A Memory Model for Visual Image Storage and Recall Based on Bayesian Decision

Ying Jiang, Yanjiang Wang*, Yujuan Qi, Baodi Liu

College of Information and Control Engineering, China University of Petroleum, Qingdao 266580, PR China
yjwang@upc.edu.cn

Abstract

Traditional pattern recognition methods mainly focus on ‘classification’ rather than ‘cognition’. A new object which has been never seen before is often falsely classified as a certain kind of category studied, while humans will just respond unknown. This is mainly due to the fact that the human recognition process is intimately associated with the human memory system. So far, most memory models are studied in word list and few are reported in visual images. In this paper, we incorporate the human memory into visual image classification and propose a memory model for visual image storage and recall based on Bayesian decision (VISRBD). All the studied images are stored in the form of feature vector in the memory space. And each image feature component is correctly copied with certain probability generated by an exponential distribution. When a probe image is presented, the likelihood ratio between it and each studied image is calculated based on probabilistic theory. Then the odd in favor of an old over a new item is computed based on the ratio values between the test image and all the studied images. According to the odd value, the Bayesian decision rule for image classification is performed. Experimental results show that the proposed VISRBD model can gain good classification performance and the false alarm rate is far lower than SVM.

Keywords

Memory Model; Image Storage; Image Recall; Image Classification; Bayesian Decision

Introduction

As an active research topic in computer vision, visual image classification and recognition have been widely applied in many areas, such as face recognition for identity authentication, pedestrian detection and tracking in visual surveillance, and vehicle tracking for traffic monitoring. Visual image classification methods have drawn a lot of attention and been studied for more than fifty years with various theories and algorithms, e.g. the nearest neighbor classifier [1], neural network classifier [2], support vector machine classifier (SVM) [3], etc.. While Most of the existing categorization methods, including the bag-of-features (BOF) [4] and spatial pyramid matching (SPM)[5], only focus on ‘classification’, which distinguishes one category from other categories studied.

However, the human cognition process can distinguish one category from infinite unknown categories, which mainly lays emphasis on ‘cognition’ first and then ‘classification’. The traditional pattern recognition methods ignore ‘cognition’, and a new object which has been never seen may be falsely classified as a certain studied category. However, in the same situation, the immediate reaction of a human may be unknown rather than identifying it as a certain studied category directly.

As we all know, the main goal of computer vision research is to enable the computers to recognize and classify visual images as easily as humans do. It is supposed in psychology, cognitive science and neuroscience that the fact that human can easily retrieve target from its surrounding environment has close relationship with the memory...
mechanism. Whatever human have seen and experienced has to be processed within the memory system. When perceiving new thing, the relevant information in memory will be recalled so as to speed up the cognitive process and adapt to the new environment. However, how human brain stores and retrieves visual images still remains unclear.

Modern memory models have their roots in the models developed in the 1950s by mathematical psychologists such as Estes, Bush, and Mosteller[6, 7], etc.. Later in 1960s, Atkinson and Shiffrin[8] proposed a psychology model for describing the memory structure, which may indeed be regarded as the first modern model for human memory. Murdock et al. [9] ever indicated that any general theory of memory must specify at least four things: how information is represented, the type of information that is stored and retrieved, the nature of the storage and retrieval operations, and the format of the storage. Concentrating on these memory issues, researchers have proposed various memory models in psychology field, mainly including episodic memory and semantic memory, such as the SAM (Search of associative memory) theory [10], MINERVA2 model [11], TCM (Temporal Context Model) model[12], REM (Retrieving Effective from Memory) model [13], BCDMEM(the Bind Cue Decide Model of Episodic Memory) model [14], distributed associative memory model(TODAM) [15], and so on.

Most of the above memory models can provide explanation for the memory effect of word list, whereas the study about how visual images are represented and stored in human brain is limited. In addition, although the recent state-of-the-art image classification methods such as bag-of-features (BOF) and spatial pyramid matching (SPM) [16] as well as some other image classification frameworks in image processing research have exhibited good performance. These algorithms lay emphasis on the classification only but lack the recognition ability.

This paper tries to model the process of visual information representation, storage, and retrieval by incorporating the human memory modeling approach in psychology into computer vision field. All the studied image feature vectors are stored in the memory space and the feature is correctly copied with certain probability generated by an exponentially distribution. When a probe image is presented, the likelihood ratio between it and each studied image is calculated based on probabilistic theory. Then the odd in favor of an old over a new item is computed. If the odd is less than 1, it is thought that the test image belongs to a new category; otherwise, it is regarded as an old one and classified as the category with the maximum category likelihood value.

The structure of the paper is as follows. In Section 2, the related work especially the REM model is briefly introduced which is studied in word list. In Section 3, the memory model for visual image storage and recall based on Bayesian decision (VISRBD) is described in detail, including the psychologically visual information representation, storage and the recall process. And the image recognition and classification rule based on Bayesian decision is presented in detail as well. In Section 4, several experiments are conducted on the benchmark datasets to evaluate the recognition and classification performance of the model. Finally, Section 5 draws the conclusions.

**Relative Work**

The REM model suggested that human memory consists of separate images; each is represented as a vector of feature values, and the stored vector is an incomplete and error prone copy of the studied feature vector. When studying a word, there exists a probability $u^*$ that some new information will be stored for each feature. Note that once some value has been stored, it is not changed any longer. If something is stored for some feature, its value is copied correctly from the studied vector with probability $c$; the stored value is chosen randomly according to $P[V = j] = (1 - g)^{j-1} g, j = 1, 2, ..., \infty$ with the probability $1 - c$, and allowing for choosing the correct value.
accidentally.

Given the probe vector that either has been studied or has not been studied, it is matched in parallel to the studied word vectors and the matching result $D = \{ D_j \}_{j=1,...,N}$ is obtained, where $D_j$ indicates the aligning result between the probe and the $j$th word feature vector. Label those positions whose values match and positions whose values do not match, and ignore the positions where the feature contains no value. To make it easier to follow, a currently presented image that has been stored during an earlier presentation is termed an $s$-image. An image that has been stored during presentation of any image other than the image currently presented is termed a $d$-image.

Then the likelihood ratio $\lambda_j$ is calculated, that is the probability that the $j$th image is an $s$-image divided by the probability that the $j$th image is a $d$-image based on observed result $D_j$:

$$\lambda_j = \frac{P(D_j|S_j)}{P(D_j|N_j)} = \frac{(1-c)^v}{\prod_{k=M}^c (1-g)^v g^{-1}}$$

(1)

Where $S_j$ represents the event that the $j$th image is an $s$-image, $N_j$ denotes the event that the $j$th image is an $d$-image, $M$ is the number of all nonzero matching features in the $j$th image, $V_k$ signifies the value of the $k$th feature in the $j$th image, $g$ is the geometric distribution parameter. And ultimately the odds is obtained in favor of an old over a new test item:

$$\Phi = \frac{1}{N} \sum_{j=1}^N \lambda_j$$

(2)

If $\Phi > 1$, the probe word is regarded as ‘old’ and matches with the $j$th word that corresponds to the maximum $\lambda_j$; otherwise, the probe word is decided to be a ‘new’ word.

The basic REM model provides a mechanism for how the memory system responds to a particular cue. Researchers have demonstrated that REM can explain a wide range of episodic memory phenomena, such as the list-length, list-strength, and word-frequency effects, etc.

**Visual Information Storage and Recall based on Bayesian Decision**

**Visual Information Representation and Storage**

Given a visual image set $I = \{ I_1, I_2, ..., I_N \}$, it is consisted of $M$ different categories denoted by $\omega = \{ \omega_1, \omega_2, ..., \omega_M \}$, $M \leq N$. The image number of the $i$th category $\omega_i$ is set to be $n_i$, i.e. $\{I_{i_1}, I_{i_2}, ..., I_{i_{n_i}}\} \in \omega_i$, $1 \leq i \leq M$. Therefore, the image set can be rewritten as $I = \{ I_{11}, I_{12}, ..., I_{n_1}, I_{21}, I_{22}, ..., I_{n_2}, ..., I_{M1}, I_{M2}, ..., I_{Mn_M} \}$, where $I_j$ signifies the $j$th image belonging to the $i$th
category, \[ \sum_{i=1}^{M} n_i = N. \]

Then, the feature vector of the image \( I_{ij} \) is extracted and subsequently signified by \( V'_{ij} \in R^{1 \times K} \), and the feature set is defined as \( V' = \{ V'_{11}, V'_{12}, \ldots, V'_{1n_1}, V'_{21}, \ldots, V'_{2n_2}, \ldots, V'_{M1}, \ldots, V'_{Mn_M} \} \), where \( V'_{ij} \) means the feature vector of the image \( I_{ij} \).

During memory storage process, each feature vector is stored with an incomplete copy. For storing the feature vector, there exists a probability \( u^* \) that some new information will be stored for each feature. Once some value has been stored, it is not changed any longer. The stored value is correctly coped with probability \( c \), and assigned to a random value with probability \( 1 - c \) according to an exponential distribution based on the parameter \( g \):

\[
P[V = x] = \begin{cases} \ e^{-gx}, & x > 0 \\ 0, & \text{otherwise} \end{cases}
\]

After the above incomplete copy procedure, the final image feature set \( V = \{ V'_{ij} \}_{i=1 \ldots M, j=1 \ldots n_j} \) is stored instead of \( V' = \{ V'_{ij} \}_{i=1 \ldots M, j=1 \ldots n_j} \).

**Recall**

When a probe image is presented, it either belongs to a certain studied category, or is a new image. The feature vector \( V_{test} \in R^{1 \times K} \) of the probe image \( I_{test} \) is then extracted and matched in parallel to the studied feature set \( V = \{ V_{11}, \ldots, V_{Mn_M} \} \) stored. It is supposed that the nonzero value in a certain position of the feature vector indicates that the corresponding position is activated. Then the matching rule is given as follows:

for \( 1 \leq k \leq K \),

\[
D_{ij}(k) = \begin{cases} V_{test}(k) & \text{if} V_{test}(k) \neq 0 \text{and} V'_{ij}(k) \neq 0 \\ 1 - c & \text{if} V_{test}(k) \neq 0 \text{and} V'_{ij}(k) = 0 \\ 1, & \text{otherwise} \end{cases}
\]

where \( D_{ij} \) stands for the matching result between \( V_{test} \) and \( V'_{ij} \), the whole matching result is signified by \( D = \{ D_{11}, \ldots, D_{Mn_M} \} \).

Similar to REM model, the likelihood ratio \( \hat{\lambda}_{ij} \) is calculated based on observed result \( D_{ij} \):
and \( ij \) is, \( \ldots, \) denote the odds that the probe image \( \mathbf{I} \) be the number of all nonzero mismatched features, respectively. When \( k \in Q \),

\[
P(V_{k,ij} \mid S_{ij}, V_k) = (1-c)P(V_{k,ij} \mid N_{ij}, V_k)
\]

Let \( n_{ij,q} \) be the number of all nonzero mismatched features of the image \( I_{ij} \), by substituting (6) into (5), we have:

\[
\lambda_{ij} = (1-c)^{n_{ij,q}} \prod_{k \in M} \frac{P(V_{k,ij} \mid S_{ij}, V_k)}{P(V_{k,ij} \mid N_{ij}, V_k)}
\]

As mentioned previously, the probability of storing value \( v \) is \( g(v) \) with an exponential distribution in (3). Therefore

\[
\lambda_{ij} = (1-c)^{n_{ij,q}} \prod_{k \in M} \frac{c + (1-c)g(V_{k,ij})}{g(V_{k,ij})} = (1-c)^{n_{ij,q}} \prod_{k \in M} \frac{c + (1-c)e^{-gV_{k,ij}}}{e^{-gV_{k,ij}}}
\]

For the given probe image \( I_{test} \), it is checked against with all the studied images \( I = \{I_{ij}\}_{i=1,...,M, j=1,...,n_i} \), and then the likelihood ratio set \( \hat{\lambda} = \{\lambda_{ij}\}_{i=1,...,M, j=1,...,n_i} \) is calculated.

**Image Recognition and Classification by Bayesian Decision**

The first issue is to determine whether the probe image \( I_{test} \) is old or new. Let \( \Phi \) denote the odds that the probe image \( I_{test} \) is in favor of an old over a new item, which equals the probability that the probe image \( I_{test} \) is old divided by the probability that it is new. The prior probabilities that the probe image is old or new are supposed to be identical. Then we have:

\[
\lambda_{ij} = \frac{P(S_{ij})P(D_{ij} \mid S_{ij})}{P(N_{ij})P(D_{ij} \mid N_{ij})} = \frac{P(D_{ij} \mid S_{ij})}{P(D_{ij} \mid N_{ij})}
\]

\[
= \prod_{k=1}^{K} \frac{P(V_{k,ij} \mid S_{ij}, V_k)}{P(V_{k,ij} \mid N_{ij}, V_k)} \prod_{k \in Q} \frac{P(V_{k,ij} \mid N_{ij}, V_k)}{P(V_{k,ij} \mid N_{ij}, V_k)}
\]

Where the upper equality in Eq.(5) is obtained by supposing that the prior probability of \( s \)-image and \( d \)-image is identical. \( S_{ij} \) and \( N_{ij} \) represent the event that image \( I_{ij} \) is an \( s \)-image and a \( d \)-image, respectively. \( V_k \) signifies the value of the \( k \)th feature in the probe image, \( V_{k,ij} \) is the value of the \( k \)th feature of the image \( I_{ij} \). \( M \) and \( Q \) are the sets of indices for the matched and mismatched nonzero features, respectively.
\[ \Phi = \frac{P(O|D)}{P(N|D)} = \frac{P(O)P(D|O)}{P(D)} = \frac{P(D|O)}{P(D|N)} = \sum_{j=1}^{M} \frac{P(D|S_{ij})P(S_{ij})}{P(D)} \]

\[ = \sum_{j=1}^{M} \frac{1}{n_{ij}} \frac{1}{M} \sum_{j=1}^{n_{ij}} \sum_{j=1}^{n_{ij}} \frac{P(D|S_{ij})}{P(D)} = \frac{1}{M} \sum_{j=1}^{n_{ij}} \sum_{j=1}^{n_{ij}} \frac{P(D|S_{ij})}{P(D)} \]

\[ = \frac{1}{M} \sum_{j=1}^{n_{ij}} \sum_{j=1}^{n_{ij}} \frac{1}{n_{ij}} \lambda_{ij} \]

(9)

If \( \Phi > 1 \), the probe image is regarded as ‘old’; otherwise, it is decided to be a ‘new’ image. The next issue is to identify which category the probe image belongs to when \( \Phi > 1 \).

In the interest of simplicity, assume there are only two different classes, \( \omega_u \) and \( \omega_v \), \( 1 \leq u, v \leq M \). Then to judge which class the probe image \( I_{test} \) belongs to. Let the posteriori probability of class \( \omega_u \) and \( \omega_v \) given \( I_{test} \) be \( P(\omega_u|I_{test}) \) and \( P(\omega_v|I_{test}) \), respectively. The pattern is classified into class \( \omega_u \) or \( \omega_v \) according to the Bayesian decision rule for minimum error:

\[ I_{test} \in \begin{cases} \omega_u, \text{if } P(\omega_u|I_{test}) > P(\omega_v|I_{test}) \\ \omega_v, \text{otherwise} \end{cases} \]  

(10)

Eq. (10) can also be converted into the following expression:

\[ I_{test} \in \begin{cases} \omega_u, \text{if } \alpha > 1 \\ \omega_v, \text{otherwise} \end{cases} \]

where \( \alpha = \frac{P(\omega_u|I_{test})}{P(\omega_v|I_{test})} \)

(11)

The studied images of class \( \omega_u \) and \( \omega_v \) are denoted by \( \{I_{uj}\}_{j=1...n_u} \) and \( \{I_{vj}\}_{j=1...n_v} \), respectively, and \( \alpha \) can be rewritten as:

\[ \alpha = \frac{P(\omega_u|I_{test})}{P(\omega_v|I_{test})} \]

\[ = \frac{\sum_{j=1}^{n_{uj}} P(D|S_{uj})P(S_{uj})}{\sum_{j=1}^{n_{vj}} P(D|S_{vj})P(S_{vj})} = \frac{1}{M} \frac{1}{n_{uj}} \sum_{j=1}^{n_{uj}} \lambda_{uj} \]

\[ \frac{1}{n_{uj}} \sum_{j=1}^{n_{uj}} \frac{1}{n_{uj}} \sum_{j=1}^{n_{uj}} \lambda_{uj} \]

(12)
Thus the final Bayesian decision rule for two-class image classification can be defined as, 

\[ I_{\text{test}} \in \begin{cases} \omega_v, & \text{if } \frac{1}{n_v} \sum_{j=1}^{n_v} \lambda_{v,j} > \frac{1}{n_v} \sum_{j=1}^{n_v} \lambda_{v,j} \\
\omega_i, & \text{otherwise} \end{cases} \quad (13) \]

Similarly, for multi-class classification, suppose the studied class set \( \omega = \{ \omega_1, \omega_2, ..., \omega_M \} \), then the final Bayesian decision rule can be defined as,

\[ I_{\text{test}} \in \omega_v, \text{ if } \frac{1}{n_v} \sum_{j=1}^{n_v} \lambda_{v,j} > \frac{1}{n_v} \sum_{j=1}^{n_v} \lambda_{v,j}, \text{ for all } v \neq u, 1 \leq u, v \leq M. \quad (14) \]

**Experiment**

To evaluate the performance of the VISRBD model, we present the experimental results on Coil-20 dataset [17] and Caltech-101 dataset[18].

**Experimental Settings**

The scale invariant feature transform (SIFT) [19] is an efficient image feature extraction method with invariance to scale, rotation, and affine transformations and much robust to noise and illumination changes. It has been applied to object detection and recognition for many years. In the following experiments, the dense SIFT features are firstly extracted with patch size 16x16 and step size 8 pixels. For each patch, 128 dimensional SIFT descriptors are obtained and normalized to 1 with \( l_2 \) norm. Assume the number of image patches is \( k \), then the size of the dense SIFT feature of an image is \( k \times 128 \).

Then sparse coding [20] is applied to obtain the sparse representation of the dense SIFT features. The basic idea of the sparse representation is that a natural image can be represented by the linear combination of learned dictionary. It has been widely used in image coding, object classification as well as pattern recognition. During dictionary learning, more than 500 samples are chosen randomly. Here the dictionary size is selected as \( 128 \times 512 \), so the size of the final feature vector of an image is \( k \times 512 \).

**Coil-20 Dataset**

The Coil-20 dataset is composed of images of 20 objects. Each object is rotated horizontally by 360 degrees and the images are taken by every 5 degrees. Thus each object consists of 72 images and each image is 128x128 pixels. Fig. 1 shows some samples of images in Coil-20 dataset.

![FIG. 1 SAMPLES OF IMAGES IN COIL-20 DATASET]
First, we select the first 15 objects for studying, and 8 rotated images per object are taken by every 0, 45, 90, 135, 180, 225, 270, 315 degrees. In addition, the probe images are selected from the dataset for test, including 360 images of 5 studied objects and 360 images of 5 new objects, as shown in Fig. 2.

![Sample Images](image)

**FIG. 2 SAMPLES OF PROBE IMAGES ON COIL-20 DATASET**

Table 1 shows the experimental result of the VISRBD model on the Coil-20 dataset with different parameters \( c \) and \( g \).

<table>
<thead>
<tr>
<th>( c )</th>
<th>( g = 0.05 )</th>
<th>( g = 0.1 )</th>
<th>( g = 0.2 )</th>
<th>( g = 0.3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>80.14</td>
<td>89.86</td>
<td>78.06</td>
<td>56.53</td>
</tr>
<tr>
<td>0.2</td>
<td>75.83</td>
<td>87.64</td>
<td>86.25</td>
<td>73.89</td>
</tr>
<tr>
<td>0.3</td>
<td>76.33</td>
<td>88.06</td>
<td>86.11</td>
<td>81.67</td>
</tr>
<tr>
<td>0.4</td>
<td>70.28</td>
<td>87.92</td>
<td>85.83</td>
<td>82.36</td>
</tr>
<tr>
<td>0.5</td>
<td>83.33</td>
<td>88.06</td>
<td>85.28</td>
<td>81.53</td>
</tr>
</tbody>
</table>

It is obvious that the recognition rate reaches the maximum 89.86% with \( c = 0.1 \) and \( g = 0.1 \). In addition, in order to observe the impact of the parameters on the hit probability and false alarm rate, the corresponding curves are presented in Fig. 3. The hit probability \( P(H) \) is the probability of saying ‘old’ when a target is presented, while the probability of a false alarm \( P(F) \) is the probability of saying ‘old’ when a distracter is presented.

It can be seen from Fig. 3(a) that the hit probability with \( g = 0.05 \) and \( g = 0.1 \) is obviously higher than that of other \( g \). While in Fig. 3(b), it can be seen that the false alarm rate of \( g = 0.05 \) is much higher than that of other \( g \).

When \( g = 0.1 \), the hit probability reaches the maximum at \( c = 0.1 \) and the false alarm rate is lower than that of other value \( c \) except for \( c = 0.5 \). Thus \( g = 0.1 \) and \( c = 0.1 \) are chosen on Coil-20 dataset.

![Experimental Curves](image)

**FIG. 3 THE EXPERIMENTAL CURVE UNDER THE VARYING VALUES OF \( c \) AND \( g \)**
To investigate the influence of the training image number on recognition performance, we choose variable number of training samples (1, 4, 6, 8, 12 and 18) from each object category. For the first 15 objects on the Coil-20 dataset, the rotated images taken by 0, 90, 60, 45, 30 and 20 degrees are selected for training. Therefore there are 15, 60, 90, 120, 180, 270 training images, respectively. And the test image set is consisted of 360 images of 5 studied objects and 360 images of 5 new objects. Fig.4 shows the experimental result on the Coil-20 dataset with the training image number increasing with $g = 0.1$ and $c = 0.1$.

As shown in Fig.4, the hit probability is increasing faster than the false alarm rate so that the recognition rate is steadily increasing as the training image number increases.

To further illustrate the performance difference between the VISRBD model and the traditional pattern recognition method such as SVM, we choose different number of training samples (6, 8, 12, 18 and 24) from each object category among 15 objects, so there are 90, 120, 180, 270, 360 training images, respectively. The test images are the same as the experiment above. Table 2 shows the comparison classification performance between VISRBD and SVM with $g = 0.1$, $c = 0.1$.

<table>
<thead>
<tr>
<th>Training number per object</th>
<th>P(H) VISRBD</th>
<th>P(H) SVM</th>
<th>P(F) VISRBD</th>
<th>P(F) SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>91.11</td>
<td>95</td>
<td>15.56</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>96.94</td>
<td>99.72</td>
<td>17.22</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>99.44</td>
<td>99.72</td>
<td>20.28</td>
<td>100</td>
</tr>
<tr>
<td>18</td>
<td>100</td>
<td>100</td>
<td>21.11</td>
<td>100</td>
</tr>
<tr>
<td>24</td>
<td>100</td>
<td>100</td>
<td>23.61</td>
<td>100</td>
</tr>
</tbody>
</table>

As mentioned above, P(H) and P(F) represent the hit probability and the false alarm rate, respectively. For the training images which should be responded as ‘old’, both methods provide excellent performance. When 12, 18, 24 training images per object are chosen, P(H) of VISRBD and SVM are almost the same. While 6, 8 images per object are trained, P(H) of SVM is a little higher than that of VISRBD. However, For the training images which should be responded as ‘new’, P(F) of SVM reaches up to 100% and VISRBD is obviously with a much lower false alarm rate.
This is the key point which VISRBD differs from the traditional pattern recognition method.

**Caltech-101 Dataset**

The Caltech-101 dataset contains 102 classes, one of which is the background, the remaining 101 classes with 8,677 images in total are used for classification test, with each class varying from 31 to 800 images. Fig.5 shows some examples of images in Caltech-101 dataset.

![Fig.5 Examples of Caltech-101 Dataset](image)

To evaluate the performance of the proposed VISRBD model, we conduct the recognition and classification task on the Caltech-101 dataset. First, 20 categories are selected while 12 images of each category are chosen as the training set. Some training objects are shown in Fig.6.

For testing, 20 categories (including 10 studied and 10 new categories) are selected while 30 images per category are chosen. Fig.7 shows some test objects.

Table 3 presents the experimental result of the VISRBD model on the Caltech-101 dataset with different $c$ and $g$.

<table>
<thead>
<tr>
<th>$c$</th>
<th>$g$ = 0.05</th>
<th>$g$ = 0.1</th>
<th>$g$ = 0.2</th>
<th>$g$ = 0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>47.33</td>
<td>71.17</td>
<td>70</td>
<td>68.17</td>
</tr>
<tr>
<td>0.2</td>
<td>53</td>
<td>70.17</td>
<td>70</td>
<td>68.5</td>
</tr>
<tr>
<td>0.3</td>
<td>64.17</td>
<td>71.17</td>
<td>70</td>
<td>68.33</td>
</tr>
<tr>
<td>0.4</td>
<td>68.83</td>
<td>71.5</td>
<td>69.33</td>
<td>68.17</td>
</tr>
<tr>
<td>0.5</td>
<td>70.5</td>
<td>70.67</td>
<td>69</td>
<td>68.17</td>
</tr>
</tbody>
</table>

It can be seen from Table 4 that the recognition and classification performance is almost the same when $g$=0.1, $g$ = 0.2 and $g$ = 0.3. In addition, the hit probability and false alarm rate curves are shown in Fig.8.

Similar to the discussion in 4.2, $c = 0.1$ and $g = 0.1$ are chosen on the Caltech-101 dataset too.

For further validation of the performance, we conduct the comparison experiment between VISRBD and SVM. And the training and test images are chosen as the above experiment. The classification performance of VISRBD and SVM on Caltech-101 dataset is illustrated in Table 4 with $c = 0.1$ and $g = 0.1$. 

242
FIG. 6 SAMPLES OF THE TRAINING IMAGES ON CALTECH-101 DATASET

FIG. 7 SAMPLES OF THE TEST IMAGES ON CALTECH-101 DATASET

(a) The Hit Probability Curve  (b) The False Alarm Rate Curve

FIG. 8 THE EXPERIMENTAL CURVE UNDER THE VARYING VALUES OF $c$ AND $g$
It can be seen from Table 4 that, the hit probability of SVM is higher than that of VISRBD, but its false alarm rate reaches 100% which is much higher than that of VISRBD (12.33%). It is the extremely low false alarm rate which makes VISRBD outperform the traditional pattern methods.

### Conclusion

Inspired by human memory model, the major contribution of this paper is to propose a memory model for visual image storage and recall based on Bayesian decision (VISRBD). Experimental results have shown that VISRBD can be applied to the recognition and classification of visual images. To make it easier to understand the proposed model, Table 5 shows the comparison of VISRBD with the REM model.

The VISRBD can first judge whether the probe image has been seen or not. If yes, the probe image is then classified. If the recognition result is new, the classification process will not be performed. This is the difference between the VISRBD model and other traditional pattern recognition methods. We compare the performance of VISRBD with SVM, and find that VISRBD can gain a far lower false alarm probability while the hit probability is almost the same or only a little lower than SVM.

### ACKNOWLEDGEMENTS

This work is supported in part by National Natural Science Foundation of China (Grant No.61271407, 61402535),
the Fundamental Research Funds for the Central Universities (Grant No.14CX06066A) and the Shandong Provincial Natural Science Foundation, China (Grant No. ZR2013FQ015, ZR2014FQ001).

**REFERENCES**


