

Introducing Memory and Association Mechanism Into a Biologically Inspired Visual Model

Hong Qiao, *Senior Member, IEEE*, Yinlin Li, *Student Member, IEEE*, Tang Tang, *Student Member, IEEE* and Peng Wang,

Abstract—A famous biologically inspired hierarchical model firstly proposed by Riesenhuber and Poggio has been successfully applied to multiple visual recognition tasks. The model is able to achieve a set of position- and scale-tolerant recognition, which is a central problem in pattern recognition.

In this paper, based on some other biological experimental results, we introduce the Memory and Association Mechanisms into the above biologically inspired model. The main motivations of the work are (a) to mimic the active memory and association mechanism and add the 'top down' adjustment to the above biologically inspired hierarchical model and (b) to build up an algorithm which can save the space and keep a good recognition performance.

More details of the work could be explained as follows:

- (1) **In objects memorizing process:** Our proposed model mimics some characteristics of human's memory mechanism as follows:
 - (a) In our model, one object is memorized by semantic attributes and special image patches (corresponding to episodic memory). The semantic attributes describe each part of the object with clear physical meaning, for example, if eyes and mouths of faces are 'big' or 'small' and so on. One special patch is selected if the value of the corresponding semantic feature is far from average one. The patch should be the most prominent part of the object.
 - (b) In our model, different features (semantic attributes and special patches) of one object are stored in distributed places and the common feature of different objects is saved aggregately, which can learn to classify the difference of similar features of different objects. The similarity thresholds to each object can be learnt when new objects are learnt.
- (2) **In object recognition process:** In biological process, the associated recognition including familiarity discrimination and recollective matching. In our proposed model, firstly mimicking familiarity discrimination ('knowing' though the episode), we compare the special patches of candidates with that of saved objects using above mentioned biologically inspired hierarchical model, where the candidates and saved objects have the same prominent semantic features. Then mimicking recollective matching, the comparison results of special patches are combined with semantic feature comparison.

The new model is also applied to object recognition processes. The primary experimental results show that our method is efficient with much less memory requirement.

Index Terms—Memory, Association, object recognition, biologically inspired visual model.

Manuscript received July 3, 2013; revised.

H. Qiao is with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: hong.qiao@ia.ac.cn).

Y.L. Li and T. Tang are with the Graduate School and the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences.

I. INTRODUCTION

MANY researchers have established a series of neural computational models for vision processes based on the biological mechanism and showed that the models can be well applied to pattern recognition [1], [2], [3], [4]. More importantly, recently, biological research with information technology integration is an important trend of research in all over the world.

In particular, in 1999, Riesenhuber and Poggio proposed a famous neural computational model for vision process [2]. This model is a hierarchical feed-forward model of the ventral stream of primate visual cortex, which is briefly called as HMAX. The model is closely related to biological results. Each level in the model was designed according to data of anatomy and physiology experiments, which mimicked the architecture and response characteristics of the visual cortex.

Furthermore, Giese and Poggio have extended the above model to biological motion recognition [5]. They established a hierarchical neural model with two parallel process streams, i.e. form pathway and motion pathway. A series of extended models have given very good performance in biological motion recognition in cluster [6] and are compared with psychophysics results [7]. The model was also extended to a computational model for general object recognition tasks [8], which can output a set of position and scale invariance features by alternating between a template matching and a maximum pooling operation corresponding to S1-C2 layers. Some other researchers introduced sparsification, feedback and lateral inhibition [9], [10] into HMAX. The series works have demonstrated very good performance in a range of recognition tasks which is competitive with the state of art approaches, such as face recognition [11], scene classification [12] and handwritten digit recognition [13].

From another point of view, Itti established a visual attention model [3], [14], which is inspired by the behavioral and neural architecture of the early primate visual system. Itti and Poggio et al [15] merged the saliency-based attention model proposed in the previous work [3] to HMAX in order to modify the activity of the S2 layer. Here the S2 layer mimics the response characters of V4 in the primate visual cortex, which shows an attention modulation in electrophysiology and psychophysics experiments.

Saliency based visual attention models are further compared with behavioral experiments [16] and showed that the models could account for a significant portion of human gaze behavior in a naturalistic, interactive setting. A series of visual attention

models [17] have demonstrated successful applications in computer vision [18], mobile robotics [19], and cognitive systems [20].

Some other work has tried to combine HMAX and Deep Belief Networks. The HMAX model could extract position and scale invariance features with a feed-forward hierarchical architecture, but the lack of feedback limits its performance in pure classification tasks. Deep Belief Networks (DBN) [21] has shown the state of the art performance in a range of recognition tasks [22], [23]. DBN uses generative learning algorithms, utilizes feedback at all levels and provides ability to reconstruct complete or sample sensory data, but DBN lacks the position and scale invariance in the HMAX. Therefore, combining DBN with the HMAX may be meaningful to extract more accurate features for pattern recognition.

The above models are all biologically inspired models and have been successfully applied in practical work. In this paper, based on other biological results, we try to introduce the Memory and Association into the HMAX. Our method can reduce memory requirement and achieve comparable accuracy to the state of art approaches.

II. RELATED WORKS

In this paper, we propose a model based on HMAX and some basic facts about the memory and association mechanism established over the last decade by several physiological studies of cortex [24], [25], [26], [27]. The accumulated evidences points are summarized as follows.

(1). *The Memory About an Object Includes the Episodic and Semantic Memory*

Tulving et al. proposed that episodic memory and semantic memory are two subsystems of declarative memory [27]. Though the two systems share many features, semantic memory is considered as the basis of episodic memory which has additional capabilities that semantic memory does not have. In the memory encoding process, information is encoded into semantic memory first and can then be encoded into episodic memory through semantic memory, as shown in Fig. 1.

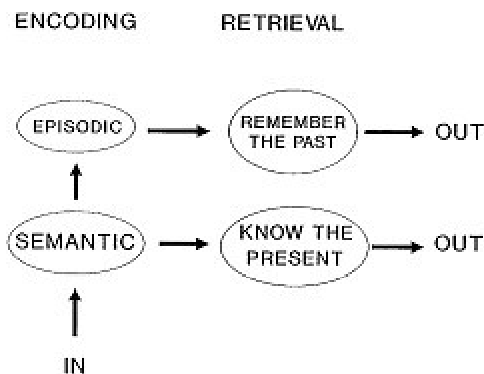


Fig. 1: Sketch of the relationship between semantic and episodic memory [24]

(2). *An Object Concept May be Represented by Discrete Cortical Regions*

Brain imaging studies show that object representation is related to different cortical regions, forming a distributed network that is parallel to the organization of sensory and motor systems [25]. For example, in the word generation experiments, ventral and lateral regions of the posterior temporal cortex have been found eliciting differentially with different types of information retrieved [28]. In other related behavioral experiments in which the subjects are required to name an action or a color, the specialization of different cortical regions has also been observed [28].

(3). *Neurons Responding to the Common Feature of Different Objects Aggregate Together*

Researchers believe that the brain tends to process common features from different objects in the same areas [25]. One supporting fact for the assertion is the existence of the cortical region responding to the shape attributes of different visual categories. In the related behavioral experiments, when the subject was required to perform same task with different object categories as stimuli, the ventral occipitotemporal cortex was consistently activated and encoded the object forms with distinct neural response patterns [25].

Another body of evidence comes from studies of cortical lesions [29], [30]. The damage to temporal lobes is found to be strongly associated with the impairment of abilities to recognize objects and retrieve information about object-specific characteristics [29], [30]. This implies that the temporal lobe is the region in which object-specific information from all categories is commonly stored.

Functional column may provide the anatomical basis accounting for the above phenomena, and recent studies suggest that the functional column is a basic processing architecture spreading over visual neural systems [31], [32], [33]. On the surface of infratemporal cortex that activated by faces, neurons with common preference were found aggregating in patches specifying different stimuli [31].

Similar organization was also seen in the area of posterior TE that is responsive to simple two-dimensional shapes [32]. It is inferred that such a columnar organization would produce stimulus invariance [34] and also provides a visual alphabet from which numerous complex objects can be efficiently represented as a distributed code [33].

(4). *Recognition Memory Includes Familiarity and Recollection Components*

As a basic functional part of memory, recognition memory has its use in identifying and judging whether the presented object has ever been captured consciously [35], [26]. It is widely accepted that two components make up the recognition memory: One is the familiarity discrimination that determines 'knowing' though the episode in which the object has appeared may not be recalled, and the other is the recollective matching that means

remembering both the presence of the object and the related episode.

Different mechanisms have been proposed to explain the two recognition experiences. Some researchers argue that the only difference between familiarity discrimination and recollection matching is the trace strength, in term of which the former is weaker than the latter [36], [37], [38], while others regard the two processes as qualitatively different modes of recognition memory, which may be executed by hippocampus and perirhinal cortex, respectively [35], [26].

(5). *Familiarity Recognition is Rapid and Accurate and Only Needs a Small Number of Neurons*

From the evolutionary perspective, owning a simplified recognition system focusing on familiarity is likely to have advantage in speed of response to novelty by saving time of deep recalling. This hypothesis has been confirmed in some related experiments, in which subjects made familiarity decisions faster than recollect decisions [39], [40], [27].

The computational efficiency of a recognition system dedicated for familiarity discrimination has also been demonstrated by the simulated neural networks [41]. Compared with systems relying on associative learning, familiarity decision can be made accurate and faster by a specially designed network with smaller size of neuron population.

All of above works form the basis of our model which would be explained in Section III.

III. THE NEW MODEL-INTRODUCTION MEMORY AND ASSOCIATION INTO BIOLOGICALLY INSPIRED MODEL

The new model is presented in Fig. 2. This section includes 5 parts, which introduce the framework of our model and algorithms related to sub processes.

A. The Framework of Our Model

Figure III1 presents our model which introduces biologically inspired memory and association model into HMAX. Our framework consists of 5parts, i.e., block 1 to block 5.

1) **Objects are Memorized Through Episodic and Semantic Features (Block1)**

O_i represents the i th object ($i = 1, \dots, n$), a_{ji} represents the j th semantic attribute ($j = 1, \dots, m$) of the i th object, and s_{xi} represents the x th special episodic patch of the i th object. In face recognition process, semantic description can be 'the eyes are large', 'the month is small' and so on. The special episodic patch can be the image of an 'eye' if the eyes are prominent on the face. These two kinds of features are just corresponding to the semantic and episodic memory of declarative memory in cognitive science [24]

2) **Features of One Object are Saved in Distributed Memory and of and Common Features of Various Objects Are Stored Aggregately (Block2)**

A_1 to A_m represent various common semantic attributes of different objects. S_x is a library for special episodic patches of different objects. A common attribute of different objects is stored in the same regions, for example, $a_{11} - a_{1n}$ is the first feature of different objects ($i = 1, \dots, n$) which are stored together, called as A_1 . A_1 not only has clear biological area but also has learning ability. For example, the similarity sensitivity would be learnt when new objects are saved. This memory mechanism has clear biological evidences [34], [25]

3) **Recognition of One Candidate (Block3)**

T_t represents a candidate. The semantic features a_{1t} to a_{mt} and episodic patch feature s_{xt} are extracted before recognition.

4) **Familiarity Discrimination (Block4)**

Familiarity discrimination is achieved through the HMAX model. Both S_x and s_{xt} are extracted in the C1 layer of HMAX model. The saved object can be ordered by comparing similarity between the candidate and the saved objects in the C2 layer of HMAX model. This process corresponds to the familiarity discrimination ('knowing' through the episode) in biological recognition memory [35].

5) **Recollective matching (Block5)**

Recollective matching is achieved through integration of semantic feature and episode feature similarity analysis. We compare the semantic features of the candidate with those of the top saved objects according to familiarity discrimination. If the difference between the candidate and the closest object does not exceed the threshold which we have learnt during the memory process, we consider that the candidate 'is' the object. This procedure is illuminated by recollective matching in recognition memory [35].

B. Encoding Episodic and Semantic Features (Block 1)

As we know, memory is the process by which information is encoded, stored, and retrieved, and the memory to an object with conscious process is called declarative memory, which can be divided into semantic memory and episodic memory [24], [42]. Semantic memory refers to the memory which is learned by people, such as meanings, understandings, concept-based knowledge and work-related skills and facts [43]. Episodic memory refers to the memory which is explicitly located in the past involving the specific events and situations, such as times, places, associated emotions and other contextual knowledge [44], [45].

In this paper, the features of a saved object i include descriptive attributes a_{1i} to a_{mi} and special patch s_{xi} (Fig. 3), where descriptive attributes correspond to semantic features, and special patches correspond to episodic features.

1) *Semantic Features*: In biological process, human would memorize an object using semantic features. In computational process, semantic features can reduce the memory size requirement. However, in order to get semantic features, a big dataset and learning ability are needed, which correspond to the prior knowledge.

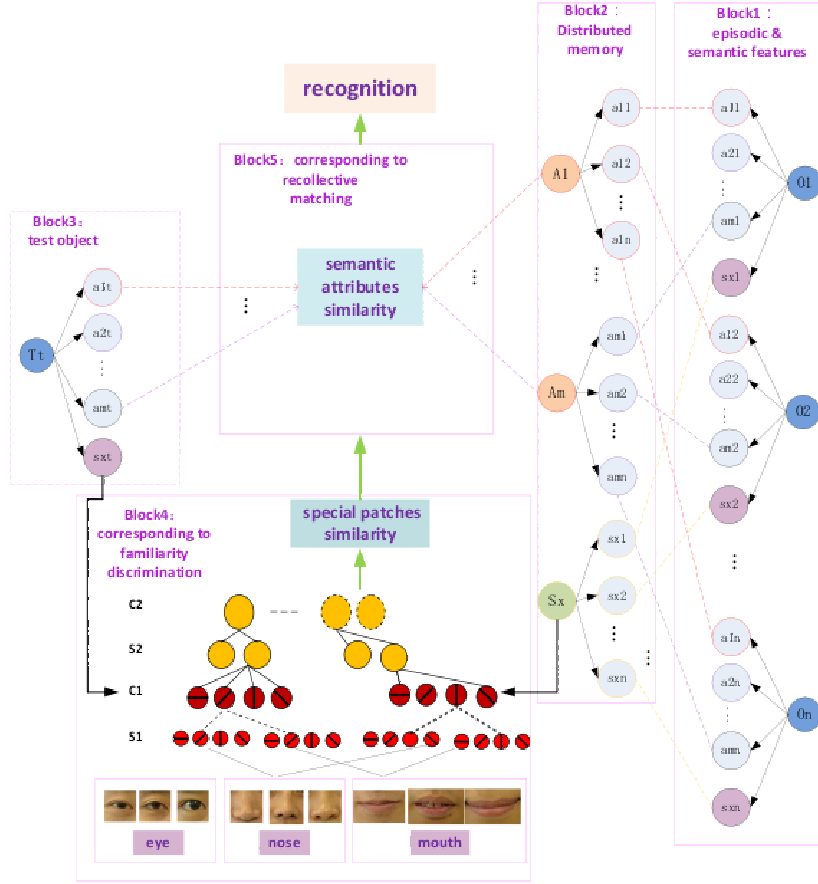


Fig. 2: The model proposed - Introducing biologically inspired memory and association into HMAX model.

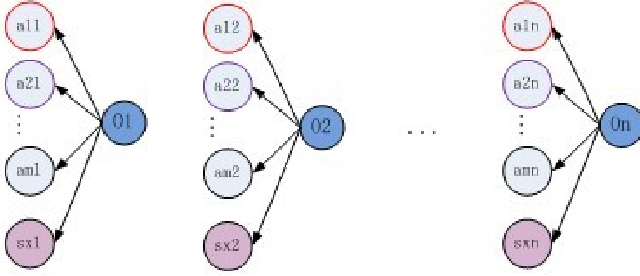


Fig. 3: Semantic attributes and special episodic patches in memory.

A simplified algorithm is presented as follows. First, we need to extract patches g_{ji} with clear physical meaning (such as the images of eyes, months of one face) for each object i . Then, we can get an average view \bar{g}_j of each feature g_{ji} . At last, we compare g_{ji} and \bar{g}_j ($j = 1, \dots, m$) to obtain the semantic attribute a_{ji} of a particular object. The detailed algorithm is given in Algorithm 1.

Take face recognition as an example. In this paper, the active shape model (ASM) is used to extract typical points from faces. Avoiding to describe one part 'big' or 'small', which can be vague, in this paper, we compute effective geometric features of each part related declarative features of eyes, nose

Algorithm 1: Extracting semantic attributes.

- Input:** objects $o_i, i = 1 \dots n$
 features $g_{ji}, j = 1 \dots m$
- 1: Extract geometric features ($j = 1, \dots, m$) from object o_i
 - 2: Compute a common view for each feature
 $\bar{g}_j = f(g_{j1}, g_{j2}, \dots, g_{jn})$
 - 3: Compute semantic attributes $a_{ji} = f(g_{ji}, \bar{g}_j)$
- Output:** Semantic attributes a_{ji} .

and mouth (Table I). We could establish an average value for each geometric feature. By comparing each individual with the average face, we can get the computational semantic attributes for each part.

eyes	eye_area eye_length eye_height eye_ratio(l/h) brow_eye_distances eye_cheek_ratio	nose	nose_width nose_ratio court_ratio	mouth	mouth_length mouth_height mouth_ratio Philtrum_length
------	---	------	---	-------	--

TABLE I: 15 geometric features of a face.

In experiments, we would show that these geometric features have coincidence with general semantic meaning, such as 'the eyes are large' and so on.

It should be noted that representing the geometric features of

a face component should consider the influences of expression, view direction, light and even subjective feeling of an observer. In order to get fair results, our current dataset is collected with frontal faces with less expression, and the geometric attributes are normalized as $g_{ji} = \frac{g_i - u_j}{\sigma_j}$, where u_j is the average of the j th geometric feature and σ_j is the standard deviation of the j th geometric feature.

For one person, many other attributes, such as gender, race and age can be included in further research. Some events and their relationships can enrich the memory processes.

2) *Episodic Patches*: In order to mimic episodic memory, we need to find out the dominant patches of an object through finding out the most prominent semantic features of the object I (Equation 3-2).

Take face recognition as an example again. First, we need to find out the most prominent semantic feature for physical parts, for example, eyes, nose and mouth:

$$J_i = \arg \max_{j=1,2,\dots,m} (\|a_{1i} - \bar{a}_1\|_2, \|a_{2i} - \bar{a}_2\|_2, \dots, \|a_{mi} - \bar{a}_m\|_2) \quad (1)$$

where J_i is the dominant semantic feature of the i th object, a_{ji} is the j th semantic feature of the i th object, and \bar{a}_{ji} is the j th average semantic feature. Special episodic patches could be extracted, which corresponds to J_i of the object O_i and are put into HMAX model proposed by [2], which can extract a set of position and scale tolerant features.

Different from the original HMAX model, only episodic patches corresponding to the prominent part of a candidate are put into the model, and only those of the 'known' objects are stored in a library.

C. Distributed Memory Structures (Block 2)

How to remember and organize different features is a key problem in the memory process. It would also influence the retrieval and association processes. This is also an important part of our framework.

The studies of cognitive sciences have indicated that the extraction, storage and retrieval of different features are realized by the distributed cortical areas. As we have listed in section II, many data and evidence proved that an object concept may be represented by the discrete cortical regions [28] and neurons responding to the common feature of different objects aggregate together [34].

In this paper, we mimic this distributed structure to memorize and retrieve features. Suppose there are m visual cortical regions A_1 to A_m (Fig. 4), which are sensitive to different kinds of features $a_{1,t}$ to $a_{m,t}$. Thus candidate which have feature $a_{i,t}$ will active a distinctive response in region A_i .

This kind of distributed structures has great advantages. The visual cortical region A_i ($i = 1, \dots, m$) can be more and more sensitive and effective through comparing different $a_{i,t}$. Two similarity thresholds for semantic attributes and episodic patches which would decide if a candidate is a 'known' object are learned when a new object is memorized. In the further work, the attributes of one object would also be connected and influence each other.

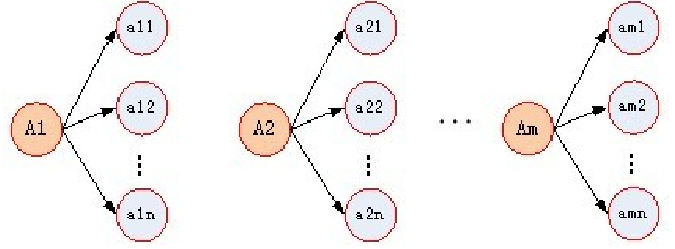


Fig. 4: Distributed memory structure.

D. Recognition based on Familiarity Discrimination and Recollective Matching (Block 3)

As shown in Section II, there are two processes in the recognition memory: familiarity discrimination and recollective matching [35], and their combination is useful for a fast and accurate human recognition task. In the proposed framework, we also have two processes, corresponding to familiarity discrimination and recollective matching.

1) *Familiarity Discrimination*: In block 1, we established a library of episodic patches $\{S_x\}$ of different 'known' objects in the memory process. In the familiarity discrimination of the association process, a special episodic patch s_{xt} of the candidate T_t is extracted and put into the C1 layer of the HMAX model. Then we compute the S2 and C2 level features of the candidate and the 'known' objects which have the same prominent semantic attribute. Finally, we sort the 'known' objects through their C2 feature similarity compared with the candidate.

This process corresponds to the familiarity discrimination in recognition memory [35].

2) *Recollective Matching*: We compare the semantic features $a_{i,t}$ ($i = 1, \dots, m$) of the candidate with those of the top 'known' objects obtained from familiarity discrimination. If the smallest dissimilarity does not exceed the thresholds which we have learnt during memory process, the candidate is 'recognized' as the closest 'known' object. Otherwise, we regard the candidate as a new one. This procedure is illuminated by re-collective matching in recognition memory [35].

The whole algorithm is given as algorithm 2:

IV. EXPERIMENT

In this section, we conduct experiments to demonstrate the effectiveness of our model through introducing biologically inspired memory and association mechanism into HMAX. We first establish a dataset which consists of front faces of 7 persons (Fig. 5). The first 5 persons' face images are used in both the memory and association phases and other 2 persons' face images are only used in the recognition phase. Some of the samples in the dataset are shown in Fig. 6:

The experimental details are given as follows.

A. Special Patches Evaluation

The first person's 8 candidate images are presented in Fig. 7. Responses of the C2 units in HMAX model are used for

Algorithm 2: Memory and retrieval process

memory process : semantic attributes, episodic patch features, dominant attributes, and similarity thresholds of the training objects

Retrieval Process: A candidate

- 1: Extracting semantic attributes, dominant attributes and the episodic patch
- 2: **Familiarity Discrimination:** finding the training objects which have the same prominent attribute j , and using their stored patches library S_j to get the C2 layer features for both f_{test}^{C2} and f_i^{C2} , and sort the candidates by $\|f_{test}^{C2} - f_i^{C2}\|_2$
- 3: **Recollective Matching:** comparing the detailed semantic features of the test object with those of the top candidate training objects

$$similarity_i = \arg \min_{i \in top3} \sum_{j=1}^m (\|a_{jtest} - a_{ji}\|_2)$$

- 4: **Zero-shooting object memorizing:** Store feature set of the test object
-



Fig. 5: 7 persons whose pictures are used in experiments.

the dissimilarity analysis of special patches.

Although light, expression, scales and other factors would influence the similarity analysis, the experimental results show that the similarity of the special patches through HAMX primarily matches the semantic description. For example, one candidate's special patch is more similar to that with the same semantic description, that is, the big eyes have high similarity with big eyes and low similarity with small eyes.

In the memory process, the special patches of the first 3 persons' images are all eyes. The eyes images are scaled with 4 and 6, and the first 2 with 'big' eyes and the third one with 'small' eyes (Fig. 8).

In recognition process, three other pictures of the first 3 persons are used (see Fig. 9).

The similarity analysis and rank are presented in Table II.

Where, $Patch_{ij}$ represents an eye patch with scale $j \times j$ of the i th person. The patches are ranked according to their average values from the C2 layer.

It is not surprising that the eye patches of the first two persons generate higher values with portraits having bigger eyes, while the patches of the third person prefer the smaller ones. This rule always holds even in some cases that the eyes of the first two identities are confused.

The results above demonstrate that the low-level features used in the model are reliable indicators of visual properties.



Fig. 6: The first person's 8 pictures used in the memory process.

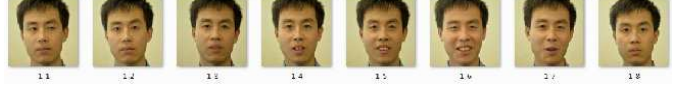


Fig. 7: The first person's 8 candidate images.

By combining a sufficient number of such features, visual objects can be efficiently represented by the response patterns, just like the neural activities in certain cortical regions by which they are encoded.

B. Familiarity Discrimination

The purpose of familiarity discrimination is to 'feel' if 'knowing' the subject. After a special patch similarity analysis, it is necessary to decide if the subject is 'known' rapidly. Then the problem is how to design the threshold of each saved subject which can be used to decide if the candidate is the saved subject.

In this paper, the threshold for a 'known' face image i is given as follows,

$$\begin{aligned} thres_{1i} &= \arg \max_i (\|f_i^q - \bar{f}_i\|_2) \\ thres_{2i} &= \arg \max_{j, j \neq i} (\|f_i^q - \bar{f}_j\|_2) \\ thres &= thres_{1i} + \|thres_{2i} - thres_{1i}\|/\lambda \end{aligned} \quad (2)$$

where f represents the semantic features or C2 features of special patches, f_i^q is the q th image feature of the i th object (i th object has q images), \bar{f}_i is the average features of q images belonging to i th object, $thres_{1i}$ is the intra-class maximum difference, $thres_{2i}$ is the inter-class minimum difference, λ is the ratio of $thres_{1i}$ and $thres_{2i}$ and $thres$ is the final similarity threshold.

Clearly, the larger the threshold is, the more insensitive the image is. $\|thres_{2i} - thres_{1i}\|/\lambda$ is used to adjust the $thres_{1i}$ to make it more flexible.

The thresholds for 5 persons' images are shown in following Table III:

The dissimilarities between 3 other persons' images and 5 'known' persons' images are given in Table IV.

It can be seen that the dissimilarities between three other persons' images and five 'known' persons' images are all larger than the threshold. Therefore the three other persons' can be recognized as 'unknown' rapidly.

However, instead of presenting the result of the familiarity process by 'yes' (the dissimilarity value is smaller than $thres$) and 'no' (the dissimilarity value is larger than $thres$), in this paper, the process result is presented by the probability of one candidate being a 'known' subject that computed by (3)



Fig. 8: The first and second persons have with 'big' eyes and the third one has with small eyes.



Fig. 9: Candidates in Special Patch Evaluation.

$$p(c_i|x) = \frac{p(x|c_i)p(c_i)}{\sum_{i=1}^n p(x|c_i)p(c_i)} \quad (3)$$

As we assume all $p(c_i)$ are the same, the identities are actually sorted by $p(x|c_i) = \exp(-d)$, where d is the Euclidean distance between features of the learned face and the incoming face.

Based on (3), the first five persons' images used for recognition are compared with their images used for memory. The probability and the corresponding ranks are given in Table V (only the results of 4 persons' 3 images are listed here).

It can be seen that each person's other images are close to themselves through familiarity matching. Three 'known' objects among top three matching probability are reserved for further recognition.

C. Recognition by recollection matching

In this process, the similarity in semantic meaning is compared.

In this paper, the controlled points of faces are obtained by the Active Shape Model (ASM) or manually. One sample (68 control points can be obtained from 1632*1632 image) is given in Fig. 10.

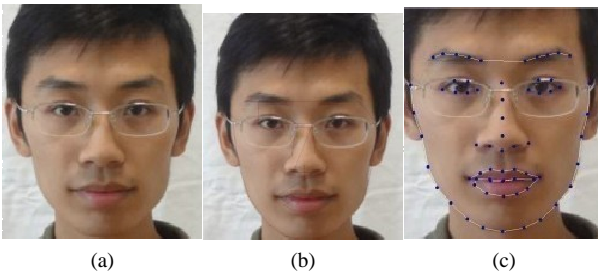


Fig. 10: (a) input image (b) compute controlled points with ASM (c) affine transformation with manual modifying.

The main parts of a face include eyes, nose and mouths. The semantic presentation has 15 attributes as shown in Section

patch 1.4	C2 value	0.0088	0.0096	0.0136
	Ranking	2	1	3
patch 2.4	C2 value	0.0032	0.0113	0.0165
	Ranking	2	1	3
patch 3.4	C2 value	0.0028	0.0088	0.0117
	Ranking	3	2	1
patch 1.6	C2 value	0.0205	0.0300	0.0359
	Ranking	1	2	3
patch 2.6	C2 value	0.0080	0.0380	0.0472
	Ranking	2	1	3
patch 3.6	C2 value	0.0055	0.0285	0.0438
	Ranking	3	2	1

TABLE II: similarity analysis and rank.

	mouth			Eyes	
identity	1	2	4	3	5
patch_thres	0.5406	0.5203	1.5492	0.4174	0.4861

TABLE III: The thresholds for 5 persons' images

III-B1. Therefore, one 'known' face is saved by 15 semantic attributes, 1 prominent attribute which determines the part of special patch, 1 special patch and thresholds (See Table VI).

The attribute thresholds computed by function 4-1 for 5 persons are shown in following Table VII:

By computing the difference between the test image and the three candidate persons' attributes in turn, and compared with the attributes thresholds. We could give the final label for each test image.

The recognition results are listed in Table VIII.

V. CONCLUSION

In this paper, the biologically inspired memory and association model has been proposed. The typical features of the model based on biological work include:

- in both the memory and recognition processes, the objects have semantic and episodic features and episodic patch is extracted according to semantic feature.
- The sensitivities to the semantic and episodic features are learnt during the process.
- In association process, the semantic and episodic features are separately compared and the results are integrated, which correspond to familiarity and cognitive matching.

Through six blocks, the above model is introduced to the HMAX and the corresponding algorithms are given. The memory and association mechanism provides 'top-down' control function in the process, which can active select the important features in memory and association and influence the recognition process.

The experimental results show that with the new method, the required memory is very small and recognition performance is very good.

In the future work, the combined model would be further improved with more learning ability.

ACKNOWLEDGMENT

The authors would gratefully thank Dr. Wei Wu and Dr. Tianshi Chen for their kindly help to improve this paper.

identity	1		2		3		4		5	
attri_thres	4.5946		4.1696		4.0039		4.4526		3.5100	
Test	label	Sim	label	Sim	label	Sim	label	Sim	label	Sim
X_1	1	2.1773	2	2.0454	3	3.2711	4	1.9933	5	2.1834
X_2	1	2.6458	2	1.7975	3	2.9305	4	1.8047	5	2.2325
X_3	1	2.3152	2	2.507	3	2.1742	4	3.8357	5	3.3121
X_4	1	2.332	2	2.0656	3	3.7265	4	3.4721	*	\
X_5	1	3.2519	2	3.9128	3	2.6567	4	1.6642	3	3.5643
X_6	0	\	2	1.5454	3	3.1565	4	2.3439	5	2.5769
X_7	1	3.5434	2	1.8358	3	2.219	4	2.9189	*	\
X_8	1	3.1574	2	2.317	0	\	4	3.1242	*	\

TABLE VIII: Recognition results. '*' indicates that the person is wrongly rejected during familiarity discrimination. '0' indicates that the person is wrongly rejected during recollection matching, and '\ ' means that the similarity is missing since the person is rejected in previous process.

	Mouth			Eyes		
	1	2	4	3	5	
Patch_thres	0.5406	0.5203	1.5492	0.4174	0.4861	
6_1	0.596875	0.923608	1.771417	0.869486	0.75737	
6_2	0.709693	0.870208	1.681255	1.155204	1.031485	
6_3	1.101026	1.380079	2.12309	0.654594	0.655111	
7_1	0.821075	0.94641	1.785884	0.922266	0.917094	
7_2	0.760753	0.780838	1.828294	0.701748	0.822338	
7_3	0.589346	0.888509	2.171345	0.44176	0.785486	

TABLE IV: The dissimilarities between 3 other persons' images and 5 'known' persons' images

1_1	Probability	0.2836	0.2534	0.2207	0.1826	0.0597
	Ranking	1	5	2	3	4
1_2	Probability	0.2975	0.2455	0.2207	0.1783	0.0579
	Ranking	1	5	2	3	4
1_3	Probability	0.3413	0.2559	0.193	0.1508	0.059
	Ranking	1	5	3	2	4
2_1	Probability	0.316	0.2547	0.2202	0.164	0.045
	Ranking	2	1	3	5	4
2_2	Probability	0.3409	0.2527	0.2115	0.1409	0.054
	Ranking	2	1	3	5	4
2_3	Probability	0.3026	0.2474	0.2144	0.1819	0.0537
	Ranking	2	3	1	5	4
3_1	Probability	0.3608	0.2202	0.2038	0.1838	0.0314
	Ranking	3	2	5	1	4
3_2	Probability	0.3407	0.2198	0.2084	0.2023	0.0288
	Ranking	3	2	1	5	4
3_3	Probability	0.3163	0.2341	0.2107	0.2099	0.029
	Ranking	3	2	1	5	4
4_1	Probability	0.4885	0.1476	0.1444	0.1238	0.0958
	Ranking	4	3	2	1	5
4_2	Probability	0.4832	0.1511	0.1468	0.1248	0.0941
	Ranking	4	2	3	1	5
4_3	Probability	0.4698	0.1595	0.1507	0.1316	0.0884
	Ranking	4	1	3	2	5

TABLE V: The probability and possibility rank of first 4 people' other images compared with their 'known' images through special patches

REFERENCES

- [1] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in cats visual cortex," *Journal of Physiology London*, vol. 160, no. 1, p. 106, 1962.
- [2] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.
- [3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [4] N. C. Rust, V. Mante, E. P. Simoncelli, and J. A. Movshon, "How mt cells analyze the motion of visual patterns," *Nature Neuroscience*, vol. 9, no. 11, pp. 1421–1431, 2006.
- [5] M. A. Giese and T. Poggio, "Neural mechanisms for the recognition of biological movements," *Nature Reviews Neuroscience*, vol. 4, no. 3, pp. 179–192, 2003.
- [6] R. Sigala, T. Serre, T. Poggio, and M. Giese, "Learning features of intermediate complexity for the recognition of biological motion," in *Artificial Neural Networks: Biological Inspirations–ICANN 2005*. Springer, 2005, pp. 241–246.
- [7] A. Casile and M. A. Giese, "Critical features for the recognition of biological motion," *Journal of Vision*, vol. 5, no. 4, pp. 348–360, 2005.
- [8] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.
- [9] Y. Huang, K. Huang, D. Tao, T. Tan, and X. Li, "Enhanced biologically inspired model for object recognition," *Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics*, vol. 41, no. 6, pp. 1668–1680, 2011.
- [10] J. Mutch and D. G. Lowe, "Multiclass object recognition with sparse, localized features," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 11–18.
- [11] E. Meyers and L. Wolf, "Using biologically inspired features for face processing," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 93–104, 2008.
- [12] K. Huang, D. Tao, Y. Yuan, X. Li, and T. Tan, "Biologically inspired features for scene classification in video surveillance," *Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics*, vol. 41, no. 1, pp. 307–313, 2011.
- [13] M. Hamidi and A. Borji, "Invariance analysis of modified c2 features: case study-handwritten digit recognition," *Machine Vision and Applications*, vol. 21, no. 6, pp. 969–979, 2010.
- [14] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.
- [15] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognition a gentle way," in *Biologically Motivated Computer Vision*. Springer, 2002, pp. 472–479.
- [16] R. J. Peters and L. Itti, "Applying computational tools to predict gaze direction in interactive visual environments," *Acm Transactions on Applied Perception*, vol. 5, no. 2, 2008.
- [17] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
- [18] Y. F. Ma, X. S. Hua, L. Lu, and H. J. Zhang, "A generic framework of

Features	Dimensions
Geometric attributes	15
Prominent attributes	1
Special patch	scale [4, 6, 8, 10] × 8
Attributes similarity threshold	1
C2 similarity threshold	1

TABLE VI: Attributes for one 'known' face.

identity	1	2	4	3	5
attri_thres	4.5946	4.1696	4.0039	4.4526	3.5100

TABLE VII: The thresholds for attributes

- user attention model and its application in video summarization,” *Ieee Transactions on Multimedia*, vol. 7, no. 5, pp. 907–919, 2005.
- [19] S. Lang, M. Kleinhagenbrock, S. Hohanner, J. Fritsch, G. A. Fink, and G. Sagerer, “Providing the basis for human-robot-interaction: A multi-modal attention system for a mobile robot,” in *Proceedings of the 5th international conference on Multimodal interfaces*. ACM, 2003, pp. 28–35.
- [20] S. Frintrop, E. Rome, and H. I. Christensen, “Computational visual attention systems and their cognitive foundations: A survey,” *Acm Transactions on Applied Perception*, vol. 7, no. 1, 2010.
- [21] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [22] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations,” in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 609–616.
- [23] A.-r. Mohamed, G. E. Dahl, and G. Hinton, “Acoustic modeling using deep belief networks,” *Ieee Transactions on Audio Speech and Language Processing*, vol. 20, no. 1, pp. 14–22, 2012.
- [24] E. Tulving and H. J. Markowitsch, “Episodic and declarative memory: Role of the hippocampus,” *Hippocampus*, vol. 8, no. 3, pp. 198–204, 1998.
- [25] A. Martin and L. L. Chao, “Semantic memory and the brain: structure and processes,” *Current Opinion in Neurobiology*, vol. 11, no. 2, pp. 194–201, 2001.
- [26] L. R. Squire, J. T. Wixted, and R. E. Clark, “Recognition memory and the medial temporal lobe: A new perspective,” *Nature Reviews Neuroscience*, vol. 8, no. 11, pp. 872–883, 2007.
- [27] B. McElree, P. O. Dolan, and L. L. Jacoby, “Isolating the contributions of familiarity and source information to item recognition: A time course analysis,” *Journal of Experimental Psychology-Learning Memory and Cognition*, vol. 25, no. 3, pp. 563–582, 1999.
- [28] A. Martin, J. V. Haxby, F. M. Lalonde, C. L. Wiggs, and L. G. Ungerleider, “Discrete cortical regions associated with knowledge of color and knowledge of action,” *Science*, vol. 270, no. 5233, pp. 102–105, 1995.
- [29] J. Hart and B. Gordon, “Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation,” *Annals of Neurology*, vol. 27, no. 3, pp. 226–231, 1990.
- [30] J. R. Hodges, K. Patterson, S. Oxbury, and E. Funnell, “Semantic dementia - progressive fluent aphasia with temporal-lobe atrophy,” *Brain*, vol. 115, pp. 1783–1806, 1992, 6.
- [31] G. Wang, K. Tanaka, and M. Tanifuji, “Optical imaging of functional organization in the monkey inferotemporal cortex,” *Science*, vol. 272, no. 5268, pp. 1665–1668, 1996.
- [32] I. Fujita, K. Tanaka, M. Ito, and K. Cheng, “Columns for visual features of objects in monkey inferotemporal cortex,” *Nature*, vol. 360, no. 6402, pp. 343–346, 1992.
- [33] M. P. Stryker, “Neurobiology - elements of visual-perception,” *Nature*, vol. 360, no. 6402, pp. 301–302, 1992.
- [34] M. J. Tovee, “Is face processing special?” *Neuron*, vol. 21, no. 6, pp. 1239–1242, 1998.
- [35] M. W. Brown and J. P. Aggleton, “Recognition memory: What are the roles of the perirhinal cortex and hippocampus?” *Nature Reviews Neuroscience*, vol. 2, no. 1, pp. 51–61, 2001.
- [36] F. Haist, A. P. Shimamura, and L. R. Squire, “On the relationship between recall and recognition memory,” *Journal of experimental psychology. Learning, memory, and cognition*, vol. 18, no. 4, pp. 691–702, 1992.
- [37] E. Hirshman and S. Master, “Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm,” *Memory and Cognition*, vol. 25, no. 3, pp. 345–351, 1997.
- [38] W. Donaldson, “The role of decision processes in remembering and knowing,” *Memory and Cognition*, vol. 24, no. 4, pp. 523–533, 1996.
- [39] M. Seeck, C. M. Michel, N. Mainwaring, R. Cosgrove, H. Blume, J. Ives, T. Landis, and D. L. Schomer, “Evidence for rapid face recognition from human scalp and intracranial electrodes,” *Neuroreport*, vol. 8, no. 12, pp. 2749–2754, 1997.
- [40] D. L. Hintzman, D. A. Caulton, and D. J. Levitin, “Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall,” *Memory and Cognition*, vol. 26, no. 3, pp. 449–462, 1998.
- [41] R. Bogacz, M. W. Brown, and C. Giraud Carrier, “High capacity neural networks for familiarity discrimination,” in *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on*, vol. 2. IET, 1999, pp. 773–778.
- [42] L. R. Squire and S. M. Zola, “Episodic memory, semantic memory, and amnesia,” *Hippocampus*, vol. 8, no. 3, pp. 205–211, 1998.
- [43] D. Saumier and H. Chertkow, “Semantic memory,” *Current neurology and neuroscience reports*, vol. 2, no. 6, pp. 516–22, 2002.
- [44] H. Eichenbaum, “A cortical-hippocampal system for declarative memory,” *Nature Reviews Neuroscience*, vol. 1, no. 1, pp. 41–50, 2000.
- [45] M. A. Conway, “Episodic memories,” *Neuropsychologia*, vol. 47, no. 11, pp. 2305–2313, 2009.