Image Retrieval Using Local Characterization

Cordelia SCHMID and Roger MOHR

INRIA
655 avenue de l'Europe
38330 Montbonnot
FRANCE
first.last-name@imag.fr

Abstract

This paper presents a general method to retrieve images from large databases using images as queries. The method is based on local characteristics which are robust to the group of similarity transformations in the image. Images can be retrieved even if they are translated, rotated or scaled. Due to the locality of the characterization, images can be retrieved even if only a small part of the image is given as well as in the presence of occlusions. A voting algorithm, following the idea of a Hough transform, and semi-local constraints allow us to develop a new method which is robust to noise, to scene clutter and small perspective deformations. Experiments show an efficient recognition for different types of images. The approach has been validated on an image database containing 1020 images, some of them being very similar by structure, texture or shape.

1 Introduction

Image retrieval is an important problem for accessing large image databases. We address the problem of retrieving any kind of images under the following conditions:

- partial visibility and complex background or clutter;
- different viewing angles,
- moderate changes in illumination;
- thousands of potential reference shapes.

Existing approaches use either geometric features of an object or rely on its luminance signature. Geometric approaches are robust to transformations and occlusions, but they only allow to deal with certain classes of objects. On the other hand, photometric approaches allow to deal with any kind of objects, but they do not work if the object is only partially visible. Furthermore, these methods are not invariant to any kind of image transformation. Recently [5] developped a method invariant to rotation using steerable filters. When considering colour, Slater and Healey [7] developped illumination invariant descriptors used for image recovery [2].

This paper presents a approach which allows to derive a set of greylevel or colour image invariant together with an indexation method allowing fast retrieval of parts of image already seen in similar condition (view point, illumination). The method uses local characteristics of the greyvalue signal which are invariant to similarity transformations. These characteristics are calculated at automatically detected keypoints, as shown in figure 1; for illustration purpose only some of the keypoints are displayed.

The originality of this work consists of several points. The use of local greyvalue differential invariants for indexing into a database presents the most important novelty. These invariants are continuous and

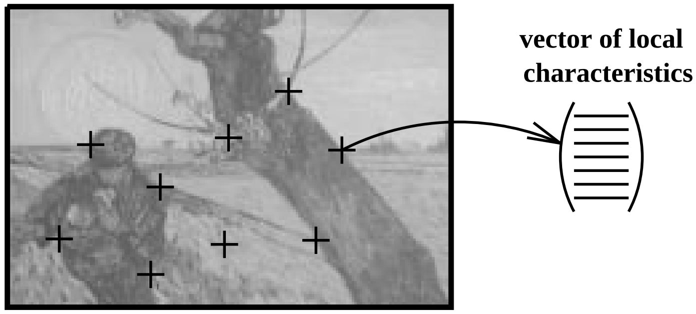

Figure 1: Representation of an image

independent of any image displacement. Another important point is the use of automatically detected keypoints which are representative of the object. Other authors [5, 9] use points fixed on a grid. As these grid points might not be significant, the vectors they use have to be much longer than ours. In case of occlusions, grid placement gets difficult and recognizing parts of images is impossible, as the grid can not be centered any longer. Our method avoids these drawbacks.

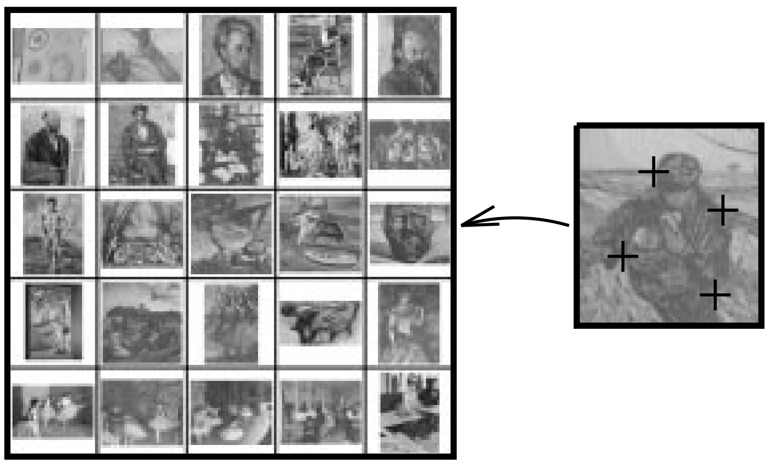

Figure 2: Research in the base, for illustration purpose only some of the keypoints are displayed

2 Method

The first step of our algorithm is the extraction of keypoints. The advantage of keypoints is that the informational content of the signal is high at their location. Furthermore, keypoints are local primitives. Standard vision algorithms exist for automatic extraction of keypoints. It is important that the detector is repeatable, that is results have to be invariant to image transformations. A comparison of different detectors in the presence of image rotation, scale change, light changes and image noise has shown a poor stability of existing methods and best results for the Harris detector. A stabilized implementation of this detector has been used in the present work.

The second step of our algorithm is to compute the local characterization. It is based on differential greyvalue invariants [3, 6] under the group of rigid motion. Due to a stable implementation of these invariants, a reliable characterization of the signal is obtained. A multi-scale approach [8, 4] makes this characterization robust to scale changes up to a factor 2. This has never been reported in the literature.

The third and final step of our algorithm is the retrieval or matching algorithm as schematized in figure 2. The image to be retrieved is compared to the images stored in the database. We therefore compared the vectors calculated for the image to be retrieved and the vectors calculated for the images in the database. The Mahalanobis distance is used to take into account uncertainties. A voting algorithm determines the most likely image. To allow fast retrieval of the image, the vectors of the database are organized in an index table : the vectors are ordered in a multi-dimensional hash table. Each level of this multi-dimensional hash table indexes one component of a characterization vector.

If we are dealing with complex scenes the voting algorithm may result in several hypotheses. We therefore add constraints of local coherence. For a given match, at least half of the neighbor keypoints have to be compatible and angular spacements have to correspond. Robust recognition is then possible even in case of important geometric transformations and with only an image fragment.

3 Experiments

The database used for our experiments contains more than 1000 images. These images are of different types : painting images, aerial images and images of 3D objects. Some images of the database are shown in figure 3. Experiments conducted for this database have shown the robustness of the method to image rotation, scale change, partial visibility and scene clutter.

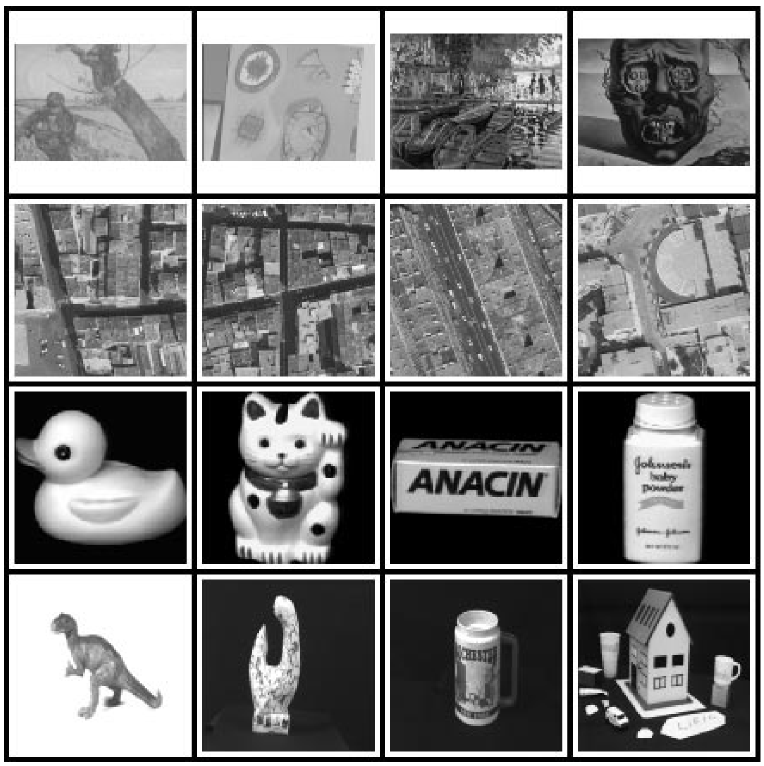

Figure 3: Some images of the database. The database contains more than 1000 images.

A set of test images to be retrieved contains 1000 images, either taken from a different point of view, under image rotation or scale change. The recognition rate obtained is 99%. Recognition for part of images has also been tested. The extracted parts cover 20 % or less of the entire image. The recognition rate is again near 100 % even if the image is rotated or scaled or if small viewpoint changes occur.

For more important viewpoints changes, for which we are able to recognize correctly being given the entire image, recognizing part of images works not as well (74%). This is due to the fact that small parts do not contain enough points, that is the number of votes is limited. In this case the robust algorithm can not overcome the uncertainty statistically.

Figure 4 shows parts of painting images which allow to correctly retrieve the entire image. Correct retrieval is also possible in case of image rotation and scale change for entire images as well as for parts of images (cf. 4).

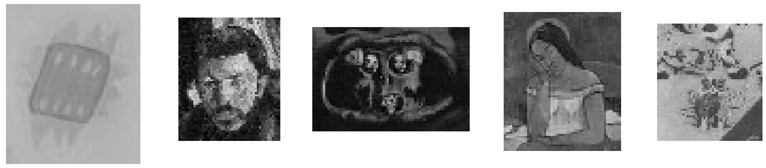

Figure 4: Parts of paintings.

Another example is displayed in figure 6 for an aerial image. On the right of figure the recognized image is shown (a black frame indicates the corresponding part of the image). Recognition is possible for a part which has been rotated and scaled. Furthermore, there is a small perspective deformation between the two images, as the airplane has moved.

Typical retrieval time is about 20 second on a Spark10 for this kind of experiments: 15 seconds for image processing and 5 seconds for searching through the huge 20 Mb index table encoding the 1020 images. Faster implementations are easy to consider, either using devoted image processing systems, or by exploiting the large extrinsic parellism of the processes.

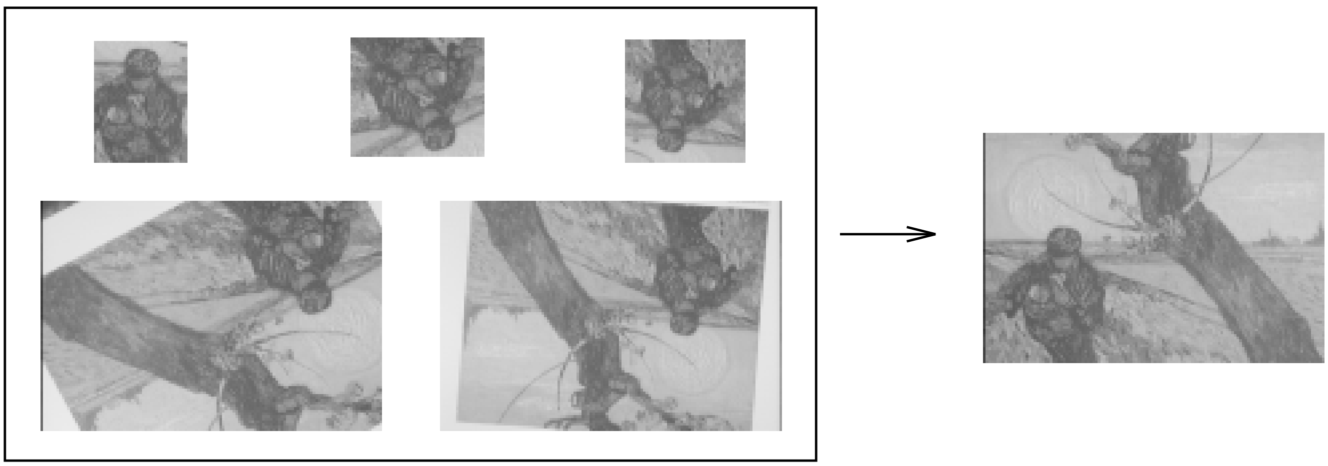

Figure 5: The image on the right is correctly retrieved using any of the images on the left.

4 Conclusion and future work

Our approach is an important contribution to image retrieval. It makes retrieval possible for images in situations which could not be dealt with before. We can identify images in case of partial visibility, image transformations and complex scenes. The success of our approach is based on the combination of differential invariants computed at keypoints with a robust voting algorithm and semi-local constraints. These invariants can be implemented with a sufficiently small filter size to capture local discriminant greylevel information. Moreover, the multi-scale approach makes our method robust to scale changes up to a factor 2.

We are presently extending these invariance to different directions:

- robust invariant in order to allow partial occlusion even in the local subwindow where the computation is done;

- colour invariance; instead following the path opened by [7], we experimented the colour invariance introduced by Funt and Finlayson [1]; in particular, this invariant seems very promising and it allows to some extend to take into account the shadows.

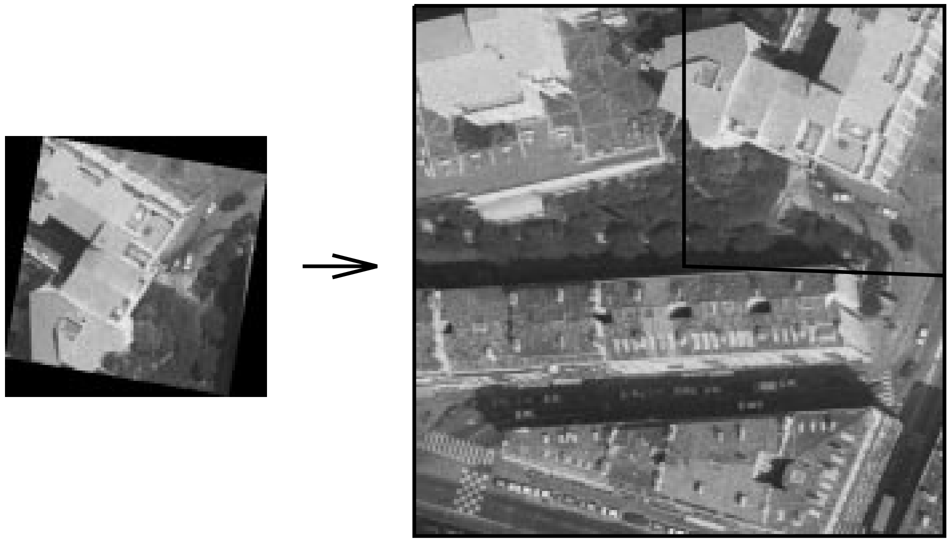

Figure 6: Recognizing part of an aerial image (courtesy of Istar).

5 Acknowledgements

Cordelia Schmid was supported by the European Community within the framework of the Human Capital and Mobility program.

References

[1] G. Finlayson, S. Chatterjee, and B. Funt. Color angular indexing. In *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, pages 16–27, 1996.

[2] G. Healey and D. Slater. Computing illumination-invariant descriptors of spatially filtered color image regions. IEEE *Transactions on Image Processing*, 6(7):1002–1013, 1997.

[3] J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[4] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[5] R.P.N. Rao and D.H. Ballard. Object indexing using an iconic sparse distributed memory. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 24–31, 1995.

[6] B.M Romeny, L.M.J. Florack, A.H. Salden, and M.A. Viergever. Higher order differential structure of images. *Image and Vision Computing*, 12(6):317–325, 1994.

[7] D. Slater and G. Healey. Combining color and geometric information for the illumination invariant recognition of 3D objects. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 563–568, 1995.

[8] A.P. Witkin. Scale-space filtering. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, Germany*, pages 1019–1023, 1983.

[9] X. Wu and B. Bhanu. Gabor wavelets for 3D object recognition. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 537–542, 1995.