

Image sampling for localization using entropy

Loic Lacheze and Ryad Benosman

loic.lacheze@isir.fr, ryad.benosman@isir.fr

ISIR, FRE2507, Université Pierre et Marie Curie -Paris 6, F-75016, France

Abstract

This paper introduces a robust adaptive patches sampling technique. The method does not rely on the use of keypoints to extract local information but all information contained in images. It performs an optimal multilayer quadtree decomposition of images driven by the quantity and homogeneity of information. Extracted patches will be of different sizes according to the covered zones in the image and the information they contain. Experimental results carried out in localization, including different cases of corrupted images, and image topology. Finally to illustrate the technique possibilities, preliminary results in object recognition are shown.

1 Introduction

In the last few years, the problem of extracting features from images has received growing attention. The majority of existing methods use derivatives approaches and rely on local image patches as basic features [1]. Recently, bag-of-features [8] representations have become popular, they are geometry free, based purely on characterizing the statistics of local patch appearances. The idea behind the method is to extract a set of local image patches which are sampled and assigned a metric description. The resulting quantified descriptors give an implicit distribution description space that can be quantified using different methods. Most of the existing work differ mainly according to the way patches are sampled and then described. They are generally selected using keypoints SIFT [4] based approaches. Codebooks [6, 7] are then produced using k-means and agglomerative clustering. Most of the cited techniques consider partial information from scenes mainly distributed around maximal gradient points, which limits the robustness of visual loops. The presented method is driven by the idea that all the information contained in images is useful. The paper introduces a new technique

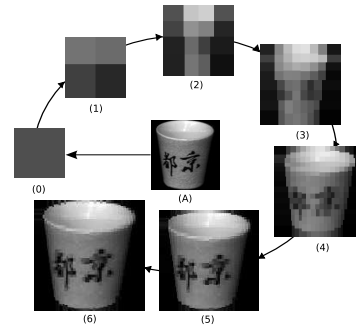


Figure 1. Examples of optimal sampling of image(A). Each patch contains his mean color in grey level.

of sampling images. It is based on a dense multilayer decomposition of the image driven by the quantity and homogeneity of the information contained within subpatches. Extracted patches will be of different sizes according to the covered zone in the image. The method can be applied to any type of images, we will present applications in both perspective and omnidirectional cases.

2 Optimal decomposition of images

An efficient decomposition must produce an optimal and possibly a unique partitioning of images. In addition it would be interesting to produce less patches, but of variable size so that they can cover homogeneous texture zones. In order to achieve an optimal generation of patches, a recursive algorithm is set up. Classic quadtree algorithm cut recursively images into subimages and so on. Starting from the initial image, each subimage is cut into four equal subimages. The idea is to use the same principle, but at the contrary of the regular quad-tree approach, the division of subimages will be driven by a measure of the information they contain.

The goal is to cut a subimage at the location where the difference of the quantity of information between possible subimages is minimal. This information is given for an image I by :

$$H(I) = - \sum_{c=0}^{c=255} Occ(I=c) \log P(I=c)$$

with $Occ(I=c)$ the number of times the pixel value c appears in I , $P(c)$ is the probability of appearance of the grey value c within I .

To estimate the optimal point minimizing the variance of the distance between the information contained in the four subimages of I , the principle of integral images introduced in [2] is used.

Let $q(i, j)$ be the quantity of information of a pixel $I(i, j)$ with $q(i, j) = \log(P(I(i, j)))$. We set the integral information of $I(x, y)$ as :

$$QI(x, y) = \sum_{i \leq x, j \leq y} q(i, j)$$

This sum is computed in one iteration on the whole image or subimage considered. We set $R(x, y-1)$ the integral quantity of information on the row x of height $y-1$. The principle of computation is presented in figure 2(A).

$$R(x, y-1) = QI(x, y-1) - Q(x-1, y-1)$$

Finally the integral quantity of information for a (x, y) (see figure 2(B)(C)) is given by:

$$QI(x, y) = QI(x-1, y) + R(x, y-1) + q(i, j)$$

Once QI is computed the variance value within each pixel becomes implicit. It is important to compute the mean value of the quantity of information contained within the four subimages. In the case where I is of size $m \times n$ we have for a cutting position $(x = m/2, y = n/2)$:

$$QI_m = QI(m, n)/4$$

The quantity of information of each zone is :

$$QI_{11}(x, y) = QI(x, y)$$

$$QI_{12}(x, y) = QI(m, y) - QI(x, y)$$

$$QI_{21}(x, y) = QI(x, n) - QI(x, y)$$

$$QI_{22}(x, y) = QI(m, n) - QI_{21} - QI_{12} + QI_{11}$$

Finally the optimal (x, y) position is the one minimizing the following sum of differences:

$$\exists(x, y) / \min_{x, y} \left(\sum_{a=1, b=1}^{a=2, b=2} (QI_m - QI_{ab}(x, y))^2 \right)$$

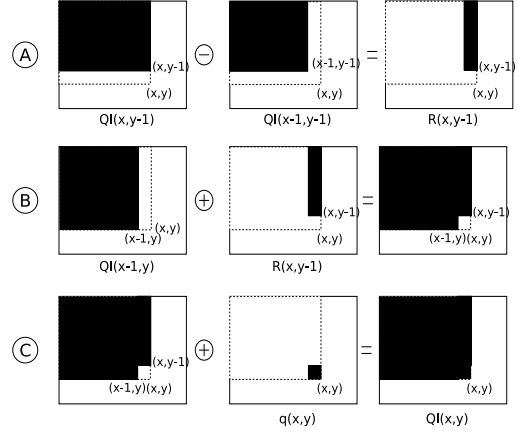


Figure 2. Computation of $QI(x, y)$ using $QI(x-1, y-1)$, $QI(x, y-1)$, $QI(x-1, y)$ and $q(x, y)$. In (A) The computation of $R(x, y-1)$ and in (B) and (C) the final computation of $QI(x, y)$.

An example of an image decomposition is shown in figure 1. In what follows the extracted patches are described using two measures, their greyscale mean value and a texture description as defined in [3]. The comparison of two images is performed by comparing their patches according to their locations, descriptions and the level at which they appear.

3 Localization

An autonomous differential drive robot equipped with an omnidirectional catadioptric sensor mounted on top is used [3]. Omnidirectional catadioptric sensors images are variant scale sensors. Pixels do not have the same resolution, the notion of neighborhood is then lost. Differential methods are no longer adapted to the geometry of images [10], most existing methods are not meant to be used on such cases. The position of the robot is constantly estimated by a camera network mounted on top, it provides a ground truth measure of the estimated positions given by the navigation system of the robot. The robot explores randomly its environment, and creates a dense topologic map. Nodes are created every 20 cm as shown by figure 3, they store the optimal sampling of its corresponding omnidirectional image and its location within the map. The optimal sampling is compared to a localization method using SIFT, the database has a total of 195 locations. The indoor arena used has a size of 4m x 4m. The optimal

| | ref. | w. noise | occl. |
|---------------|-------|----------|-------|
| SIFT | 0.586 | 0.591 | 0.558 |
| opt. sampling | 0.758 | 0.781 | 0.746 |

Table 1. Indoor localization rates.

sampling is applied on raw images, two spatially close positions generate two images which optimal sampling provides very close decompositions. In order to test different scenarios, the content of acquired omnidirectional images are modified in the three following manner. Virtual occlusions are added (from 1 to 10 squares) at random positions, their individual size never exceeding 10% of the size of the original image. White noise is added having a uniform distribution between 0 and 255 concerning 10% of the maximum number of pixels.

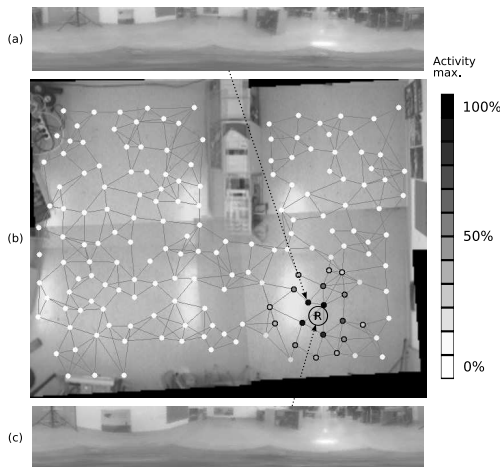


Figure 3. Activity of the current position compared to known locations.

The localization results are shown in Table 1 and have required a precise orientation for each view. It is interesting to notice that the optimal sampling produces better results than SIFT this due to two main reasons. The first one is connected to the geometry of images, SIFT can not extract reliable corner points due to the non linear resolution. The second reason is linked to the fact that the optimal sampling uses the whole image to localize, some indoor scenes do not contain sufficient corner points, thus SIFT fails to extract sufficient features to ensure localization. We have performed the same experiments with a perspective cameras and the two methods provided similar results due to the same reason. The trajectory of the robot is shown in figure 4, it appears that the robots succeeds in locating it-

self within the scene as both the estimated and ground truth trajectories are close. The location is estimated using a weighted triangulation of most active nodes and provides a mean error of 9 cm.

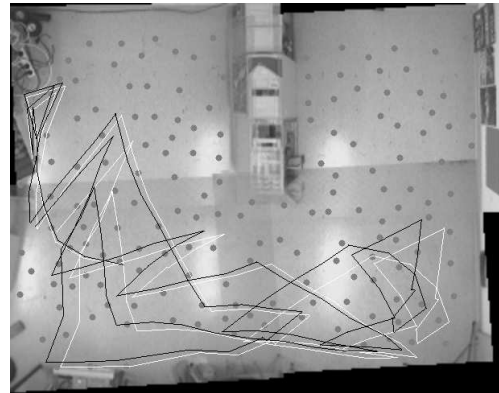


Figure 4. Upper view of the arena showing the estimated trajectory (white) of the real tracked positions (black).

4 Objects recognition

The initial aim of the method being extracting features for localization, we carried out in a second stage preliminary experiments of the optimal sampling in the context of object recognition. In order to recognize objects their must be a stability of decomposition so that the object can be described in a unique manner. Experiments are carried out using COIL-100. Each object is characterized by 72 images, 8 are used for generating codebooks corresponding to a sampling of $1/40$. The codebook is generated using the selected images and is set to a maximal size of 32 patches. Images were artificially corrupted, the recognition rate for each case is : (a) objects where randomly translated within the image : 0.963, (b) white noise is added : 0.954, (c) the scale of the objects is modified with a uniform probabilistic distribution of scales between 0.3 and 1.9 of the original size : 0.853, (d) all corruptions simultaneously: 0.846. These results are due to the fact that decomposition of the objects is very stable, as shown in by figure 5, the vocabulary associated with recognized patches is very stable, even in case of severe corruption of the original image. Objects that are poorly textured tend to produce the worst results which is an expected result as the description function is texture. We compared the optimal sampling to the methods described in [9, 11] and SIFT. Results are presented in Table 2, they

| learning size | 18 im. | 8 im. | 4 im. |
|--------------------------|--------|-------|-------|
| LAFs [9] | 0.999 | 0.994 | 0.947 |
| SNoW/edges [11] | 0.941 | 0.892 | 0.883 |
| opt. sampling - texture | 0.991 | 0.953 | 0.833 |
| opt. sampling - color | 0.995 | 0.974 | 0.886 |
| opt. sampling - position | 0.862 | 0.768 | 0.658 |

Table 2. Recognition rates on COIL-100.

correspond to the recognition rate per number of learned images of each object, corresponding respectively to an image each 20, 40 and 90. The optimal sampling used several patch description, the best results were given by color that achieves recognition rates very close to those given by [9]. One major additional advantage is the memory load to store the extracted features to index the database COIL-100, SIFT used 130 Mb whereas the optimal sampling needed only 2 Mb.

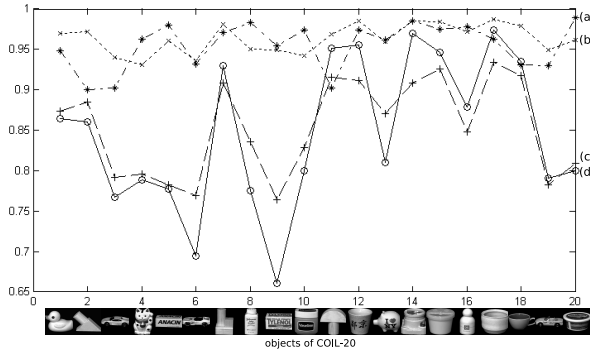


Figure 5. Rate of patch stability for each object, for different image corruptions.

5 Conclusion

This paper introduced a robust adaptive sampling method to extract patches from perspective and omnidirectional images. All the content of the images is used. We have shown that the presented sampling performs robust localization, which is the initial aim of the approach. Preliminary results showed that the technique can be applied to object recognition, it provided similar results to most used techniques with a very low memory load. Current work is adding new strategies of sampling using gradient information, to ensure a better stability and extend the method to detect objects within unknown scenes.

References

- [1] Frederic Jurie and Bill Triggs, Creating Efficient Codebooks for Visual Recognition, International Conference on Computer Vision (ICCV), 2005.
- [2] Viola, P. and Jones, M. , Rapid object detection using a boosted cascade of simple features, on Computer Vision and Pattern Recognition (CVPR), 2001.
- [3] Lacheze, L. and Benosman, R., Visual localization using an optimal sampling of Bags-of-words with entropy, International Conference On Intelligent Robots and Systems (IROS) ,2007.
- [4] Lowe, D.,Distinctive image features from scale-invariant keypoints,International Journal of Computer Vision (IJCV), 2004.
- [5] Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D. , Shape distributions, on ACM Transactions on Graphics, 2002.
- [6] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. on Conference on Computer Vision and Pattern Recognition (CVPR), 2003.
- [7] Filliat, D., A visual bag of words method for interactive qualitative localization and mapping, International Conference on Robotics and Automation (ICRA), 2007.
- [8] Nowak, E., Jurie, F. and Triggs, B., Sampling strategies for bag-of-features image classification, European Conference on Computer Vision (ECCV), 2006.
- [9] Stepan Obdrzalek and Jiri Matas. Object recognition using local affine frames on distinguished regions. In Paul L. Rosin and David Marshall, editors, Proceedings British Machine Vision Conference (BMVA), 2002.
- [10] Se, S. and Lowe, D. and Little, J., Vision-based Mobile robot localization and mapping using scale-invariant features, Proceedings of the IEEE International Conference on Robotics and Automation (ICRA),2001.
- [11] Yang, M.-H., Roth, D. and Ahuja, N. , Learning to Recognize 3D Objects with SNoW in European Conference on Computer Vision (ECCV), 2000.