Some features of creating a computer vision system

https://doi.org/10.31713/MCIT.2021.22

Petrivskyi Yaroslav
Further Mathematics Department
Rivne State Humanitarian University
Rivne, Ukraine
prorectorsgu@ukr.net

Petrivskyi Volodymyr
Program system and technologies Department
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
vovapetrivskyi@gmail.com

Oleksiy Bychkov
Program system and technologies Department
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
bos.knu@gmail.com

Oleksandr Pyzh
Program system and technologies Department
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
iryna.888@ukr.net

***Abstract* — In the paper some features of models and algorithms of computer vision are presented. An algorithm for training the neural network of object recognition is proposed and described. The peculiarity of the proposed approach is the parallel training of networks with the subsequent selection of the most accurate. The presented results of experiments confirm the effectiveness of the proposed approach.**

***Keywords* — computer vision; objects identification; neural networks; differential equations systems; principal component analysis.**

I. Introduction

According to the level of modern technological progress, more and more problems are solved using neural networks. One such task is the task of computer vision or computer identification, which is to identify people, objects, and phenomena of various natures using the camera of any device [1]. A key aspect of this task is the selection of the most accurate neural network training algorithm. Of particular interest is the identification of objects in real time.

In recent years, computer vision has become a key technology in many areas of science, technology, engineering, society as a whole.It is impossible to imagine the automation of industrial production without the use of computer vision technology for quality control and simplification with an increase in the accuracy and quality of technological processes.

It is quite informative to get acquainted with the history, development and current state of the issue of theoretical and applied aspects of computer vision technology, for example, in works where, together with a description of the mathematical apparatus, all other components of computer vision systems integral to the development and operation are shown quite fully. The theoretical aspects of processing visual data are considered using a large number of examples from practical problems [2–8]

Along with classical topics – description of scene images and pattern recognition, Bayesian classification, pattern matching, camera calibration and perception of three-dimensional scenes, the books deal with the issues of image databases and systems of virtual and augmented reality, image search based on content, image segmentation, three-dimensional models, recognition of objects in images based on models. Examples of applications in industry, medicine, land use, multimedia and computer graphics are given.

Thus, the combination of mathematical apparatus in combination with computer technology and the development of appropriate software for solving problems of computer vision and computer identification is an urgent problem today.

II. Main part

In the world and national practice of designing and developing computer vision systems, a three-level sequence of research has been selected. This architecture is based on a bionic approach to the human perception process [9–12].

Within the framework of this architecture, images go through three stages of transformations: low, medium and high. In this case, the following tasks are sequentially solved in the indicated order:

In general, the process of creating a recognition algorithm can be represented in the form of the following sequential operations: scanning followed by transformation of the original object; extraction of representative parameters of the object from the data stream; identification of the object according to the selected classification criteria.

1) input of information, i.e. obtaining an image of the working scene using sensors;

2) image preprocessing using noise suppression techniques;

3) segmentation, i.e. highlighting one or more scene objects of interest in the image;

4) description, i.e. determination of the characteristic parameters of each object necessary for its selection on the scene;

5) recognition, or identification, of an object, i.e. establishing it belonging to a certain class;

6) interpretation, i.e. identifying the belonging of an object to a group of recognizable ones or compiling the semantics of a scene.

The intensive development of the technical capabilities of computer technology (large memory, high-bandwidth communication channels, and multimedia computer systems) made it possible to develop, create and process arrays of large databases, including image databases.

Modern video systems make it possible to implement the initial stage of classification quite well.

The problem of detecting the desired characteristic features of objects in images and video of the scene of observation is an important applied problem and represents the relationship of two classification problems, namely, the segmentation of the object as a whole, and the recognition of the class to which the object belongs. At the moment, there are many approaches and developments, methods and technologies for solving segmentation and recognition problems. Segmentation is a problem that is addressed at the middle and upper levels of the three-tier architecture of the computer vision system. At the same time, recognition remains a priority task only at the top level of this architecture. Segmentation tools include methods of grouping, clustering, classifying and categorizing pixels, lines and regions (patterns). Most of them are based on the principle of the feature approach, when a predetermined set of features is measured and the structural elements are divided into groups. Segmentation at the upper level of the computer vision system is more aimed at measuring the relationship between previously identified structural and geometric elements, which are used along with features. At the top level of the computer vision system, solving the segmentation problem is more focused on recognition than just dividing into groups.

At the same time, the process of fixing its key characteristics in the data stream about the object of classification remains quite important. For this purpose, in our studies, we have proposed and successfully implemented an algorithm for the principal component analysis method (PCA) in the classical formulation of Pearson about the problem of approximating a finite set of data in the absence of a hypothesis of their statistical distribution. The statistical PCA method is determined by a linear transformation, which allows you to transform a stationary stochastic process, represented by the original vector, into a vector using a matrix at $K < N$ so that the resulting reduced size space preserves the most important information about the original data of the object. Maximum variance taking into account their deviations, namely, the first independent component determines the direction in which the maximum possible deviation is recorded, the second independent component determines (in the flow orthogonal subspace) the direction of the maximum variance, etc. Thus, the transformation by the PCA method allows replacing a large amount of information mutually correlated input data, a set of statistically independent components taking into account their significance.In communication (coding) theory, this method is known as the Karjunen-Leve transformation, which allows compression with loss of information [13].

It is known that a neural network is a multidimensional function whose arguments belong to the space of the domain of definition, and the result to the space of the domain of values. Components - a set of neurons, perform the operation of summing signals coming from other neurons that have gone through the corresponding transformation. In the general case, dependencies are approximated that are functions of many arguments, where the sum of functions is performed, each of which is a function of only one argument. The Hecht-Nielsen theorem generalized the Kolmogorov-Arnold theorem that any continuous function of many variables $f(x_1, x_2, \dots, x_n)$ can always be represented as a sum of continuous functions of one variable and approximated using a two-layer neural network with lines complete connections with n neurons of the input layer, (2n +1) neurons of the hidden layer with previously known limited activation functions) and m neurons of the output layer with unknown activation functions [14]. Also, when creating a neural network, an important task is to find the optimal size of the network - such a number of hidden layers and neurons in the layers that will give a minimum of generalization error. In our case, two approaches were used: theoretical - the result of the Arnold-Kolmogorov and Hecht-Nielsen theorems, practical, where the dependences of learning and generalization errors on the size of the neural network were used, and the optimal value corresponds to a local minimum.

The required number of neurons in the hidden layers according to the theorems of Arnold-Kolmogorov and Hecht-Nielsen was determined by the formula [15]:

$$\frac{N_y Q}{1+log_2(Q)} \le N_w \le N_y K + N_y, \qquad (1)$$

where N_y – dimension of the output signal y, Q – number of training examples, N_w – the required number of synaptic connections, $K = \left(\frac{Q}{N_x} + 1\right)\left(N_x + N_y + 1\right)$, N_x – dimension of the input signal x.

According to the Hecht-Nielsen theorem, sigmoidal transfer functions with adjustable parameters are used in neural networks for both the first (hidden) and second (output) layers.

Having estimated the required number of synaptic connections N_w using formula (1), the required number of neurons in the hidden layers was calculated, which is equal to:

$$N = \frac{N_w}{N_x + N_y}. \qquad (2)$$

A fairly large number of applied software developments are known that implement the concept of a neural network algorithm, which allows a wide range of tasks.

The developed identification algorithm can be divided into two parts: the choice of model and the identification process. In turn, part of the model selection consists of pre-processing the data, training the models, selecting and saving the required model. Under pre-processing of data we understand preparation of a set of photos of necessary subjects with the subsequent sorting on folders with the name of a subject. The next stage is the learning process that takes place in parallel. The use of parallel calculations speeds up the production of trained models by 67%. After receiving the trained models, the required model is selected and saved. The selection criterion is the accuracy of identification. The selection is made automatically by the software or by an expert or a group of experts. The saved model is loaded into the appropriate item identification software. A feature is the ability to use an already saved model without prior training. A prerequisite for using the software is a connected camera. The image read from the camera is transmitted to the identifier for further identification. The result is a corresponding message identifying the item with the percentage of compliance. A feature of the described software is network training on all known models in parallel. Schematically described above, the software algorithm can be presented in the following form (Fig. 1):

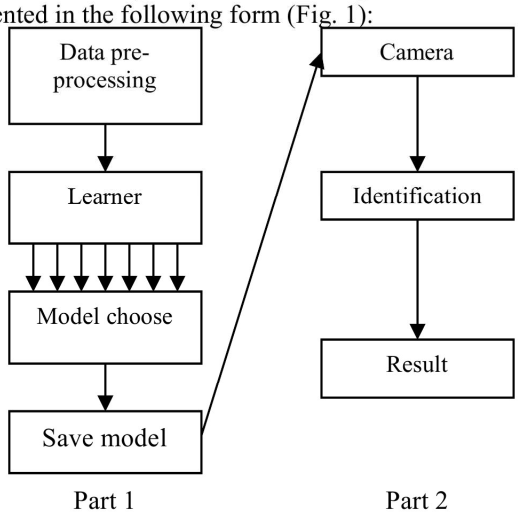

Figure 1. Objects identification algorithm

III. COMPUTER SIMULATION RESULTS

During the computer experiment, the identification of 2 objects was studied: a pen and glasses. The input data set consisted of 300 photographs of each object. The process of learning and choosing a model lasted 5 minutes 47 seconds. Of all the models, the classification tree and fine tree were the most accurate. The identification results are presented in the following figures (Fig. 2–4):

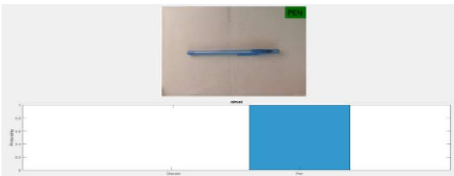

Figure 2. Computer simulation results

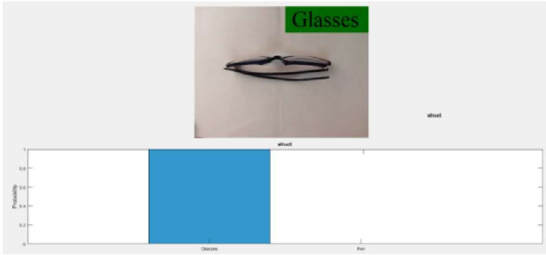

Figure 3. Computer simulation results

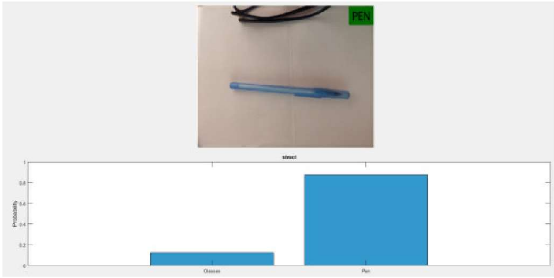

Figure 4. Computer simulation results

IV. CONCLUSION

Thus, we can conclude that the use of mathematical apparatus in combination with modern information technology can effectively solve the problem of object recognition. developed software with the proposed algorithm using parallel computing technologies accelerates the total training time of the network by 58%.

REFERENCES

[1] A. Zeyegh and N. Bassam, “Neural Network Principles and Applications,” Digital Systems, 2018.

[2] L. Shapiro and D. Stockman, “Computer vision,” Binom, 2013, 752 p.

[3] R. Klette and A. Rosenfeld, “Digital Geometry,” Morgan Kaufmann, San Francisco, 2004.

[4] F. Huang, R. Klette and K Scheibe, “Panoramic Imaging,” Wiley, West Sussex, 2008.

[5] R. Klette, “Computer vision. Theory and algorithms,” DMK Press, 2019, 506 p.

[6] Y. Furman, “Introduction to contour analysis and its applications for image and signal processing,” Fizmatlit, 2002.

[7] D. Murrey and U. van Raiper, “Encyclopedia of graphic file formats,” Kiev : BHV, 1997.

[8] U. Pret, “Digital images processing,” Moskva : Mir, 1982.

[9] D. Foly and A. Dem, “Fundamentals of interactive machine graphics,” Moskva : Mir, 1985.

[10] A. Rossenfeld, “Image recognition and processing,” Moskva : Mir, 1972.

[11] L-J. Lorier, “Artificial intelligence systems,” Moskva : Mir, 1991.

[12] J. Lugger, “Artificial intelligence: strategies and methods for solving complex problems,” 2002, 864 p.

[13] D. Marr, “Information approach to the study of the representation and processing of visual images,” Radio and connection, 1987, 400 p.

[14] I. Jolliffe, “Principal Component Analysis,” Springer Series in Statistics, 2008, 324 p.

[15] “Arnold: Swimming Against the Tide,” American Mathematical Society, 2014, p. 165.