A NOVEL PLAUSIBLE MODEL FOR VISUAL PERCEPTION

Zhiwei Shi  Zhongzhi Shi  Hong Hu
Key Laboratory of Intelligent Information Processing,
Institute of Computing Technology, CAS, Beijing 100080, China
shizw@ics.ict.ac.cn

Abstract
Traditionally, how to bridge the gap between the low level visual features and the high level semantic concepts has been a tough task for the researchers. In this paper, we propose a novel plausible model, namely globally connected and locally autonomic Bayesian network (GCLABN), to model the process of visual perception. The new model takes advantage of both the low level visual features, such as colors, textures and shapes, of the target object and the interrelationship between the known objects, and integrates them into a Bayesian framework, which possesses both firm theoretical foundation and wide practical applications. According to our meticulous analysis, in many aspects, the novel model theoretically outperforms the original Bayesian network, which has been successfully applied to many related areas, such as object detection, scene analysis and other similar tasks. Finally, although the GCLABN is designed for the visual perception, it also has great potential to be applied to other areas.

Keywords: Visual perception; Bayesian network.

1. INTRODUCTION
The primary task of perceptual organization is to organize the visual features of an image into some already known objects. Yet, how to bridge the gap between the low level visual features and the high level semantic concept has long been a tough problem, which puzzles the researchers all along. Until now, most of the proposed algorithms just focus on some particular objects, such as human faces, cars, people, and so on (see e.g., [19, 22, 25]). Researchers utilize various schemes (see e.g., [2, 14, 24]) to integrate the low level visual features, including colors, textures and shapes into the profile of the target object. Although some people [18] exploit the background, or scene, information to improve the recognition, they do not take advantage of the interrelationship between objects to help the detection process.

As a matter of fact, the interrelationship between all the objects is of great importance to the perceptual organization. Researchers have shown that there might exist some visual patterns in the human brains, which enable human beings to recognize some simple objects, e.g. English letters, as soon as they see them. Besides, they also point out that the speed of processing visual information is very limited in the human brain [10]. Consequently, it is highly plausible that human beings utilize the interrelationships to facilitate the recognition of the complex, unfamiliar objects based on the recognition of some simple, familiar objects. Obviously, the interrelationships will largely reduce the information required for the recognition, and improve the effectiveness and efficiency as well. Further more, the lateral connections, which widely exist in the cortex of human brains [5], also provide the biological support of the usage of the interrelationship between objects.

Meanwhile, due to the large amount of uncertainty in the process of perceptual organization, Bayesian methods are widely used in the modeling of perceptual organization or object identification (see e.g., [9, 11, 16]. Typically, uncertainties can rise in a large amount of cases. For example, the target object is too small in the visual scene, or only part of the target object is visible due to some obstructions, or the visual scene is vague in some bad weather. As a result, Bayesian method, which can deal with uncertainties and make use of uncertainties, is obviously a favorite option for many researchers.

In this paper, we propose a variation of the Bayesian network, a globally connected and locally autonomic Bayesian network (GCLABN), which makes use of both the low level visual features and the interrelationships between high level objects under a Bayesian framework to model the visual perception.

The rest of this paper is organized as follows: in Section 2, we will briefly review the related works, mainly the Bayesian network, and analyze the weakness of the traditional Bayesian network in the task of visual perception; next, in Section 3, we propose our novel model, the globally connected and locally autonomic Bayesian network (GCLABN), including the definitions, the learning methods and the inference using GCLABN;
then, in section 4, we discuss the merits and demerits of
the GCLABN for the task of visual perception; finally we
conclude our paper with a short review of our work in
this paper and possible directions in the future in the last
section.

2. RELATED WORK

2.1 Bayesian Network

Bayesian network, as a probabilistic graphical model,
is an increasingly popular tool for encoding uncertain
knowledge in complex domains, and, in the last decade,
it gains wide applications in many areas, such as decision
support systems, information retrieval, discovery of gene
regulatory pathways, natural language processing (see
e.g., [1, 8, 15]), and especially object detection (see e.g.,
[12, 13, 23]). Mathematically, the Bayesian network can
be defined as follows:

A Bayesian network is a pair \((G, P)\), where \(G = (V, E)\)
is a directed acyclic graph (DAG). Here \(V\) is the node set,
where each node represents the variables in the problem
domain, and \(E\) is the edge set, where each edge denotes
the direct dependent relationship between nodes. Another
component of the Bayesian network, \(P\), is the set of
conditional probabilistic distribution (CPD) of the child
node, given the parent nodes. Additionally, \(G\) and \(P\) must
satisfy the Markov condition: every variable, \(X \in V\), is
independent of any subset of its non-descendant variables
conditioned on the set of its parents [20]. In many cases,
a directed edge from one node to another can be
interpreted as a causal relationship although the absence
of the edge does not mean the independence between the
two nodes. So, in some cases, a Bayesian network is also
called a causal network.

Let \(pa(X)\) denote the set of parents of \(X\), then the
conditional independence property can be represented as
follows:

\[
P(X \mid V \setminus X) = P(X \mid pa(X))
\]

(1)

This property largely relaxes the strict memory and
computing power requirements by simplifying the
computation in a Bayesian network. For example, the
joint distribution of the set of all the variables in \(V\) can be
expressed as a product of conditional probabilities as
follows:

\[
P(V) = \prod_{X \in V} P(X \mid pa(X))
\]

(2)

2.2 The Weakness Of Bayesian Network

Although the Bayesian network has so many attractive
properties, such as the firm theoretical foundation, the
simple representation and the capability of encoding the
interrelationships between objects (variables), it cannot
be employed as the model of perceptual organization
directly without any modification. Here are some reasons.

The most important reason is about the incomplete
data. As we all know, each piece of our observation
contains only a tiny part of the real world, as well as its
reflection in our mind. It is absolutely impossible to have
a completed sample of all the objects that we know
within just one observation. This means that the task of
learning the Bayesian network will inevitably depend on
the techniques of learning with incomplete data.

Although, in the last decade, with the rapid progress in
the research on the Bayesian network, numerous
algorithms [7, 21, 26] have been developed to deal with
the incomplete data in the learning of Bayesian network,
the proportion of the missing data in their researches is
relative small. Typically, their missing rate is less than
50%. Therefore, in the cases that only several objects are
observed with millions of objects missing, their methods
will definitely become powerless.

The second problem is that, in the Bayesian network,
the edges between the nodes are one-way directed. Yet,
the interrelationships between the objects, which exist in
the real world and can be reflected in the human brains,
sometimes are bidirectional. In these cases, the original
Bayesian network is incapable for the task of perceptual
organization due to its inherent limitation.

The last point comes from the feasibility. The problem
of learning an exact Bayesian network from data under
certain broad conditions has been proven worst-case NP-
hard [4]. Although researchers have proposed heuristic
polynomial learning algorithms [3] to alleviate the work,
it is, computationally, still intractable for dealing with the
problem with millions of variables in the domain.

Additionally, researches have shown that making
inference on such a large-scale network is also infeasible
[6].

Based on the analyses above, we get that it is not wise
to exploit a huge global Bayesian network to construct a
perceptual model. This might be the reason why the
Bayesian network is merely applied to single object
detection. Consequently, we modify the original
Bayesian network according to the observations of the
order of nature, and propose a novel perceptual model,
globally connected and locally autonomic Bayesian
network (GCLABN), in accordance. In the following
section, we will discuss the model in detail.

3. GLOBALLY CONNECTED AND
LOCALLY AUTONOMIC BAYESIAN
NETWORK (GCLABN)

3.1 Motivation

---

\[
\text{follows:}
\]

\[
\text{expressed as a product of conditional probabilities as}
\]

\[
\text{joint distribution of the set of all the variables in}
\]

\[
\text{computing power requirements by simplifying the}
\]

\[
\text{simple representation and the capability of encoding the}
\]

\[
\text{properties, such as the firm theoretical foundation, the}
\]

\[
\text{simple representation and the capability of encoding the}
\]

\[
\text{interrelationships between objects (variables), it cannot}
\]

\[
\text{be employed as the model of perceptual organization}
\]

\[
directly without any modification. Here are some reasons.
\]

\[
\text{The most important reason is about the incomplete}
\]

\[
data. As we all know, each piece of our observation
\]

\[
\text{contains only a tiny part of the real world, as well as its}
\]

\[
\text{reflection in our mind. It is absolutely impossible to have}
\]

\[
a completed sample of all the objects that we know
\]

\[
within just one observation. This means that the task of
\]

\[
learning the Bayesian network will inevitably depend on
\]

\[
the techniques of learning with incomplete data.}
\]

\[
Although, in the last decade, with the rapid progress in
\]

\[
the research on the Bayesian network, numerous
\]

\[
algorithms [7, 21, 26] have been developed to deal with
\]

\[
the incomplete data in the learning of Bayesian network,
\]

\[
the proportion of the missing data in their researches is
\]

\[
relative small. Typically, their missing rate is less than
\]

\[
50%. Therefore, in the cases that only several objects are
\]

\[
observed with millions of objects missing, their methods
\]

\[
will definitely become powerless.
\]

\[

\]

\[
\text{edges between the nodes are one-way directed. Yet,}
\]

\[
the interrelationships between the objects, which exist in
\]

\[
the real world and can be reflected in the human brains,
\]

\[
sometimes are bidirectional. In these cases, the original
\]

\[
Bayesian network is incapable for the task of perceptual
\]

\[
organization due to its inherent limitation.
\]

\[

\]

\[
\text{the feasibility. The problem}
\]

\[
of learning an exact Bayesian network from data under
certain broad conditions has been proven worst-case NP-
hard [4]. Although researchers have proposed heuristic
polynomial learning algorithms [3] to alleviate the work,
it is, computationally, still intractable for dealing with the
problem with millions of variables in the domain.

Additionally, researches have shown that making
inference on such a large-scale network is also infeasible
[6].

Based on the analyses above, we get that it is not wise
to exploit a huge global Bayesian network to construct a
perceptual model. This might be the reason why the
Bayesian network is merely applied to single object
detection. Consequently, we modify the original
Bayesian network according to the observations of the
order of nature, and propose a novel perceptual model,
globally connected and locally autonomic Bayesian
network (GCLABN), in accordance. In the following
section, we will discuss the model in detail.

3. GLOBALLY CONNECTED AND
LOCALLY AUTONOMIC BAYESIAN
NETWORK (GCLABN)

3.1 Motivation

---
Before we depict our model, we will examine some nature principles, which inspire us to construct our model.

Basically, our model comes out from the following observation of the order of nature, by which everything in the word exists and functions.

Firstly, everything in the world is unique. Everything has its distinctive features that make it different from others. Many object detection algorithms (see e.g., [2, 14, 24]) are developed based on this principle. Secondly, the appearance of each object has variety. One object will present various different appearances due to different illumination, or different visual angles of the observer, or some other factors. That is why many researchers intend to exploit Bayesian method to deal with the uncertainty in the observation. Thirdly, everything exists in some particular local circumstances, or scenes, and has more or less relationships with its neighbors. It is a pity that researchers [18] take the circumstance information into account, while leave the relationships between objects alone.

The rules above describe the way in which physical world exists. According to the isomorphism\(^1\) of the Gestalt Theory, the brain field of human beings should have the similar structure with that of the external world exists. According to the isomorphism of the Gestalt Theory, the brain field of human beings should have the similar structure with that of the external world exists. According to the isomorphism of the Gestalt Theory, the brain field of human beings should have the similar structure with that of the external world exists. According to the isomorphism of the Gestalt Theory, the brain field of human beings should have the similar structure with that of the external circumstances, which is reflected by the inside perceptual model. Accordingly, it is highly reasonable to make a perceptual model that takes advantage of both the unique visual features of individual objects and the interrelationships between objects.

### 3.2 Model Description

As we have seen, a global Bayesian network is not capable to model the perceptual organization. So, in our model, we interconnect numerous overlapped local Bayesian network into a whole to avoid the problems listed in the previous section.

**Definition 3.1** In a directed acyclic graph \( G = (V, E) \), a node \( C \in V \) is called a **center node** if and only if for any node \( X \in V \setminus \{C\} \), the edge \( <X, C> \in E \) is satisfied\(^2\). For example, in Figure1, the node \( n_2 \) in the graph \( G_1 \) and the node \( n_4 \) in the graph \( G_2 \) are all center nodes in their graphs respectively.

**Definition 3.2** A directed acyclic graph is a **centered graph** if and only if it has only one center node, and there is no edges ending at the non-center nodes. For example, the graph \( G_1 \) is a centered graph, and the graph \( G_2 \) is not due to the extra edges between non-center nodes. The center node of a centered graph is specified shortly to \( C(G) \). So, in Figure1, we have \( C(G_1) = n_2 \).

\begin{itemize}
    \item \( n_1 \)
    \item \( n_2 \)
    \item \( n_3 \)
    \item \( n_4 \)
    \item \( n_5 \)
\end{itemize}

**Fig.1. The Sample Graphs**

**Definition 3.3** A centered Bayesian network (CBN) is a pair \((G, P)\), where \( G = (V, E) \) is a centered graph, and \( P \) is a set of conditional probabilistic distribution (CPD) of the center node of \( G \) given its parents. For convenience, we also call the center of \( G \), \( C(G) \), the center of the CBN, shortly for CBN).

**Definition 3.4** For any two given centered Bayesian network \( B_1 = (G_1, P_1) \) and \( B_2 = (G_2, P_2) \), if \( C(B_1) \neq C(B_2) \), the union \( B = B_1 \cup B_2 \) of the CBN \( B_1 \) and CBN \( B_2 \) is a pair \((G, P)\), where \( G \) is the graph sum\(^3\) of \( G_1 \) and \( G_2 \), and \( P \) is the union of \( P_1 \) and \( P_2 \).

Figure2 illustrates a sample union of two CBNs. From the figure, we can see that, in the left oval is one CBN \( B_1 \), whose edge is solid arrowed lines, and in the right oval is the other CBN \( B_2 \), whose edge is dot arrowed line. The union of \( B_1 \) and \( B_2 \) is also a labeled graph model, whose graph structure is just the graph sum of the graphs of the two CBNs, and its notation \( P \) is the union of the notations of \( B_1 \) and \( B_2 \). Obviously, the union of two CBNs is not a CBN any more, because the center node is not guaranteed, and more CPDs are included.

\begin{itemize}
    \item \( n_1 \)
    \item \( n_2 \)
    \item \( n_3 \)
    \item \( n_4 \)
    \item \( n_5 \)
\end{itemize}

**Fig.2. The union of two CBNs**

\(^1\) In mathematics an isomorphism between two systems requires a one-to-one correspondence between their elements (that is, each element of one system corresponds to one and only one element of the other system, and conversely), which also preserves structures. In Gestalt psychology, the one-to-one correspondence between elements is not required; similarity of structures is required. [17]

\(^2\) Here \( X \) is the start point of the edge.

\(^3\) The definition of “Graph Sum” can be found from MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/GraphSum.html
Definition 3.5 A globally connected and locally autonomic Bayesian network (GCLABN) is a union, \( GCLABN = B_1 \cup B_2 \cup \ldots \cup B_n \), where \( B_i \) (\( i = 1, \ldots, n \)) are centered Bayesian networks, and they are subject to the following constraints:

\[
C(B_i) \neq C(B_j) \quad 1 \leq i < j \leq n
\] (3)

and

\[
C(B_j) \in \left( \bigcup_{j \neq i} V_j \right) \quad i = 1, \ldots, n
\] (4)

where \( V_j \) is the node set of \( B_j \).

The first constraint makes every component network unique. In the later section, we will find that each center node of a CBN represents a complicated object to be recognized. The second constraint makes all the component networks connected in terms of center nodes. If the center node corresponds to the complicated object, the second constraint just encodes the interrelationship between these objects and indicates that no complicated object is isolated from other complicated ones.

Any common Bayesian network can be viewed as a GCLABN, if we regard each family of a child node, including the child node and its parent nodes, as a CBN. Yet, we cannot transfer some GCLABN into a common Bayesian network, if there exists directed loop in the GCLABN.

3.3 Learning A GCLABN

Since a GCLABN is consisted of many interconnected but autonomic component Bayesian networks, the learning task of the model can be divided into many independent parallelizable tiny learning task. Each task just focuses on one component network. Compared with the common Bayesian network, a centered Bayesian network is much simpler, because, for a given problem domain, or a set of variables, and a center node, the network structure can be fixed (all edges pointing to the center node from the non-center nodes), and only the CPD of the center node needs to be learned. As a result, only the parameter learning technique of Bayesian network is needed, and the learning process can be much more efficient than the ordinary Bayesian network due to the simplicity of the CBN.

3.4 Inference On A GCLABN

The inference of the model is a little complicated. As we presented above, a GCLABN is composed of more than one overlapped CBNs. It is highly possible that there exist directed circles among the nodes, though the graph model of each CBN is acyclic. In these cases, traditional inference algorithms will not work, because the inference process among the nodes will be recursive. One feasible solution is to update the state, or value, of the center node of each CBN recursively, until there is no state changing or the number of recursion is beyond the fixed threshold. The detailed algorithm is listed in Figure 3.

\[
num = 1
\]

while \( \text{Ev} \neq \Phi \) or \( \text{num} < \text{Threshold} \) \{

\[
\text{num}++;\]

\[
\text{NewEv} = \Phi;\]

For each CBNi \{

\[
\text{Ev}_i = \text{Ev}_i \cap (V_i \setminus C(CBN));\]

If \( \text{Ev}_i \neq \Phi \) \{

\[
\text{Update the state of } C(CBN) \text{ according to } \text{Ev}_i;\]

If (the state of \( C(CBN) \) changed)

\[
\text{NewEv} = \text{NewEv} \cup C(CBN);\]

\}

\[
\text{Ev} = \text{NewEv};\]
\}

Fig. 3. The Algorithm for Inference on The GCLABN

In this algorithm, \( \text{Ev} \) denotes the set of evidences, \( \text{NewEv} \) specifies the new set of evidences after each iteration, \( V_i \) is the node set of the \( i \)th CBN (CBNi), and \( C(CBN) \) is the center node of CBNi. In the boldfaced line, any practical method can be used to determine the state of the center node. As a matter of fact, this operation includes two parts: Firstly, the posterior distribution is updated according to the evidences. Any feasible inference algorithm for the common Bayesian network will be competent for this simple task. Then we will sample a state for the center node based on its posterior distribution via some predefined process, e.g. Monte Carlo method or some other similar methods.

3.5 Model The Visual Perception With GCLABN

In this subsection, we will exploit the GCLABN to model the visual perception.

The fundamental issue is to map the objects of problem domain into the variables, or nodes, in the GCLABN. Typically, the objects, or variables, in a perceptual system refer to any perceptible stuff in the real world. Yet, because the identification of any target object involves both the basic visual features, such as colors, textures and shapes, and other related objects, which have more or less relationship with the target object. To make the model uniform, we regard the basic visual features, including colors, textures and shapes, as the objects in the GCLABN as well. Thus, the entire set of objects can be divided into two categories. One is for the basic objects, which can be detected by the visual system directly without any inference. This category will include, but not limited to, the basic visual features. The other is
for the high level objects, which is detected based on the
identification of other objects and some necessary
inference. The variables representing the objects in the
first category, namely elementary objects, will
correspond to the non-center nodes in the GCLABN. In
contrast, the variables representing the objects in the
second category, namely complicated objects, will
become the center nodes in the GCLABN.

Another key issue of constructing the model is how to
decide the node set of each component network $CBN_i$, $i = 1, \ldots, n$. As we mentioned previously, everything in
the world exist in some particular circumstances and often
co-occur with some other objects. It is nature to use the
frequency of the co-occurrence in a same visual scene of
the two objects to quantify the relationship between them.
Since each CBN corresponds to a complicated object, we
can figure out that the node set of this CBN will include
all the elementary features of its center object, and all the
complicated objects that have ever co-occurred with the
center object in a same scene.

After the GCLABN for visual perception is defined.
The learning method mentioned in section 3.3 can be
used to learn the parameters of the GCLABN. Then, the
inference algorithm, described in section 3.4, will be
utilized to recognize objects according to some visual
input.

4. DISCUSSION

In this section, we will examine the merits and
demerits of the GCLABN for the task of visual
perception. As we show in the previous section, the
GCLABN employs Bayesian network as its component,
so it gains all the merits of Bayesian network, such as the
theoretical foundation and the simple representation.
Besides, the GCLABN also possesses the merits that
ordinary Bayesian network does not have.

The most distinctive merit of the GCLABN is that its
structure fully represents the order of nature about how
the things in the physical world exist. For it encodes both
the distinctive visual features of the perceptual objects
and the interrelationship between them. Furthermore, it
exploits the Bayesian framework to deal with the
uncertainty in the perceptual process. The model is
greatly consistent with the nature of the physical world.

Another significant merit of the GCLABN is that it
has predefined network structure (all non-center nodes
pointing to the center node), which encodes more prior
knowledge than the common Bayesian network does.
The prior knowledge not only makes the model more
reasonable, but also largely reduces the time consumed in
the learning process.

In addition, the GCLABN supports both batch
learning and online learning. As we all know, it is hard to
perform online learning for common Bayesian network,
because the fresh incoming data will modify both the
parameters and the structure of the Bayesian network.
And the searching of proper structure is extremely time
consuming [4]. On the contrary, for GCLABN, the
property of predefined network structure enables the
GCLABN to have more stationary structure. In most
cases, only the parameters need to be updated. The
possible modification of structure involves only putting
on new edges between the objects in the new sample. As
the learning process is limited to a relative small area, the
learning is actually very efficient. Therefore, the online
learning is more feasible for GCLABN.

The last but not least merit of the GCLABN is its
capability for scaling up. Because the GCLABN is an
aggregation of numerous autonomic CBNs, the learning
process, as well as the inference process, can be
decomposed into many isolated tasks that can be
performed at each CBN. Namely, the GCLABN is a
distributed system. Any fresh CBN can be added into the
model at any moment, without much modification in the
existing model. Actually, only several neighbor CBNs
need slight modification. That means the system can be
easily scale up. Further more, with more CBNs coming
into the system, the system will accumulate more
knowledge about the external world. In this sense, the
GCLABN is a real learning system that can make
accumulations of knowledge.

Certainly, the GCLABN does have its weaknesses,
especially the inference process. Due to its high
possibility to have directed circles in the model, the
inference will, theoretically, be more time consuming
then the common Bayesian network. So, more
mechanisms should be added to the model in the future to
speed up the inference process. In addition, there also
needs some limitation on the number of the connections
from the neighbor center nodes to prevent too much
complicated network structure.

5. CONCLUSIONS

How to integrate various visual features from scenes
to form known objects has traditionally been a tough task
for the researchers. In this paper, we propose a novel
model, namely globally connected and locally autonomic
Bayesian network (GCLABN), to deal with the visual
perceptual task. The proposed model takes both the
distinctive visual features of the target objects and the
interrelationship between them into account. Meanwhile,
the Bayesian framework is employed to handle the
uncertainty during the perceptual process. From the
meticulous analysis we can gain that the GCLABN is
much more suitable for the perception task than its
counterpart, Bayesian network due to its prominent
features. Yet, no practical test is reported in this paper,
mainly because the prototype system of the GCLABN is
still under development. So, one of our future work is to fully develop the system and make a practical comparison with other existing algorithm to verify the effectiveness and the efficiency of the GCLABN. Additionally, although the GCLABN is designed for the visual perception, it also has great potential to be applied to other areas.

Acknowledgements

This work is supported by the National Science Foundation of China No. 60435010.

References