

THE COMPONENT METHOD OF SCENE ANALYSIS AND OBJECT RECOGNITION

Aleksander Vechur, Aleksey Chayka, Yuriy Chemodakov

Kharkov National University of Radioelectronics, 14 Lenin Avenue,
 61166 Kharkov, Ukraine, vechur@ieec.org

Abstract: *The main aim of this work is to propose method that allows some machine to determine surrounding objects. It is important that algorithm must use small computer resources so that machine could work in real time. To improve segmentation on natural images, it is necessary to combine multiple features effectively. Our experimental results are consistent with the theoretical analysis.*

Keywords: - Neural networks, Bayesian classifier, Pattern recognition, Grammar for syntax analysis

1. PROBLEM DESCRIPTION

Human is not a machine. Sometimes he is getting tired and makes errors. Therefore people try to give the machine all routine work and to automatize any possible process. When anybody looks at some object he does not pay attention to all things. Our eye and memory just unable to see all things and keep them in mind. That's why people often can lose the way; they can forget exterior of another people. There are moments when you need to evaluate situation as soon as possible and to make a resolve to the corresponding operations. For instance self-learning machine can work much more easily and faster in space than human. But such machine with all requirements must be created. For example, Moon research vehicle. It can spend on the moon's surface much time and will not be tired of data processing. The machine receives mainly information from video cameras that are placed on its board. Therefore the machine must process the information as quickly as possible.

Image understanding is one of the key areas of research where the main objective is to understand the components of an image and interpret its semantic meaning. Image component recognition is the basic building block of most image processing based researches that involves image understanding. What comprises an image must be first identified before it can be analyzed any further. In some applications, the recognition of image objects may be enough in itself. For example, in medical images, the identification of a tumour is enough and no further processing is needed. On the other hand, in

some cases it is need to identify image objects as a basis of a more detailed understanding of the image. Also, in some applications the knowledge of object category may be defined in term of their properties. For example, in the case of content based image retrieval, it is the content that defines the object category.

The recognition process of available algorithms includes two main stages: "training" stage and "control" stage. At the first stage the algorithm should recognize relationship between values of the describing characteristics x and value of the target characteristic Y . This regularity is expressed in "decision rule". At the control stage uses attributes of q objects. It allows making a decision if any object q belongs to one of k images that are available. The example is shown graphically in Fig. 1.

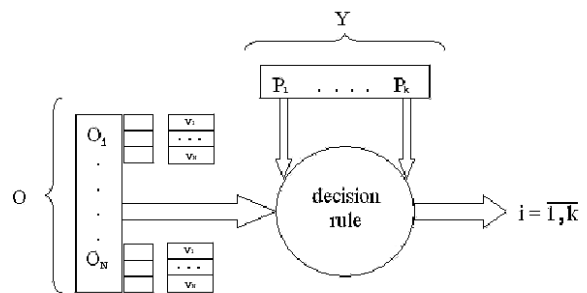


Fig. 1 – Example of the recognition process.

In a classic recognition problem universal set separates into parts that called images. The image of any object is setting by a set of its parts. In a case

with text recognition there will be signs that all are available in universal set. There must be all typefaces of each letter. In that case recognition program with a few examples of typefaces of each letter from learning sample can recognize what letter was symbolized by writing symbol. But it is used the exhaustive search of matching of a character with each character class. This means impossibility of quick machine reaction with plenty of learning samples.

There are also model optimization methods of computation estimate of parameter that describes observational accuracy feature; methods and algorithms of precedent recognition based on continuous mapping such as Peano curves; many methods and algorithms of recognition process that are based on expansions and modifications of the recurrent form of least-squares method; methods that construct hierarchic rules in recognition problems; effective mathematical models and methods that describe images based on fractal coding of hierarchic structures; base models and technology of computer research of systems that process complex signals. There are algorithms of the optimum recognizing operators of construction by the polynomials no more than quadratic power. Bayesian classifier has good results, but it is necessary to have prior probabilities of relation between given image and classes that machine has for classification (it is used in algorithm at the step Initial suppositions).

Some recognition algorithms use neural networks. Neural network is a big class of various systems that allows making a data analysis. The classic neural networks are capable to process only information, given in a simple form, otherwise big computer resources are needed. Here you can see the most using neural network in image recognition:

- The Hopfield networks capable to store 15% images from input neurons in theory. But really images must differ very much.
- The Hamming networks have better practical results and quantity of images that storing depends on quantity of output neurons [1]. Dependence of quantity of connections is linear and a Hopfield network has quadratic dependence (it is used in algorithm at the step Initial suppositions).
- Multilayer perceptron works well enough in recognition algorithms. But it is necessary to calculate deviation between given image and image that is sample for this network. Plenty of elementary images need much of computing resources.

But all these algorithms can be applied only in highly specialized scopes of activity if you use only one of the methods. If the best sides of many

methods be gained together the more good results can be reached with a few computer resources. That is why it is the best to use the algorithm that is described below. Image analysis is a complex task; therefore it is necessary to divide it into parts.

2. IMAGE TRANSFORMATION

Normalization. Natural object images are highly dependent on the quality of outdoor sunlight. Therefore object images must be converted to a natural view for simplification of their identification. There must be used different filters (Gabor, wavelet etc. [2]) the relative brightness, contrast, precision and washout.

The filter results are obtained by a discrete cyclic convolution. Using quadratic images with N pixels per axis the filter-response is obtained by

$$P'_{mn} = \sum_{m',n'=0}^N W_{m',n'} P_{m-\frac{N}{2}+m',n-\frac{N}{2}+n'} \quad (1)$$

where $W_{m',n'}$ is the amplitude of the two-dimensional filter-mask at (m', n') and the dimension of the filter is the dimension of the image. The indices $(m-\frac{N}{2}+m', n-\frac{N}{2}+n')$ of the original image are defined modulo N . This operation with N^2 multiplications and additions has to be carried out for each pixel P_{mn} in the filtered image, so it has a time-complexity of M^2 , where $M = N^2$ is the total amount of pixels of the image. For a typical image of 256×256 pixels, this yields $4.3 \cdot 10^9$ additions and multiplications. Even a 1 GHz microprocessor capable of doing one multiplication or addition per clock-cycle would need almost 10 seconds to perform such an operation, which is far from real-time.

Looping. This algorithm is based on the syntactic method. This method is based on structural relations that describe an identifiable image. Image (scene) is represented as some structure that consists of graphic primitives [5]. First the contours chooses to find objects on the image. It is possible to divide the image into contours set while image processing with sharp distinctions of brightness and color. Each contour mainly consists of one color and illumination that differs a little [2]. Therefore principally these areas of the image will belong to one object (you can see it in Fig. 3.).

Color removal. The filter makes a gray-scale picture using information on brightness and contrast. For example vehicle that is studying planet's surface. Color probably will not give much

information for image processing as against image precision.

3. SEMANTIC PROCESSING

Vectoring. It takes much computer resources and time if image is taken as a set of pixels. Contour is vectoring for simplification of image processing. Besides it is needed less memory size for image store.

Initial suppositions. During object recognition process it would be effective to use the prior information about robot's (cameras') location and object that was recognized before. For example if object of recognition is somewhere in desert there is no sense to look for a bear on the picture. Let's assume that after picture processing with filters, the water image and clouds image will be similar. However there is no sense to suppose that it was water if the camera was directed upward. It is possible to calculate probabilities that object (x) belong to the (k) class using Bayesian approach [4]:

$$p(k_i | \bar{x}) = \frac{p(\bar{x} | k_i) * P(k_i)}{\sum_{j=1}^M p(\bar{x} | k_j) * P(k_j)} \quad (2)$$

Object segmentation. Each natural object can be presented as set of elementary objects that are called primitives. It can be circles, squares and triangles. When these primitives are found during image processing they are placed in the table with information about their sizes and location at image. It is possible that it will be necessary to apply approximation of definite power at the step 2.1 for segmentation of sharper primitives. The primitives segmentation can be made using Hough's transform. It allows finding out curves with defined shape at the image, even if curves are fragmented. It is important that Hough transform needs much computer resources even for circles detection. Therefore it is desired that it will use more optimized "snakes" method in a future:

$$E_{snake} = \int_0^1 E_{int}(r(s))ds + \int_0^1 E_{image}(r(s))ds + \int_0^1 E_{ext}(r(s))ds \quad (3)$$

where $r(s)$ describes image; the first integral takes into account length and smoothness of a curve; the second integral has small values if segmented contour passes through image areas with heavy gradients; the third integral allows taking into account influence of image elements that are geometrically located outside of "snake".

Object recognition must be steady to any object transformation. Context including is obligatory

condition for effective object recognition. Hanson and Riseman [6] proposed expert method that create rules if-then for hypothesis about object presence in the image.

Another method is statistical. It connects N objects of the scene O_1, \dots, O_N (Song [7]). Then it determine set of local dimensions V_1, \dots, V_N that corresponds to image regions.

$$P(O_1, \dots, O_N, V_1, \dots, V_N) = [\prod_i P(V_i | O_i)] * P(O_1, \dots, O_N) \quad (4)$$

Probability density function (PDF) is producing through getting of conditional independence between local dimensions.

Torralla [8] showed that simple substances could be recognized through centering of image scene. It allows getting through object component identification.

$$P(O|V) = \frac{P(V|O)}{P(V)} P(O) \quad (5)$$

where $P(O|V)$ is conditional probability of PDF of object O representation that is presented by set of dimensions V . Set V may contain pixels depth or color allocation or multichannel band-pass filter exit. O represents total properties of object in the scene.

$$O\{o, x, \sigma, \dots\} \quad (6)$$

where o is category number (for example human or car); x is value of object coordinates in the image; σ is object size. For more details it is possible to use additional parameters. The purpose is to display O from object features. Under real conditions (7) formula has been used:

$$P(O|V) \cong P(O|V_L) = \frac{P(V_L|O)}{P(V_L)} P(O) \quad (7)$$

where $V_L = V_B(x, \varepsilon)$ is set of local dimensions of image near B point with position x and measurement $\varepsilon = g(\sigma)$. V_L has low number of dimension.

Relevant object features for object recognition is that potentially has relation with object and is not a background. For example the pattern based paradigm compares pattern obtained from image objects with other image objects that have various sizes and positions. Image that wasn't detected is given up and assigns as noise with particular statistical properties.

These object-oriented methods became ineffective when image is noisy or has bad resolution. Besides it will take much time to search and compare in all model space.

Synthesis of interdependent objects. Algorithm checks possible syntactical interdependences between obtained primitives on the basis of segmented primitives and constructed grammar. It gives defined objects [3]. The grammar is constructed by the man and it is represented by records of the model objects that consist of primitives. It is planning to automatize process of grammar creation as much as possible. The searching on grammar implements by comparison of neighboring primitives. It is based on property of human visual system. It is capable to group image details in some structures without any priori knowledge of image content (perceptual grouping). For example, if you use an ascending analysis with such grammar rules, it is possible to conclude that primitives a, b make a rectangle S (p is right angle):

$$\begin{aligned} S &\rightarrow ApBpApB \\ A &\rightarrow aA | a \\ B &\rightarrow bB | b \end{aligned} \quad (6)$$

Then recognized object is cutting from the image and the machine starts to search for relations for the rest primitives. All recognized objects are transmitted to a neural network.

4. OBJECT RECOGNITION

Check for validity of categorized object. The specially trained Hamming neural network is capable to recognize the image that has been compounded of primitives [4], therefore, it will be used for confirmation of classification on grammar [5].

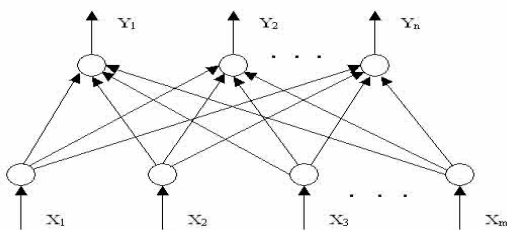


Fig. 2 – Example of Hamming neural network.

The example is presented in Fig. 2. where \bar{x} is image of object that is represented as vector; \bar{y} is output vector, that indicates what class is the model for given object.

In case of error there is a rollback of operations up to step Synthesis of interdependent objects. Then interdependences search is prolonged except for the given combination. If result is positive then works step Synthesis of interdependent objects.

The stop descends if none of the objects can be composed from the primitives. In that case when

none of the objects can be recognized probably machine need to change place of view or angle of view.



Fig. 3 – Example of algorithm working.

5. CONCLUSION

The developed algorithm can be used for quick decision making about environment objects in real situation. For example vehicle that is studying unknown planet's surface. But this algorithm application could be found in Earth conditions: it can be built in autodriver in any car. It can be more effective than nowadays autodriver decisions, because they use boundary line detection and nearest car detection. But these methods can't make decision without human.

This work made attempt to realize algorithm that gets images from surrounding space and after recognition make decision what kind of object gather surround vehicle or car.

6. REFERENCES

- [1] S.Y. Kung. *Digital Neural Networks*. Engewood Cliffs. New Jersey: PTR Prentice Hall, 1994. p. 418.
- [2] Mona Sharma. *Performance Evaluation of Image Segmentation and Texture Extraction Methods in Scene Analysis*, to the University of Exeter as a thesis for the degree of Master of Philosophy in Computer Science, 2001.
- [3] C.T. Lin. *Neural fuzzy systems: a neuro – fuzzy synergism to intelligent systems*. C.S.G.Lee. Upper Saddle River, New Jersey: PTR Prentice Hall, 1997. p. 786.
- [4] T. Kohonen. *Self-organizing Maps*. Berlin: Springer-Verlag, 1995. p. 363.
- [5] P. Rummel, W. Beutel. *Workpiece Recognition and Inspection by a Model-based Scene Analysis System*, *Pattern Recognition*, Vol. 17, No. 1 (1984). pp.141-148.
- [6] A. R. Hanson, E.M. Riseman. *VISIONS: A computer system for interpreting scenes*, *Computer Visions Systems* (1978). Academic Press, New York. pp. 303-333.

- [7] X. Song, J. Sill, Y. Abu-Mostafa, H. Kasdan. *Image recognition in context: application to microscopic urinalysis*, *Advances in Neural Information Processing Systems* 963-969 (2000). MIT press, Cambridge, MA.
- [8] A. Torralba, P. Sinha. *Statistical context priming for object detection*. *IEEE. Proc. Of Int. Conf. in Comp. Vision, Vol. 1 (2001)*. pp. 763-770.



Alexander V. Vechur

Date of birth: 11.09.1977.

Place of birth: Chuguev, Kharkov reg.

Current position: teaching Assistant at Kharkiv National University of Radioelectronics.

Current work: "The Component Method of Scene Analysis and Object

Recognition".

Scientific Interests: Advanced Process Systems and Control Engineering and also in the use of technology in Learning. Analysis and synthesis of continuous-time and discrete-time polynomial control systems; applications of algebraic geometry in control theory; applications of symbolic computation in control theory; controller design for sampled-data non-linear systems; analysis and applications of the input-to-state stability property to control systems; averaging methods in control theory

Other interests and hobby: travel, my friends.

Aleksey A. Chayka

Date of birth: 27.05.1983.

Place of birth: Semipalatinsk.

Current position: bachelor at Kharkiv National University of Radioelectronics.

Other education: Kharkiv economical luceym No.161.

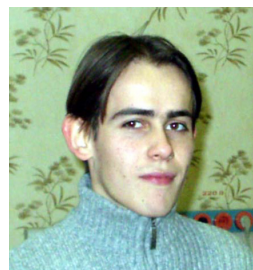
Areas of interest: application of Artificial intelligence in industry.

Current work: "The Component Method of Scene Analysis and Object Recognition".

Further work: advanced object recognition; applying AI in ultradeep drilling.

Other passions: volleyball, football.

Future plans: cooperation with one of European University in object recognition using AI methods area.



Yuriy A. Chemodakov

Date of birth: 12.03.1984.

Place of birth: Illichivsk, Odessa reg.

Current position: bachelor student at Kharkiv National University of Radioelectronics.

Current work: "The Component Method of Scene Analysis and Object Recognition".

AI Interests: Network applications (Adaptive peer-to-peer systems, Web crawling), Pattern recognition, Agents, Data Mining.

Other interests and hobby: I like drum and bass, classical and jazz music, opera and painting. Also I'm keen on philosophy, mathematics and astrology.

Comments: for any information mail to kri.ks@mail.ru