

# Method for Clusters and Classifying the Objects in Images

Natalya Shakhovska<sup>1</sup>, Oleksii Shamuratov<sup>2</sup>

<sup>a</sup> Lviv Polytechnic National University, 12 Bandera str., Lviv, 79013, Ukraine

## Abstract

This paper presents an algorithm by which it is possible to determine the main features of objects such as area and perimeter, radii of inscribed and circumscribed circles, the sides of the circumscribed rectangle, the number and relative position of angles, histogram gradient of object. On the basis of these features to carry out clustering and classification of the image. The success of the recognition based on the convolutional neural network was evaluated. According to the results, it is possible to conclude that the smaller the invariance of the class, the greater the accuracy of recognition. The amount of data in the training sample has little effect on the accuracy of the algorithm.

## Keywords<sup>1</sup>

machine translation, clustering, object classification, convolution artificial neural networks

## Introduction

In today's world, the entertainment industry in the Internet is developing rapidly, creating a demand for better products. This in turn has led to the use of artificial intelligence not only in science but also in entertainment. At the moment, applications that allow you to create animations of objects in photos are gaining popularity. This article presents an approach to solving the problem of defining objects for animation.

To classify and further identify objects, you must first determine their features. This is one of the abstraction options in which the input set of properties of the object is reduced to the minimum required number of features by which it will be possible to identify the object. The features selection begins with the input data set, which is divided into primary and secondary. It is assumed that the primary signs should be informative and should not be redundant [1]. This approach helps to speed up the process of learning object recognition algorithms and their more accurate operation. When analyzing complex objects, one of the main problems is the number of variables. This approach consumes more computing resources and can also cause a collision with respect to training data sampling, leading to a decrease in recognition accuracy. Isolating features is the main method of constructing combinations of variables to avoid these problems, but it is enough to describe the object accurately. Properly optimized feature allocation is the key to building an effective model [2].

After defining the features of the object, it will be necessary to cluster and classify the data, transfer objects to exemplar. Exemplar means the name of the area in the space of signs, which displays many objects or phenomena of the material world. Sign is a quantitative description of a property of the object or phenomenon [3]. A feature space is an N-dimensional space defined for a given recognition problem, where N is a fixed number of measured features for any object. The vector from the feature space  $x$  corresponding to the object of the recognition task is an N-dimensional vector with components  $(x_1, x_2, \dots, x_N)$ , which are the values of the features for this object. In other words, pattern recognition can be defined as the assignment of source data to a particular class by

---

International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2021), September 20-21, 2021, Kharkiv, Ukraine

EMAIL: jonburned@gmail.com (O.Shamuratov); nataliya.b.shakhovska@lpnu.ua (N.Shakhovska)

ORCID: 0000-0003-1913-5362 (O.Shamuratov); 0000-0002-6875-8534 (N.Shakhovska)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

selecting the essential features or properties that characterize this data from the total mass of insignificant details [4].

After analyzing the general information about the classification, we can describe the formal task of classification. Many objects are given. They need to be classified. The set is represented by subsets called classes. Specified: information about classes, a description of the whole set, and a description of information about an object whose membership in a certain class is unknown. Based on the available information about the classes and the description of the object, you need to determine to which class this object belongs [5].

The clustering problem is similar to the classification problem, but the difference between the method is that the classes are not defined. Clustering does not make any statistical conclusions, but makes it possible to analyze the data structure, i.e. the purpose of clustering. there is a search for similar structures.

Artificial neural networks are one of the most effective and common ways to represent and solve image recognition problems. Neural networks perform very well in pattern recognition problems because it combines mathematical and logical calculations. The neural network allows you to process a large number of factors, regardless of their origin, this is a stable universal algorithm. Neural networks allow to build dependencies of parameters in the form of a polynomial on the basis of educational sample that very simplifies realization of recognition of objects [6].

Color image segmentation is the process of selecting one or more areas in an image that meet some of the homogeneity criteria.

Image segmentation divides the image into areas with similar characteristics. The main feature for segmentation is the brightness for the monochrome image and the color component for color image. Also, image boundaries and textures are used for segmentation. Consider the main groups of segmentation methods, used in the analysis of cytological images.

One of the most well-known methods is threshold segmentation. This method is used for contrasting images. So, threshold segmentation used in conjunction with morphological operations of dilatation and erosion, and the choice of threshold is usually based on a priori information about micro-objects, stand out. The advantages of these methods are the high speed of image processing and the presence of theoretical justification. However, in more complex images with weakly contrasting boundaries or with an uneven background, these methods cannot be used [7].

Morphological segmentation is another approach to solving the segmentation problem. It comes down to finding the contour of the cell that would best fit its boundaries. This is achieved using a morphological gradient. However, the presence of a large number of false branches and internal contours does not allow to solve the problem. The high complexity of such methods is obvious, and, as a consequence, low speed and high complexity.

The third group is the methods of segmentation of building areas. Area scaling is widely used in image analysis, because often the structure is the background and does not allow you to select the object, because its elements have brightness and brightness levels, which match the brightness of the background. Typically, the selection of segmentation starting points occurs either randomly or with the help of a human operator on the ground certain a priori information [8].

Another group of methods is based on the principles of the cluster approach. During segmentation in the image, the centers of the clusters are selected, and then sequentially all points are checked for the distance to the centers in some metric.

The disadvantage of this group of algorithms is the need to pre-specify the number of clusters. It can be overcome using a neural network approach. The disadvantage of using neural networks is the learning stage, which significantly increases the analysis time [9].

The paper is organized as follows: in the next section the procedure of object detection and classification is developed. The Results section presents the described dataset and the accuracy of the procedure. The Conclusion section summarizes the work.

## **Materials and Methods**

The procedure of image analysis is considered in accordance with the so-called Marr paradigm. This paradigm states that image analysis is based on several successive levels of the ascending

information line: from point representation (raster image, unstructured information) to symbolic (vector and attributive data in a structured form, relational structures) and should be carried out on a modular basis using the following stages of analysis:

- image pre-processing;
- primary image segmentation;
- selection of the geometric structure of the visible field;
- definition of the relative structure and semantics of the visible scene.

The levels of processing associated with these stages are called analysis of the lower (noise filtering, analysis of histograms, etc.), medium (segmentation, contour analysis, etc.) and high (geometric transformations, classification, recognition, etc.) levels. While there is a lot of scientific work to solve low-level image processing problems, medium- and high-level algorithms continue to be a central field of research.

To find an object in an image database of objects must exist. It is planned to fill the database at the expense of system users, who will select objects for further animation. For such objects, the system must select the features, classify the object, and save it in the database. The database of classified objects will be used in the future to search for similar objects in the image and allow the user to automatically select objects for subsequent animation.

Therefore, the definition of the object in the image must be stable to the rotation of the object, the scale of the image and the variety of perspectives of the image. The following features of the object are selected:

- Determination of the area and perimeter;
- Determination of radii of inscribed and described circles;
- Definition of the sides of the described rectangle;
- Determining the number and relative position of the angles;
- Construction of a histogram of object gradients.

Clustering (segmentation) is procedure for grouping objects based on the proximity of their properties so that each cluster consists of similar objects, and objects of different clusters differ significantly. Clustering helps to understand the natural grouping or structure of a data set. The purpose of clustering is to define internal grouping in a set of unlabeled data segments. With the help of clustering solve the following problems of data analysis: search and discovery of knowledge, grouping and recognition of objects: search for homogeneous groups (reduction of data dimension), search for natural clusters and description of their unknown properties, search for useful and appropriate grouping, search for unusual objects. projects (emission detection).

The main methods of clustering include hierarchical, partitioning, density-based or grid-based, and artificial neural networks (ANN).

Before clustering the object should be detected on the image. The different images can be analyses with different objects. That is why the object segmentation should be provided. For object segmentation, the following steps are used: determination of the perimeter, the histogram filter usage, feature selection.

The task of classification in machine learning is the task of assigning an object to one of the predefined classes on the basis of its formalized features. Each of the objects in this problem is represented as a vector in N-dimensional space, each dimension of which has a description of one of the features of the object. Suppose we need to classify monitors: measurements in our parameter space will be the size of the diagonal in inches, the aspect ratio, maximum resolution, the presence of HDMI-interface, cost, etc. The case of classification of texts is a bit more complicated, for them the term-document matrix is usually used.

To teach a classifier, you must have a set of objects for which classes are predefined. This set is called a training sample, its marking is done manually, with the involvement of specialists in the study area.

### **3.1. Determination of area and perimeter**

The standard formula for calculating the area of an object is calculated by counting the number of elements related to the object. In the case of images entering the system, the area can be calculated by the number of pixels in the image area.

$$S = \sum_x \sum_y \begin{cases} 1, A(x, y) \in O \\ 0, A(x, y) \notin O \end{cases}$$

where  $O$  is the set of an array of pixels  $A(x, y)$  belonging to the object [7].

To find the perimeter of the object  $P$ , it is enough to count the number of pixels belonging to the contour of the object.

In order to ensure invariance to the scale of the object, a normalized sign  $U$  is added.

$$U = \frac{A}{p^2}.$$

### 3.2. Determining the radii of inscribed and circumscribed circles

First the determination of the geometric center of the object is provided,

$$C_{(x,y)} = \left( \frac{\sum_x \sum_y x A(x,y)}{\sum_x \sum_y A(x,y)}, \frac{\sum_x \sum_y y A(x,y)}{\sum_x \sum_y A(x,y)} \right),$$

where  $x$  and  $y$  are the row and column numbers of all pixels  $A(x,y)$  belonging to the object.

At this stage, the radius of the inscribed and circumscribed circles can be determined. This is done by choosing the minimum and maximum length of the vector, which starts at the center of the image and ends at points inside the perimeter.

$$R_{max} = \max_{0 \leq i \leq N} \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2},$$

$$R_{min} = \min_{0 \leq i \leq N} \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2},$$

where  $x_c$  and  $y_c$  are the coordinates of the center,  $N$  is the number of pixels in the perimeter, and  $(x_i, y_i) \in P$  [10].

The normalized parameter  $R$  is added to the formula.

$$R = \frac{R_{max}}{R_{min}}.$$

### 3.3. Determining the sides of the circumscribed rectangle

Determine the maximum and minimum values of the abscissa and ordinate of the image of the object  $x_{max}$  i  $x_{min}$ ,  $y_{max}$  i  $y_{min}$ , and then determine the height  $H$  and length  $L$  of the rectangle:

$$L = \max_{0 \leq i \leq N} x_i - \min_{0 \leq i \leq N} x_i,$$

$$H = \max_{0 \leq i \leq N} y_i - \min_{0 \leq i \leq N} y_i,$$

where  $N$  is the number of pixels in the perimeter, and  $(x_i, y_i) \in P$  [12].

### 3.4. Determining the number and relative position of angles

To do this, estimate the distance  $l$  between the start and end points of the contour fragment.

$$l_i = x_i |x_{i-2} - x_{i+2}| + |y_{i-2} - y_{i+2}|.$$

Then the condition  $l_i \leq H$ , is checked, where  $H$  is the conditional threshold value. You will need to implement its definition based on the object's past properties. If the condition is satisfied, then the point belongs to the set of angular points [13].

### 3.5. Construction histogram gradient of object

The basic idea is the assumption that the appearance and shape of the objects in the area of digital image can be described by the distribution of intensity gradients. The direction of the gradient of each pixel is calculated. For a two-dimensional function  $f(x,y)$  the gradient vector has the following form:

$$\nabla f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right].$$

The partial derivatives of the intensity functions at the  $x$  and  $y$  coordinates are the estimates of the contrast in the direction of the corresponding coordinate axes [11]. As an accurate assessment of

contrast in a direction corresponding coordinate axis can average the three different estimates of contrast around the pixel of the formulas:

$$\nabla_x = \frac{\partial f}{\partial x} \approx \frac{1}{3} \left( \frac{f[x+1,y]-f[x-1,y]}{2} + \frac{f[x+1,y-1]-f[x-1,y-1]}{2} + \frac{f[x+1,y+1]-f[x-1,y+1]}{2} \right),$$

$$\nabla_y = \frac{\partial f}{\partial y} \approx \frac{1}{3} \left( \frac{f[x,y+1]-f[x,y-1]}{2} + \frac{f[x-1,y+1]-f[x-1,y-1]}{2} + \frac{f[x+1,y+1]-f[x+1,y-1]}{2} \right).$$

For data evaluation, follow minimize operator mask image of the object contrast. To reduce computational complexity is best to use a differential-dimensional mask [14].

Once counted intensity gradient should split the image at the cell and construct a histogram of gradients object for each cell pocket  $G$  histogram module corresponds to the intensity gradient at point.

$$G = \sqrt{\nabla_x^2 + \nabla_y^2}.$$

Then carry out smoothing histograms in cells to ensure invariance relative change of anger and contrast.

### 3.6. Object features selection

To find the similarity between objects the semantic distance between signs should be calculated.

$$D_{ij} = \sqrt{\sum_{k=1}^N (X_{ik} - X_{jk})^2},$$

where  $D_{ij}$  - is the Euclidean distance between points  $i$  and  $j$ , which characterize one feature of different objects..  $N$  is the smallest number of values taken into account for a feature between two objects,  $k$  is the ordinal number of the feature value.

The procedure for determining the number of clusters consists of the following steps:

1. The k-means algorithm for  $K$  clusters is started and the corresponding internal variance

$$\delta_k = \sum_{i=1}^N \sum_{x \in c_i} \|x - m_i\|^2.$$

$x$  is the vector that characterizes the object included in the cluster  $c_i$ , the total distance between the features and the center of the cluster. The  $m_i$  - parameter is the center of the current cluster. A set  $\delta_k$  is constructed for different values of  $K$ .

2. Choosing the degree of transformation

$$Y = \frac{p}{2}.$$

3. Calculate the jumps according to the formula

$$J_k = \delta_k^{-Y} - \delta_{k-1}^{-Y}.$$

4. For the final number of clusters choose a value that is equal to the maximum [15].

### 3.7. Object recognition

The convolutional neural network architecture was chosen to search for objects in the image for the following reasons:

- images have a large dimension;
- a large number of parameters and classes of objects;

invariance to image zoom changes, camera shooting angles and other geometric distortions of the input signal [16]:

- Input layer. Will contain the input value of the pixels of the image, the dimension of the layer [64x64x3].
- Convolutional layer. The size of the filter is 7x7x3. The analysis will be performed on 5 features, so the layer will consist of 5 planes with dimensions [62x62x1].
- Sublayer layer. The dimension of the mask is 2x2. It consists of 5 planes measuring [31x31x1].
- Source layer. Dimension [1x1xN], where N is the number of clusters in the database.

The hyperbolic tangent was chosen as the activation function:

$$f(a) = A \cdot \tanh(S_a),$$

where  $f(a)$  - is the value of the searched element,  $a$  is the weighted sum of the signals of the previous layer,  $A$  is the amplitude of this function,  $S$  is its position relative to the reference point [17].

This feature has a number of advantages to solve the issue:

- symmetric activation functions, such as hyperbolic tangent, provide faster convergence than the standard logistics function;
- function has a continuous first derivative;
- function has a simple derivative that can be calculated because of its significance, giving a lower computational complexity.

## Results

5 different object classes were used to test the classification. Samples were taken from the open access of the TensorFlow project [18]. This dataset consists of the ImageNet dataset resized to fixed size.

An artificial neural network was trained on image samples, each class contained from 2528 to 16185 images with a size of 64x64 pixels. 1000 images of objects of each class were then selected for testing. The results are shown in the Table 1.

**Table 1.** Result of objects classification and recognition

Class name	Number of objects in the training sample	The average percentage of successful recognition
Cars	16185	83%
Flowers	3670	68%
Apple	2528	85%
Leaves	6732	71%
Cats	12361	85%

So the lowest result was for the Flowers class, due to the fact that the training sample was not large and the images of objects were flowers of more than one species. In the sample of the Cars class there were also cars of different brands and cab types, but due to the large number of images in the sample the final test result was higher. The highest result was obtained for the Apple class. This is because the sample contained an object whose structure and shape differed little.

## Conclusion

This paper presents an algorithm by which it is possible to determine the main features of objects such as area and perimeter, radii of inscribed and circumscribed circles, the sides of the circumscribed rectangle, the number and relative position of angles, histogram gradient of object. On the basis of these features to carry out clustering and classification of the image. The success of the recognition based on the convolutional neural network was evaluated. According to the results, it is possible to conclude that the smaller the invariance of the class, the greater the accuracy of recognition. The amount of data in the training sample has little effect on the accuracy of the algorithm.

The task of classification in machine learning is the task of assigning an object to one of the predefined classes on the basis of its formalized features. Each of the objects in this problem is represented as a vector in N-dimensional space, each dimension of which has a description of one of the features of the object. Suppose we need to classify monitors: measurements in our parameter space will be the size of the diagonal in inches, the aspect ratio, maximum resolution, the presence of HDMI-interface, cost, etc. The case of classification of texts is a bit more complicated, for them the term-document matrix is usually used.

Using pixel relationships in a neighborhood has a number of advantages over using individual point characteristics:

- the ability to use for images of any type;
- increased resistance to segmentation of images in which objects are in close proximity to each other. This advantage allows you to use this method for segmentation of cytological images;
- reducing the impact of noise and distortion of the input image on the overall result of segmentation by analyzing the image by several algorithms of previous markup;
- reducing the number of "undefined" points, ie points that are on the boundaries of areas and with equal probability may belong to two areas.

The disadvantages are:

- complexity of the segmentation process;
- influence of previous marking on work results.

But this effect can be reduced by increasing the number of previous markups

## References

- [1] Tien D. B., Ching Y. S., Zi-Cai L., Yuan Y. T., "Computer Transformation of Digital Images and Patterns", p. 276, 1989.
- [2] Y.-Q. Wang, "An Analysis of Viola-Jones Face Detection Algorithm", IPOL Journal, 2013.
- [3] Khan H. Abdullah и M. Shamian Bin Zainal, «Efficient eyes and mouth detection algorithm using combination of viola jones and skin color pixel detection» International Journal of Engineering and Applied Sciences, № Vol. 3 № 4, 2013.
- [4] V. Gaede и O. Gunther, "Multidimensional Access Methods", ACM Computing Surveys, pp. 170-231, 1998.
- [5] S. Khan, H. Rahmani, Syed Afaq Ali Shah, M. Bennamoun, G. Medioni, S. Medioni, "A Guide to Convolutional Neural Networks for Computer Vision", Morgan & Claypool, p. 207, 2018.
- [6] Sibte ul Hussain, "Machine Learning Methods for Visual Object Detection". p. 160, 2012.
- [7] Zhang, Y. J. (2006). An overview of image and video segmentation in the last 40 years. *Advances in Image and Video Segmentation*, 1-16.
- [8] Ngugi, L. C., Abelwahab, M., & Abo-Zahhad, M. (2021). Recent advances in image processing techniques for automated leaf pest and disease recognition—A review. *Information processing in agriculture*, 8(1), 27-51.
- [9] Liu, Y., Shen, C., Yu, C., & Wang, J. (2020, August). Efficient semantic video segmentation with per-frame inference. In *European Conference on Computer Vision* (pp. 352-368). Springer, Cham.
- [10] P. Arabie, L. J. Hubert, G. De Soete, "Clustering and Classification", p. 500, 1996.
- [11] D. Parks, "Object Detection and Analysis: A Coherency Filtering Approach", p. 172, 2008.
- [12] Yongqiang Z., Chen Y., Seong G. K., Quan P., Yongmei C., "Multi-band Polarization Imaging and Applications" 1st ed., p. 204, 2016.
- [13] Manikandan S., "Vision Based Assistive System for Label and Object Detection", p. 64, 2015.
- [14] Salma H., "Object Detection Using Histogram Of Gradients", p. 52, 2018.
- [15] Wu J., "Advances in K-means Clustering: A Data Mining Thinking", Springer Science & Business Media, p. 180, 2021.
- [16] J.Loy, "Neural Network Projects with Python: The ultimate guide to using Python to explore the true power of neural networks through six projects", Packt Publishing, p. 308, 2019.
- [17] Brannon W. C., "Object Detection in Low-spatial-resolution Aerial Imagery Using Convolutional Neural Networks", p. 79, 2019.
- [18] Dataset <https://knowyourdata-tfds.withgoogle.com>