Visual Awareness in Mind Model CAM

Zhongzhi Shi[1], Jinpeng Yue[1,2], Gang Ma[1,2]
[1]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology
Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
e-mail: { shizz, yuejp, mag}@ics.ict.ac.cn

***Abstract*—Visual awareness is an important function in mind model CAM (Consciousness And Memory). In this paper，we construct a visual awareness component from two respects, namely, objective processing and spatial processing. The Conditional Random fields based Feature Binding computational model (CRFB) is applied to visual objective processing. For visual spatial processing, we explore three important kinds of relationships between objects that can be queried: topology, distance, and direction. The details of object processing and spatial processing are presented.**

Keywords-visual awareness; CAM; visual objective processing; visual spatial processing

I. INTRODUCTION

The visual system is characterized by functional specialization, and each different visual attribute is processed by a different specialized system. Psychophysical experiments have demonstrated that different visual attributes are perceived at different times and independently from each other. Most scientists have the common understanding that primary visual cortex (V1) perceives any visual feature, while higher brain areas may perceive particular visual features. A ventral pathway leading from V1 to the temporal lobe is for representing 'what' objects are. A dorsal pathway leading from V1 to the parietal lobe is for representing 'where' objects are located. The Middle-Temporal (MT) area is for motion perception.

Mental imagery resembles perceptual experience, but occurs in the absence of the appropriate external stimuli. Visual mental imagery is the thought to be caused by the presence of picture-like representations in the mind, soul, or brain. Kosslyn [1] proposed the processing system of visual mental imagery mainly consists of visual buffer, object properties processing, spatial properties processing, associative memories, information shunting and attention shifting.

Inspired by Kosslyn's visual mental imagery, Laird et al. [2] have added memory and processing structures to directly support perception-based representation. The Spatial-Visual Imagery (SVI) focused on modeling the characteristics of human mental imagery [3]. A Spatial Visual System (SVS) combines the concrete spatial representations and abstract symbolic representations [4]. All of these expanded Soar's capabilities [2] with human reasoning.

Object properties processing and spatial properties processing are two important processing issues in visual awareness. The paper will present the main principles for handling visual awareness in CAM (Consciousness And Memory) [5].

In this paper, Section II will outline the architecture of CAM. Object properties processing is discussed in Section III. Sections IV will explore the spatial properties processing. Finally, the conclusions of this paper are drawn and future works are pointed out.

II. CAM ARCHITECTURE

A mind model entitled CAM is proposed by Intelligence Science Laboratory of Institute of Computing Technology, Chinese Academy of Sciences [5]. Comparing with other mind models CAM has several important distinct characteristics, such as unique and sophisticated computational models for perception and cognition; complete memory system including working memory, short-term memory, and long-term memory which has semantic memory, episodic memory and procedural memory; the global workspace and motivation-model based consciousness. The architecture of CAM is illustrated in Figure 1 and organized into ten modules.

A. Visual Module

Visual module is the part of the central nervous system, which gives organisms the ability to process visual detail, as well as enabling the formation of several non-image photo response functions. It detects and interprets information from visible light to build a representation of the surrounding environment. The visual system carries out a number of complex tasks, including the reception of light and the formation of monocular representations; the buildup of a binocular perception from a pair of two dimensional projections; the identification and categorization of visual objects; assessing distances to and between objects; and guiding body movements in relation to visual objects. From Lateral Geniculate Nucleus (LGN) neuron send their signals to the primary visual cortex V1. About 90% of the outputs from the retina project to the LGN and then onward to V1. In the ventral pathway, many signals from V1 travel to ventral extra striate area V2, V3 and V4 and onward to many areas of the temporal lobe.

B. Aural Module

The auditory module is comprised of many stages and pathways that range from ear, to the brainstem, to subcortical nuclei, and to cortex. The advent of neuroimaging techniques has provided a wealth of new data for understanding the cortical auditory system.

 ISBN: 978-1-61208-340-7

C. Sensory Buffers

Each of the classical senses is believed to have a brief storage ability called a sensory buffer.

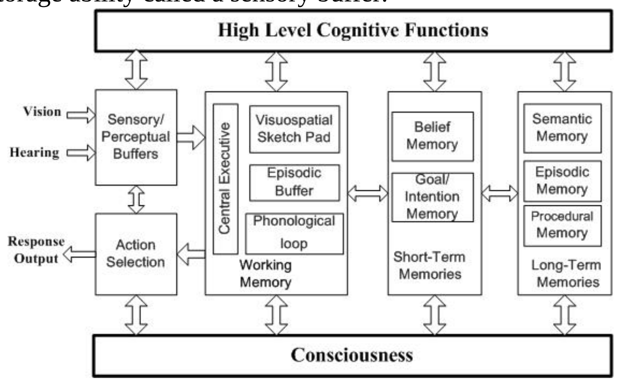

Figure 1. CAM architecture [6]

D. Working Memory

Baddeley [8] presented that working memory includes the central executive, visuo-spatial sketch pad, phonological loop and episodic buffer as illustrated in Figure 1 [7, 8]. The central executive is future directed and goal oriented in effective, flexible and adaptive. At the basic level the working memory is located in the prefrontal cortex. Working memory provides temporary storage and manipulation for language comprehension, reasoning, problem solving, reading, planning, learning and abstraction.

The ability to mentally maintain information in an active and readily accessible state, while concurrently and selectively process new information is one of the greatest accomplishments of the human mind. Working memory provides temporary storage and manipulation for language comprehension, reasoning, problem solving, reading, planning, learning and abstraction.

In working memory, the central executive is the core component. It drives and coordinates other subcomponents in working memory to accomplish cognitive tasks. The visuo-spatial sketch pad holds the visual information about what the cognitive system had seen. The phonological loop deals with the sound or phonological information. The episodic buffer stores the linking information across domains to form integrated units of visual, spatial and verbal information with time sequencing (or chronological ordering), such as the memory of a story or a movie scene. The episodic buffer is also assumed to have links to long-term memory and semantic meaning.

E. Short-Term Memory

Short-term memory stores agent's beliefs, goals and intention contents, which are change rapidly in response to environmental conditions and agent's agenda. Perceptual short-term memory stores the pre-knowledge of objects coded in relational coding scheme and empirical expectations of correlated objects.

F. Long-Term Memory

Long-term memory contains semantic, episodic and procedural knowledge which change gradually or not at all.

1) Semantic memory stores general facts which are represented as ontology. In philosophy, ontology is a theory about the nature of existence. In information science, ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the semantic Web has a taxonomy and a set of inference rules. In CAM, ontology specifies a conceptualization of a domain in terms of concepts, attributes, and relations in the domain. Dynamic Description Logic (DDL) is used to define ontology [9].

2) Episodic memory is one part of long-term memory that involves the recollection of specific events, situations and experiences which are snapshots of working memory. Nuxoll and Laird demonstrated that an episodic memory can support an intelligent agent to own a multitude of cognitive capabilities [10].

In CAM, the episode is an elementary unit that stores previous scene in episodic memory where an episode is divided into two levels: one is an abstract level in terms of logic, another is a primitive level which includes perception information correlated to abstract level of the described object.

3) Procedural memory is a type of long-term memory for the performance of particular types of action. Procedural memory stores knowledge about what to do and when to do it. In ACT-R, 4CAPS, SOAR [2], etc., procedure knowledge is encoded as situation-action rules which provide an efficient and scalable representation. In CAM, procedural knowledge is represented in DDL logic.

G. Action Selection

Action selection is the process of constructing a complex composite action from atomic actions to achieve a specific task. Action selection can be divided into two steps, first is atomic action selection, i.e., select related atomic action from action library. Then, selected atomic actions are composed together using a planning strategy. One of action selection mechanism is based on a spiking basal ganglia model.

H. Response Output

The motor hierarchy begins with general goals, influenced by emotional and motivational input from limbic regions. The primary cortical motor region directly generates muscle based control signals that realize a given internal movement command.

I. Consciousness

The primary focus is on global workspace theory, motivation model, attention, and the executive control system of the mind in CAM. Baars [11] proposed the global workspace theory which all the elements have reasonable brain interpretations, allowing us to generate a set of specific, testable brain hypotheses about consciousness and its many roles in the brain. We presented a new motivation model which is 3-tuples {N, G, I}, where N means needs, G is goal, I means the motivation intensity [12].

J. High Level Cognitive Functions

It includes a class of high level cognitive functions, such as reasoning, planning, learning, etc., which perform cognitive activities based on the basic cognitive functions supported by the memory and consciousness components of CAM.

K. High Level Cognitive Functions

It includes a class of high level cognitive functions, such as reasoning, planning, learning, and etc., which perform cognitive activities based on the basic cognitive functions supported by the memory and consciousness components of CAM.

III. VISUAL OBJECTIVE PROCESSING

Many findings have shown that the ventral pathway acts as an object-properties processing subsystem, whereas the dorsal pathway acts as a spatial-properties processing subsystem. By "object properties" means shape, color, and texture; by "spatial properties" means relative positions in space of two or more objects or parts. Objective processing is to understand what the object is and spatial processing is to know where the object locates. Visual object processing is discussed in this section, whereas next section will explore visual spatial processing.

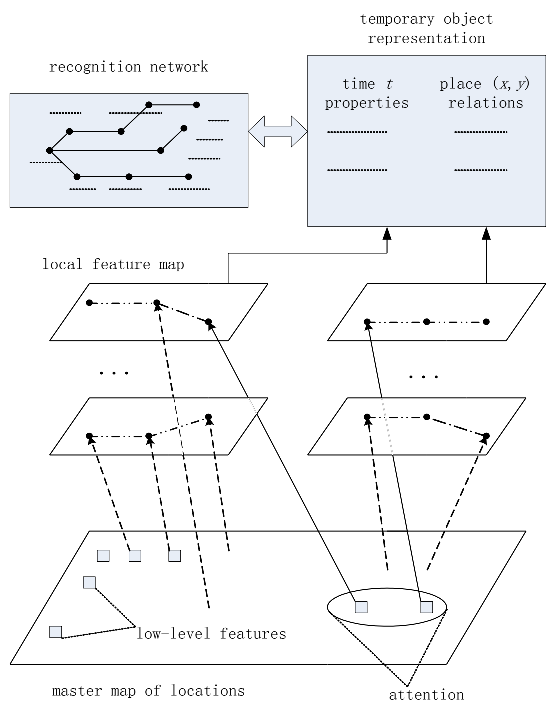

Figure 2. The framework of visual objective processing

We have proposed CRFB in 2010 [13]. Now we apply CRFB for visual object processing. Figure 2 shows the framework of CRFB model. We regard a whole image as a master map of locations [14]. With respect to the precise position tags on the master map, low-level image features are extracted. According to the relational coding schemes, the position tags and low-level image features are locally combined into feature maps [14]. When attentional window [15] scans the master map, the being scanned locations stimulate their corresponding feature maps, forming a temporary object representation, which describe the object in relational coding scheme with no object name. Then, we search the recognition network in perceptual short-term memory, which stores the pre-knowledge of objects coded in relational coding scheme and empirical expectations of correlated objects. When serial scan [15] (for a 2-D image, horizontal and vertical scans are sufficient) of the master map finished, we accomplish the binding process. In the whole process, the fundamental concepts and principles of random fields enlighten us a lot to model the binding problem.

Following the framework, we detail the image features extraction in subsection A. Next, we stipulate the relational coding schemes in subsection B. In subsection C, we learn the recognition network according to the maximum entropy principle. At last, we clarify visual conjunction search process in subsection D.

A. Low-level Feature Extraction

Feature extraction constructs combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Low-level image features are cornerstones for feature binding process. It is desired that we could extract the feature of single object one by one, but to our disappointment, up till now, there are no methods could perfectly sketch the contour of single object among interacted ones. So, we synthesize the latest effective methods to extract image features to represent the low-level visual information to our best.

We present a Coding and Combing Feature framework (CCF) in multi-scale space. We first partition an image into square grids with the same size. For each grid, we compute complement features, i.e., the gradient texture histogram, color histogram and normalized intensity histogram. After various features are extracted, we compact them into efficient and effective codes. The coding process preserves as much information as possible and repesents features with effecive codes. The codes generated in the compact coding step are then combined by multi-kernel hashing to make a more discriminative feature representation. The final representation is effective in different situations. Then, we enlarge the size of the grid and obtain the features for the larger grid as the former procedure and we regard the feature of the larger grid as regional feature. At last, we normalize the local features for an image into a vector. For many images, vectors are clustered and each image is assigned an index of cluster center. We call the index global feature of an image.

The coding and combining framework is evaluated on the UKbench image retrieval dataset [16]. From Figure 3, we can see that the combined feature achieves a higher recall than the color histogram and the texture feature. The combined feature utilizes complementary features and benefit from compact coding; thus, the features generated by our CCF framework is more discriminative and robust.

The multi-scale feature is proposed for the following two reasons: a) It is intractable to represent the feature of an object for difficult outlining objects. Features from different scales could compensate the above problem on some extent; b) Multi-scale feature enrich our association function in relational coding scheme, which we will see in subsection B.

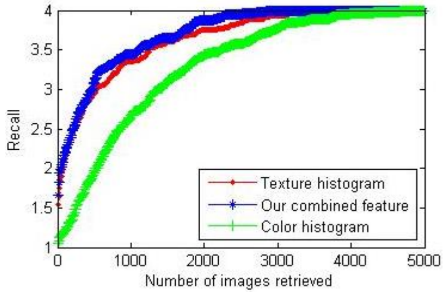

Figure 3. Combined feature evaluation

B. Relational Coding Scheme

We stipulate the relational coding scheme for three functions. First, we lock the low-level features to the locations of the master map. Secondly, correlations between low-level features and high-level knowledge are coded. Thirdly, transfer relations among knowledge are coded. These functions also construct the local feature map. For the first function, we build a 2-D coordinates for the image, then, assign each local grid a precise position (x, y), so we can map the location to the corresponding low-level feature.

We implement the second function under the following considerations: *a*) a longer coding scheme is expected to manage the unstructured image information; *b*) the coding length is restricted by the attentional window. So, we make a compromise to design a reasonable length of the coding scheme. We introduce the association function, inspired by the state feature function [17], to encode the correlations. Let o be an observed low-level local feature, s_t be a state s at time t and l be a name of some object. Association function is defined as follows,

$$f_i(o, s_t) = \delta\left(x_i(o,t)\right)\delta\left(s_t = l\right) \tag{1}$$

where $s_t=l$ means that current state s_t is associate with the object name l. $x_i(o,t)$ is a logical function to judge whether or not a specific serial of low-level features. $\delta(e)$ is equal to 1 if the logical expression e is true, and 0 otherwise. We design the logical function $x_i(o,t)$ within the attentional window. Figure 2 shows the procedure.

In Figure 4, the attentional window size is $2n+1$, so the maximum length of our logical function is n+1. The arrow indicates the current state. o represents the low-level local feature, r represents the combination of low-level regional and global feature and l represents the high-level knowledge. For a current state, there are some fixed schemes to construct $x_i(o,t)$, such as o_0, $o_{-1}o_0$, $r_{-1}r_0$, $r_{-1}r_0o_0$ and so on. We can see that multi-scale low-level features are engaged to construct $x_i(o,t)$, which gives more feature combination schemes for logical function than with the low-level local features only. For a 2-D image, adjacent low-level features are correlated by the logical function. We associate the current observed low-level feature with its horizontal and vertical neighbors within length n to construct $x_i(o, t)$. The 2-D structural correlation coding scheme could better reflect the image semantics.

We introduce the interaction function, similar to the edge feature function [16], to realize the third function. Interaction function is defined as follows,

$$g_j(s_{t-1}, s_t) = \delta\left(s_{t-1} = l'\right)\delta\left(s_t = l\right) \tag{2}$$

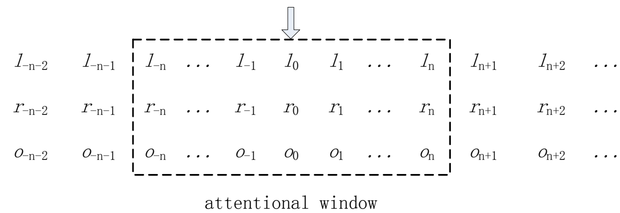

Figure 4. Designing the logical function $x_i(o,t)$.

We introduce the interaction function, similar to the edge feature function [17], to realize the third function. Interaction function is defined as follows,

$$g_j(s_{t-1}, s_t) = \delta\left(s_{t-1} = l'\right)\delta\left(s_t = l\right) \tag{2}$$

This formula suggests that the interaction function presents the object name transfers from previous state to current state, which indeed indicates the transfer scheme of high-level knowledge. Similar to the association function, the knowledge transfer scheme also employed a 2-D structure, coding the relational knowledge horizontally and vertically.

We assemble the association and interaction functions as relation functions, which construct the local feature map. With the ensemble of local feature maps, we can represent the relationships among low-level features and high-level knowledge.

C. Learning the Recognition Network

Relational Coding associate the local feature maps with the master map of locations. At the same time, abundant chaotic correlations between low-level feature and high-level knowledge are supplied. How do we organize the association functions and the interaction functions to gain maximal empirical pre-knowledge for recognizing object?

We need to build a probability distribution p as general as possible to give a maximum entropy of the relation functions which contains the maximal quantity of information [18]. We apply certain expectation constraints with respect to feature functions:

$$E\left(f_i(o, s_t)\right) = \alpha_i \tag{3}$$

$$E\left(g_j(s_{t-1}, s_t)\right) = \beta_j \tag{4}$$

For the association function $f_i(o,s_t)$, if certain serial of observed low-level features associate with some object name, it equals 1, otherwise 0; for the interaction function $g_j(s_{t-1},s_t)$,if the previous object transfers to the current one, it equals 1, otherwise 0. In enormous observing samples, we may set the expectation value $\alpha_i = \infty, \beta_j = \infty$. We assume p_x an element of finite length vector P. Now we may formalize our target as a convex optimization problem subject to linear constraints:

$$\max_{\mathrm{p}} \quad -\sum_{x} p_x \log p_x \tag{5}$$

subject to

$$\sum_{x} p_x f_i\left(o, s_t\right) = \alpha_i \tag{6}$$

$$\sum_{x} p_x g_j\left(s_{t-1}, s_t\right) = \beta_j \tag{7}$$

$$\sum_{x} p_x = 1 \tag{8}$$

We write the Lagrangian as

$$\begin{aligned} J = & -\sum_{x} p_x \log p_x + \sum_{i} \lambda_i \left(\sum_{x} p_x f_i\left(o, s_t\right) - \alpha_i \right) \\ & + \sum_{j} \mu_j \left(\sum_{x} p_x \left(s_{t-1}, s_t\right) - \beta_j \right) + \theta \left(\sum_{x} p_x - 1 \right) \end{aligned} \tag{9}$$

Taking derivatives with respect to a specific element p_x

$$\frac{\partial J}{\partial p_x} = -1 - \log p_x + \sum_{i} \lambda_i f_i\left(o, s_t\right) + \sum_{j} \mu_j g_j\left(s_{t-1}, s_t\right) + \theta = 0 \tag{10}$$

Hence

$$p_x = \frac{1}{Z} \exp\left(\sum_{i} \lambda_i f_i\left(o, s_t\right) + \sum_{j} \mu_j g_j\left(s_{t-1}, s_t\right) \right) \tag{11}$$

That is the distribution we are pursuing. In learning process, we estimate the weight parameters of the relation functions through the K training image samples by maximum the log-likelihood of the objective function [18]:

$$L\ \left(\lambda, \mu\right) = \sum\nolimits_{k} \log p_x\left(\mathbf{s_k}, \mathbf{o_k}\right) \tag{12}$$

We define

$$F\left(\mathbf{s_k}, \mathbf{o_k}\right) = \sum_{i} \lambda_i f_i\left(o, s_t\right) + \sum_{j} \mu_j g_j\left(s_{t-1}, s_t\right) \tag{13}$$

Then, we seek the zero of gradient

$$\nabla L\ \left(\lambda, \mu\right) = \sum\nolimits_{k} \left[F\left(\mathbf{s_k}, \mathbf{o_k}\right) - E_{p_x(\mathbf{S}|\mathbf{o_k})} F\left(\mathbf{S}, \mathbf{o_k}\right) \right] \tag{14}$$

The expectation $E_{p_x(\mathbf{S}|\mathbf{o_k})} F\left(\mathbf{S}, \mathbf{o_k}\right)$ can be efficiently computed with the forward-backward algorithm in [17]. For the 2-D structural of our relation functions, the transition matrix differs [17] in the following formula

$$M_t\left(l', l\right) = \exp\left(\sum\nolimits_{m} \lambda_m f_m\left(s_t, o\right) + \sum\nolimits_{n} \mu_n g_n\left(s_{t-1}, s_t\right) \right) \tag{15}$$

which means that the current matrix $M_t(l', l)$ sums up all the association functions and interaction functions of current state s_t in both horizontal and vertical direction. We penalize the likelihood with a spherical Gaussian weight prior to avoid overfitting [20]. We input the gradient of the log-likelihood to the L-BFGS [21] algorithm for an iteration process which gives the values of parameters $(\lambda_1, \lambda_2, \cdots \mu_1, \mu_2, \cdots)$. With the weighed relation functions, we build the recognition network. Suppose we have learnt $|L|$ categories of object, we can build a $|L| \times |L|$ knowledge transfer matrix by summing up the weighted interaction functions across the overall learning samples where $g_j(s_{t-1}, s_t)$ satisfies $s_{t-1} = l'$ and $s_t = l$.

$$M_i\left(l', l\right) = \exp\left(\sum\nolimits_{j} \mu_j g_j\left(s_{t-1}, s_t\right) \right) \tag{16}$$

The transfer matrix implies the empirical knowledge that on what extent object *B* co-occurs with the object *A*. The association functions with weights, which indicate how much low-level features relate to a specific object, are stored in the recognition network, ready for retrieval.

D. Conjunction Search

The binding process is driven by both attention and particular expectations. Attention works through the attentional window scanning the image and expectations mean predicting the presence of a particular object by contextual constraints, which can be represented by the knowledge transfer matrix. With the preparations in the former sections, we clarify the conjunction search below.

When a new image comes, low-level features are first extracted in a set $S = \{o_1, o_2, \cdots r_1, r_2 \cdots\}$. Then, local feature maps are generated in relational coding scheme. Here, our local feature maps only code the location information and the related serial of low-level features. Location information is coded in a map where each local feature o_{i} in S corresponds to an only precise position coordinate $\left(x_{\mathrm{i}}, y_{\mathrm{i}}\right)$, formulated as $o_{\mathrm{i}} \leftrightarrow \left(x_{\mathrm{i}}, y_{\mathrm{i}}\right)$. Positional adjacent features are coded in the form: $\mathbf{c} = o_1 \cdots o_{\mathrm{s}} r_1 \cdots r_{\mathrm{k}} (s, k \le n+1)$. Next, at time t, the attentional window whose maximum size is $2n$+1, scans the master map of locations with respect to both horizontal and vertical directions. Low-level features within attentional window are cared, and the outside features are temporarily neglected. The relevant feature maps are activated by the window to form a temporary object representation, which contains the location of low-level features and the low-level features combination schemes set $C = \{c_1, c_2, \cdots\}$,of the current state s_t. Now, we search the recognition network to find the association functions which contain the element of set ***C***. Related weighted association functions containing the same object name l are added up. Looking up to the transfer matrix, we obtain the transfer probability from every other object l' to l. We pick the maximal exponential sum of the two factors to be the object l associated with state s_{t} as formula (17).

$$\max \left\{ \underbrace{\exp\left(\sum\nolimits_{i} \lambda_i f_i\left(s_t, o\right) + \sum\nolimits_{j} u_j g_j\left(s_{t-1}, s_t\right) \right), \ldots}_{\text{total number } |L|} \right\} \tag{17}$$

As the attentional window moving forward step by step, the procedure we discussed above is repeated. With regards to the previous states, we get a recursive expression as follows

$$\varphi_{t+1}\left(s_i\right) = \max\nolimits_{s_j} \left(\varphi_t\left(s_j\right) \exp F\left(\mathbf{s}, \mathbf{o}, t+1\right) \right) \tag{18}$$

where $\varphi_{\mathrm{t}}\left(s_{\mathrm{i}}\right)$ denotes the probability that state s_{i} associate with a particular object l. When attentional window scans over the whole image, we obtain the conjunction search result as follows

$$\mathbf{s}^* = \arg\max\nolimits_{\mathbf{s}} \exp\left(\sum\nolimits_{t} F\left(\mathbf{s}, \mathbf{o}, t\right) \right) \tag{19}$$

 ISBN: 978-1-61208-340-7

We apply Viterbi algorithm to the above process to perform an efficient computing. When conjunction search finished, we bind the features to get an integrated understanding of the image.

Conjunction search procedure can be accelerated by the hint of feature inhibition [15], which means that the active features suppress the non-target features. In the low-level local features, adjacent features share similar characteristics causing identical local feature maps. We define the highly similar local features as non-target features. In the conjunction search process, when we come across non-target features, we ignore them and move the attentional window to the next target features. By doing so, we obtain the same binding result but compute less.

Conjunction search associates the every element of low-level local feature set *S* with object name *l*. Looking up to the map of relational coding scheme, using $o_i \leftrightarrow (x_i, y_i)$, we can locate the objects in one image. Up till now, the recognition task is accomplished.

IV. VISUAL SPATIAL PROCESSING

Visuo-spatial sketch pad holds the information it gathers during the initial processing and often in normal thought processes with visualization and conscious effort. Logie has proposed that the visuo-spatial sketch pad can be subdivided into two components [22]: a) The visual cache, which stores information about form and color. b) The inner scribe deals with spatial and movement information.

In order to do visual spatial processing, we follow SVS idea which is proposed by Wintermute [4]. CAM adds a quantitative spatial representation in the spatial scene short-term memory and a visual depictive presentation in the visual buffer short-term memory. In addition to the two short-term memories, there is a long-term memory in CAM for visual, spatial, and motion data and it is called Perceptual LTM. Visual imagery is cognitively useful and can be implemented without true perception. Predicate extraction provides symbolic processing with qualitative properties of the contents of the spatial scene and visual buffer.

For the spatial system, there are three important kinds of relationships between objects that can be queried: topology, distance and direction. Topological relationships describe how the surfaces of objects relate to one another. Distance queries are similarly simple. Currently, the system can query for the distance between any two objects in the scene along the closest line connecting them. Direction queries are implemented as in the approach of Hernández [23].

V. CONCLUSIONS AND FUTURE WORKS

Visual awareness is an important function in mind model CAM. This paper has presented to apply the conditional random fields based feature binding computational model (CRFB) for visual objective processing. We have also shown the main idea for visual spatial processing in the paper.

Visual spatial processing is more difficult and we will continue to look for good representation and algorithm to solve the problem.

ACKNOWLEDGMENT

This work is supported by the National Program on Key Basic Research Project (973) (No. 2013CB329502), National Natural Science Foundation of China (No. 61035003, 61202212, 61072085), National High-tech R&D Program of China (863 Program) (No.2012AA011003), National Science and Technology Support Program(2012BA107B02).

REFERENCES

[1] S. M. Kosslyn, “Mental Images and the Brain.”, Cognitive Neuropsychology, 2005, 22 (3/4): pp. 333–347.

[2] J. E. Laird, “The Soar Cognitive Architecture.” MIT Press, 2012.

[3] S. D. Lathrop, “Extending Cognitive Architectures with Spatial and Visual Imagery Mechanisms.” Ph.D. dissertation, University of Michigan, 2008.

[4] S. Wintermute, “An Architecture for General Spatial Reasoning.” Ph.D. Dissertation. University of Michigan, 2009.

[5] Z. Shi, “On Intelligence Science and Recent Progress.” Invited Speaker, ICCI2006, Beijing, 2006.

[6] Z. Shi, J. Yue, J. Zhang, “Mind Modeling In Intelligence Science.” IJCAI WIS-2013, 30-35, 2013.

[7] A. D. Baddeley, “Working Memory.” Oxford: Oxford University Press, 1986.

[8] A. D. Baddeley, “The episodic buffer: a new component of working memory?” Trends Cogn. Sci. (Regul. Ed.), 2000, 4 (11): pp. 417–423.

[9] L. Chang, Z. Shi, T. Gu, and L. Zhao, “A Family of Dynamic Description Logics for Representing and Reasoning About Actions”, J. Autom. Reasoning, 2012, 49(1): pp. 1-52.

[10] A. Nuxoll and J. E. Larid, “Extending Cognitive Architecture with Episodic Memory”, In Proceedings of the 22nd AAAI Conference on Artificial Intelligence, 2007, pp. 1560-1564.

[11] B. J. Baars, “In the Theatre of Consciousness: Global Workspace Theory, A Rigorous Scientific Theory of Consciousness”. Journal of Consciousness Studies 4, 1997, 4: pp. 292-309.

[12] Z. Shi, J. Zhang, J. Yue, and B. Qi, “A Motivational System For Mind Model CAM”, AAAI Symposium on Integrated Cognition, pp. 79-86, Virginia, USA, 2013.

[13] X. Wang, X. Liu, Z. Shi, and H. Sui, “A feature binding computational model for multi-class object categorization and recognition.” Neural Computing and Applications, 2012, 21(6): pp. 1297-1306.

[14] A. Treisman and G. Gelede, “A feature-integration theory of attention”. Cognit Psychol, 1980, 12: pp. 97-136.

[15] A. Treisman, “Feature binding, attention and object perception” Phil. Trans. R. Soc. Lond. B, 1998, 353, pp. 1295-1306.

[16] D. Nister and H. Stewenius, “Scalable Recognition with a Vocabulary Tree”. In CVPR, 2006, pp. 2161-2168.

[17] J. Lafferty, A. McCallum, and F. Pereira. “Conditional random fields: Probabilistic models for segmenting and labelling sequence data”, In Proceedings of the Eighteenth International Conference on Machine Learning (ICML), 2001, pp. 282–289.

[18] H. Wallach, “Efficient Training of Conditional Random Fields”, M.Sc. thesis, Division of Informatics, University of Edinburgh, 2002.

[19] F. Sha and F. Pereira. “Shallow Parsing with Conditional Random Fields”. Proceedings of HLT- NAACL 2003 213-220N.

[20] S. F. Chen and R. Rosenfeld. “A Gaussian prior for smoothing maximum entropy models”. Technical Report CMU-CS-99-108, Carnegie Mellon University, 1999.

[21] R. H. Byrd, J. Nocedal, and R.B. Schnabel. “Representations of quasi-newton matrices and their use in limited memory methods”, Mathematical Programming, 1994, 63:129–156.

[22] R. H. Logie, “Visuo-Spatial Working Memory.” Lawrence Brlbaum Associates Ltd., Publishers, East Sussex, UK, 1995.

[23] D. Hernández, "Qualitative Representation of Spatial Knowledge ." Springer-Verlag, 1994 .