BAYESIAN NETWORK MODEL FOR OBJECT CONCEPT

Yasuhito Shinchi, Yosuke Sato, Takayuki Nagai

Department of Electronic Engineering
The University of Electro-Communications
1-5-1 Chofugaoka Chofu-shi, Tokyo, 182-8585 Japan
{shinchi, satouyousuke, tnagai}@apple.ee.uec.ac.jp

ABSTRACT

This paper discusses a computational model for object concept formation. We propose a model of object concept based on the relationship between shape and function. Implementation of the proposed framework using Bayesian Network is presented. At this point we need an explicit definition of object function. In the proposed model each function is defined as certain changes in a target object caused by the object. Therefore each function is represented by a feature vector which quantifies the changes in the target. Then the function is abstracted from these feature vectors using the bayesian learning approach. The system can form object concept by observing the human tool use based on the abstract function and shape information. Furthermore, it is demonstrated that the learned model (object concept) enables the system to infer the property of unseen object. The system is evaluated using 35 hand tools, which reveals validity of the proposed framework.

Index Terms— Bayesian Network, object concept, object function, object recognition

1. INTRODUCTION

As the recent developments in humanoid robotics, there is growing interest in object recognition and learning, since they are essential tasks for robots to work in our surrounding environments. Most frameworks for recognition and learning are based only on visual features. It seems that those are insufficient for 'understanding' of objects, since each object has its own intended use leading to the function, which is the key to object concept[2][9]. Of course the appearance depends on its function, since many objects have certain forms resulting from their functions. This fact is especially-pronounced in hand tools. Thus the visual learning and recognition of hand tools may succeed to some extent. However, such classification does not give any information on their functions. The important point is not classification in its own right but rather inference of the function through the classification. We believe that must be the basis of 'understanding' (object concept). Therefore objects must be learned, categorized and recognized through their functions.

In this paper objects (hand tools) are modeled as the relationship between their shapes and functions. The proposed approach uses the model, which relates shape and function, for learning and recognizing objects. The shape is defined as contour of the object, while the function is defined as certain changes in target object caused by the object. Each function is represented by a feature vector which quantifies the changes in the target. Then the function is abstracted from these feature vectors using the bayesian learning approach[3]. All information can be obtained by observing the scene, in which a man uses the hand tool. For the model of object concept, Bayesian Network is utilized. The conditional probability tables, which are parameters of the model, are estimated by applying EM algorithm to the observed shape and function information. This process can be seen as the learning of objects based on their functions. Since the function and shape are stochastically connected in the model, inference of unseen object's function is possible as well as recognizing its category.

Related works are roughly classified into three categories. One of these is an attempt to recognize objects through their functions [8][9][10]. Although those works share the same idea of us, the authors do not consider the learning process of object function. Thus the function of each object must be defined and programmed manually. Secondly, unsupervised visual categorization of objects has been studied extensively[4][5]. However, function is not taken into consideration. Thirdly, there has been a research on object recognition through human action[6]. The authors relate object recognition with human action, which represents how to use it, rather than the object function itself. In [7], authors have reported the model for robot tool use. However, they do not consider categorization and the robot can not cope with unknown objects. The proposed framework differs from those works in important ways. The key point of the proposed approach is learning of the relationship between shape and function. This approach may lead to a computational model for the affordance[1].

2. FORMING OBJECT CONCEPT

2.1. Bayesian Network for Object Concept

To 'understand' objects a novel framework, which differs from conventional matching-based 'recognition' approach, is required. Here we define 'understanding' for an object as inference of its function. For example, to understand 'scissors' is to infer their function, that is, cutting their target objects. Here is the problem to be considered, that is, what is the definition of the function? Especially by almost all hand tools, target object undergoes some physical change. For example, scissors change the shape and number of target object, and pens can change target's surface brightness. These various changes in a scene can be observed as a feature vector, which results in our definition of function. A detail description of the function will be given in the following section.

The schematic diagram of the above discussion is shown in Fig.1 (a). Then Fig.1(a) can be rewritten using graphical model as in Fig.1(b). It should be noted that the following relationship is used to rewrite Fig.1(a) to Fig.1(b).

$$P(S)P(O|S)P(F|O) = P(O)P(S|O)P(F|O). \qquad (1)$$

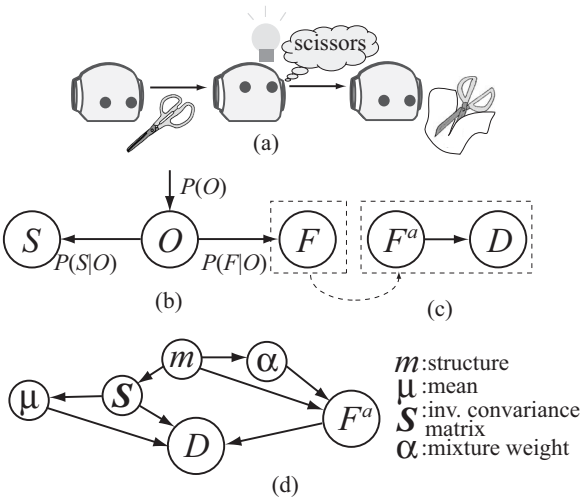


Fig. 1. A model of object concept. (a)Schematic diagram. (b)Graphical model representation of (a). (c)Details of the node F in (b). (d)Gaussian Mixture Model of (c). It should be noted that each parameter has its conjugate prior.

Thus the problem considered in this paper results in the parameter estimation and inference using the graphical model in Fig.1(b). Of course the model is too simple to explain all aspects of object understanding. In fact, more complex factors such as usage of the tool etc. are important and should be taken into account. This is an issue in the future and now we focus the discussion on the implementation of the system based on the model in Fig.1(b). The Bayesian Network in Fig.1(b) has three nodes; one of these is unobservable object concept O and the other nodes are observable shape S and function F. To be precise F is not observable. In Fig.1(c), details of the node F is illustrated. In the figure D and F^a represent observable feature vector and 'abstract function', which is abstracted from feature vectors using the bayesian learning approach, respectively. Thus the proposed model consists of two steps. First step is to estimate abstract function F^a from the observable feature vector$D$. In the object concept model, abstract function F^a is used as F.

2.2. Learning Algorithm

The joint probability of shape S, function F and object O can be written as

$$P(S,F,O) = P(O)P(S|O)P(F|O). \tag{2}$$

The parameters in the above equation $P(O)$, $P(S|O)$ and $P(F|O)$ are estimated using the EM algorithm, as the model contains unobserved latent variable. Let the parameter be θ, the problem is a maximization of the following equation:

$$\begin{aligned}\log \int P(S,F,O|\theta)dO &\geq F(q(O),\theta) \\ &= \int q(O|S,F,\hat{\theta}) \log \frac{P(S,F,O|\theta)}{q(O|S,F,\hat{\theta})} dO. \end{aligned} \tag{3}$$

Then the lower limit $F(q(O),\theta)$ is maximized iteratively with respect to $q(O)$ and θ one after the other. The maximization with respect to $q(O)$ is to compute

$$P(O|S,F) = \frac{P(O)P(S|O)P(F|O)}{\sum_o P(O)P(S|O)P(F|O)}. \tag{4}$$

On the other hand the maximization with respect to θ is equivalent to maximize the Q-function;

$$Q(\theta) = \langle \log P(S,F,O|\theta) \rangle_{q(O|S,F,\hat{\theta})}. \tag{5}$$

The parameter θ can be updated by solving $\partial Q(\theta)/\partial\theta = 0$. The EM algorithm alternates following two steps stating from initial values and converges to local minima.

[E-step] Compute Eq.(4).

[M-step]

$$P(O) \propto \sum_S \sum_F N(S,F)P(O|S,F), \tag{6}$$

$$P(S|O) \propto \sum_F N(S,F)P(O|S,F), \tag{7}$$

$$P(F|O) \propto \sum_S N(S,F)P(O|S,F), \tag{8}$$

where $N(S,F)$ denotes how many times $\{S,F\}$ occurred in the observations.

2.3. Inference

An object (category) can be recognized from observed shape and function using the learned model as

$$\begin{aligned}&\operatorname*{argmax}_O P(O|S,F) = \\ &\quad \operatorname*{argmax}_O \frac{P(O)P(S|O)P(F|O)}{\sum_O P(O)P(S|O)P(F|O)}. \end{aligned} \tag{9}$$

It is possible to infer the unseen object function only from the observed shape information. Inversely typical shape of the object that has a specific function can be derived. Inference of object function can be carried out by

$$\begin{aligned}&\operatorname*{argmax}_F P(F|S) = \\ &\quad \operatorname*{argmax}_F \frac{\sum_O P(O)P(F|O)P(S|O)}{\sum_O \sum_F P(O)P(F|O)P(S|O)}. \end{aligned} \tag{10}$$

3. SHAPE AND FUNCTIONS

3.1. Object Shape

There are two different attributes of object parts. One is functional parts and the other is non-functional ones. The clipper blade and scissors handle are examples of functional parts, which are requisite for scissors. The relative location of these parts is also important. On the other hand, non-functional parts are not directly linked to the object function. The object shape reflects these two types of parts. Therefore, only functional parts should be extracted to capture the relationships between shapes and functions correctly. However simple object contour is used as a first step in this paper.

The lower part of Fig.2 illustrates the processing for computing object shape. At first the object region is extracted from images as

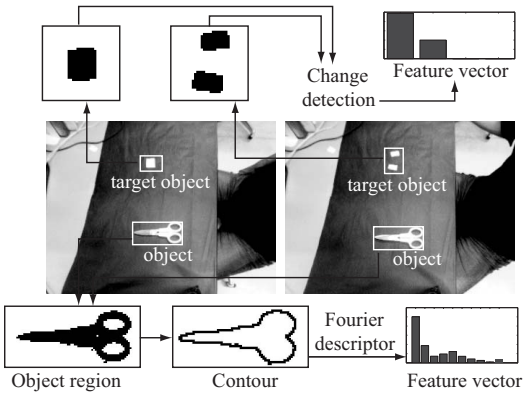


Fig. 2. Processings for object shape and function.

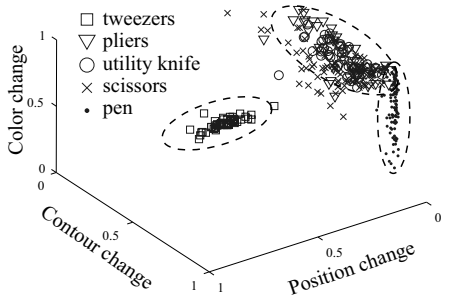


Fig. 3. 3D-plot of feature vectors of object functions.

shown in the figure. The contour is then transformed into frequency domain using Fourier descriptor and high frequency components are omitted for compact representation of the feature vector. Finally the feature vector is vector quantized using the code book, which is generated in advance by k-means clustering of many object shapes. In the examples below the number of clusters is defined as 10.

3.2. Feature Extraction for Functions

As we mentioned earlier, function of a tool is defined as the pattern of certain changes in its target object. It is very important to select changes to be observed, since it directly affects the ability of the system to discover object functions.

Here four features are computed considering properties of general hand tools. (1)Color change on the surface of target object; this change can be captured by computing the correlation coefficient between color histograms of target object before and after manipulation. (2)Contour change of target object; to capture this change the correlation coefficient between Fourier descriptors of target object before and after manipulation is computed. (3)Barycentric position change of target object; the relative distance between barycentric positions of target object before and after manipulation is computed. (4)Change in number of target object; this can be detected by counting the connected components relevant to target object.

The upper part of Fig.2 illustrates an example of the feature extraction. As shown in the figure, above four features are extracted from images before and after manipulation.

Fig. 4. Hand tools used in the experiment. (a)Set A. (b) Set B.

3.3. Bayesian Learning of Functions

The object functions are modeled by GMM (Gaussian Mixture Model) as in Fig.1(d) using the feature vectors described above. This modeling process corresponds to abstraction of object functions. Figure 3 shows 3D-plot of features that motivates us to use GMM. Three clusters, which represent different functions, can clearly be seen in the figure. The Variational Bayes (VB) framework[3] is used for the parameter estimation, since the number of abstract functions can be estimated as an optimal model structure.

In the VB approach the following marginal likelihood is maximized:

$$\mathcal{L}(D) = \log P(D) = \log \sum_m \sum_{F^a} \int_\theta P(D, F^a, \boldsymbol{\theta}, m) d\boldsymbol{\theta}, \quad (11)$$

where $\boldsymbol{\theta} = \{\alpha, \mu, \boldsymbol{S}\}$ denotes model parameters as shown in Fig.1(d). Now the factorizable variational posterior $q(F^a, \boldsymbol{\theta}, m)$ is introduced to make the problem tractable. Then the problem becomes the maximization of the free energy $F[q]$ with respect to q.

$$F[q] = \sum_m q(m) \left\{ \left\langle \log \frac{P(D, F^a|\boldsymbol{\theta}, m)}{q(F^a|m)} \right\rangle_{q(F^a|m), q(\theta|m)} + \sum_i \left\langle \log \frac{P(\boldsymbol{\theta}_i|m)}{q(\boldsymbol{\theta}_i|m)} \right\rangle_{q(\theta_i|m)} + \frac{P(m)}{q(m)} \right\}. \quad (12)$$

Thus we finally obtain,

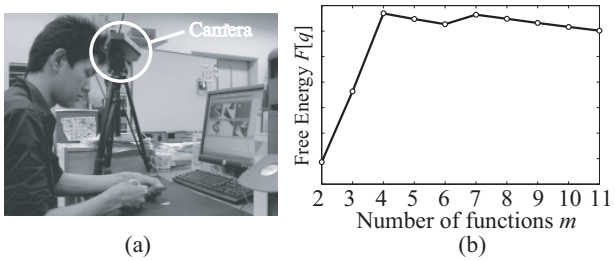
$$q(F^a|m) = C \exp \langle \log P(D, F^a|\boldsymbol{\theta}, m) \rangle_{q(\theta|m)},$$
$$q(\alpha|m) = C' P(\alpha|m) \exp \langle \log P(D, F^a|\boldsymbol{\theta}, m) \rangle_{q(F^a|m), q(\mu, S|m)},$$
$$q(\mu, \boldsymbol{S}|m) = C'' P(\mu, \boldsymbol{S}|m) \exp \langle \log P(D, F^a|\boldsymbol{\theta}, m) \rangle_{q(F^a|m), q(\alpha|m)}.$$

We solve these equations iteratively. The optimal model structure can be obtained by finding the maximum $F[q]$ among various m.

4. EXPERIMENTS

4.1. Experimental Setup

A total of 35 objects with 5 categories are employed in the experiments. These 35 objects are divided into two groups. Figure 4(a) is the set A containing a total of 23 hand tools (7 scissors, 8 pens, 2 pliers, 3 tweezers, and 3 utility knives). The set B consists of a total of 12 hand tools (3 scissors, 3 pens, 2 pliers, 2 tweezers and 2 utility



(a) (b)

Fig. 5. (a)A snapshot of the system. (b)Structure vs. free energy.

Table 1. Learning results for 350 data.

	scissors	pens	pliers	tweezers	knives
O_1	95	0	0	0	1
O_2	0	109	3	0	0
O_3	5	1	37	0	0
O_4	0	0	0	50	0
O_5	0	0	0	0	49

knives), which are shown in Fig.4(b). Figure 5(a) shows the actual system setup. The camera is fixed to capture the user's hands and takes images during the manipulation. The tool and its target object are extracted based on background difference method, and then the system computes shape and function information as we mentioned earlier. Three experiments were conducted using this system.

4.2. Finding Abstract Functions

At first the function model in Fig.1(d) was trained. Each of 35 hand tools used 10 times and total of 350 feature vectors were obtained. The VB algorithm was applied to the data to estimate the parameters and optimal structure, i.e. number of abstract functions. Figure 5(b) shows free energy over the number of functions m. The figure implies that four functions explains the data best. In fact, we have confirmed these four functions correspond to 'cut', 'write', 'move' and 'deform'. In the following experiments, the abstract function model, which was obtained in this experiment, is used.

4.3. Results of Learning

The tools in both A and B sets are used in the second experiment for the training of Fig.1(b). Each of 35 hand tools was used 10 times; hence the model was trained using a total of 350 data. Then Eq.(9) was used to classify 350 data. The classification result is compared with ground truth to evaluate how well the objects are categorized. The result is shown in Tab.1. From the table one can see that the system has reasonably categorized the objects.

4.4. Results of Inference

A total of 230 data, which consists of 10 times use of each hand tool in the A set, were used to train the model. And then, the system observed unseen objects in the B set and inferred their functions from the observable shape. Equation(10) was used to identify the function. The result is given in Tab.2. It can be seen that the system inferred object functions correctly except for 'knife2', whose contour is very close to that of a pen. This similarity in appearance leads to the misrecognition of its function. In order to avoid this difficulty, we need more sophisticated visual features rather than the simple contour.

Table 2. Inference results for unseen objects.

object	function	object	function
scissors1	cut	pliers1	deform
scissors2	cut	pliers2	deform
scissors3	cut	tweezers1	move
pen1	write	tweezers2	move
pen2	write	knife1	cut
pen3	write	knife2	write

5. CONCLUSIONS

This paper has proposed a novel framework for object understanding. Implementation of the proposed framework using Bayesian Network has been presented. Although the result given in this paper is preliminary one, we have shown that the system can form object concept by observing the performance by human hands. The online learning is left for the future works. Moreover the model should be extended so that it can represent the object usage and target object.

6. REFERENCES

[1] J. J. Gibbson. *The Ecological Approach to Visual Perception*. Lawrence Eribaum, Hillsdale, NJ, 1979.

[2] B. Landau, L. Smith, and S. Jones. Object shape, object function, and object name. *Journal of Memory and Language*, 38(ML972533):1–27, 1998.

[3] H. Attias. Infering parameters and structure of latent variable models by variational bayes. *in Proc. of Uncertainty in Artificial Intelligence*, 1999.

[4] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman. Discovering object categories in image collections. *AI Memo*, 2005-005:1–12, February 2005.

[5] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. *in Proc. CVPR*, February 2003.

[6] A. Kojima, M. Higuchi, T. Kitahashi, and K. Fukunaga. Toward a cooperative recognition of human behaviors and related objects. *Proc. of Scecond Int. Workshop on Man-Machine Symbiotic Systems*, pages 195–206, November 2004.

[7] T. Ogura, K. Okada, and M. Inaba. Humanoid tool operating motion generation. *in Proc. of The 23rd Annual Conf. of the Robotics Society of Japan*, 1F15, 2005 (in japanese).

[8] E. Rivlin, S. J. Dickinson, and A. Rosenfeld. Recognition by functional parts. *Computer Vision and Image Understanding: CVIU*, 62(2):164–176, 1995.

[9] L. Stark, K. Bowyer, A. Hoover, and D. B. Goldgof. Recognizing object function through reasoning about partial shape descriptions and dynamic physical properties. *Roceedings of The IEEE*, 84(11):1640–1656, November 1996.

[10] K. Woods, D. Cook, L. Hall, K. W. Bowyer, and L. Stark. Learning membership functions in a function-based object recognition system. *Journal of Artificial Intelligence Research*, 3:187–222, 1995.