Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Robotics and Automation, Madrid, Spain, February 15-17, 2006 (pp218-224)

SUS: Learning of the Categories of Visual Objects

Zbigniew Les Magdalena Les
The Queen Jadwiga Research Institute of Understanding
P.O. Box 654, Toorak, VIC 3241
AUSTRALIA
http://www.qjf.pl.org

Abstract: - In this paper the new method of learning of the visual objects, called the categorical learning, is presented. In Shape Understanding System (SUS) learning is seen as a continuous process where the knowledge can be acquired on many levels of the categorical hierarchies. Categorical learning that is part of the shape understanding method was applied to learn selected categories of visual objects. In the categorical learning the visual information is learned independently as a visual concept. In the first stage of learning, the visual concept is learned whereas in the second stage the metalingual or phenomenological concept is learned. The visual concept is used to recognize object as a member of one of the categories of the visual object. The meaning of the object is obtained in the interpretational process based on the metalingual or phenomenal concept. Learned knowledge is stored as part of the conceptual structure of the categories that represent knowledge about the world. Categories have hierarchical structure; at the bottom of each category is the prototype of the category. In contrary to other methods of learning where only a small number of categories is learned (e.g. different fonts of the letter, mechanical tools) in the categorical learning objects that can belong to any category are learned as a part of the SUS knowledge. The proposed method can be applied in designing the new generation of robots that will be acting based on the results of the visual thinking (imagination).

Key-Words: - image understanding**,** shape understanding, categorical learning, visual concept

1 Introduction

Robot that is able to solve complex visual problems need to have ability to learn visual information that could be accessible during reasoning about the visual aspect of the world. Learning should include categories of all visual objects and learned visual information should be understandable by human. Existing machine learning methods are focused on learning of the description of selected classes of objects. Machine learning techniques are often used to learn shape representation of the given set of real world objects. However not all characteristics learned by the proposed method are understandable by human. For example, in [1] the inductive learning techniques were used to discover knowledge from shape contours. In the first step the contour was divided into segments based on the k-curvature algorithm and properties of each segment were computed. Next the inductive logic programming was used to obtain a set of rules which reflected the properties of contours. The learning a prior knowledge (models) for a system which has capability to recognize objects in image is described in [2]. The prior knowledge is concerned with the class of possible inputs. Objects' features are first extracted and the relations between them are found. These relations are then converted to the symbolic form and FOIL-the relational learning system which produces definitions of the object.

Existing methods of machine learning use the concept learning to learn the concept of the selected areas of interest. For example, concept learning from examples (concept acquisition) is the task to induce general description of concepts from specific instances of the concepts. An important variant is the incremental learning where the input information includes, in addition to the training examples, previously learned hypotheses or human-provided initial hypothesis that may be partially incorrect or incomplete [3]. Proposed categorical learning is focused on learning the concepts of different categories of objects rather than the concept of objects.

2 Generalization

In the categorical learning, the concept is given as the definition expressed in terms of symbolic names that refers to the shape classes [4]. In SUS a symbolic name is given in the form of the SUS representation. The symbolic name can be expressed in the different forms that represent the different aspects of the visual object. For example, Fig.1 can be interpreted as a complex object given as $C(L^3, L^3)$, or as a concave object given as $Q^2_{L^4}(L^3_O, L^3_O)$. The symbolic name $C(L^3, L^3)$ denotes an archetype of the complex class (composed of two triangles). The symbolic name $Q^2_{L^4}(L^3_O, L^3_O)$ on the detail level of description has additional parts {mmslle} and {apaoao}. The term L{mmslle} denotes the normalized size of the sides (l - large, m-medium, s-small and e - very small). The term {apaoao} denotes angles (a - acute, o - obtuse and p - right).

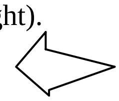


Figure 1. The archetype of the complex class

Generalization that is performed by translating a symbolic name into a string form requires including all details of the symbolic name. The level of the detail is marked by introducing the symbol ”_”. The symbolic name is translated into the string form L0_L1_...Ln, where the level Ln denotes the level of the detailed description of the archetype of the class. For example, the triangle class L^3_A (m,m,m) is translated into the form L_3_A_mmm. During generalization the symbol is dropped from the left to the right e.g. for the symbol L_3_A, the two generalizations are possible: L_3 and L, where “L_3_A” is any acute triangle, “L_3” is any triangle, and “L” is any polygon.

3 Phantom Concept

In the shape understanding method 2D visual objects are called phantoms. Phantoms are divided into four groups: the figure, the letter, the pictogram and the icon [5]. The description in this chapter is focused on the icon and the iconic concept. The icon refers to the real world object. The concept of the real world object ρ includes all possible descriptions of the real world object. The visual object (phantom) $u_i \in U$ that refers to real world object ρ is given by its name α and the iconic concept Θ^I that consists of two components: the phenomenal concept Ξ and the visual concept φ. The iconic concept is given by the name α, similarly to the phantom and the real world object to which a phantom refers. The iconic concept is obtained during the learning process. The iconic concept Θ^I consists of two components: the phenomenal concept Ξ and the visual concept. Both these components are acquired in the learning process. To learn the concept Θ^I we need to learn both the phenomenal and visual concept. Each concept Θ^I is given by its name α^I. The name of the concept is given by one of the languages L. To understand meaning of the real world object SUS transfers the perceptual data into the visual concept during the visual reasoning process. The visual concept is used to find the name of the perceived object and the iconic concept Θ^I that is “the mental representative” of the perceived object. The phenomenal concept Ξ that is part of the iconic concept, given by the name α includes possible description of the object given by the name α. The visual concept is responsible for recognition of the object and attaching name to it, whereas the meaning of the object is given by the phenomenal concept and its conceptual link with categorical structure of the knowledge about the world.

The real world object that is perceived by SUS is transformed into 2D representation during the segmentation process. The segmented object has its parts marked by the different colors. Parts of the segmented object do not necessarily refer to the different parts of the object. The proper interpretation of the segmented parts is obtained during the visual reasoning. In the visual reasoning the segmented object is transformed into black and white object. During visual reasoning at first the object is classified to one of the categories and next, parts are interpreted based on the knowledge of these categories. As the result of the final stage of the reasoning the structural archetypes are obtained. The structural archetypes that capture the main visual characteristic of the object are used to perform the ‘mental transformation’ during the visual thinking process. The process of understanding of the real world object is shown in Fig.2. The photograph of the object is segmented and next the black and white object (shape) is obtained. During the visual reasoning the shape is divided into parts based on the visual attributes of the shape. Next the iconic concept is obtained and the segmented parts are interpreted based on the knowledge of the object. The segmented parts are matched with the parts of the photograph of the object. The structural archetype is used in the process called the conceptual magnification to check if the part is interpreted correctly.

Figure 2. Process of interpretation of the visual object

Visual conceptualization is connected with finding the suitable visual representation of the object such as the realistic representation of the object, the conventional representation of the object (Fig. 3a), the schematic representation of the object (Fig. 3b), the sign representation of the object (Fig. 3c), and the conceptual structural skeleton of the object (Fig. 3d,e).

a b c d e

Figure 3. Different types of the visual representations of the object

4 Visual Categories

Categories that are established for the purpose of this research refer to the all visible objects. The introduction of the general categories facilitates process of interpretation of the perceived object by classifying it to one of the general categories. The knowledge that is given by a general category is used to find the specific interpretation of the perceived object. Learning of the definition of the category supplies the knowledge that is needed during the understanding process. One aspect of understanding connected with perception of the input data is classification to the specific category based on the utilization of the context information. Context information is used to 'filter' the possible interpretations using the knowledge about the world. For example, it is impossible to find the fish on the moon. Each category has its name as the result of the naming process.

The general categories are defined based on the general knowledge which can be acquired from available sources. The most general category is the category of the visible object. Categories are defined in such a way that they can be easy to modify and there is a possibility to include more general or more specific category. The categories are implemented as a hierarchical structure of the experts.

Each category is part of the conceptual structure of the knowledge about the world. Categories have the hierarchical structure. At the bottom of each category is the prototype of the category. The category of a higher level includes categories of the lower level. The relationships between categories of the different levels are denoted by applying a bracket $\nu^i \equiv \left\langle \nu_0^{i-1}, \nu_1^{i-1}, \ldots, \nu_n^{i-1} \right\rangle$. Categories refer to the taxonomy of phantoms. There are four groups of phantoms: the icon, the pictogram, the letter and the figure. The higher category of the visual object includes the category of the real world object (icon), the category of the letter, the category of the pictogram and the category of the figure. It is written as $\nu_O \equiv \left\langle \nu_{realO}, \nu_{letter}, \nu_{sign}, \nu_{fig} \right\rangle$. The category of the figure consists of the definition category, the formula category and specification of generation of the figure category. The category of the letter consists of the category of the type of alphabet, the category of meaning of the letter, and the category of rules of composition of the words. The category of pictogram consists of the category of meaning of the visible object and the category of the second meaning as a symbol or a sign.

The category of the real world object ν_R includes the category of the micro-scale object ν_{mic}, the category of the macro-scale object ν_{mac}, and the category of the earthy object ν_{ertc}. It is written as $\nu_R \equiv \left\langle \nu_{mic}, \nu_{mac}, \nu_{eartc} \right\rangle$.The category of the earthy object ν_{ertc} consists of the category of the living object ν_{liv} and the category of the non-living object ν_{mat}. It is written as $\nu_{ert} \equiv \left\langle \nu_{liv}, \nu_{mat} \right\rangle$. The category of the non-living object ν_{mat} includes the category of the man-made object ν_{mad} and the category of the non-man-made object ν_{cre}. It is written as $\nu_{mat} \equiv \left\langle \nu_{mad}, \nu_{cre} \right\rangle$. The category of the man-made object ν_M includes the category tools ν_{tool}, the category vehicles ν_{veh} and the category buildings ν_{buil},. This research is focused on the learning of the figures, letters, pictograms and the visual objects that can be termed as tools. Tools' categories are derived from the process category. The process category ν_x^P consists of: the category of worker ν_x^W, the category of tools ν_x^T, the category of material ν_x^{Mat}, the category of knowledge ν_x^W and the category of result ν_x^{Res}. The category of processes is written as $\left\langle \nu_{man}^P, \nu_{tools}^P, \nu_{mat}^P, \nu_{kno}^P, \nu_{res}^P \right\rangle$. As an example of the category of process a mason category is given. The mason category is written as $\nu^{mason} \equiv \left\langle \nu_{man}^{mason}, \nu_{tools}^{mason}, \nu_{mat}^{mason}, \nu_{kno}^{mason}, \nu_{res}^{mason} \right\rangle$, where the worker is represented by the category

Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Robotics and Automation, Madrid, Spain, February 15-17, 2006 (pp218-224)

worker-mason $v_{man}^{mason} \equiv \langle v_{man}^{mason} \rangle$, tools are represented by the category mason-tools $v_{tools}^{mason} \equiv \langle v_{trowel}^{m/t}, v_{hamer}^{m/t}, \ldots \rangle$, material is represented by the category mason-material $v_{mat}^{mason} \equiv \langle v_{brick}^{m/m}, v_{stone}^{m/m}, \ldots \rangle$, and the result is represented by the category mason-results $v_{res}^{mason} \equiv \langle v_{\mathrm{hom}e}^{m/r}, v_{office}^{m/r}, \ldots \rangle$.

5 Categorical Learning

The proposed categorical learning is the task to induce the general description of the categories from the specific instances (phantoms) of the concepts. The input information in the categorical learning includes, in addition to training examples, previously learned categories and human-provided initial hypothesis that may be partially incorrect or incomplete. The learned concept on the bottom of the hierarchy of the categories is called a prototype. The prototype is the definition of the learned phantom (the visual concept) in terms of the symbolic names and characteristic features of the category given by its name. Categorical learning takes into account interpretation of the visual object in the context of all categories. During learning the new case is evaluated in the context of all learned categories. The visual concept of the general category includes all prototypes of the specific categories. The prototype as a definition of the category (an object) depends on the type of category. For example, the geometrical figure such as a circle or a convex polygon can be defined using only a few symbolic names whereas the complex mechanical tools such as a car needs a large number of symbolic names and characteristic features to define it.

Learning of the visual object consists of the two stages. At the first stage the visual concept is learned. At the second stage the metalingual or phenomenological concept is learned. In this paper the learning of the visual concept is presented. Learning of the visual concept of the object independently from other conceptual ingredients is a new approach in machine learning methods. All visual information that is extracted from the object is transformed into the symbolic representation called the visual concept. Such an approach makes it possible to concentrate on the visual aspect of the learned object. The categories are represented by their names and all knowledge that is learned by learning of the visual concept is the knowledge about the visual appearances of objects.

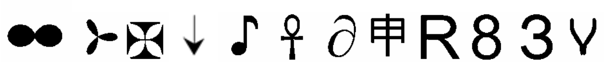
Learning of the visual concept depends on the visual complexity of the learned object. The visual complexity refers to the number of parameters that are needed to fully describe the visual object and variability among the different phantoms of the same concept of the visual object. In the case when there is a big number of phantoms needed as a training set, or not all phantoms can be available during the learning process, the learning of the new concept is based on the set of rules that define the general concept. The visual concept φ is obtained during the learning process. It is assumed that the visual concept is uniquely described by the name α. During the learning process the set of phantoms $U^P \in U^O$ that are representatives of a given visual category is selected and next for each phantom $u_i \in U^P$ the symbolic name η_i is obtained. As the result of the learning process a set of symbolic names $\varphi_\alpha = \{\eta_1, \eta_2, \ldots, \eta_n\}$ that represents the visual concept φ_α is obtained. In general, to find the visual concept φ_α a set of phantoms $\mathbf{u}$ is used as a training set in the process of learning. Each phantom is transformed into its digital representation using a perceptual transformation and next into the symbolic name η_i during reasoning process. The visual concept represented by the category v^p is called a prototype. An object that belongs to the category v^p is defined by application of the set of rules. For example,

$$[a^0 = A_0] \Rightarrow u > v^p$$

$$([a^1 = A_1].\wedge.[a^2 = A_1]).\vee.([a^1 = A_2].\wedge.[a^2 = A_2]) \Rightarrow \quad > v^p$$

$$([a^3 = A_3].\vee.[a^4 = A_4]\ldots) \Rightarrow \quad > v^p$$

where a^i denotes the symbolic name of the 'parts' of the visual object or the characteristic feature. The prototype is learned starting from the definition of the general category. The definition of the general category is expressed in terms of symbolic names. During generalization the symbolic name is translated into the string form L0_L1_...Ln, where the level Ln denotes the nth level of description of the archetype of the class. Learning of the figure pentagon is given as an example of the learning process. The symbolic name WL5[aaaaa][sssss] obtained during the reasoning process is transformed into the string form W_L_5_[a]_[s]. The concept defined by a set of rules is learned by starting from the definition of the general concept. The general concept is defined in the context of all learned prototypes. Let's assume that the first learned prototypes are

.

The general concept of the learned figure  is defined as

HUL="W", NAME ="ConvexObject",
if[m_CHul=HUL]
{m_Name= NAME}

The variable m_CHul denotes the symbolic name of the examined object obtained during the process of visual reasoning. The variable m_Name is the name of the prototype defined by the definition of that prototype. This definition well describes the differences (dissimilarity) among objects. All learned prototypes are concave objects. In the categorical learning, testing and learning processes are complementary. During testing of the learned categories where the figures  are given as an input these figures will be assigned to the name 'convexobject'. The answer given by SUS is correct however the definition given in the previous stage is too general. The figure  is not distinguished from another two figures. In this situation there is the need to use a symbolic name in the definition of the prototype at the more specific level. The definition is as follows:

HUL="W_L", NAME ="Polygon",
if[m_CHul=HUL]
{m_Name= NAME}

In the next stage the additional figures are learned . In this case there is a need to use the symbolic name in the definition of the prototype at the more specific level. The definition is as follows:

HUL="W_L_5", NAME ="Pentagon",
if[m_CHul=HUL]
{m_Name= NAME}

In the next stage the additional figures are learned

.

Their symbolic name at the specific level is given in the form W_L_5_[a]_[s]. The symbol [s] denotes the term {sssss}, where the symbol s can have one of the values from a set of normalized sides (l-large, m-medium, s-small and e-very small). The symbol [a] denotes the term {aaaaa}, where the symbol a can have one of the values from the set of normalized angles (a - acute, o - obtuse and p - right).

In one approach the specific description in the form _[a]_[s] is attached into the name of the object for further reasoning. For example,

HUL="L5", HULSIDES="[mmmmm]", HULANGLE="[oaapo]", m_Nazwa="Pentagon ",
if[m_CHul=HUL]
{m_Name= NAME+ HULSIDES+ HULANGLE}

There is also possibility to define all pentagons that are given by the combination of the symbols {sssss} and {aaaaa}. For example, the object called "Pentagon_Ideal" is given by the symbolic name L5[mmmmm][ooooo]. However the number of definitions will grow very fast and there is also problem with the error when the values of parameters are misinterpreted.

In second approach the new sub-specific classes are derived from the pentagon class. For example, by applying description 'L5[nP], where nP denotes a number of right angles of the pentagon, the pentagons will be divided into five groups described by symbolic names 'L5[0P]', 'L5[1P]', 'L5[2P]', 'L5[3P]', 'L5[4P]'. The new characteristic feature, such as 'symmetry', can be used to derive the additional sub-specific classes. The rules are given in the form:

HUL="L5", NAME0='pentagon', NAME1='pentagonNS', NAME2='pentagonS', NAMETYPE[0]='P0', NAMETYPE[1]='P1', NAMETYPE[2]='P2', NAMETYPE[3]='P3', NAMETYPE[4]='P4',
if[m_CHul=HUL]
{ m_Name= NAME0
if(m_Sym== 0)
{ m_Name= NAME1+NAMETYPE[i]}
else
{ m_Name= NAME0+NAMETYPE[i]}
}}
.

6 Experiment

To perform understanding tasks SUS needs to learn the knowledge about the world. In SUS learning is seen as a continuous process where the knowledge can be acquired on many levels of the categorical hierarchies. Categorical learning that is part of the shape understanding method was applied to learn selected categories. The shape understanding method is implemented as a shape understanding system (SUS) [6]. The shape understanding system is implemented in C++ under Windows 2000. The visual object is extracted from the background (scene) by the application of one of the existing segmentation methods. Object that consists of the different parts is represented by different color. Object that is perceived by SUS is called an exemplar and is given as a set of pixels. Each pixel can have a value from 0 to 256. The set of pixels is divided into two subsets, the background and the figure called a set of critical points. In the case when pixel has more than two values each area of different color is

interpreted as a part of the visual object.
The aim of the experiment was to test that the proposed categorical learning method can learn concepts of the objects of different categories. In the experiment the phantoms that are representative of the different categories are used. In contrary to other methods of learning where only a small number of categories is learned (e.g. different fonts of the letter, mechanical tools) in the categorical learning objects that can belong to any category are learned as a part of the SUS knowledge. Examples of visual objects used in the experiment are shown in Fig. 4. Phantoms represent the following categories: mathematical symbols, road signs, musical symbols, hieroglyphs, card symbols, flag symbols, logos, signs of crosses, letters (Latin, Hebrew, Greek, Arabic, Japanese, Cyrillic), geometrical figures (polygons, 2D figures, graph of functions), mechanical parts, mechanical tools, tools "to eat" (knives, spoons, glasses, mugs, bottles, glasses) and living objects (lives, apple, flower). The categories such as letters were learned by application of the phantoms that represent the different fonts of the letter as well as phantoms that can be 'recognized' as a given letter. Examples of the letters (M, N, T) are shown in Fig. 4. The category 'glass' is represented by different types of glasses (see Fig. 4). Each object named a 'glass' was defined by application of the symbolic names of the complex class. This type of definition makes it possible to interpret a part of the object as a category 'broken glass' or 'occluded glass'.
Objects were generated by application of the MathLink and Mathematica, scanned from available literature, acquired from Microsoft Word (different fonts of letters) or designed by the application of TurboCAD. After preprocessing all objects (phantoms) were stored as 256x256 binary images. During learning and testing stages the learned object was presented to SUS and SUS assigned it to one of the categories. It was assumed that the world that can be understood by SUS consists of categories that SUS has learned. During the testing stage the SUS was given a set of phantoms from each category and a set of phantoms from categories that were not used during the learning stage. In the case when SUS did not understand the phantom it gave the answer 'I do not understand?". In this case the description of the object can be added to the system. In the case when there is more than one interpretation of the phantom presented to SUS, the unique interpretation is obtained based on the derivation of the sub-specific classes.

Figure 4. Examples of visual objects used in the experiment

As the results show the categorical learning gives very good results in learning of the different categories of the object. SUS assumed that the world consists of learned objects and other not known objects. In the case when an examined object belonged to the general category that was learned based on a few examples, SUS interpreted it as a possible object "it can be x". When an examined object belonged to the learned specific category the answer was 'this is x'.
SUS is able to find the occluded object or the incomplete figure by learning the visual concept from the partially occluded objects. The occluded objects are interpreted in the same way as a part of the object. For example, a triangle can be interpreted

as a part of the arrow, or as an occluded arrow. The learning of the occluded objects is the topic of the research focused on the understanding of the distorted objects.

In the experiment the concepts of the phantoms were obtained based on the learning from the selected examples. In SUS understanding is related to the body of knowledge that SUS has learned. For example, SUS understands the concept of a hammer as a tool that is used by man to do a certain kind of work. SUS understands the construction of the tool and can interpret the visual object as a hammer. SUS understands that the concept of a hammer belongs to the category of the man-made object and is a real world object. Based on these categories there is relatively easy to find the conceptual similarities with other categories. For example, a hammer, a nail, and an anvil have the conceptual similarity. The visual similarity refers to the visual concept and describes the objects that look similar. For example, visual aspects of the hammer and specific fonts of the letter ‘T’ look similar. The visual similarity is responsible for obtaining the different results of interpretation of the visual object.

5. Conclusion

In this paper a new method of learning, called the categorical learning, where visual information is learned independently from other conceptual ingredients, is presented. Such an approach makes it possible to concentrate on the visual aspects of the learned objects. The learned information is stored as a hierarchical structure of the categories. At the first stage of learning the categories are represented by their names and all knowledge that is learned is the knowledge about the visual appearances of objects. In the second stage of learning the non-visual information is learned. The non-visual information is used for interpretation of the meaning of the perceived object The proposed method can be applied in designing the new generation of robots that will be acting based on the results of the visual thinking (imagination).

References:

[1] Hsu, J.C., and Hwang, S-Y., A Machine Learning Approach for Acquiring Descriptive Classification Rules of Shape Contours. *Pattern Recognition*, 1997. **30**(2): p. 245-252.

[2] Pahlang, M., and A. Sowmya, A. *Learning Object Models from Real Images*. in *First International Conference on Visual Systems*. 1996. Melbourne.

[3] Mitchell, T.M., *Machine Learning*. 1997, New York: McGraw-Hill.

[4] Les, Z., Shape Understanding. Possible Classes of Shapes. *International Journal of Shape Modelling*, 2001. **7**(1): p. 75-109.

[5] Les, Z., and Les, M., Understanding in Shape Understanding System. *International Journal of Pattern Recognition and Artificial Intelligence*, 2004. **18**(4): p. 727-745.

[6] Les, Z., and Les, M., Shape Understanding System-the System of Experts. *International Journal of Intelligent Systems*, 2004. **19**(10): p. 949-978.