Neurocomputing 173 (2016) 2041–2048

Contents lists available at ScienceDirect

Neurocomputing

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

Brief Papers

A PMJ-inspired cognitive framework for natural scene categorization in line drawings

Minjing Yu [a], Yong-Jin Liu [a,*], Su-Jing Wang [b], Qiufang Fu [b], Xiaolan Fu [b]

[a] Tsinghua National Lab for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[b] State Key Lab of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form
7 August 2015
Accepted 16 September 2015
Communicated by Xu Zhao
Available online 28 September 2015

Keywords:
Visual perception
Line drawings
Natural scene categorization

ABSTRACT

Humans' remarkable capacity on rapid natural scene categorization has been widely studied in neuroscience. Recently, a functional MRI (fMRI) study showed that in human brain, decoding of natural scenes from line drawings was very similar to those from color photographs. In this paper, based on recently proposed computational cognition model of Perception, Memory and Judgement (PMJ model), we investigate the computational model of line drawings and propose a PMJ-inspired cognitive framework for natural scene categorization in line drawings. The Ohio State University (OSU) dataset was used, which included 475 color photographs in six categories, i.e., beaches, city streets, forests, highways, mountains and offices, as well as 475 corresponding line drawings produced by trained artists. Experimental results show that our proposed cognitive framework achieves 48.4% recognition rate in leave-one-out cross-validation, which is much higher than fMRI-data-driven decoding accuracy in the visual-processing hierarchy (29% in V1, 27% in V2+VP, 26% in V4, 29% in PPA and 23% in RSC).

1. Introduction

It is well known that humans had remarkable capacity at perceiving and categorizing natural scenes, and the neural mechanisms of rapid natural scene categorization in human brain had been widely investigated (e.g., [1]). It was reported [2] that in human brain, scene category information can be encoded in patterns of functional MRI (fMRI) activity in the parahippocampal place area (PPA), the retrosplenial cortex (RSC), the lateral occipital complex (LOC), and the primary visual cortex (V1). Visual areas V2, VP and V4 are also interested since they build representations based on V1 information.

Recently, a study on natural scene categorization by reducing the scenes to mere lines was presented in [3]. In this study, color photographs in six categories (beaches, city streets, forests, highways, mountains and offices) were collected. Then line drawings were created by trained artists who traced those contours in the photographs that best captured the scene. We call these data *the Ohio State University (OSU) dataset*. An elaborated experiment was performed in [3] in which color photographs and line drawings in OSU dataset were presented alternatively to participants. Functional MRI images were recorded when participants passively viewed the OSU dataset. Experimental results showed that despite the marked difference in scene statistics and consideration degradation of information, scene category can be decoded from fMRI data for line drawings just as well as from activity for color photographs.

In this paper, we propose a computational cognitive framework that investigates how a computer program can be used to simulate the human vision system that categorizes natural scenes from line drawings. Our work is based on a recently proposed computational cognition model of *Perception, Memory and Judgement* (PMJ model) [4]. In the perception stage, we compute a saliency map on color photograph, and map the salient region onto the corresponding line drawing. In the memory stage, we apply a local histogram with circular bins to extract feature instantiations from perceived line drawings. The collected feature instantiations are clustered in a bag-of-word model and forms a visual vocabulary. Then each line drawing is presented by a feature vector that is a set of visual words in the vocabulary. In the judgement stage, a SVM classifier is applied and optimal parameters are trained from feature vectors represented line drawings in six categories in leave-one-out cross-validation. Experimental results show that our proposed cognitive framework in machine vision has a consistent performance in the line with the pattern of brain activity measured with fMRI in observers who viewed line drawings of natural scenes. Our PMJ-inspired cognitive framework achieves 48.4% recognition rate in OSU dataset, which is much higher than

* Corresponding author.
E-mail address: liuyongjin@tsinghua.edu.cn (Y.-J. Liu).

http://dx.doi.org/10.1016/j.neucom.2015.09.046
0925-2312/

fMRI-data-driven decoding accuracy in the visual-processing hierarchy (29% in V1, 27% in V2+VP, 26% in V4, 29% in PPA and 23% in RSC).

2. Related work

A line drawing is usually referred to as a set of sparse, simple two-dimensional feature lines without hatching or stippling for shading/tone effects [5]. Humans have an innate ability to perceive, recognize and interpret line drawings, e.g., children's sketches and line arts by ink on paper. Line drawings had been used for studying objects and scenes in human cognition for more than three decades [6,7]. Recently a fMRI study found that the neural activation in response to line drawings was similar to that in color photographs [3] and a study that investigated how the nature scene categorization of line-drawings and color photographs was reflected in event-related potentials (ERPs) was presented in [8].

Due to less information stored in line drawings when compared to the color photographs, the usage of line drawings in intelligent process of visual media, including the retrieval and reuse of images [9,10], videos [11,12], 3D graphical models [13–15] and conceptual design in industry [16,17], has attracted considerable attention recently. The reader is referred to [5] for a survey. However, up to now, there are very few algorithms that directly simulated the patterns of brain activity due to its extreme complexity [18]. Since object and natural scene categorization is one of the fundamental problems that find a wide range of applications in computer vision, it is much desired to develop practical algorithms that make use of the cognitive mechanism in rapid categorization in linear drawings. These practical algorithms may also shed some lights on solving other more complex problems such as scene understanding and intelligent decision making.

Line drawings include hand-drawn figures, symbols or handwritten texts. In particular, as a kind of pictographs, Chinese characters had been widely studied in its own right. The reader is referred to [19] for a survey in the field of handwritten Chinese character recognition (HCCR). Similar to the intelligent process of visual media using line drawings, state-of-the-art HCCR technique has been extended to build a personalized handwritten Chinese recognition engine [20]. Here we emphasize that the line drawings studied in this paper are general, i.e., depicting one of the six natural scene contents including beaches, city streets, forests, highways, mountains and offices.

3. PMJ-inspired cognitive framework

Consensus had been reached in psychological research that most cognitive processes consist of several successive processing stages [22]. In our study, we apply the PMJ model [4] that partitions the cognitive process into the stages of perception, memory and judgement, corresponding to the stages of analysis, modeling and decision in the computation process. In the stage of perception, through pre-attention selection and selective attention, the cognitive load of cognition system is reduced and salient visual

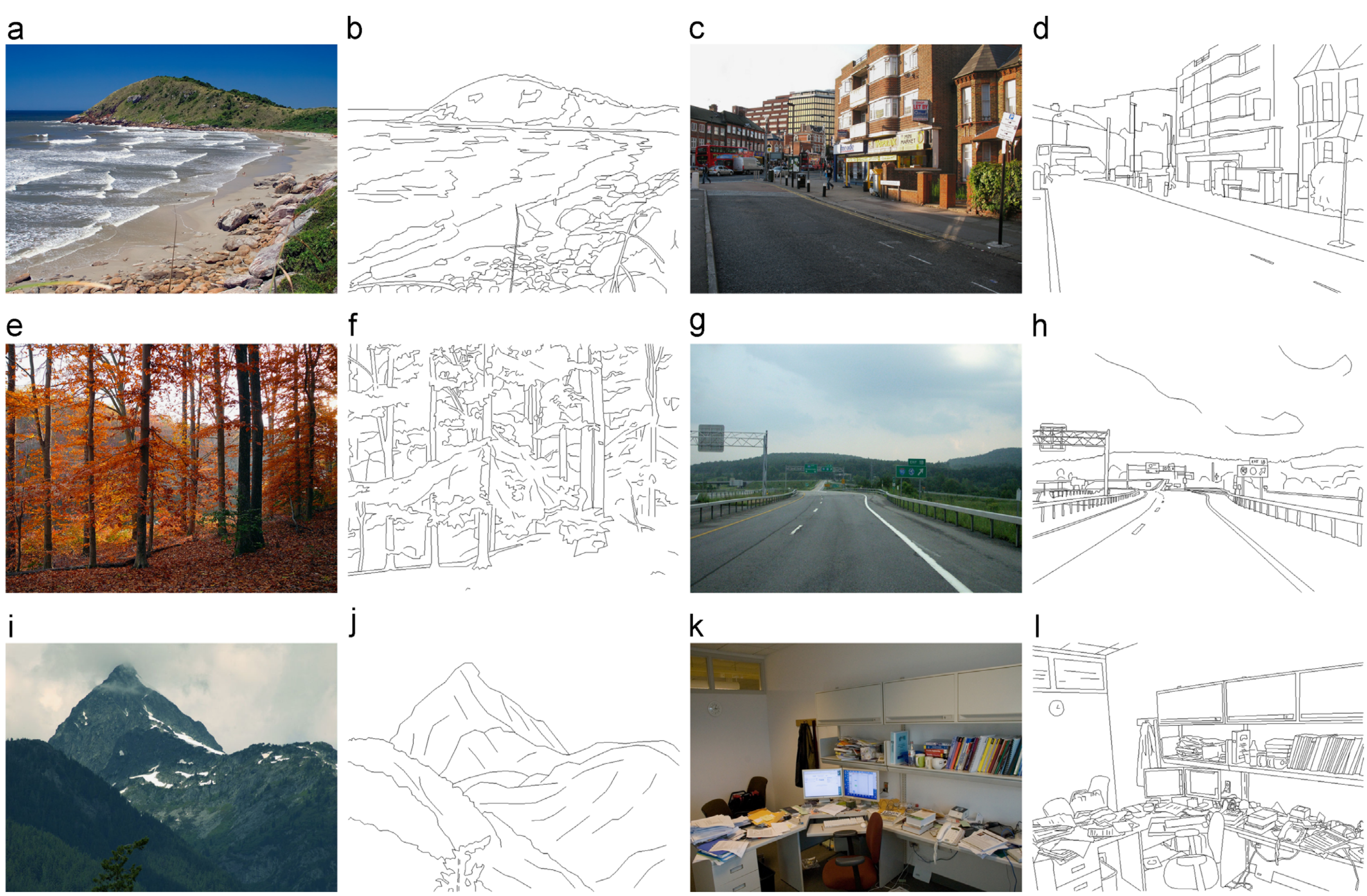

Fig. 1. Examples of color photographs (CP) and corresponding line drawings (LD) in six categories (beaches, city streets, forests, highways, mountains and offices) in the OSU dataset. (a) Beach (CP). (b) Beach (LD). (c) City street (CP). (d) City street (LD). (e) Forest (CP). (f) Forest (LD). (g) Highway (CP). (h) Highway (LD). (i) Mountain (CP). (j) Mountain (LD). (k) Office (CP). (l) Office (LD). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

features are extracted. In the stage of memory, dynamic memory system is achieved through the mechanisms of encoding and storing processes (short-term memory) and the mechanisms of updating and consolidation (long-term memory). In the stage of judgment, judgments or decisions are made in an efficient way through categorization learning and action coding. These three stages can interact with each other to complete the cognitive processing tasks.

We apply the PMJ model to develop a practical algorithm for natural scene categorization depicting by line drawings. The OSU dataset is used for supervised learning, which contains 475 color photographs in six categories: beaches (80), city streets (79), forests (80), highways (80), mountains (76) and offices (80). Each color photograph has a corresponding line drawing that was produced by trained artists at the LotusHill Research Institute by tracing contours in the color photographs via a custom graphical user interface. The order and coordinates of all line strokes were recorded digitally to allow for later reconstruction of the line drawings at any resolution in terms of preserving long or short contours, in order to reflect the global or local structures. All color photographs and line drawings were rendered at a resolution of 800×600 pixels. Line drawings are represented by black lines on a white background. See Fig. 1 for an example.

3.1. Perception

In the perception stage, selective attention is applied and salient visual information is extracted. We compute a saliency map for each color photograph using the GBVS algorithm [21]. Refer to Fig. 2. The GBVS algorithm assigns a saliency value between 0 and 1 to each pixel in the color photograph, where 1 means the most significance and 0 means the least significance (Fig. 2b). We determine a salient region in the image by setting a threshold τ in saliency map. The choice of an optimal threshold τ is discussed in Section 4 and experimental results show that our cognitive framework is not sensitive to the threshold value.

We store the salient region in a binary image, in which each white pixel has a saliency value larger than the threshold τ (Fig. 2c). Then we generate a perceived line drawing (Fig. 2e) by only preserving in full line drawing (Fig. 2d) the lines falling into the salient region. The memory stage presented in the next section utilizes a set of sample points (Fig. 2f) in perceived line drawing to generate a vocabulary representation of categorized line drawings.

3.2. Memory

The memory stage contains two types of memories: short-term and long-term memories. Short-term memory of new learned information is created instantly and can be easily disrupted by learning other information. In other words, the short-term memory is fragile. Meanwhile, the memories are consolidated over time from short-term (seconds to hours) to long-term (days to months) and the memory enhancement by long-term training has been demonstrated to be a biological mechanism [23].

We apply the distributed memory computation model proposed in [15] to compute a vocabulary-based memory representation from perceived line drawings. The memory model in [15] is briefly summarized below:

- Short-term memory: After randomly sampling the perceived line drawings, a histogram with circular bins is generated as an instantiation of features. All instantiations in perceived line drawings are then clustered into a vocabulary.
- Long-term memory: The vocabulary is not static but only a transience (short term) and the stationary distribution (long term) in vocabulary is modeled by a state space in a discrete-time Markov chain.

Fig. 3 illustrates the histogram of circular bins. For each line drawing of 800×600 pixels, we generate $PtNum = 500$ random points by applying Halton's quasi-random point sequence. Then averagely there are $\frac{Area_{perceived}}{800 \times 600} \times PtNum$ random points inside a perceived line drawing (Fig. 2f). For each random point in a perceived line drawing, a histogram with circular bins ($NumCir = 15$) is generated (Fig. 3a). The radius of the maximum circle is $l_{digonal}/5$, where $l_{digonal}$ is the diagonal length of the bounding box of the

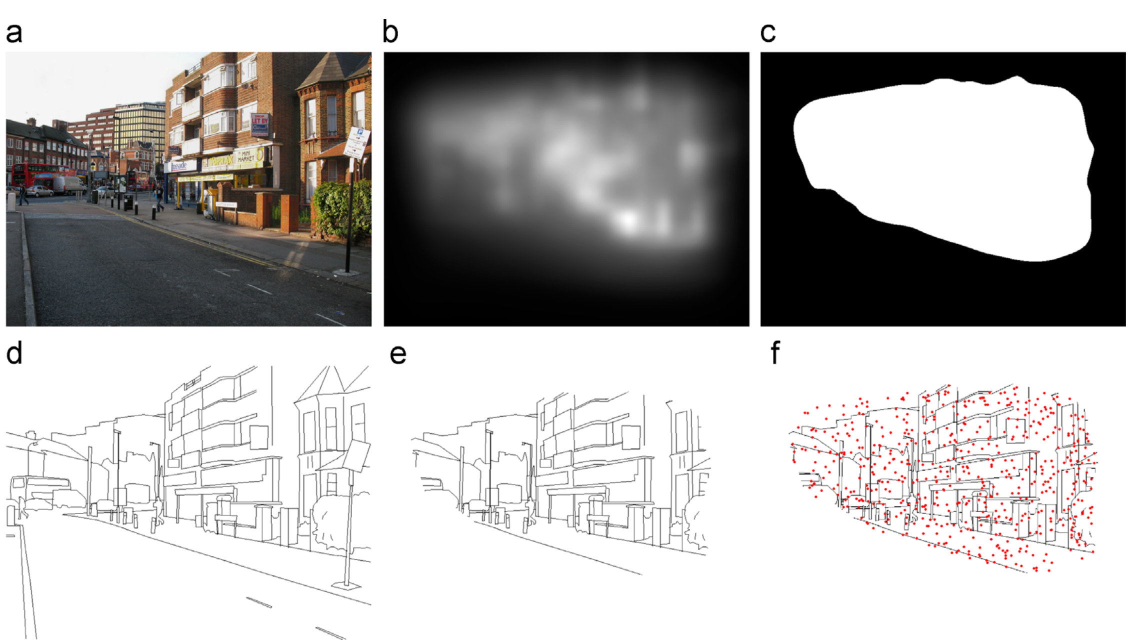

Fig. 2. The perception stage in PMJ-inspired cognitive framework. (b) is the saliency map of the image (a) using the GBVS algorithm [21]. (c) is the salient region (indicated by white) by setting a threshold $\tau = 0.2$ in (b). (e) is the perceived line drawing by applying the salient region (c) into the full line drawing (d). (f) shows sampling points (in red) in perceived line drawing. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

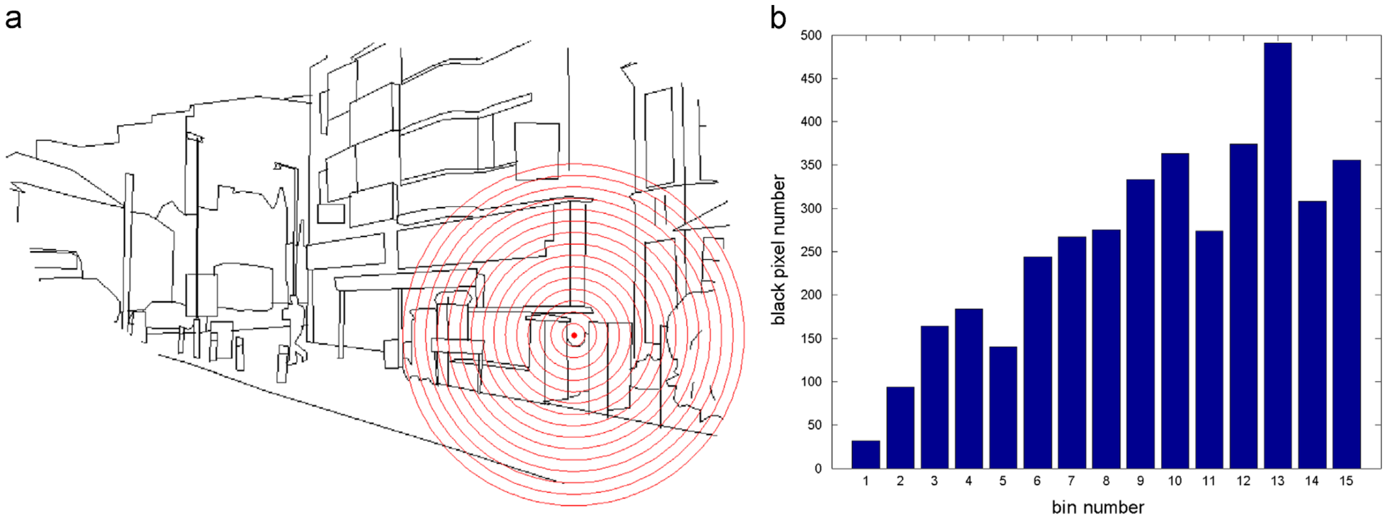

Fig. 3. A histogram with circular bins as feature instantiation. (a) For each sample point shown in Fig. 2f, a histogram with circular bins ($NumCir=15$) is generated. (b) The value of each bin is the number of black pixels falling into that bin.

Table 1
The confusion matrix of the recognition results obtained by our PMJ-inspired cognitive framework (PMJCF) by leave-one-out cross-validation.

		Ground truth					
		Beach	City	Forest	Highway	Mountain	Office
Prediction	Beach	**66.3%**	7.6%	11.3%	33.8%	25.0%	12.5%
	City	3.8%	**58.2%**	22.5%	18.8%	1.3%	26.3%
	Forest	0.0%	11.4%	**41.3%**	13.8%	4.0%	12.5%
	Highway	12.5%	10.1%	2.5%	**25.0%**	4.0%	12.5%
	Mountain	10.0%	0.0%	11.3%	1.3%	**65.8%**	1.3%
	Office	7.5%	12.7%	11.3%	7.5%	0.0%	**35.0%**

Table 2
The confusion matrix of the recognition results obtained by PMJCF in two classes: man-made scenes (city streets, highways and offices) and non-man-made scenes (beaches, forests and mountains).

		Ground truth	
		Man-made scenes	Non-man-made scenes
Predict	Man-made scenes	**78.0%**	31.4%
	Non-man-made scenes	22.0%	**68.6%**

perceived area. The differences of radii of remaining circles are of equal-size. Then the value of each bin in this histogram is the number of black pixels falling into that bin (Fig. 3b).

Each histogram (corresponding to a random sample) can be represented by a vector in a space $\mathbb{R}^{NumCir}$ of dimension $NumCir=15$. All the random samples in all perceived line drawings in the OSU dataset generate a point cloud in space $\mathbb{R}^{NumCir}$. We apply the K-means clustering algorithm in $\mathbb{R}^{NumCir}$ and for each cluster its center is treated as a visual word v_i. In our experiment, we choose $K=2000$ and generate a vocabulary $V=\{v_1, v_2, \dots, v_{2000}\}$.

Each visual word v_i may not have equal representative capacity for classifying categories of line drawings. In other words, some visual words may be more presentative than the others in the vocabulary. We apply the Markov chain model in [15] to select the most representative words in vocabulary and put them in long-term memory.

Assume that a person watch and learn the line drawings one-by-one. During the watch process, more and more visual words from new observed line drawings are generated and entered into short-term memory. Then we regard the vocabulary $V=\{v_1, v_2, \dots, v_K\}$ as a state space and each word v_i is a state. Denote by $\lambda=\{\lambda_i : v_i \in V\}$ the possibility measure on V, where $\lambda_i \geq 0$ is the possibility of word v_i be representative. We maintain the constraint $\sum_{v_i \in V} \lambda_i = 1$ such that λ defines a distribution of a random state V.

To model the dynamic watching process, we introduce an artificial time t into the distribution $(\lambda_t)_{t \geq 0} = (V_t : 0 \leq t < \infty)$. To model the memory consolidation, let p_{ij} be the transition probability between any two words v_i and v_j, defined by

$$p_{ij} = \frac{s_{ij}}{\sum_{j=1}^{K} s_{ij}}$$

where

$$s_{ij} = D_{ij} - \frac{1}{n_i}\frac{1}{n_j}\sum_{I_u \in \mathbf{v}_i}\sum_{I_v \in \mathbf{v}_j} D(I_u, I_v),$$

$$D_{ij} = \max\{D(I_u, I_v), \forall I_u \in \mathbf{v}_i, \forall I_v \in \mathbf{v}_j\},$$

$\mathbf{v}_i$ and $\mathbf{v}_j$ are the clusters of v_i and v_j respectively, I_u and I_v are the instantiations (a histogram represented by a point in $\mathbb{R}^{NumCir}$) in $\mathbf{v}_i$ and $\mathbf{v}_j$ respectively, $D(I_u, I_v)$ is the distance between I_u and I_v, and n_i and n_j are the numbers of instantiations in $\mathbf{v}_i$ and $\mathbf{v}_j$ respectively.

We characterize V_t as a Markov chain with transition matrix $\mathbf{P} = p_{ij} : 1 \leq i, j \leq K$ and initial distribution $\lambda_0 = \left\{\lambda_i = \frac{n_i}{n_{all}}\right\}$, where $n_{all} = n_1 + n_2 + \dots + n_K$. Then we can infer the long-term behavior from the stationary distribution of the Markov chain model. It has been shown [15] that the matrix $\mathbf{P}$ is irreducible and the discrete-time process V_t is ergodic. Accordingly, the stationary distribution λ_∞ can be directly computed as the solution to the left-eigenvector problem $\lambda_\infty \mathbf{P} = \lambda_\infty$.

The stationary distribution λ_∞ gives a representativeness measure on the words in vocabulary V. We sort all words in V using a decreasing order of λ_i and select the top 50% words as representative words to consist a filtered vocabulary $\tilde{V}$ in long-term memory.

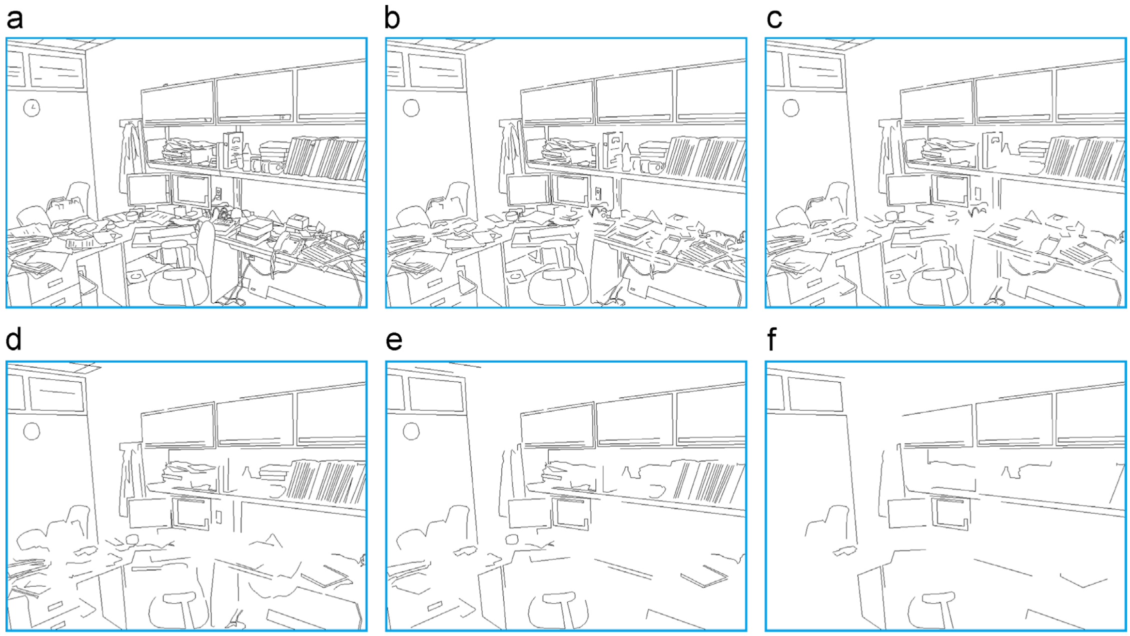

Fig. 4. Global-structure-preserving simplification of a line drawing shown in Fig. 1(l), by iteratively removing contours of shortest length. The resolution is defined by the percentage of pixels in original line drawing that were preserved in the simplified line drawing. (a) Resolution: 100%. (b) Resolution: 85%. (c) Resolution: 70%. (d) Resolution: 55%. (e) Resolution: 40%. (f) Resolution: 25%.

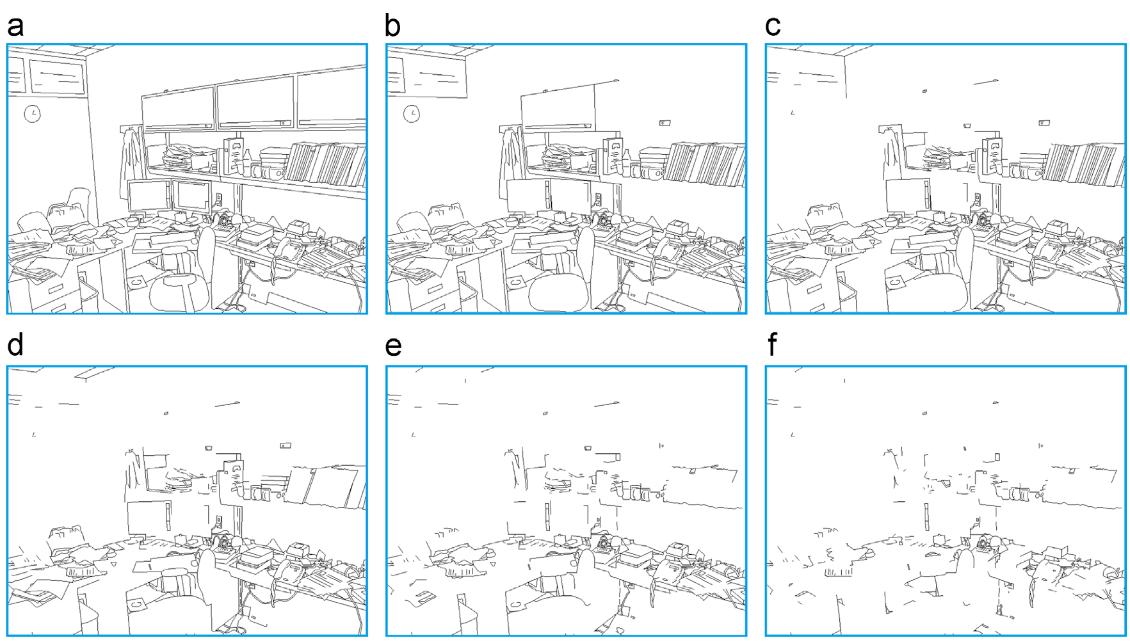

Fig. 5. Local-structure-preserving simplification of a line drawing shown in Fig. 1(l), by iteratively removing contours of longest length. The resolution is defined by the percentage of pixels in original line drawing that were preserved in the simplified line drawing. (a) Resolution: 100%. (b) Resolution: 85%. (c) Resolution: 70%. (d) Resolution: 55%. (e) Resolution: 40%. (f) Resolution: 25%.

3.3. *Judgment*

Based on the filtered vocabulary $\tilde{V} = \{v_1, v_2, \ldots, v_{K/2}\}$, each perceived line drawing LD can be presented by a feature vector $LD = (x_1, x_2, \ldots, x_{K/2})$, where $x_i = 1$ if there is an instantiation (corresponding to the histogram of a random sample) in LD which falls into the cluster of v_i, otherwise $x_i = 0$. Then we normalize the vector to have a unit length $\overline{LD} = \frac{LD}{\|LD\|_2}$, where $\|LD\|_2$ is the 2-norm length of LD.

Given the normalized feature vectors $\overline{LD}$ of all line drawings in six categories in the OSU dataset, we train a SVM classifier with a linear kernel $\mathcal{K}(x_i, x_j) = x_i^T x_j$. A penalty parameter in the kernel is optimized by a LIBSVM optimization function as presented in the next section.

4. Experiments

To the authors' best knowledge, the proposed method in this paper is the first computer program that can recognize the category of nature scenes in which a given line drawing belongs to. A previous work [3] that recognized categories of nature scenes in line drawings was based on the fMRI data taken from human observers. The same SVM classifier as in ours was used in [3].

Therefore, when we compared our recognition results with those in [3], the difference in performance was due to the recognition features we proposed in this paper.

4.1. Experiment setting

There are four parameters in our recognition features:

- The threshold τ in Section 3.1 to compute a binary image from a saliency map.
- The number *PtNum* of random points in Section 3.2 to be distributed in a line drawing of 800×600 pixels.
- The number *NumCir* of circular bins in the histogram in Section 3.2.
- The number K of clusters in the K-means clustering algorithm in Section 3.2.

In our implementation, we chose $\tau = 0.2$, $PtNum = 500$, $NumCir = 15$ and $K = 2000$. Note that our method was not sensitive to the parameters and the experimental results were stable when $0.1 \le \tau \le 0.3$, $300 \le PtNum \le 600$, $15 \le NumCir \le 20$ and $1500 \le K \le 2500$.

We used LIBSVM [24] with a linear kernel $\mathcal{K}(x_i, x_j) = x_i^T x_j$ for classification. In our application, there are six categories in the OSU dataset and LIBSVM can perform multiclass classification; see Section 7 in [24] for full details. The penalty parameter inside the implementation of the linear kernel was optimized by a LIBSVM function `SVMcgForClass`.

4.2. Recognition performance and comparison

We use leave-one-out cross-validation (LOOCV) for recognition performance evaluation. That is, in each fold, one line drawing was used as the test set and the others are used as the training set. After 475 folds, each subject has been used as the test set once and the final recognition accuracy was computed based on all the results.

Our experimental results with the setting presented in Section 4.1 showed that our PMJ-inspired cognitive framework (PMJCF) achieved a recognition rate 48.4% for line drawings in six categories (beaches, city streets, forests, highways, mountains and offices), which was significantly higher than the chance level 16.7%. As a comparison, the recognition rate in [3] that used the same OSU dataset and the same SVM classifier was only 29% by working with the patterns extracted from fMRI data in the primary visual cortex (V1), and was 27% in V2+VP, 26% in V4, 29% in PPA and 23% in RSC, where V1, V2, VP, V4, PPA and RSC are brain area in the visual-processing hierarchy.

The confusion matrix of our PMJCF results was summarized in Table 1. The results showed that the classes of beaches, city streets and mountains can be recognized very well. The class of highways had the worst recognition performance and were frequently recognized as beaches. This may be possibly because many highways images also contained parts of sky, mountain and sea which were also appeared in beaches images. The class of offices were frequently recognized as cities, possibly because these two classes of images contains blocks of straight line segments. The class of forests were sometimes recognized as cities. If we re-organized these images into man-made scenes (city streets, highways and offices) and non-man-made scenes (beaches, forests and mountains), the recognition rate was improved to 73.3% and the corresponding confusion matrix was summarized in Table 2.

4.3. Global vs. local structures in line drawings

The fMRI experimental results [3] revealed that line drawings can be decoded as accurately as photographs, although line drawings had remarkable difference in scene statistics and considerable degradation of information. Observing that color and texture information are lost in line drawings, a possible reason is that the geometric structure preserved in line drawings plays a primary role in representing scene categories. It is then interesting to ask whether long contours (reflecting global structure) or short contours (reflecting local structure) in line drawings is important to representing the geometric structure in images. In this section, we showed that long contours were more important than short contours in PMJCF if sufficient local details were provided; otherwise short contours were more important.

Experiment I: *Global-structure-preserving line drawing simplification*. In the OSU dataset, all line drawings were created by artists who traced contours in color photographs. For each line drawing, the order and coordinates of all line strokes were provided in the OSU dataset. To preserve the global structure, from a line drawing we iteratively removed contours of shortest length. This led to a hierarchy of line drawings with different resolutions. Here the resolution was defined by the percentage of black pixels in original line drawing that were preserved in the simplified line drawing. An example was illustrated in Fig. 4. Let x be a percentage ranged from 0% to 100%. We denoted by $OSU_{global}(x)$ be the set of line drawings, each one of which is a global-structure-preserving simplification at resolution x of a line drawing in the OSU dataset. We applied the proposed PMJCF with LOOCV on $OSU_{global}(x)$, x ranged from 5% to 100%, and the curve of recognition rates was shown in Fig. 6 (red curve).

Experiment II: *Local-structure-preserving line drawing simplification*. For each line drawing in OSU, we iteratively removed contours of longest length to preserve local structures. An example was illustrated in Fig. 5. We denoted by $OSU_{local}(x)$ be the set of line drawings, each one of which is a local-structure-preserving simplification at resolution x of a line drawing in the OSU dataset. We applied the proposed PMJCF with LOOCV on $OSU_{local}(x)$, x ranged from 5% to 100%, and the curve of recognition rates was shown in Fig. 6 (green curve).

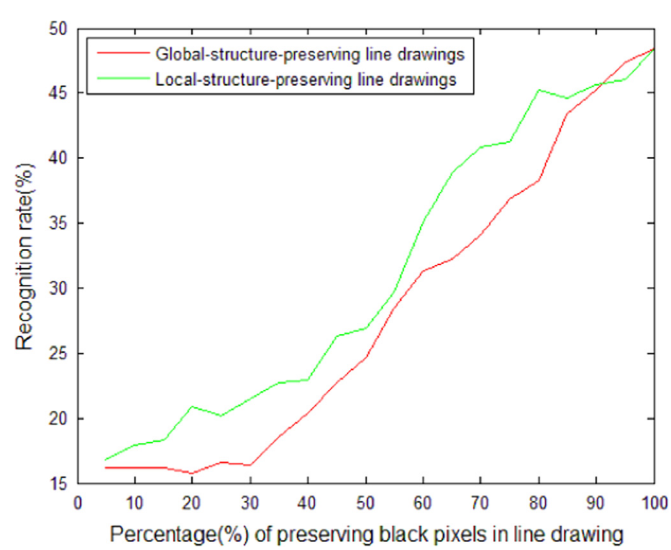

Fig. 6. Recognition rates in PMJCF of global or local-structure-preserving line drawing simplification. Red curve: recognition rates of $OSU_{global}(x)$, x ranged from 5% to 100%. Green curve: recognition rates of $OSU_{local}(x)$, x ranged from 5% to 100%. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

The results summarized in Fig. 6 showed that when the resolution x reached 5%, the sets $OSU_{global}(x)$ and $OSU_{local}(x)$ contain very few scene information and cannot be used for scene categorization; accordingly, their recognition rates both reached the chance level 16.7%. When the resolution $x \geq 90\%$, the global structure contained in long contours (which also contains sufficient local details indicated by $x \geq 90\%$) has a better performance than the local structure contained in short contours. When the resolution $x < 90\%$, local structures contained in short contours outperformed global structures contained in long contours; this possibly because PMJCF made use of histograms with local circular bins in the memory stage.

5. Conclusion

Natural scene categorization had drawn considerable attention from neuroscience and computer science. In this paper, a PMJCF computational model is proposed to recognize the categories of natural scenes based on line drawings. PMJCF consists of three stages of computation. At the first stage of perception, a saliency map is extracted from color photographs and applied to obtain perceived line drawings. At the second stage of memory, a vocabulary of visual words is obtained by clustering instantiations of features (short-term memory) followed by a stationary distribution analysis of a discrete-time Markov chain (long-term memory). At the third stage of judgement, a SVM classifier with a linear kernel is applied. Experimental results show that PMJCF can achieve above-chance recognition rate 48.4%, which is much better than the SVM-classifier-based recognition rates (the best is 29%) in [3].

Acknowledgments

The authors would like to thank Dr. Dirk Bernhardt-Walther for providing us the Ohio State University (OSU) dataset. This work was supported by the Natural Science Foundation of China (61322206, 61379095, 61272228, 61472138), the Beijing Natural Science Foundation (4152055), the Open Projects Program of National Laboratory of Pattern Recognition (201306295) and the National Basic Research Program of China (2011CB302200). The work of Y.J. Liu was supported in part by Tsinghua national laboratory for information science and technology cross-discipline foundation, and by Beijing Higher Institution Engineering Research Center of Visual Media Intelligent Processing and Security.

References

[1] M.V. Peelen, L. Fei-Fei, S. Kastner, Neural mechanisms of rapid natural scene categorization in human visual cortex, Nature 460 (2009) 94–97.
[2] D.B. Walther, E. Caddigan, L. Fei-Fei, D.M. Beck, Natural scene categories revealed in distributed patterns of activity in the human brain, J. Neurosci. 29 (34) (2009) 10573–10581.
[3] D.B. Walther, B. Chai, E. Caddigan, D.M. Beck, L. Fei-Fei, Simple line drawings suffice for functional mri decoding of natural scene categories, Proc. Natl. Acad. Sci. U.S.A. 108 (23) (2011) 9661–9666.
[4] X. Fu, L.-H. Cai, Y. Liu, J. Jia, W.-F. Chen, Y. Zhang, G.-Z. Zhao, Y.-J. Liu, C.-X. Wu, A computational cognition model of perception, memory and judgment, Sci. China Inf. Sci. 57 (3) (2014) 032114:1–032114:15.
[5] Y.-J. Liu, M. Yu, Q. Fu, W. Chen, Y. Liu, L. Xie, Cognitive mechanism related to line drawings and its applications in intelligent process of visual media: a survey, Front. Comput. Sci. 9, 2015, http://dx.doi.org/10.1007/s11704-015-4450-1.
[6] J.G. Snodgrass, M. Vanderwart, A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity, J. Exp. Psychol.: Hum. Learn. Memory 6 (2) (1980) 174–215.
[7] I. Biederman, R.J. Mezzanotte, J.C. Rabinowitz, Scene perception: detecting and judging objects undergoing relational violations, Cogn. Psychol. 14 (2) (1982) 143–177.
[8] Q. Fu, Y.-J. Liu, W. Chen, X. Fu, The time course of natural scene categorization in human brain: simple line-drawings vs. color photographs, J. Vis. 13 (9), 2013, http://dx.doi.org/10.1167/13.9.1060.
[9] Y. Cao, C. Wang, L. Zhang, L. Zhang, Edgel index for large-scale sketch-based image search, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado, USA, 2011, pp. 761–768.
[10] C.-C. Yu, Y.-J. Liu, T. Wu, K.-Y. Li, X. Fu, A global energy optimization framework for 2.1d sketch extraction from monocular images, Graph. Models 76 (5) (2014) 507–521.
[11] C.-X. Ma, Y.-J. Liu, H.-A. Wang, D.-X. Teng, G.-Z. Dai, Sketch-based annotation and visualization in video authoring, IEEE Trans. Multimed. 14 (4) (2012) 1153–1165.
[12] Y.-J. Liu, C.-X. Ma, Q. Fu, X. Fu, S.-F. Qin, L. Xie, A sketch-based approach for interactive organization of video clips, ACM Trans. Multimed. Comput. Commun. Appl. 11 (1) (2014) 2:1–2:21.
[13] Y.-J. Liu, C.-X. Ma, D.-L. Zhang, Easytoy: plush toy design using editable sketching curves, IEEE Comput. Graph. Appl. 31 (2) (2011) 49–57.
[14] Y.-J. Liu, X. Luo, A. Joneja, C.-X. Ma, X.-L. Fu, D.-W. Song, User-adaptive sketch-based 3-d cad model retrieval, IEEE Trans. Autom. Sci. Eng. 10 (3) (2013) 783–795.
[15] Y.-J. Liu, Q.-F. Fu, Y. Liu, X. Fu, A distributed computational cognitive model for object recognition, Sci. China Inf. Sci. 56(9), Article No. 092101(1-13), 2013.
[16] C.-X. Ma, Y.-J. Liu, H.-Y. Yang, D.-X. Teng, H.-A. Wang, G.-Z. Dai, Knitsketch: a sketch pad for conceptual design of 2d garment patterns, IEEE Trans. Autom. Sci. Eng. 8 (2) (2011) 431–437.
[17] L. Zeng, Y.-J. Liu, J. Wang, D.-L. Zhang, M.-F. Yuen, Sketch2jewelry: semantic feature modeling for sketch-based jewelry design, Comput. Graph. 38 (2014) 69–77.
[18] F. Cole, K. Sanik, D. DeCarlo, A. Finkelstein, T. Funkhouser, S. Rusinkiewicz, M. Singh, How well do line drawings depict shape?, in: ACM SIGGRAPH '09, 2009, pp. 28:1–28:9.
[19] C.-L. Liu, S. Jaeger, M. Nakagawa, Online recognition of Chinese characters: the state-of-the-art, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2) (2004) 198–213.
[20] X.-Y. Zhang, C.-L. Liu, Writer adaptation with style transfer mapping, IEEE Trans. Pattern Anal. Mach. Intell. 35 (7) (2013) 1773–1787.
[21] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Proceedings of Neural Information Processing Systems (NIPS), 2006, pp. 545–552.
[22] M.W. Eysenck, M.T. Keane, Cognitive Psychology: A Student's Handbook, 6th edition, Psychology Press, New York, 2010.
[23] J.L. McGaugh, Memory – a century of consolidation, Science 287 (5451) (2000) 248–251.
[24] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 27:1–27:27.

Minjing Yu received her B.Eng. degree from Wuhan University, China, and she is now a Ph.D. student at Department of Computer Science and Technology, Tsinghua University, China. Her research interests include image processing, computer graphics and cognitive science.

Yong-Jin Liu received his B.Eng. degree from Tianjin University, China, in 1998, and his M.Phil. and Ph.D. degrees from the Hong Kong University of Science and Technology, Hong Kong, China, in 2000 and 2004, respectively. He is now an Associate Professor with the Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, China. His research interests include computational geometry, multimedia, computer graphics and computer-aided design. For more information, visit http://cg.cs.tsinghua.edu.cn/people/Yongjin/yongjin.htm.

Su-Jing Wang received the Master's degree from the Software College of Jilin University, China, in 2007. He received the Ph.D. degree from the College of Computer Science and Technology of Jilin University in 2012. He is a Postdoctoral Researcher in Institute of Psychology, Chinese Academy of Sciences. He is One of the Ten Selectees of the Doctoral Consortium at International Joint Conference on Biometrics 2011. He was named *Chinese Hawkin* by the Xinhua News Agency. His current research interests include pattern recognition, computer vision and machine learning. For more information, visit http://sujingwang.name.

Qiufang Fu received her Ph.D. degree from Institute of Psychology, Chinese Academy of Sciences, in 2006. She is now an Associate Professor in Institute of Psychology, CAS, with interests in implicit learning and unconscious knowledge. She intends to explore the neural and cognitive mechanisms responsible for the dissociation of implicit and explicit processes.

Xiaolan Fu received the B.S. and M.S. degrees in psychology from Peking University, Peking, China, in 1984 and 1987, respectively, and the Ph.D. degree from the Institute of Psychology, Chinese Academy of Sciences, Beijing, China, in 1990. Currently she is the Director of Institute of Psychology, Chinese Academy of Sciences and Vice Director of State Key Laboratory of Brain and Cognitive Science. Her research interests focus on visual and computational cognition, including attention and perception, learning and memory, and affective computing. She serves as an Associate Editor of PsyCH Journal and journal of Protein & Cell.