Dependency of detectors' and descriptors' efficiency on image resolution for SIFT, SURF and ORB

Victor Eruhimov, Anna Kogan
Itseez Ltd, Nizhny Novgorod, Russia
Department of Computational Mathematics and Cybernetics
Nizhny Novgorod State University, Nizhny Novgorod, Russia
victor.eruhimov@itseez.com, anna.kogan@itseez.com

Abstract

In this paper we present a method for evaluation of dependency of object recognition efficiency by state-of-the-art methods on initial image size. By means of this method comparison of efficiency of various state-of-the-art image detectors and descriptors is made, and parameters providing best quality of ORB method are found.

Keywords: *detectors, descriptors, performance evaluation.*

1. INTRODUCTION

Detectors and descriptors of images are widely applied in the field of computer vision in such tasks as recognition of objects on images, stitching, visual odometry, splitting of objects into categories, creation of augmented reality systems and other tasks.

Often in practice it is necessary to analyze images of low resolution or quality. It happens, first, because of technical capabilities of the cameras installed, for example, on robots. Secondly, big dimension of images demands big computing expenses and memory sizes for information storage. We will investigate dependence of quality of recognition on resolution of the image and will try to find existing methods that allow achieving the best results at low resolutions.

In this work some detectors and the descriptors possessing high quality of recognition of the image are chosen. We will compare their efficiency using the benchmark offered by us. Criterion for comparison is the recognition accuracy, i.e. percent of the correct matches between keypoints among all established matches. For detectors we also will consider repeatability of detection. We will try to find parameters for the detector and a descriptor ORB providing high precision of recognition.

1.1 Related work

A huge number of affine regions detectors and local descriptors exist nowadays. To limit the comparison of quality made in this paper, we study only methods that seem to outperform others on recognition accuracy. First is SIFT [4], presented in 1999 by David Lowe, which is generally used as a benchmark when speaking about object recognition. Also we study SURF [1] which has been widely applied in computer vision and showed good performance. Third method to study is ORB [9], a new method presented in 2009 by group of researchers (Ethan Rublee, Vincent Rabaud, Kurt Konolige and Gary Gary Bradski). It is designed to perform good recognition quality in real-times.

Evaluation of detectors' and descriptors' efficiency on has been a popular topic of investigation in computer vision recently. [6] contains an exhaustive comparison of state-of-the-art affine region detectors in context of object recognition, introducing a method to compare efficiency on which we base in our studies. The paper studies dependencies of detectors repeatabilities on different scene transformations, such as changes of viewpoint, illumination, scale, blur, JPEG compression. No single method seems to outperform others in all the situations.

[5] presents an extensive research on quality of local descriptors and present a novel descriptor GLOH, based on SIFT method. The research shows GLOH and SIFT to give the best performance almost in any case.

In previous works only a little attention was given to dependency of recognition quality on initial image size. We evaluate the dependency in current work.

1.2 Overview

In chapters 2 and 3 we review existing state-of-the-art detectors and descriptors. Chapter 4 describes implementation details of a benchmark and dataset used for comparison. We present results of conducted experiments in chapter 5 and discuss them in chapter 6.

2. DETECTORS

Detector is an algorithm that finds image areas which contain some special points, constant characteristics of the image. These characteristics have to be robust to changes in the scale, noise and illumination of images to make detector repeatable, i. e. for detector to find keypoints with the same elements on different images.

The area found by the detector is represented by circle and is characterized by three parameters: coordinates of center, radius and angle of rotation.

2.1 SIFT (Scale-invariant feature transform)

SIFT is considered to be a reference algorithm since it possesses one of the highest repeatabilities among detectors.

SIFT consists of several stages, on each of which the number of considered areas decreases. Thus operations with big computing complexity are applied only to the areas which have passed preliminary tests.

1) Detection of the characteristic points robust to scaling of the image using scale-space extrema in Difference-of-Gaussians images.
2) Determination of radius of the area including the characteristic found earlier.
3) Determination of keypoint orientation.

2.2 SURF (Speeded Up Robust Feature)

SURF was conceived as more productive replacement to already existing effective detectors, for example, SIFT detector.

The algorithm is in many respects similar to SIFT. However, in SURF are realized rather exact and fast approximate calculations methods. It allows SURF to be 2–3 times faster than SIFT without significant quality loss.

2.3 ORB (Oriented FAST and Rotated BRIEF)

ORB also was designed as faster alternative of SIFT and SURF.

ORB detector is based on FAST detector of corners presented in 2006 by Edward Rosten ([7], [8]). Center of a circle is considered to be a vertex of angle, if not less than ½ and no more than ¾ the pixels lying on a circle are significantly more intensive than central pixel (or vice versa). In practice radius of circle equals 9 or 16 pixels.

ORB is modification of FAST which also calculates orientation (angle of rotation) of each corner. Besides, ORB looks through several scales of the initial image.

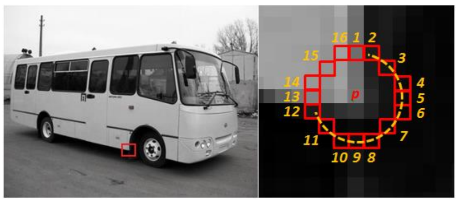

Figure 1: Definition of corners by means of the detector FAST with the radius of a considered circle of 3 pixels. Intensity of the pixels lying on a dashed line are less than central pixel p intensity by a set threshold t, $0 < t < 255$.

3. DESCRIPTORS

Descriptor is a set of numbers characterizing a keypoint on the image. The descriptor is to have diverse values for keypoints containing different objects, and identical values for same objects, and to be robust to changes of lighting, affine and perspective transformations, noise and other image deformations. Thus, the descriptor is to use some invariant characteristics of the image.

3.1 SIFT

SIFT descriptor uses the algorithm based on model of human vision. Intensity gradients of pixels in 16 squares of 4×4 pixels are summed and splitted into 8 bins. To store the sum of gradients in a 4×4 square, we need 8 numbers describing magnitude of gradient in each of 8 directions, so the descriptor is stored by a 16 x 8 = 128-component vector. Big dimension of a vector causes difficulties in search of the descriptor closest to a selected one.

3.2 SURF

SURF descriptor of is based on the same algorithm as SIFT descriptor. However, for reduction of calculations volume SURF descriptor is stored by a 64-component vector. SURF contains some other improvements aimed on reduction of time of comparison of couple of descriptors.

3.3 ORB

ORB descriptor represents modified BRIEF descriptor ([2]).

BRIEF descriptor describes image area by means of a set of binary intensity tests in a 31x31 image patch. 256 couples of pixels chosen by a fixed pattern on a patch are compared by intensity. If first pixel in a couple is brighter than the other, we store 1 to describe the couple, otherwise — 0. Thus we get 256 binary digits describing all the couples.

The descriptor of BRIEF is unstable to image rotations on more than 10°. ORB is its modification robust to rotation of a scene. ORB detector is also called rBrief - BRIEF robust to rotations.

4 BENCHMARK

In this chapter we discuss a benchmark for comparison of detectors and descriptors efficiency: how experiments are made, images used, criterion of correctness of recognition. We will use realizations of detectors and descriptors from OpenCV library.

Benchmark is designed to evaluate efficiency of various detectors and descriptors on image characteristics, such as size. While other benchmarks used to compare recognition efficiency an approximate homography matrix is used (computed on feature matches), we use exact homography matrix as with it we synthetically generate the image of object from the scene image.

4.1 Studied images

For establishment of dependences we will make series of experiments. Each experiment corresponds to a pair of resolutions of object and a scene.

In this work the object is represented by a part of a scene to which homography transformation is applied. Thus, we know the real homography converting object to a scene part.

Note that homography transformation can be applied to describe transform between two images, only if the translation between cameras used to get these images is much less, than distance between the cameras and a scene. In our case we use homography as some approximation of real distortions ([5], [6]).

Examples of objects and scenes are presented in fig. 2.

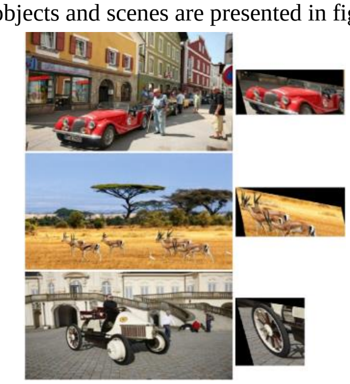

Figure 2. Examples of the images participating in experiments: images on the left — initial scenes, on the right — objects.

Minimum resolution of object is not less than 9000 pixels. 10–18 various images were used for experiments.

4.2 Repeatability

On the first stage we will detect areas on object and a scene. We will evaluate repeatability, the main quality characteristic of the detector — ability to detect same element on two and more various images. For calculation of repeatability we will use the following formula:

$$\text{repeatability} = \frac{\#\text{overlapping keypoints}}{\#\text{keypoints on object}}$$

Since we know how the real transformation for the scene and the object (ground truth homography matrix), we can project keypoints from scene to object and estimate overlap. We consider keypoints from different images to be equivalent if area of their overlap is not less than 50% of area of their union.

4.3 Accuracy

For detected keypoints we compute descriptors, and for each keypoint from object we find a match on a scene with the closest descriptor.

We can estimate a relative pose of corresponding keypoints again. Similarly, we will consider match to be correct if keypoints overlap is not less than 50% of their union. Thus we can determine recognition accuracy:

$$\text{repeatability} = \frac{\#\text{correct matches}}{\#\text{all matches}}$$

5 EXPERIMENTAL RESULTS

In this chapter we will discuss obtained results and we make conclusions on quality of various detectors and descriptors.

5.1 SIFT, SURF

Dependencies of repeatability and accuracy of SIFT and SURF on image resolutions are shown in fig. 3. Each of curves corresponds to a specific test image pair.

It is seen that both repeatability and accuracy of SIFT and SURF have almost no dependency on resolution of the image as corresponding curves are almost horizontal. At low resolutions on there are the some fluctuations caused by small absolute quantity of areas on images. Conducted measurements confirm that SIFT possesses higher repeatability and accuracy than SURF.

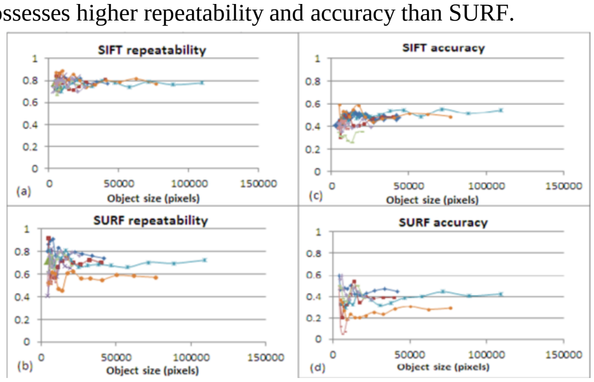

Figure 3. Dependence of repeatability of SIFT (**3a**) and SURF (**3b**) detectors, dependency of accuracy of SIFT (**3c**) and SURF (**3d**) descriptors on resolution of object.

Figure 4. Dependence of absolute number of correct matches on resolution of the image (SIFT).

The conclusion that neither accuracy nor repeatability do not depend on image resolution was not an expected result: we expected to gain direct correlation between them. The more is resolution of the image, the more information it brings to us and to detector; despite that, keypoints detected by SIFT and SURF appear to be scalable, and percent of correct matched doesn't significantly change from scale to scale. Note that the absolute number of found keypoints linearly increases with increase of the image resolution, as shows fig. 4.

5.2 ORB

In ORB implementation in OpenCV we can find method parameters significantly affecting the algorithm quality. First is *nfeatures* – number of keypoints with maximal response than are returned by detector; all the other keypoints are rejected. By default it equals 500. At such nfeatures value it is possible to receive recognition accuracy not higher than 20%.

We found that to achieve maximal quality it is needed to leave all the keypoints in consideration, even with the small response. Dependency of recognition accuracy on resolution for different nfeatures values given in fig. 5. Therefore experiments were made with the nfeatures value of 150000 – such to exceed quantity of the recognizable areas on scenes of the highest resolutions. Note that this, of course, affects computation times, which we don't study in this work.

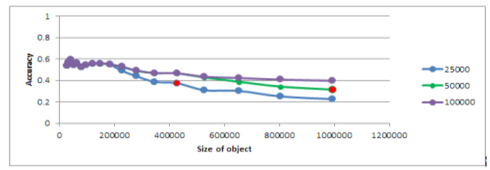

Figure 5. Dependence of accuracy of ORB on resolution of the image for different nfeatures (corresponding to different curves).

Note that at low resolutions of ORB detects very small quantity of areas. That is caused by an analog of FAST detector ORB uses: circle of 16 radius must be applied to detect corner in the center of a circle; on small images only a few such circles may fit the image. It causes poor quality of recognition at the lowest resolutions. However with increase of resolution of object ORB starts showing a good accuracy of recognition. On that reason we run experiments only on images with size of object not less than 9000 pixels.

Dependence of repeatability and accuracy of ORB on resolution of the image are given in fig. 6. All parameters are default, except nfeatures value.

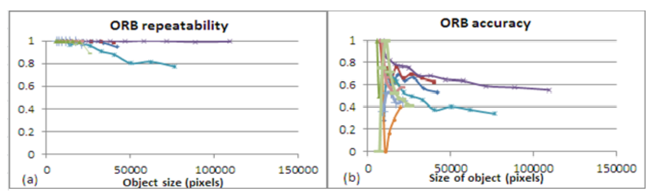

Figure 6. Dependence of repeatability of the ORB detector repeatability (8a) and ORB descriptor accuracy (8b) on resolution of the image.

It is visible that ORB detector (unlike SIFT or SURF) possesses repeatability of almost 100% — highest of repeatabilities of considered detectors. ORB accuracy also seems to be the highest among the studied methods, which is surprising as SIFT was always thought to be the best recognition method.

Almost absolute repeatability of ORB is achieved not only by high ability to detect same objects (corners, as ORB detects corners using oFast), but also by huge amount of found keypoints. Indeed, ORB detects up to 40 times more keypoints than SIFT or SURF. The keypoints are strictly located around the corners on image, but may have different radiuses and slightly different centers and orientations. Usually there are 5–15 keypoints around a single corner.

Besides, an inverse dependency on resolution of the image is observed. There may be a number of root causes for this effect, but the main one is that rBrief descriptor by default uses a fixed patch size of 31x31 pixels to calculate the descriptor. A fixed size of patch on which a descriptor is computed is not a good solution to serve both for big and small images, and thus big and small keypoints.

To avoid it, we can set *patchSize* parameter of ORB to depend on image size: for example, 5% or 10% of object width or height. In this case we get more slight and non-linear inverse dependency on image resolution, as seen on fig. 7.

The value of patchSize giving the best accuracy results is 10% of object width. Inverse dependency on image resolution is still present, but now shows off only on the biggest resolutions.

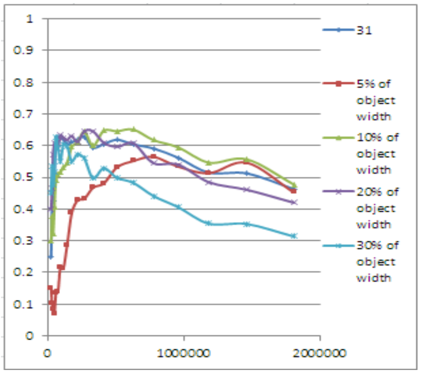

Figure 7. Dependence of accuracy of ORB on resolution of the image for different patchSize values (corresponding different curves).

Other parameters of ORB also can affect quality significantly, like edgeThreshold – size of image border in which no detection is made for noise reduction. To study it, we made experiments with various parameters of ORB detector and descriptor. Results of these experiments are presented in fig. 8.

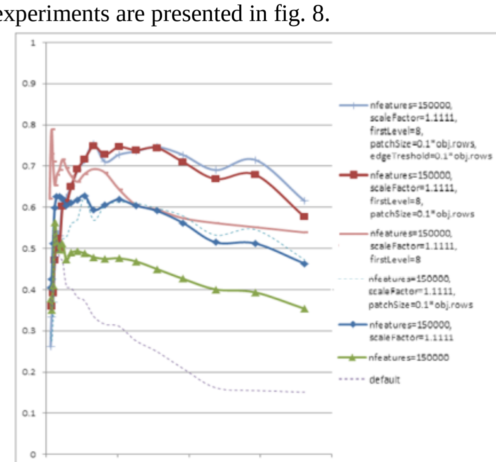

Figure 8. Dependence of accuracy of ORB on resolution for different parameters.

The combination of parameters providing the maximum accuracy:

• nfeatures = 150000 (as big as possible not to reject any keypoints);™

• scaleFactor (coefficient of reduction of height and image width for receiving a pyramid) = 1.1111;

• firstLevel (the first level of a pyramid used for detecting) = 8;

• patchSize (the size of a patch of rBrief) = 10% from object height;

• edgeThreshold = 10% from object height aren't detected.

This combination of parameters allows to receive the accuracy of recognition of 65-70% higher than ORB accuracy when using parameters by default.

6 DISCUSSION

In this work we presented a method to evaluate the dependency of objects recognition accuracy by SIFT, SURF and ORB methods on resolution of the initial image. By means of this method we compared efficiency of various detectors and descriptors of images. We managed to choose the parameters of the ORB method providing the best quality of recognition of the image.

Accuracy and repeatability of SIFT and SURF don't depend on resolution of the studied image. SIFT possesses slightly higher quality of recognition than SURF. Neither SIFT nor SURF show dependency of quality on image size.

ORB seems to outperform both SIFT and SURF. Repeatability and accuracy of ORB have weak inverse dependency on resolution of the image: at high resolutions ORB has 10-15% smaller accuracy, than at low. To investigate the root cause of this effect we experimented with parameters of this method and managed to find optimal parameters that are able to significantly increase quality of object recognition. In that case, 640x480 pixels (300000 pixels) resolution is enough to achieve 65-70% recognition accuracy.

Obviously, the comparison carried out by us isn't full. We used 10-20 images of various categories, however their quantity can be increased for receiving more reliable results. Research of this sort can be applied to various detectors and descriptors, and also to various combinations of detectors and descriptors.

7 REFERENCES

[1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool. "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008.

[2] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features // European Conference on Computer Vision, 2010.

[3] Martin A. Fischler and Robert C. Bolles. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". June 1981, Comm. Of the ACM 24: 381—395.

[4] David G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, 60, 2 (2004), pp. 91-110, 2004.

[5] K. Mikolajczyk, C. Schmid. "A performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10): pp. 1615-1630, 2005.

[6] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. Van Gool. "A comparison of affine region detectors", International Journal of Computer Vision, 65(1/2): pp. 43-72, 2005.

[7] Edward Rosten, Tom Drummond. Fusing points and lines for high performance tracking // IEEE International Conference on Computer Vision, 2005.

[8] Edward Rosten, Tom Drummond. Machine learning for high-speed corner detection // European Conference on Computer Vision, 2006.

[9] Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski. ORB: An Efficient Alternative to SIFT or SURF // International Conference on Computer Vision, Barcelona, Spain, November, 2011.