International Journal of Imaging and Robotics,
[Formerly known as the "International Journal of Imaging" (ISSN 0974-0627)]
Volume 18; Issue No. 2; Year 2018; Int. J. Imag. Robot.
ISSN 2231–525X; Copyright © 2018 [International Journal of Imaging and Robotics]

Object detection in images from the perspective of the active perception theory

Andrey A. Filyakov[1], Vasily E. Gai[1], Alla A. Koshurina[2], Maxim S. Krasheninnikov[2], Maxim O. Derbasov[1], Igor V. Polyakov[1] and Dmitry I. Zhurba[1]

[1]Institute of Radioelectronics and Information Technologies
[2]Institute of Transport Systems
Nizhny Novgorod State Technical University n.a. R.E.Alekseev
K. Minina, 24, Nizhny Novgorod, Russian Federation, 603950
andreyfilyakov@gmail.com
iamuser@inbox.ru

ABSTRACT

Object detection in images is a highly topical problem of theoretical computer science. The effectiveness of object detection depends on the chosen method of presenting the image, i.e. on its feature description. This article describes a new method of feature descriptor construction based on the active perception theory (APT). The results of a computational experiment for evaluating the accuracy of object localization under different conditions are presented.

Keywords: object detection, active perception theory, pattern recognition.

2000 Mathematics Subject Classification: 68T10, 68U10.

ACM Computing Classification System: C.3, I.4.0, I.5.4.

1 Introduction

Problem of object detection in images is one of the computer vision problems. Known solutions of this problem may be decomposed into three stages (see fig. 1):

- Pre-processing of an image;
- Construction of a feature descriptor;
- Object detection in an image (decision-making).

The pre-processing stage is presenting an image in a form suitable for further construction of its feature descriptor. The part of this stage is filtration or noise removal [14].
Feature descriptor construction creates a set of features that allows to unambiguously describe the object. The crucial part of this stage is the detection of key points that are distinguishing for this object. Representing the object as a set of key points reduces the size of a feature descriptor. There are a lot of methods that can be used at this stage, including:

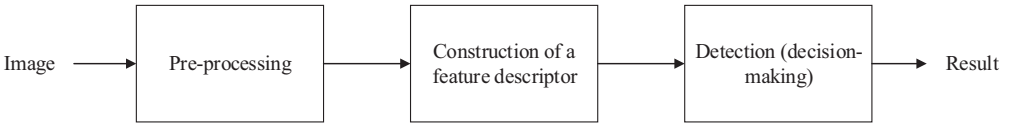

Figure 1: The process of object detection in images

- Scale-invariant feature transform (SIFT) [9] [11];
- Speeded up robust features (SURF) [1] [12];
- Binary robust independent elementary features (BRIEF) [2].

Feature descriptor obtained using these methods is a vector of real numbers. All those methods have one common disadvantage i.e. the big size of feature descriptor (a vector can be up to 256 elements in length). Object detection is finding the object where the key points are concentrated in the processed image. This stage can utilize the following methods:

- Region merging segmentation algorithm [9];
- Hierarchical pyramid scheme [12];
- Randomized trees classifier [8];
- *K*-nearest classifier [7];
- Naive Bayes classifier [10];
- Neural networks [13].

This article describes a method of object detection in images which uses the active perception theory (APT) [14]. APT is used in the stages of image pre-processing and feature descriptor construction, while the stage of decision-making utilizes the clustering methods. If compared with known approaches to the construction of an object feature descriptor, APT exhibits the following characteristics:

- Unlike the methods that use training, APT calculates feature descriptor using predefined templates;
- The calculation of *U*-transformation uses only addition and subtraction operations.

We suggest that detection of obstructions for autonomous vehicles operating in difficult climatic conditions could be one of the possible applications of the developed method.

2 Information model of object detection on image

The process of object detection in images is described in fig. 2. in a form of entity-relationship model. Input data for this problem is a digital image *I* of the size *N x M* pixels.

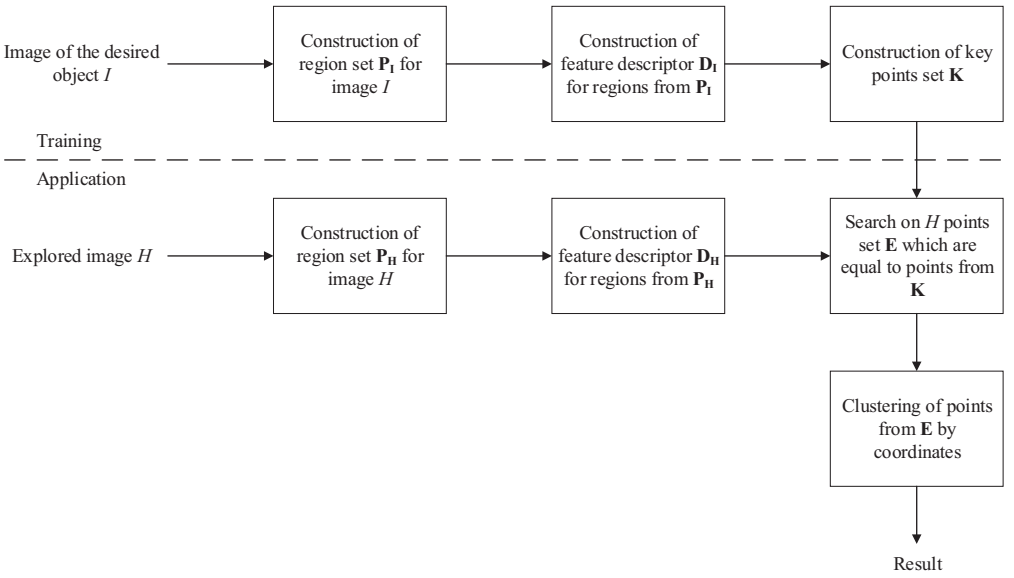

Figure 2: Entity-relationship model of object detection in images process

2.1 Pre-processing of an image

Pre-processing of an image is its processing using a sliding window of the size n n with incremental intervals (steps) horizontally and vertically. As a result, the image is presented as a set of regions:

$$\mathbf{P_I} = \{P_i, i = \overline{1, Q}\},$$

where P_i is a region of the image, Q is count of regions in the image I.
The algorithm for constructing those regions of set $\mathbf{P_I}$ is the following:

$\forall x = \overline{1, N - n}$
  $\forall y = \overline{1, M - n}$
    $P_i = I[x + n, y + n]$
    $i = i + 1$
    $y = y + s_h$
  $x = x + s_w$,

where *N x M* is the size of the image, *n x n* is the size of the sliding window, s_h is vertical step, s_w is horizontal step, *I* is processed image.
Under the conditions of prior uncertainty, horizontal and vertical steps are taken equal $s_w = s_h$.

2.2 Construction of an object feature descriptor

The construction of the feature descriptor of the object starts with creating descriptors for all the image regions of this object. Spectral representation coefficients of the image obtained as

a result of *U*-transformation are used as a descriptor for each region.

$$D_i = U[P_i] = \{d_k, k = \overline{1, L}\},$$

where *U* is operator performing *U*-transformation [14], P_i is image region for which a descriptor is being built, d_k is *k* spectral representation coefficient for the image region, *L* is count of filters used for *U*-transformation (*L* = 16).

Image feature descriptor of the whole object is represented as a set of all its regions feature descriptors:

$$\mathbf{D_I} = \{D_i, i = \overline{1, Q}\}.$$

The next step in constructing a feature descriptor is the detection of its key points. The point of an object is considered key if the region where this point is located contains a stark difference in brightness. We assume that differences in brightness correspond to the contours of the object.

The selection of regions that contain key points starts with calculating a standard deviation of its feature descriptors spectral coefficients for each image region:

$$s_i = \sigma[D_i] = \sqrt{\frac{1}{L-1}\sum_{k=1}^{L-1}(d_k - \overline{d})^2},$$

where σ is operator that calculates standard deviation, *L* is count of descriptors spectral representation coefficients (*L* = 16); d_k is *k* spectral representation coefficient; $\overline{d}$ is the mean of the region non-zero spectral representation coefficients.

We propose the following criterion to determine whether the image region P_i contains a key point of the desired object. It will help us build the set of key points descriptors **K**:

$$\mathbf{S} = \{S_i, i = \overline{1, Q}\},$$

$$\mathbf{K} = \varnothing,$$

$$\mathbf{K} = \begin{cases} \mathbf{K} \cup D_i, \text{ if } s_i > k \text{ x } max(\mathbf{S}), \\ \mathbf{K}, \text{ else.} \end{cases}$$

where *k* is a threshold value.

Thus, the image region contains a key point if the standard deviation of this regions spectral representation coefficients is greater than the maximum standard deviation for all regions of the image multiplied by threshold *k*. Here are examples of key points regions found using different values of threshold *k* (*N* count of detected key points) on the fig. 3.

2.3 Object detection

On this stage, we compare the point descriptors set of the processed image (set $\mathbf{D_H}$) and regions with the key points descriptors (set **K**). Then, we detect regions of $\mathbf{D_H}$ that are equal to regions from **K** in terms of their feature descriptor. Then the set **E** is built from such regions.

$k = 0.6$
$N = 517$

$k = 0.7$
$N = 351$

$k = 0.8$
$N = 173$

Figure 3: Examples of detection regions which contain key points

Let two points P_1 and P_2 have feature descriptors D_1 and D_2. Then the equivalence criterion of points P_1 and P_2 has the following form:

$$P_1 = P_2, \text{ if } sign(D_{1k}) = sign(D_{2k}), k = \overline{1, L},$$

where *sign(x)* is the operator that calculates the sign of expression *x*, D_{1k}, D_{2k} are elements of spectral representation coefficients vector of points P_1 and P_1, *L* is count of spectral coefficients in points feature descriptors.
So, two image points are considered as equal if and only if the relevant spectral coefficients of their feature descriptors have the same sign. In this case the algorithm constructing the set **E** is the following:

$$\mathbf{E} = \varnothing,$$

$$\mathbf{E} = \begin{cases} \mathbf{E} \cup D_{Hi}, \text{ if } D_{Hi} \text{ is equal } K_j, \\ \mathbf{E}, \text{ else.} \end{cases}$$

After finding the points that are equal to the object key we can attempt to find an object where those points are concentrated. Suggested method is based on classifying points into clusters according to the coordinate principle. We assume that the cluster that contains the largest count of points is located in the image in the same place with the desired object.
There are different known methods used for clusterization [6]. These methods may give different results with the same input data. Comparing the quality of the cluster solution is difficult because there is no universal criterion of clustering quality.
This work uses the following clustering methods:

- Method based on clustering algorithm *k*-means [5];
- Clustering method Mean shift [4] [3].

3 Computational experiment

For the computational experiment, the set of 100 different objects was prepared. The set contains 4 templates for each object with a different signal-noise relation (SNR) and 100 images for the object detection. Thus, the whole set contained 10000 images. The average image size was 1000 1000 pixels.

Tabl. 1 contains the results of developed method testing in normal conditions (without noise) for different values of the input parameters. As it can be seen from the tabl. 1, the proposed

Table 1: Algorithm testing under normal conditions

Treshold k	Region side length n, pixel	Offset s_h, s_w, pixel	Clustering method	Detection accuracy, %
0.6	16	4	K-means	96.5
0.6	32	4	K-means	96.5
0.6	32	4	Mean shift	96.5
0.65	16	4	K-means	92.0
0.65	16	4	Mean shift	92.0
0.65	32	4	K-means	94.0
0.65	32	4	Mean shift	95.0
0.7	16	4	Mean shift	85.0
0.7	32	4	K-means	89.0
0.7	32	4	Mean shift	89.0

algorithm exhibited the highest object detection accuracy (96,5 %) for three combinations of input parameters. Those parameters were used to test the algorithm with noise.

Considering that the image set contains images that are different from the template in scale and rotation angle we can conclude that the developed algorithm is resistant to such changes in the object. It is due to the fact that the proposed algorithm doesnt take into account relative positions of the key points on the object, but looks for the region with the greatest concentration of points that are equal to key points. Object template contains the same count of key points regardless of the rotation angle, thats why the object can be detected on the image with the same accuracy in the cases of different rotation angles. The similar statement is valid for templates which have different scale excluding the cases where the object size on the image is so small that enough of the key points cannot be detected.

Tabl. 2 contains the results of the algorithm testing under the conditions of noise. As can be seen from the tabl. 2, the proposed method is resistant to noise. The best results were obtained using region side length n = 32 pixels.

Tabl. 3 contains the object detection accuracy that can be obtained by using other known methods. Comparing the data from the tabl. 3 and obtained results of algorithm testing in different conditions (see tabl. 1 and tabl. 2), we can conclude that the developed method has the object detection accuracy comparable with the accuracy of the other known methods. Compared to

Table 2: The results of algorithm testing in the conditions of noise

Treshold k	Region side length n, pixel	Offset s_h, s_w, pixel	Clustering method	SNR, Db	Detection accuracy, %
0.6	16	4	Mean shift	0	54.5
0.6	16	4	Mean shift	10	87.5
0.6	16	4	Mean shift	20	96.5
0.6	32	4	K-means	0	74.0
0.6	32	4	K-means	10	92.0
0.6	32	4	K-means	20	94.0
0.6	32	4	Mean shift	0	79.0
0.6	32	4	Mean shift	10	90.5
0.6	32	4	Mean shift	20	95.0

Table 3: The results of other object detection methods

Construction of object feature descriptor	Making the decision of the object detection	Detection accuracy, %
Scale-invariant feature transform (SIFT)	Region merging segmentation algorithm	83.33 - 91.66
Speeded up robust features (SURF)	Hierarchical pyramid scheme	95
Binary robust independent elementary features (BRIEF)	Randomized trees classifier	94 - 98

some known methods the accuracy of suggested method is even higher.

Tabl. 4 compares the object feature descriptor construction methods by time and by size [15]. It is difficult to compare quantitatively the time of one object feature descriptor construction which was obtained using different methods because the testing of these methods was done on the computers with different configurations. But in the case of qualitative comparison the time of developed method is comparable to its analogues.

Developed method has the size of one region feature descriptor which is equal to descriptor size of BRIEF method and is much smaller than descriptor size of methods SIFT and SURF.

Fig. 4 represents the output of the program which implements the developed object detection algorithm (solid lines are input objects locations, dotted lines are the results of object detection).

Table 4: Comparison of the feature descriptor methods

Method	SIFT	SURF	BRIEF	Developed method
Descriptor construction time, ms	0.93	1.72	0.5	1.4
Descriptor size, byte	512	256	64	64

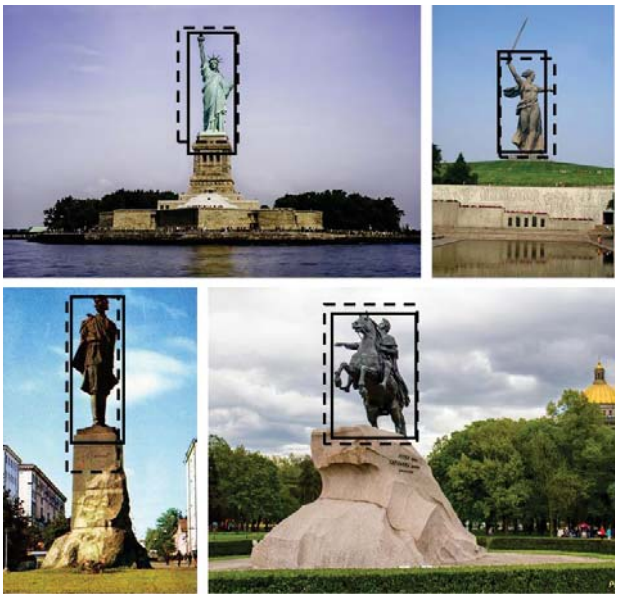

Figure 4: The results of work

4 Conclusion

In this article, we describe a new method of object detection in images. This method uses a new combination of approaches to image localization on different stages. The stage of the object feature descriptor construction uses the algorithms which are based on the active perception theory (APT). The stage of decision-making utilizes methods based on the object key points clustering by k-means and Mean shift algorithms.

The developed method was tested by a computational experiment. Obtained experiment results indicate the correct work of suggested method. The object detection accuracy is comparable with the accuracy of previously known methods, and is even superior in some cases. The developed method was tested in conditions of noise with different values of SNR. The results in conditions of noise showed that the method is stable to such kind of impact. These facts

mean that the method is quite competitive and usable in practise.
Practical application of the proposed method is the analysis of images obtained from external sensors of the vehicle which is moving in autopilot mode in harsh climatic conditions of Arctic and Antarctic. The algorithm can be used to detect possible obstructions and adjust the trajectory to avoid them.

References

[1] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Computer vision–ECCV 2006*, pages 404–417, 2006.

[2] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. *Computer Vision–ECCV 2010*, pages 778–792, 2010.

[3] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.

[4] Keinosuke Fukunaga and Larry Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40, 1975.

[5] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[6] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[7] Daniel T Larose. K-nearest neighbor algorithm. *Discovering Knowledge in Data: An Introduction to Data Mining*, pages 90–106, 2005.

[8] Vincent Lepetit and Pascal Fua. Keypoint recognition using randomized trees. *IEEE transactions on pattern analysis and machine intelligence*, 28(9):1465–1479, 2006.

[9] Reza Oji and Farshad Tajeripour. Full object boundary detection by applying scale invariant features in a region merging segmentation algorithm. *International Journal of Artificial Intelligence and Applications*, 3(5):41, 2012.

[10] Mustafa Ozuysal, Michael Calonder, Vincent Lepetit, and Pascal Fua. Fast keypoint recognition using random ferns. *IEEE transactions on pattern analysis and machine intelligence*, 32(3):448–461, 2010.

[11] Jeong-Tak Ryu and Donghwoon Kwon. An analysis of the surveillance image monitoring system using multi-image stitching. *International Journal of Imaging and Robotics*, 17(3):31–40, 2017.

[12] Drew Schmitt and Nicholas McCoy. Object classification and localization using surf descriptors. *CS*, 229:1–5, 2011.

[13] Christian Szegedy, Alexander Toshev, and Dumitru Erhan. Deep neural networks for object detection. In *Advances in neural information processing systems*, pages 2553–2561, 2013.

[14] Vladimir Aleksandrovich Utrobin. Physical interpretation of the elements of image algebra. *Physics-Uspekhi*, 47(10):1017–1032, 2004.

[15] Jian Wu, Zhiming Cui, Victor S Sheng, Pengpeng Zhao, Dongliang Su, and Shengrong Gong. A comparative study of sift and its variants. *Measurement Science Review*, 13(3):122–131, 2013.

Acknowledgment

The work was carried out at the NNSTU named after R. E. Alekseev, with the financial support of the Ministry of Education and Science of the Russian Federation under the agreement 14.577.21.0222 of 03.10.2016. Identification number of the project: RFMEFI57716X0222. Theme: ”Creation of an experimental sample of an amphibious autonomous transport and technological complex with an intelligent control and navigation system for year-round exploration and drilling operations on the Arctic shelf.”