

This paper reports a study into the methods for recognizing the type of an air object on a digital image acquired from an air situation video monitoring system. A method has been proposed that is based on the application of a specific neural network, which solves the problem of categorizing multidimensional complex vectors of objects' features based on complex calculations. In this case, a feature vector for recognizing the type of an air object is built on the basis of a Fourier transform for the sequence of coordinates of its two-dimensional contour. A technique has been proposed to train a neural network to recognize the type of an air object based on three image classes corresponding to three projections. This makes it easier to solve the classification problem owing to a more compact arrangement of the multidimensional feature vectors. The architecture of an air situation video monitoring system has been suggested, which includes an image preprocessing module and a module of a complex-valued neural network. Pre-processing makes it possible to identify an object's contour and build a sequence of normalized descriptors, which are partially independent of the spatial position of the object and the contour processing technique. Existing methods of air object recognition require significant computational resources and do not take into consideration the specificity of recognizing objects with three degrees of freedom or do not account for the complex nature of the numerical representation of a contour. This study has shown that the reported results make it easier to train a neural network and reduce the hardware requirements in order to solve the task of air situation video monitoring. The proposed solution leads to increased mobility and extends the scope of application of such systems, including individual devices

Keywords: air object recognition, contour analysis, Fourier descriptors, complex-valued neural network

UDC 004.931:519.688

DOI: 10.15587/1729-4061.2020.220035

A METHOD OF AIR OBJECT RECOGNITION BASED ON THE NORMALIZED CONTOUR DESCRIPTORS AND A COMPLEX-VALUED NEURAL NETWORK

V. Yesilevskiy

PhD, Associate Professor*

E-mail: yes_v_s@ukr.net

A. Tevyashev

Doctor of Technical Sciences,
Professor, Head of Department*

E-mail: tad45ua@gmail.com

A. Koliadin

Junior Researcher*

E-mail: antonkoliadin@gmail.com

*Department of Applied Mathematics

Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166

Received date 25.09.2020

Accepted date 06.11.2020

Published date 25.12.2020

Copyright © 2020, V. Yesilevskiy, A. Tevyashev, A. Koliadin

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0>)

1. Introduction

Advanced development of air objects increases the complexity of tasks related to their detection and recognition. Solving these tasks fast at high quality is of great importance for civilian applications in air traffic control, airport air situation monitoring, as well as military activities. Until recently, various classes of aircraft were the target of air recognition tasks. Currently, the list of types of aircraft objects has been significantly expanded through the use of unmanned aerial vehicles (UAVs), quadcopters, cruise missiles, and helicopters. That has fundamentally changed the range of detectable parameters of air objects, starting from shape and size, to the dynamic characteristics of the movement. Therefore, optical video surveillance systems are increasingly being used to detect aerial objects. Such an approach makes it possible to reduce the dimensions of detection systems, produce them at lower costs, and, therefore, to render more mobility to them. Video monitoring also avoids problems related to the masking of characteristics of air objects in the radar detection area (stealth technology, small size, etc.). Harnessing new

technologies shifts the weight of air-recognition tasks into the field of digital video processing and automated object recognition.

In the area of computer vision, the tasks of determining the type of image refer to classification problems. Solving them is typically based on the use of deep convolutional networks. Such a solution is too universal, not taking into consideration the specificity of the subject area. It should be taken into consideration that in the area of recognition of the type of air objects, the contours of the object yield enough information to solve the problem. The issue is that the resulting numerical characteristics of the contour should be invariant relative to the geometric distortion (displacement, orientation, scale) of the object's image. For air objects, this is especially relevant due to the three degrees of freedom in determining the position and is an important area to study.

One of the most effective and commonly used approaches to numerical description of the geometric shape of a flat object's contour is the application of a Fourier transform procedure. This approach is particularly interesting because it generates a unique one-dimensional identification sequence of the

standard size, termed as Fourier descriptors, for all examined objects. It solves not only the problems of invariance but also filtering for noisy images.

However, the Fourier descriptors that are complex in their nature are used to solve the task of classifying multidimensional vectors on multi-layered neural networks, which, in the classical version, are based on real calculus. Recently, there has been an increased interest in the application of complex-valued neural networks because they operate with a related phase and the amplitude of the input signal. Studying the applicability of such neural networks for specific classification tasks is important not only because it is most adequate to the representation of input vectors. It appears to be modeling a process that is close to the actual mechanisms of brain image recognition.

The deep convolutional networks that are used in laboratory image recognition experiments are too heavy both at the training stage and at the operational stage. For actual application, including mobile devices, methods that require fewer computing resources are in demand. This is confirmed by works [1–3] that address the use of object recognition methods based on feature vectors, including Fourier descriptors, acquired from a contour analysis. However, known solutions involving these methods either do not use complex-valued neural networks or, at the stage of training a neural network, do not parse the contours of objects into classes corresponding to three different projections. This suggests the relevance of our research into the application of complex-valued neural networks to recognize air objects in an arbitrary spatial position, based on the Fourier descriptors of the contour.

2. Literature review and problem statement

Study [4] gives a detailed overview of the significant advances in the use of deep convolutional neural networks to recognize arbitrary images. However, as shown in papers [5, 6], deep learning methods require a large number of training sets and significant computer time to train a network. Work [7] described the application of these approaches in the field of air object recognition but it does not overcome these shortcomings.

It is shown in [8] that the use of transfer training may become an option to overcome the heaviness of deep network training from scratch. This method assumes that a pre-trained universal network additionally learns from specific images of air objects.

The application of transfer training makes it possible to bypass the limitations associated with the training time. However, some questions remain open. First, there is the issue of the redundancy of a universal approach based on deep neural networks when it comes to the recognition of types of air objects. Second, there is an issue related to the division of object classes, each of which is represented by too different (due to spatial orientation) images of objects on a flat image.

It should also be taken into consideration that explaining the recognition process in deep networks is based on highlighting the hierarchy of attributes in the image. In the first layers of the network, these features are special points (including the points of the contour). In the subsequent layers, these attributes are merged in groups that become a feature of the next level, etc. The hierarchy of attributes makes it possible to generate numerical descriptors of objects in the image. These descriptors underlie the classification problem tackled

by the last layers of the network, which, in essence, are a classic multi-layered perceptron.

This means that the network learning process involves over-training the network, including for generating the contours' attributes, although this can be solved more effectively, as shown in work [9], by classical methods of image processing.

Describing an image of an object by its contour is sufficient for the task of recognizing the types of air objects; it utilizes much less information than when analyzed using deep neural networks, which allows for a series of advantages. Studies [10, 11] report various methods for the mathematical notation of a contour; work [12] shows the use of these methods to recognize the types of air objects.

When recognizing 3-dimensional objects by their 2-dimensional image, there is an issue related to deriving a numerical descriptor invariant relative to the orientation of the object and its size in the image. The work cited earlier [1] applied various characteristics of a plane's contour, acquired by using Hu moments, Zernike moments, as well as wavelet moments, to solve this task. However, the most interesting for solving the problem of air object type recognition is the use of Fourier descriptors, based on the application of a discrete Fourier transform to the function that describes the contour of the object. Fourier transforms in the tasks of describing geometric shapes were first applied by Cosgriff in 1960. As shown in papers [10, 11], this method is based on the representation of a closed flat curve by a sequence of points whose coordinates are considered complex numbers. The application of a Fourier discrete transform to this sequence generates a unique one-dimensional identification sequence of standard size values called Fourier descriptors, which possesses a series of interesting properties that are discussed in detail below in terms of contour description. Similar to how the spectrum of an audio signal, derived from a Fourier transform, identifies this signal, the Fourier descriptors identify the closed contour of the flat shape. In fact, this sequence is a digital passport of the shape.

Fourier descriptors partially resolve the issue of invariance in relation to the rescaling and rotation in the image plane. However, direct use of all the information derived from a Fourier transform is not possible as the phase component of the descriptors depends on the choice of the starting point of the contour. This issue requires a separate study.

The task of recognizing aircraft types based on feature vectors derived from the Fourier descriptors has been solved since the 1980s by using a variety of classification methods. Work [13] applies methods of correlational analysis, distance assessment, support vector machine; paper [14] explored as a separate direction the use of neural networks to solve the task of categorizing objects by their descriptors of a Fourier contour.

A conventional method for solving the problems to classify objects described by the multidimensional vector of numerical values is the use of classic multi-layered neural networks. In the case of Fourier descriptors, the characteristics vector is described in a complex space while the network operates with physical arithmetic.

The idea of using complex neural networks whose operation involves the transformation of input values within a complex space is more appropriate for the task being solved. Complex neural networks have been studied since the 1970s, starting from work [15]. There are two main approaches to this area of study. The first (complex-valued neural network –

CVNN) follows the same principles as modern real-valued neural networks. In this case, they choose, as shown in [16, 17], a specific activation function, and use Clifford's algebra to train a network by a gradient method of back-propagation. Paper [18] reports a comparative analysis of classification tasks for the real-valued and complex-valued multi-layered neural networks; the authors show the advantages and limitations of this type of network.

The second approach, proposed in [19–21] for complex-valued neural networks with multi-valued neurons (MLMVN), employs a specific activation function and does not need to be differentiated for training. This approach to solving classification problems has shown some computational advantages over real-valued neural networks. An important factor in the application of this approach, as shown in [19], is also that its computational model more accurately describes physiological processes in the brain when solving a recognition task.

Work [21] demonstrates the application of complex-valued neural networks to model tasks that have a known solution for real-valued networks. In the cited work, the author shows that the complex-valued networks successfully categorize all the Boolean functions with two inputs, as well as a classification task for a two-spiral problem.

A review of the results of the above studies reveals the following shortcomings in existing approaches to solving a task of air object recognition. First, deep convolutional network methods are excessively heavy and do not take into consideration the specificity of air object recognition. Second, methods that classify feature vectors, based on the contouring of an object, do not use all the information received from Fourier descriptors. Third, the neural-network classification methods do not take advantage of complex-valued neural networks and do not take into consideration the possibility to simplify learning when dividing the training sample into subclasses corresponding to three different projections.

All this suggests that it is appropriate to construct an air object recognition method that would combine the advantages of a contour analysis based on Fourier descriptors and complex-valued neural networks in order to overcome the above shortcomings.

3. The aim and objectives of the study

The aim of this study is to construct an air object recognition method based on the Fourier descriptors and neural networks with a complex calculus, which would make it possible to effectively recognize the type of air object without the excessive cost of computer resources.

To accomplish the aim, the following tasks have been set:

- to investigate the dependence of changes in different attributes formed from a contour analysis of images, and determine the structure and size of feature vectors based on the processing of a digital representation of the contour of a 2-dimensional object projection that would be best suited to the task of recognizing the examined set of types of aerial objects;
- to define the architecture of a complex-valued neural network, develop a training sample, and train the network;
- to develop algorithmic maintenance for an air situation video monitoring system, which includes auxiliary modules to generate a training sample and the main module that solves the task of recognizing types of air objects based on the normalized contour descriptors and a complex-valued neural network.

4. Studying the dependence of changes in the descriptors of the contour of air objects on an angle

4.1. The study input data

The task of analyzing the air situation by means of visual observation is composed of a series of tasks: the detection of a mobile flying object; determining the characteristics of an object (range, size, speed, maneuverability, etc.); the recognition of the type of air objects. These tasks can be addressed both in stages and at the same time. This study assumes that the moving object would first be detected and localized in a video image, and then the task of categorizing it as a certain class could be solved. It is also assumed that the image of an air object can be represented as an ordered sequence of $z(k) = (x_k, y_k)$, $k=0, \dots, N-1$ points that describe the contour.

4.2. Exploring the properties of Fourier descriptors

A one-dimensional Fourier transform of the real function $f(t)$:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (1)$$

makes it possible to derive a continuous spectrum of this function in the frequency domain normally used to analyze time signals. For the case of the discrete sequence N of points $z(k)$ of the examined signal (1) takes the form of a discrete Fourier transform and leads to the calculation of the discrete spectrum:

$$F(n) = \sum_k^{N-1} z(k) e^{-i \frac{2\pi nk}{N}}, \quad (2)$$

where $n = -N, \dots, 0, 1, \dots, N-1$.

In this case, formulae (1) and (2) can be used for the case of complex $f(t)$ and $z(k)$, respectively. The result of the Fourier transform is also complex. It is important that this transform allows for a reverse operation – an inverse discrete Fourier transform:

$$z(k) = \frac{1}{N} \sum_n^{N-1} F(n) e^{i \frac{2\pi nk}{N}}, \quad (3)$$

that restores the original sequence.

The sequence of complex values derived from decomposing (2) is termed the $z(k)$ sequence Fourier descriptors. They possess a series of important properties that explain their use in describing the contours of objects in two-dimensional images: invariant under shift, rotation, and scaling, as well as orderliness.

A two-dimensional contour can be represented by an ordered sequence of the coordinates of its points $\{(x_k, y_k)\}$ and reduced to the complex form $z(k) = (x_k, y_k) = x_k + iy_k$. Assuming that some reference view of the object has been selected, the shift, rotation, and scaling of the object in a video frame can be interpreted as an arithmetic operation involving the Fourier descriptors of the reference image contour. The contour scaling and rotation can be reduced to multiplying by a certain complex number $\tau = r e^{i\varphi}$, where r determines a scaling factor, and φ is the rotation angle.

Indeed, if $z'(k)$ and $z(k)$ are the Fourier descriptors for the transformed and reference contour, respectively, and $z'(k) = \tau z(k)$, then the descriptors are additionally multiplied by the same value of τ :

$$\begin{aligned}
 F'(n) &= \sum_k^{N-1} z'(k) e^{-i \frac{2\pi nk}{N}} = \\
 &= \sum_k^{N-1} \tau \cdot z(k) e^{-i \frac{2\pi nk}{N}} = \tau \cdot F(n) = r e^{i\varphi} \cdot F(n).
 \end{aligned}
 \tag{4}$$

The contour shift can be reduced by adding the contour coordinates to the complex number $\sigma = \sigma_x + i\sigma_y$, where σ_x and σ_y is the image shift along the x and y axes, respectively, that is, $z'(k) = z(k) + \sigma$. Such a transformation leads to a corresponding additive change (shift) of the zero-frequency descriptor:

$$z'(k) = z(k) + \sigma = F(0) + \sigma + \frac{1}{N} \sum_{n \neq 0}^{N-1} F(n) e^{i \frac{2\pi nk}{N}}.
 \tag{5}$$

Thus, all coefficients except $F(0)$ are invariant to the shift, and the coefficient itself, as demonstrated in (2), shows the position of the center of gravity of the contour.

Fourier descriptors are sensitive to the choice of the starting point of the contour during the procedure of a contour's discrete line processing. When the starting point of the contour changes, the phase spectrum of the curve changes, although the amplitude spectrum does not change. One can show that the initial point shift by k_0 results in the following:

$$F'(n) = \sum_k^{N-1} z(k - k_0) e^{-i \frac{2\pi nk}{N}}.
 \tag{6}$$

Let $m = k - k_0$. Then

$$F'(n) = \sum_m^{N-1} z(m) e^{-i \frac{2\pi n(m+k_0)}{N}} = e^{-i \frac{2\pi nk_0}{N}} \cdot F(n).
 \tag{7}$$

This means that the change in the phase spectrum depends not only on the selection of the starting point but also on the descriptor number. This effect makes it difficult to directly use the phase component of the descriptors to identify the contour.

There are also problems with arbitrary affine transformation. Such a transformation is one of the simplest, which roughly describes the distortion of the shape of a flat image of a contour when a 3-dimensional object is at an arbitrary angle relative to the plane of the camera. These cases are the most interesting when recognizing real objects. The affine transformation of a flat shape can be recorded as $c' = Ac + b$, where $c, c' \in R^2$ are the two-dimensional vectors of the coordinates of the object's points before and after the transformation, respectively, A is the non-degenerate matrix 2×2 with constant coefficients ($\det A \neq 0$), b is the two-dimensional shift vector. It is easy to see that because of the linearity of transform (2), the Fourier descriptors are also subjected to the same affine transformation, but, in this case, the module and phase in relation to the reference contour change non-linearly.

The most important useful feature in terms of contour identification is the orderliness of the descriptors by the degree of their importance to the image. It is known that in the frequency spectrum, such as an audio signal, the high-frequency components do not have much influence on the shape of the signal and the sound quality. This makes it possible to discard (zero) the high-frequency components

without losing signal quality. This signal processing process is termed filtering. Similarly, in the sequence of the Fourier descriptors for a flat shape, the information about the shape of the contour is delivered by the first few elements of the sequence.

An inverse Fourier transform makes it possible to assess the degree of recognition of the shape at zeroed high-frequency components. The image contour shown in Fig. 1, a contains about 2,000 points. Accordingly, a Fourier discrete transform would result in the same number of descriptors. The inverse transform would result in an accurate (probably with few computational errors, given the integer representation of pixel coordinates in the image) contour restoration.

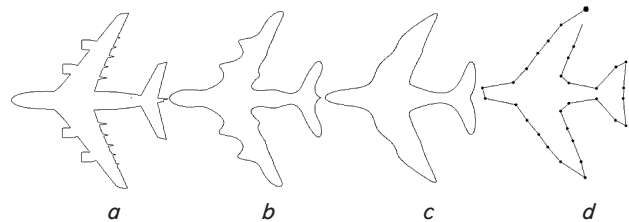


Fig. 1. Illustration of the image contour filtering effect involving a Fourier transform: a – original object contour; b, c – the object contour, restored after zeroing the descriptors of high frequencies, except for 64 and 32, respectively; d – the object's contour, restored by 32 low-frequency descriptors

The zeroing of the descriptors above some of the filter's chosen boundary frequency results in a decrease in the details in the image at inverse transform while the object's recognition is retained. Shifting the filter's boundary frequency to the high-frequency domain leads to an increasingly accurate display of the contour. A shift of the boundary frequency to the low-frequency domain would make the image rougher still, reducing it to the image of an ellipse in the most extreme informative case. Fig. 1, b, c shows the contour restored after zeroing all but 64 and 32 low-frequency descriptors, respectively. Fig. 1, d shows an image obtained for the same original contour by discarding all descriptors except 32 for the rapid derivation of only 32 points of the restored image with linear interpolation between them.

4. 3. Choosing a feature vector

In a general case, the Fourier descriptors $F(n)$ can accept different values for different images of the flat contour of the same object. Their magnitudes depend on the scale r , the rotation angle φ , and the choice of the contour starting point k_0 , as shown by (4) and (7).

Assuming that some reference value of the descriptors $F^*(n)$ has been selected, the calculated $F(n)$ sequence for the recognized object can be represented as follows:

$$F(n) = r e^{i\varphi} e^{-i \frac{2\pi nk_0}{N}} \cdot F^*(n).
 \tag{8}$$

One can convert the Fourier descriptors to a form that lacks the influence of these factors. Consider the normalized descriptors according to [22]:

$$N(n) = \frac{F(1+n)F(1-n)}{F^2(1)}.
 \tag{9}$$

It is then possible to show that:

$$N(n) = \frac{F^*(1+n)re^{i\theta}e^{-\frac{2\pi(1+n)k_0}{N}} F^*(1-n)re^{i\theta}e^{-\frac{2\pi(1-n)k_0}{N}}}{\left(re^{i\theta}e^{-\frac{2\pi k_0}{N}} F^*(1) \right)} = \frac{F^*(1+n)F^*(1-n)}{(F^*(1))^2} = N^*(n), \tag{10}$$

where $N^*(n)$ are the normalized coefficients, corresponding to the reference set of the Fourier descriptors $F^*(n)$.

It is easy to see that the normalized descriptors $N(n)$ (9), unlike Fourier descriptors, are independent of the above factors and can serve as an adequate sequence to describe the reference image of the contour.

4. 4. Computational experiment to investigate the descriptors of the contour of a 2-dimensional projection of an air object in an arbitrary position

We have investigated the dependence of Fourier descriptors (2) and the normalized descriptors (9) on the angles of spatial position for four different types of aerial objects: aircraft, unmanned aerial vehicle, helicopter, quadcopter. In accordance with the filtering properties of a Fourier transform shown in Fig. 1, it was determined that the first 32 descriptors would suffice to represent the shape of the air objects being examined to solve a classification problem.

One can see (Fig. 2) that even similar profiles of the aircraft and helicopter have different «patterns» in terms of the amplitude component for their Fourier descriptors. The phase component, as discussed above, is not representative because of the uncertainty of choosing the starting point of the contour. However, this difference is especially noticeable for normalized descriptors where both the amplitude and the phase component are important.

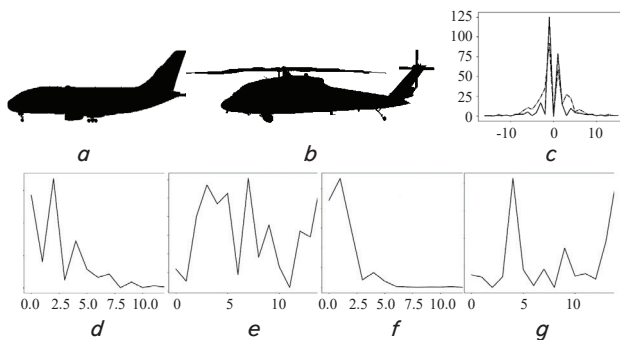


Fig. 2. Difference in the shape of the objects represented by their Fourier descriptors and normalized descriptors: *a* – a binary image of the A380 aircraft profile; *b* – the same for the Apache helicopter; *c* – amplitude characteristics of Fourier descriptors (solid line – for A380, dotted line – for Apache); *d*, *e* – the amplitude and phase characteristics of the normalized descriptors for A380; *f*, *g* – the same for the normalized descriptors for Apache

When recognizing the type of any three-dimensional object in a two-dimensional image, there is an issue of the mutual location of the object and the video camera, that is, the angle of the image. Regardless of the classification method, whether it is a method of comparison with the reference or classification methods involving neural networks, the

proximity of the view of the contour of the real image (or describing its descriptors) and some reference remains important. It is obvious that for most objects that do not have specific symmetry properties, images and corresponding descriptors can vary greatly from different angles. This is a particularly important issue for aerial objects that have the freedom to rotate around any of the 3 axes in three-dimensional space.

Specific terms are typically used for aircraft rotation angles. The rotation around the longitudinal axis when one wing falls and the other rises is termed a roll. The rotation around the vertical axis at which the aircraft turns the nose left or right is termed yaw. The rotation around the transverse axis when the plane lowers and lifts the nose is called a pitch.

There are a series of simplifications. One can assume that the video plane is parallel to the *XY* plane of the system of coordinates of the aircraft (the *x* axis coincides with the flight direction; the *y* axis passes along the wings (Fig. 3)). At the same time, the aircraft can be simplified to present as a model in the form of flat surfaces (Fig. 3, *a*). Then it is obvious that the roll of the aircraft (rotation around the *x* axis) would result in an easy computed change in the contour of the wing projection in the image. In this case, for each point of this contour relative to the contour at a zero roll, the *y*-th coordinate would change proportionally to the cosine of the roll angle.

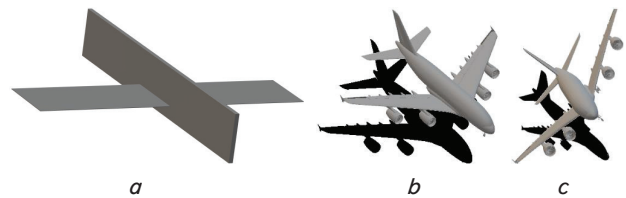


Fig. 3. Change in the aircraft projection when its position changes: *a* – model representation of the aircraft by two planes; *b* – an aircraft projection at a zero roll; *c* – change in the aircraft projection at roll

One can more accurately describe the change in the coordinates of a rotating object. If one designates via α and β a rotation angle along the *x* and *y* axes, respectively, the 3-dimensional transformation matrices can be described as follows:

$$T_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix}, T_\beta = \begin{pmatrix} \cos \beta & 0 & -\sin \beta \\ 0 & 1 & 0 \\ \sin \beta & 0 & \cos \beta \end{pmatrix}. \tag{11}$$

Then

$$T_{\alpha\beta} = T_\alpha \cdot T_\beta = \begin{pmatrix} \cos \beta & 0 & -\sin \beta \\ -\sin \alpha \sin \beta & \cos \alpha & -\sin \alpha \cos \beta \\ \cos \alpha \sin \beta & \sin \alpha & \cos \alpha \cos \beta \end{pmatrix}. \tag{12}$$

And since of interest is a change in the *x* and *y* coordinates, one can finally record:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \beta & 0 \\ -\sin \alpha \cos \beta & \cos \alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{13}$$

where x' and y' are the coordinates of the new position of the arbitrary point of the object at coordinates x and y .

If we follow the assumption that an object can be represented by flat surfaces, and, when the angles change, the projection contour is still determined by the faces of these

surfaces, this transformation can roughly represent a change in the observed contour in the image.

If one uses the matrix version of ratios (7) and (9), it is easy to show that due to the linearity of a Fourier transform, the proposed normalized descriptors (9) do not depend not only on such a change of angles (10) but, in general, on an arbitrary affine transformation.

Studying the dependence of Fourier descriptors (2) and normalized descriptors (9) on the spatial angles for the examined air objects has shown that when the angle of the roll increases from 0 to 90 degrees, the side image (profile projection) smoothly transitions into the image from below (horizontal projection) (Fig. 4), while the Fourier descriptors and the normalized descriptors are smoothly transformed into the horizontal projection descriptors.



Fig. 4. Change in the shape of the A380 aircraft’s projection when the angle of the roll changes from 0° to 90°

However, in practice, the representation of an aircraft in the form of flat surfaces is too rough; the descriptors of different frequencies perform differently during this transformation. Table 1 gives the results of a change in the amplitude of the first few normalized descriptors.

Similar studies have been conducted for other types of aircraft (Fig. 5).

It is generally accepted that the sequence of Fourier descriptors is invariant not only to the change in scale and displacement but also to the rotation of the yaw if the shooting is carried out from the surface of the earth, as this rotation can be compensated for by the argument of a complex multiplier. And only the roll and pitch influence the change in the two-dimensional projection of the three-dimensional silhouette of an aircraft.

Our analysis revealed that in the case when the shooting point is not directly under the plane, even yaw is not a pure image rotation, which can be compensated for when processing the descriptors.

Changing the yaw angle distorts the aircraft contour as the silhouette is determined by the projection of a 3-dimensional curve – the boundary of the observation cone. And the simplistic idea of changing the shape of the contour when one changes the angle of rotation according to the affinity transformations is too rough.

Thus, even in the trivial case, if one studies the contour of a cube or parallelepiped, at a yaw rotation the quadrangle contour can change to hexagonal.

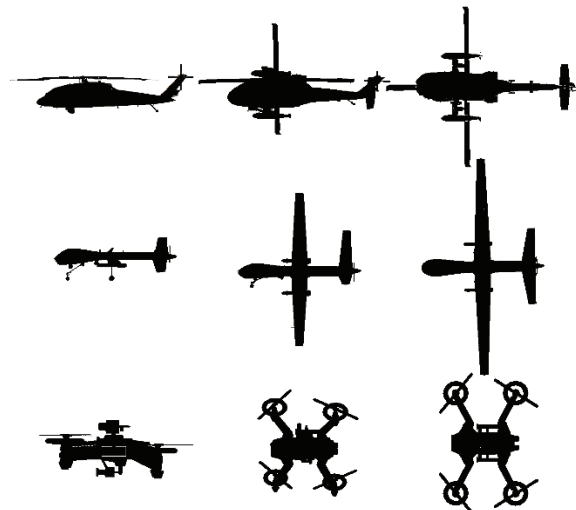


Fig. 5. Change in the projection shape for the helicopter, UAV, and quadcopter

The result of studying the contours of 4 types of objects: passenger plane, helicopter, unmanned aerial vehicle, quadcopter has established that the normalized descriptors $N(n)$ (9) even for low frequencies are sensitive to changes in the position of the object in space. That necessitates the use of separately normalized descriptors for each of the three orthogonal projections of the object as class references.

Therefore, three orthogonal projections as separate classes are used when solving a classification task and building a training set in this study, although this increases the number of classes to 11 for the 4 selected types of objects (the quadcopter has two projections that match).

Table 1

Change in the amplitude of the normalized contour descriptors when the angle of the A380 aircraft roll changes

Roll angle	0°	20°	40°	60°	90°
Amplitude of the first few low-frequency descriptors	0.07135509	0.04448492	0.01938495	0.00362799	0.00180103
	0.00028499	0.00136822	0.00125024	0.00160868	0.00307285
	0.00008098	0.00202573	0.00415826	0.00566903	0.00723830
	0.00033568	0.00129310	0.00155052	0.00151954	0.00264811
	0.00002451	0.00018371	0.00073380	0.00074591	0.00054625
	0.00014459	0.00047944	0.00052269	0.00094067	0.00106715
Amplitude charts starting at 2					

It should be taken into consideration that both the amplitude and the phase components are stable for the normalized descriptors, which makes the complex-valued neural network an adequate classification module.

5. Studying complex-valued neural networks in the task of recognizing air objects by contour descriptors

5. 1. Complex-valued neural networks in a task of multi-dimensional classification

The development of complex neural networks has recently shown good results in working with complex real signals that have a phase and amplitude (for example, images or audio signals). Therefore, it is natural to study such networks to analyze data that are complex numbers in the sequence of Fourier descriptors for the contour of an object.

The mathematical model of a complex-valued neuron can be built in accordance with the classical real analog:

$$y = S\left(\sum_{j=1}^n z_j w_j + w_0\right), \tag{14}$$

where S is the activation function applied to the linear combination of the input vector $\mathbf{z} \in C^n$ and the integrated weights vector $\mathbf{w} \in C^n$ with a $w_0 \in C$ shift.

This change in the model leads to new properties of its performance. Even a single neuron does not just scale the input vector, as in a real case. It turns it around, due to a well-known geometric interpretation of complex multiplication. And this cannot be achieved by a single neuron with a real calculus if one represents a complex number by a pair of real values. In the real case, to rotate a two-dimensional vector, you would have to use multiplication by a 2x2 transformation matrix.

There are two fundamentally different approaches to determining the activation function.

That leads to issues related to differentiation when building training using gradient descent.

The first approach to determining the activation function [17] follows the classic scheme with a sigmoid function or hyperbolic tangent but applied separately to the real and imaginary part. The second approach [21] employs a fundamental property of rotating complex multiplication and introduces the concept of a multi-valued neuron (MVN) with an activation function that maps a linear combination of the input signal with weight coefficients:

$$net = \sum_{j=1}^n z_j w_j + w_0 \tag{15}$$

onto a point at a unit circle:

$$S(net) = e^{i \arg(net)} = \frac{net}{|net|}. \tag{16}$$

To avoid the redundant term w_0 in designations (15) and to be able to consider the weighted sum as a scalar product of two $(n+1)$ -dimensional vectors, one can write down:

$$net = \sum_{j=0}^n z_j w_j = \mathbf{z} \cdot \mathbf{w}, \tag{17}$$

assuming the use of an extended input $n+1$ -dimensional vector \mathbf{z} at $z_0 = 1$.

In the case of the discrete activation function, the mapping is rendered onto one of the k sectors in a unit circle:

$$S(net) = e^{i \frac{2\pi j}{k}} \equiv \varepsilon_k^j, \quad \frac{2\pi j}{k} \leq \arg(net) < \frac{2\pi(j+1)}{k}, \tag{18}$$

where $\varepsilon_k = e^{i \frac{2\pi}{k}}$ is the main value; the set $\{1, \varepsilon_k, \varepsilon_k^2, \dots, \varepsilon_k^{k-1}\}$ forms a complete group of the complex root of power k from unity. The selection of k determines the order of the so-called « k -values threshold function», which makes it possible to solve classification problems with k classes by comparing the circle sector to one of the classes. The activation function (18), as in the first approach, is not differentiated, which does not make it possible to use gradient descent as a training procedure. Instead, two non-differentiated learning rules are proposed, which is another advantage of this approach.

Inference of one of the two rules is based on the following considerations. Assuming that \mathbf{w}^* is the proper value of weights, the *err* error in the value of the argument would be:

$$err = \sum_{j=0}^n z_j w_j^* - \sum_{j=0}^n z_j w_j = \sum_{j=0}^n z_j \Delta w_j, \tag{19}$$

which means that each component of the vector $\Delta \mathbf{w} = \mathbf{w}^* - \mathbf{w}$ contributes to the overall error. Assuming (for the lack of more accurate information) that each component makes an equal contribution, we obtain:

$$z_j \Delta w_j = \frac{err}{n+1}. \tag{20}$$

Then one can obtain a simple ratio:

$$\Delta w_j = \frac{err}{n+1} z_j^{-1}, \tag{21}$$

that, given that $|z_j| = 1$, and, therefore, $z_j^{-1} = \bar{z}_j$ of the conjugated value, we finally obtain the learning rule:

$$\Delta w_j = \eta \frac{err}{n+1} \cdot \bar{z}_j, \tag{22}$$

where η is the «learning speed» factor, as in the real case, used to control the adjustment step.

The proposed MVN learning algorithms are not considered to be a task of minimizing error functionality. For the same reason (the absence of the minimization task), the training does not face the problem of getting stuck in local minima, which is typical for the gradient optimization rules of training in a real case.

Based on the mathematical model of a single complex multi-valued neuron, it is possible to build multi-layered neural networks (MLMVN) with the rule of learning using a method of error backpropagation without a gradient descent.

5. 2. Training a complex-valued neural network

To train a neural network, we used grid three-dimensional models of air objects, generated in the editor of 3-dimensional graphics. The prepared training set included images to recognize 11 classes of images for 4 types of air objects: aircraft, UAV, helicopter, quadcopter, for 3 classes for each type according to three orthogonal projections (the quadcopter has two projections that match).

For each orthogonal projection, which corresponds to the direction along one axis, we used a variation of rotation angles around the other two, ranging from -45° to $+45^\circ$.

Training dynamics metrics were applied to assess the quality of classification: training accuracy and validation accuracy.

Training accuracy is the accuracy of image classification, which the neural network used for training; validation accuracy is the accuracy on images that were not used to train the neural network. Therefore, validation accuracy is a more reliable measure of how accurate the model is.

The architecture of a multi-layered neural network with multi-values coding of neurons MLMVN has a topology of 15-100-1, that is, 15 input neurons, 100 neurons in one hidden layer, and an output layer with one output neuron, the complex output of which is divided into 11 sectors ($k=11$). As a set of reference outputs, we apply $R_k = \{1, \varepsilon_k, \varepsilon_k^2, \dots, \varepsilon_k^{k-1}\}$, where $\varepsilon_k = e^{i\frac{2\pi}{k}}$.

All neurons have continuous ins and outs. Therefore, (16) is used as an activation function. As the initializing values of all neurons (real and imaginary parts), random numbers were taken from the interval $[0, 1]$. MLMVN training is based on an error backpropagation algorithm that differs from the classic method for real networks in that there is no need to calculate the gradient to change the weights.

The training process is as follows:

1. Perform the procedure of direct transmission of the signal. To this end, for each input vector from the training sample, first for the neurons from the hidden layer, one calculates, according to (15), a weighted sum, and then the outputs according to (16). After that, one performs the same procedure for a single output neuron.

2. Form the value of a network error. To this end, after passing through the network of each training vector, one forms a partial error:

$$err_j = \left(\arg \varepsilon^{\alpha_j} - \arg \varepsilon^\alpha \right) \bmod 2\pi, \quad (23)$$

where ε^{α_j} is the expected output for the j class; ε^α is the output obtained as a result of the direct transmission of the signal. After a batch of 32 instances of training vectors, the total batch error err is calculated as the rms value of all partial errors.

3. Update the weights of the neural network. To this end, according to (21), one determines separately the value for adjusting the weights of the output neuron and the neurons from the hidden layer. In the error backpropagation procedure, the weight adjustment is performed in accordance with the principle of equal separation of responsibility for error between neurons [21]. This means that the error must be multiplied by the inverse values of the corresponding weight. This is an important difference between the MLMVN training procedure and the classical backpropagation algorithm for real neural networks.

4. Terminate a training procedure. The criterion for stopping is either a set number of learning epochs or an achievement of the predefined error.

The process of assessing the quality of classification and error during the training is shown in Fig. 6.

The classification accuracy chart confirms the correctness of the proposed approach to solving the problem of recognizing the types of air objects.

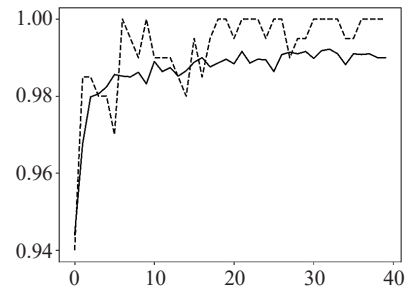


Fig. 6. Assessment chart of the accuracy of classification during training (dashed line – the score on the training set, solid line – on the validation set)

6. Algorithmic maintenance of an air object recognition system

The recognition task solved in this work is an integral part of the hardware and software complex of detecting moving aerial objects and determining their characteristics. The algorithmic maintenance, suggested as a study result, includes an auxiliary module to train a neural network at the preparation stage and the main module. An auxiliary module is needed to form a training set for neural network training. In addition, during the study phase of the behavior of contour descriptors, it was necessary to have an extensive set of model images of air objects from different angles. To solve this task, a tool has been developed for computer-generated images of aerial objects based on grid 3D models, employing the three-dimensional animation library.

The support module software’s algorithm is as follows:

1. The initial position of an air object model is set, corresponding to one of its three orthogonal projections. This fixes one of the three rotation axes.

2. Change the position of the image around the other two axes in the predefined range at the preset angle step.

3. A two-dimensional image on the screen is copied to an external file with a name that includes information about the angles of the current position.

For a given task, we used three-dimensional images of various models of air objects to determine one of the classes: aircraft, UAVs, helicopter, quadcopter.

The main module is designed to recognize types of air objects in real time as part of the hardware and software complex for detecting moving air objects. In this module, the recognition task is solved in such a way that in the first stage an object or objects must first be detected as moving in video sequences and localized in video images. In the second stage, the task of categorizing each object on the frame to a certain class is solved.

To obtain the contour’s numerical characteristics, a localized image (Fig. 7, *a*) must first be brought to a black-and-white view (Fig. 7, *b*). One then needs to highlight a set of the contour points (Fig. 7, *c*) and represent it as a sequence. This problem, while not trivial, can be solved by standard image processing tools included in the OpenCV library.

The result of our analysis of different approaches to the task of air object recognition described above is the following proposed algorithm for solving it:

1. Download a color image and highlight a rectangle covering the region of interest (in OpenCV terms) based on the coordinates determined from the task of detecting a moving object.

2. Transform it into a black-and-white image with an adaptive cut-off threshold, which is calculated according to the Gaussian function.

3. Search for the object's contours under the mode of object external contour search as an array of two-dimensional points. The result of this processing is the image of an air object represented as an ordered sequence of points, which is a discrete representation of a continuous two-dimensional curve describing the contour.

4. Derive a sequence of Fourier descriptors for the contour and truncate it to 32 complex values.

5. Generate 15 complex normalized descriptors.

6. Solve a classification problem using a pre-trained complex-valued neural network that assigns a 15-dimensional complex vector to one of 11 classes. Each of the four types of aerial objects is matched with 3 contour classes for 3 orthogonal projections, except for a quadcopter whose two projections match.

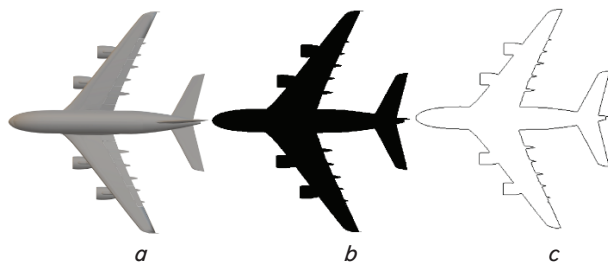


Fig. 7. Image pre-processing steps: *a* – the original image of a localized object; *b* – black-and-white image; *c* – a contour image of an object

7. Discussion of results of studying the proposed method for recognizing the type of images of air objects

Our results of the convergence of the training procedure (Fig. 6) indicate that the proposed approach to the selection of normalized contour descriptors as the vectors of object attributes in order to solve the task of recognizing the types of air objects using a complex-valued neural network makes it possible to build an effective recognition subsystem for the complex of moving air object detection. Indeed, Fig. 6 shows that the validation accuracy achieved on images that were not used to train the neural network reaches 99 %. It should be noted, however, that the effectiveness of solving a classification problem is attained by dividing the training set into classes that correspond to 3 different projections of the object. This leads to a simplification of the training procedure as the neural network does not have to attribute to the same class different images of the same object in different angles and, therefore, having very different multidimensional features, as it follows from Table 1. In addition, the choice of a complex-valued neural network makes it possible to conduct the training procedure without the use of gradient methods, which does not lead to the effect of getting stuck in local minima. This is due to the choice of network architecture, built on a complex calculus, which is most consistent with the representation of feature vectors in a complex space. This representation is dictated by a comprehensive form of a Fourier transform method used to calculate normalized descriptors. The application of normalized descriptors avoids the issues commonly encountered in the contour analysis of Fourier descriptors, depending on factors such as distortion of the shape

of the object's contour due to the mutual spatial location of the video camera and aerial objects, as well as the choice of the initial point of the contour when processing the image. That is why the proposed variant of air object class recognition, described in chapter 6, should be considered promising.

The proposed approach to solving the task of recognizing digital images of air objects, based on the use of normalized contour descriptors as the feature vectors of an object, in conjunction with the application of a complex-valued neural network, has the advantage that it avoids the use of a heavy apparatus of deep neural networks, which is typically used to solve this problem.

The application of the proposed method is associated with a series of limitations in terms of its practical use. For example, solving the task implies, first, that an air object is large enough to detect a contour that can be used to calculate 15 normalized descriptors. Second, it is assumed that the image of the object is not distorted by fog or partially closed by clouds. The removal of these restrictions could be partially implemented by the pre-processing of the image and is the subject of further research.

It should be noted that the use of a complex-valued neural network to solve the recognition task is a new direction in this subject area and requires additional research. Another promising direction is research aimed at calculating the parameters of affine transformation and determining the real spatial location of air objects on a separate frame based on the derived descriptors, which would make it possible to determine the nature of their maneuver in a video sequence.

8. Conclusions

1. A computational experiment involving model images was performed to investigate the dependence of change in Fourier descriptors and the normalized contour descriptors of images of 4 types of air objects on the angle of the object's rotation relative to 3 axes. It is shown that the use of normalized descriptors has a series of advantages over the Fourier descriptors. It has been established that to simplify the recognition task, one needs to parse the training set for each type of object into 3 classes corresponding to 3 orthogonal projections. This makes it easier to solve the classification problem owing to a more compact arrangement of multidimensional feature vectors for shapes with similar images.

2. We have studied the possibility of using a complex-valued neural network to solve the task of recognizing the types of air objects. The study results have made it possible to propose the configuration of a multi-layered neural network with multi-valued coding of neurons MLMVN with 15 input neurons, 100 neurons in one hidden layer, and one output neuron. The validation accuracy of recognition is 99 % at the training stage.

3. Algorithmic maintenance for an air situation video monitoring system has been developed. Imaging techniques that pre-process the image in the following sequence have been investigated. The examined localized image of the object is reduced to a binary view (a black-and-white image) to make it easier to detect its contour. The object's contour is highlighted in the form of a sequence of the coordinates of points. A Fourier transform of this sequence is calculated, followed by the selection of 32 low-frequency Fourier descriptors. They are used to compute 15 complex-valued normalized descriptors, which are a numerical representation

of a two-dimensional contour suitable for use in the classification problem. The task of recognizing the type of an air object is tackled by a pre-trained neural network with a specific architecture based on complex calculus. The validation detection accuracy is as high as 99 %. This confirms that the proposed method of building a system for recognizing the

types of air objects could simplify the requirements for the implementation of hardware while improving the accuracy when solving a task of air situation recognition. That, in turn, creates the preconditions for the increased mobility and extended scope of application of such systems, including individual detection devices.

References

1. Strotov, V. V., Babyan, P. V., Smirnov, S. A. (2017). Aerial object recognition algorithm based on contour descriptor. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W4, 91–95. doi: <https://doi.org/10.5194/isprs-archives-xlii-2-w4-91-2017>
2. Costa, L. da F., Cesar, Jr., R. M. (2018). *Shape Classification and Analysis*. CRC Press, 685. doi: <https://doi.org/10.1201/9781315222325>
3. Hirose, A. (Ed.) (2013). *Complex-Valued Neural Networks*. Wiley. doi: <https://doi.org/10.1002/9781118590072>
3. Sharma, N., Jain, V., Mishra, A. (2018). An Analysis Of Convolutional Neural Networks For Image Classification. *Procedia Computer Science*, 132, 377–384. doi: <https://doi.org/10.1016/j.procs.2018.05.198>
4. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60 (6), 84–90. doi: <https://doi.org/10.1145/3065386>
5. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. et. al. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2015.7298594>
6. Mash, R., Becherer, N., Woolley, B., Pecarina, J. (2016). Toward aircraft recognition with convolutional neural networks. 2016 IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovatoin Summit (OIS). doi: <https://doi.org/10.1109/naecon.2016.7856803>
7. Yesilevskiy, V., Teviashev, A., Koliadin, A. Transfer learning in aircraft classification. Available at: https://openarchive.nure.ua/bitstream/document/11942/1/3_IST.pdf
8. Bradski, G., Kaehler, A. (2008). *Learning OpenCV. Computer Vision with the OpenCV Library*. O'Reilly Media, 555.
9. Chaki, J., Dey, N. (2019). *A Beginner's Guide to Image Shape Feature Extraction Techniques*. CRC Press, 152. doi: <https://doi.org/10.1201/9780429287794>
10. Yang, M., Kpalma, K., Ronsin, J. (2012). Shape-Based Invariant Feature Extraction for Object Recognition. *Intelligent Systems Reference Library*, 255–314. doi: https://doi.org/10.1007/978-3-642-24693-7_9
11. Rong, H.-J., Jia, Y.-X., Zhao, G.-S. (2014). Aircraft recognition using modular extreme learning machine. *Neurocomputing*, 128, 166–174. doi: <https://doi.org/10.1016/j.neucom.2012.12.064>
12. Makarov, M. A., Berestneva, O. G., Andreev, S. Yu. (2014). Solving the problem of moving objects contour classification and recognition on video frame. *Izvestiya Tomskogo politehnicheskogo universiteta*, 325 (5), 77–83.
13. Nguen, T. T. (2010). Algoritmicheskoe i programmnoe obespechenie dlya raspoznavaniya figur s pomoshch'yu Fur'e-deskriptorov i neyronnoy seti. *Izvestiya Tomskogo politehnicheskogo universiteta*, 317 (5), 122–125.
14. Aizenberg, N. N., Ivaskiv, Yu. L., Pospelov, D. A. (1971). A certain generalization of threshold functions. *Dokl. Akad. Nauk SSSR*, 196 (6), 1287–1290. Available at: http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=dan&paperid=35992&option_lang=eng
15. Guberman, N. (2016). On Complex Valued Convolutional Neural Networks. *arXiv.org*. Available at: <https://arxiv.org/pdf/1602.09046.pdf>
16. Nitta, T. (2011). Ability of the 1-n-1 Complex-Valued Neural Network to Learn Transformations. *Computational Modeling and Simulation of Intellect*, 566–596. doi: <https://doi.org/10.4018/978-1-60960-551-3.ch022>
17. Mönning, N., Manandhar, S. (2018). Evaluation of Complex-Valued Neural Networks on Real-Valued Classification Tasks. *arXiv.org*. Available at: <https://arxiv.org/pdf/1811.12351.pdf>
18. Aizenberg, I. (2011). *Complex-Valued Neural Networks with Multi-Valued Neurons*. Springer. doi: <https://doi.org/10.1007/978-3-642-20353-4>
19. Fajjul Amin, M., Murase, K. (2009). Single-layered complex-valued neural network for real-valued classification problems. *Neurocomputing*, 72 (4-6), 945–955. doi: <https://doi.org/10.1016/j.neucom.2008.04.006>
20. Aizenberg, I., Moraga, C. (2006). Multilayer Feedforward Neural Network Based on Multi-valued Neurons (MLMVN) and a Back-propagation Learning Algorithm. *Soft Computing*, 11 (2), 169–183. doi: <https://doi.org/10.1007/s00500-006-0075-5>
21. Granlund, G. H. (1972). Fourier Preprocessing for Hand Print Character Recognition. *IEEE Transactions on Computers*, C-21 (2), 195–201. doi: <https://doi.org/10.1109/tc.1972.5008926>