Reconstructing an image from its local descriptors

Philippe Weinzaepfel, Hervé Jégou, Patrick Pérez

Reconstructing an image from its local descriptors

Weinzaepfel
ENS Cachan Bretagne

Hervé Jégou
INRIA

Patrick Pérez
Technicolor

Figure 1. *Top*: the original picture. *Bottom*: the image reconstructed by our approach, which uses as only input the output of a standard local description software (position and shape of a number of "regions of interest" and for each of them a SIFT descriptor or the like).

Abstract

This paper shows that an image can be approximately reconstructed based on the output of a blackbox local description software such as those classically used for image indexing. Our approach consists first in using an off-the-shelf image database to find patches which are visually similar to each region of interest of the unknown input image, according to associated local descriptors. These patches are then warped into input image domain according to interest region geometry and seamlessly stitched together. Final completion of still missing texture-free regions is obtained by smooth interpolation. As demonstrated in our experiments, visually meaningful reconstructions are obtained just based on image local descriptors like SIFT, provided the geometry of regions of interest is known. The reconstruction allows most often the clear interpretation of the semantic image content. As a result, this work raises critical issues of privacy and rights when local descriptors of photos or videos are given away for indexing and search purpose.

1. Introduction

Image indexing and retrieval have received a considerable attention in the last few years, thanks to the generalization of digital personal devices. Existing systems now search in millions [17, 14] to hundred millions [8] of images on a single machine.

The most successful frameworks rely on local descriptors, such as the popular scale invariant feature transform (SIFT) [12], attached to a number of "interest" regions extracted beforehand. For large scale image search, efficiency requires that geometry information (position and shape of interest regions) is first ignored and local appearance descriptors are aggregated, e.g., within a bag of visual words [18]. However, the precision of search based on such global representations is often improved subsequently by getting back to the geometry information in a post-verification step that filters out wrong matches [16].

One key application of such systems is near-duplicate detection [2, 22], which is used in particular for detection of illegal copies. It is worth noting that the best performing approaches (e.g., [3]) in the copy detection task of TRECVID [19] rely on local descriptors and use a post-verification scheme. Recently, following a trend in watermarking system design, researchers realized that copy detection is a non-collaborative game: a pirate can lure the system by altering an image in a proper manner [10].

In this paper, we address another security aspect raised by image indexing: the privacy of images. When a copyright holder resorts to a third party content based image retrieval system, it would rather pass images' fingerprints to this third party than share its sensitive contents. This image information that allows efficient and accurate indexing and comparison of contents is typically composed of local appearance descriptors and additional meta-data, including

geometrical information to allow post-verification, associated to detected regions of interest.

Nevertheless, a privacy threat remains: since the descriptors extracted from an image provide a "summary" of its visual properties over its most informative fragments, it might be possible to use them to interpret the image content or, even, to create a pirated copy. Such reconstructions were evidenced for features like filters, where they were used [9, 20] to visually represent the information conveyed by the features.

In this paper, we show that this is also the case for state-of-the-art descriptors. We present and demonstrate an algorithm for reconstructing an image from its local descriptors, here based on SIFT, and associated geometric information. Similar to [5] and [21], we use an external image database to help the reconstruction. However, in contrast to these approaches that are geared towards image editing and image generation, we are not interested in completing or inpainting images in a convincing way. Our main objective is to show how a simple technique could permit an "image hacker" to interpret the content of the original image, and this even if there is no similar image in his external database.

Our technique progressively builds an approximation of the unknown image by reconstructing its regions of interest one by one. First, for each region the nearest descriptor is searched in the external database and the corresponding image patch is extracted and reshaped according to geometric information attached to the region. This patch is subsequently pasted, after receiving a correction that makes it blend seamlessly in current reconstruction of the image. After all regions of interest have thus been approximately recovered and glued together, some image portions might still be missing. The absence of detected interest regions indicates that these portions should be fairly deprived of texture and structure. They are reasonably reconstructed by smooth interpolation from reconstructed boundary conditions.

Not surprisingly, the resulting image is not perfect (see Figure 1 for a first illustration, and other examples presented in experimental section). However, it is visually close enough to the original unknown image to interpret its content. This makes explicit the privacy threat that lies in local image descriptions. Copyright holders, in particular, should thus be aware of the surprising amount of visual information given away when passing such descriptions to third party indexing and search systems.

The paper is organized as follows. Section 2 states more precisely the reconstruction problem that we address, with discussion of related works. The actual reconstruction algorithm is presented in Section 3. Experiments reported in Section 4 demonstrate the strengths of the algorithm as well as some of its limitations, which are discussed.

2. Problem statement

In this section, we first introduce the image information provided by a typical local description software. We then proceed with defining and analyzing the image reconstruction problem we want to solve based on such an information. Finally, we briefly discuss the relationship between this problem with other image manipulation tasks that aim at building images out of fragments.

2.1. Image description

A local description scheme, as depicted by Figure 2, is assumed in the rest of the paper. More precisely, an image with support Ω is described by a set of "interest regions" extracted by a detector. Each region is mapped to an intensity patch (a disc in our case, with regions being elliptically shaped[1]), whose appearance is summarized by a d-dimensional descriptor $\mathbf{v}_i \in \mathbb{R}^d$. The number of regions of interest depends on the image size and content.

As the output of the local description program, we typically have the following information for i-th interest region:

- appearance descriptor $\mathbf{v}_i$;
- coordinates $\mathbf{x}_i = (x_i, y_i) \in \Omega$ of region center;
- scale s_i and dominant gradient orientation o_i;
- symmetric definite matrix A_i that defines the elliptic support of the region.

Geometric information $(\mathbf{x}_i, o_i, A_i)$ uniquely defines the affinity W_i that maps elliptic image fragment to a centered normalized circular patch with dominant orientation 0.

All or part of above quantities are needed by the indexing system, in particular geometrical information is used during the geometrical verification stage, as done in [12, 16]. In the following, we denote $R_i = \{\mathbf{v}_i, \mathbf{x}_i, s_i, o_i, A_i\}$ this set of quantities associated with an interest region. Abusing the terminology, we refer to it as the "descriptor" of this region.

In this paper, we focus on SIFT descriptors [12], though our method should work for most local descriptors. SIFT descriptors are invariant to image orientation and scale, and are robust to affine and perspective distortions. They come in the form of normalized positive 128-dimensional vectors (d=128). As for region detector, we use an Hessian-affine technique [13]. Both detection and description are provided by Mikolajczyk's software, which is used in many works on image indexing, e.g., in [16, 7]. Given its output on an image, we aim at reconstructing this image approximately.

2.2. The reconstruction challenge

Reconstructing a single image patch from its local descriptor is impossible because local description drastically

[1] Note that the descriptor $\mathbf{v}_i$ is computed on a normalized square patch. However, to avoid corner artifacts, only the pixels of the inscribed disc are used at reconstruction time.

Figure 2. Image analysis stage, as done by a typical local description software. Regions of interest are first detected and affine-normalized to a fixed-size square patch, which is subsequently described by a local descriptors and additional meta-data used in geometrical verification. Note that in our reconstruction approach (see Section 3), we will reconstruct elliptic patches out of normalized circular patches.

compacts appearance information for both invariance and robustness purposes. It typically captures some aspects of local contrast distribution. As a consequence the description function is a many-to-one mapping whose inversion is ill-posed, unless appropriate additional priors or constraints are used. An exemplar-based prior will be obtained in our approach, thanks to an external image database. From these images, possibly very different from the unknown image of interest, a large number of image patches and associated descriptors will be extracted to allow approximate and simple inversion of local description function.

In the case of SIFTs, only weighted histograms of intensity gradient orientations computed over a partition of the normalized image patch are accounted for. Hence, the following difficulties have to be overcome when using the regions of interest that have been approximately recovered from this type of descriptors:

- There is no chrominance information.
- Since the descriptors are normalized according to the Euclidean norm, the absolute contrast of a given interest region is not known.

Also, the image is unevenly described, see Figure 3. Indeed, interest region detectors aim at selecting informative image fragments, typically those with specific contrast patterns. Texture and structure free regions of the image are thus devoid of interest regions and do not get described at all. These regions are usually very smooth in intensity, uniform sky portions for instance. Conversely, structure and/or texture regions trigger lots of overlapping region detections. Some pixels in such regions can get covered by more than one hundred interest regions spanning a large range of scales, shapes and positions.

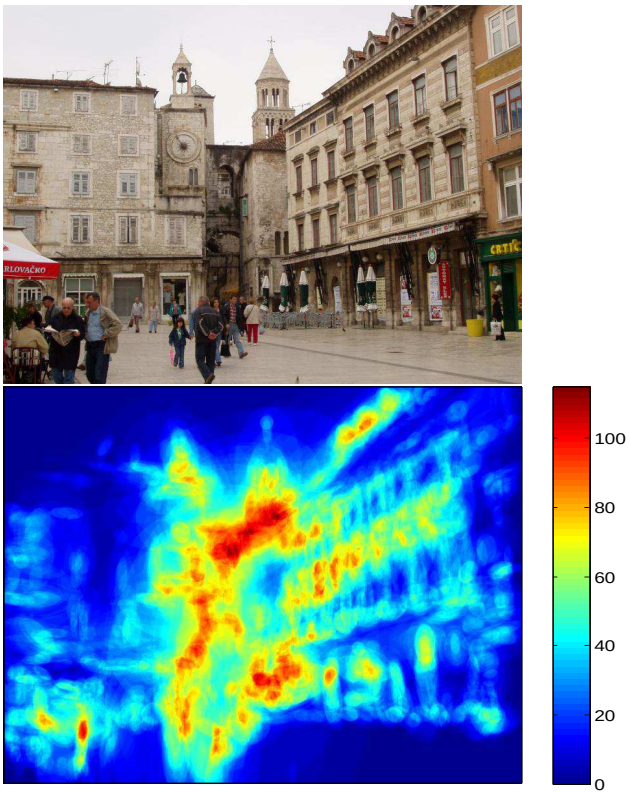

Figure 3. Number of (elliptical) regions of interest covering each pixel of the original image. Some pixels belong to many regions, but many locations are not or poorly described.

2.3. Link to image editing

Our image reconstruction problem bears connection with a number of image editing tasks where an image is built out of multiple images or image fragments, sometimes with some amount of interactivity: image composing and cloning where image cutouts are pasted in a new background; image completion and inpainting for restoration, correction or editing; stitching of multiple views from a scene to create panoramas for instance; automatic collage of multiple photos; "image analogy" (example-based image generation from auxiliary information such as low-res images for super-resolution or semantic segmentation for texture-by-number).

As we shall see, our reconstruction algorithm resorts to basic tools (harmonic correction or interpolation) that have extensively used within aforementioned image editing tasks. Addressing a completely different problem though, our work drastically departs from these research trends in several ways:

- Whereas in image editing problems information mostly remains attached to pixel grid (either in the

color domain or in the gradient domain), the input data is, in our case, of a very different nature (compacted descriptor domain with spatial information partly lost), which makes the task especially challenging as explained earlier. In particular, as opposed to image completion and image inpainting contexts, we do not have initial image data to start with.

- Image stitching and exemplar-based image generation techniques, whether for inpainting, image analogy or photo-collage, rely on collections of images or of image fragments that are quite homogeneous: in patch-based inpainting, all patches come from the image itself and they are not normalized to a fixed size irrespective of their original size; in image completion from large photo collections and in photo collages, large fragments of real photos are used, which helps preserving both visual and semantic quality of final composite; in image analogy, example patches with auxiliary information have to be very consistent with the input to be processed; panorama stitching concerns only few images of very similar content. In our case, there is a very large number of fragments (several thousands), which drastically differ in size, detail level and color, to be assembled.
- The amount of overlap between fragments can be extremely large at some places in our case, with nested inclusions; in contrast, in all exemplar-based image generation or completion techniques, fragments only overlap over thin borders such that most pixels of final image belong to only one source fragment, others rarely belonging to more than two.
- If final evaluation remains subjective for both our work and mentioned tools, we are not aiming at excellent visual quality of assembled image, simply at semantic recovery of most visual content.

3. Reconstruction algorithm

3.1. Overview

As explained in Section 2, we consider an external database of color images $\mathbf{I}_k$, k=1$\cdots M$, from which a set of interest regions are extracted off-line and described as $R_j = \{\mathbf{v}_j, \mathbf{x}_j, s_j, o_j, A_j\}$, j=1$\cdots m$. We shall denote S_j the pixel support of j-th region (ellipse centered at $\mathbf{x}_j$ and with shape defined by A_j) and $k(j)$ the index of the database image it stems from.

These regions and associated image patches will be used as a prior to invert local description function. Given a set of query descriptors $R_i = \{\mathbf{v}_i, \mathbf{x}_i, s_i, o_i, A_i\}$, $i = 1\cdots n$, extracted in the same way from an unknown color image $\mathbf{I}$ with support Ω, we aim at reconstructing this image approximately. The reconstruction proceeds as follows:

Figure 5. Reconstruction without blending: the patches are here copied without being adapted, yielding poor reconstruction. See Figure 4 for the original image.

1. For each query appearance descriptor $\mathbf{v}_i$, search its nearest neighbor in the descriptor database

$$j^* = \arg \max_{j \in \{1\cdots m\}} \|\mathbf{v}_i - \mathbf{v}_j\|_2, \tag{1}$$

and recover the corresponding elliptic image patch

$$\mathbf{q}_j^* = \mathbf{I}_{k(j^*)}(S_{j^*}). \tag{2}$$

Warp this patch such that it fits into the destination ellipse $S_i \subset \Omega$:

$$\mathbf{p}_i = W_i^{-1} \circ W_{j*}(\mathbf{q}_{j^*}). \tag{3}$$

2. Seamlessly stitch all patches $\mathbf{p}_i$, $i = 1\cdots n$, together to obtain a partial reconstruction with support $S = \cup_{i=1}^n S_i \subset \Omega$ (see details below) .
3. Complete remaining empty zone $\bar{S} = \Omega \setminus S$ by smooth interpolation, as shown in Figure 4 (see details below).

3.2. Seamless stitching of patches

Recovered image patches are numerous, they span a large range of sizes and shapes and they overlap a lot. This makes their joint stitching difficult. We take instead a "dead leaves" approach by stacking patches one after another, newly added patch partly occluding the current reconstruction if it overlaps it. Since large patches are more likely to exhibit visual artifacts due to extreme stretching of original source patch, we want to favor the contribution of smaller patches. The order of sequential stitching is thus chosen according to decreasing support's sizes.

Such a simple stacking is not sufficient though to get a satisfactory reconstruction. Indeed, since patches originate

Figure 4. From left to right: the original picture and the reconstruction before and after completion of uncovered regions.

from a large number of unrelated images, they usually exhibit very different appearances (chrominance, intensity and texture). This results in a disruptive patchwork effect as illustrated in Figure 5. If removing texture and structure discontinuities is difficult, this is not the case for color and intensity discontinuities. Such seams are easily concealed by Poisson image editing [15], as routinely done for cloning, composing and stitching.

Consider the stage of sequential stitching where i-th image patch $\mathbf{p}_i$ with support S_i has to be incorporated. Denote $\{\mathbf{I}(\mathbf{x}),\ \mathbf{x} \in T\}$ the image reconstructed up to that point with T the union of supports of patches used so far. If $S_i \cap T = \emptyset$, the new patch is simply copied in place: $\mathbf{I}(\mathbf{x}) = \mathbf{p}_i(\mathbf{x}),\ \forall \mathbf{x} \in S_i$. If S_i overlaps T, the imported patch is modified by an additive harmonic (null Laplacian) correction such that it fits exactly current image $\mathbf{I}$ at the border of the overlapping regions. More precisely, let's denote $\partial S_i = \{\mathbf{x} \in S_i \cap T : N(\mathbf{x}) \cap (T \setminus S_i) \neq \emptyset\}$, the intersection of the inner border of S_i (according to 4-nearest neighborhood $N(.)$) with T. Color values over ∂S_i stay as in current reconstruction, whereas new values are computed over $\tilde{S}_i = S_i \setminus \partial S_i$ such that:

$$\forall \mathbf{x} \in \tilde{S}_i,\ |N(\mathbf{x}) \cap S_i| \mathbf{I}(\mathbf{x}) - \sum_{\mathbf{y} \in N(\mathbf{x}) \cap \tilde{S}_i} \mathbf{I}(\mathbf{y}) = \sum_{\mathbf{y} \in N(\mathbf{x}) \cap \partial S_i} \mathbf{I}(\mathbf{y}) + \sum_{\mathbf{y} \in N(\mathbf{x}) \cap S_i} [\mathbf{p}_i(\mathbf{x}) - \mathbf{p}_i(\mathbf{y})]. \quad (4)$$

These three discrete Poisson equations (one per channel) on domain $\tilde{S}_i$ with Dirichlet boundary conditions have unique solutions that are obtained efficiently with either direct or iterative sparse solvers. Note that for better results, we work in CIE-Lab color space that separates image intensity from chrominance while exhibiting good perceptual regularity. Clamping is performed if resulting values are not in the admissible range.

3.3. Final completion by interpolation

When all patches recovered from descriptors have been stitched together, the image reconstruction is complete over $S = \cup_{i=1}^{n} S_i \subset \Omega$. Fragments that are still missing are likely to exhibit only little texture and structure in the original image. Hence, they are simply approximated by harmonic interpolation of known reconstruction over ∂S. Mathematically it is the same problem as before but without imported patch information:

$$\forall \mathbf{x} \in \bar{S},\ |N(\mathbf{x})| \mathbf{I}(\mathbf{x}) - \sum_{\mathbf{y} \in N(\mathbf{x}) \cap \bar{S}} \mathbf{I}(\mathbf{y}) = \sum_{\mathbf{y} \in N(\mathbf{x}) \cap \partial S} \mathbf{I}(\mathbf{y}). \quad (5)$$

This system is solved as previous ones. Note however that, if $\bar{S}$ has more than one connected component, the system can be split into several independent subproblems, one per component, for sake of efficiency.

4. Experiments

In this section, after introducing the datasets used in our experiments, we present the reconstruction results for several images and underline the remaining weaknesses of the reconstruction. We then analyze the impact of the external database size on the final reconstruction.

4.1. Datasets

The experiments are carried out using two image datasets introduced for the evaluation of indexing systems: the INRIA Holidays [6] and Copydays [4] datasets. Both are composed of holiday snapshots. The first one contains 1491 images grouped in 500 distinct sets, each of which is associated with the same scene or object. Different types of photos are included: natural images, man-made, crowd, etc. The entire dataset is described by 6 756 563 SIFT descriptors. The Copydays dataset is composed of 157 independent photos and several (artificially) transformed versions of these images, which were used as queries in [4]. We do not use these synthetic transformed images in our paper.

Our goal is to measure to what extent the interpretation of the reconstructed image content is possible. Therefore, the performance of the algorithm is judged in a subjective manner, based on the quality of the reconstruction with respect to a possible interpretation by human. We there-

fore present reconstruction results, and underline failure and pathological cases. Two scenarios are considered:

Scenario I: The images of Copydays are reconstructed using Holidays. As there is no intersection (no common object or scene) between the two datasets, this scenario corresponds to the case where the image to be reconstructed has no corresponding image in the external database used to support the reconstruction. A few images of monuments downloaded from the web are also considered.

Scenario II: The queries of Holidays are reconstructed using the Holidays dataset. Each query is removed in a leave-one-out manner. This scenario reflects the case where the image to reconstruct is similar to some images of the external dataset, which is likely to happen on common objects, logos, or famous places if we use a large external set. There is at least one image similar to the image to reconstruct in the dataset.

4.2. Reconstruction results

Hereafter, we analyze the impact of the evaluation scenario and of the external database size on the reconstruction. We also underline some limitations of the algorithm, most of which are inherent to the description used as input.

Scenario I *vs* Scenario II. Reconstruction results for various types of scenes are presented in Figure 6 and Figure 7, for Scenario I and II, respectively. In both cases, the humans (the man on the bike in Figure 6, and the Asian women in Figure 7) are not very well reconstructed, and require an interpretation effort to distinguish the person. However, it is still possible to recognize the person if the face is large enough, see the dictator in Figure 10. Natural (trees, leaves) and man-made objects can be recognized to some extent: the cars can be localized in the left-most image of Figure 6. However it is not possible to recognize their brands. Overall, the buildings and text look better than vegetation or human beings. Famous monuments are easily recognized from the reconstructed images, see Figure 8.

Comparing the results from Scenario I (Figure 6) to those of Scenario II (Figure 7), the reconstruction is slightly improved if similar images are contained in the database. Note that if the image to reconstruct is in the external database, then the reconstructed image is almost perfect (except for the uncovered areas, which are interpolated): only interpolation artifacts are observed.

Limitations: Although our method reconstructs an interpretable image, the reconstruction is imperfect. Hereafter we focus on the main artifacts and why they appear.

Color. The fact that color is poorly reproduced is not surprising because SIFT descriptors do not contain any color or absolute photometric information. In some cases, as for vegetation, texture and color are related and the dominant color remains satisfactory. However, this is not the case in general. Looking at Figure 10, it appears that a black and white image is reconstructed with colors, and the dominant color varies from one reconstruction to another. This is because the algorithm is sensitive to the selection of the first patches (which are different in Figure 10 because the database is not the same), which have a significant impact on the dominant color. In our opinion, the best way to handle color would be to insert the user in the loop by using a weakly supervised reconstruction, or by using a dedicated colorization technique, as the one proposed in [11].

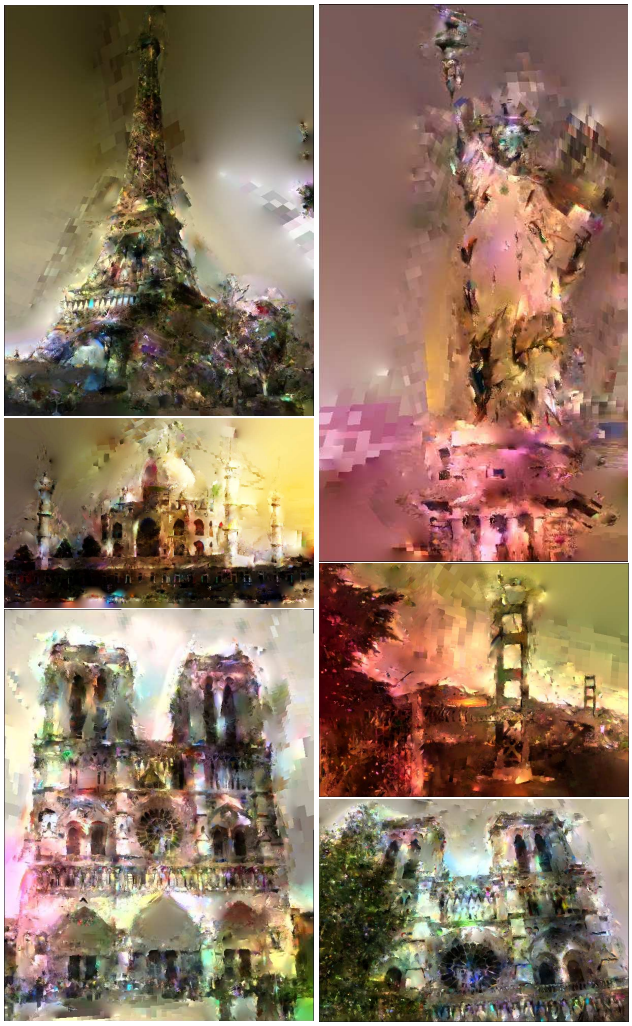

Figure 8. Reconstruction of famous buildings (Scenario I).

The richness of the image description has a strong impact on the reconstruction quality. Images described by few descriptors (less than 100) are poorly reconstructed in general, as shown by Figure 9, where the original image is described by a limited number of salient regions. Moreover, most of the image pixels are not covered at all. In that case, interpolation over large areas with few boundary conditions fails to

Figure 6. Scenario I: Reconstructions of images from Copydays using the external dataset Holidays.

Figure 7. Scenario II. Reconstructions of images from Holidays using Holidays deprived of query image as external dataset.

invent the missing area, for instance the clouds in the sky. Even if the uncovered regions contain a limited amount of visual information, the overall rendering severely impacts the interpretation of the image.

Finally, *pixelization* occurs when large regions are reconstructed from small ones. The absolute photometric intensity is often quite different from the original image, and spurious edges and lines appear.

Impact of the size of the database. Intuitively, the larger the external database, the better the reconstruction: in that case more tuples (descriptors,patches) are available and the probability to find a better patch is higher. This is confirmed by Figure 10, where two images are reconstructed using an external database of increasing size. As the database grows, the artifacts tend to disappear and the details are reproduced with higher fidelity.

Figure 9. An example of a picture with only 63 regions of interest: (left) original picture; (right) reconstruction before completion

5. Conclusion

This paper, by showing that an image can be reconstructed from its local descriptors in a way that allows interpretation of its content by human, raises the problem of privacy of image description by state-of-the-art local descriptors. To our knowledge, this issue is ignored in existing indexing systems, despite the value of the indexed content.

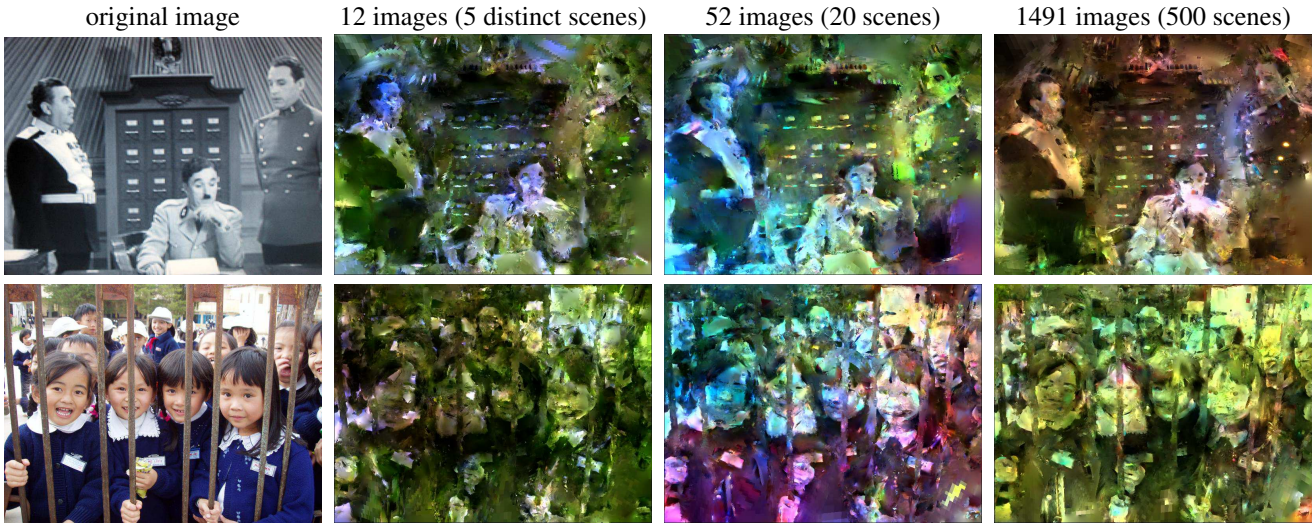

Figure 10. *Left to Right*: original image and its reconstruction based on an external image set of increasing size: 12, 52 and 1491 images.

The proposed method is entirely automatic, which suggests that much better reconstruction could probably be obtained with user interaction, in particular to overcome the lack of color information. Doing so, it is still not clear, however, to which extent the images could be reconstructed with sufficient commercial value and become pirated copies. Content providers should take care of this issue.

References

[1] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *IJCV*, 74(1):59–73, 2007.

[2] O. Chum, J. Philbin, and A. Zisserman. Near duplicate image detection: min-hash and tf-idf weighting. In *BMVC*, September 2008.

[3] M. Douze, H. Jégou, and C. Schmid. An image-based approach to video copy detection with spatio-temporal post-filtering. *IEEE Trans. on Multimedia*, 12(4):257–266, jun 2010.

[4] M. Douze, H. Jégou, H. Singh, L. Amsaleg, and C. Schmid. Evaluation of GIST descriptors for web-scale image search. In *CIVR*, July 2009.

[5] J. Hayes and A. Efros. Scene completion using millions of photographs. In *SIGGRAPH*, 2007.

[6] H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*, October 2008.

[7] H. Jégou, M. Douze, and C. Schmid. Improving bag-of-features for large scale image search. *IJCV*, 87(3), May 2010.

[8] H. Jégou, M. Douze, C. Schmid, and P. Pérez. Aggregating local descriptors into a compact image representation. In *CVPR*, jun 2010.

[9] D. Jones and J. Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *ECCV*, pages 395–410, 1992.

[10] E. Kijak, T. Furon, and L. Amsaleg. Challenging the security of cbir systems. Technical Report RR-7153, INRIA, December 2009.

[11] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. *ACM. Trans. Graph. (SIGGRAPH)*, 23(3):689–694, 2004.

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[13] K. Mikolajczyk and c. schmid. scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[14] D. Nistér and H. Stewénius. Scalable recognition with a vocabulary tree. In *CVPR*, June 2006.

[15] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM. Trans. Graph. (SIGGRAPH)*, 22(3):313–318, 2003.

[16] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, June 2007.

[17] T. Quack, U. Mönich, L. Thiele, and Manjunath. Cortina: a system for large-scale, content-based web image retrieval. In *ACM Multimedia*, pages 508–511, 2004.

[18] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *ICCV*, October 2003.

[19] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *Intl Work. Multi. Inform. Retrieval (MIR)*, pages 321–330, 2006.

[20] A. Torralba and A. Oliva. Depth estimation from image structure. *PAMI*, 24(9):1226–1238, 2003.

[21] O. Whyte, J. Sivic, and A. Zisserman. Get out of my picture! internet-based inpainting. In *BMVC*, 2009.

[22] Z. Wu, Q. Ke, M. Isard, and J. Sun. Bundling features for large scale partial-duplicate web image search. *CVPR*, 0:25–32, 2009.