Methods for Searching Key Points in Images

Mikhaylov A.V., Kovalenko P.O., Kataev A.V., Rozaliev V.L.
Volgograd State Technical University, CAD department
Volgograd sity, Russia
Email: a.mikhaylov.v@gmail.com
Email: p.kovalenko.o@gmail.com
Email: alexander.kataev@singularis-lab.com
Email: vladimir.rozaliev@gmail.com

***Abstract*—The paper presents the interpretation of the concept - "key points", classification approaches to their definition. Described descriptors key points such as SIFT, SURF, BRIEF, ORB. Presents the comparative analysis of the data descriptors on such parameters as time and detection accuracy.**

***Keywords*—Computer vision, key point, descriptor, image, detection.**

I. Introduction

In recent time a lot of attention in the field of computer vision given to the task of visual tracking of objects. It consists in the consecutive determining the location target for each frame of the video stream. This task finds practical applications in many fields, anyway associated with video processing to obtain some information from them.

Examples of systems that use tracking technologies, are analysis of the system of sports matches that have appeared recently a system of augmented reality, security organizations, law enforcement agencies. It is also an object tracking system can be designed for use on quadrocopters, through which it will be possible to observe the object from a height.

Search and comparison of the key points in the images is an important task of computer vision. At first glance, the respective sets of points on the images give very little information about the images and the observed scene, but in fact it is not. If multiple images of the same scene, and a set of corresponding points in these images becomes possible to determine the camera position and settings for each image.

The relevance of this work is justified by the fact that, despite the huge progress in research, visual detection of objects remains a challenge.

II. The concept of key points

Of particular importance in the construction of feature image description is the selection (detection) specific to the parts of the image, as which it is possible to consider, for example, corners, edges, regions corresponding intensity extremums, etc.

The key point of the stage or point feature - it is the image point (pixel) with a characteristic neighborhood, that is, other than their neighborhood from all the neighboring points. Describes the feature vector is calculated based on the intensity / gradients, or other characteristics of the neighborhoods of the points. Using the key points may be analyzed as a whole picture and the objects to them. Good key points allow to cope with the change in scale, perspective and overlay a scene or object.

The key point of such properties as the repeatability (key point is in the same location of the scene or object image, in spite of the changes in viewpoint and illumination), focus (neighborhood points should be big differences from one another), locality (a key point should occupy a small area of the image to be to reduce the likelihood of sensitivity to the geometric and photometric distortion between two images taken at different points of the review), the number (the number of points should be large enough so that they have enough to detect even small objects), accuracy (in terms must be localized exactly as in the original image, and taken on a different scale), efficiency (time of detection points in the image should be admissible in a time-critical applications).

III. Classification of approaches to identify key points

Approaches to determining the key points can be divided into the following categories:

1) based on the intensity of the image: the key points are computed directly from the image pixel intensity values;
2) using the contours of the image: the methods extract contours and looking for a place with a maximum value of curvature or make a polygonal approximation of the contour and define the intersection. These methods are sensitive to the neighborhood of the intersection;
3) based on the use models: models are used with intensity as parameters that are adjusted to the images-patterns to sub-pixel accuracy, depending on the template.

In practice, for a broad application of the most common methods based on image intensity.

IV. Descriptors key points in the image

Descriptor key points is the numeric vector characterizing features of the image in the vicinity of this point.

There are a number of requirements to the descriptors [1]: they must be invariant (the description of the same point on the two different images should be identical), unique (descriptors two different features should be markedly different from each other), stable (for geometric transformations descriptor the same point does not vary greatly). In a real situation the descriptor do not have all of these properties simultaneously and have to choose the one descriptor that best fits the task.

The general scheme of the algorithm for calculating descriptors is shown in Figure 1

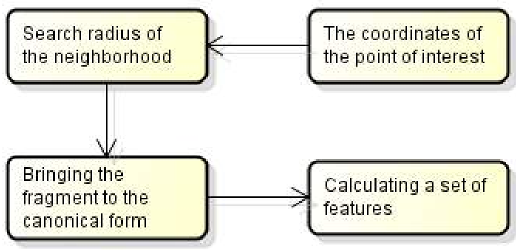

Figure 1. The algorithm for calculation descriptors

One of the most famous descriptors key points is a SIFT (Scale Invariant Feature Transform) [2], which is based on the idea of calculating the histogram of oriented gradients in the vicinity of the key points (figure 2). Neighborhood feature point divided into four square sectors. In every pixel inside each sector gradient image is calculated as well as its direction and module. Then the modulus of the gradient are multiplied by the weight decreases exponentially with distance from the point of interest. For each sector going histogram gradient direction, and each entry is weighted gradient module [3].

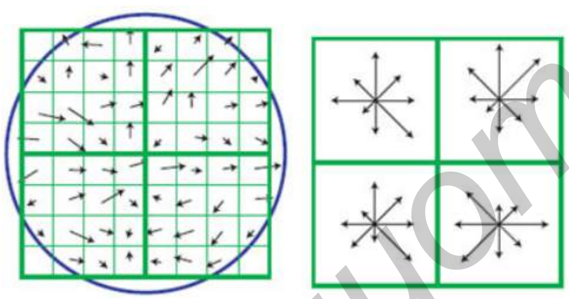

Figure 2. Building a SIFT descriptor

The closest competitor is the descriptor SIFT-SURF-descriptor (Speeded up Robust Features) [4]. SURF similar to its predecessor, but the procedure is to describe the key points of the neighborhood is somewhat different, because it does not use the histogram-weighted gradient, and the responses of the original image on Haar wavelets (figure 3).

The first step in getting descriptors around points of interest builds a square area that is oriented with respect to some preferred direction. Then, the square area is divided into sectors. In each of the sectors in the points belonging to a regular grid, calculated responses to the two types of wavelets - horizontal and vertical directions. The responses are weighted by a Gaussian, summed for each sector, and constitute the first part of the descriptor (figure 4).

The second part consists of sums feedback modules. This is done in order to take into account not only the fact that changes in brightness from one point to another, but also to

Figure 3. Haar wavelets

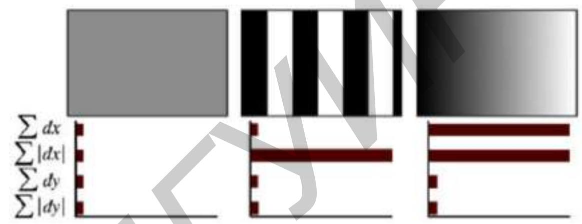

Figure 4. SURF-descriptor components point to the structure of the image. From left to right: a homogeneous region, strongly anisotropic texture, gradient texture

keep information about changes in direction. SURF-descriptor has a length of 64. As with SIFT, SURF-descriptor is invariant to changes in the brightness of the additive.

Among the key points of search engine algorithms there is a division on a math-based, but the relatively slow (Harris detector, the SIFT detector), and a heuristic, but rapid (SURF, FAST). The same division applies to the description of point features, but here as a separating sign stands compact and easy to handle computation. The smaller the descriptor length, the less memory is required to store it, less time and its comparison with others. This feature is very important when processing large numbers of images.

The most compact descriptor refers BRIEF (Binary Robust Independent Elementary Features) [5]. To calculate the descriptor at matching the brightness values of points located in its vicinity. This compares luminance values are not all the points with all but analyzed only a small subset of adjacent pairs of points, the coordinates of which are distributed randomly (but in the same way for each of the analyzed points).

Themselves BRIEF descriptors are not invariant to rotation. However, this invariance can be achieved if the pre-turn fragment around a point of interest on the angle corresponding to, for example, the dominant direction of the luminance gradient, as is done for SIFT and SURF descriptor.

The purpose of the descriptor BRIEF [6] was to provide recognition sites identical images, which were obtained from different viewing angles. The objective is to reduce the number of calculations performed. A more effective alternative is a binary descriptor BRIEF is descriptor ORB (Oriented FAST and Rotated BRIEF) [6] - an improved version of the detector and the combination of points and FAST BRIEF binary descriptors.

SIFT algorithm shows excellent results in problems of detection, identification and localization of objects in the image, but as is evident from the present disclosure requires significant computation on volumes. In this connection, efforts continue to create simpler algorithms points detection and calculation of descriptors, thus providing sufficient invariance of the distortion. One of them is considered in this paper, the above-mentioned ORB.

In the first step for the detection of large-scale construction of the singular points of the Gaussian image pyramid. Then extremes of brightness functions are defined at each scale level. To solve this problem, use the FAST algorithm, according to which each point of the image is formed by a circle of radius, and count the number of points lying on the circle and having a brightness value less than or greater than the luminance of its center.

The structure of the ORB algorithm shows that it is less demanding of computing resources in comparison with the algorithm described earlier SIFT. The gain in speed of calculation is determined, above all, a simpler procedure for constructing descriptors and the mechanism of calculating the norm.

V. Comparison key points

There is an approach associated with the search for the image of the marker on a picture on his key points. This process can be divided into four stages:

1) highlight key points using the detector;
2) description points found using a descriptor;
3) getting a set of relations (correspondence) between the key points using matcher;
4) analysis of the connections to determine whether location marker in the image.

The implementation of the first three points is not difficult because of the presence of a broad class of computer vision library. Among them stands out OpenCV library (which we will use in the future), is written in C ++ and has the most extensive feature set. First of all, it contains a module Feature Points Detection and Description, contains the implementation of the various detectors and descriptors matchers. In view of this, the first three points of the search marker in the image are resolved simply and without any serious difficulties.

The main difficulty is in the fourth step, as there is some special features, which could confidently say - is the desired object in the image or not.

Famous matching algorithm between the key points of the two images often show a high percentage of false links (figure 5). This leads to the fact that they can not correctly match the tested image even though they are actually the same, only removed from several different angles.

Sets the corresponding points in the images give very little information about the images and the observed scene. For example, if we have multiple images of the same scene and sets the corresponding points in these images we can determine the configuration and position of the camera for each image. This requires to match each image key point.

An important property of any features matching algorithm is a set of image distortion key point with which it is able to cope. Typically, the following types of distortion:

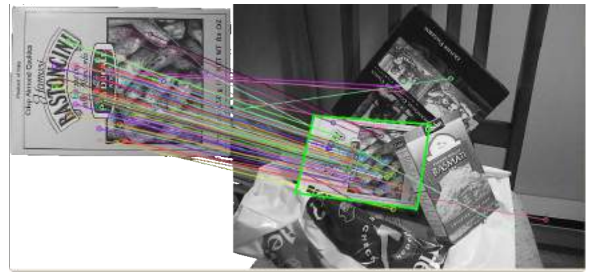

Figure 5. Comparison of key points on the two images

1) changes in the light key point;
2) zooming;
3) rotation of the key point;
4) projective image distortion characteristics associated with such rotation and movement of the camera in space.

VI. Description of research methodology

OpenCV library has a fairly wide range of detectors, descriptors and matchers. At the same time there are possibilities of different combinations with each other. They all differ in speed, the number of allocated points, as well as resistance to image transformations: rotations, shifts angles, changes of scale.

As a criteria comparison were selected: settling time key points in the image (in milliseconds), accuracy (percentage) defined as a percentage of the right set points to their total amount.

The research was conducted in a Microsoft Visual Studio 2012 Ultimate with the help of OpenCV 3.0 library.

To determine the accuracy of the establishment of the key points of the two images were selected with a slightly shifted scene.

To start searches for the key points and the calculation of descriptors (SIFT / SURF / BRISK / ORB) on both images, producing a description of found points, evaluating their position through the description of the neighborhoods.

Then the comparison of them, in other words, the construction of correspondences between two sets of image points.

To improve accuracy, the screening was carried out points that were obviously wrong compared. This is done as follows: were calculated minimum and maximum distance between the points, after comparing them. Then select only those points, the distance between them was less than half the sum of the maximum and minimum distance between them.

The result of the image will be built with matching lines between key points of two images.

VII. The research results

A result of research has been established settling time key points in the image and accuracy:

1) Descriptor SIFT load of 417 ms with an accuracy of 50 per cent (figure 6);

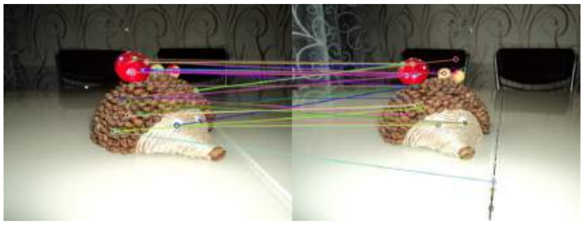

Figure 6. The result of the SIFT descriptor

2) Descriptor SURF load of 226 ms with an accuracy of 55 per cent (figure 7);

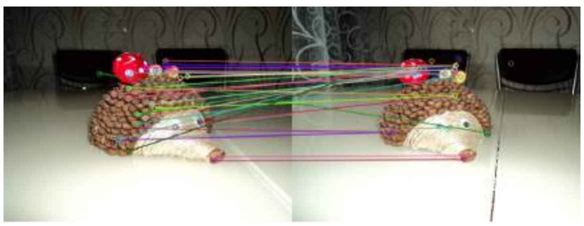

Figure 7. The result of the SURF descriptor

3) Descriptor BRISK load of 126 ms with an accuracy of 31 per cent (figure 8);

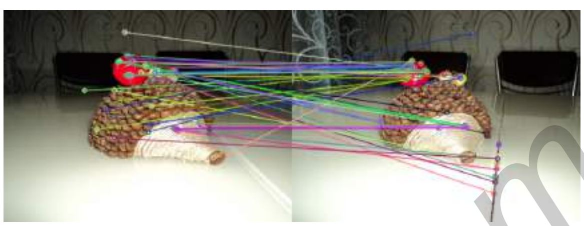

Figure 8. The result of the BRISK descriptor

4) Descriptor ORB load of 154 ms with an accuracy of 61 percent (figure 9).

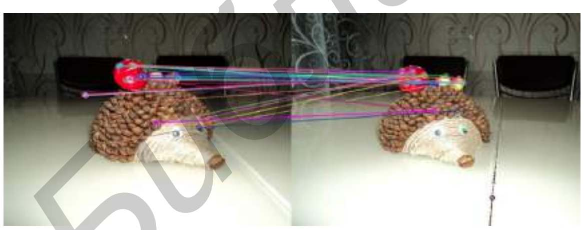

Figure 9. The result of the ORB descriptor

To establish the object in the image is best suited ORB descriptor, which showed low enough settling time points in the image (154 milliseconds) and high accuracy of establishing (61 percent) as compared with other descriptors. The lowest accuracy identify key points has a descriptor BRISK (31 percent). At the time of establishing the key points of the least expensive proved descriptor BRISK (126 milliseconds), and the longest - SIFT a result of 417 milliseconds.

VIII. Conclusion

Based on the analysis of literature related to the recognition scene in the video stream, it can be concluded that to effectively search for the object in the sequence of frames on a sample can be applied algorithms based on the calculation descriptors key points of the image.

The results of the work has been have explained to identify the key points on the image, a classification of approaches to their definition, are investigated descriptors key points as SIFT, SURF, BRIEF, ORB; identified and analyzed the statistics of their work.

Local features are a popular tool for image description nowadays. They are the standard representation for wide baseline matching and object recognition, both for specific objects as well as for category-level schemes. In this survey, we gave an overview of some of the most widely used detectors, with a qualitative evaluation of their respective strengths and weaknesses, which can be found at the end of the sections and chapters. Discussion and Conclusion that can inspire future research on local features and avoid a waste of resources by reinventing the wheel. The literature is huge, and we could only touch the different contributions without going into details. Yet, we hope to provide the right pointers so those who are interested have a starting point and can delve deeper if they want to.

The research work carried out as part of the grant to the winner of the program "UMNIK MIPT-autumn 2015" with the project "Development of methods for visual search and track targets using quadrocopters".

References

[1] Zitova B., Flusser J. Image registration methods: a survey // Image and Vision Computing. – 2003. – V. 21. – P. 977–1000

[2] Lowe D. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. 2004. N 60. P. 91–110

[3] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. – Springer, 2008

[4] Bay H., Ess A., Tuytelaars T., Gool L.V. SURF: speed up robust features // Computer Vision and Image Understanding (CVIU). 2008. V.110, N 3. P. 346-359

[5] Calonder M., Lepetit V., Strecha C., Fua P. BRIEF: Binary Robust Independent Elementary Features // Proc. European Conference on Computer Vision. – 2010. – P. 778–792

[6] Calonder M., Lepetit V., Strecha C., Fua P. BRIEF: Binary Robust Independent Elementary Features // In Proceedings of the 11th European Conference on Computer Vision (ECCV'10). 2010

МЕТОДЫ ПОИСКА КЛЮЧЕВЫХ ТОЧЕК НА ИЗОБРАЖЕНИЯХ

Михайлов А.В., Коваленко П.О., Катаев А.В., Розалиев В.Л.

В работе рассматривается понятие «ключевые точки» и их применение для анализа видеоизображений. Приведена классификация подходов к их определению. Описаны дескрипторы ключевых точек, такие как SIFT, SURF, BRIEF, ORB. Описана методика проводимых исследований. Представлен сравнительный анализ работы данных дескрипторов по таким параметрам как время работы и точность детектирования.