

# Detection of mine-like objects in high-resolution aerial images

Andrey V. Zhuravlev\*<sup>a</sup>, Alexander S. Bugaev<sup>b</sup>, Sergey I. Ivashov<sup>a</sup>, Vladimir V. Razevig<sup>a</sup>

<sup>a</sup>Remote Sensing Laboratory, Bauman Moscow State Technical University,  
5, 2<sup>nd</sup> Baumanskaya str., Moscow, 105005, Russia;

<sup>b</sup>Dept. of Physical and Quantum Electronics, Moscow Inst. of Physics and Technology,  
9, Institutskii per., Dolgoprudny, Moscow Region, 141700, Russia

## ABSTRACT

A system for mine detection in aerial images is considered as an interactive system in which the operator is responsible for making iterative queries to the database of images and analyzing the results. Preliminary, each image undergoes formal decomposition into a set of feature vectors. Each feature vector is calculated for every irregularity found at a scale-invariant salient point detection stage where a blob detector is used. Assuming that every small object in the image can be described by a single invariant feature vector calculated on a patch around the salient point, formalization of search algorithms is feasible. While the template-based search is straightforward in the terms of one object - one feature vector, we focus on another important option when searching for mines – similar object search. For similar object search a hierarchical clustering algorithm is considered. The mentioned steps of image processing as well as similar object search are illustrated by performing on aerial mine field images taken by an electronic camera from a height of 27 meters. Encouraging preliminary results lead to the formulated plan for future research. The developed algorithms are planned to be used in an image search engine allowing the operator to interactively search for mines in a database of aerial images in humanitarian de-mining operations.

**Keywords:** Airborne mine detection, minefield detection, salient points, feature vector, hierarchical clustering

## 1. INTRODUCTION

Using aerial images in reconnaissance of minefields gives ability to observe large areas and imposes no hazard for the equipment and personnel when compared to contact methods with the help of a radar or metal detector. The main problem when dealing with aerial images in the context of detecting mines is the performance of computer algorithms assisting the operator. The ultimate goal of such algorithms is to facilitate the search in the database of aerial images for potential mines [1, 2].

Most of the algorithms for mines and mine-like object search can be roughly described by the scheme in Fig. 1. The input image is passed to a salient point detection algorithm such as an anomaly detection algorithm, blob detection, etc. At this stage a set of interesting points is extracted from the input image reducing the amount of information passed to the next stages of image processing. Among the requirements imposed on this stage is high probability to respond to objects of interest (mines) which means that each mine, ideally, will not be missed and passed to the next stage of image processing. The second desirable requirement is to reduce the amount of interesting points falling into background or belonging to irrelevant objects. The first requirement is more important as it may limit the performance of the system. Without prior information on imaging conditions it is good to extract salient points in scale- or affine-invariant manner.

On the next stage feature vector is calculated on a patch of the image around each salient point enabling them to be compared. There exist many ways to construct a feature vector. It is desirable that feature vectors should be stable under varying imaging conditions: illumination, viewpoint direction, object rotation, scale or affine transform, etc.

In the task of detecting mines in aerial images two approaches were suggested, viz. a template base search or similar object search. The former may perform when some templates exist from previous interactive search through the database of images.

\*azhuravlev@rslab.ru; phone +7 495 632-2219; fax +7 495 632-2219; www.english.rslab.ru

When searching by a template the quality of a search depends on how stable objects of the same type in different images obtained allegedly under different imaging conditions, varying point of view, etc. The task to deal under such circumstances is laid upon feature vector representation. Object model depicted in Fig. 1 is supposed to encompass probabilistic properties of feature vectors representing same-class objects on the basis of previous queries to the image database and user input.

Without having a template of interesting object there is one visible option to be exploited, viz. to search for a compact class of feature vectors representing candidates for mines. On the final stage the results given by a template search or similar object search are viewed by the user who corrects the input query and is responsible for providing optional input in the template database and object model. In addition to these actions the user can exclude irrelevant objects from subsequent search queries. The search for mines is repeated in the loop in which the user provides input after analyzing results at the end of each search query. Ultimately, a report is generated summarizing the results of a series of search queries.

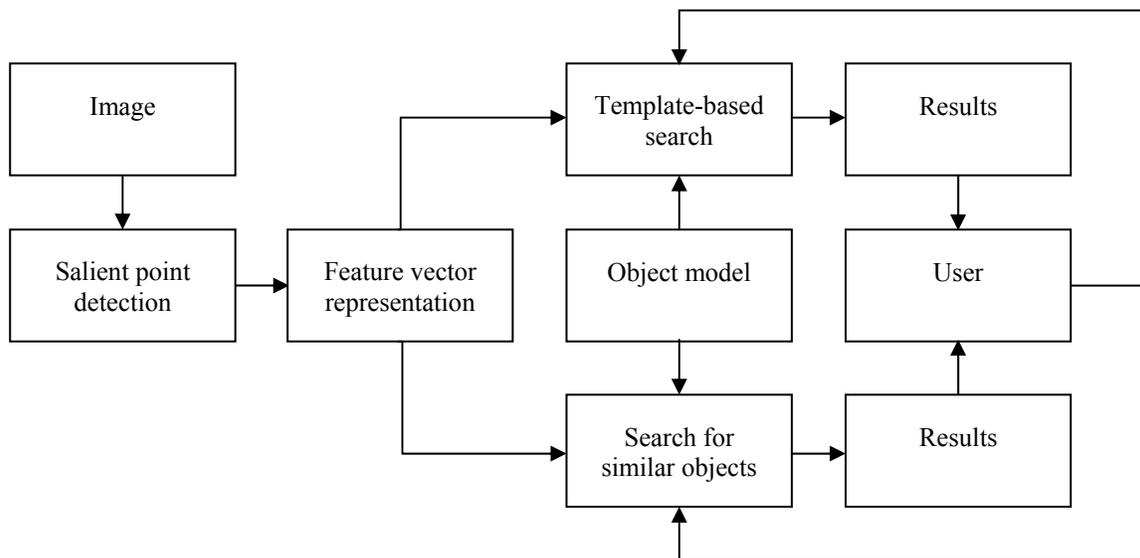


Fig. 1. Block diagram of a potential system for detecting mines in aerial images.

The main objective of this paper is to point out to the methods that can be used on each considered stage and which comply with mentioned requirements imposed on each stage. These methods are used in modern pattern recognition algorithms were they give state-of-the-art results. The rest of the paper is organized as follows: the second section describes salient point detector, for which we used the blob detector proposed in the paper by T. Lindeberg [3], in the third section we use the SIFT descriptor proposed by D. Lowe [4] to produce feature vector representation, the fourth section illustrates application of a hierarchical clustering algorithm to search for similar objects, and the last section summarizes the results and outlines directions for future research. Examples of algorithm performance are showed in each section on the nadir images of mines obtained by an electronic camera from a funicular railway at a height of 27 meters.

## 2. SALIENT POINT DETECTION

At this stage of image processing we follow the steps described in [3]. Initial image is successfully convolved with a

Gaussian kernel  $g(x_1, x_2; t) = \frac{1}{2\pi t} \exp\left(-\frac{x_1^2 + x_2^2}{2t}\right)$  where parameter  $t$  is the scale parameter. Producing for a

single image a series of images for a set of  $t$  in succession gives a scale-space representation for an image making salient points extraction scale-invariant. The scale space representation for an image can be expressed by the following formula:

$$L_k(x_1, x_2) \equiv L(x_1, x_2; t_k) = I(\cdot) * g(\cdot; t_k), k = 0, 1, \dots, N-1. \quad (1)$$

In Equation (1)  $I(x, y)$  is the initial image,  $g(\cdot; t_k)$  is a Gaussian kernel,  $L(x, y; t_k)$  is the scale space representation for the image, and  $x_1, x_2$  are spatial coordinates.

At the next step, a combination of spatial derivatives  $\partial_{x_i}$  is applied to each scale level of the image. This combination of derivatives (possibly non-linear) is chosen to have maximum response to a desired discontinuity type in the image. To detect irregularities looking like blobs in an image the Laplace operator is used. To automatically detect salient points in the image, spatial derivatives are normalized by multiplying by scale factor  $t$  in a power of  $\gamma$ , with the magnitude of the latter being defined experimentally to achieve desired performance. At the final step within obtained volume of normalized derivatives, local maxima and minima of their magnitude are sought to mark salient points in an image with associated scale parameter  $t_k$  of a layer where this extremum was found. Parameter  $t_k$  is related to the size of visible spot on an image.

An example of image fragments illustrating the outlined scheme is shown in Figures 2-3. In Figure 2 a round mine is shown in the center. The center of the mine is marked by a white arrow. The absolute value of scale space normalized derivatives calculated at this point is shown in the same figure on the right as a function of  $\sigma = \sqrt{t}$ .

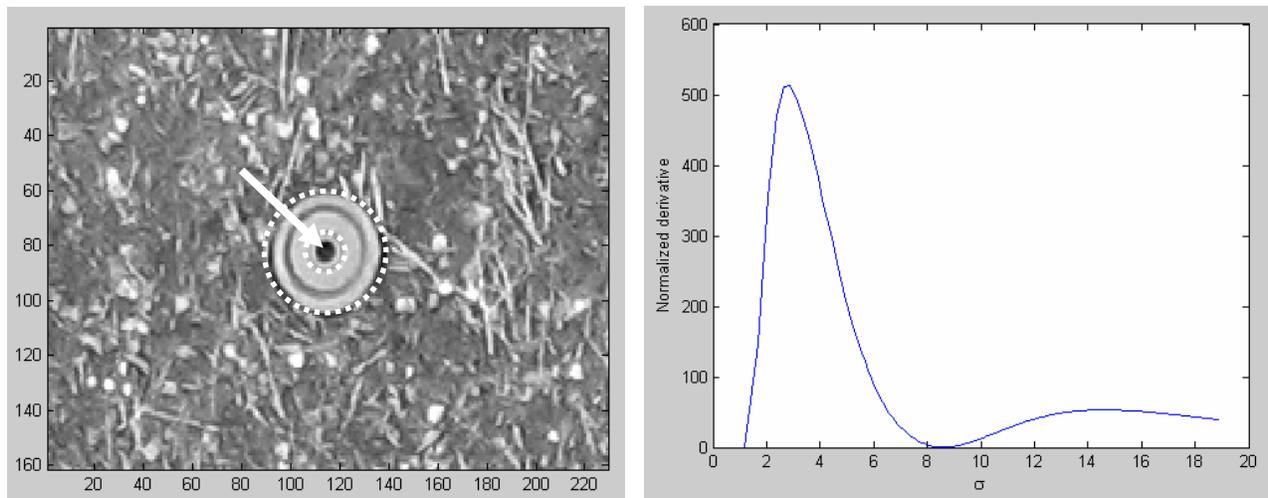


Fig. 2. Detection of scale space extrema of normalized derivatives. A round mine (left) and the absolute value of normalized derivative calculated at the center of the mine as a function of  $\sigma$  related to the scale parameter as  $\sigma = \sqrt{t}$  (right).

As it can be seen in the right part of Figure 2, the absolute value of normalized derivatives achieves two local maxima: the strongest maximum at  $\sigma = 3.5$  and the minor local maximum at  $\sigma = 14.5$ . The strongest maximum corresponds to the black speckle at the center of the mine and the minor local maximum corresponds to the whole mine (these are marked as dotted circles). Finding such extrema in the scale space allows detecting all speckles in an image. Dominant speckles in the image can be picked out by setting up a threshold on the absolute value of normalized derivatives.

A result of salient point detection algorithm is shown in Figure 3. The output data set is a set of blobs, each of them is characterized by size and strength (normalized derivative value). Visual assessment of the result on the right side of Figure 3 allows claiming that each visually perceived circle is attended and extracted by the algorithm with its size (shown by the diameter of a circle) and strength (not shown). Extracted by such a procedure salient points are fed to the next stage of image processing where a patch around each salient point is encoded into an invariant multi-dimensional vector to be compared at further steps. Such multi-dimensional vector is referred to as a feature vector or local descriptor.

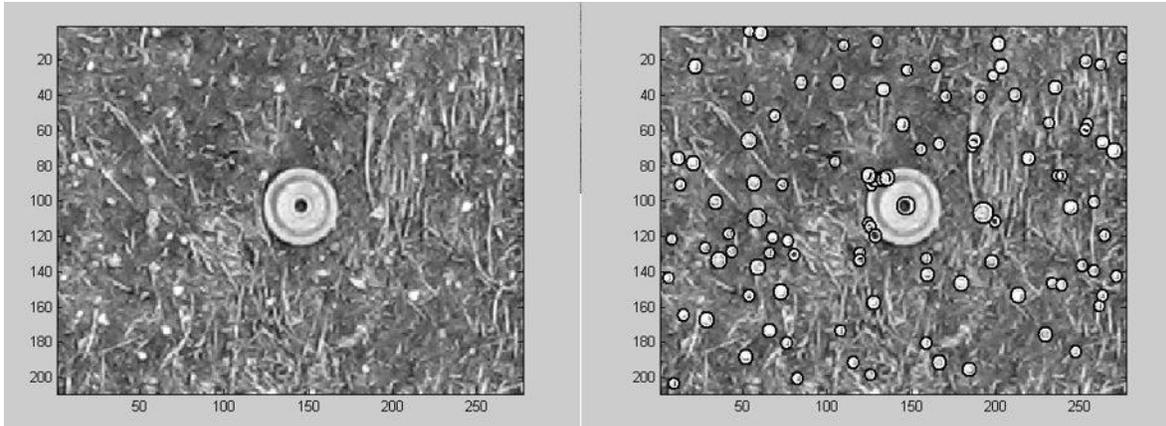


Fig. 3. The initial image (left) and the result of the salient point detection algorithm. The size of each detected speckle is detected automatically and shown on the right image as a circle.

### 3. FEATURE VECTOR REPRESENTATION

As it was mentioned earlier the main property to which a feature vector should conform is its ability to distinctively capture image structure in an image patch around a certain point. At the same time a feature vector should be stable when extracted at the same point of the same entity viewed under varying imaging conditions. Present day feature vectors are extracted in scale invariant manner, stable under varying imaging conditions, noise, illumination, and geometric transforms [5].

In the present work we use SIFT feature vector proposed by D. Lowe [4]. This vector is build upon image gradient direction distribution histograms within a set of patches around a certain point and encompasses image fragment into a vector with 128 components. This number can vary depending on desired performance.

At first, for each salient point from the previous stage one or several directions are defined by finding maxima in gradient direction histogram with the gradients sampled on a predefined grid covering a region around the salient point. Finding one or more dominant directions for each salient point and aligning the local reference frame with this direction to calculate a feature vector makes the latter rotationally invariant.

To construct a feature vector 16 gradient directional histograms are calculated on patches aligned with the local reference frame found at the previous step. The gradient vectors are sampled on grids covering patches. To make the feature vector less sensitive to distant gradients from the sample point, the gradients are weighed by a Gaussian function with its maximum at the salient point and falling down to the peripheral sample grid nodes. In such an approach there are many loose parameters which need to be tuned according to the problem at hand. The best combination of parameters should be defined by running tests on images with the objects to be detected. But in this work we use the parameters reported as giving good results with ordinary real world images [4].

To illustrate how it works on an image with mines we refer to Figure 4. There are several salient points detected at the salient point detection step. Among these salient points there are two of them that fall into two mines and more of them that fall into the background. The feature vectors calculated for four salient points are presented as grayscale images where white cells correspond to larger values of feature vector components. As can be seen the feature vectors falling into background are diverse more significantly than the feature vectors corresponding to the mines while the feature vectors corresponding to the mines look similar to each other. Given an array of such feature vectors extracted from an image, one could cluster a compact class of similar vectors and outline this set as mine candidates. In the next section we consider a clustering algorithm to obtain a set of closest feature vectors.

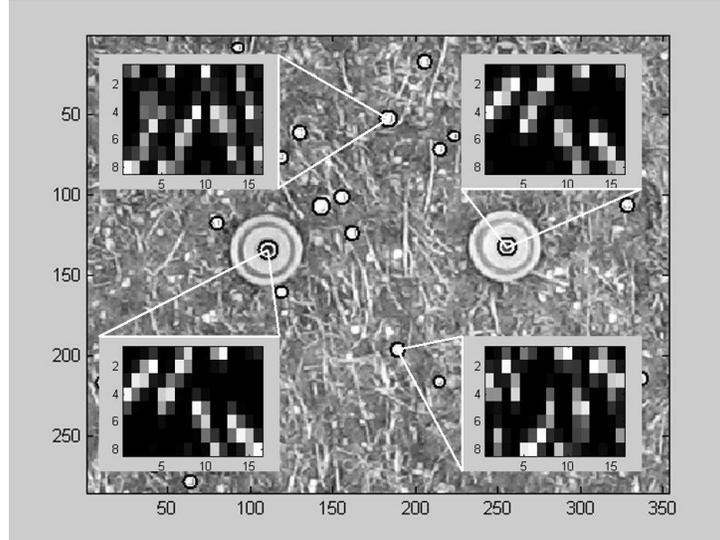


Fig. 4. Four salient points with relevant feature vectors presented as 8x16 gray scale images. The whiter the cell the higher the value of relevant component. Feature vectors representing the same entities in the image are similar to each other and at the same time they are different from the feature vectors falling into the background.

#### 4. SIMILAR OBJECT CLUSTERING

To find a cluster of closest feature vectors we use a hierarchical clustering algorithm. A binary tree is built for all feature vectors extracted from an image. In such a representation any pair of vectors will be invariantly connected to each other at a certain level. The tree begins by enumerating the elements (leaves) and is grown by joining the closest vectors in the beginning and by joining the closest clusters at higher levels. The tree is finished when the last two clusters are joined at the highest level.

To calculate the distance between a pair of vectors we used Euclidean distance, although other metrics: standardized Euclidean distance, Mahalanobis distance, etc. are also possible. Another option to be defined at this stage is the choice of distance between clusters of vectors. By varying these parameters in the hierarchical clustering algorithm, slightly different results can be obtained.

A panoramic nadir image obtained at a height of 27 meters is shown in Figure 5. There are five round dummy mines in the image. Salient points were extracted from the image and only 30 most dominant of them passed through a predefined threshold to the feature vector extraction stage. The dendrogram corresponding to this set of feature vectors is shown in Figure 6. Salient points and relevant feature vectors corresponding to the five round mines in Figure 5 were renumbered to 1-5 for convenience. In the dendrogram the horizontal axis shows index numbers of extracted feature vectors and the vertical axis shows the distance at which two objects are joined. The bigger the value the more distant the objects (leaves or clusters) are.

The clustering algorithm starts by linking (link 1) a pair of closest vectors (1 and 5 in Figure 6) and advances upwards by calculating inconsistency of each link from the first two joined objects to the end of the hierarchy tree (link 5). The inconsistency index is calculated by the following formula:

$$c_n = \frac{h_n - \bar{h}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (h_i - \bar{h})^2}}, \quad (2)$$

where  $\bar{h} = \frac{1}{n} \sum_{i=1}^n h_i$  is the average height of the links below the current link. The result of applying this operation to the links joining vectors 1 – 5 and the rest of the vectors is shown in Table 1. As it can be seen from this table, link 5 has a

maximum inconsistency index. Breaking up link 5 gives a cluster of closest feature vectors. Highlighting an approximate area around each salient point belonging to the cluster that was used to calculate the feature vector may present the result of a similar object search. The result of similar object search is shown in Figure 7 where five round dummy mines are marked by squares.



Fig. 5. A panoramic nadir image taken at a height of 27 meters by a digital camera. Five round dummy mines are scattered in the image.

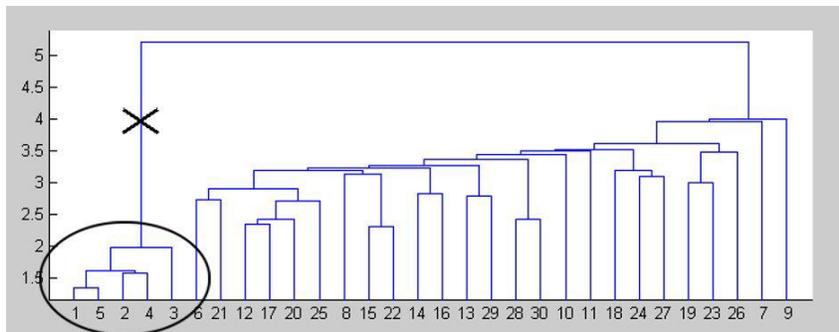


Fig. 6. A dendrogram built for the feature vectors extracted from the previous figure. Numbers 1-5 correspond to round dummy mines shown in the figure above. The inconsistent link is marked by a cross. A cluster of similar vectors is shown by an ellipse.

Table 1. Inconsistency index

Link Number	Feature vectors belonging to the link	Links used in the formula	Inconsistency index
1	1, 5	-	0
2	2, 4	-	0
3	1, 2, 4, 5	1-3	0.8775
4	1-4	1-4	1.5498
5	1-5	1-5	1.98



Fig. 7. A result of the similar object search algorithm.

Upon obtaining mine candidates by a similar object search it is possible to interactively extract feature vectors and make input to the object model for a certain mine type in the database. Subsequently, the user can perform a template-based search if object model database is not empty or a similar object search with the found objects omitted (Figure 1).

## 5. CONCLUSION AND FUTURE WORK

A mine detection system is viewed as an interactive system with a user responsible to make queries to the system, similar to a text search engine but different in input method. Instead of using words the user should interactively define salient points and feature vector parameters taken either from the object model or set manually.

In this paper we tried to show that today paradigm in pattern recognition society may be implemented in a system aimed at facilitating the search for mine-like objects in aerial high resolution images. The key moments in such a system are: user interaction in loop with the system, two stage prior image processing. At the first stage salient points extracted in

scale invariant manner from each image in the database no matter what object a given irregularity belongs. At the second stage a feature vector invariantly encoding a patch around each salient point is calculated. This decomposition of images into feature vectors allows similar object search for which we propose a simple method based on hierarchical clustering. A template-based search is also straightforward in terms of single feature vector for each object in the image. The conducted experiments on aerial images taken from a height of 27 meters showed encouraging results. The following tasks are foreseen for future work:

- \* Implement salient point detection and feature vector extraction algorithms in C/C++ language to achieve better performance with large panoramic images (the current realization is written in Matlab).
- \* Consider different types of salient points in addition to mentioned blobs such as: ridges, steps, etc. Consider other local descriptors compliant with asymmetric not blob like salient points: ridges, curves, etc. Compare their performance.
- \* Develop an interactive application to implement these algorithms interactively via graphical user interface. Perform experiments with a database of images to obtain statistical data and evaluate the performance of the operator using the application in the task of detecting mines in aerial images.

### **ACKNOWLEDGEMENT**

This work is supported by Russian Foundation for Basic Research grant #06-07-08001 and the President of Russian Federation Young Scientist Grant #MK-1136.2007.9

### **BIBLIOGRAPHY**

1. S.I. Ivashov, V.V. Razevig, R.R. Ilushchenko, I.A. Kharchenko, D.V. Tsvetkov. Airborne Minefield Detection for Humanitarian Operations, High Technologies, V. 3, No. 6, 2002, pp. 27-37. (in Russian)
2. S.I. Ivashov, V.V. Razevig, A.P. Sheyko, I.A. Vasilyev, A Review of the Remote Sensing Laboratory's Techniques for Humanitarian Demining. Proceedings of the International Conference on Requirements and Technologies for the Detection, Removal and Neutralization of Landmines and UXO, EUDEM2-SCOT-2003, 15-18 September 2003, Vrije Universiteit Brussel, Brussels, Belgium, 2003, Vol. 1, pp 3-8.
3. T. Lindeberg. Feature Detection with Automatic Scale Selection. Int. J. of Computer Vision, vol. 30, number 2, 1998.
4. David Lowe. Distinctive image features from scale-invariant keypoints. Int. J. of Computer Vision, vol. 60, number 2, 2004.
5. Krystian Mikolajczyk, Bastian Leibe, and Bernt Schiele. Local Features for Object Class Recognition. Proc. of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 2005.