ISSN 0027-1349, Moscow University Physics Bulletin, 2015, Vol. 70, No. 6, pp. 479–483. © Allerton Press, Inc., 2015.
Original Russian Text © A.V. Zubyuk, A.B. Fedotov, 2015, published in Vestnik Moskovskogo Universiteta. Fizika, 2015, No. 6, pp. 49–54.

An Algorithm for Object Identification Based on Hyperplane Separation But Tolerant to Illumination and Viewing Angle Variations[1]

A. V. Zubyuk and A. B. Fedotov

Department of Physics, Moscow State University, Moscow, 119991 Russia
e-mail: zubuk@cmpd2.phys.msu.ru, toivo@ya.ru

Received June 13, 2015; in final form, August 21, 2015

Abstract—A new algorithm for object identification by their images is proposed. It combines tolerance against imaging conditions variations proper to morphological methods for image analysis by Pyt'yev with computational simplicity of methods based on separating classes of images with hyperplane. The algorithm was tested on images taken in different illuminations and views. It can be applied to analyse microscopic image in biophysics, acoustic signals in geophysics, Earth resources using satellite data, etc.

Keywords: morphological image analysis, pattern recognition, hyperplane separation, support vector machine.

DOI: 10.3103/S0027134915060193

INTRODUCTION

Many problems of image and signal analysis are difficult to solve because aside from the useful information any image or signal carries the information about unknown, uncontrolled, and varying registration conditions. For example the image of any object carries not only information about geometric shape, orientation and other properties of this object, but also the information about lighting conditions under which the image is taken, about the parameters and settings of the camera used etc. If one considers the problem of object recognition then the lighting conditions and the camera parameters shall be considered as "nuisance" parameters and their variations shall not affect the recognition results. As another example consider an acoustic signal from a remote source such as nuclear explosion or rock fall. When such a signal is propagating through the atmosphere it undergoes significant distortions. So aside from the information about its source it carries the information about acoustic properties of the atmosphere [1]. If one solves the problem of the signal source identification then acoustic properties of the atmosphere shall be considered as "nuisance" parameters and their variations shall not affect the identification results.

Methods for morphological image and signal analysis [2–8] were developed exactly for solving such problems. They consider images and signals of various nature as elements of function spaces, which makes it possible to use general methods of linear algebra and functional analysis to describe and analyze images and signals. So in this paper we use the term "image analysis" implying that the same approach can be applied to analyze various signals. The morphological image analysis is based on the principle of invariance under variations of unknown and uncontrolled registration conditions. The maximum invariant of the set of all such variations is called the *image shape* and can be defined as the set of all images carrying the same useful information but taken under different registration conditions. The techniques of orthogonal and oblique projection onto image shape are used to solve applied problems of image analysis with the morphological methods. This determines their computational efficiency if the image shape is a fairly simple set, for example a linear subspace of small dimension. However if the image shape is not a linear subspace and has large dimension, the problem of projection becomes very complex both in mathematical and computational contexts.

Another widely used approach to some problems of image and signal analysis such as object recognition or identification of sound sources is separating the classes of images or signals using quite simple manifolds (such as hyperplanes) in the feature space [9–14]. This approach is simple to implement and allows developing more computationally efficient algorithms compared to the morphological algorithms based on the projection technique. The main difficulty of this approach is to select the features (i.e. the numerical characteristics of the image) that make it possible to separate the classes of images. There are some meth-

[1] The article was translated by the authors.

ods selecting such features automatically [15–19]. However in most cases success in selecting the features depends on the intuition and experience of the researcher [20–27]. In addition such features are often not invariant under all possible variations of registration conditions.

We consider the model of image registration based on the assumption that the objects are illuminated by incoherent light sources and each photodetector of the camera used to registration produces an output proportional to incident light flux. We demonstrated that in the context of this model the sets of images of different objects taken in all possible illuminations and views are separable with hyperplanes directly in the image space. This allowed us to develop the object identification algorithm with ability to learn based on hyperplane separation and tolerant against variations of illumination and view. The paper contains the results of tests performed on the real images and the results of comparison between the developed algorithm and well-known support vector machine [13, 14].

1. UNDERLYING IMAGE REGISTRATION MODEL

In this paper a finite subset of coordinate plane $(-\infty,\infty)\times(-\infty,\infty)$ is called *field of vision*, and is denoted by $X = \{x_1,\dots,x_n\}$. It consists of the points $x_1,\dots,x_n$ called *pixels*.

We say that any function $f : X \to (-\infty,\infty)$ is the *image*, and its value $f(x)$ is *brightness* of image f, $x \in X$. The image

$$x \mapsto \alpha f(x) + \beta g(x)$$

is the linear combination of images f and g with coefficients α and β, and the following number is their scalar product:

$$(f,g) = \sum_{x\in X} f(x)g(x).$$

Let R be the set of all images with the linear operations and scalar product determined above. The space R is n-dimensional Euclidean one called *image space* in this paper.

Consider the model of image registration defined by the following constraints on the camera, objects and light sources:

1. The Camera Used

Brightness $f(x)$ of the image f at point $x \in X$ is proportional to the light flux incident upon a photodetector corresponding to the pixel x.

2. Optical Properties of the Objects

The objects don't emit their own light but only reflect it. In the case of illumination produced by a point light source the relation between intensities of incident and reflected light is as follows:

$$I_{\mathrm{ref}(\vec{e})} = kI_{\mathrm{inc}(\vec{s})}, \qquad (1)$$

where $\vec{s}$ and $\vec{e}$ are the unit vectors of incident and reflected light directions, $I_{\mathrm{inc}(\vec{s})}$ and $I_{\mathrm{ref}(\vec{e})}$ are the corresponding intensities, and k is the reflectance which doesn't depend on $I_{\mathrm{inc}(\vec{s})}$.

3. Light Sources

Illumination is produced by incoherent point light sources.

2. SHAPE OF THE OBJECT IMAGE TAKEN FROM THE CONSTANT CAMERA VIEWPOINT

Consider registration of object images from a fixed viewpoint. Let ϕ be a set of scalar parameters that define this viewpoint. For example two angles defining orientation of the object can be used as such parameters. Assume that only the lighting conditions can vary. Constraints 1–3 above determine the set of all object images taken under all possible lighting conditions which is the *shape* of the object image [6].

Indeed, let the image $f \in R$ of the object is taken under the lighting conditions L. Let's multiply the brightness of all the light sources producing illumination L by $\alpha \geq 0$ and denote the new lighting conditions by αL. Changing the light conditions from L to αL results in multiplying intensity of the light incident upon the object by α. Therefore according to (1) intensity of the reflected light is also multiplied by α which results in the same change of the light flux incident upon each photodetector. So according to constraint 1 on the camera used the image of the object taken under the light conditions αL equals αf. Thus if f is the object image, αf is the image of the same object taken probably under another lighting conditions, $\alpha \geq 0$.

Consider now two images $f_1 \in R$ and $f_2 \in R$ of the same object taken under the lighting conditions L_1 and L_2 (in this case illuminations L_1 and L_2 are assumed to be independent). Let illumination L_{1+2} be superposition of the illuminations L_1 and L_2, i.e. L_{1+2} is produced by the light sources producing both L_1 and L_2 switched on simultaneously. As (1) is linear and the light sources are incoherent (constraint 3) the light flux incident upon each photodetector is the sum of fluxes produced by the illuminations L_1 and L_2. So the image of the object taken under the illumination L_{1+2}

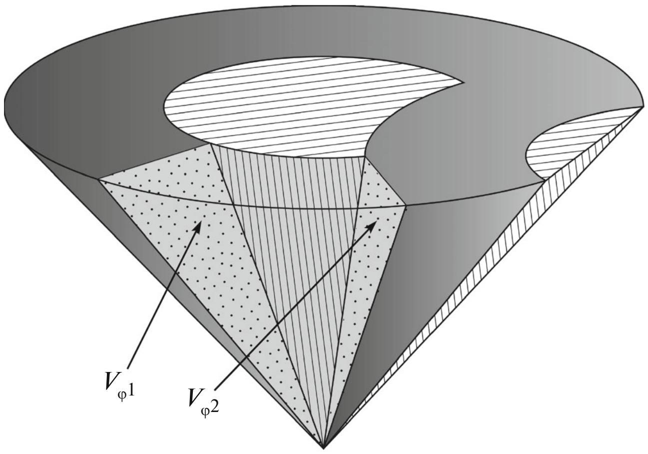

Fig. 1. Visual representation of the image shape of the object photographed under various illuminations and views. Two convex cones V_{ϕ_1} and V_{ϕ_2} marked with dots are the sets of images taken under various illuminations from two different camera viewpoints. Image shape of the object (marked with gray) is the union of convex cones corresponding to various camera viewpoints; this union is a cone but isn't necessarily a convex one. Its complement to the convex hull (marked with hatching) consists of unreal images.

is $f_1 + f_2$. Thus if f_1 and f_2 are the images of some object, $f_1 + f_2$ is the image of the same object.

Consequently we have the following. Shape of the object image taken from the fixed viewpoint ϕ is the set $V_\phi \subset R$:

$$f \in V_\phi \Rightarrow \alpha f \in V_\phi, \quad \alpha \geq 0,$$
$$f_1, f_2 \in V_\phi \Rightarrow f_1 + f_2 \in V_\phi.$$

It means that the image shape V_ϕ is a convex cone [28].

3. SHAPE OF THE OBJECT IMAGE TAKEN FROM AN UNKNOWN CAMERA VIEWPOINT

Suppose that the viewpoint determined by the parameter ϕ can vary in addition to the lighting conditions. Let Φ be the set of values of vector parameter ϕ. In this case the *shape* V of the object image is represented as follows:

$$V = \bigcup_{\phi \in \Phi} V_\phi,$$

where V_ϕ is the convex cone containing all the images of the object taken from the viewpoint $\phi \in \Phi$ under all possible lighting conditions. The set V is a cone but, generally speaking, not a convex one (the area marked with gray in Fig. 1).

Let $\overline{V}$ be the convex hull of the shape V. Since dimension $\dim R = n$ of image space R is finite, according to Caratheodory's theorem [28] the set $\overline{V}$ can be represented as follows:

$$\overline{V} = \left\{ \sum_{i=1}^{n+1} \alpha_i f_i | \alpha_i 0, f_i \in V, i = 1, \dots, n+1, \sum_{i=1}^{n+1} \alpha_i = 1 \right\}, \quad (2)$$

and $\overline{V}$ is a convex cone. Let δV be the complement of the shape V to its convex hull $\overline{V}$, i.e. $\delta V = \overline{V} \setminus V$. The set δV is marked with hatching in Fig. 1. According to (2) any image f from δV can be represented as a convex combination of some images of the object, and some of that images are taken from different viewpoints and occur in the combination with positive coefficients (otherwise $f \in V_\phi \subset V$ for some $\phi \in \Phi$). Figure 2 represents an example of such an image provided the object is a plaster figurine of a bear. As one can see such images look like a superposition of significantly different images, therefore we assume below that they are not images of any real object and can't be taken at all.

Consequently we have the following. The shape of the object image taken from unknown viewpoint is a cone V, and its convex hull $\overline{V}$ contains no real images except the images of this object.

4. ADAPTIVE MACHINE LEARNING IDENTIFICATION ALGORITHM

Let O_1 and O_2 be the different physical objects and their image taken from any viewpoint and under any lighting conditions (except absolute darkness) differ

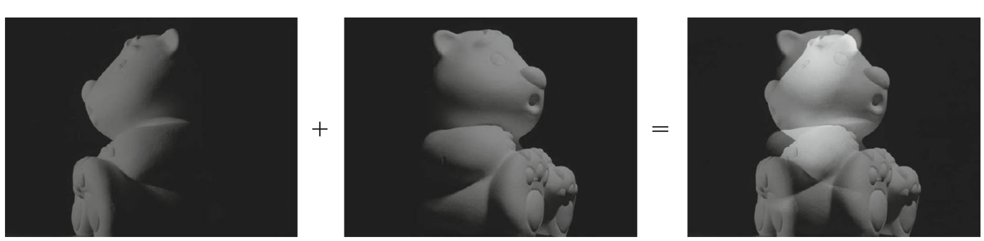

Fig. 2. Example of an unreal image produced by summing of two images of the same object taken from the different camera viewpoints.

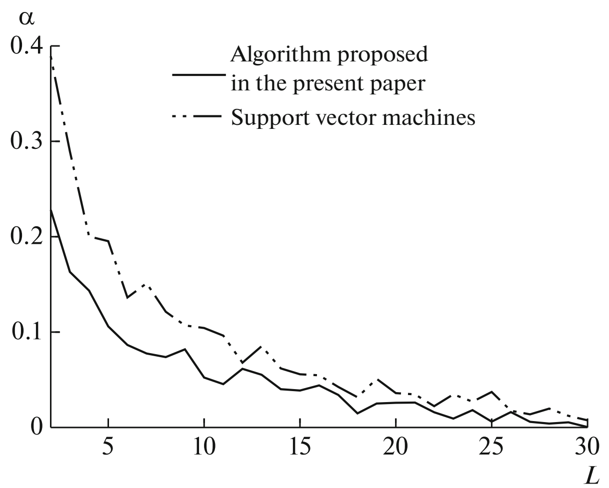

Fig. 3. Graph of average error rate α versus size L of training image set for algorithm proposed in this paper (solid line) and support vector machine (dash-dotted line).

from each other. Consider the problem of object identification by their images. Thus we have the image $g \in R$ of one of the objects O_1, O_2 and need to determine what object is depicted in the image.

Let $V^{(i)}$ be the image shape of the object O_i photographed from various viewpoints and under various lightning conditions, $i = 1, 2$. According to the previous section the convex hull $\overline{V}^{(i)}$ of the shape $V^{(i)}$ is a convex cone that contains real images only of the object O_i and contains no images of the object O_j, i.e. $\overline{V}^{(i)} \cap V^{(j)} = \{0\}, j \neq i$. Let's make stronger but also realistic assumption that any convex combination of images of the object O_1 cannot be represented as a convex combination of images of the object O_2 and vise versa, i.e. $\overline{V}^{(i)} \cap \overline{V}^{(j)} = \{0\}, j \neq i$. According to the above conclusions and assumptions $\overline{V}^{(1)}$ and $\overline{V}^{(2)}$ can be separated with a hyperplane passing through the origin directly in the image space R. Thus, there is such an image $w \in R, \|w\| > 0$:

$$(g, w) \leq 0 \;\; at \;\; g \in \overline{V}^{(1)}, \\ (g, w) \geq 0 \;\; at \;\; g \in \overline{V}^{(2)}, \tag{3}$$

moreover $(g, w) = 0$ if and only if $g \in \overline{V}^{(1)} \cap \overline{V}^{(2)} = \{0\}$.

Since $V^{(i)} \subset \overline{V}^{(i)}$ and the image $g = 0$ corresponds to the case of absolute darkness (which is of no interest), inequalities (3) can be used to identify the object as follows:

$$\textit{if } (g, w) < 0, \quad g \textit{ is the image } O_1, \\ \textit{if } (g, w) > 0, \quad g \textit{ is the image } O_2. \tag{4}$$

To build the image $w \in R, \|w\| > 0$ in (4) one can use the following heuristic algorithm. Let

$$f_1^{(i)}, \ldots, f_{L_i}^{(i)} \text{-} \tag{5}$$

be the images of object O_i taken from various viewpoints and under different lightning conditions, $i = 1, 2$. We say the set (5) is *training sample*. Define w as follows:

$$w = e^{(2)} - e^{(1)}, \tag{6}$$

where

$$e^{(i)} = \frac{1}{L_i} \sum_{j=1}^{L_i} f_j^{(i)} / \left\| f_j^{(i)} \right\|, \quad i = 1, 2.$$

Such heuristic is based on the following considerations. If the images from the training set are taken from such viewpoints and under such lighting conditions that $f_j^{(i)} / \left\| f_j^{(i)} \right\|, j = 1, \ldots, L_i$, are "uniformly scattered" by intersection of the cone $V^{(i)}$ with the unit sphere, then the vector $e^{(i)}$ belongs to the "center" ray of cone $V^{(i)}$. Therefore we can expect that the hyperplane orthogonal to the image w connecting[2] $e^{(1)}$ and $e^{(2)}$ separates the cones $V^{(1)}$ and $V^{(2)}$.

5. THE TESTS OF THE PROPOSED ALGORITHM

To test the proposed algorithm we used 240 images of plaster figurines similar to the figurines shown in Fig. 2 (left and center): 120 images of the plaster figurine of a bear (object O_1), and 120 images of the plaster figurine of a girl (object O_2). Images of those figurines were taken under different lighting conditions and from different viewpoints.

The image set was randomly divided into two parts:

1. Training Set

To build the image w according to (6) $L = L_1 = L_2$ images of each figurine were used.

2. Test Set

The remaining $240 - 2L$ images were used to test the identification algorithm based on (4).

Such random division was performed many times. As a result of the tests average identification error rate α was calculated as the number of erroneously identified test images (when a bear was identified as a girl or

[2] We use the term "connecting" in regard with geometric interpretation of vector difference in three-dimensional space. According to that interpretation the vector $\vec{c} = \vec{b} - \vec{a}$ connects the ends of the vectors $\vec{a}$ and $\vec{b}$.

vise versa) divided by the total number of images used for the tests. A graph of average error rate α versus size L of the training image set is shown in Fig. 3 (solid line).

In addition dash-dotted line in Fig. 3 shows a similar graph for another widely used identification algorithm with ability to learn based on separation with hyperplanes and called Support Vector Machine (SVM) [13, 14]. Unlike the algorithm proposed in this paper SVM doesn't use any underlying model of image registration and separates classes of images with hyperplanes not necessarily passing through the origin.

According to the results shown in Fig. 3, the algorithm proposed in this paper demonstrates lower average error rate than support vector machine.

CONCLUSIONS

A new adaptive machine learning algorithm for object identification by their images taken under various illuminations and views is proposed. The efficiency of the algorithm is proved both theoretically by analyzing the model of image registration and experimentally by testing it on the set of real images. During the tests the proposed algorithm demonstrated higher identification quality in comparison with support vector machine.

Further investigation may be aimed to adaptation of the proposed algorithm to the practically important case of unknown and uncontrollable parameters of the camera photodetectors.

ACKNOWLEDGMENTS

The research was supported by the Russian Foundation for Basic Research, grant no. 14-07-00409A.

REFERENCES

1. S. N. Kulichkov, A. I. Chulichkov, and D. S. Demin, *Morphological Analysis of Infrasound Signals in Acoustics* (Moscow, 2010) [in Russian].
2. Yu. P. Pyt'ev, Dokl. Akad. Nauk SSSR **224**, 1283 (1975).
3. Yu. P. Pyt'ev, Dokl. Akad. Nauk SSSR **269**, 1061 (1983).
4. Yu. P. Pyt'ev, Pattern Recognit. Image Anal. **3**, 19 (1993).
5. Yu. P. Pyt'ev, Pattern Recognit. Image Anal. **8**, 517 (1998).
6. Yu. P. Pyt'ev and A. I. Chulichkov, *Methods for Morphological Analysis of Images* (Moscow, 2010) [in Russian].
7. Yu. V. Vizil'ter and S. Yu. Zheltov, J. Comput. Syst. Sci. Int. **47**, 944 (2008).
8. Yu. V. Vizil'ter and S. Yu. Zheltov, J. Comput. Syst. Sci. Int. **48**, 282 (2009).
9. J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles* (Addison-Wesley, 1974; Mir, Moscow, 1978).
10. V. N. Vapnik and A. Ya. Chervonenkis, *Pattern Recognition Theory* (Moscow, 1974) [in Russian].
11. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
12. R. A. Fisher, Ann. Eugen. **7**, 179 (1936).
13. B. E. Boser, I. Guyon, and V. N. Vapnik, in *Proc. 5th Annual Workshop on Computational Learning Theory, Pittsburgh, USA, 1992* (ACM, New York, 1992), Vol. 5, p. 144.
14. C. Cortes and V. Vapnik, Mach. Learn. **20**, 273 (1995).
15. K. Pearson, London, Edinburgh Dublin Philos. Mag. J. Sci. **6**, 559 (1901).
16. H. Hotelling, J. Educ. Psychol. **24**, 417 (1933).
17. I. Jolliffe, *Principal Component Analysis* (Springer, 2002).
18. J. B. Tenenbaum, V. De Silva, and J. C. Langford, Science **290**, 2319 (2000).
19. J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction* (Springer, 2007).
20. R. Chellappa and R. Bagdazian, IEEE Trans. Pattern Anal. Mach. Intell. **6**, 102 (1984).
21. G. Zhang, Z. M. Ma, Q. Tong, et al., in *Proc. Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing, Harbin, China, 2008*, p. 71.
22. W. W. Boles, in *Proc. 4th Int. Symp. on Signal Processing and Its Applications, Gold Coast, Australia, 1996* (IEEE, 1996), Vol. 2, p. 601.
23. R. Szewczyk, K. Grabowski, M. Napieralska, et al., Pattern Recognit. Lett. **33**, 1019 (2012).
24. S. Abbasi, F. Mokhtarian, and J. Kittler, Multimedia Syst. **7**, 467 (1999).
25. M. Daoudi and S. Matusiak, J. Visual Lang. Comput. **11**, 287 (2000).
26. L. M. Mestetskii, Mash. Obuchenie Anal. Dannykh **1**, 1291 (2014).
27. S. V. Sidyakin and Yu. V. Vizil'ter, Komp'yut. Opt. **38**, 511 (2014).
28. R. T. Rockafellar, *Convex Analysis* (Princeton Univ. Press, 1970; Mir, Moscow, 1973).