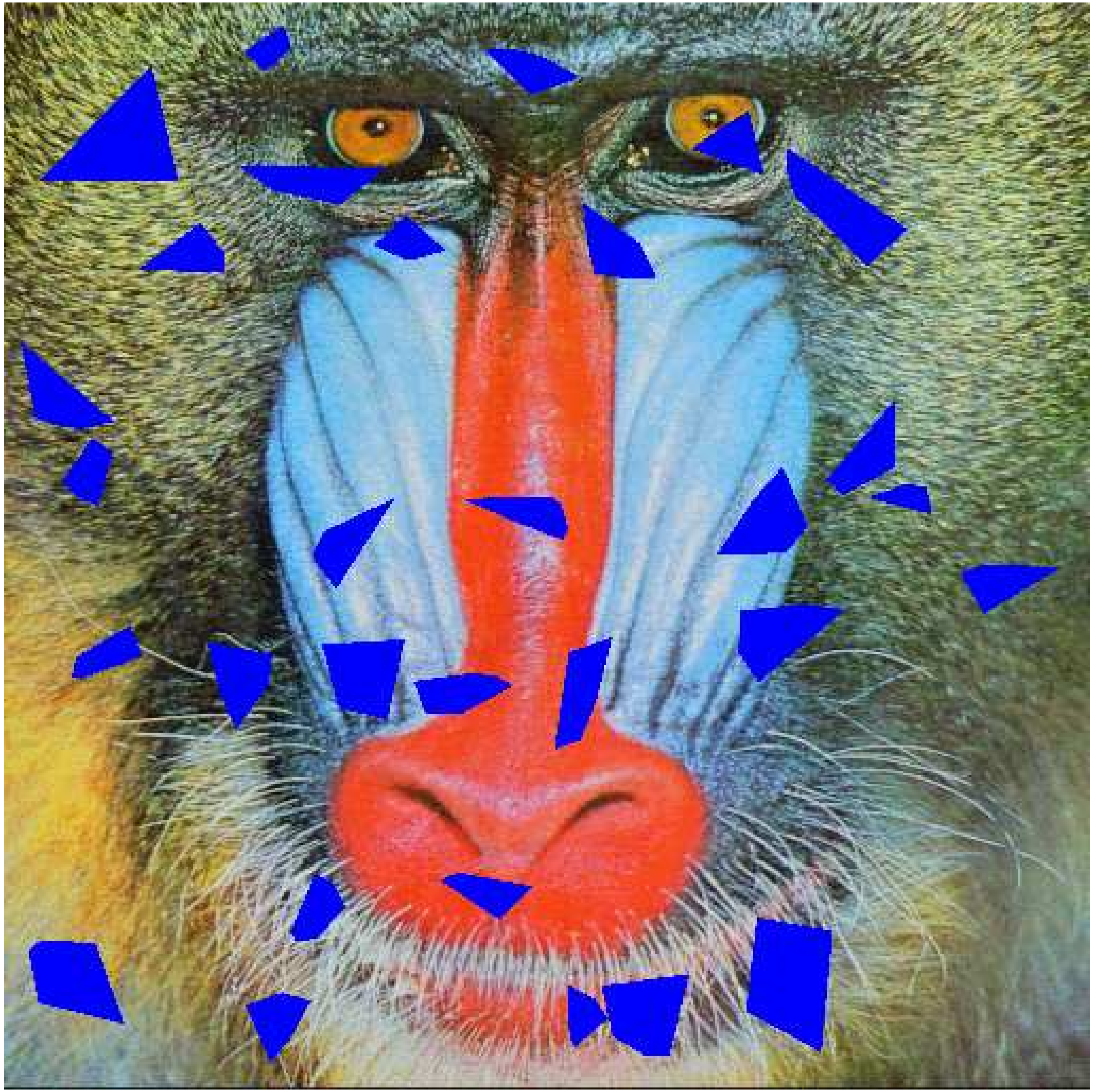
There are many variants of this algorithm (e.g. one randomly picks a patch among all those that are close enough). The results may be very different depending on

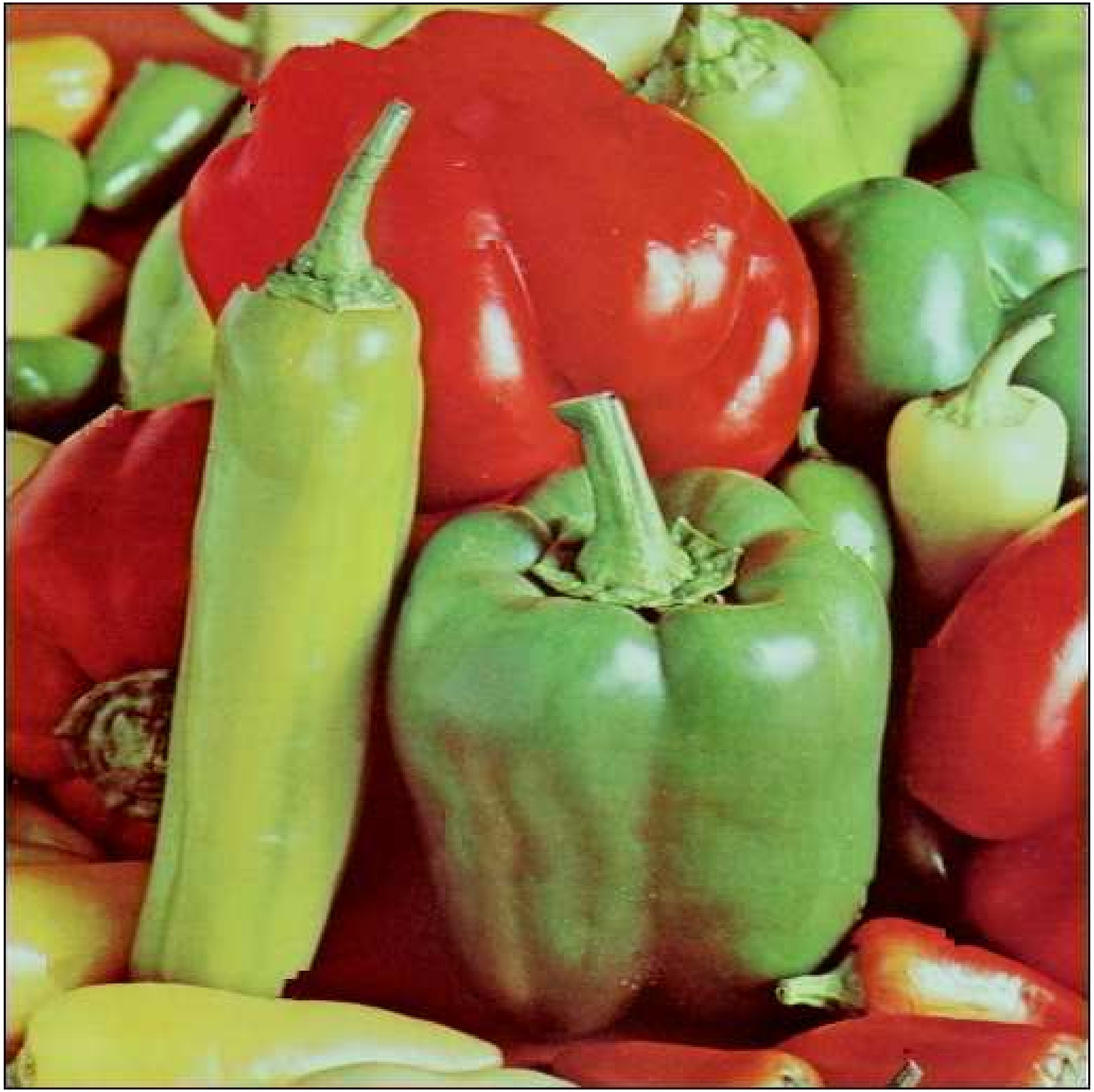
1. the choice of p, q. In particular, it is always better to take $p > 0$, i.e. to copy patches rather than points, and to choose $q > p$ in order to ensure good transitions;
2. the scanning order, i.e. which x_0 should be chosen at every iteration. In [Criminisi-Pérez-Toyama 2004], x_0 is taken as the minimizer of a functional depending on
    i) the local geometry of the hole,
    ii) the local geometry of the level lines around x_0.

These methods give very impressive results on texture...



but not only...





However this process is too low-level for those situations where one needs a non local interpretation of image.

Better results can be obtained if

(i) the dictionary of patches in image is enlarged by taking all rotated patches

(ii) the process is iterated

but still some geometric failures can be observed.

The previous method is able to restore

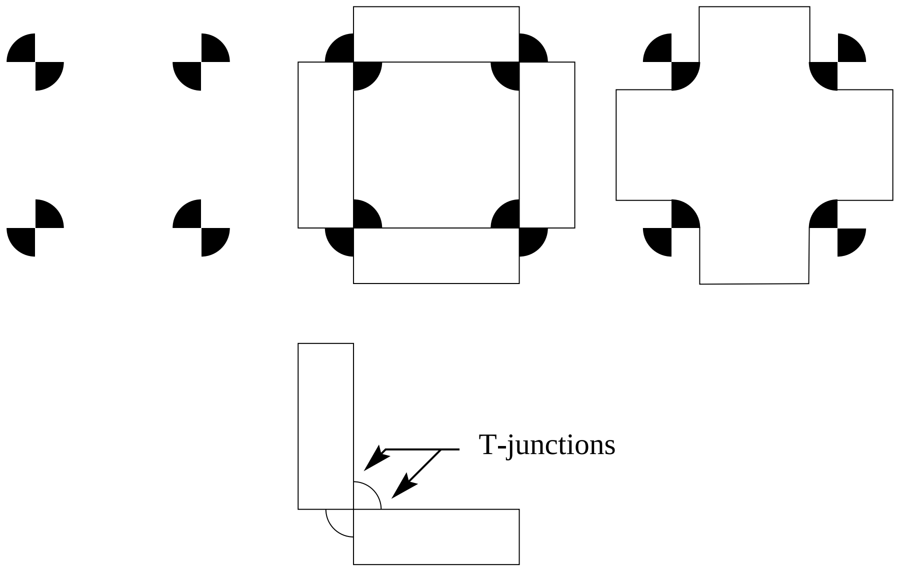
1. the texture
2. the stationary geometric structures

but fails at recovering long-range, unstationary geometric structures.

$\Rightarrow$ how can we introduce long-range geometric constraints ?

Amodal completion

Human visual system is able to reconstruct partially occluded objects by amodal completion

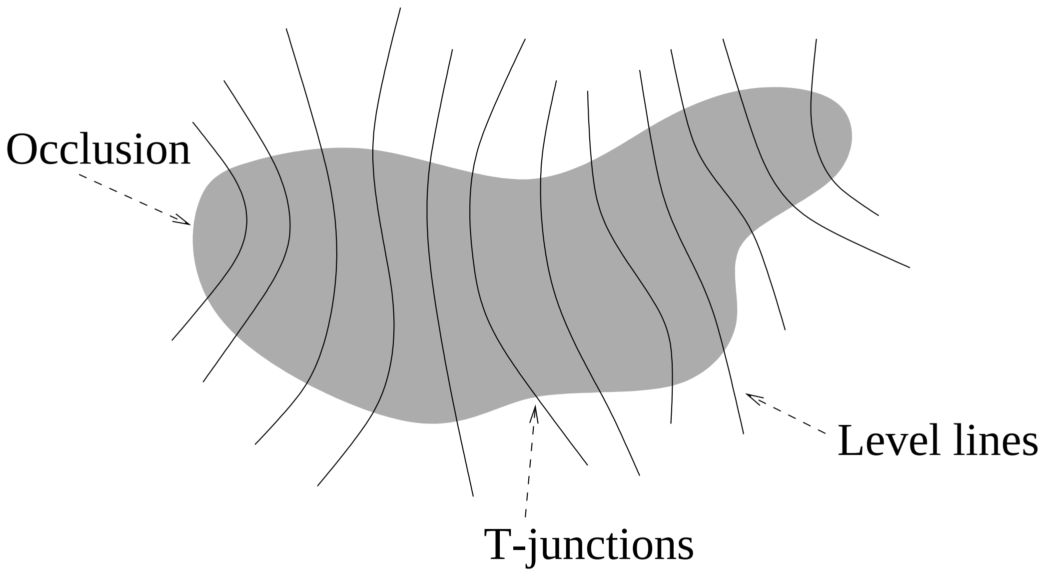


G. Kanizsa ⇒ our visual system performs edge continuation between T-junctions

⇒ the restored edges are generally short and not too curvy.

Adaptation to image interpolation

[Masnou-Morel '98]



1) Decomposition of image into its level lines = boundaries of level sets $\{u \geq t\}$ = complete and reliable image representation, well-suited for BV functions.

2) Level lines intersect the occlusion at **T-junctions**.

3) Application of **Kanizsa's principle** : level lines are continued between T-junctions.

4) **Minimize** among all possible pairwise connections of the T-junctions in order to obtain short and not too curvy curves that do not cross each other.

$$E(u) = \int_{-\infty}^{+\infty} \left(\sum_{T-junctions} \int_0^{\mathcal{L}} (1 + |\kappa(s)|^p) ds \right) d\lambda$$

with $\kappa =$ level line curvature.

5) Get the restored image by **reconstruction**

$$u(x) = \sup\{\lambda : x \in \{u \geq \lambda\}\}$$

This is “equivalent” to minimizing

$$\int_{\Omega} |\nabla u| \left(1 + \left|\operatorname{div} \frac{\nabla u}{|\nabla u|}\right|^{p}\right) dx$$

among all functions that coincide with the original image outside the hole.

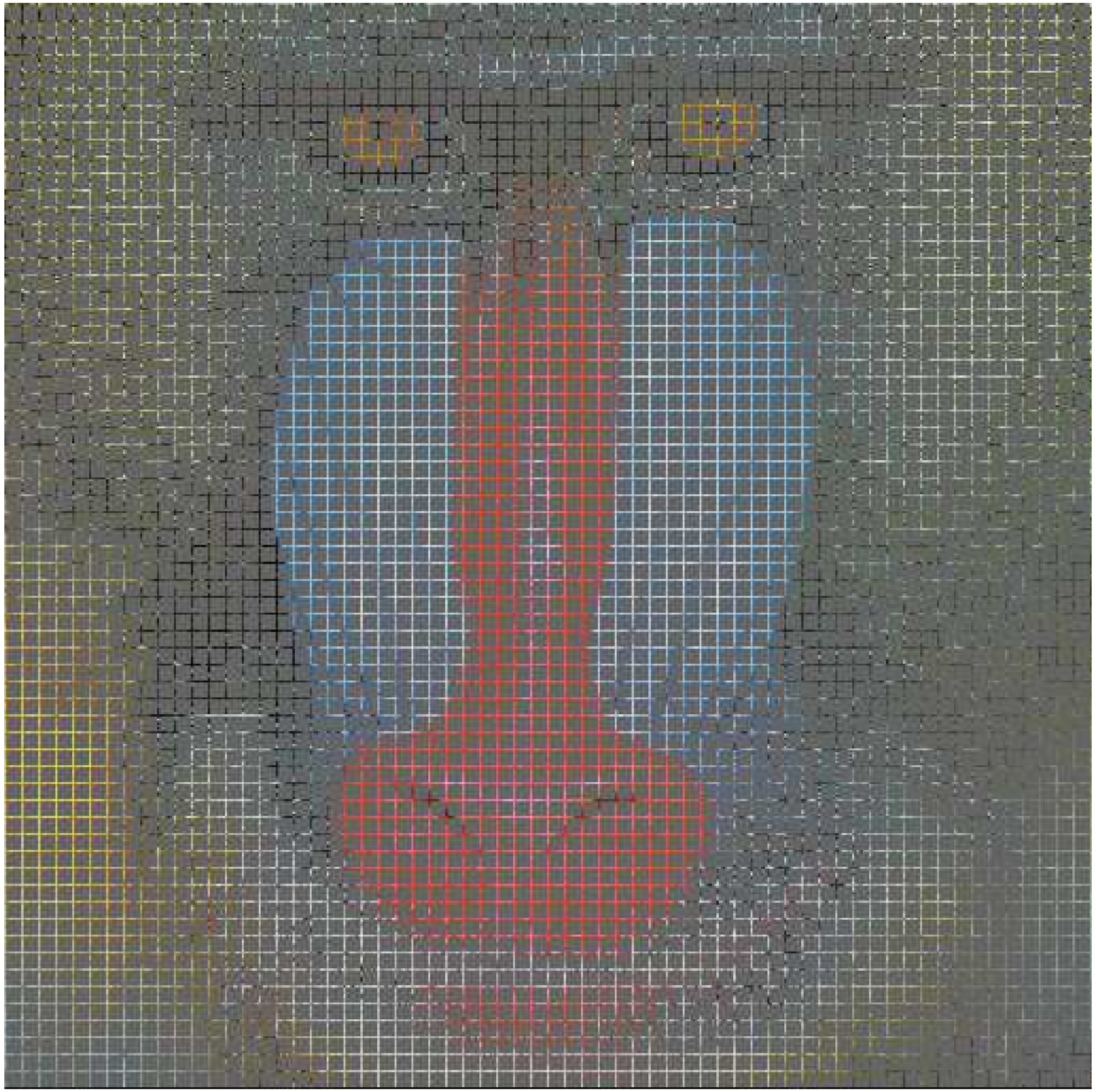


Figure 1: Squares 7x7 removed

Figure 2: After disocclusion (7x7)

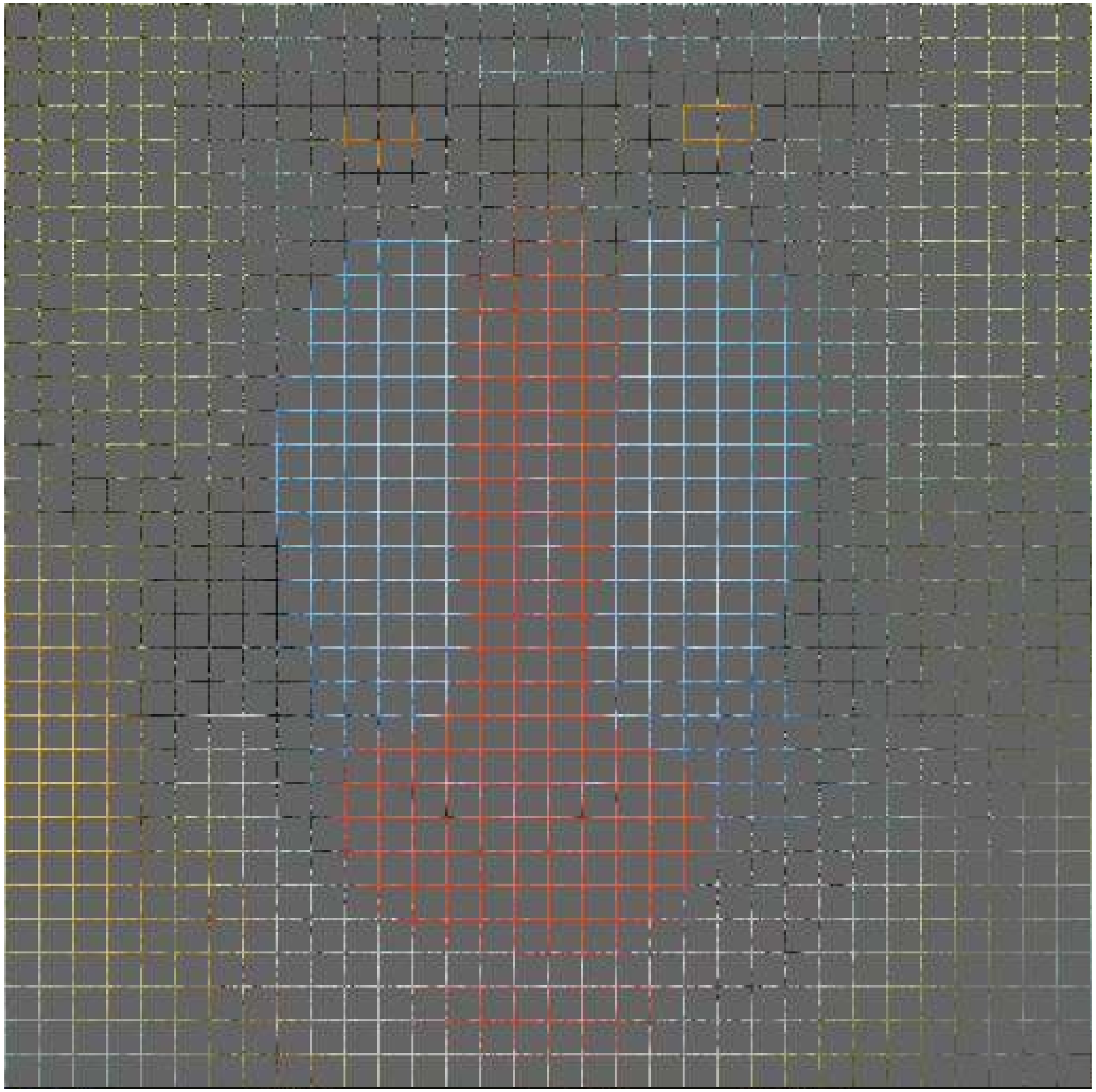


Figure 3: Squares 15x15 removed

Figure 4: After disocclusion (15x15)

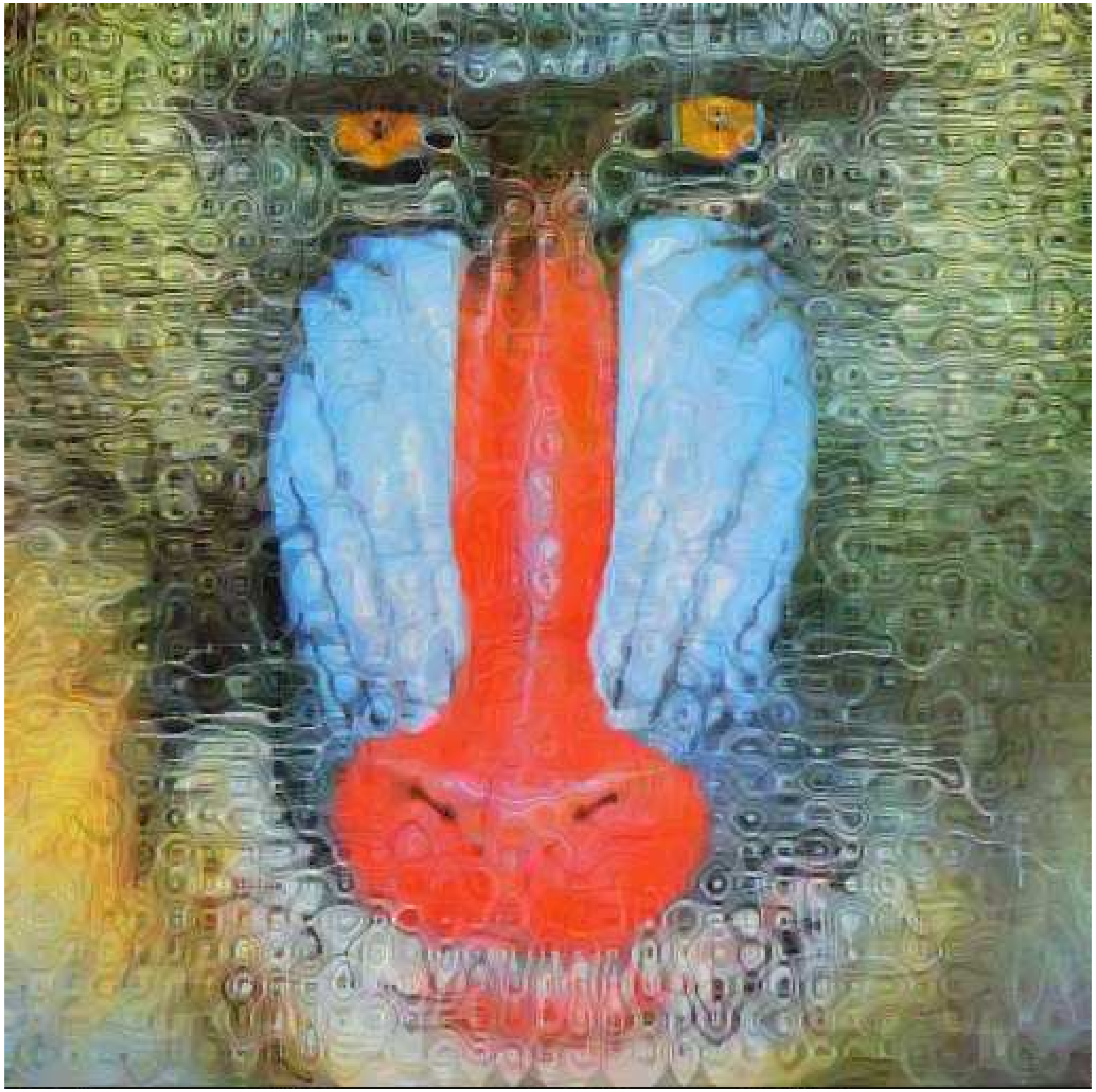


Figure 5: After disocclusion (15x15) using Euler spirals

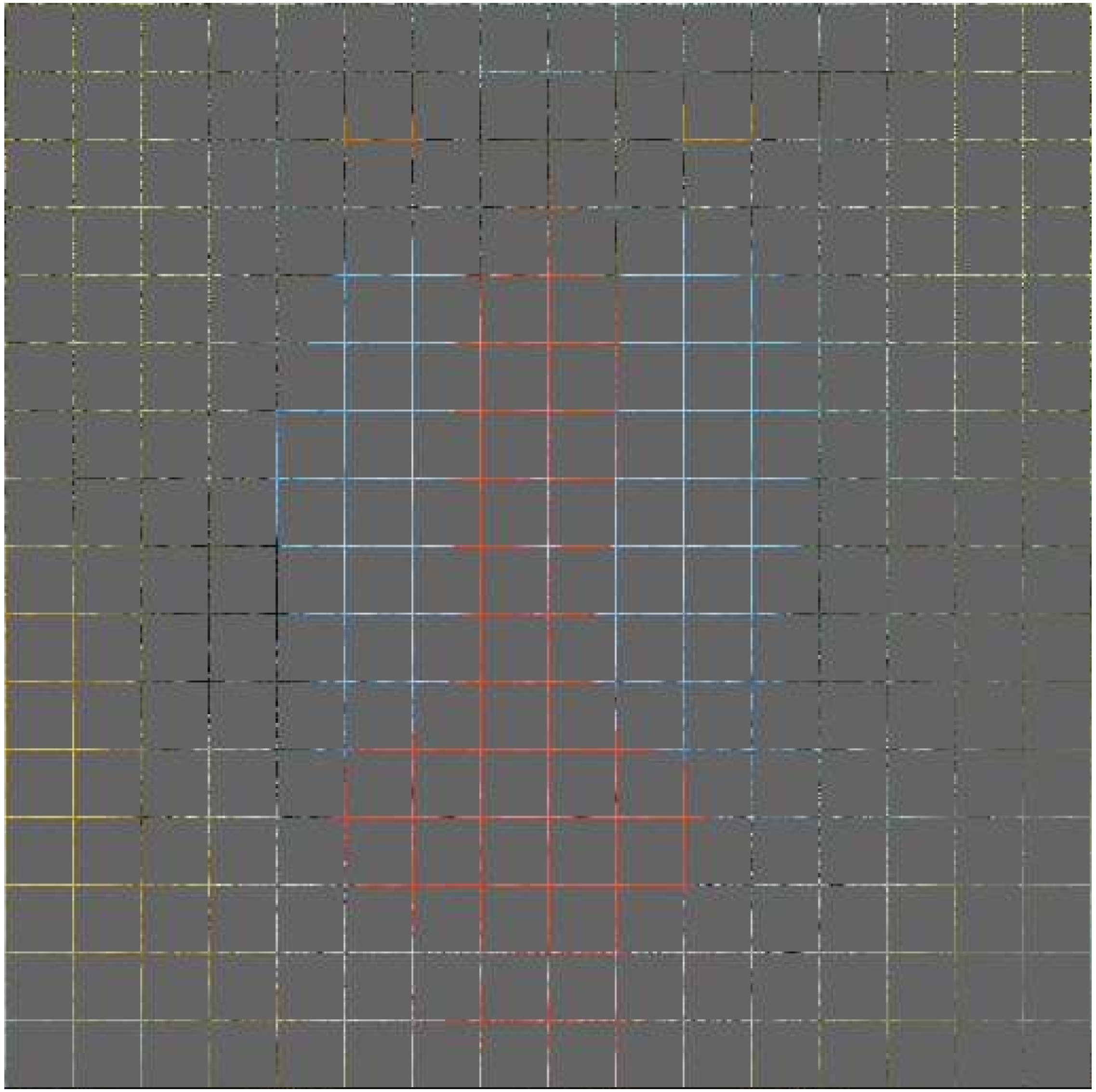


Figure 6: Squares 31x31 removed

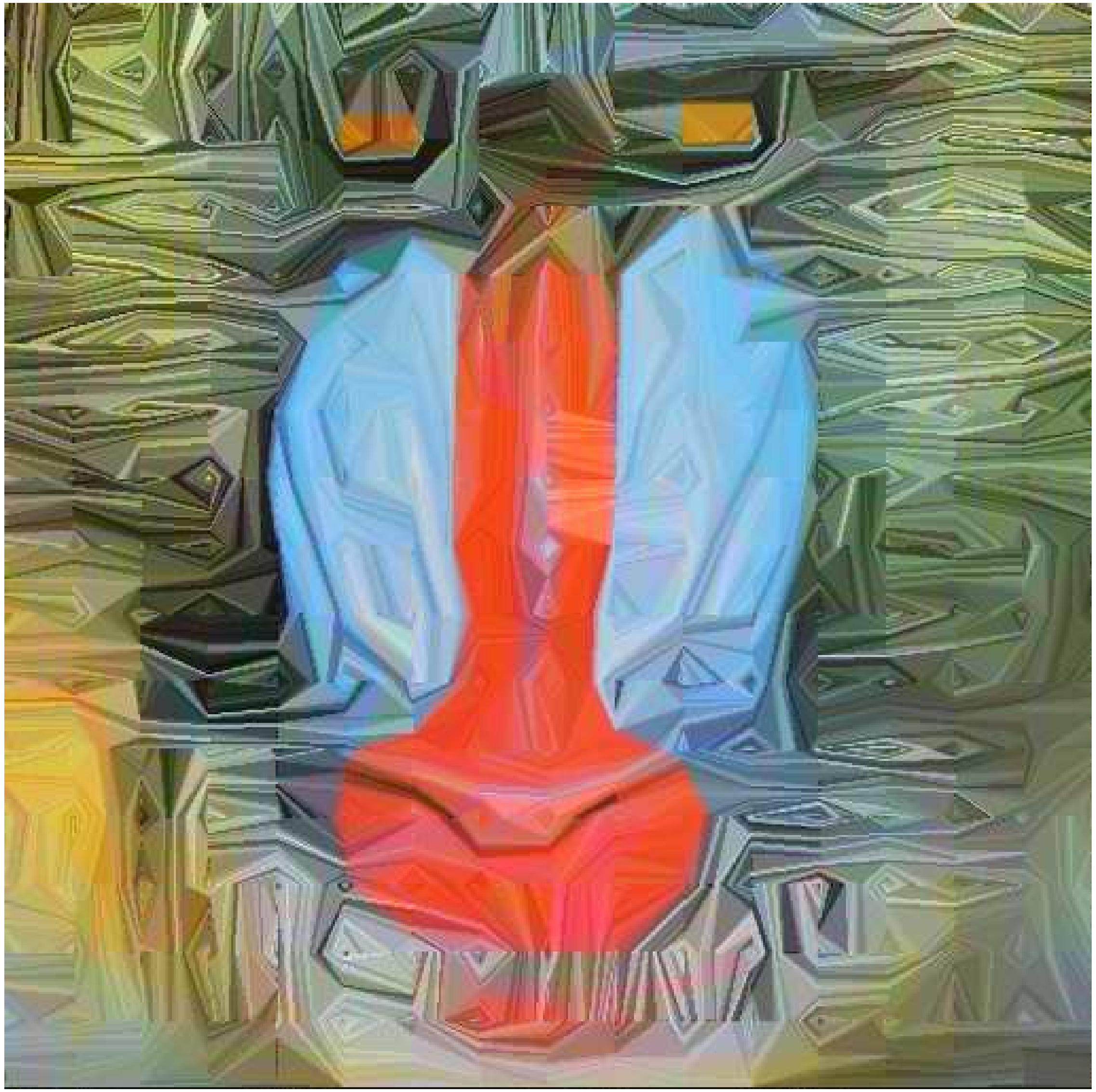


Figure 7: After disocclusion (31x31)

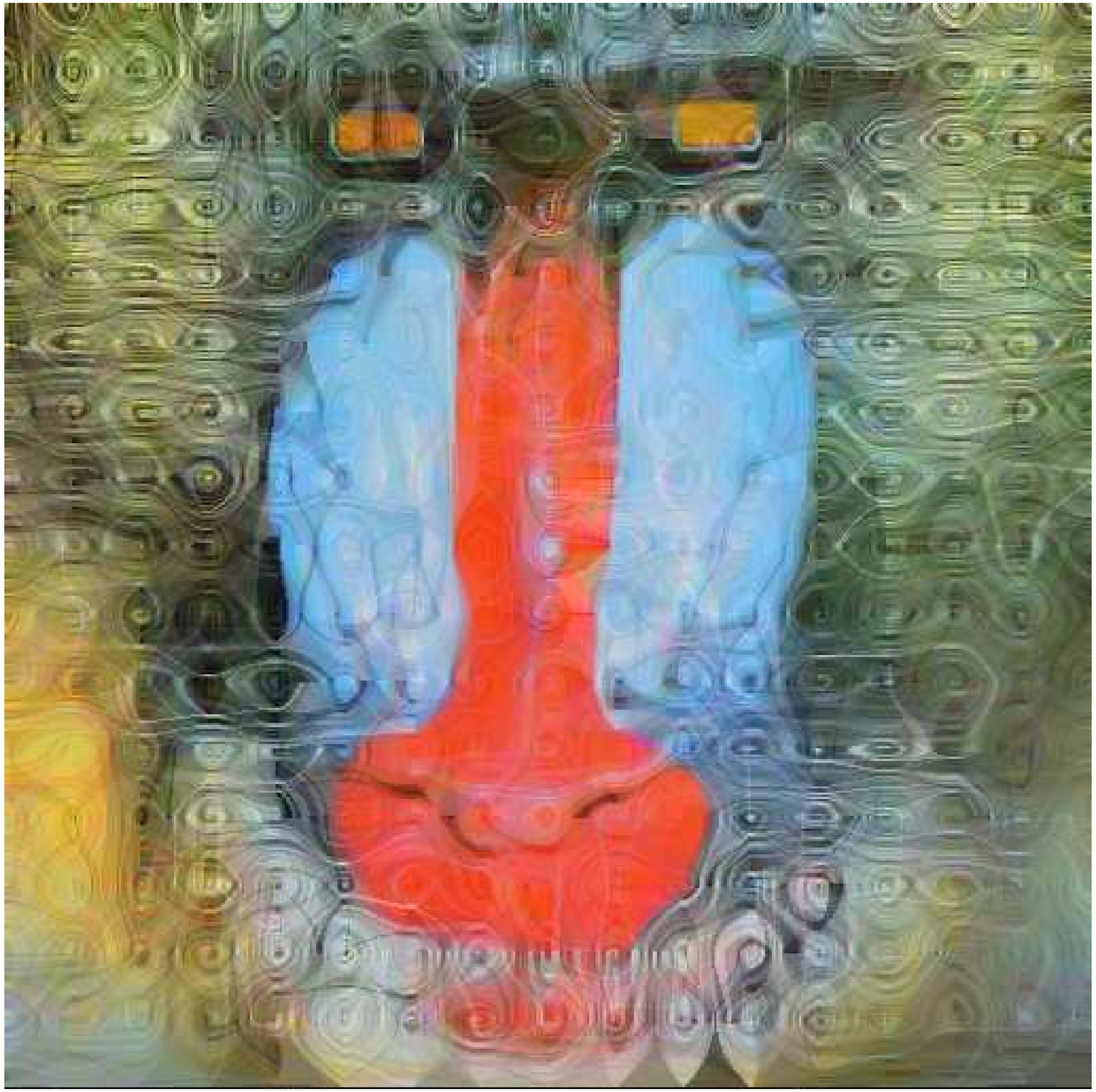


Figure 8: After disocclusion (31x31) using Euler spirals

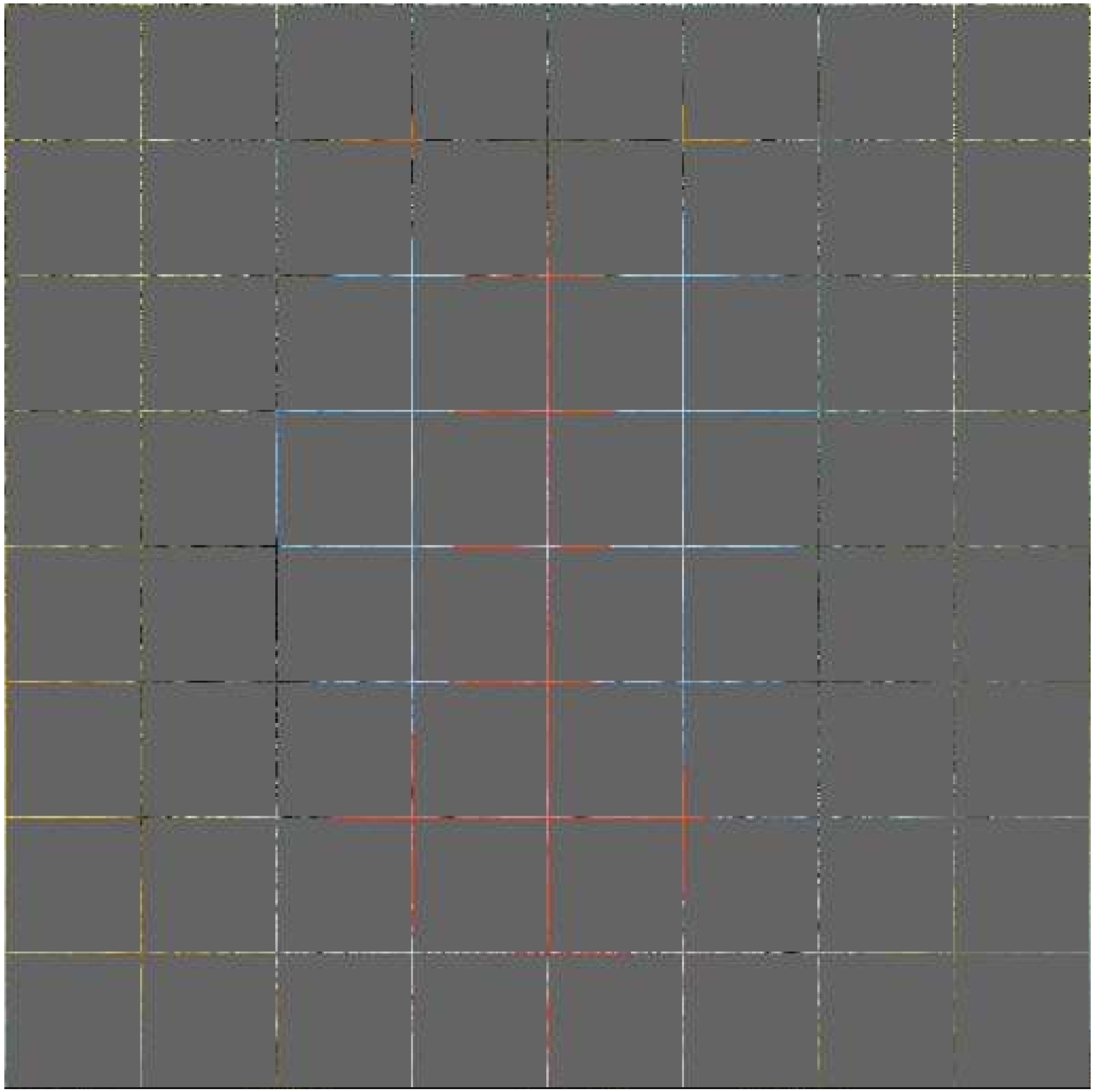


Figure 9: Squares 63x63 removed

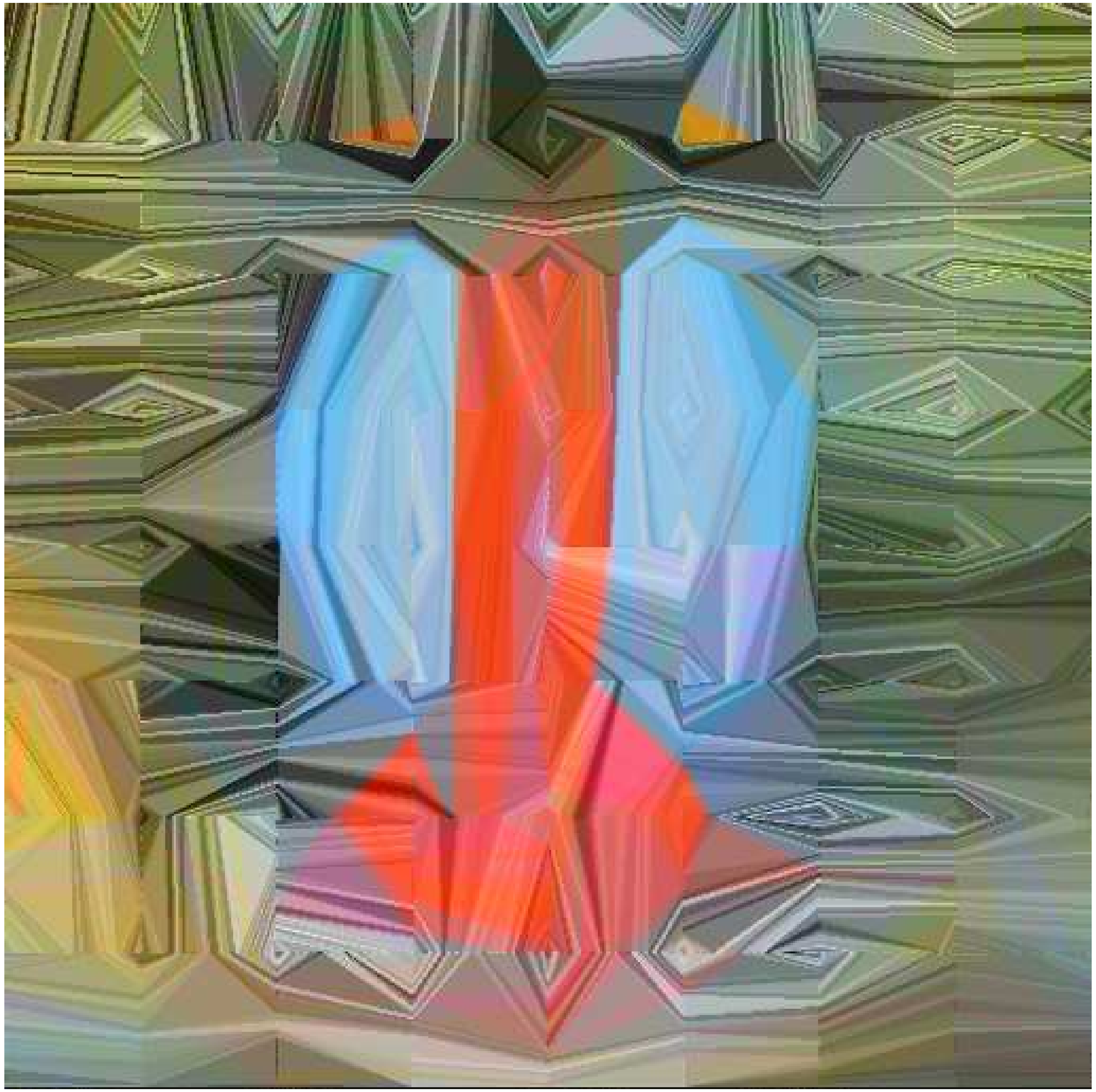


Figure 10: After disocclusion (63x63)

Figure 11: After disocclusion (63x63) using Euler spirals

Incorporation of a geometric constraint

1) compute a sketch S of the original image, i.e. a piecewise constant approximation of the image obtained by keeping only the level lines that are very contrasted, in the sense that their contrast could very unlikely be found in a white noise image (*a contrario* model, cf. A. Desolneux's talk);

2) interpolate S by disocclusion;

3) apply the copy-paste procedure with the new metric

$$d(x_0, y_0) = \alpha\, d_{\mathrm{L}^2}(B_I(x_0, p), B_I(y_0, p)) + (1-\alpha)\, d_{\mathrm{L}^2}(B_S(x_0, p), B_S(y_0, p))$$

where B_I denotes a patch in the original image and B_S a patch in the sketch image.

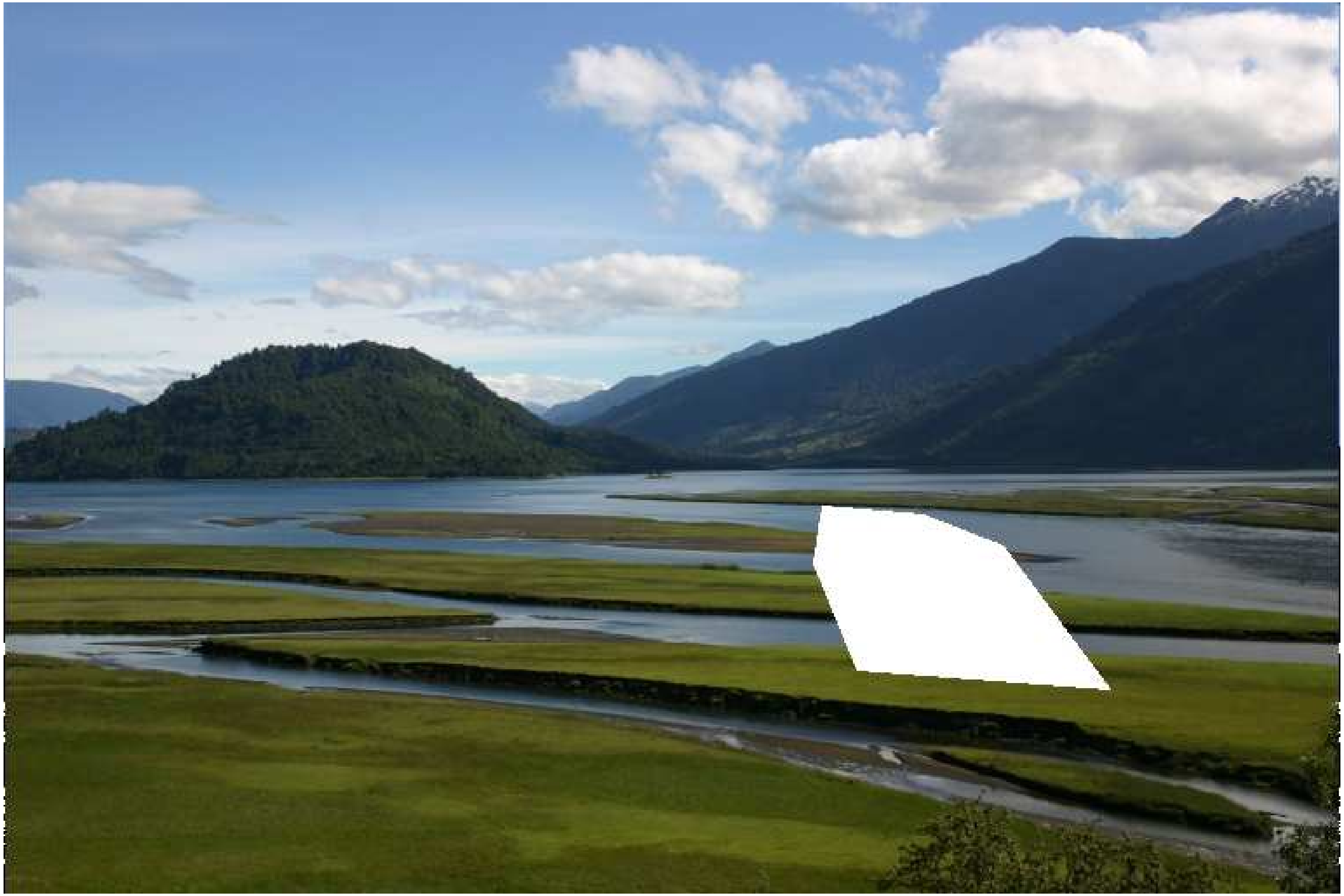


Figure 12: Original image with holes

Figure 13: After classical copy-paste

Figure 14: After copy-paste with constrained geometry

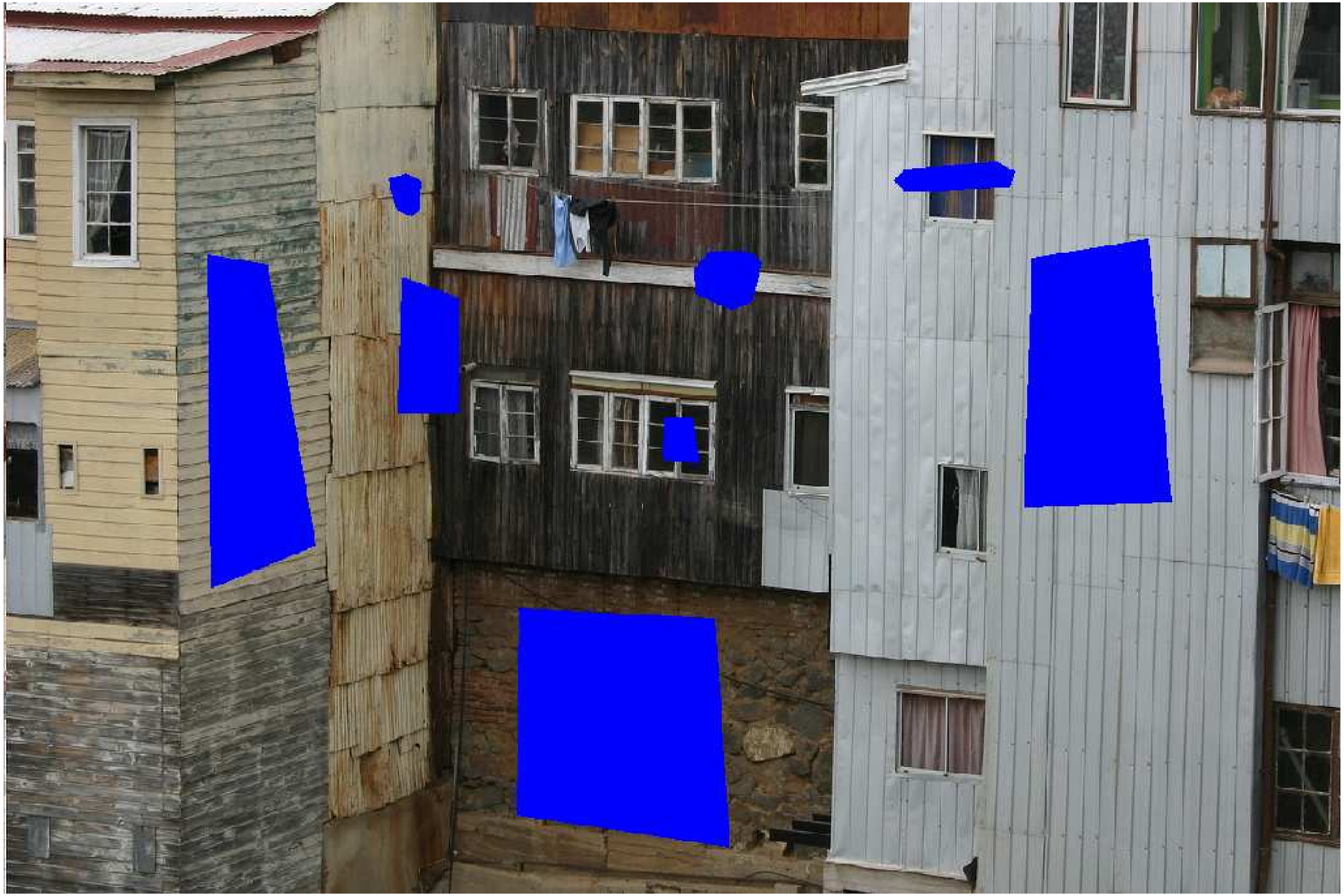


Figure 15: Original image with holes

Figure 16: After application of D. Tschumperlé's anisotropic diffusion PDE

Figure 17: After copy-paste with constrained geometry

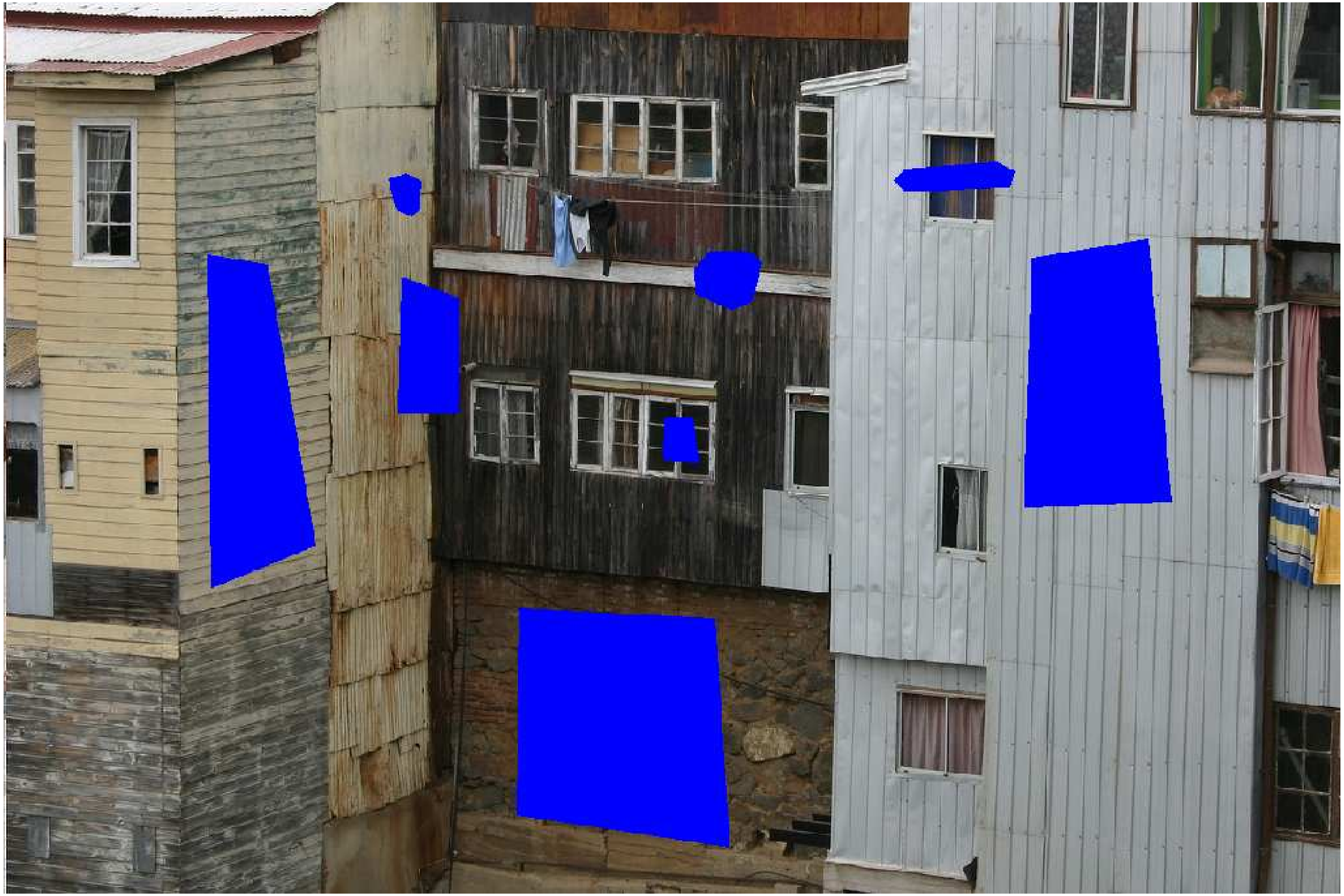


Figure 18: Original image with holes

Conclusions

The combination of the copy-paste procedure and the disocclusion algorithm can sometimes be interesting. Yet many questions remain open:

1. Parameters p, q are related to texture scale. Is there a way to fix them locally ?

2. The parameter α is global. Is there a way to let it depend locally on the relative strength of geometry vs texture ?

3. Allowing rotated patches introduces more flexibility but also more instability. Can this be controlled ?

From a more theoretical viewpoint :

1. What happens when p, q tend to 0 ? Is our process related to a front evolution with nonlocal term (work in progress with J.-F. Aujol and S. Ladjal)?
2. What is the link between this implicit model of image formation and the texture/geometry decomposition of Rudin-Osher and Y. Meyer ?