

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**4,800**

Open access books available

**122,000**

International authors and editors

**135M**

Downloads

Our authors are among the

**154**

Countries delivered to

**TOP 1%**

most cited scientists

**12.2%**

Contributors from top 500 universities



**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



## Forming object concept using Bayesian network

Tomoaki Nakamura and Takayuki Nagai  
*The University of Electro-Communications*  
Japan

### 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 (Landau et al., 1998; Stark et al., 1996). Of course, appearance is deeply related with functions, 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', which we call object concept. Therefore objects must be learned, i.e. categorized, and recognized through their functions. In this chapter objects (hand tools) are modeled as the relationship between appearance and functions. The proposed approach uses the model, which relates appearance and functions, for learning and recognizing objects.

The appearance is defined as a visual feature of the object, while the function is defined as certain changes in work objects caused by a tool. Each function is represented by a feature vector which quantifies the changes in the work object. Then the function is abstracted from these feature vectors using the Bayesian learning approach (Attias, 1999). 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 visual features and function information. This process can be seen as the learning of objects based on their functions. Since the function and appearance 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 (Rivlin et al., 1995; Stark et al., 1996; Woods et al., 1995). Although those works share the same idea with 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 (Fergus et al., 2003; Sivic et al., 2005). However, function is not taken into consideration. Thirdly, there has been research on object recognition through human action (Kojima et al., 2004). The authors relate object recognition with human action, which represents how to use it, rather than the object function itself. In (Ogura et al., 2005), 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 appearance and function. This approach may lead to a computational model for the affordance (Gibson, 1979).

This chapter is organized as follows: the following section discusses an object concept model based on the Bayesian Network. Then, the details of the model such as object appearance and the function model are described in section 3. Section 4 shows some experimental results to validate the proposed framework and this chapter is concluded in section 5.

## 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' of an object as inference of its function. For example, to understand 'scissors' is to infer their function, that is, cutting the work objects. Here is the problem to be considered, that is, what is the definition of the function? Especially by almost all hand tools, the work object undergoes some physical change. For example, scissors change shape and number of the work object, and pens can change 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).

$$P(I)P(O|I)P(X_V|O)P(F|O) = P(O)P(I|O)P(X_V|O)P(F|O). \quad (1)$$

Thus the problem considered in this chapter 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 our discussion on the implementation of the system based on the model in Fig.1(b).

The Bayesian Network in Fig.1(b) has four nodes; one of these is unobservable object concept  $O$  and the other nodes are observable object(scene) ID  $I$ , visual feature  $X_V$  and function  $F$ . To be precise  $F$  is not observable. In Fig.1(c), details of the node  $F$  is illustrated. In the figure  $X_F$  and  $Z_F$  represent observable feature vector and 'abstract function', which is abstracted from feature vectors using Bayesian learning approach, respectively.

### 2.2 Learning Algorithm

From Fig.1(c), the joint probability of  $I$ ,  $X_V$ ,  $X_F$ ,  $O$  and  $Z_F$  can be written as

$$P(I, X_V, X_F, O, Z_F) = P(O)P(I|O)P(X_V|O)P(Z_F|O)P(X_F|Z_F). \quad (2)$$

The parameters in the above equation  $P(O)$ ,  $P(I|O)$ ,  $P(X_V|O)$  and  $P(Z_F|O)$  are estimated using the EM algorithm, as the model contains unobserved latent variable. It should be noted that  $P(X_F|Z_F)$  is given by the abstract function model(Gaussian Mixture Model) as we describe later. Let the parameters be  $\theta$ , the problem is a maximization of the following equation:

$$L(D) = \log \sum_{Z_F} \sum_O P(I, X_V, X_F, O, Z_F | \theta). \quad (3)$$

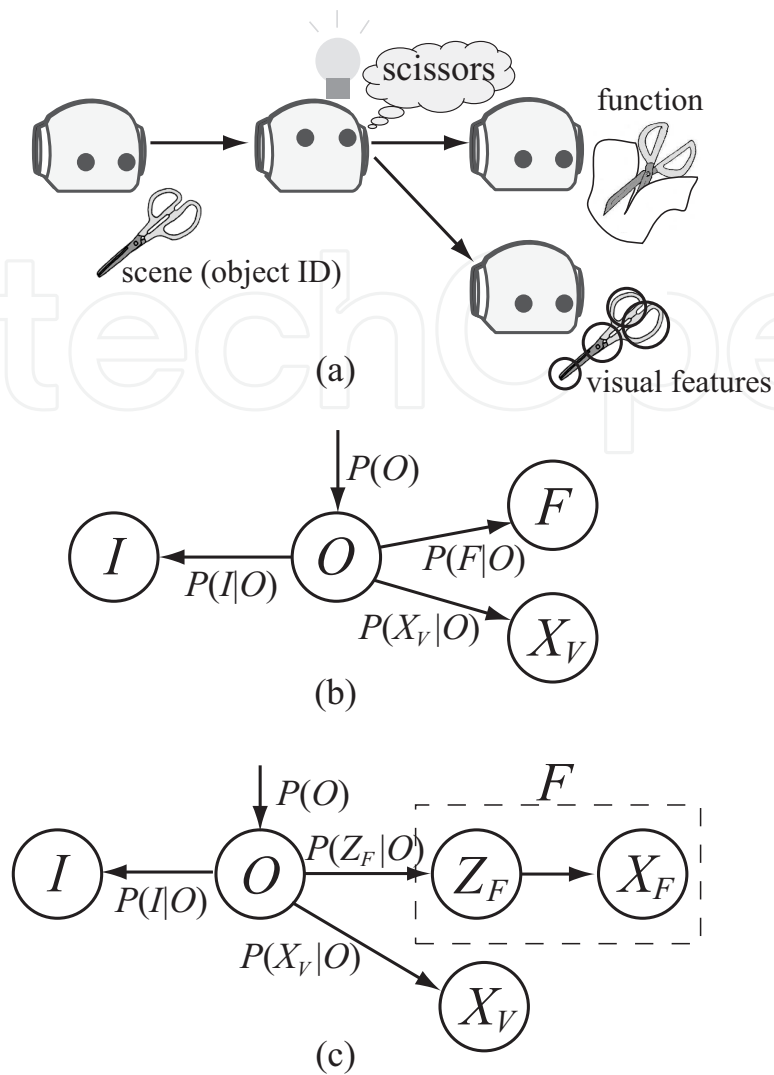


Fig. 1. A model of object concept. (a)Schematic diagram. (b)Graphical model representation of (a). (c)Details of the node  $F$  in (b).

By applying Jensen’s inequality, we obtain

$$\begin{aligned}
 L(D) &= \log \sum_{Z_F} \sum_O q(O, Z_F | I, X_V, X_F, \hat{\theta}) \frac{P(I, X_V, X_F, O, Z_F | \theta)}{q(O, Z_F | I, X_V, X_F, \hat{\theta})} \\
 &\geq F(q, \theta) = \sum_{Z_F} \sum_O q(O, Z_F | I, X_V, X_F, \hat{\theta}) \log \frac{P(I, X_V, X_F, O, Z_F | \theta)}{q(O, Z_F | I, X_V, X_F, \hat{\theta})}.
 \end{aligned} \tag{4}$$

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

$$q(O, Z_F | I, X_V, X_F, \hat{\theta}) = \frac{P(O)P(I|O)P(X_V|O)P(Z_F|O)P(X_F|Z_F)}{\sum_{Z_F} \sum_O P(O)P(I|O)P(X_V|O)P(Z_F|O)P(X_F|Z_F)}. \tag{5}$$

On the other hand the maximization with respect to  $\theta$  is equivalent to maximize the Q-function;

$$Q(\theta) = \langle P(I, X_V, X_F, Z_F, O | \theta) \rangle_{q(O, Z_F | I, X_V, X_F, \hat{\theta})}. \quad (6)$$

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

[E-step] Compute Equation 5.

[M-step]

$$P(O) \propto \sum_I \sum_j \sum_{Z_F} \{n(I, X_{Vj})q(Z_F, O | I, X_{Vj}, X_F, \hat{\theta})\}, \quad (7)$$

$$P(I|O) \propto \sum_j \sum_{Z_F} \{n(I, X_{Vj})q(Z_F, O | I, X_{Vj}, X_F, \hat{\theta})\}, \quad (8)$$

$$P(X_{Vj}|O) \propto \sum_I \sum_{Z_F} \{n(I, X_{Vj})q(Z_F, O | I, X_{Vj}, X_F, \hat{\theta})\}, \quad (9)$$

$$P(Z_F|O) \propto \sum_j \sum_I \{n(I, X_{Vj})q(Z_F, O | I, X_{Vj}, X_F, \hat{\theta})\}, \quad (10)$$

where  $X_{Vj}$  represents the  $j$ -th dimension of  $X_V$  and  $n(I, X_{Vj})$  denotes how many times  $\{I, X_{Vj}\}$  occurred in the observations. It should be noted that  $P(X_V|O)$  can be written as  $P(X_V|O) = \prod_j P(X_{Vj}|O)$ .

### 2.3 Inference

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

$$\operatorname{argmax}_O P(O | X_V, X_F, I) = \operatorname{argmax}_O \frac{P(O)P(I|O)P(X_V|O) \sum_{Z_F} \{P(Z_F|O)P(X_F|Z_F)\}}{\sum_O [P(O)P(I|O)P(X_V|O) \sum_{Z_F} \{P(Z_F|O)P(X_F|Z_F)\}]}. \quad (11)$$

It is worth noting that Equation 11 is for the known object  $I$ . In order to apply Equation 11 to the unseen object  $\hat{I}$ ,  $P(\hat{I}|O)$  and  $P(O)$  must be recalculated using the EM-algorithm described in the foregoing section. At this time  $P(X_V|O)$  and  $P(Z_F|O)$  are fixed. This idea is called fold-in heuristics described in (Hofmann, 2001).

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

$$\operatorname{argmax}_{Z_F} P(Z_F | X_V, \hat{I}) = \operatorname{argmax}_{Z_F} \frac{\sum_O P(O)P(\hat{I}|O)P(X_V|O)P(Z_F|O)}{\sum_O P(O)P(\hat{I}|O)P(X_V|O)}. \quad (12)$$

The fold-in heuristics should be applied to the calculation of  $P(\hat{I}|O)$  and  $P(O)$ .

## 3. Visual Information and Functions

### 3.1 Object appearance ( $X_V$ )

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

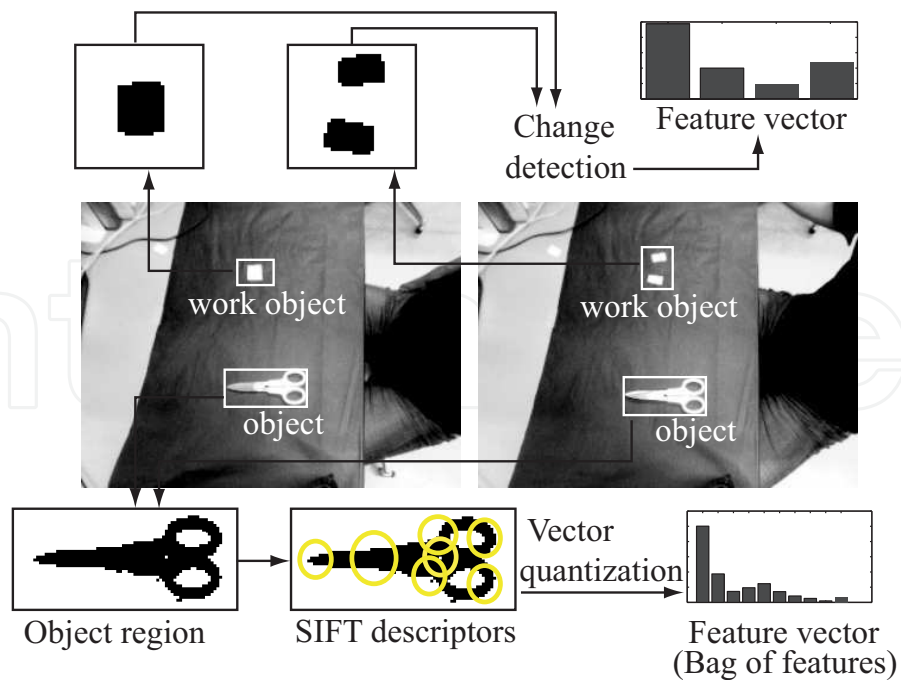


Fig. 2. The image processing for the object appearance and functions.

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 relationship between appearances and functions correctly. We use SIFT descriptors (Lowe, 2004) in order to extract parts of the object and then the object appearance is represented by bag of features model.

The lower part of Fig.2 illustrates the processing for computing visual information. At first the object region is extracted from images as shown in the figure. The SIFT descriptors are computed in the object region. These computed descriptors are vector quantized using the pre-defined code book and frequency count is taken for the bag of features representation.

In Fig.3, some examples of the actual SIFT key points and histograms. Each histogram given in the figure is only a part of whole 500 dimensional information. One can see the similarities between within class objects.

### 3.2 Feature Extraction for Functions ( $X_F$ )

As we mentioned earlier, the function of a tool is defined as the pattern of certain changes in its work 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 the work object; this change can be captured by computing the correlation coefficient between color histograms of the work object before and after manipulation ( $C_{before}$  and  $C_{after}$ ).

$$x_C = Cr(C_{after}, C_{before}), \quad (13)$$

where  $Cr(a, b) = (a \cdot b) / |a||b|$ .

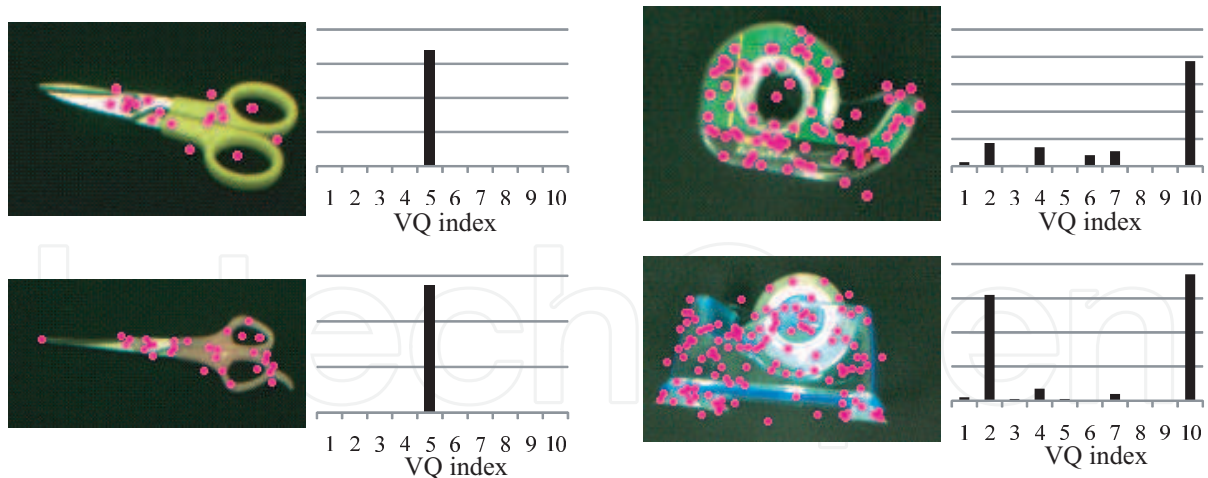


Fig. 3. Examples of SIFT keypoints and histograms.

(2) Contour change of the work object; to capture this change the correlation coefficient between Fourier descriptors of the work object before and after manipulation is computed ( $\mathbf{F}_{before}$  and  $\mathbf{F}_{after}$ ).

$$x_F = Cr(\mathbf{F}_{after}, \mathbf{F}_{before}). \quad (14)$$

(3) Barycentric position change of the work object; the relative distance between barycentric positions of a work object before and after manipulation ( $\mathbf{G}_{before}$  and  $\mathbf{G}_{after}$ ) is computed.

$$x_G = |\mathbf{G}_{after} - \mathbf{G}_{before}|. \quad (15)$$

(4) Change in number of the work object; this can be detected by counting the connected components relevant to a work object.

$$x_N = N_{after} - N_{before}, \quad (16)$$

where  $N_{before}$  and  $N_{after}$  represent the numbers of connected components of the work object before and after manipulation, respectively.

Then, the feature vector can be written as  $\mathbf{x} = (x_N, x_G, x_C, x_F)$ . 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.

### 3.3 Bayesian Learning of Functions

#### 3.3.1 The graphical model for GMM

The object functions are modeled by Gaussian Mixture Model (GMM) as in Fig.4 using the feature vectors described above. This modeling process corresponds to abstraction of object functions. Figure 5 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 (Attias, 1999) is used for the parameter estimation, since the number of abstract functions can be estimated as an optimal model structure.

Figure 4 illustrates the graphical model for the GMM. This model corresponds to  $F$ , which is denoted by the dashed box, in the whole model Fig.1(c). In Fig.4,  $\mathbf{x}_n (n = 1, \dots, N)$  represents the observable change vectors of the work objects and  $N$  is the number of training samples.

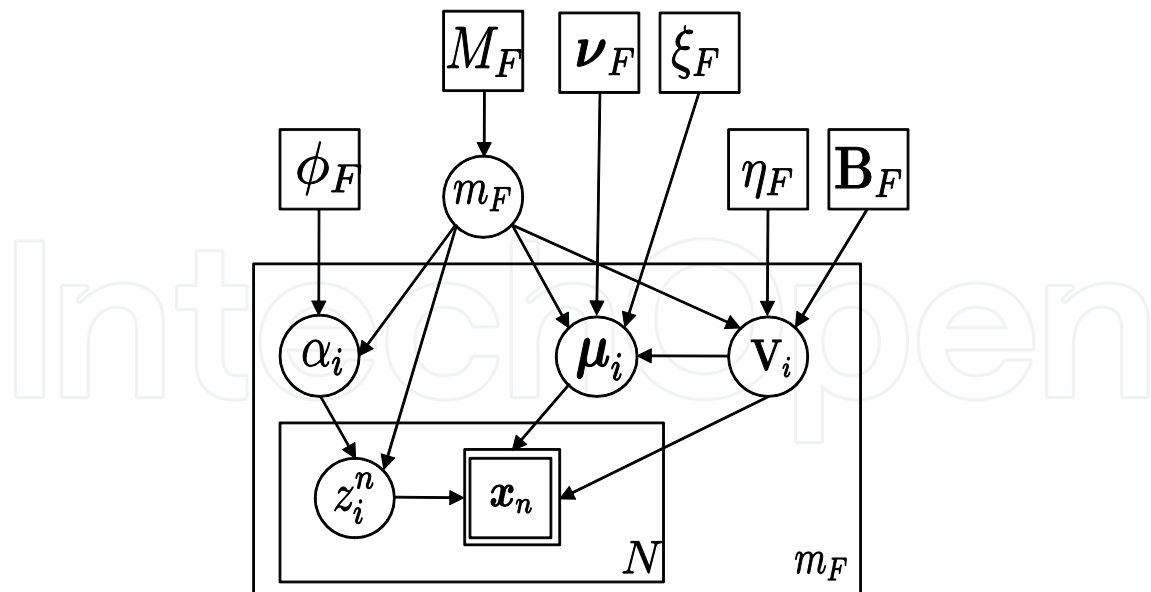


Fig. 4. The detailed graphical model for functions.

$m_F$ ,  $\mu_i$ ,  $V_i$  and  $\alpha_i$  denote the number of functions,  $i$ -th component of  $m_F$  sets of mean vectors, precision matrices and mixture ratios, respectively. All of these parameters have their own prior distributions. The prior distribution of the multinomial distribution  $\alpha = \{\alpha_1, \dots, \alpha_{m_F}\}$  is the Dirichlet distribution with the degrees of freedom  $\phi_F$ . The prior distribution of the mean vector  $\mu_i$  is the Gaussian distribution with the mean vector  $\nu_F$  and the precision matrix  $\xi_F V_i$ .  $V_i$  has the following Wishart distribution, which is parameterized by  $\eta_F$  and  $B_F$ , as the prior distribution:

$$\mathcal{W}(V_i | \eta_F, B_F) \propto |V_i|^{1/2(\eta_F - d - 1)} \exp\left(-\frac{1}{2} \text{Tr}(V_i B_F)\right), \quad (17)$$

where  $\mathcal{W}()$  represents the Wishart distribution. The model structure  $m_F$  also has the uniform distribution  $M_F$  as its prior distribution.  $z_i^n$  denotes a latent variable, which represents the functions.

In the variational Bayesian approach, the following marginal likelihood of the observations  $D = \{x_1, \dots, x_N\}$  is considered:

$$\mathcal{L}(D) = \log P(D) = \log \sum_{m_F} \sum_{Z_F} \int_{\theta} P(D, Z_F, \theta, m_F) d\theta, \quad (18)$$

where  $Z_F = \{z_i^n\}_{n=1, i=1}^{N, m_F}$  and  $\theta$  represent latent variables and a set of model parameters, respectively. Now the variational posterior  $q$  is introduced to make the problem tractable.

$$q(Z_F, \theta, m_F) = q(m_F) q(Z_F | m_F) \prod_k q(\theta_k | m_F), \quad (19)$$



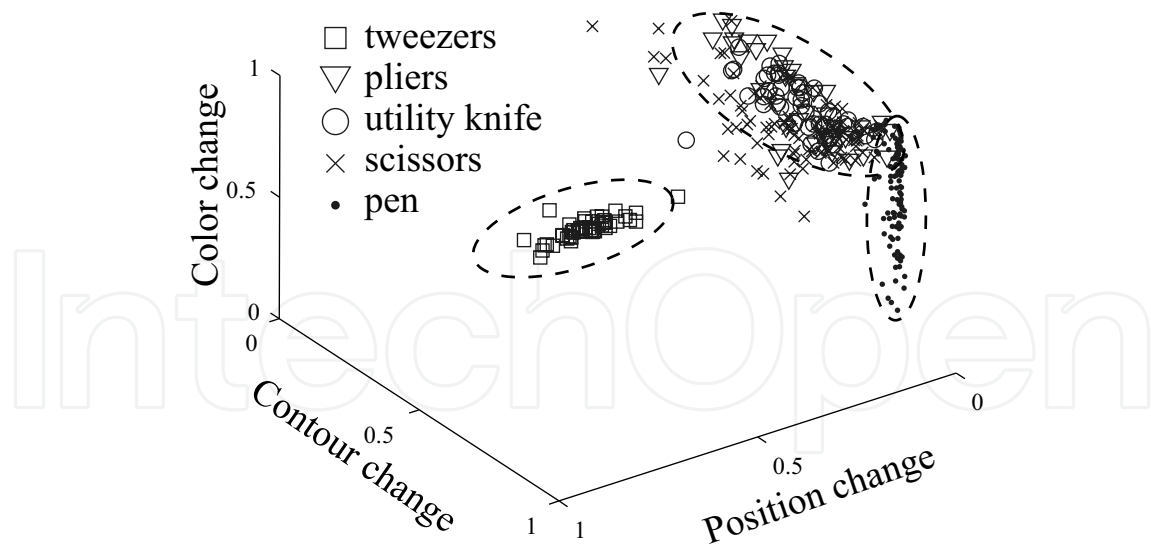


Fig. 5. 3D-plot of feature vectors of object functions.

where  $\theta$  is assumed to be decomposed into  $k$  independent parameters  $\theta_k (k = 1, \dots, K)$ . Then,  $\mathcal{L}(\mathbf{D})$  can be written as follows:

$$\begin{aligned}
 \mathcal{L}(\mathbf{D}) &= \log \sum_{m_F} \sum_{\mathbf{Z}_F} \int_{\theta} q(\mathbf{Z}_F, \theta, m_F) \frac{P(\mathbf{D}, \mathbf{Z}_F, \theta, m_F)}{q(\mathbf{Z}_F, \theta, m_F)} d\theta \\
 &= \sum_{m_F} \sum_{\mathbf{Z}_F} \int_{\theta} q(\mathbf{Z}_F, \theta, m_F) \log \frac{q(\mathbf{Z}_F, \theta, m_F)}{P(\mathbf{Z}_F, \theta, m_F | \mathbf{D})} d\theta \\
 &\quad + \sum_{m_F} \sum_{\mathbf{Z}_F} \int_{\theta} q(\mathbf{Z}_F, \theta, m_F) \log \frac{P(\mathbf{D}, \mathbf{Z}_F, \theta, m_F)}{q(\mathbf{Z}_F, \theta, m_F)} d\theta \\
 &\equiv \text{KL}(q(\mathbf{Z}_F, \theta, m_F), P(\mathbf{Z}_F, \theta, m_F | \mathbf{D})) + \mathcal{F}[q],
 \end{aligned} \tag{20}$$

where  $F[q]$  and  $KL$  denote free energy and Kullback-Leibler divergence, respectively. Since  $L(\mathbf{D})$  does not depend on  $q$ , the maximization of  $F[q]$  with respect to  $q$  is equivalent to the minimization of Kullback-Leibler divergence between  $q$  and true posterior  $P$ . Therefore, variational posterior  $q$ , which maximizes  $F[q]$ , is the best approximation to the true posterior  $P$ . The optimum variational posterior of each parameter can be obtained by maximizing  $\mathcal{F}[q]$  with respect to  $\theta_k$  using Lagrange multipliers method.

$$\begin{aligned}
 \mathcal{F}[q] &= \sum_{m_F} q(m_F) \left\{ \left\langle \log \frac{P(\mathbf{D}, \mathbf{Z}_F | \theta, m_F)}{q(\mathbf{Z}_F | m_F)} \right\rangle_{q(\mathbf{Z}_F | m_F), q(\theta | m_F)} \right. \\
 &\quad \left. + \sum_{k=1}^K \left\langle \log \frac{P(\theta_k | m_F)}{q(\theta_k | m_F)} \right\rangle_{q(\theta_k | m_F)} + \log \frac{P(m_F)}{q(m_F)} \right\}.
 \end{aligned} \tag{21}$$

Derivation of  $q(\mathbf{Z}_F | m_F)$ ,  $q(\theta_k | m_F)$  and  $q(m_F)$  are given hereafter.

### 3.3.2 Derivation of $q(\mathbf{Z}_F|m_F)$

$q(\mathbf{Z}_F|m_F)$  can be obtained by maximizing  $\mathcal{F}(q)$  under the constraint  $\sum_{\mathbf{Z}_F} q(\mathbf{Z}_F|m_F) = 1$  using the Lagrange multiplier method. From Equation 21,

$$\mathcal{F}[q(\mathbf{Z}_F|m_F)] = \left\langle \log \frac{P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F)}{q(\mathbf{Z}_F|m_F)} \right\rangle_{q(\mathbf{Z}_F|m_F), q(\boldsymbol{\theta}|m_F)}. \quad (22)$$

Let  $\lambda$  be a Lagrange multiplier. Then the problem becomes the extreme value problem of the following functional  $\mathcal{J}[q(\mathbf{Z}_F|m_F)]$ ,

$$\mathcal{J}[q(\mathbf{Z}_F|m_F)] = \mathcal{F}[q(\mathbf{Z}_F|m_F)] + \lambda \left( \sum_{\mathbf{Z}_F} q(\mathbf{Z}_F|m_F) - 1 \right) \quad (23)$$

$$\begin{aligned} \frac{\partial \mathcal{J}[q(\mathbf{Z}_F|m_F)]}{\partial q(\mathbf{Z}_F|m_F)} &= \frac{\partial}{\partial q(\mathbf{Z}_F|m_F)} \left[ \sum_{\mathbf{Z}_F} \left\{ q(\mathbf{Z}_F|m_F) \int q(\boldsymbol{\theta}|m_F) \log \frac{P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F)}{q(\mathbf{Z}_F|m_F)} d\boldsymbol{\theta} \right\} \right. \\ &\quad \left. + \lambda \left( \sum_{\mathbf{Z}_F} q(\mathbf{Z}_F|m_F) - 1 \right) \right] \\ &= \frac{\partial}{\partial q(\mathbf{Z}_F|m_F)} \int q(\boldsymbol{\theta}|m_F) \{ q(\mathbf{Z}_F|m_F) \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \\ &\quad - q(\mathbf{Z}_F|m_F) \log q(\mathbf{Z}_F|m_F) \} d\boldsymbol{\theta} + \lambda \\ &= \int q(\boldsymbol{\theta}|m_F) \{ \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) - \log q(\mathbf{Z}_F|m_F) - 1 \} d\boldsymbol{\theta} + \lambda \\ &= \langle \log p(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\boldsymbol{\theta}|m_F)} - \log q(\mathbf{Z}_F|m_F) - 1 + \lambda \\ &= 0 \\ &\Rightarrow q(\mathbf{Z}_F|m_F) = \exp \left\{ \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\boldsymbol{\theta}|m_F)} + \lambda - 1 \right\} \end{aligned} \quad (24)$$

$$\begin{aligned} \frac{\partial \mathcal{J}[q(\mathbf{Z}_F|m_F)]}{\partial \lambda} &= \sum_{\mathbf{Z}_F} q(\mathbf{Z}_F|m_F) - 1 \\ &= \sum_{\mathbf{Z}_F} \exp \left\{ \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\boldsymbol{\theta}|m_F)} + \lambda - 1 \right\} - 1 = 0 \\ &\Rightarrow \exp(\lambda - 1) = \frac{1}{\sum_{\mathbf{Z}_F} \exp \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\boldsymbol{\theta}|m_F)}} \end{aligned} \quad (25)$$

From Equations 24 and 25,  $q(\mathbf{Z}_F|m_F)$  can be obtained by

$$\begin{aligned} q(\mathbf{Z}_F|m_F) &= \frac{\exp \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\boldsymbol{\theta}|m_F)}}{\sum_{\mathbf{Z}_F} \exp \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\boldsymbol{\theta}|m_F)}} \\ &= C \exp \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\boldsymbol{\theta}|m_F)}, \end{aligned} \quad (26)$$

where C represents a normalizing constant.

### 3.3.3 Derivation of $q(\theta_i|m_F)$

$q(\theta_i|m_F)$  can be also obtained through the maximization of  $\mathcal{F}(q)$  under the constraint of  $\sum_{\theta_i} q(\theta_i|m_F) = 1$ . From Equation 21, we can write the terms, which are dependent on  $q(\theta_i|m_F)$ ,

$$\mathcal{F}[q(\theta_i|m_F)] = \left\langle \log \frac{P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F)}{q(\mathbf{Z}_F|m_F)} \right\rangle_{q(\mathbf{Z}|m_F), q(\boldsymbol{\theta}|m_F)} + \sum_j \left\langle \log \frac{P(\theta_j|m_F)}{q(\theta_j|m_F)} \right\rangle_{q(\theta_j|m_F)}. \quad (27)$$

Let  $\lambda$  be a Lagrange multiplier. Then the problem becomes the extreme value problem of the following functional  $\mathcal{J}[q(\theta_i|m_F)]$ ,

$$\begin{aligned} \mathcal{J}[q(\theta_i|m_F)] &= \mathcal{F}[q(\theta_i|m_F)] + \lambda \left( \int q(\theta_i|m_F) d\boldsymbol{\theta}_i - 1 \right) \\ \frac{\partial \mathcal{J}[q(\theta_i|m_F)]}{\partial q(\theta_i|m_F)} &= \frac{\partial}{\partial q(\theta_i|m_F)} \left[ \sum_{\mathbf{Z}_F} \left\{ q(\mathbf{Z}_F|m_F) \int q(\boldsymbol{\theta}|m_F) \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \right. \right. \\ &\quad \left. \left. - q(\boldsymbol{\theta}|m_F) \log q(\mathbf{Z}|m_F) d\boldsymbol{\theta} \right\} \right. \\ &\quad \left. + q(\theta_i|m_F) \log P(\theta_i|m_F) - q(\theta_i|m_F) \log q(\theta_i|m_F) \right] + \lambda \\ &= \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\mathbf{Z}_F|m_F), q(\boldsymbol{\theta}_{-i}|m_F)} + \log P(\theta_i|m_F) \\ &\quad - \log q(\theta_i|m_F) - 1 = 0, \end{aligned} \quad (28)$$

$$\Rightarrow q(\theta_i|m_F) = P(\theta_i|m_F) \exp \left( \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\mathbf{Z}_F|m_F), q(\boldsymbol{\theta}_{-i}|m_F)} + \lambda - 1 \right), \quad (29)$$

$$\begin{aligned} \frac{\partial \mathcal{J}[q(\theta_i|m_F)]}{\partial \lambda} &= \int q(\theta_i|m_F) d\boldsymbol{\theta}_i - 1 \\ &= \int P(\theta_i|m_F) \exp \left( \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\mathbf{Z}_F|m_F), q(\boldsymbol{\theta}_{-i}|m_F)} \right. \\ &\quad \left. + \lambda - 1 \right) d\boldsymbol{\theta}_i - 1 = 0, \end{aligned} \quad (30)$$

$$\Rightarrow \exp(\lambda - 1) = \frac{1}{\int P(\theta_i|m_F) \exp \left( \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\mathbf{Z}_F|m_F), q(\boldsymbol{\theta}_{-i}|m_F)} \right) d\boldsymbol{\theta}_i}, \quad (31)$$

where  $\boldsymbol{\theta}_{-i}$  represents all of parameters  $\boldsymbol{\theta}$  except  $\theta_i$ . Substituting Equation 31 into 29, we obtain the following variational posterior  $q(\theta_i|m_F)$ :

$$\begin{aligned} q(\theta_i|m_F) &= \frac{P(\theta_i|m_F) \exp \left( \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\mathbf{Z}_F|m_F), q(\boldsymbol{\theta}_{-i}|m_F)} \right)}{\int P(\theta_i|m_F) \exp \left( \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\mathbf{Z}_F|m_F), q(\boldsymbol{\theta}_{-i}|m_F)} \right) d\boldsymbol{\theta}_i} \\ &= C_i P(\theta_i|m_F) \exp \left( \langle \log P(\mathbf{D}, \mathbf{Z}_F|\boldsymbol{\theta}, m_F) \rangle_{q(\mathbf{Z}_F|m_F), q(\boldsymbol{\theta}_{-i}|m_F)} \right), \end{aligned} \quad (32)$$

where  $C_i$  represents a normalizing constant.

### 3.3.4 Derivation of $q(m_F)$

The maximization of  $F[q]$  with respect to  $q(m_F)$  results in the optimum variational posterior of the model structure  $q(m_F)$ . Equation 21 can be rewritten as

$$\mathcal{F}[q] = \langle \mathcal{F}_{m_F} \rangle_{q(m_F)} + \left\langle \log \frac{P(m_F)}{q(m_F)} \right\rangle_{q(m_F)}, \quad (33)$$

where  $\mathcal{F}_{m_F}$  represents the sum of terms, which do not contain  $q(m_F)$ . Since  $q(\mathbf{Z}_F|m_F)$  and  $q(\boldsymbol{\theta}|m_F)$  affect only  $\mathcal{F}_{m_F}$ , the maximization of  $\mathcal{F}[q]$  with respect to  $q(\mathbf{Z}_F|m_F)$  and  $q(\boldsymbol{\theta}|m_F)$  is equivalent to the maximization of  $\mathcal{F}_{m_F}$  with respect to  $q(\mathbf{Z}_F|m_F)$  and  $q(\boldsymbol{\theta}|m_F)$ . Now let a maximum value of  $\mathcal{F}_{m_F}$  with respect to  $q(\mathbf{Z}_F|m_F)$  and  $q(\boldsymbol{\theta}|m_F)$  be  $\mathcal{F}_{m_F}^*$ . Then, the optimum variational posterior of the model structure  $q(m_F)$  can be written as

$$\mathcal{F}[q] = \langle \mathcal{F}_{m_F}^* \rangle_{q(m_F)} + \left\langle \log \frac{P(m_F)}{q(m_F)} \right\rangle_{q(m_F)}. \quad (34)$$

The above equation should be maximized with respect to  $q(m_F)$  under the constraint of  $\sum_{m_F} q(m_F) = 1$ . A Lagrange multiplier  $\lambda$  is again introduced and then, it becomes the extreme value problem of the functional  $\mathcal{J}[q(m_F)]$ ,

$$\begin{aligned} \mathcal{J}[q(m_F)] &= \langle \mathcal{F}_{m_F}^* \rangle_{q(m_F)} + \left\langle \log \frac{P(m_F)}{q(m_F)} \right\rangle_{q(m_F)} + \lambda \left( \sum_{m_F} q(m_F) - 1 \right) \\ &= \sum_{m_F} q(m_F) \left( \mathcal{F}_{m_F}^* + \log \frac{P(m_F)}{q(m_F)} \right) + \lambda \left( \sum_{m_F} q(m_F) - 1 \right), \\ \frac{\partial \mathcal{J}[q(m_F)]}{\partial q(m_F)} &= \mathcal{F}_{m_F}^* + \log P(m_F) - \log q(m_F) - 1 + \lambda = 0, \\ &\Rightarrow q(m_F) = P(m_F) \exp(\mathcal{F}_{m_F}^* - 1 + \lambda), \end{aligned} \quad (35)$$

$$\begin{aligned} \frac{\partial \mathcal{J}[q(m_F)]}{\partial \lambda} &= \sum_{m_F} q(m_F) - 1 = 0 \\ &\Rightarrow \exp(\lambda - 1) = \frac{1}{\sum_{m_F} P(m_F) \exp(\mathcal{F}_{m_F}^*)}, \end{aligned} \quad (36)$$

Substituting Equation 36 into 35, the following  $q(m_F)$  is obtained:

$$q(m_F) = \frac{P(m_F) \exp(\mathcal{F}_{m_F}^*)}{\sum_{m_F} P(m_F) \exp(\mathcal{F}_{m_F}^*)} = C_{m_F} P(m_F) \exp(\mathcal{F}_{m_F}^*), \quad (37)$$

where  $C_{m_F}$  is a normalizing constant ensuring  $\sum_{m_F} q(m_F) = 1$ . The maximization of  $q(m_F)$  is equivalent to the maximization of  $\mathcal{F}_{m_F}^*$ , since a uniform distribution  $P(m_F) = M_F$  is assumed as the prior distribution of  $m_F$ . Therefore an optimum model structure  $m_F$  can be estimated through the maximization of  $\mathcal{F}_{m_F}^*$  with respect to  $q(\mathbf{Z}|m_F)$  and  $q(\boldsymbol{\theta}|m_F)$ .

### 3.3.5 Variational posteriors for GMM

Finally, we can obtain the variational posterior of latent variables for given  $m_F$  as follows:

$$q(\mathbf{Z}_F|m_F) = C \prod_{i=1}^{m_F} \prod_{n=1}^N \exp \left\{ z_i^n \left( \langle \log \alpha_i \rangle_{q(\boldsymbol{\alpha}|m_F)} + \frac{1}{2} \langle \log |\mathbf{V}_i| \rangle_{q(\mathbf{V}_i|m_F)} - \frac{1}{2} \text{Tr} \left\{ \langle \mathbf{V}_i \rangle_{q(\mathbf{V}_i|m_F)} \langle (\mathbf{x}_n - \boldsymbol{\mu}_i)(\mathbf{x}_n - \boldsymbol{\mu}_i)^T \rangle_{q(\boldsymbol{\mu}_i|m_F)} \right\} \right) \right\}. \quad (38)$$

The variational posterior of  $\boldsymbol{\alpha}$  is the following Dirichlet distribution parameterized by  $\{\phi_0 + \bar{N}_i\}_{i=1}^{m_F}$ .

$$q(\boldsymbol{\alpha}|m_F) = \mathcal{D}(\{\alpha_i\}_{i=1}^{m_F} | \{\phi_0 + \bar{N}_i\}_{i=1}^{m_F}), \quad (39)$$

where  $\mathcal{D}()$  denotes Dirichlet distribution and

$$\bar{N}_i = \sum_{n=1}^N \bar{z}_i^n, \quad \bar{z}_i^n = \langle z_i^n \rangle_{q(z_i^n|m_F)}. \quad (40)$$

The variational posteriors of  $\mathbf{V}_i$  is the Wishart distribution, which is parameterized by  $\mathbf{B}_i$  and  $\eta_F + \bar{N}_i$  as follows:

$$q(\mathbf{V}_i|m_F) = \mathcal{W}(\mathbf{V}_i | \eta_F + \bar{N}_i, \mathbf{B}_i), \quad (41)$$

where

$$\mathbf{B}_i = \mathbf{B}_F + \bar{\mathbf{C}}_i + \frac{\bar{N}_i \bar{\zeta}_F}{\bar{N}_i + \bar{\zeta}_F} (\bar{\mathbf{x}}_i - \boldsymbol{\nu}_F)(\bar{\mathbf{x}}_i - \boldsymbol{\nu}_F)^T, \quad (42)$$

$$\bar{\mathbf{x}}_i = \frac{1}{\bar{N}_i} \sum_{n=1}^N \bar{z}_i^n \mathbf{x}_n, \quad \bar{\mathbf{C}}_i = \sum_{n=1}^N \bar{z}_i^n (\mathbf{x}_n - \bar{\mathbf{x}}_i)(\mathbf{x}_n - \bar{\mathbf{x}}_i)^T. \quad (43)$$

The marginalization of  $q(\boldsymbol{\mu}_i, \mathbf{V}_i|m_F)$  with respect to  $\mathbf{V}_i$  yields the variational posterior of  $\boldsymbol{\mu}_i$ . It turns out that the variational posterior becomes the Student's t-distribution parameterized by  $\bar{\boldsymbol{\mu}}_i$ ,  $\boldsymbol{\Sigma}_{\boldsymbol{\mu}_i}$ , and  $f_{\boldsymbol{\mu}_i}$ .

$$q(\boldsymbol{\mu}_i|m_F) = \mathcal{T}(\boldsymbol{\mu}_i | \bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\boldsymbol{\mu}_i}, f_{\boldsymbol{\mu}_i}), \quad (44)$$

where  $\mathcal{T}()$  represents the Student's t-distribution as follows:

$$\mathcal{T}(\boldsymbol{\mu}_i | \bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\boldsymbol{\mu}_i}, f_{\boldsymbol{\mu}_i}) \propto \left\{ 1 + (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T (\boldsymbol{\Sigma}_{\boldsymbol{\mu}_i} f_{\boldsymbol{\mu}_i})^{-1} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i) \right\}^{-\frac{d+f_{\boldsymbol{\mu}_i}}{2}}, \quad (45)$$

where  $d$  represents the dimension of the input vector. Moreover each parameter can be written as

$$\bar{\boldsymbol{\mu}}_i = \frac{\bar{N}_i \bar{\mathbf{x}}_i + \bar{\zeta}_F \boldsymbol{\nu}_F}{\bar{N}_i + \bar{\zeta}_F}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\mu}_i} = \frac{\mathbf{B}_i}{(\bar{N}_i + \bar{\zeta}_F) f_{\boldsymbol{\mu}_i}}, \quad f_{\boldsymbol{\mu}_i} = \eta_F + \bar{N}_i + 1 - d. \quad (46)$$

The optimization process starts from an initial guess and iterates the E-step (Equation 38) and M-step (Equations 39, 41, and 44) until it converges. This variational EM-algorithm gives a local maximum of  $\mathcal{F}[q]$ . The optimum structure  $m_F$ , which represents the number of functions,

object category	ID	A set	B set	total
scissors	T1	7	3	10
pen	T2	8	3	11
pliers	T3	2	2	4
tweezers	T4	3	2	5
utility knife	T5	3	1	4
stapler	T6	4	1	5
tape	T7	4	2	6
colored vinyl tape	T8	2	2	4

Table 1. Number of tools in the experiment.

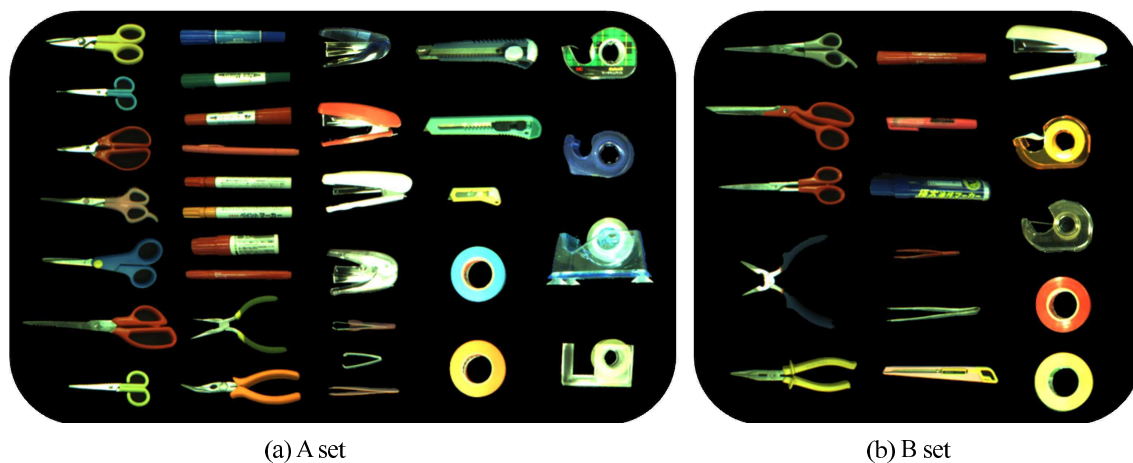


Fig. 6. Hand tools used in the experiment. (a)Set A. (b)Set B.

can be also obtained by selecting  $m_F$  that maximizes  $\mathcal{F}[q]$ . In the later experiment, the model structure is selected in the range of  $2 \leq m_F \leq 8$ .

When a novel observation  $X_F$  is given to the learned model, the function  $Z_F^*$  can be estimated as follows:

$$Z_F^* = \underset{Z_F}{\operatorname{argmax}} P(X_F|Z_F) = \underset{j}{\operatorname{argmax}} \hat{\alpha}_j \mathcal{N}(X_F|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{V}}_j), \quad (47)$$

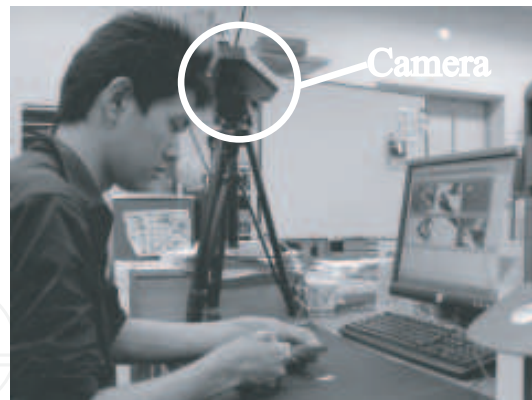
where  $\mathcal{N}$  represents the Gaussian distribution.  $\hat{\alpha}_j$  is the  $j$ -th component of a mode of the variational posterior  $q(\boldsymbol{\alpha}|m_F)$ .  $\hat{\boldsymbol{\mu}}_j$  and  $\hat{\mathbf{V}}_j$  denote modes of the variational posteriors  $q(\boldsymbol{\mu}_j|m_F)$  and  $q(\mathbf{V}_j|m_F)$ , respectively. For given  $X_F$ ,  $P(X_F|Z_F)$  in Fig.1(c) is derived as

$$P(X_F|Z_F) = \hat{\alpha}_{Z_F} \mathcal{N}(X_F|\hat{\boldsymbol{\mu}}_{Z_F}, \hat{\mathbf{V}}_{Z_F}). \quad (48)$$

## 4. Experiments

### 4.1 Experimental Setup

A total of 49 objects with 8 categories, which are given in Tab.1, are employed in the experiments. These 49 objects are divided into two groups. Figure 6(a) is the set A containing a total of 33 hand tools (7 scissors, 8 pens, 2 pliers, 4 staplers, 3 tweezers, 4 tapes, 2 colored vinyl



(a)



(b)

Fig. 7. (a)A snapshot of the system. (b)Structure vs. free energy.

tapes and 3 utility knives). The set B consists of a total of 16 hand tools (3 scissors, 3 pens, 2 pliers, 2 tweezers, 1 stapler, 2 tapes, 2 colored vinyl tapes and 1 utility knife), which are shown in Fig.6(b). Figure 7(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 the work object are extracted based on background difference method, and then the system computes appearance 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.4 was trained. Each of 33 hand tools (set A) used 10 times and a total of 330 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 7(b) shows free energy over the number of functions  $m_F$ . The figure implies that six functions explain the data best. In fact, we have confirmed these six functions correspond to 'cut', 'write',

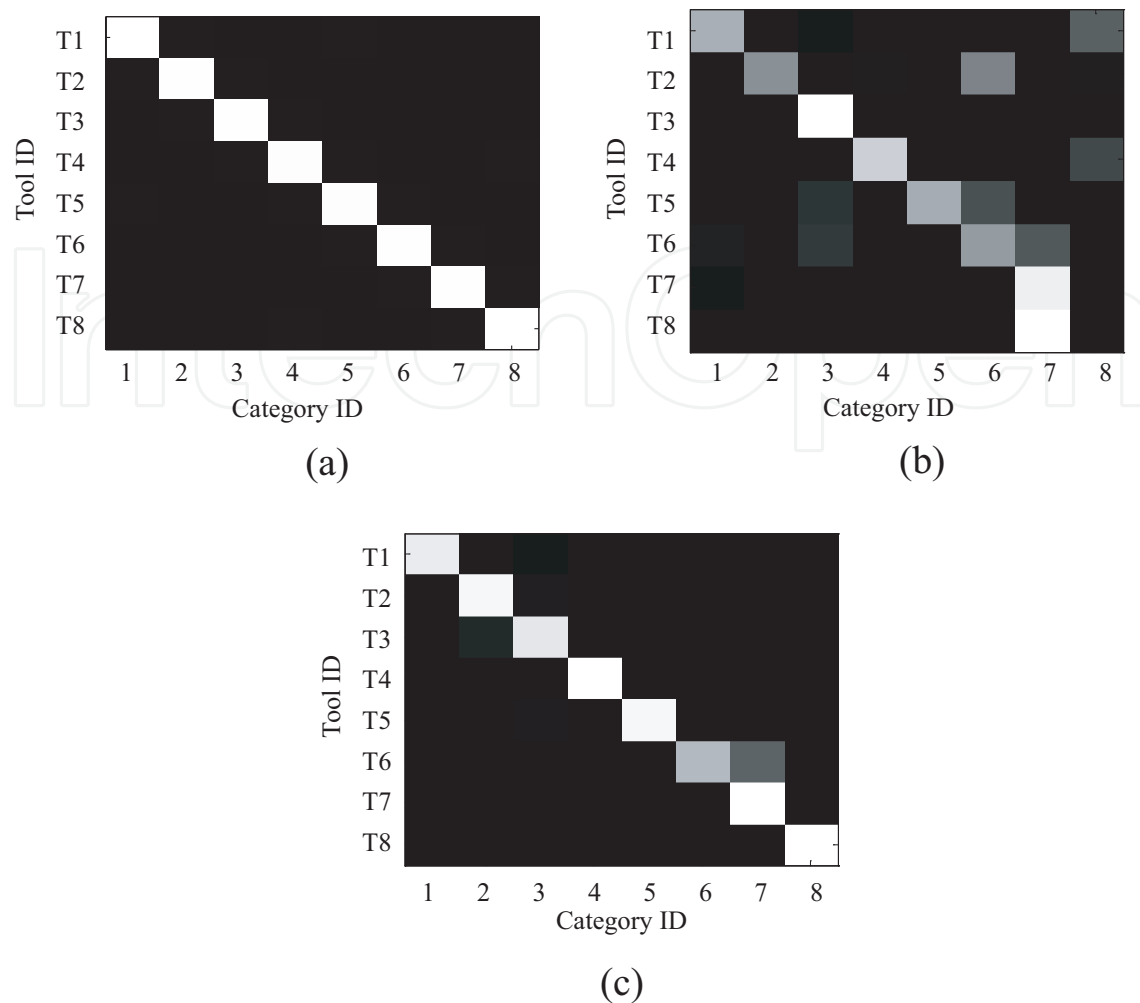


Fig. 8. (a)Correct categorization. (b)Categorization with only visual information. (c)Categorization with visual and function information.

'move'. 'deform', 'adhesion' and 'adhesion with color change'. In the following experiments, the abstract function model, which was obtained in this experiment, is used.

#### 4.3 Results of Learning

The tools in the A set are used in the second experiment for the training of Fig.1(c). Each of 33 hand tools was used 10 times; hence the model was trained using a total of 330 data. Then Equation 11 was used to classify 330 data. The classification result is compared with ground truth to evaluate how well the objects are categorized. The result is shown in Fig.8. In these figures the horizontal and vertical axes indicate category and object indices, respectively. The white bar in the figure represents that the object is classified into the category. From the figure one can see that the system has reasonably categorized the objects by using both appearance and function information.



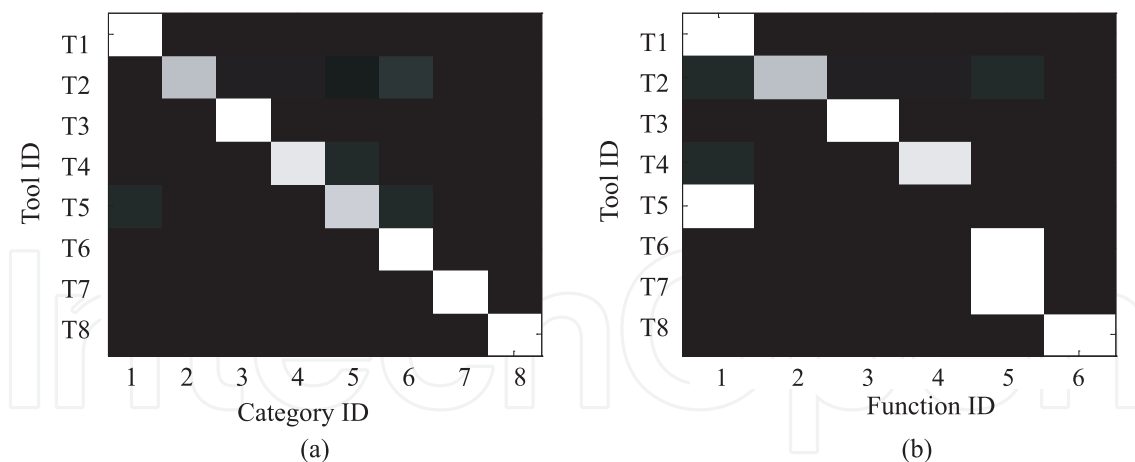


Fig. 9. Result of recognition. (a) Tool recognition from visual and function information. (b) Function recognition from only visual information.

#### 4.4 Results of Inference

After the training in the foregoing subsection, the system observed unseen objects in the B set and recognized their categories from the observable visual information and functions. The result is given in Fig.9 (a). Then the system inferred their functions only from appearance. Equation(12) was used to identify the function. The result is given in Fig.9 (b). It can be seen that the system inferred object functions almost perfectly. In fact, inference accuracy is 95.4%.

### 5. Conclusions

This chapter has discussed 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 on-line learning is left for the future works. Moreover the model should be extended so that it can represent the object usage and work objects.

### 6. Acknowledgements

This work was supported by Grant-in-Aid for Scientific Research (C) (20500186, 20500179) and Grant-in-Aid for Scientific Research on Innovative Areas (The study on the neural dynamics for understanding communication in terms of complex hetero systems).

### 7. References

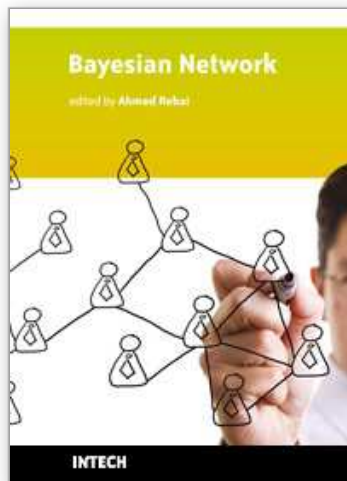
- Attias, H. (1999). Inferring parameters and structure of latent variable models by variational bayes, *Proceedings of 15th Conference on Uncertainty in Artificial Intelligence*, pp.21–30.
- Fergus, R., Perona, P., & Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol.2, pp.264–271.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*, Lawrence Erlbaum, Hillsdale, NJ.

- Hofmann, T. (2001). Unsupervised Learning by Probabilistic Latent Semantic Analysis, *Machine Learning*, vol.42, pp.177–196.
- Kojima A., Higuchi, M., Kitahashi, T., & Fukunaga, K. (2004). Toward a cooperative recognition of human behaviors and related objects, *Proceedings of Second International Workshop on Man-Machine Symbiotic Systems*, pp.195–206.
- Landau, B., Smith, L., & Jones, S. (1998). Object shape, object function, and object name, *Journal of Memory and Language*, 38(ML972533):1–27.
- Lowe, D. G. (2004) Distinctive image features from scale-invariant keypoints, *Int. Journal of Computer Vision*, 60(2):91–110.
- Ogura, T., Okada, K., & Inaba, M. (2005). Humanoid tool operating motion generation, *Proceedings of The 23rd Annual Conf. of the Robotics Society of Japan*, 1F15, (in japanese).
- Rivlin, E., Dickinson, S. J., & Rosenfeld, A. (1995). Recognition by functional parts, *Computer Vision and Image Understanding: CVIU*, 62(2), pp.164–176.
- Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A. & Freeman, W.T. (2005). Discovering object categories in image collections, *AI Memo*, 2005-005, pp.1–12.
- Stark, L., Bowyer, K., Hoover, A., & Goldgof, D. B. (1996). Recognizing object function through reasoning about partial shape descriptions and dynamic physical properties, *Proceedings of The IEEE*, 84(11):1640–1656.
- Woods, K., Cook, D., Hall, L., Bowyer, K. W., & Stark, L. (1995) Learning membership functions in a function-based object recognition system, *Journal of Artificial Intelligence Research*, 3:187–222.

IntechOpen

IntechOpen

IntechOpen



## **Bayesian Network**

Edited by Ahmed Rebai

ISBN 978-953-307-124-4

Hard cover, 432 pages

**Publisher** Sciyo

**Published online** 18, August, 2010

**Published in print edition** August, 2010

Bayesian networks are a very general and powerful tool that can be used for a large number of problems involving uncertainty: reasoning, learning, planning and perception. They provide a language that supports efficient algorithms for the automatic construction of expert systems in several different contexts. The range of applications of Bayesian networks currently extends over almost all fields including engineering, biology and medicine, information and communication technologies and finance. This book is a collection of original contributions to the methodology and applications of Bayesian networks. It contains recent developments in the field and illustrates, on a sample of applications, the power of Bayesian networks in dealing the modeling of complex systems. Readers that are not familiar with this tool, but have some technical background, will find in this book all necessary theoretical and practical information on how to use and implement Bayesian networks in their own work. There is no doubt that this book constitutes a valuable resource for engineers, researchers, students and all those who are interested in discovering and experiencing the potential of this major tool of the century.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Tomoaki Nakamura and Takayuki Nagai (2010). Forming Object Concept Using Bayesian Network, Bayesian Network, Ahmed Rebai (Ed.), ISBN: 978-953-307-124-4, InTech, Available from:

<http://www.intechopen.com/books/bayesian-network/forming-object-concept-using-bayesian-network>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen