Stochastic Image Reconstruction from Local Histograms of Gradient Orientation

Agnès Desolneux, Arthur Leclaire
{agnes.desolneux,arthur.leclaire}@cmla.ens-cachan.fr

CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235 Cachan, France.

Abstract. Many image processing algorithms rely on local descriptors extracted around selected points of interest. Motivated by privacy issues, several authors have recently studied the possibility of image reconstruction from these descriptors, and proposed reconstruction methods performing local inference using a database of images. In this paper we tackle the problem of image reconstruction from local histograms of gradient orientation, obtained from simplified SIFT descriptors. We propose two reconstruction models based on Poisson editing and on the combination of multiscale orientation fields. These models are able to recover global shapes and many geometric details of images. They compare well to state of the art results, without requiring the use of any external database.

Keywords: Image Synthesis, Reconstruction from Features, SIFT, Poisson Editing, Maximum Entropy Distributions, Exponential Models

1 Introduction

Extraction of local features constitutes a first step for many image analysis tasks, e.g. image matching and rectification, object detection and tracking, image recognition, image classification, image understanding, see for instance [1] and references therein. Depending on the application, one should use local features reflecting some kind of geometric information while being invariant with respect to several image transformations. For example, in order to match two images of the same distant scene taken under different view points and different illumination conditions, one should compute local features that are invariant to homography and contrast change. It is thus an interesting question to ask how much we can retrieve from the initial image based only on these image descriptors. In other words, what information is really contained in these descriptors.

Since the pioneering work by Attneave [2], many techniques have been proposed to extract points of interest in images and information around them. Here we will only mention a few descriptors and we refer to [3, 4, 1] for a more comprehensive survey. An early descriptor of a patch is given by the Local Binary Descriptors [5]: it extracts the signs of difference between values of Gaussian windows applied at different locations in the patch. Perhaps one of the most famous and powerful feature extraction technique is the Scale-Invariant Feature Transform (SIFT) [6, 3]. As for most image description techniques, this

algorithm first extracts a set of points of interest (called *keypoints*) and then computes for each keypoint a descriptor based on the local behavior of the image around this keypoint. We give a brief description of the SIFT method in Section 2. Another popular local descriptor is given by the histograms of gradient orientations (HOG) [7]. "Dense" HOG descriptors are obtained as quantized histograms of the gradient orientations in each patch of a non-overlapping patch set. At a higher semantic level, local image behavior can be also represented as visual words [8] which are obtained as cluster points in a feature space.

Recently, Weinzaepfel et al. [9] have raised the issue of the reconstruction from image descriptors (in particular SIFT). This work was essentially motivated by privacy issues which appear when image descriptors are clearly transmitted on an unsecured network, e.g. for image recognition or classification. Their image reconstruction method consists in pasting image parts with similar features taken from a database. This has triggered several works based on other descriptors. Vondrick et al. [10] address reconstruction from dense HOG by relying on dictionary representation of HOG and patches. Also, d'Angelo et al. [11] address reconstruction from local binary descriptors by relying on primal-dual optimization techniques. Kato and Harada address reconstruction from bag of visual words [12]. Reconstruction problems have also been addressed with methods based on neural networks [13–15]; these methods are generic and can be applied to any image representation that can be approximated by the output of a convolutional neural network (in particular HOG as suggested in [14]). Finally, generative *a contrario* models [16] have been proposed to recover an image from a set of detected features (segments for example).

In all these works, the objective is essentially to reconstruct *one* image from the descriptors. But in many cases, this description transform has no reason to be invertible: for example with the SIFT descriptors, since it does not retain information outside the SIFT cells, many images could have the same SIFT output. Thus we propose here, rather than unique reconstruction, to sample from a stochastic image model whose statistics comply with the information contained in the local descriptors. We study in particular the case of simplified SIFT transforms which extract multiscale HOG from regions around the SIFT keypoints. We will essentially propose two models that can be adapted to several transforms based on local HOG. They both rely on Poisson editing [17] which is a way to compute an image whose gradient is as close as possible to a given vector field. Also, our second model is an instance of exponential models [18, 19] which provide maximum entropy stochastic models having prescribed average values for a given set of numerical features.

Our paper is organized as follows. Section 2 contains a description of the features on which our image reconstruction relies. In Section 3 we introduce a model (called MS-Poisson) that combines gradient orientations at different scales by solving a multiscale Poisson problem. In Section 4 we propose a maximum entropy model (called MaxEnt) on orientation fields which respects the statistical distribution of gradient orientation in the SIFT regions. Finally, in Section 5, we give and comment several reconstruction results, and compare with [9, 13].

2 Overview and notations

In all the paper we consider a gray level image $u_0 : \Omega \to \mathbb{R}$ defined on a discrete rectangle of $\mathbb{Z}^2$. The Gaussian scale-space of u_0 is denoted by $g_\sigma * u_0$ where g_σ is the Gaussian kernel of standard deviation σ. We will denote by $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ the set of angles (possible orientation values).

In this paper, we propose to reconstruct an image from a simplified SIFT transform detailed below. Let us begin with a brief explanation of the main steps of the SIFT methods; for technical details we refer the reader to [20].

1. Computing SIFT keypoints:
   (a) Extract local extrema of a discrete version of $(\boldsymbol{x}, \sigma) \mapsto \sigma^2 \Delta g_\sigma * u_0(\boldsymbol{x})$.
   (b) Refine the positions of the local extrema (sub-pixel precision).
   (c) Discard extrema with low contrast and extrema located on edges.
2. Computing SIFT local descriptors associated to the keypoint $(\boldsymbol{x}, \sigma)$:
   (a) Compute one or several principal orientations θ.
   (b) For each detected orientation θ, consider a grid of 4×4 square regions around $(\boldsymbol{x}, \sigma)$. These square regions, which we call SIFT subcells are of size $3\sigma \times 3\sigma$ with one side parallel to θ. In each subcell compute the histogram of $\mathrm{Angle}(\nabla g_\sigma * u_0) - \theta$ quantized on 8 values $(k\frac{\pi}{4}, 1 \le k \le 8)$.
   (c) Normalization: the 16 histograms are concatenated to obtain a feature vector $f \in \mathbb{R}^{128}$, which is then normalized and quantized to 8-bit integers.

When computing orientation histograms in steps 2(a), 2(b), each pixel votes with a weight that depends on the value of the gradient norm at scale σ and of its distance to the keypoint center $\boldsymbol{x}$. Also in step 2(b), there is a linear splitting of the vote of an angle between the two adjacent quantized angle values.

The reconstruction method described in this paper relies on the SIFT keypoints $(\boldsymbol{x}, \sigma)$ and on 8-quantized histograms orientations of $\mathrm{Angle}(\nabla g_\sigma * u_0) - \theta$ around $\boldsymbol{x}$. In other words, we do not include the vote weights nor the normalization step 2(c).

We denote by $(s_j)_{j \in \mathcal{J}}$ the collection of SIFT subcells, $s_j \subset \Omega$. In a SIFT cell, there are 16 SIFT subcells so that different s_j can correspond to the same keypoint. We will denote by $(\boldsymbol{x}_j, \sigma_j)$ the keypoint associated to s_j, and θ_j the principal orientation. For $x \in \Omega$, we will denote by $\mathcal{J}(\boldsymbol{x}) = \{j \in \mathcal{J} \mid \boldsymbol{x} \in s_j\}$ the set of indices of SIFT subcells containing $\boldsymbol{x}$. In s_j we extract the quantized HOG H_j at scale σ_j which is identified to the piecewise constant density function

$$h_j = \sum_{\ell=1}^{8} H_j^\ell \mathbf{1}_{[\theta_j + (\ell-1)\frac{\pi}{4}, \theta_j + \ell\frac{\pi}{4}[}, \tag{1}$$

$$\text{where} \quad H_j^\ell = \frac{1}{|s_j|} \Big| \big\{ \boldsymbol{x} \in s_j \ ; \ \mathrm{Angle}(\nabla g_{\sigma_j} * u_0)(\boldsymbol{x}) - \theta_j \in [(\ell-1)\tfrac{\pi}{4}, \ell\tfrac{\pi}{4}] \big\} \Big|.$$

The main purpose of this paper is to reconstruct an image whose content agrees with the multiscale HOG h_j in the SIFT subcells s_j. One main difficulty is that the gradient magnitude is *a priori* completely lost during the extraction of these features. Another issue is the fact that a point $\boldsymbol{x}$ can belong to

several subcells s_j. In that case we need to find a way to combine the different information given by the histograms h_j to recover the image orientation at $\boldsymbol{x}$.

3 Poisson Reconstruction of an Image

For the first proposed reconstruction method, we define vector fields $V_j : \Omega \to \mathbb{R}^2$ that will serve as objective gradient at scale σ_j in the SIFT subcell s_j. We propose to set $V_j = w_j e^{i\gamma_j}$ where $w_j = \frac{1}{\sigma_j}\mathbf{1}_{s_j}$ (whose choice is motivated by the homogeneity argument $\nabla\big(u(\frac{\boldsymbol{x}}{\sigma})\big) = \frac{1}{\sigma}\nabla u(\frac{\boldsymbol{x}}{\sigma})$) and where γ_j is a random orientation field sampled from the probability density function h_j. In order to obtain a reconstructed image U, these objective vector fields V_j for all the scales σ_j are combined by solving a multiscale Poisson problem explained in Section 3.2.

3.1 Classical Poisson Reconstruction

The aim of Poisson reconstruction is to look for an image $u : \Omega \to \mathbb{R}$ whose gradient is as close as possible to an objective vector field $V = (v_1, v_2)^T : \Omega \to \mathbb{R}^2$. In the case of image editing, this technique has been proposed by Pérez et al. [17] in order to copy pieces of an image into another one in a seamless way. More precisely, the goal is to minimize the functional

$$F(u) = \sum_{\boldsymbol{x}\in\Omega} \|\nabla u(\boldsymbol{x}) - V(\boldsymbol{x})\|_2^2. \tag{2}$$

Since $F(c+u) = F(u)$ for any constant c, we can impose $\sum_{\boldsymbol{x}\in\Omega} u(\boldsymbol{x}) = 0$. If we use periodic boundary conditions for the gradient, we can solve this problem with the Discrete Fourier Transforms. Indeed, if we use the simple derivation scheme based on periodic convolutions

$$\nabla u(\boldsymbol{x}) = \begin{pmatrix} \partial_1 * u(\boldsymbol{x}) \\ \partial_2 * u(\boldsymbol{x}) \end{pmatrix} \qquad \text{where} \quad \begin{cases} \partial_1 = \delta_{(0,0)} - \delta_{(1,0)} \\ \partial_2 = \delta_{(0,0)} - \delta_{(0,1)} \end{cases}, \tag{3}$$

this problem can be expressed in the Fourier domain with Parseval formula

$$F(u) = \frac{1}{|\Omega|}\sum_{\boldsymbol{\xi}\neq 0} |\widehat{\partial}_1(\boldsymbol{\xi})\widehat{u}(\boldsymbol{\xi}) - \widehat{v}_1(\boldsymbol{\xi})|_2^2 + |\widehat{\partial}_2(\boldsymbol{\xi})\widehat{u}(\boldsymbol{\xi}) - \widehat{v}_2(\boldsymbol{\xi})|_2^2. \tag{4}$$

Thus, for each $\boldsymbol{\xi}$ we have a barycenter problem which is simply solved by

$$\forall \boldsymbol{\xi} \neq 0, \quad \widehat{u}(\boldsymbol{\xi}) = \frac{\overline{\widehat{\partial}_1}(\boldsymbol{\xi})\widehat{v}_1(\boldsymbol{\xi}) + \overline{\widehat{\partial}_2}(\boldsymbol{\xi})\widehat{v}_2(\boldsymbol{\xi})}{|\widehat{\partial}_1(\boldsymbol{\xi})|^2 + |\widehat{\partial}_2(\boldsymbol{\xi})|^2}. \tag{5}$$

3.2 Multiscale Poisson Reconstruction

In order to simultaneously constrain the gradient at several scales $(\sigma_j)_{j\in\mathcal{J}}$, we propose here to consider the following multiscale Poisson energy

$$G(u) = \sum_{j\in\mathcal{J}} w(\sigma_j) \sum_{\boldsymbol{x}\in\Omega} \|\nabla(g_{\sigma_j} * u)(\boldsymbol{x}) - V_j(\boldsymbol{x})\|_2^2, \tag{6}$$

where g_σ is the Gaussian kernel of standard deviation σ, $V_j = (v_{j,1}, v_{j,2})^T$ is the objective gradient at scale σ_j, and $w(\sigma_j)$ is a set of weights. In our application, since there are more keypoints in the fine scales (i.e. when σ_j is small), and since the keypoints at fine scales are generally more informative, a reasonable choice is to take all weights $w(\sigma_j) = 1$. But we keep these weights in the formula for the sake of generality.

Again, with periodic boundary conditions, this problem can be expressed in Fourier domain as

$$G(u) = \frac{1}{|\Omega|} \sum_{j \in \mathcal{J}} \sum_{\boldsymbol{\xi} \neq 0} w(\sigma_j) \Big(|\widehat{g}_{\sigma_j}(\boldsymbol{\xi}) \widehat{\partial}_1(\boldsymbol{\xi}) \widehat{u}(\boldsymbol{\xi}) - \widehat{v}_{j,1}(\boldsymbol{\xi})|_2^2 + |\widehat{g}_{\sigma_j}(\boldsymbol{\xi}) \widehat{\partial}_2(\boldsymbol{\xi}) \widehat{u}(\boldsymbol{\xi}) - \widehat{v}_{j,2}(\boldsymbol{\xi})|_2^2 \Big). \tag{7}$$

The solution is still a barycenter. It is given by (recall that $\widehat{g}_{\sigma_j}(\boldsymbol{\xi}) \in \mathbb{R}$ since g_{σ_j} is even):

$$\forall \boldsymbol{\xi} \neq 0, \quad \widehat{u}(\boldsymbol{\xi}) = \frac{\displaystyle\sum_{j \in \mathcal{J}} w(\sigma_j) \widehat{g}_{\sigma_j}(\boldsymbol{\xi}) \Big(\overline{\widehat{\partial}_1(\boldsymbol{\xi})} \widehat{v}_{j,1}(\boldsymbol{\xi}) + \overline{\widehat{\partial}_2(\boldsymbol{\xi})} \widehat{v}_{j,2}(\boldsymbol{\xi}) \Big)}{\displaystyle\sum_{j \in \mathcal{J}} w(\sigma_j) |\widehat{g}_{\sigma_j}(\boldsymbol{\xi})|^2 \Big(|\widehat{\partial}_1(\boldsymbol{\xi})|^2 + |\widehat{\partial}_2(\boldsymbol{\xi})|^2 \Big)}. \tag{8}$$

Notice that, depending on the finest scale, the denominator can vanish in the high frequencies because of the term $\widehat{g}_{\sigma_j}(\boldsymbol{\xi})$ (as it is the case in a deconvolution problem). Therefore, it may be useful to add a regularization term controlled by a parameter $\mu > 0$. Then, if we minimize

$$G(u) + \mu \|\nabla u\|_2^2, \tag{9}$$

we get the well-defined solution

$$\widehat{u}(\boldsymbol{\xi}) = \frac{\displaystyle\sum_{j \in \mathcal{J}} w(\sigma_j) \widehat{g}_{\sigma_j}(\boldsymbol{\xi}) \Big(\overline{\widehat{\partial}_1(\boldsymbol{\xi})} \widehat{v}_{j,1}(\boldsymbol{\xi}) + \overline{\widehat{\partial}_2(\boldsymbol{\xi})} \widehat{v}_{j,2}(\boldsymbol{\xi}) \Big)}{\left(\mu + \displaystyle\sum_{j \in \mathcal{J}} w(\sigma_j) |\widehat{g}_{\sigma_j}(\boldsymbol{\xi})|^2 \right) \Big(|\widehat{\partial}_1(\boldsymbol{\xi})|^2 + |\widehat{\partial}_2(\boldsymbol{\xi})|^2 \Big)}. \tag{10}$$

4 Stochastic Models for Gradient Orientations

The second reconstruction method will consist in solving a classical Poisson problem (Section 3.1) with one single objective gradient $V = we^{i\gamma}$ where $\gamma : \Omega \to \mathbb{T}$ is now a sample of an orientation field which is inherently designed to combine the local HOG at the scale $\sigma = 0$.

In this section, we aim at defining stochastic models of orientation fields that fit the distributions of oriented gradients in the SIFT subcells $(s_j)_{j \in \mathcal{J}}$. In contrast with the usual SIFT method, for simplicity we will consider the orientations in all subcells with the same quantization bins

$$B_\ell = \big[(\ell - 1)\tfrac{\pi}{4}, \ell \tfrac{\pi}{4} \big[, \quad (1 \leqslant \ell \leqslant 8). \tag{11}$$

We then consider for all $j \in \mathcal{J}$ and $1 \leqslant \ell \leqslant 8$, the real-valued feature function given by

$$\forall \theta \in \mathbb{T}^\Omega, \quad f_{j,\ell}(\theta) = \frac{1}{|s_j|} \sum_{\boldsymbol{x} \in s_j} \mathbf{1}_{B_\ell}(\theta(\boldsymbol{x})) \tag{12}$$

We are then interested in probability distributions P on $\mathbb{T}^\Omega$ such that

$$\forall j \in \mathcal{J}, \forall \ell \in \{1, \dots, 8\}, \quad \mathbb{E}_P\big(f_{j,\ell}(\theta)\big) = f_{j,\ell}(\theta_0) \ , \tag{13}$$

where $\theta_0 = \mathrm{Angle}(\nabla u_0)$ is the orientation field of the original image u_0. There are many probability distributions P on $\mathbb{T}^\Omega$ that satisfy (13), and we will be mainly interested in the ones that are at the same time as "random" as possible, in the sense that they are of maximal entropy. Let us notice that the gradient orientation extracted from the MS-Poisson model (10) is not ensured to satisfy the constraints (13).

Theorem 1 ([19]). *There exists a family of numbers $\lambda = (\lambda_{j,\ell})_{j \in \mathcal{J}, 1 \leqslant \ell \leqslant 8}$ such that the probability distribution*

$$dP_\lambda = \frac{1}{Z_\lambda} \exp\left(- \sum_{j,\ell} \lambda_{j,\ell} f_{j,\ell}(\theta) \right) d\theta, \tag{14}$$

where the partition function Z_λ is given by $Z_\lambda = \int_{\mathbb{T}^\Omega} \exp\big(- \sum_{j,\ell} \lambda_{j,\ell} f_{j,\ell}(\theta)\big) d\theta$, satisfies the constraints (13) *and is of maximal entropy among all absolutely continuous probability distributions w.r.t. $d\theta$ satisfying the constraints* (13)*.*

Proof. This result directly follows from the general theorem given in [19]. The only difficulty is to handle the technical hypothesis of linear independence of the $f_{j,\ell}$. In our framework, the $f_{j,\ell}$ are not independent (in particular because $\sum_{\ell=1}^8 f_{j,\ell} = 1$, and also because there may be other dependencies for instance when one subcell is exactly the union of two smaller subcells). But one can still apply the theorem to an extracted linearly independent subfamily. This gives existence of the solution for the initial family $(f_{j,\ell})$ (but of course not the unicity).

One can show (see [19]) that the solutions P_λ are obtained by minimizing the smooth convex function

$$\Phi(\lambda) = \log Z_\lambda + \sum_{j,\ell} \lambda_{j,\ell} f_{j,\ell}(\theta_0). \tag{15}$$

Let us examine P_λ from closer. For that we can write

$$- \log \frac{dP_\lambda}{d\theta} - \log Z_\lambda = \sum_{j \in \mathcal{J}, 1 \leqslant \ell \leqslant 8} \lambda_{j,\ell} f_{j,\ell}(\theta) = \sum_{\boldsymbol{x} \in \Omega} \varphi_{\lambda,\boldsymbol{x}}(\theta(\boldsymbol{x})), \tag{16}$$

$$\text{where} \quad \varphi_{\lambda,\boldsymbol{x}} = \sum_{\ell=1}^{8} \left(\sum_{j \in \mathcal{J}(\boldsymbol{x})} \frac{\lambda_{j,\ell}}{|s_j|} \right) \mathbf{1}_{B_\ell}. \tag{17}$$

This proves that under P_λ the values $\theta(\boldsymbol{x})$ are independent and have a probability distribution $\frac{1}{Z_{\lambda,\boldsymbol{x}}} e^{-\varphi_{\lambda,\boldsymbol{x}}}$ where $Z_{\lambda,\boldsymbol{x}} = \sum_{\ell=1}^{8} \exp\Big(- \sum_{j \in \mathcal{J}(\boldsymbol{x})} \frac{\lambda_{j,\ell}}{|s_j|} \Big) |B_\ell|$. Therefore, the constraints (13) can be written in terms of λ as

$$\forall j, \ell, \quad \sum_{\boldsymbol{x} \in s_j} \frac{1}{Z_{\lambda,\boldsymbol{x}}} \exp\left(- \sum_{k \in \mathcal{J}(\boldsymbol{x})} \frac{\lambda_{k,\ell}}{|s_k|} \right) = |\{ \boldsymbol{x} \in s_j \ ; \ \theta_0(\boldsymbol{x}) \in B_\ell \ \}| \ . \tag{18}$$

Notice that this system is highly non-linear and is in general difficult to solve. However, if the SIFT subcells do not overlap (which is very rare), then the distribution on $\theta(\boldsymbol{x})$ simplifies to: the uniform distribution if $\boldsymbol{x}$ doesn't belong to any s_j, and the distribution given by the empirical quantized histogram of θ_0 on s_j if $\boldsymbol{x} \in s_j$.

Now, if the SIFT subcells intersect, there is no explicit solution anymore. To cope with that, as in [18] we use a numerical scheme to find the maximum entropy distribution P_λ based on the minimization of (15). For that, we recall the gradient of $\log Z_\lambda$, obtained by differentiating the partition function:

$$\frac{\partial \log Z_\lambda}{\partial \lambda_{j,\ell}} = \frac{1}{Z_\lambda} \frac{\partial Z_\lambda}{\partial \lambda_{j,\ell}} = -\mathbb{E}_{P_\lambda}\big[f_{j,\ell}(\theta) \big]. \tag{19}$$

This expression allows us to minimize Φ with a gradient descent with constant (sufficiently small) step size, which converges [21] since Φ is a smooth convex function with Lipschitz gradient (as can be seen on the Hessian matrix).

5 Results and discussion

In this section, for several images, we show reconstructions obtained as samples of the models proposed in the two previous sections. We also compare with other reconstruction algorithms described in [9, 10] (the test images are taken from these articles).

The first model (denoted by MS-Poisson) consists in computing the solution of regularized multiscale Poisson reconstruction (10) with objective vector fields $V_j = \frac{1}{\sigma_j} e^{i\gamma_j} \mathbf{1}_{s_j}$ where γ_j is composed of independent samples of the probability distribution h_j given by (1), that is the available HOG at scale σ_j in the subcell s_j. In this case, the orientation fields (γ_j) are independent. The weights are set to $w(\sigma_j) = 1$ for all j, and the regularization parameter is set to $\mu = 50$.

The second model (denoted by MaxEnt) is obtained as the solution of a classical Poisson problem associated to the objective vector field

$$V(\boldsymbol{x}) = \Big(\max_{j \in \mathcal{J}(\boldsymbol{x})} \frac{1}{\sigma_j} \Big) e^{i\gamma(\boldsymbol{x})} \mathbf{1}_{\mathcal{J}(\boldsymbol{x}) \neq \emptyset}. \tag{20}$$

where γ is a sample of the maximum entropy model of Section 4. Here again, we chose a magnitude function $|V(\boldsymbol{x})|$ that favors the locations where there is information at fine scale (note that by definition of the keypoints, $\sigma_j > 0$).

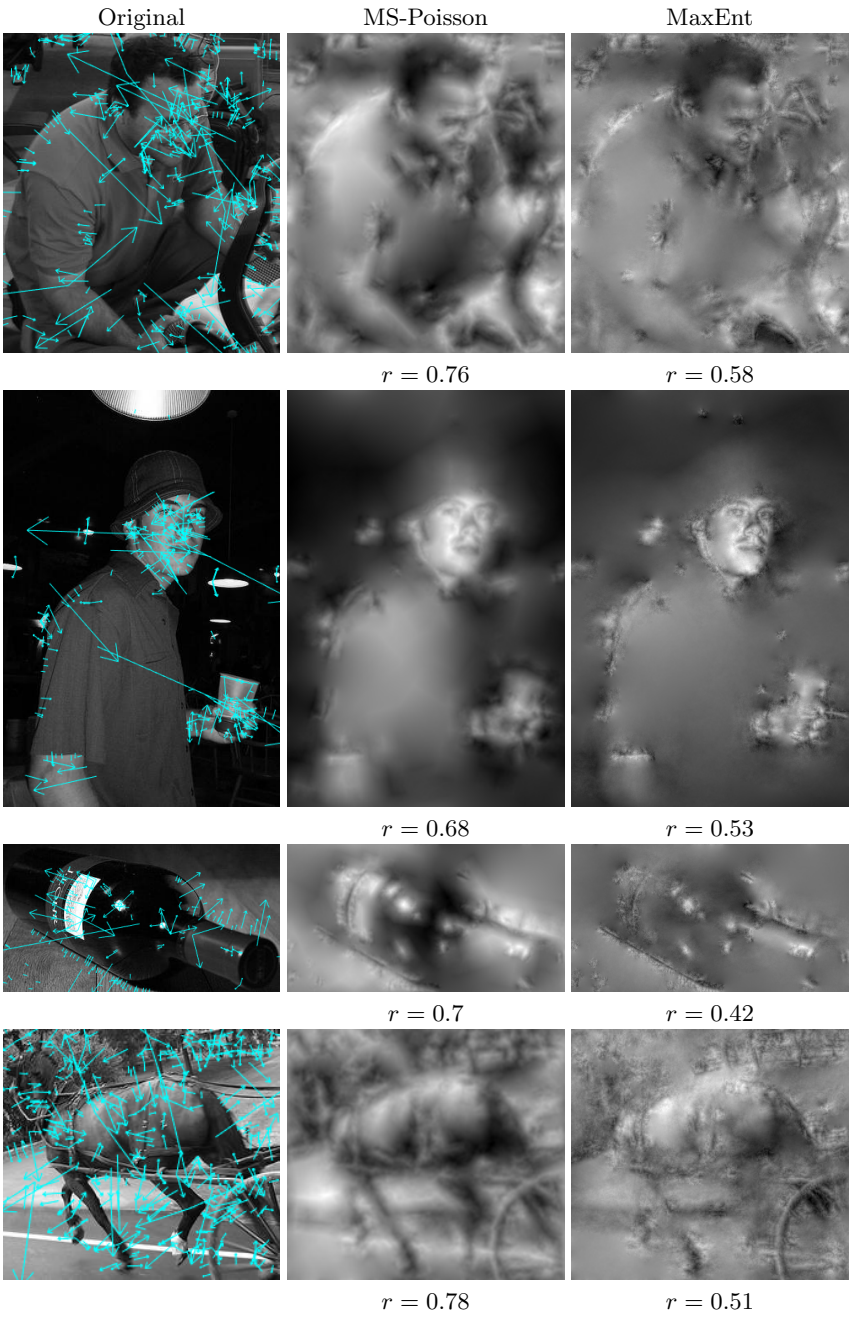

Fig. 1. Reconstruction results. For each row, from left to right, we display the original image with over-imposed arrows representing the keypoints (the length of the arrow is 6σ which is the half-side of the SIFT cell, and twice the size of the corresponding SIFT subcells), the reconstruction with MS-Poisson, and the reconstruction with MaxEnt. (Image Credits [10]).

In Fig. 1 we give reconstruction results obtained with the models MS-Poisson and MaxEnt. These results show some similarity: both these algorithms are able to recover content in the SIFT subcells. They recover geometric details from the fine scale subcells while global shapes are constrained thanks to the large scale information; many objects contained in the image are easily recognizable. These reconstruction results may seem surprisingly good at first but can be explained by the fact that the local HOG in the SIFT subcells is not as sparse as one could imagine. On the left of Fig. 1 one can see that there are many keypoints in these images and in particular keypoints at fine scales which give a precise local information.

Notice that the result of MaxEnt appears a bit sharper, as expected: on the one hand, MaxEnt relies on local HOG computed at the scale of the image ($\sigma = 0$) while MS-Poisson relies on multiscale HOG, and on the other hand, MS-Poisson realizes a compromise between many objective vector fields at different scales, which introduces some blur. Of course, these algorithms cannot recover precisely the content outside the SIFT cells: this completion is essentially obtained by regularization. For example, the top of the bottle in the first row of Fig. 1 is not properly restored.

These two reconstruction methods draw a sample of a random model. But for both of them, since Poisson reconstruction is a linear operation, it is also possible to compute a reconstruction which is the expectation of these random models. It turns out that these mean images are very similar (though a bit more regular) to the random samples. In future work we will investigate the variance of these models to better explain this.

Assessing the quality of reconstruction is a very ill-posed task, in particular because 1) it is a very subjective question, 2) such a measure should be invariant to affine contrast changes, and 3) there is no information outside the SIFT subcells. In the lack of anything better, we give the values r of the normalized cross-correlation between the reconstruction and the original image (which is contrast-affine-invariant). In Fig. 1 we observe that MS-Poisson leads to much better cross-correlation values which reflects the fact that large scale contrasted regions are better retrieved (while not accounted for with MaxEnt). For that reason, in the following experiments we only show results obtained with MS-Poisson.

On Fig. 2, we show comparisons between the results of the Poisson reconstruction from dense HOG (computed on 5×5 non-overlapping patches), and the results of Hoggles [10]. The results of [10] look cleaner because they exploit redundancies between overlapping HOG templates (via a learned pair-dictionary representation of HOG templates and patches) whereas Poisson reconstruction only performs independent sampling of pixel orientations.

The two proposed reconstruction methods can also be applied by using the SIFT feature vectors as substitute to the input local HOG. For that, for each subcell s_j we extract the part of the descriptor that is relative to this subcell, and we normalize its sum to 1 (in order to get a probability distribution function). Such a result of SIFT reconstruction can be seen in Fig. 3. The result obtained

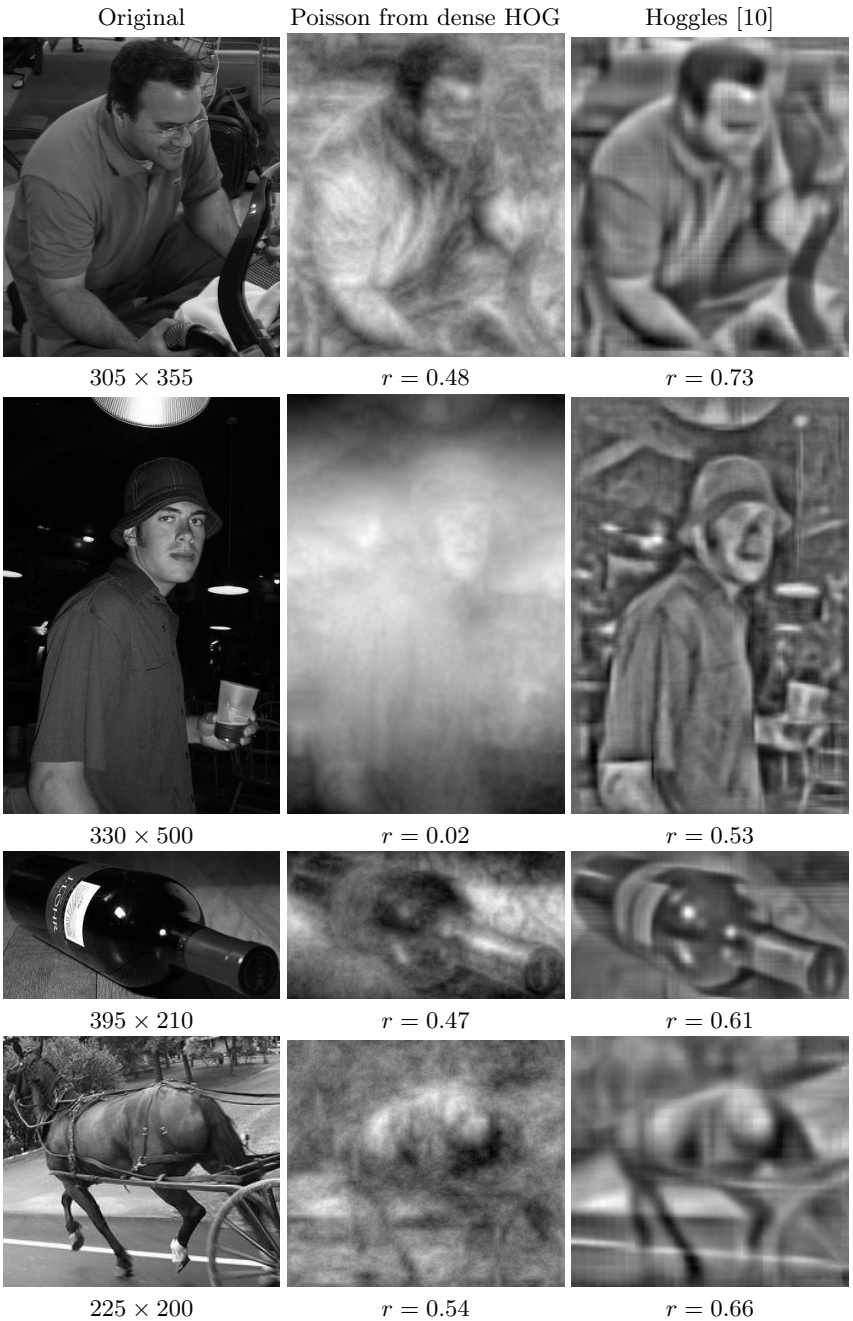

Fig. 2. Reconstruction from dense HOG. For each row, from left to right we display the original image, the result of Poisson reconstruction from dense HOG (computed on 5×5 patches with no overlap), and the result of [10]. The levels of recovered details with the two methods are comparable. However, the result of Poisson reconstruction looks less clean because of the independent sampling of all pixels. (Image Credits [10]).

with the SIFT descriptors retrieves the global shapes of the image but is in general less precise than the one obtained with the local HOG, because of the several normalization steps applied in SIFT.

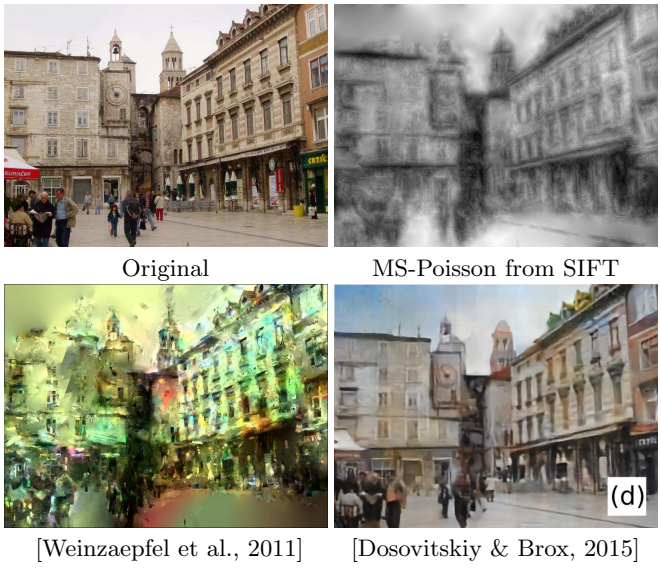

Fig. 3. Comparison for SIFT reconstruction. In the first row we display the original image and the reconstruction results from the SIFT descriptors obtained with MS-Poisson. In the second row we display the results obtained with the methods of [9] and [13]. (Image Credits [9]).

In Fig. 3 we compare our reconstruction results with the methods by Weinzaepfel et al. [9] and Dosovitskiy & Brox [13]. The main difference is that our method relies only on the content provided in the SIFT subcells while these methods use an external database either to copy local information from patches with similar SIFT descriptors in [9] or to build an up-convolutional neural network for reconstruction in [13]. Thus our work has no intention to outperform these methods in terms of visual quality of reconstruction (in particular, our method has absolutely no possibility of recovering the color information). But still, one can see that MS-Poisson can recover the global shapes with no stitching artifacts (compared to [9]). However, MS-Poisson gives much blurrier results; indeed the method of [9] is able to copy patches of sharp images while the method of [13] in some sense includes a deconvolutional step through the up-convolutional network. In order to obtain sharper results, one could restrict the inner sums of the MS-Poisson energy (6) to the SIFT subcells: this new method will certainly avoid over-regularization (but the solution will not be explicit anymore).

References

1. Morel, J.M., Yu, G.: ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences **2**(2) (2009) 438–469
2. Attneave, F.: Some informational aspects of visual perception. Psychological review **61**(3) (1954) 183
3. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on PAMI **27**(10) (2005) 1615–1630
4. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. Foundations and trends in computer graphics and vision **3**(3) (2008) 177–280
5. Ojala, T., Pietikinen, M., Menp, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on PAMI **24**(7) (2002) 971–987
6. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2) (2004) 91–110
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE CVPR. Volume 1. (2005) 886–893
8. Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: Proceedings of the IEEE ICCV. (2003) 1470–1477
9. Weinzaepfel, P., Jégou, H., Pérez, P.: Reconstructing an image from its local descriptors. In: Proceedings of the IEEE CVPR. (2011) 337–344
10. Vondrick, C., Khosla, A., Malisiewicz, T., Torralba, A.: Hoggles: Visualizing object detection features. In: Proceedings of the IEEE ICCV. (2013) 1–8
11. d'Angelo, E., Jacques, L., Alahi, A., Vandergheynst, P.: From bits to images: Inversion of local binary descriptors. IEEE Transactions on PAMI **36**(5) (2014) 874–887
12. Kato, H., Harada, T.: Image reconstruction from bag-of-visual-words. In: Proceedings of the IEEE CVPR. (2014) 955–962
13. Dosovitskiy, A., Brox, T.: Inverting Visual Representations with Convolutional Networks. arXiv:1506.02753 [cs] (2015)
14. Mahendran, A., Vedaldi, A.: Understanding deep image representations by inverting them. In: IEEE CVPR. (2015) 5188–5196
15. Mahendran, A., Vedaldi, A.: Visualizing deep convolutional neural networks using natural pre-images. International Journal of Computer Vision **120**(3) (2016) 233–255
16. Desolneux, A.: When the a contrario approach becomes generative. International Journal of Computer Vision **116**(1) (2016) 46–65
17. Pérez, P., Gangnet, M., Blake, A.: Poisson Image Editing. In: ACM SIGGRAPH 2003 Papers. SIGGRAPH '03 (2003) 313–318
18. Zhu, S., Wu, Y., Mumford, D.: Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. International Journal of Computer Vision **27**(2) (1998) 107–126
19. Mumford, D., Desolneux, A.: Pattern Theory: The Stochastic Analysis of Real-World Signals. A K Peters/CRC Press, Natick, Mass (2010)
20. Rey Otero, I., Delbracio, M.: Anatomy of the SIFT Method. Image Processing On Line **4** (2014) 370–396
21. Nesterov, Y.: Introductory lectures on convex optimization: A basic course. Volume 87. Springer (2004)