

Using of Viola and Jones Method to Localize Objects in Multispectral Aerospace Images based on Multichannel Features

Sergey Usilin^{1,2}, Oleg Slavin^{1,2,*}

¹Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russian Federation

²Smart Engines Service LLC, Moscow, Russian Federation

Abstract. A new algorithm for localizing engineering objects on multispectral images based on the Viola and Jones method has been developed. The proposed algorithm uses multichannel features allowing to construct classifiers that are sensitive to features of joint brightness distribution and the brightness distribution in different channels. The algorithm described in the paper provides a precision value of 0.96 and a recall value of 0.99 in the problem of localizing oil storage tank images in a set of aerospace images. The proposed algorithm can be used for visual analytics and automatic detection of various critical objects in aerospace images.

Keywords. Viola and Jones method, multispectral aerospace, multichannel features

1 Introduction

To date, visual data obtained from various remote sensing systems are widespread. As a result, remote diagnostics of aerospace monitoring objects is gaining popularity due to modern computing tools for processing remote sensing data [1-3]. In this situation, models and algorithms for recognizing object images observed on remote sensing data that is usually represented as digital multispectral images become fundamental [4].

In the field of digital image processing and pattern recognition, machine learning methods are often used to solve the problem of object localization and identification. One of such methods is the Viola and Jones method [5] allowing to efficiently solve the problem of detecting rigid images. However, despite its universality, the Viola and Jones method should be adapted to solve a specific problem and to achieve high precision indicators.

In this paper, we propose an original method for localizing objects on multispectral images obtained by terrain aerospace images. The proposed method is based on the Viola and Jones method and uses multispectral features to build a stable object detector that is invariant to possible changes in brightness.

2 Problem Statement

In addition to direct aerospace photography, remote sensing includes the decryption of the information received. This information includes knowledge about spatial coordinates of interesting objects.

As initial data, remote sensing systems have images that were obtained in different visibility ranges. Most remote sensing systems return images in the visible

range, including near infrared area. In some complexes, images may be also obtained in thermal and radio ranges (see Fig. 1).

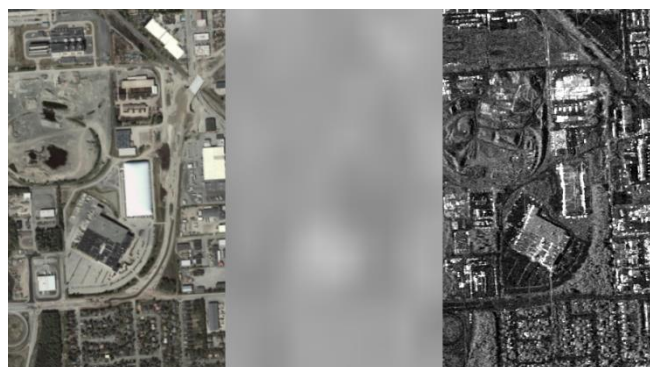


Fig. 1. A fragment of an aerospace probing image in the visible, infrared and radio ranges.

In this paper, we shall consider the problem of localizing geometric coordinates of interesting objects on multispectral images that were obtained during aerospace probing of the earth's surface. The multispectral image was obtained in the visible, infrared and radio ranges.

3 Proposed Method Description

The work proposes an original algorithm for localizing objects based on the Viola and Jones method [5], a statistical construction scheme for object detectors with rigid geometry (based on precedents). The Viola and Jones method uses Haar-like features as a feature space. Their value is based on the difference between the sums of image area pixel brightness inside black and white

* Corresponding author: oslavin@smartengines.ru

rectangles. To efficiently calculate the value of Haar-like features, an integral representation of the image $I_f(y, x)$ is used. For a grayscale image $f(y, x)$ with dimensions $\times N$, it is determined as follows:

$$I_f(y, x) = \sum_{i < y, j < x} f(i, j).$$

A binary weak classifier $h(x): \mathbb{X} \rightarrow \{-1, +1\}$ represented by a recognizing tree with one branch associates the Viola and Jones method with each feature. Such classifiers demonstrate weak localization power. So, the Viola and Jones method uses the AdaBoost algorithm to make a “strong” classifier based on a linear combination of the most powerful weak classifiers:

$$S(x) = \left[\sum_{t=1}^T \alpha_t \cdot h_t(x) > 0 \right],$$

where $[\cdot]$ – indicator function. High performance in the Viola and Jones method is additionally ensured by cascade classifiers that are based on strong classifiers and allow to quickly (in early estimation stages) recognize “empty” images (images without the target object):

$$Cascade(x) = \prod_{i=1}^N [S_i(x) > 0].$$

An object in the image is searched with a built cascade classifier and the sliding window method.

The Viola and Jones algorithm was modified to search for oil storage tanks. An oil storage tank image example is shown in Fig. 2.



Fig. 2. Oil storage tank image example

To effectively train the Viola and Jones detector for the type of objects specified, a feature space allowing to use geometric features of an object, rather than brightness characteristics, should be chosen [6-7]. So, Haar-like features estimated over gradient norm images were used as the space of a feature.

Image processing and analyzing tasks usually considered the term “gradient norm” as its L_2 norm that is estimated by the following formula:

$$G_{L_2} = \sqrt{\frac{\partial f^2}{\partial x} + \frac{\partial f^2}{\partial y}},$$

where $\frac{\partial f}{\partial x}$ – partial derivative of an image $f(y, x)$ by x , and $\frac{\partial f}{\partial y}$ – partial derivative of an image $f(y, x)$ by y .

In recognition problems, from a productive point of view, it turns out to be beneficial to determine its L_1 norm [8] considered as the sum of individual component modules, instead of determining L_2 the gradient norm:

$$G_{L_1} = \left| \frac{\partial f}{\partial x} \right| + \left| \frac{\partial f}{\partial y} \right|.$$

The monotonicity of the difference between rectangle sum values, and not absolute values of these sums, is an important aspect of Haar-like features. So, in this case, the use of the norm L_1 for determining gradient norm is a justified Viola and Jones method improvement.

Regarding the multispectral information, a modification of the Haar-like features is used to determine the difference between total brightness values in different channel sub-windows (each channel of a multispectral image contains an image belonging to one range) [7].

4 Experiments

A training dataset consisting of 20×20 px 70 oil storage tank images, as well as 22 full-size remote probing images without oil tanks was prepared to train the oil storage tank detector. Since the initial number of the training dataset was small, we used augmentation [8-10]. A cascade classifier was used as a high-level classifier structure. The trained cascade diagram is given in Fig. 3.

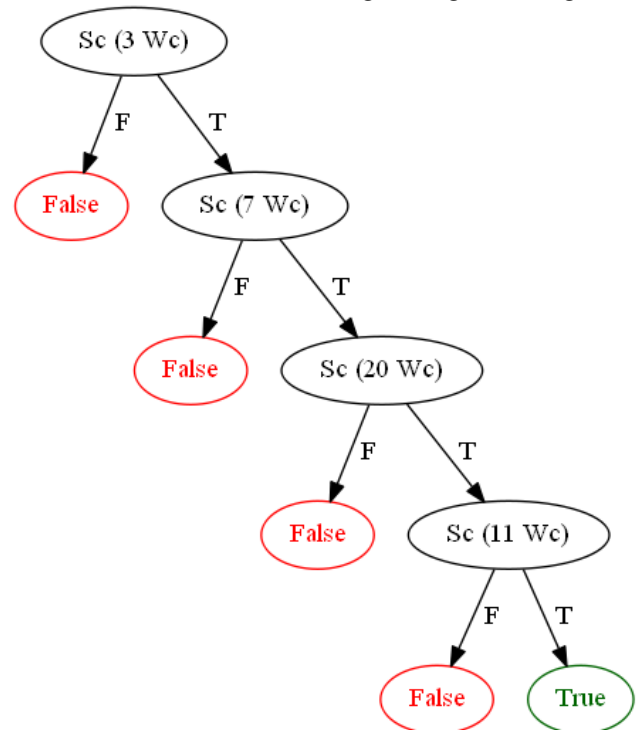


Fig. 3. The trained cascade scheme

The distribution of features can be analyzed with the information saturation map (see Fig. 4), a digital image meeting a classifier in size, where each feature is assigned to a pixel covered by this feature. The more features cover an image area, the brighter this area is on the information saturation map.

From the information saturation map, it follows that trained classifier features are mainly concentrated around the object perimeter. In addition, the information map shows that the classifier uses target object features -

axial symmetry, color uniformity, the border around the perimeter.

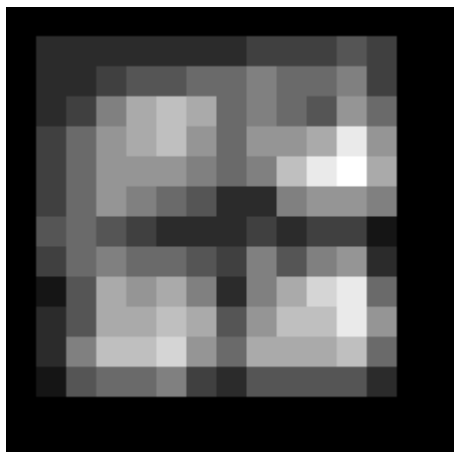


Fig. 4. Trained cascade information saturation map

The trained cascade was applied on test images to assess the classifier's quality. The test set of images contained 73 target objects.

To determine the number of correctly localized objects, we shall use the technique proposed in the PASCAL Visual Object Classes (VOC) Challenge framework [11]. As an answer, a trained cascade returns a framing rectangle, as well as its confidence value that can be used to find a compromise between the first and second kind of errors. Table 1 shows experimental results with quantities for given self-confidence values

Table 1 gives experimental results. The following statistics was determined for the given confidence degree values: the number of correct localizations (true positive, TP), the number of false localizations (false positive, FP), the number of false omissions (false negative, FN), as well as the precision, recall and F-measure were determined.

Table 1. Experimental Results

Conf.	TP	FP	FN	Precision	Recall	F ₁
0.0	73	4	0	0.948	1.000	0.973
0.3	72	3	1	0.960	0.986	0.973
0.6	70	2	3	0.972	0.959	0.965
0.9	66	0	7	1.000	0.904	0.950

Table 1 shows that the highest F-measures are achieved in the first two lines, while the highest precision is achieved with a confidence level of 0.9 or higher.

The average number of determined features in the image is 3.1908. With regard to the trained classifier structure (with three features at the first cascade level), this means that, in case of the majority of analyzed image sections, at the first cascade level, the trained classifier presented a confident answer that the specified region does not belong to the target object.

The detector's operation is given in Fig. 5.



Fig. 5. An example of oil storage localization with a trained cascade

5 Conclusion

Nowadays various remote sensing systems are widespread. As a result, different methods of automatic object detection and recognition of aerospace images are very actual. In this paper, a new algorithm for localizing engineering objects on such multispectral images are proposed. The algorithm is based on the Viola and Jones method and adapted for efficient work on multispectral images.

We have applied the proposed algorithm to solve the problem of localization of oil tank storages on aerospace images. The algorithm provides a precision value of 0.96 and a recall value of 0.99.

The proposed algorithm can be used for visual analytics and automatic detection of various critical objects in aerospace images.

This work is partially supported by Russian Foundation for Basic Research (17-29-03161 and 17-29-03297).

References

1. M.S. Boori, A.V. Kuznetsov, K.K. Choudhary etc, *Satellite image analysis to evaluate the urban growth and land use changes in the city of Samara from 1975 to 2015*, Computer Optics, **39**, 5 (2015)
2. A.M. Chandra, S.K. Gosh, *Distancionnoe zondirovanie i geograficheskie informacionnye sistemy* (2008, in Russian)
3. A.V. Kuznetsov, V.V. Myasnikov, *A comparison of algorithms for supervised classification using hyperspectral data*, Computer Optics, **38**, 3 (2014)
4. K.B. Bulatov, M.V. Chukalina, D.P. Nikolaev, *Fast X-Ray Sum Calculation Algorithm for Computed Tomography Problem*, Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software (Bulletin SUSU MMCS), **13**, 1 (2019)

5. P. Viola, M. Jones, *Robust real-time object detection*, International Journal of Computer Vision (2001)
6. A.A. Kotov, S.A. Usilin, S.A. Gladilin, D.P. Nikolaev, *Construction of robust features for detection and classification of objects without characteristic brightness contrasts*, Journal of information technologies and computing systems, **1** (2014)
7. E. G. Kuznetsova, S.A. Usilin, A.G. Minkina, D.P. Nikolaev, *Modification of weak classifiers of the Viola-Jones method for multispectral images*, Proceedings of the Information Technology and Systems (2015, in Russian)
8. D.P. Matalov, S.A. Usilin, V.V. Arlazarov, *Modification of the Viola-Jones approach for the detection of the government seal stamp of the russian federation*, the Proceedings of the Eleventh International Conference on Machine Vision (2019), doi:10.1117/12.2522793
9. A.V. Gayer, Y.S. Chernyshova, A.V. Sheshkus, *Effective real-time augmentation of training dataset for the neural networks learning*, Eleventh International Conference on Machine Vision (2019), doi: 10.1117/12.2522969
10. S.A. Usilin, P.V. Bezmaternykh, V.V. Arlazarov, *Fast approach for QR code localization on images using Viola-Jones method*, the Proceedings of the Twelfth International Conference on Machine Vision, (2020)
11. M. Everingham, *The pascal visual object classes (VOC) challenge*, Int. J. Comput. Vis, **88**, 2 (2010)