Electronic Notes in Theoretical Computer Science 46 (2001)
URL: http://www.elsevier.nl/locate/entcs/volume46.html 20 pages

Spatial Pattern Discovering by Learning the Isomorphic Subgraph from Multiple Attributed Relational Graphs.

Pengyu Hong [1] and Thomas S. Huang [2]

Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana Champaign
Urbana, IL61801, USA

Abstract

Inexact graph matching has been widely investigated to relate a set of object/scene primitives extracted from an image to a set of counterparts representing a model or reference. However, little has been done to address how to build such a model or reference. This paper develops the theory for automatic contextual pattern modelling to automatically learn a parametric pattern ARG model from multiple sample ARGs. The learned pattern ARG characterizes the sample ARGs, which represent a pattern observed under different conditions. The maximum-likelihood parameters of the pattern ARG model are estimated via the Expectation-Maximization algorithm. Particularly, for Gaussian attributed and relational density distribution assumptions, analytical expressions are derived to estimate the density parameters of the pattern ARG model. The pattern ARG model with Gaussian distribution assumptions is therefore called the Contextual Gaussian Mixture model. The theory and methodology is applied to the problems of unsupervised spatial pattern extraction from multiple images. The extracted spatial pattern can be used for data summarization, graph matching, and pattern detection. One immediate application of this newly developed theory will be information summarization and retrieval in digital image and video libraries.

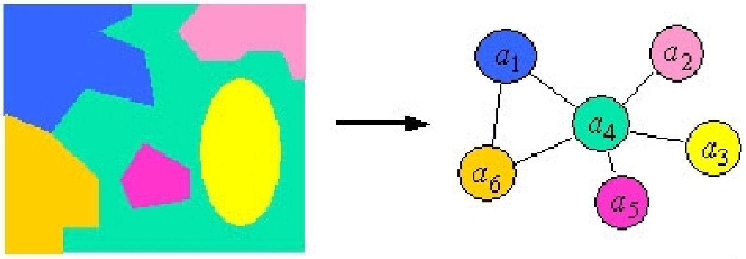
1 Introduction

Contextual pattern modelling has been an important task in computer vision and pattern recognition [2,3,5,10,13]. It is fundamental to image registration, recognition, and classification. In contextual pattern modelling research, a model object/scene is usually represented as an attributed relational graph

[1] Email: hong@ifp.uiuc.edu

[2] Email: huang@ifp.uiuc.edu

(ARG) [10] that consists of a set of nodes and arcs. An example of the ARG is illustrated at Fig. 1. The nodes of an ARG represent the object/scene primitives in the images. The attributes of the nodes encode the appearance properties of the object/scene primitives. The relations among the nodes specify the contextual information of nodes. Example of such relations are the relative distance between two primitives, the relative orientations of one primitive to the others, and so on. The relations of a node uniquely specifies that node given the identity of other nodes. In the rest of the paper, we assume that the observed images are processed and represented as sample ARGs.



(a) (b)

Fig. 1. An ARG that represents a group of image segments. (a) A group of image segments with different colors, (b) the ARG representation of (a). The color of a node represents the color attribute of its corresponding image segment. The arcs represent the adjacent relations among the image segments.

Recently, ARG and graph matching techniques have begun to attract great attentions in content-based image/video retrieval community [12,15,16]. Basically, the user submits a set of sample images (usually more than two) to the system. The system first summarizes the sample images in some ways. Then, the system uses the summarized information to search through its database and return a set of images, which are similar to the sample images based on its similarity measurement method.

Most image/video classification and retrieval approaches use the global features (texture, color, etc.) of the images [18]. The global features is a mixture of the features of the image primitives. This one of the main reasons that lead to ambiguities in image classification and retrieval applications. The introduction of ARG representation enables the system to examine images at a finer and more meaningful level. Consequentially, techniques and algorithms for summarizing multiple sample ARGs are required. Although two-graph matching as a fundamental problem has been widely investigated [1,4,6,17,19,20,21], little has been done for summarizing multiple sample ARGs.

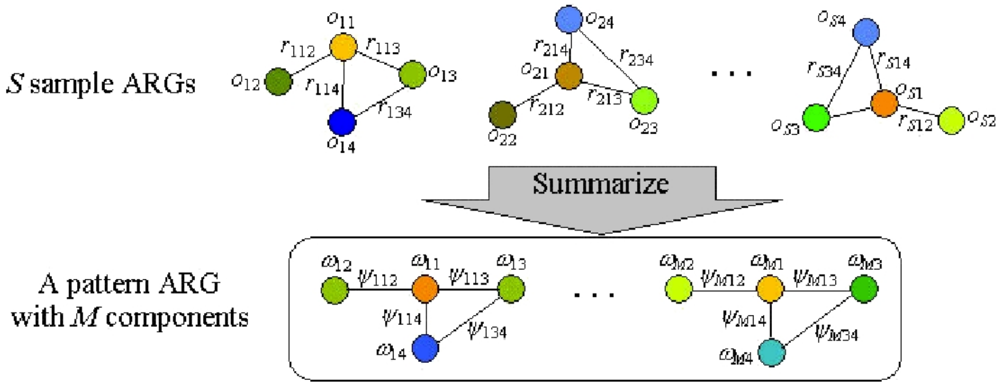
Some merely use ARG to represent samples and apply two-graph matching techniques to measuring the similarity between the samples in the database and the query for retrieval purpose. Huet and Hancock [12] used ARG to represent the geometric attributes and structural information of line-patterns. Ozer [15] used relational graph to annotate the images where the object of interest is present. However, both [12] and [15] only use ARG for information representation and apply two-graph matching algorithms for information retrieval.

Recently, some approaches try to learn pattern from multiple instances. Ratan et al [16] used Diverse Density algorithm to learn “visual concepts”, or pattern in the context of this paper, from multiple sample images. The learned “visual concepts” can be used to classify new images. In [16], a concept is a pre-specified conjunction of several image primitives (e.g. image segments). The representation of the concept is similar to ARG. Nonetheless, the relational information of the image primitives, which is essential for distinguishing image primitives, is not utilized in [16].

Frey and Jojic [9] defined transformation function as a discrete latent variable in the probabilistic graphical model and use the Expectation-Maximization (EM) algorithm [7] to learn patterns from images with clutter backgrounds. In their approach, transformation function is defined on the image pixel level. The values of the transformation functions are selected from a predefined discrete transformation set. The image pixels are just like the nodes of an ARG. The transformation set is similar to the value range of the matching function in graph matching algorithm. However, their approach does not consider the contextual information of image pixels. Therefore, the matching ambiguity of image pixels is left as an open problem. In addition, operating at pixel level limits the possible transformation set (many transformations are defined on pixel group and are continuous), increases pixel matching ambiguity, and brings about high computational complexity.

Multiple sample ARGs should provide more information about the pattern than two sample ARGs do. It is not appropriate to use one of the sample ARGs as the model or reference. The relations and the attributes of a sample ARG may not best represent those of other sample ARGs. This is due to the variance of the relations and attributes caused by noise, lighting conditions, transformations, and so on. We propose to learn a parametric pattern ARG from a set of sample ARGs. The learned pattern ARG model can be further used for object detection.

The rest of the paper is organized as below. Section 2 formulates the automatic contextual pattern modelling problem. Section 3 derives expressions for estimating the parameters of the pattern ARG via the EM algorithm. Analytic expressions can be obtained for some parameters regardless the attributed and relational density distribution functions. Section 4 derives analytical expressions for estimating the density parameters of a special case of the pattern ARG, called the Contextual Gaussian Mixture model. Section 5 discusses how to used the learned pattern ARG model for pattern detection. Implementation issues are presented in Section 6. Experimental results are provided in Section 7. Finally, the paper closes with summary and discussions in Section 8.



Modeling correspondences	Pattern ARG		Sample ARGs
Node correspondences:	$\omega_{11}, \omega_{21}, \ldots, \omega_{M1}$	–	$o_{11}, o_{21}, \ldots, o_{S1}$
	$\omega_{12}, \omega_{22}, \ldots, \omega_{M2}$	–	$o_{12}, o_{22}, \ldots, o_{S2}$
	$\omega_{13}, \omega_{23}, \ldots, \omega_{M3}$	–	$o_{13}, o_{23}, \ldots, o_{S3}$
	$\omega_{14}, \omega_{24}, \ldots, \omega_{M4}$	–	$o_{14}, o_{24}, \ldots, o_{S4}$
Relation correspondences:	$\psi_{112}, \psi_{212}, \ldots, \psi_{M12}$	–	$r_{112}, r_{212}, \ldots, r_{S12}$
	$\psi_{113}, \psi_{213}, \ldots, \psi_{M13}$	–	$r_{113}, r_{213}, \ldots, r_{S13}$
	$\psi_{114}, \psi_{214}, \ldots, \psi_{M14}$	–	$r_{114}, r_{214}, \ldots, r_{S14}$
	$\psi_{134}, \psi_{134}, \ldots, \psi_{M34}$	–	$r_{134}, r_{234}, \ldots, r_{S34}$

Fig. 2. Modelling S sample ARGs $\{G_i\}$ by a pattern ARG with M model components ($M << S$). o_{ij} ($1 \leq i \leq S, 1 \leq j \leq 4$) represents a sample node. r_{mk} ($1 \leq m \leq S, 1 \leq k \leq 4$) represents a sample relation. ω_{pq} ($1 \leq p \leq M, 1 \leq q \leq 4$) represents a pattern node. ψ_{abc} ($1 \leq a \leq M, 1 \leq b \leq 4, 1 \leq c \leq 4$) represents a pattern relation.

2 Automatic Contextual Pattern Modelling

In [11], we used a pattern ARG model, which contains only one parametric ARG model, to model a set of sample ARGs. Here, we generalize the pattern ARG model so that it consists of a small number of parametric ARG models. The expanded pattern ARG model has larger modelling capacity and can be used to effectively model a larger set of sample ARGs that are observed under more diverse conditions. Fig. 2 illustrates the idea of modelling a large set of sample ARGs with a pattern ARG. The pattern ARG model is a compact representation of the sample ARGs. It explains each sample ARG on two scales. On the macro scale, a sample ARG is a linear combination of the model components. On the micro scale, if a model component is specified, a node of the sample ARG is a linear combination of the nodes of the model component in terms of node matching probabilities.

The contextual pattern modelling problem becomes straightforward if the node and relation correspondences between the sample ARGs and the components of the pattern ARG are specified. However, it is tedious and labor intensive to manually specify the node and relation correspondences for a large set of sample ARGs. Moreover, the observed images always contain the pattern and its backgrounds. It is also tedious and labor intensive to manually label the pattern out of its backgrounds.

This paper is interested in automatic contextual pattern modelling that does not require to manually specify the correspondences and manually extract the pattern from the observed images. The automatic learning procedure should calculate: (a) the attributed parameters (appearance information) of the pattern ARG, (b) the relational parameters (contextual information) of the pattern ARG, (c) the structure (the number of nodes and that of the relations) of the pattern ARG, and (d) the node and relation correspondences between the components of the pattern ARG and the sample ARGs.

We first define the notations that will be used in the rest of the paper.

(a) The observed sample ARGs is represented as $G = \{G_1, ..., G_S\}$, where S is the number of the sample ARGs. Each sample ARG $G_i = \langle O_i, A_i, R_i, B_i \rangle$ $(1 \leq i \leq S)$ has: (1) U_i data nodes [3] $O_i = \{o_{ik}\}_{k=1}^{U_i}$; (2) the attribute set of the data nodes $A_i = \{\overrightarrow{a}_{ik}\}_{k=1}^{U_i}$ and $\overrightarrow{a}_{ik}$ is the attribute vector of data node o_{ik}; (3) $U_i \times U_i$ data relations $R_i = \{r_{icd}\}$ $(1 \leq c, d \leq U_i)$; and (4) the feature vector set of the data relation set $B_i = \{\overrightarrow{b}_{icd}\}_{c,d=1}^{U_i}$ and $\overrightarrow{a}_{icd}$ is the feature vector of the data relation r_{icd}. R_i and B_i define the contextual information of the nodes in G_i. Self-relation, which represents the relation between a node and itself, is allowed. For example, the distance relation between a node and itself is 0. We have $r_{icd} = r_{idc}$ and $\overrightarrow{b}_{icd} = \overrightarrow{b}_{idc}$ if the relations are unidirectional.

(b) The pattern ARG model Γ has M components and $\Gamma = \{\Phi_w\}_{w=1}^{M}$. Each component $\Phi_w = \langle \Omega_w, \Psi_w, \Theta_w \rangle$ consists of: (1) N model nodes $\Omega_w = \{\omega_{wk}\}_{k=1}^{N}$; (2) $N \times N$ model relations $\Psi_w = \{\psi_{wcd}\}$ $(1 \leq c, d \leq N)$; and (3) the parameter set Θ_w. We will discuss the details of the parameter set Θ_w when we derive the expressions for estimating Θ_w. Let Θ denote $\{\Theta_w\}$.

(c) The correspondences between the sample ARGs and the pattern ARG is denoted by $Y = \{\overrightarrow{Y}_i\}_{i=1}^{S}$. Y specify the ways that the pattern ARG model Γ generates G. Y is a random variable governed by the distribution $f(y|G, \Gamma) = f(y|G, \Theta)$. Each element of Y, say $\overrightarrow{Y}_i$, is also a random variable. And let $\overrightarrow{y}_i = [q_i, y_{i1}, ..., y_{iU_i}]$ denote an instance of $\overrightarrow{Y}_i$. The value of q_i denotes that G_i matches with the model component Φ_{q_i} or G_i is generated by Φ_{q_i}. The value of y_{ij} denotes that the data node o_{ij} matches with the model node $\omega_{q_i y_{ij}}$ or o_{ij} is generated by $\omega_{q_i y_{ij}}$. We also have $1 \leq q_i \leq M$ and $1 \leq y_{ij} \leq N$. Once the value of Y is decided, the correspondences between the sample relations and the model relations are fixed. Let $\Re_{y_i}(r_{icd}) \in \Psi_{q_i}$ denote the corresponding model relation of the data relation r_{icd} respect to $\overrightarrow{y}_i$.

[3] We allow $U_i \neq U_j$ if $i \neq j$ because the observed images may have different numbers of image primitives. This is due to noise or the fact that the pattern being placed in different backgrounds. For example, if a node represents a line, the lines tend to get broken during the process of line detection, which results in extraneous nodes.

Overall, Θ and Y are unknowns to be estimated.

3 Finding the Maximum Likelihood Parameters of the Pattern ARG via the EM Algorithm

An algorithm that simultaneously considers all the sample ARGs is needed to estimate the parameters of the pattern ARG and calculate the correspondences between the sample ARGs and the pattern ARG. In this paper, the EM algorithm [7] is used.

3.1 The Basic EM Algorithm

The EM algorithm is a technique for finding the maximum-likelihood estimate of the parameters of underlying distributions from a training data set, which is incomplete or has missing values. The EM algorithm works iteratively in two steps: Expectation and Maximization. The algorithm defines the function:

$$Q(\Theta; \Theta^{(t)}) = E[\log p(D_o, D_m; \Theta) | D_o, \Theta^{(t)}] \tag{1}$$

where Θ is the parameter to be estimated, D_o is the observed data, D_m is the missing information, and t is the number of the iteration of the EM algorithm. $Q(\Theta; \Theta^{(t)})$ is a function of Θ under the assumption that $\Theta = \Theta^{(t)}$. The right hand side of eq. (1) denotes that the expected value of the complete data log-likelihood $\log p(D_o, D_m; \Theta)$ with respect to D_m and D_o and assuming $\Theta = \Theta^{(t)}$. In the Expectation step, $Q(\Theta; \Theta^{(t)})$, is computed. In the Maximization step, the algorithm updates Θ by $\Theta^{(t+1)} = \arg\max_{\Theta} Q(\Theta; \Theta^{(t)})$.

3.2 Derive Expressions for Estimating the Parameters of the pattern ARG via the EM algorithm

In the context of learning the pattern ARG model, the observed data D_o is $G = \{G_i\}_{i=1}^{S}$ and the missing data D_m is $Y = \{\overrightarrow{Y}_i\}_{i=1}^{S}$. We can rewrite eq. (1) as:

$$\begin{aligned} Q(\Theta; \Theta^{(t)}) &= E[\log p(G, Y; \Theta) | G, \Theta^{(t)}] \\ &= \Sigma_y f(y|G, \Theta^{(t)}) f(G|\Theta^{(t)}) \log p(G, y|\Theta) \end{aligned} \tag{2}$$

where $f(y|G, \Theta^{(t)})$ is the marginal distribution of the unobserved data Y and is dependent on the observed data G and the current values of the parameter set Θ. Since $f(G|\Theta^{(t)})$ is not dependent on Θ and will not effect the final results, we can take it out. Without losing the generality, we can assume that G_i is independent to each other, and consequently $\overrightarrow{y}_i$ is independent to each other. Therefore, eq. (2) can be rewritten as:

$$Q(\Theta;\Theta^{(t)}) = \sum_{\overrightarrow{y}_1} \cdots \sum_{\overrightarrow{y}_S} \sum_{i=1}^{S} \log(p(G_i|\overrightarrow{y}_i,\Theta)P(\overrightarrow{y}_i)) \prod_{j=1}^{S} f(\overrightarrow{y}_j|G_j,\Theta^{(t)}) \quad (3)$$

where $p(G_i|\overrightarrow{y}_i,\Theta)$ is the density function of G_i given the pattern ARG model and the match $\overrightarrow{y}_i$, and

$$\begin{aligned} p(G_i|\overrightarrow{y}_i,\Theta) &= p(G_i|[y_{i1}\cdots y_{iU_i}],\Theta_{q_i}) \\ &= \prod_{m=1}^{U_i} p(o_{im}|\omega_{q_i y_{im}}) \prod_{c=1}^{U_i}\prod_{d=1}^{U_i} p(r_{icd}|\Re_{\overrightarrow{y}_i}(r_{icd})) \end{aligned} \quad (4)$$

where $p(o_{im}|\omega_{q_i y_{im}})$ is the attributed distribution function and $p(r_{icd}|\Re_{\overrightarrow{y}_i}(r_{icd}))$ is the relational distribution function. If the relation is unidirectional, eq. (4) should be written as:

$$p(G_i|\overrightarrow{y}_i,\Theta) = \prod_{m=1}^{U_i} p(o_{im}|\omega_{q_i y_{im}}) \big(\prod_{c=1}^{U_i}\prod_{d=1}^{U_i} p(r_{icd}|\Re_{\overrightarrow{y}_i}(r_{icd}))\big)^{1/2} \quad (5)$$

In the following derivation, we use eq. (4). It can be easily shown the eq. (5) will only affect part of the final results by a scale of 1/2.

We can also write down the term $P(\overrightarrow{y}_i)$ in eq. (3) as:

$$P(\overrightarrow{y}_i) = P(q_i) \prod_{n=1}^{U_i} P(y_{in}|q_i) \quad (6)$$

Let $P(q_i = h) = \alpha_h$ $(1 \le h \le M)$, such that $\Sigma_{h=1}^{M}\alpha_h = 1$. Let $P(y_{in} = \eta|q_i = h) = \beta_{h\eta}$ $(1 \le \eta \le N)$, such that $\Sigma_{\eta=1}^{N}\beta_{h\eta} = 1$. The underlying intuitions of eq. (6) are: (1) On the macro scale, a sample ARG is a linear combination of the model components weighted by α_h; (2) On the micro scale, given the fact that G_i match with the model component Φ_h, a data node is a linear combination of the model nodes in Φ_h weighted by $\beta_{h\eta}$. $\{\alpha_h\} \cup \{\beta_{h\eta}\}$ is part of the parameter set to be estimated, or $\{\alpha_h\} \cup \{\beta_{h\eta}\} \subset \Theta$.

The term $f(\overrightarrow{y}_j|G_j,\Theta^{(t)})$ in eq.(3) is the marginal distribution of $\overrightarrow{y}_j$. Since the contextual information is fully described in G_j, y_{jk} is independent to each other. Hence,

$$f(\overrightarrow{y}_j|G_j,\Theta^{(t)}) = P(q_j|G_j,\Theta^{(t)}) \prod_{k=1}^{U_j} f(y_{jk}|G_j,\Theta_{q_j}^{(t)}) \quad (7)$$

Submitting eq.(4),(6),(7) into eq. (3), we have

$$\begin{aligned}
Q(\Theta;\Theta^{(t)}) &= \sum_{\overrightarrow{y}_i}\cdots\sum_{\overrightarrow{y}_S}\sum_{i=1}^{S}\log(p(G_i|\overrightarrow{y}_i,\Theta)P(\overrightarrow{y}_i))\prod_{j=1}^{S}f(\overrightarrow{y}_j|G_j,\Theta^{(t)})\\
&= \sum_{q_1=1}^{M}\sum_{y_{11}=1}^{N}\cdots\sum_{y_{1U_1}=1}^{N}\cdots\sum_{q_S=1}^{M}\sum_{y_{S1}=1}^{N}\cdots\sum_{y_{SU_S}=1}^{N}\sum_{i=1}^{S}log\Bigg(\prod_{m=1}^{U_i}p(o_{im}|\omega_{q_iy_{im}})\\
&\quad\prod_{c=1}^{U_i}\prod_{d=1}^{U_i}p(r_{icd}|\Re_{\overrightarrow{y}_i}(r_{icd}))P(q_i)\prod_{n=1}^{U_i}P(y_{in}|q_i)\Bigg)\prod_{j=1}^{S}\Bigg(P(q_j|G_j,\Theta^{(t)})\\
&\quad\prod_{k=1}^{U_j}f(y_{jk}|G_j,\theta_{q_j}^{(t)}\Bigg)
\end{aligned}\tag{8}$$

Replacing $log(\prod g(x))$ with $\sum(log(g(x)))$ in eq.(8), we have:

$$\begin{aligned}
Q(\Theta;\Theta^{(t)}) &= \sum_{q_1=1}^{M}\sum_{y_{11}=1}^{N}\cdots\sum_{y_{1U_1}=1}^{N}\cdots\sum_{q_S=1}^{M}\sum_{y_{S1}=1}^{N}\cdots\sum_{y_{SU_S}=1}^{N}\sum_{i=1}^{S}\Bigg[\log P(q_i)+\\
&\quad\sum_{m=1}^{U_i}\log\Bigg(p(o_{im}|\omega_{q_iy_{im}})P(y_{im}|q_i)\Bigg)+\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\log p(r_{icd}|\Re_{\overrightarrow{y}_i}(r_{icd}))\Bigg]\\
&\quad\prod_{j=1}^{S}\Bigg(P(q_j|G_j,\Theta^{(t)})\prod_{k=1}^{U_j}f(y_{jk}|G_j,\theta_{q_j}^{(t)})\Bigg)
\end{aligned}\tag{9}$$

Eq. (9) can be greatly simplified into

$$\begin{aligned}
Q(\Theta;\Theta^{(t)}) &= \sum_{i=1}^{S}\sum_{h=1}^{M}\Bigg[\log(\alpha_h)P_{q_i}(h|G_i,\Theta^{(t)})+\\
&\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\log(\beta_{h\eta})f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})+\\
&\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\log(p(o_{im}|\omega_{h\eta}))f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})+\\
&\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\sum_{\sigma=1}^{N}\sum_{\tau=1}^{N}\log(p(r_{icd}|\psi_{i\sigma\tau}))f_{y_{ic}}(\sigma|G_i,\Theta_h^{(t)})f_{y_{id}}(\tau|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})\Bigg]
\end{aligned}\tag{10}$$

where $P_{q_i}(h|G_i,\Theta^{(t)})$ denotes $P(q_i = h|G_i,\Theta^{(t)})$ and $f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})$ denotes $f(y_{im} = \eta|G_i,\Theta_h^{(t)})$. Note that $f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})$ can be calculated using inexact graph matching techniques. Readers are asked to refer to Appendix A for the details about simplifying eq. (9).

It is now clear that $\Theta = \{\alpha_h\} \cup \{\beta_{h\eta}\} \cup$ {the parameters of the attributed distribution } $\cup$ { the parameters of the relational distribution }. In the Maximization step, Θ is updated by $\Theta^{(t+1)} = \arg\max_{\Theta} Q(\Theta; \Theta^{(t)})$. Both the parameters of the attributed distribution and those of the relational distribution depend on the forms of the distribution functions, and so are their update expressions. The expressions for updating α_h and $\beta_{h\eta}$ can however be obtained as below regardless the forms of the attributed and relational distributions:

$$\alpha_h^{(t+1)} = \frac{\sum_{i=1}^{S} P_{q_i}(h|G_i, \Theta^{(t)})}{S} \tag{11}$$

$$\beta_{h\eta}^{(t+1)} = \frac{M \sum_{i=1}^{S} \sum_{m=1}^{U_i} f_{y_{im}}(\eta|G_i, \Theta_h^{(t)}) P_{q_i}(h|G_i, \Theta^{(t)})}{\sum_{i=1}^{S} U_i} \tag{12}$$

where $f_{y_{im}}(\eta|G_i, \Theta_h^{(t)})$ and $P_{q_i}(h|G_i, \Theta^{(t)})$ can be calculated using graph matching techniques, which will be discussed in Section 6.1. Readers are asked to refer to Appendix B for the details of deriving eq. (11) and (12).

4 Contextual Gaussian Mixture Model

In most applications, the attributed distribution and the relational distribution are likely to be assumed to be Gaussian. This kind of pattern ARG model is called the Contextual Gaussian Mixture (CGM) model. Analytical expressions for estimating the distribution parameters of the CGM in the Maximization step of the EM algorithm can be derived.

Assume the attributed distribution is

$$p(o_{im}|\omega_{h\eta}) = \frac{exp(-\frac{1}{2}(\overrightarrow{a}_{im} - \overrightarrow{\mu}_{h\eta})^T \Sigma_{h\eta}^{-1} (\overrightarrow{a}_{im} - \overrightarrow{\mu}_{h\eta}))}{(2\pi)^{\xi/2} |\Sigma_{h\eta}|^{1/2}} \tag{13}$$

where $\overrightarrow{\mu}_{h\eta}$ and $\Sigma_{h\eta}$ are the mean and covariance matrix of the attribute of the model node $\omega_{h\eta}$, and ξ is the dimension of the attribute vector. We can obtain the expressions for updating $\overrightarrow{\mu}_{h\eta}$ and $\Sigma_{h\eta}$ as below:

$$\overrightarrow{\mu}_{h\eta}^{(t+1)} = \frac{\sum_{i=1}^{S} \sum_{m=1}^{U_i} \overrightarrow{a}_{im} f_{y_{im}}(\eta|G_i, \Theta_h^{(t)}) P_{q_i}(h|G_i, \Theta^{(t)})}{\sum_{i=1}^{S} \sum_{m=1}^{U_i} f_{y_{im}}(\eta|G_i, \Theta_h^{(t)}) P_{q_i}(h|G_i, \Theta^{(t)})} \tag{14}$$

$$\overrightarrow{\Sigma}_{h\eta}^{(t+1)} = \frac{\sum_{i=1}^{S} \sum_{m=1}^{U_i} \overrightarrow{x}_{im}^{(t)} {\overrightarrow{x}_{im}^{(t)}}^T f_{y_{im}}(\eta|G_i, \Theta_h^{(t)}) P_{q_i}(h|G_i, \Theta^{(t)})}{\sum_{i=1}^{S} \sum_{m=1}^{U_i} f_{y_{im}}(\eta|G_i, \Theta_h^{(t)}) P_{q_i}(h|G_i, \Theta^{(t)})} \tag{15}$$

where $\overrightarrow{x}_{im}^{(t)} = \overrightarrow{a}_{im} - \overrightarrow{\mu}_{h\eta}^{(t)}$. Readers are asked to refer to Appendix C for the details of deriving eq. (14) and (15).

Assume the relational distribution is

$$p(r_{icd}|\psi_{h\sigma\tau}) = \frac{exp(-\frac{1}{2}(\overrightarrow{b}_{icd} - \overrightarrow{\gamma}_{h\sigma\tau})^T \Lambda^{-1}_{h\sigma\tau}(\overrightarrow{b}_{icd} - \overrightarrow{\gamma}_{h\sigma\tau}))}{(2\pi)^{\kappa/2}|\Lambda_{h\sigma\tau}|^{1/2}} \tag{16}$$

where $\overrightarrow{\gamma}_{h\sigma\tau}$ and $\Lambda_{h\sigma\tau}$ are the mean and covariance matrix of the feature vector of the model relation $\psi_{h\sigma\tau}$, and κ is the dimension of the relational feature vector. We can obtain the expressions for updating $\overrightarrow{\gamma}_{h\sigma\tau}$ and $\Lambda_{h\sigma\tau}$ as below:

$$\overrightarrow{\gamma}^{(t+1)}_{h\sigma\tau} = \frac{\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\overrightarrow{b}_{icd}\vartheta_h(y_{ic}, y_{id}, \sigma, \tau)P_{q_i}(h|G_i, \Theta^{(t)})}{\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\vartheta_h(y_{ic}, y_{id}, \sigma, \tau)P_{q_i}(h|G_i, \Theta^{(t)})} \tag{17}$$

$$\overrightarrow{\Lambda}^{(t+1)}_{h\sigma\tau} = \frac{\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\overrightarrow{z}^{(t)}_{icd}{\overrightarrow{z}^{(t)}_{icd}}^{T}\vartheta_h(y_{ic}, y_{id}, \sigma, \tau)P_{q_i}(h|G_i, \Theta^{(t)})}{\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\vartheta_h(y_{ic}, y_{id}, \sigma, \tau)P_{q_i}(h|G_i, \Theta^{(t)})} \tag{18}$$

where $\vartheta_h(y_{ic}, y_{id}, \sigma, \tau) = f_{y_{ic}}(\sigma|G_i, \Theta^{(t)}_h)f_{y_{id}}(\tau|G_i, \Theta^{(t)}_h)$ and $\overrightarrow{z}^{(t)}_{icd} = \overrightarrow{b}_{icd} - \overrightarrow{\gamma}^{(t)}_{h\sigma\tau}$. Readers are asked to refer to Appendix C for the details of deriving eq. (17) and (18).

5 Use the Learned Pattern ARG Model to Detect the Pattern

The learned pattern ARG captures the characteristics of a pattern observed under various conditions. It can be further used to detect the pattern in a new ARG, say G_{new}. Firstly, $P_{q_{new}}(h|G_{new}, \Theta)$ $(1 \leq h \leq M)$ is computed using the learned pattern ARG. The details of how to calculate $P_{q_{new}}(h|G_{new}, \Theta)$ will be discussed in Section 6.1. We select a component from the pattern ARG by picking up the component whose index $\ell = \arg\underset{h}{\min}P_{q_{new}}(h|G_{new}, \Theta)$. We then use two-graph matching technique to match G_{new} against the component Φ_ℓ of the pattern ARG. Those nodes of G_{new} that match with the non-null model nodes are selected. The relations among those selected nodes are preserved. The selected nodes and relations form an instance of the pattern ARG in G_{new}.

6 Implementation Issues

6.1 Match the Sample ARGs with the Pattern ARG

We use an implementation of probabilistic relaxation graph matching algorithm [6] to match each sample ARG against each component of the pattern ARG model. The matching results immediately provide $f_{y_{im}}(\eta|G_i, \Theta^{(t)}_h)$. The matching algorithm decides which model node to match with a data node o_{im} by:

$$\Upsilon_h(o_{im}) = \arg\underset{\eta}{\max}f_{y_{im}}(\eta|G_i, \Theta^{(t)}_h) \tag{19}$$

The sample images are usually noisy and contain backgrounds. This will not only affect the feature extracted for the object primitives but also create spurious nodes in the sample ARGs. To handle this problem, a null model node is generally used in the graph matching algorithms. We add a null node ω_{h0} to each model component Φ_h. The null node has no physical instance. Therefore, it neither has attributes nor has relations with other model nodes to be estimated. The null node provide a matching destination for the spurious nodes. We then define $P_{q_i}(h|G_i, \Theta^{(t)})$ as:

$$P_{q_i}(h|G_i, \Theta^{(t)}) = \frac{\sum_m f_{y_{im}}(\Upsilon_h(o_{im})|G_i, \Theta_h^{(t)})\aleph(\Upsilon_h(o_{im}))}{\sum_{k=1}^{M}\sum_m f_{y_{im}}(\Upsilon_h(o_{im})|G_i, \Theta_k^{(t)})\aleph(\Upsilon_h(o_{im}))} \tag{20}$$

$\aleph(\Upsilon_h(o_{im})) = 1$ if and only if $\Upsilon_h(o_{im}) \neq 0$. $\aleph(\Upsilon_h(o_{im}))$ encourages the case in which a data node matches with a non-null model node.

6.2 Initialize the Pattern ARG Model

Initializing the pattern ARG model is the first step of the learning procedure and is very important. The number of the model components is decided by the user or the applications. The average number of the nodes of the sample ARGs is calculated. A sample ARG whose number of nodes is the closest to the average number is selected. Let G_1 denote the selected sample ARG. The structure of G_1 is used to initialize that of one component of the pattern ARG model. In the case of a CGM model, the attributes and relations of the selected sample ARG are used to initialize the corresponding attributed means and relational means of the model component. The attributed covariances and relational covariances of the component are initialized as identical matrixes. The rest components of the pattern ARG model are set as NULL graphs and will be initialized by the following algorithm.

Algorithm 1. Initialize the Pattern ARG Model.

```
(a) for $K = 2$ to $M$
(b)     Calculate eq.(20) for each sample ARG using current pattern ARG.
(c)     Select a sample ARG $G_x = \arg\min_{G_i} \sum_{h=1}^{K-1} P_{q_i}(h|G_i, \Theta^{(t)})$.
(d)     Initialize the component $\Phi_K$ of the pattern ARG using $G_x$.
(e) endfor
```

6.3 Modify the Pattern ARG Model

It is likely to initialize the components of the pattern ARG model with spurious nodes and relations because the sample ARGs include backgrounds . To achieve better modelling results, those spurious nodes and relations should be detected and trimmed. Otherwise, they may cause serious mismatch problem if we keep updating their parameters. During the iterations of the EM algorithm, the graph matching results are examined. For each model nodes, we calculate the number of the data nodes that match with it. If the number is smaller than a threshold $\epsilon S/M$, the model node and its relations will be removed. ϵ can be a constant or a user-defined ascendant function of the iteration number of the EM algorithm.

7 Experimental Results

We take the pictures of the **MacDonald** $^{\text{TM}}$ sign in various backgrounds, from different viewpoints, and under two different light conditions. Ten images are taken under each lighting condition. Some of the images are shown in Fig. 3. The images are segmented and represented as ARGs. The node of each ARG represents an image segment and the attribute of the node is the mean color feature vector (RGB) of the segment. The adjacent relations among the segments are considered. The attributes of the sign under two different lighting conditions are greatly different from each other. Even under the same lighting condition, the attributes of the signs are different from each other due to different viewpoints. For example, the attribute vectors of **'m'** in the middle of the sign are (208, 150, 69), (202, 138, 60), (206, 144, 71), (240, 173, 116), (240, 180, 109), (241, 192, 120) in Fig. 3 (a), (b), (c), (d), (e) and (f) respectively.

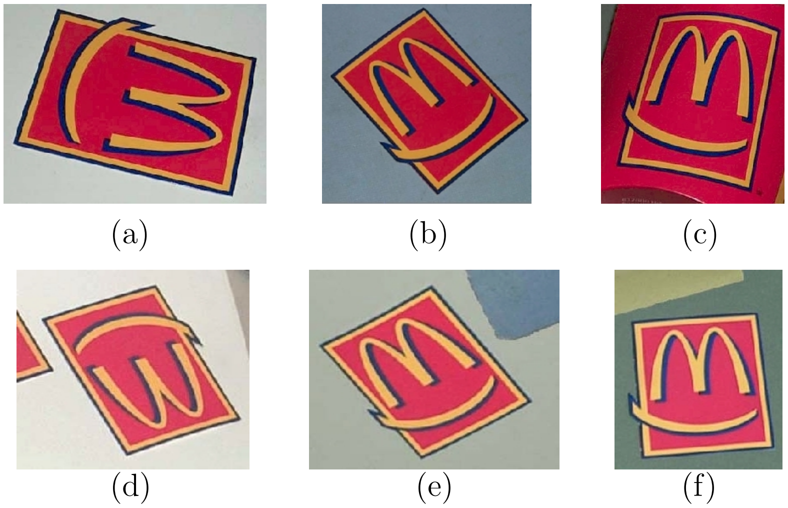
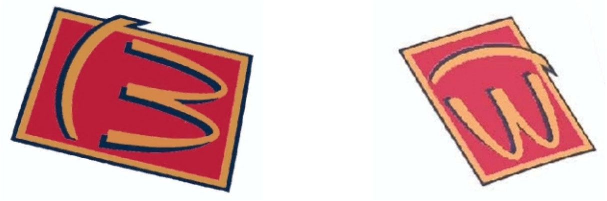

(a) (b) (c)
(d) (e) (f)

Fig. 3. The **MacDonald** $^{\text{TM}}$ sign.

A CGM model with two components is used. After learning, the training data set is summarized as two model components in the CGM model. Both of them have 8 nodes and 11 adjacent relations. To illustrate the learning results, we use the learned model to detect its isomorphic subgraph in the ARG of Fig. 3(a) and repaint the corresponding image segments using the means of the attributes of the corresponding model nodes. The same process is repeated on Fig. 3(d). The detection results are shown in Fig. 4. The mean color vectors of the model nodes that corresponding to **'m'** in the middle of the sign are (207.5, 140.3, 68.6) and (240.2, 179.7, 117.1) respectively.



(a) (b)

Fig. 4. The components of the learned pattern ARG model. (a) Model component 1, (b) model component 2.

An experiment is also conducted to match the ARG of Fig. 3(a) against that of Fig. 3(e). We modify eq. (13) and (16) to measure the attributed and relational similarity between the two ARGs. The covariance matrix $\Lambda_{h\sigma\tau}$ in eq. (16) is replaced by the identical matrix I. And the covariance matrix $\Sigma_{h\eta}$ in eq. (13) is replaced by a matrix ρI. We increase ρ from 1.0 with a step of 0.1 and calculate the matching between those two ARG for each value of ρ. If $1.0 \leq \rho < 2.3$, no correct node matching is found. If $2.3 \leq \rho < 2.5$, partial correct matching is found. If $\rho \geq 2.5$, correct matching is achieved. The **MacDonald** $^{\text{TM}}$ signs in the Fig. 3(a) and (e) share part of background, which is in light blue color. If $\rho \geq 2.5$, the light blue background, which is not part of the pattern, is however also correctly matched or extracted as part of the pattern.

Comparing the above experimental results, it won't be difficult to realize two main advantages of the our framework. An implementation of the framework can start with same initializations as those in the experiment that is just been described above, and automatic calculate the best means and covariance matrixes of the attributes and relations instead of changing them manually and blindly. Moreover, it considers multiple samples simultaneously. If the pattern is not always observed in the same background, it can learn the pattern out of its backgrounds.

8 Summary and Discussions

This paper develops theory for evidence combining that fuses the observed attributed information and contextual information of the objects. The theory is applied to unsupervised spatial pattern extraction from sample ARGs. The extracted pattern summarize the sample images and can be used for pattern detection in new images. Although the proposed theory is applied to two dimensional images in this paper, it is in its nature suitable for general spatial pattern learning and discovery because ARG can be used to represent concepts in higher dimension.

However, the learning results depend on the quality of the results of the low-level image processing. Low-level image processing must be applied to the sample images before representing them as ARGs. Currently, not effective enough high-level knowledge can be utilized in low-level image processing step. Therefore, the processing results might not be good enough under some conditions. A possible improvement would be building a feedback loop between the high-level pattern learning step and the low-level image processing step. The learned pattern encode some reliable high-level knowledge that can be applied back to doing model-based low-level image processing. The pattern can then be refined given the new results of low-level image processing. Another way to improve it is to take advantage of user interaction. For example, in relevance feedback content-based image retrieval [18], the user tries to tell the computer their information need by iteratively providing some sample images as relevance feedbacks. Our theory can be used to develop algorithm to automatic learn what the user wants based on the feedback. The user then provides the correction to the automatic learning results as relevance feedbacks to make the learning procedure more purposeful and effective.

References

[1] Almohamad, H. A., and S. O. Duffuaa, *A Linear Programming Approach for the Weighted Graph Matching Problem*, IEEE Trans. Pattern Analysis and Machine Intelligence **15** (1993), 522-525.

[2] Barrow, H. G., and R.J. Popplestone, *Relational Descriptions in Picture Processing*, Machine Intelligence **6** (1971), 377-396.

[3] Besl, P. J., and R. C. Jain, *Three-dimensional object rec-ognition*, Computing Surveys, **17** (1985), 75-145.

[4] Bhanu, B., and O. D. Faugeras, *Shape Matching of Two-Dimensional Objects*, IEEE Trans. Pattern Analysis and Machine Intelligence, **6** (1984), 137-156.

[5] Bledsoe W. W., and I. Browning, *Pattern recognition and reading by machine*, Proc. Eastern Joint Computer Conference, **16** (1959), 225-232.

[6] Christmas, W. J., J. Kittler, and M. Petrou, *Structural Matching in Computer Vision Using Probabilistic Relaxation*, IEEE Trans. Pattern Analysis and Machine Intelligence **17** (1995), no 8., 749-764.

[7] Dempster, A. P., N. M. Laird, and D. B. Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*, J. Royal Stat. Soc. Ser. B, **39** (1977), no. 1, 1-38.

[8] Felzenszwalb, P. F., and D. O. Huttenlocher, *Image Segmentation Using Local Variation*, in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, (1998), 98-104.

[9] Frey, B. J. and N. Jojic, *Transformed component analysis: Joint estimation of spatial transformations and image components*, International Conference on Computer Vision, (1999).

[10] Fu, K. S., *A Step towards Unification of Syntactic and Statistical Pattern Recognition*, IEEE Trans. Pattern Analysis and Machine Intelligence, **5** (1983), 200-205.

[11] Hong, P., R. Wang, and T. S. Huang, *Learning Patterns from Images by Combining Soft Decisions and Hard Decisions*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2000), Hilton Head Island, South Carolina.

[12] Huet, B., and E. R. Hancock, "Inexact Graph Retrieval," In Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries, pp. 40-44. 1999. Colorado. USA.

[13] Kittler, J., and J. Föglein, *Contextual classification of multispectral pixel data*, Image and Vision Computing, **2** (1984), 13-29.

[14] Li, S. Z., *Matching: Invariant to Translations, Rotations and Scale Changes*, Pattern Recognition **25** (1992) 583-594.

[15] Ozer, B., W. Wolf, and A. N. Akansu, *A Graph Based Object Description for Information Retrieval in Digital Image and Video Libraries*, In Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries, pp. 79-83. 1999.

[16] Ratan, A. L., O. Maron, *et al.*, *A framework for learning query concepts in image classification*, In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, (1999), 423-429.

[17] Rosenfeld, A., R. Hummel and S. Zucker, *Scene Labeling by Relaxation Operations*, IEEE Trans. Systems, Man and Cybernetics **6** (1976), 420-433.

[18] Rui, Y., T. S. Huang, *et al.*, *Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval*, IEEE Trans. on Circuits and Video Tech., Special Issue on Segmentation Description, and Retrieval of Video Content, **8(5)**, 1998.

[19] Shapiro, L. G., and R. M. Haralick, *Structural Descriptions and Inexact Matching*, IEEE Trans. Pattern Analysis and Machine Intelligence **3** (1981), 504-519.

[20] Umeyama, S., *An Eigen-Decomposition Approach to Weighted Graph Matching Problems*, IEEE Trans. Pattern Analysis and Machine Intelligence, **10** (1988), 695-703.

[21] Wilson, R. C., and E. R. Hancock, *Structural Matching by Discrete Relaxation*, IEEE Trans. Pattern Analysis and Machine Intelligence **19** (1997), no. 6, pp. 634-648.

APPENDIX

A Simplify the Maximum-Likelihood Function

Here, we simplify eq.(9) by simplifying its three terms separately. In the following derivations, we use the function $\delta_{m,n}$ and the fact that $\sum_{j=1}^{M} P(q_j|G_j, \Theta^{(t)}) = 1$ and $\sum_{y_{jk}=1}^{N} f(y_{jk}|G_j, \Theta_h^{(t)}) = 1$ from time to time. Note that $\delta_{m,n} = 1$ if $m = n$ and $\delta_{m,n} = 0$ if $m \neq n$.

(1) Simplify the first term of eq.(9).

$$\sum_{q_1=1}^{M} \sum_{y_{11}=1}^{N} \cdots \sum_{y_{1U_1}=1}^{N} \cdots \sum_{q_S=1}^{M} \sum_{y_{S1}=1}^{N} \cdots \sum_{y_{SU_S}=1}^{N} \sum_{i=1}^{S} (\log P(q_i)) \prod_{j=1}^{S} \Big(P(q_j|G_j, \Theta^{(t)}) \prod_{k=1}^{U_j} f(y_{jk}|G_j, \Theta_{q_j}^{(t)}) \Big)$$

$$= \sum_{q_1=1}^{M} \sum_{y_{11}=1}^{N} \cdots \sum_{y_{1U_1}=1}^{N} \cdots \sum_{q_S=1}^{M} \sum_{y_{S1}=1}^{N} \cdots \sum_{y_{SU_S}=1}^{N} \sum_{i=1}^{S} \sum_{h=1}^{M} \delta_{q_i,h} (\log P(q_i)) \prod_{j=1}^{S} \Big(P(q_j|G_j, \Theta^{(t)}) \prod_{k=1}^{U_j} f(y_{jk}|G_j, \Theta_{q_j}^{(t)}) \Big)$$

$$= \sum_{i=1}^{S} \sum_{h=1}^{M} (\log P(h)) \sum_{q_1=1}^{M} \sum_{y_{11}=1}^{N} \cdots \sum_{y_{1U_1}=1}^{N} \cdots \sum_{q_S=1}^{M} \sum_{y_{S1}=1}^{N} \cdots \sum_{y_{SU_S}=1}^{N} \delta_{q_i,h} \prod_{j=1}^{S} \Big(p(q_j|G_j, \Theta^{(t)}) \prod_{k=1}^{U_j} f(y_{jk}|G_j, \Theta_{q_j}^{(t)}) \Big)$$

$$= \sum_{i=1}^{S} \sum_{h=1}^{M} (\log P(h)) \sum_{q_1=1}^{M} \cdots \sum_{q_S=1}^{M} \prod_{j=1}^{S} \Big(\delta_{q_i,h} P(q_j|G_j, \Theta^{(t)}) \prod_{k=1}^{U_j} \sum_{y_{jk}}^{N} f(y_{jk}|G_j, \Theta_{q_j}^{(t)}) \Big)$$

$$= \sum_{i=1}^{S} \sum_{h=1}^{M} (\log P(h)) \sum_{q_1=1}^{M} \cdots \sum_{q_S=1}^{M} \prod_{j=1}^{S} \Big(\delta_{q_i,h} P(q_j|G_j, \Theta^{(t)}) \Big)$$

$$= \sum_{i=1}^{S} \sum_{h=1}^{M} \log(P(h)) \prod_{j=1}^{S} \delta_{q_i,h} \sum_{q_j=1}^{M} P(q_j|G_j, \Theta^{(t)})$$

$$= \sum_{i=1}^{S} \sum_{h=1}^{M} \log(P(h)) P_{q_i}(h|G_i, \Theta^{(t)}) = \sum_{i=1}^{S} \sum_{h=1}^{M} \log(\alpha_h) P_{q_i}(h|G_i, \Theta^{(t)}) \tag{A.1}$$

(2) Simplify the second term of eq.(9). The same method for simplifying the

first term of eq.(9) is used here.

$$
\begin{aligned}
&\sum_{q_1=1}^{M}\sum_{y_{11}=1}^{N}\cdots\sum_{y_{1U_1}=1}^{N}\cdots\sum_{q_S=1}^{M}\sum_{y_{S1}=1}^{N}\cdots\sum_{y_{SU_S}=1}^{N}\sum_{i=1}^{S}\sum_{m=1}^{U_i}\log\Big(p(o_{im}|\omega_{q_i y_{im}})P(y_{im}|q_i)\Big)\\
&\quad\prod_{j=1}^{S}\Big(P(q_j|G_j,\Theta^{(t)})\prod_{k=1}^{U_j}f(y_{jk}|G_j,\Theta^{(t)}_{q_j})\Big)\\
&=\sum_{q_1=1}^{M}\sum_{y_{11}=1}^{N}\cdots\sum_{y_{1U_1}=1}^{N}\cdots\sum_{q_S=1}^{M}\sum_{y_{S1}=1}^{N}\cdots\sum_{y_{SU_S}=1}^{N}\sum_{i=1}^{S}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\delta_{y_{im},\eta}\\
&\quad\log\Big(p(o_{im}|\omega_{q_i y_{im}})P(y_{im}|q_i)\Big)\prod_{j=1}^{S}\Big(P(q_j|G_j,\Theta^{(t)})\prod_{k=1}^{U_j}f(y_{jk}|G_j,\Theta^{(t)}_{q_j})\Big)\\
&=\sum_{q_1=1}^{M}\cdots\sum_{q_S=1}^{M}\sum_{i=1}^{S}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\log\Big(p(o_{im}|\omega_{q_i\eta})P(\eta|q_i)\Big)\\
&\quad\sum_{y_{11}=1}^{N}\cdots\sum_{y_{1U_1}=1}^{N}\cdots\sum_{y_{S1}=1}^{N}\cdots\sum_{y_{SU_S}=1}^{N}\delta_{y_{im},\eta}\prod_{j=1}^{S}\Big(P(q_j|G_j,\Theta^{(t)})\prod_{k=1}^{U_j}f(y_{jk}|G_j,\Theta^{(t)}_{q_j})\Big)\\
&=\sum_{q_1=1}^{M}\cdots\sum_{q_S=1}^{M}\sum_{i=1}^{S}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\log\Big(p(o_{im}|\omega_{q_i\eta})P(\eta|q_i)\Big)f_{y_{im}}(\eta|G_i,\Theta^{(t)}_{q_i})\\
&\quad\prod_{j=1}^{S}P(q_j|G_j,\Theta^{(t)})\\
&=\sum_{q_1=1}^{M}\cdots\sum_{q_S=1}^{M}\sum_{i=1}^{S}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\sum_{h=1}^{M}\delta_{q_i,h}\log\Big(p(o_{im}|\omega_{q_i\eta})P(\eta|q_i)\Big)f_{y_{im}}(\eta|G_i,\Theta^{(t)}_{q_i})\\
&\quad\prod_{j=1}^{S}P(q_j|G_j,\Theta^{(t)})\\
&=\sum_{i=1}^{S}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\sum_{h=1}^{M}\log\Big(p(o_{im}|\omega_{h\eta})P(\eta|h)\Big)f_{y_{im}}(\eta|G_i,\Theta^{(t)}_{h})P_{q_i}(h|G_i,\Theta^{(t)})\\
&=\sum_{i=1}^{S}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\sum_{h=1}^{M}\log\Big(p(o_{im}|\omega_{h\eta})\Big)f_{y_{im}}(\eta|G_i,\Theta^{(t)}_{h})P_{q_i}(h|G_i,\Theta^{(t)})+\\
&\quad\sum_{i=1}^{S}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\sum_{h=1}^{M}\log(\beta_{h\eta})f_{y_{im}}(\eta|G_i,\Theta^{(t)}_{h})P_{q_i}(h|G_i,\Theta^{(t)})
\end{aligned}
\tag{A.2}
$$

(3) Simplify the third term of eq.(9). Again, the same simplification method

is used.

$$\begin{aligned}
&\sum_{q_1=1}^{M}\sum_{y_{11}=1}^{N}\cdots\sum_{y_1U_1=1}^{N}\cdots\sum_{q_S=1}^{M}\sum_{y_S1=1}^{N}\cdots\sum_{y_{SU_S}=1}^{N}\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\log\Big(p(r_{icd}|\Re_{\overrightarrow{y}_i}(r_{icd}))\Big)\\
&\quad\prod_{j=1}^{S}\Big(P(q_j|G_j,\Theta^{(t)})\prod_{k=1}^{U_j}f(y_{jk}|G_j,\Theta_{q_j}^{(t)})\Big)\\
&=\sum_{q_1=1}^{M}\sum_{y_{11}=1}^{N}\cdots\sum_{y_1U_1=1}^{N}\cdots\sum_{q_S=1}^{M}\sum_{y_S1=1}^{N}\cdots\sum_{y_{SU_S}=1}^{N}\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\sum_{\sigma=1}^{N}\sum_{\tau=1}^{N}\delta_{y_{ic},\sigma}\delta_{y_{id},\tau}\\
&\quad\log\Big(p(r_{icd}|\Re_{\overrightarrow{y}_i}(r_{icd}))\Big)\prod_{j=1}^{S}\Big(P(q_j|G_j,\Theta^{(t)})\prod_{k=1}^{U_j}f(y_{jk}|G_j,\Theta_{q_j}^{(t)})\Big)\\
&=\sum_{q_1=1}^{M}\cdots\sum_{q_S=1}^{M}\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\sum_{\sigma=1}^{N}\sum_{\tau=1}^{N}\log\Big(p(r_{icd}|\psi_{q_i\sigma\tau})\Big)\\
&\quad f_{y_{ic}}(\sigma|G_i,\Theta_{q_i}^{(t)})f_{y_{id}}(\tau|G_i,\Theta_{q_i}^{(t)})\prod_{j=1}^{S}P(q_j|G_j,\Theta^{(t)})\\
&=\sum_{q_1=1}^{M}\cdots\sum_{q_S=1}^{M}\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\sum_{\sigma=1}^{N}\sum_{\tau=1}^{N}\sum_{h=1}^{M}\delta_{q_i,h}\log\Big(p(r_{icd}|\psi_{q_i\sigma\tau})\Big)\\
&\quad f_{y_{ic}}(\sigma|G_i,\Theta_{q_i}^{(t)})f_{y_{id}}(\tau|G_i,\Theta_{q_i}^{(t)})\prod_{j=1}^{S}P(q_j|G_j,\Theta^{(t)})\\
&=\sum_{i=1}^{S}\sum_{c=1}^{U_i}\sum_{d=1}^{U_i}\sum_{\sigma=1}^{N}\sum_{\tau=1}^{N}\sum_{h=1}^{M}\log\Big(p(r_{icd}|\psi_{h\sigma\tau})\Big)f_{y_{ic}}(\sigma|G_i,\Theta_h^{(t)})f_{y_{id}}(\tau|G_i,\Theta_h^{(t)})\\
&\quad P_{q_i}(h|G_i,\Theta^{(t)})
\end{aligned}\tag{A.3}$$

Finally, we can obtain eq.(10) by submitting eq.(A.1), (A.2), (A.3) into eq.(9).

B Derive Expressions for Updating α_h and $\beta_{h\eta}$

The four terms of the eq.(10) can be maximized separately while we try to calculate $\Theta^{(t+1)}$.

(1) First, we derive the update expression for α_h by maximizing the first term of eq.(10). We introduce the Lagrange multiplier λ with the constraint that $\Sigma_h\alpha_h = 1$, and solve the following equation:

$$
\begin{aligned}
&\frac{\partial}{\partial \alpha_h}\Bigg[\sum_{i=1}^{S}\sum_{h=1}^{M}\log(\alpha_h)P_{q_i}(h|G_i,\Theta^{(t)})+\lambda\Big(\sum_{h=1}^{M}\alpha_h-1\Big)\Bigg]\\
&=\sum_{i=1}^{S}\frac{1}{\alpha_h}P_{q_i}(h|G_i,\Theta^{(t)})+\lambda=0\Longrightarrow\\
&\sum_{h=1}^{M}\Bigg[\sum_{i=1}^{S}\frac{1}{\alpha_h}P_{q_i}(h|G_i,\Theta^{(t)})+\lambda\Bigg]=0\Longrightarrow\lambda=-S\Longrightarrow\\
&\alpha_h=\frac{\sum_{i=1}^{S}P_{q_i}(h|G_i,\Theta^{(t)})}{S}
\end{aligned}
\tag{B.1}
$$

(2) Secondly, we derive the update expression for $\beta_{h\eta}$ by maximizing the second term of eq.(10). Again , we introduce the Lagrange multiplier λ with the constraint that $\Sigma_\eta \beta_{h\eta}=1$, and solve the following equation:

$$
\begin{aligned}
&\frac{\partial}{\partial \beta_{h\eta}}\Bigg[\sum_{i=1}^{S}\sum_{h=1}^{M}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\log(\beta_{h\eta})f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})+\lambda\Big(\sum_{\eta=1}^{N}\beta_{h\eta}-1\Big)\Bigg]\\
&=\sum_{i=1}^{S}\sum_{m=1}^{U_i}\frac{1}{\beta_{h\eta}}f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})+\lambda=0\\
&\Longrightarrow\sum_{h=1}^{M}\sum_{\eta=1}^{N}\Bigg[\sum_{i=1}^{S}\sum_{m=1}^{U_i}\frac{1}{\beta_{h\eta}}f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})+\lambda\Bigg]=0\Longrightarrow\\
&\lambda=\frac{\sum_{i=1}^{S}U_i}{M}\Longrightarrow\beta_{h\eta}=\frac{M\sum_{i=1}^{S}\sum_{m=1}^{U_i}f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})}{\sum_{i=1}^{S}U_i}
\end{aligned}
\tag{B.2}
$$

It will not be difficult to find out that eq.(11) and eq.(12) are eq.(B.1) and eq.(B.2) with the iteration index added respectively.

C Derive Expressions for Updating the Parameters of the Gaussian Attributed and Relational Distributions

For Gaussian distribution assumptions, we obtain analytical expressions for updating the parameters of the distribution functions. (1) For Gaussian attributed distribution (eq.(13)), the update expression for the attributed distribution parameters are derived by maximizing the third term of eq.(10).

$$\begin{aligned}&\sum_{i=1}^{S}\sum_{h=1}^{M}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\log(p(o_{im}|\omega_{h\eta}))f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})\\&=\sum_{i=1}^{S}\sum_{h=1}^{M}\sum_{m=1}^{U_i}\sum_{\eta=1}^{N}\frac{-1}{2}\Bigg[\xi\log(2\pi)+\log(|\Sigma_{h\eta}^{(t+1)}|)+(\overrightarrow{a}_{im}-\overrightarrow{\mu}_{h\eta}^{(t+1)})^T\\&\quad{\Sigma_{h\eta}^{(t+1)}}^{-1}(\overrightarrow{a}_{im}-\overrightarrow{\mu}_{h\eta}^{(t+1)})\Bigg]f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})\end{aligned}\tag{C.1}$$

Take the derivative of eq.(C.1) with respect to $\overrightarrow{\mu}_{h\eta}^{(t+1)}$ and set it equal to zero, we can have:

$$\begin{aligned}&\sum_{i=1}^{S}\sum_{m=1}^{U_i}{\Sigma_{h\eta}^{(t+1)}}^{-1}(\overrightarrow{a}_{im}-\overrightarrow{\mu}_{h\eta}^{(t+1)})f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})=0\\&\Longrightarrow\overrightarrow{\mu}_{h\eta}^{(t+1)}=\frac{\sum_{i=1}^{S}\sum_{m=1}^{U_i}\overrightarrow{a}_{im}f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})}{\sum_{i=1}^{S}\sum_{m=1}^{U_i}f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})}\end{aligned}\tag{C.2}$$

Take the derivative of eq.(C.1) with respect to $\Sigma_{h\eta}^{(t+1)}$ and set it equal to zero, we can have:

$$\begin{aligned}&\sum_{i=1}^{S}\sum_{m=1}^{U_i}f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})\Bigg(2(\Sigma_{h\eta}^{(t+1)}-\overrightarrow{x}_{im}^{(t)}{\overrightarrow{x}_{im}^{(t)}}^T)-\\&\qquad diag(\Sigma_{h\eta}^{(t+1)}-\overrightarrow{x}_{im}^{(t)}{\overrightarrow{x}_{im}^{(t)}}^T)\Bigg)=0\\&\Rightarrow\sum_{i=1}^{S}\sum_{m=1}^{U_i}f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})(\Sigma_{h\eta}^{(t+1)}-\overrightarrow{x}_{im}^{(t)}{\overrightarrow{x}_{im}^{(t)}}^T)\\&\Rightarrow\Sigma_{h\eta}^{(t+1)}=\frac{\sum_{i=1}^{S}\sum_{m=1}^{U_i}\overrightarrow{x}_{im}^{(t)}{\overrightarrow{x}_{im}^{(t)}}^Tf_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})}{\sum_{i=1}^{S}\sum_{m=1}^{U_i}f_{y_{im}}(\eta|G_i,\Theta_h^{(t)})P_{q_i}(h|G_i,\Theta^{(t)})}\end{aligned}\tag{C.3}$$

where $\overrightarrow{x}_{im}^{(t)}=\overrightarrow{a}_{im}-\overrightarrow{\mu}_{h\eta}^{(t)}$.

(2) For Gaussian relational distribution (eq.(16)), the update expressions for the relational distribution parameters are derived by maximizing the fourth term of eq.(10). Using the same procedure for deriving the expressions for $\overrightarrow{\mu}_{h\eta}^{(t+1)}$ and $\Sigma_{h\eta}^{(t+1)}$, we can obtain the update expressions for $\overrightarrow{\gamma}_{h\sigma\tau}$ and $\Lambda_{h\sigma\tau}$ as eq.(17) and eq.(18) respectively. Due to the space problem, the details will be neglected here.