Neurocomputing 259 (2017) 194–200

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Psychologically inspired visual information storage and retrieval modeling for multiclass image classification

Ying Jiang, Yanjiang Wang*

*College of Information and Control Engineering, China University of Petroleum (East China), Qingdao 266580, PR China*

ARTICLE INFO

*Article history:*
Received 24 January 2016
Revised 14 September 2016
Accepted 18 September 2016
Available online 8 February 2017

*Keywords:*
SIFT
Sparse coding
Psychological memory modeling
Information storage
Information retrieval
Image classification

ABSTRACT

The computer vision research aims to enable the computers to recognize visual images as easily as human. Studies have shown that human could segregate target from its surrounding environment, which is intimately associated with the brain memory mechanism. However, it is not quite clear about how the visual images are stored and retrieved in the human brain. In this paper, we propose a psychologically visual information storage and retrieval model (PVISRM) based on sparse coding and probabilistic decision theory. First, the dense scale invariant feature transform (SIFT) algorithm is applied to extract the features of visual images and then the extracted features are represented by sparse coding. In the storage procedure, each component of the feature vector is correctly copied with certain probability generated by an exponential distribution. For retrieval, the likelihood ratio between the probe image feature vector and that of each studied image is calculated based on probabilistic theory. Then the category likelihood ratio between the probe image and each category is obtained by adding the ratio values of all images belonging to the same category. Finally, the Bayesian decision rule for image classification is presented. Experimental results show that the proposed PVISRM model can obtain good classification performance and outperforms the SVM approach.



## 1. Introduction

Computer vision tasks aim to make the computers recognize visual images as easily as human. It is supposed in psychology, cognitive science and neuroscience [1] that human can easily retrieve target from its surrounding environment, which has close relationship with human memory mechanism. All things which human has seen and experienced must be processed within the memory system. Once perceiving new thing, the memory information relating to it will be retrieved in order to speed up the cognitive process and adapt to the new environment. However, it is not quite clear about how the visual images are stored and retrieved in the human brain. Murdock [2] 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 the issues how information is stored and retrieved in memory, since the 1980s, researchers have proposed many memory modeling theories in psychology field, mainly including episodic memory and semantic memory. For example, the SAM (Search of Associative Memory) theory, based on earlier work by Raaijmakers and Shiffrin [3], was initially developed as a model for free recall. The MINERVA2 model developed by Hintzman [4] has been applied primarily to category learning and recognition memory. It was the first explicit mathematical model that incorporated the episodic memory with semantic memory to be applied to retrieval modeling. The TCM (Temporal Context Model) model, first proposed by Howard and Kahana [5], made the memory contents change gradually over time. It linked memory performance with time factor, and could explain two basic principles of human episodic memory recency and contiguity. The REM model (Retrieving Effective from Memory), first published by Shiffrin and Steyvers [6], started out as a model for recognition memory, and could account for some phenomena in many episodic memory researches, such as list-strength, list-length, word frequency, mirror effects and the normal ROC slope effect in recognition memory. In addition to the above memory models, there are also other various of memory models, such as the BCDMEM model (the Bind Cue Decide Model of Episodic Memory) by Dennis and Humphreys [7], the spreading activation model by Anderson and Bower [8], the distributed associative memory model (TODAM)by Murdock [2,9], as well as the SLiM model (Subjective Likelihood Model) by McClelland and Chappell [10], etc.

Most episodic memory models, such as SAM, REM, and MINERVA2 models, suppose that the recognition judgments are completed based on the global match familiarity signal strength.

* Corresponding author.
*E-mail address:* yjwang@upc.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.neucom.2016.09.126
0925-2312/

REM is highlighted not only because of its principled mathematical foundation, but also for its capacity of accounting for some phenomena in many episodic memory research. Therefore it is being developed and applied most actively [11–14].

However, most of the memory models including REM model can only account for the memory effect of word list, whereas the study about how visual images are represented and stored in human brain has been not reported so far. A recent study proposed a probabilistic clustering theory of the organization of visual short-term memory [15], however, it cannot model the content of a subject's visual memory for a natural scene. Although the recent state-of-the-art image classification methods such as bag-of-features (BoF) [16–18] in image processing research have exhibited good performance, they cannot explain how the images are organized and processed from the perspective of memory modeling.

In this paper, therefore, we try to model the process of visual information representation, storage, and retrieval by introducing the human memory modeling approach in psychology into computer vision field. First, the scale invariant feature transform (SIFT) method [19] is applied to extract features of visual images and then the extracted features are organized by sparse coding [20]. Then, the Bayesian decision based probabilistic theory is applied to calculate the likelihood ratio between the test image feature and that of each studied category. Finally, category of the test image is decided by the probabilities between the test image and all studied categories.

The rest of the paper is organized as follows. In Section 2, the related work about REM model is introduced in brief. In Section 3, the psychological visual information storage and retrieval model is described in detail, and the Bayesian discrimination method for image classification is introduced as well. In Section 4, we conduct several experiments on the benchmark datasets to test the model. Finally, Section 5 draws the conclusions.

## 2. Related work

Learning and memory research aims to understand how human store and retrieve information. Computational models of memory provide formal implementations of memory theories. It is well known that long-term human memory is often divided into two further main types: declarative (or explicit) memory and non-declarative (or implicit) memory. Nondeclarative memory is memory that has been learned without phenomenal awareness and pops into mind as illustrated earlier [21]. Declarative (or explicit) memory can be divided into two categories: episodic memory, which stores specific personal experiences, and semantic memory, which stores factual information [22].

Episodic memory experiments typically consist of a study phase followed by a test phase, where some subjects are exposed to a set of stimuli. The test phase is either a recognition memory test or a recall test. Episodic memory models try to describe the mental algorithms which support performance on recognition and recall tests, without addressing how the algorithms can be implemented in human brain.

Individual memories are typically represented as vectors, and each element of this vector denotes a particular feature of the memory. During study, memory traces are placed separately in a long-term store. It is supposed that new memory traces do not affect the previously stored memory traces. During test, the memory model calculates the match between the probe cue and all of the items stored in memory. The item-by-item match information can be summed across all items to calculate a global match familiarity signal. Most episodic memory theories posit that the recognition judgments are made based on the global match familiarity signal strength. There are some models which conform to this structure, such as SAM, REM, and MINERVA2.

REM is highlighted because of its principled mathematical foundation, therefore it is being developed and applied most actively [11–14]. REM, which was proposed by Shiffrin and Steyvers, can account for some phenomena, for instance the list length, list strength, and word frequency effects in many episodic memory research [6]. The list length effect in recognition memory refers to the finding that recognition performance for a short list is superior to that for a long list. The list strength effect refers to the finding that strengthening some list items (by repeating the items or presenting them for longer periods of time) does not impair recognition of other, non-strengthened items. The word-frequency effect is the finding that words of higher frequency are recognized less well than words of lower frequency.

It is suggested in the REM model that human memory contains separate images; each image is expressed as a feature value vector, and the stored vector is an incomplete and error prone copy of the studied feature vector. During 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 in some feature location, its value is copied correctly from the studied vector with probability $c$; some random value is chosen to be stored according to $P[V=j]=(1-g)^{j-1}g,\ j=1,2,\ldots,\infty$ with the probability $1-c$ , and allowing for selecting the correct value by accident.

Given the test vector, it either has been studied or is new. We match it in parallel to the studied word vectors and denote the aligning result as $D=\{D_j\}_{j=1,\ldots,N}$, where $D_j$ represents the matching result between the test and the $j$th word feature vector. Label those positions where the value matches and positions where the value do not match, and the positions where the feature contains no value is ignored. To illustrate this further, 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, which is in fact 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 $D_j$:

$$\lambda_j = \frac{P(D_j|S_j)}{P(D_j|N_j)} = (1-c)^{n_{jq}} \prod_{k\in M} \frac{c+(1-c)g(1-g)^{V_{kj}-1}}{g(1-g)^{V_{kj}-1}} \tag{1}$$

Where $S_j$ denotes the event that the $j$th image is an $s$-image, $N_j$ represents the event that the $j$th image is a $d$-image, $n_{jq}$ is the number of all nonzero mismatching features in the $j$th image, $M$ refers to the sequence number set of all nonzero matching features in the $j$th image, $V_{kj}$ indicates the $k$th feature value in the $j$th image, $g$ signifies the geometric distribution parameter. And the odds are ultimately obtained in favor of an old over a new probe item:

$$\Phi = \frac{1}{N}\sum_{j=1}^{N} \lambda_j \tag{2}$$

If $\Phi > 1$, the probe word is decided to be old and regarded to match with the $j$th word which corresponds to the maximum $\lambda_j$; otherwise, the probe word is regarded to be a new word. The basic REM model provides a mechanism for how the memory system responds to a particular cue.

## 3. The psychological visual information representation, storage and retrieval modeling

### 3.1. Image representation and storage

Given the image set $I=\{I_1, I_2, \ldots, I_K\}$, it is supposed that the set is consisted of $M$ different categories denoted by