International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 17, No. 1 (2019) 1940007 (16 pages)
© World Scientific Publishing Company
DOI: 10.1142/S0219691319400071

Enhanced biologically inspired model for image recognition based on a novel patch selection method with moment

Yanfeng Lu*,[‡] and Lihao Jia[†],[‡]

Research Center for Brain-Inspired Intelligence
Institute of Automation, Chinese Academy of Sciences
Beijing 100190, P. R. China
**yanfeng.lv@ia.ac.cn*
[†]*lihao.jia@ia.ac.cn*

Hong Qiao

The State Key Laboratory of Management
and Control for Complex Systems, Institute of Automation
Chinese Academy of Sciences, Beijing 100190, P. R. China
hong.qiao@ia.ac.cn

Yi Li

School of Information Engineering, Nanchang University
Nanchang 330031, P. R. China
littlepear@ncu.edu.cn

Zongshuai Qi

School of Computer and Communication Engineering
University of Science and Technology
Beijing 100083, P. R. China
qizongshuai@163.com

Received 15 September 2017
Revised 20 December 2017
Accepted 25 January 2018
Published 28 June 2018

Biologically inspired model (BIM) for image recognition is a robust computational architecture, which has attracted widespread attention. BIM can be described as a four-layer structure based on the mechanisms of the visual cortex. Although the performance of BIM for image recognition is robust, it takes the randomly selected ways for the patch selection, which is sightless, and results in heavy computing burden. To address this issue, we propose a novel patch selection method with oriented Gaussian–Hermite moment (PSGHM), and we enhanced the BIM based on the proposed PSGHM, named as PBIM. In contrast to the conventional BIM which adopts the random method to

[‡]These authors contributed equally to this work.

select patches within the feature representation layers processed by multi-scale Gabor filter banks, the proposed PBIM takes the PSGHM way to extract a small number of representation features while offering promising distinctiveness. To show the effectiveness of the proposed PBIM, experimental studies on object categorization are conducted on the CalTech05, TU Darmstadt (TUD) and GRAZ01 databases. Experimental results demonstrate that the performance of PBIM is a significant improvement on that of the conventional BIM.

Keywords: Image recognition; classification; BIM; oriented Gaussian–Hermite moment; Gabor features; patch selection.

AMS Subject Classification 2010: 22E46, 53C35, 57S20

1. Introduction

Image recognition has been widely applied in the applications of computer vision, such as robot navigation, pedestrian detection and clinical diagnosis.[16,17,28] In practical applications, the difficulties that arise in the image recognition are typically caused by variations in the appearance of the objects and the background complexity of the input images. The scale, rotation and illumination variability, especially in the cluttered backgrounds, disturb the recognition performance strongly. For instance, various human postures (e.g., squatting, stooping, running, or standing) in a real environment make accurate recognition a difficult task. To address this issue, lots of methods have been proposed in the past years.[29–31,34–37]

Conventional appearance-based methods often use the global low-level visual features, e.g., gray value, color, border and texture.[19] These methods usually take the extracted features into account equably; they do not selectively put particular emphasis on local discriminative features. Moreover, they are sensitive to occlusion, scale, illumination deformations. Local features-based methods mix local descriptors and key point detectors with spatial information. The representational methods, e.g., scale-invariant feature transform (SIFT),[12] gradient location and orientation histogram (GLOH),[18] histogram of gradients (HOG),[4] and speeded up robust features (SURF)[2] have been proposed. Although these methods are effective in representing the local discriminative features, they lack directional information. Even though bag-of-words (BoW)[9] and bag-of-features[11] are effective for resolving this issue; the amount of structure information still falls short.

Recently, significant advances have been made in the research of brain science.[6,20,24,32] The findings in the primary visual cortex V1 area are of significance. While researching the V1 area, Hubel and Wiesel discovered that the visual cortex analyzes features into various ways with different spatial orientations and frequencies.[6] The discovery gives an important support to early neuroscience theories. Based on these theories, Riesenhuber and Poggio described an original calculation framework for object recognition, called biologically inspired model (BIM) that tends to model the cognitive mechanism of the visual cortex.[24] Serre *et al.* upgraded the original BIM model and presented the standard BIM,[26] which shows that the visual framework significantly improves the performance of object recognition.

Lu *et al.* proposed a novel receptive field in the S1 layers and upgraded the framework by novel patch selection and matching processes.[13] Qiao *et al.* developed a updated BIM model (UBIM), and employed it in a robot system.[22,23] In conclusion, these mentioned approaches get remarkable performance by fusing certain biologically motivated mechanisms.

The traditional BIM model uses patches that are randomly selected in the second (C1) layers, which generates a huge amount of redundant information and also prevents robustness against rotational deformation. The stored patches in the C1 layers are the key components of the discriminative and robust abilities of BIM. Superior features extracted by the stored patches determine the feature invariance and selectivity, preserving BIM performance in the cases of object appearance variation and cluttered backgrounds. The majority of patches selected by the random method, however, are redundant and not discriminative for the recognition task, which results in performance degradation and high computational cost. These drawbacks seriously limit the overall performance of BIM. We propose a solution to this issue, a novel patch selection method with oriented Gaussian–Hermite moment called PSGHM. In the PSGHM, we employ the oriented Gaussian–Hermite moment (OGHM) to represent the first layers (S1) of the BIM,[24] and then the multi-scale keypoints are employed to locate the key regions of the S1, which aims to reduce the number of patches chosen, but keep those with better discrimination than those chosen by random selection. We further propose a PSGHM-based BIM model (PBIM). We show its effectiveness, by applying it to object categorization and by conducting experimental studies on the CalTech05, TU Darmstadt (TUD), and GRAZ01 databases.

The remaining part of the paper is organized as follows: in Sec. 2, we give an introduction about the conventional BIM; in Sec. 3, we describe the PSGHM method and PBIM model. In Sec. 4, we present experimental results based on three public databases. Finally, in Sec. 5, we give our conclusions.

2. Related Work

2.1. *BIM review*

Conventional BIM is a computational framework with four layers: S1, C1, S2 and C2, which follows the mechanisms of the primary visual cortex and builds feature representation by patch matching and maximum pooling operations.[26]

S1 layers: the units in the S1 layers correspond to simple cells in V1. The S1 units take the form of Gabor functions,[13] that model cortical simple cell receptive fields. Gabor functions are defined as

$$G(x,y) = \exp\left(-\frac{{x_o}^2 + \gamma^2 {y_o}^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi x_o}{\lambda}\right), \tag{2.1}$$

$$\text{s.t. } x_o = x\cos\theta + y\sin\theta \quad \text{and} \quad y_o = -x\sin\theta + y\cos\theta, \tag{2.2}$$

where θ represents orientation, λ is wavelength, σ is scale and γ indicates the spatial aspect ratio.

Given an input image, the S1 layer with orientation θ and scale σ is calculated by

$$S1_{\sigma,\theta} = |G_{\sigma,\theta} * I|,$$

where $*$ denotes convolution, I is the input image and $G_{\sigma,\theta}$ is a Gabor function with specific parameters.

C1 layers: These layers describe the complex cells in V1. The layers are the dimensionally reduced S1 layers obtained by selecting the maximum over local spatial neighborhoods. This maximum pooling operation over local neighborhoods increases invariance (providing some robustness to shift and scale transformations).

S2 layers: In these layers, S2 units pool over afferent C1 units from a local spatial neighborhood across all four orientations. The S2 layers describe the similarity between the C1 layers and stored patches in a Gaussian-like way using the Euclidean distance. The responses of the corresponding S2 layers are calculated by

$$S2 = \exp(-\beta\|C1(\sigma,\theta) - P_i\|^2), \tag{2.3}$$

where β is the sharpness of the exponential function, $C1(\sigma,\theta)$ denotes the afferent C1 layer with scale σ and orientation θ and P_i is the ith patch from the previous C1 layers.

C2 layers: The final set of shift- and scale-invariant C2 responses is computed by taking a global maximum of afferent S2 units across all scales and positions. The responses of the C2 layers are calculated by

$$C2 = \max_{(i,j,\sigma)}(S2(i,j,\sigma)), \tag{2.4}$$

where (i,j) is the position of S2 unit. The output is a vector of N C2 values, where N corresponds to the number of patches. The vector is used as the C2 feature in the recognition task. In contrast to the conventional BIM, the PBIM takes the PSGHM way to refine the representation features while offering promising distinctiveness.

2.2. *Modified discrete Gaussian–Hermite moment review*

Let the standard Hermite polynomials of order p be defined as

$$H_p(x) = (-1)^p \exp(x^2)\frac{d^p}{dt^p}\exp(-x^2). \tag{2.5}$$

Hermite polynomials satisfy the following orthogonality relation with respect to the weight function $\exp(-x^2)$:

$$\int_{-\infty}^{\infty} \exp(-x^2)H_p(x)H_q(x)dx = 2^p p!\sqrt{\pi}\delta_{pq}, \tag{2.6}$$

where p and q are the orders of derivative, and δ_{pq} is the Kronecker delta.

To obtain the orthonormal version, the normalized Hermite polynomial is calculated by using Eq. (2.6) as

$$\hat{H}_p(x) = \frac{1}{\sqrt{2^p p! \sqrt{\pi}\sigma}} \exp\left(\frac{-x^2}{2}\right) H_p(x). \tag{2.7}$$

The Gaussian–Hermite moment (GHM)[23] was first introduced by Shen, which with the order (p, q) of the continuous image function $f(x, y)$ can be defined as

$$M_{p,q} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\hat{H}_p\left(\frac{x}{\sigma}\right)\hat{H}_q\left(\frac{y}{\sigma}\right) dxdy. \tag{2.8}$$

Discrete Gaussian–Hermite moment (DGHM) is a feature representation method for global image features, which means that it cannot be applied directly to local feature representation. To represent the local image features, a modified DGHM (MDGHM) was proposed.[22] In MDGHM, a movable mask is devised as

$$t(a,b)_{(i,j)} = I\left(\frac{a+i-m}{2+1}, \frac{b+j-n}{2+1}\right), \tag{2.9}$$

and the discrete Gaussian–Hermite functions of the mask are calculated as

$$\begin{cases} \bar{H}_p(x,\sigma) = \dfrac{2}{m-1}\hat{H}_p(x,\sigma) = \dfrac{2}{m-1}\dfrac{1}{\sqrt{2^p p!\sqrt{\pi}\sigma}}\exp\left(\dfrac{-x^2}{2\sigma^2}\right) H_p\left(\dfrac{x}{\sigma}\right), \\ \bar{H}_q(y,\sigma) = \dfrac{2}{n-1}\hat{H}_q(y,\sigma) = \dfrac{2}{n-1}\dfrac{1}{\sqrt{2^q q!\sqrt{\pi}\sigma}}\exp\left(\dfrac{-y^2}{2\sigma^2}\right) H_q\left(\dfrac{y}{\sigma}\right). \end{cases} \tag{2.10}$$

Therefore, the MDGHM at any point (i, j) on the input image I is given as

$$\bar{M}_{p,q}(i,j) = \frac{4}{(m-1)(n-1)}\sum_{a=0}^{m-1}\sum_{b=0}^{n-1} I\left(\frac{a+i-m}{2+1}, \frac{b+j-n}{2+1}\right)\bar{H}_p(x,\sigma)\bar{H}_q(y,\sigma), \tag{2.11}$$

where m and n are the size parameters of the mask, and a and b are the coordinates corresponding to the mask.

3. Enhanced BIM Model Based on PSGHM (PBIM)

The BIM model is an appearance-based descriptor that focuses on the invariance and selectivity of extracted features. Although conventional BIM is more flexible than some relevant descriptors[2,4] and its recognition performance is robust, it brings in huge numbers of redundant features by the random way, and results in heavy computing burden. In addition, the Gabor model has a high level of error in matching the experimental physiological data.[33] To improve the performance of BIM, we proposed a novel patch selection way on the OGHM (PSGHM), and we enhanced the BIM model by the PSGHM to refine the representational features, named as PBIM.

The stored patches in the C1 layers are the key components of the discriminative and robust abilities of BIM; thus, the construction of a proper patch set is very

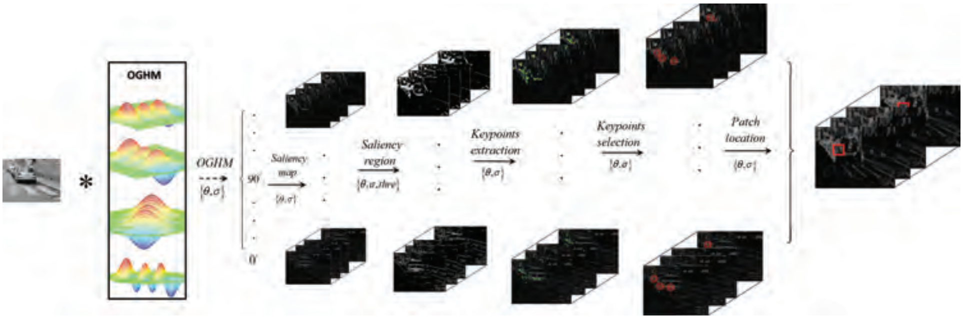


Fig. 1. Overview of PSGHM of the enhanced BIM model.

important for the visual recognition task. A random selection of patches from the universal training set is an option, but that option is prone to bringing in huge amounts of redundant information and is sensitive to rotation. We address this by proposing a novel PSGHM, which is based on a saliency mechanism and multi-scale keypoints on the OGHM layers. OGHM is a modified Gaussian–Hermite moment proposed by Lu *et al.*,[13] whose properties are effective against scale change, image rotation and illumination change.[13,27]

PSGHM consists of the following three steps: (1) processing layer extraction, (2) salient regions construction, (3) keypoint and patch localization. The overview of PSGHM of the enhanced BIM model is shown in Fig. 1.

3.1. *Processing layers*

Input images are processed by OGHM filters with different directions and scales. We obtain OGHM scale pyramids as per the method in Ref. 13. We make the directional multi-scale information tractable, by considering four orientations and 16 scales for further processing in BIM. The processing layers can be calculated by

$$M_{p,q}^{\theta}(i,j) = \frac{4}{(m-1)(n-1)} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I\left(\frac{i+u-m}{2+1}, \frac{j+v-n}{2+1}\right) \times H_p^{\theta}(X,\sigma) H_q^{\theta}(Y,\sigma), \tag{3.1}$$

where a digital image $I(i,j)$ whose size is $w \times h$ $[0 \le i \le w-1, 0 \le j \le h-1]$, $t(u,v)$ be a mask whose size is $m \times n$ $[0 \le u \le m-1, 0 \le v \le n-1]$, the oriented Gaussian–Hermite functions of the mask can be written as

$$\begin{cases} H_p^{\theta}(X,\sigma) = \dfrac{2}{m-1} \dfrac{1}{\sqrt{2^p p! \sqrt{\pi}\sigma}} \exp\left(\dfrac{-X^2}{2\sigma^2}\right) H_p\left(\dfrac{X}{\sigma}\right), \\ H_q^{\theta}(Y,\sigma) = \dfrac{2}{n-1} \dfrac{1}{\sqrt{2^q q! \sqrt{\pi}\sigma}} \exp\left(\dfrac{-Y^2}{2\sigma^2}\right) H_q\left(\dfrac{Y}{\sigma}\right). \end{cases} \tag{3.2}$$

$H_p(X/\sigma)$ and $H_q(Y/\sigma)$ is the Gaussian–Hermite function on X and Y; the rotation variables X and Y can be calculated by

$$\begin{cases} X = x\cos\theta + y\sin\theta, \\ Y = -x\sin\theta + y\cos\theta. \end{cases} \tag{3.3}$$

3.2. *Salient regions*

A huge amount of irrelevant information exists in the processing layers, which complicates locating the more discriminative regions in the whole image. Obtaining dense distinctive features requires the construction of a salient region with rich discriminative information. Based on a biological visual perception mechanism, attention is an important visual processing stage that guides the gaze towards objects of interest in a visual scene.[7] This ability to orientate towards salient objects in a cluttered visual environment is of great significance, because it allows rapid and accurate detection and tracking of prey or predators by organisms in the visual world. Itti and Koch first introduced a BIM to generate a saliency map.[8] In our paper, the saliency map is constructed in the processing layers based on a simple saliency model in Ref. 5.

The saliency map of the input image can be calculated as

$$\mathrm{Sal}(I) = |(F^{-1}(e^{R(f)+i\cdot P(f)}))^2|, \tag{3.4}$$

where F is the Fourier transform, f is frequency, $R(f)$ is the spectral residual, $P(f)$ is the phase spectrum of the image. For more details can refer to Ref. 25.

We inhibit the non-dominant information by adopting a simple version of the saliency map. We segment the constructed saliency map to obtain the salient region, i.e. where the distinctive features and patch extraction areas are concentrated. Given the saliency map of the input image, the salient region at location (x, y) can be obtained:

$$\mathrm{SR}(x, y) = \begin{cases} 1 & \text{if Sal } (I(x, y)) > \text{threshold}, \\ 0 & \text{otherwise}. \end{cases} \tag{3.5}$$

In general, we set *threshold* $= M(\mathrm{Sal}(I)) \times 2$, where $M(\mathrm{Sal}(I))$ is the mean value of every pixel in the saliency map. (PSGHM experimentally shows the best performance when *threshold* $= M(\mathrm{Sal}(I)) \times 2$, therefore we chose this value). The construction of the salient region is illustrated in Fig. 2.

3.3. *Keypoint candidate localization*

In the constructed salient regions, we locate the keypoint candidates in each layer with their corresponding directions. In the conventional BIM model, patches are randomly extracted from the overall C1 layers to form the vocabulary of visual features. However, these visual features are neither refined nor discriminative; they include irrelevant and redundant information and degrade performance. Achieving

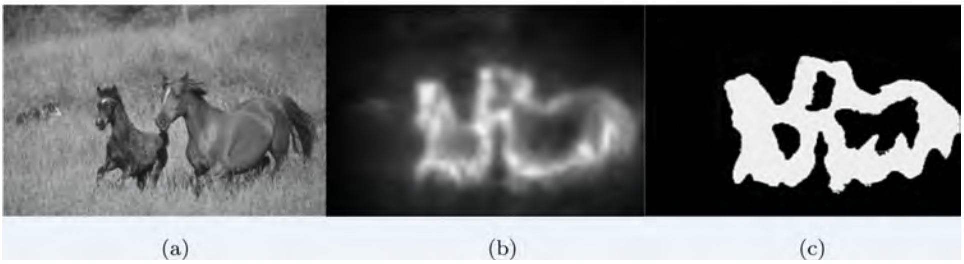


Fig. 2. Construction of the salient region: (a) original image, (b) saliency map of the input image, (c) salient region.

a reasonable recognition performance with BIM requires matching many patches, which results in a high computational cost. PSGHM locates the keypoint candidates within the salient region, which are identified by a keypoint detection method named FAST.[25] FAST is widely used because of its accuracy and speed; however, it does not have an orientation component nor does it produce multi-scale features. Hence, we employ processing layers that are processed by Gabor scale pyramids at certain angles and produce FAST keypoints at each layer. In this way, we extract multi-scale keypoint candidates with specific angles. The keypoint candidate position *key* can be localized by

$$\text{key} = \text{FAST}(P_{\theta,\sigma}(x,y)), \quad (x,y) \in \text{SR}. \tag{3.6}$$

Here, FAST is the keypoint detection method, $P_{\theta,\sigma}$ denotes the processing layer with orientation θ and scale σ, (x,y) are pixel coordinates in the layer and SR is the salient region. We preferentially extract image patches around these detected keypoint candidates.

4. Experiments

We evaluate the performance of PBIM in several recognition tasks. In Sec. 4.1, we give the experiment setup. In Sec. 4.2, we evaluate the PBIM model under conditions of under normal circumstances using three datasets (Caltech5, TUD and GRAZ01).

4.1. *Experiment setup*

Given the various appearance transformation of the images, we applied the position-scale-invariant C2 features of PBIM, and passed the features to a classifier to execute classification. (In the experiments of this paper, we select the linear Lib-SVM[3] as the classifier). The other layers of PBIM are similar to those of the standard BIM, except for the obtained OGHM-based features in the S1 layers. We chose the evaluation metrics classification accuracy (acc), recall (r) and 1-precision$(1-p)$.

$$1 - p = \frac{\text{FP}}{\text{TP} + \text{FP}}, \tag{4.1}$$

$$r = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{4.2}$$

$$\mathrm{acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}, \tag{4.3}$$

where true positive (TP) is a correct classification of a positive (an object or scene), true negative (TN) is a correct classification of a negative (background), false positive (FP) is an incorrect positive classification, false negative (FN) is an incorrect negative classification.

4.2. *Experiment evaluation*

To evaluate the performance of PBIM, we compared it with that of other related algorithms on three public image datasets: CalTech5,[26] TUD[1] and GRAZ01.[21]

4.2.1. *Caltech*5

The CalTech5 dataset contains the cars, frontal faces, aeroplanes, leaves and motorcycles, as shown in Fig. 3. We applied this database to evaluate BIHM and make comparisons with the conventional BIM and the SIFT algorithm.[12]

To make the experiment at a feature level and ensure a fair comparison between the methods, we compared the scale- and position-invariant C2 features produced by the standard BIM, and PBIM with SIFT features by passing the features to an SVM, which was trained to perform the object present/absent recognition task. We chose the classification rate for various numbers of features as the evaluation criterion. In the experiment, we randomly chose 15 images from each category of the CaltTech5 dataset as positive training images and 15 images from backgrounds as

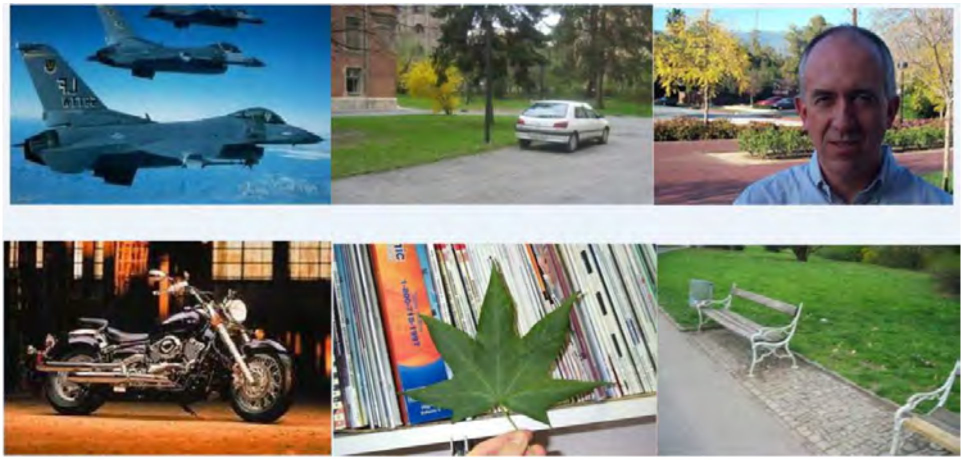
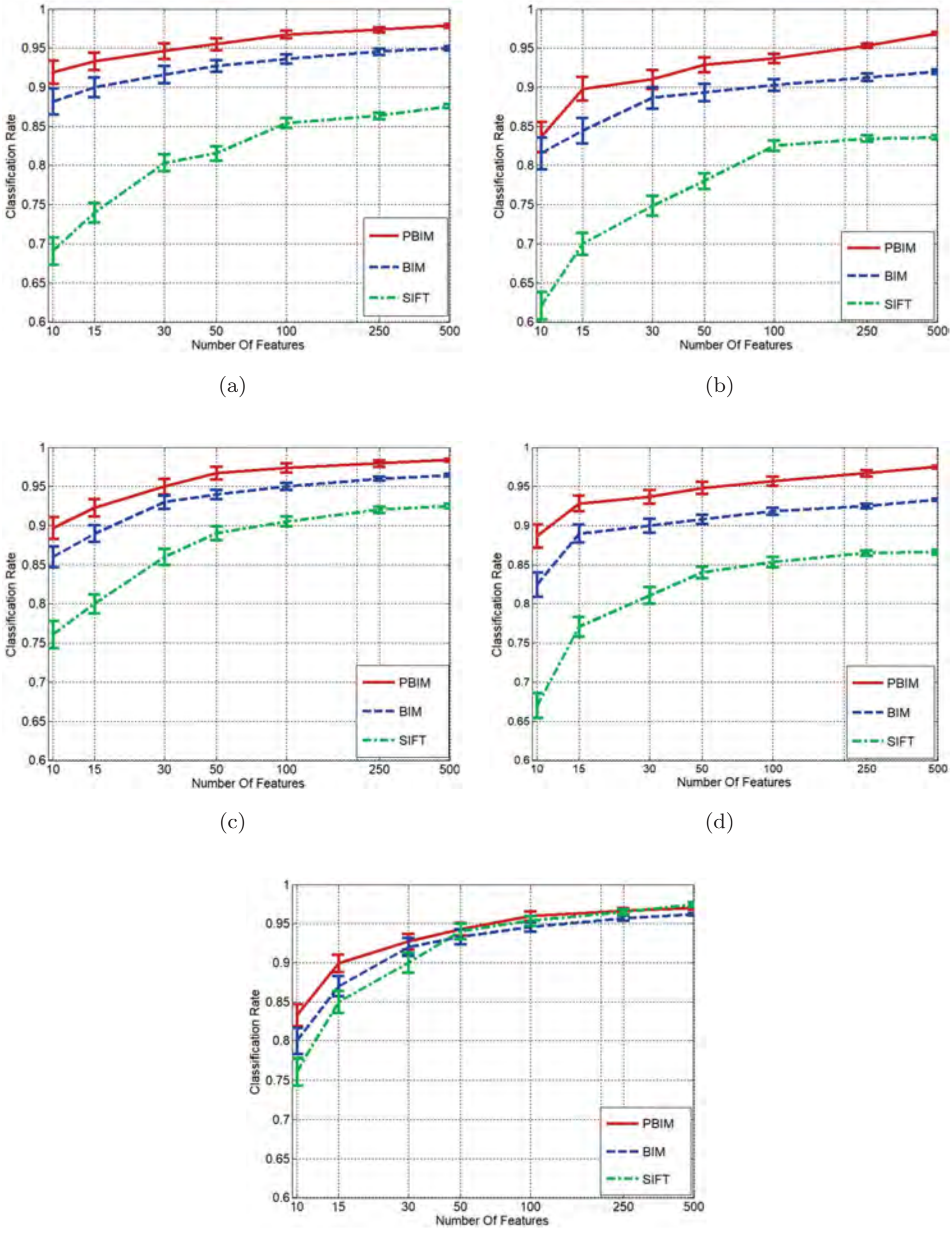


Fig. 3. Sampling images of the CalTech5 dataset. The last image is a background image.



(a)

(b)

(c)

(d)

(e)

Fig. 4. Comparison of PBIM with SIFT and standard BIM on the CalTech5 database: (a) airplanes, (b) faces, (c) cars, (d) leaves and (e) motorcycles.

the negative training set. For the tests, other images (each category of the CaltTech5 dataset) and 50 other images (backgrounds) was randomly chosen as a testing set. It should be noted that a different number of features were randomly chosen from the C2 layers and SIFT features set (the SIFT features were obtained as in Ref. 26) to train the models.

Figure 4 shows the simulation results on the CalTech5 dataset for different numbers of features. In general, it has been shown that PBIM outperforms BIM and SIFT in terms of accuracy for the most categories in the dataset. PBIM and BIM significantly outperform SIFT for the airplanes, faces, leaves and cars; for the airplanes and leaves, PBIM is clearly superior to BIM, whereas for the faces and cars, PBIM can be competitive with BIM; PBIM did not achieve superior results in the motorcycles test.

4.2.2. *TUD*

The TUD database (formerly the ETHZ database) contains side views of cars, motorcycles and cows, as shown in Fig. 5. We evaluated the PBIM model, the conventional BIM that uses the random patch selection method, and a modified biologically inspired model (MBIM) based on OGHM with random patches.[13] In addition, we also compared SIFT[12] and spatial pyramid matching (SPM) using sparse coding[1] in the experiment.

To make the comparison at the feature level, we compared the scale and position-invariant C2 features of BIM models with the features produced by SIFT and

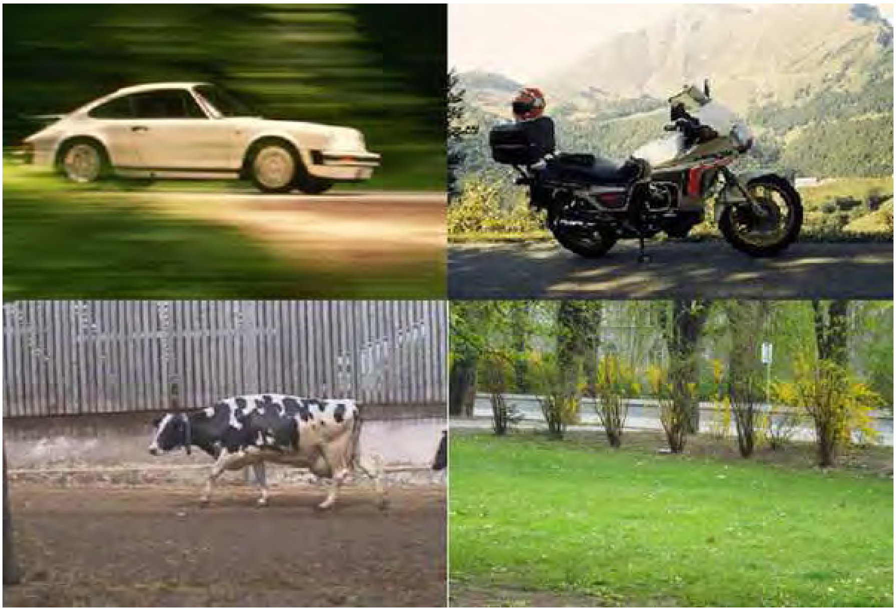
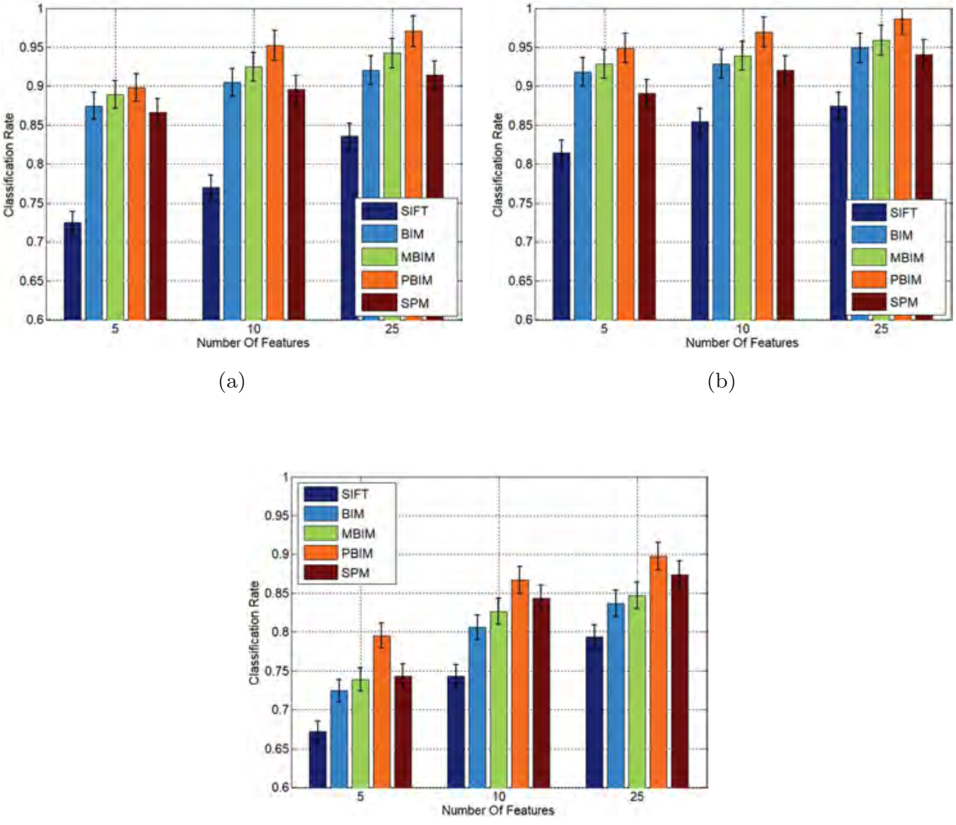


Fig. 5. Sampling images from the TUD dataset. The last image is a background image.

SPM, by passing them to a linear SVM that was trained to perform the object present/absent recognition task. We compared the classification rate for various numbers of features (5, 10 and 25). In the experiment, we randomly chose 15 images from each category of the TUD database as positive training images and 15 background images as the negative training set. For the tests, 50 images from each category of the TUD dataset and 50 images from backgrounds were randomly chosen as a test set. The results were generated from 10 independent trials. We report the mean and standard deviation of the classification across all classes.

Figure 6 shows the simulation results on the TUD dataset for different numbers of features. In general, PBIM clearly outperforms SIFT, SPM, BIM and MBIM in terms of accuracy for most of the categories in the dataset. In particular, PBIM significantly outperforms the other methods for the cars and cows.



(a)

(b)

(c)

Fig. 6. Comparison of PBIM with standard BIM, MBIM, SIFT and SPM on the TUD database: (a) cows, (b) cars and (c) motorcycles.

4.2.3. *GRAZ*-01

GRAZ-01[3] is a challenging dataset with high intra-class variability on highly cluttered backgrounds, containing persons, bikes and backgrounds. The sampling images of the GRAZ-01 dataset are shown in Fig. 7. For the GRAZ-01 dataset, we followed the method presented in Ref. 3: 100 images (bike or person) and 100 images (backgrounds) were randomly chosen as the training set; 50 other images (bike or person) and other images (backgrounds) were chosen as the testing set. Fifteen hundred initial patches (features) were used for the experiment. We repeated the experiment 10 times and reported the averaged values of the test results. For effective evaluation of the PBIM model, we also tested the ROC and recall-precision (RP) curves and compared the performance of the proposed model with that of the

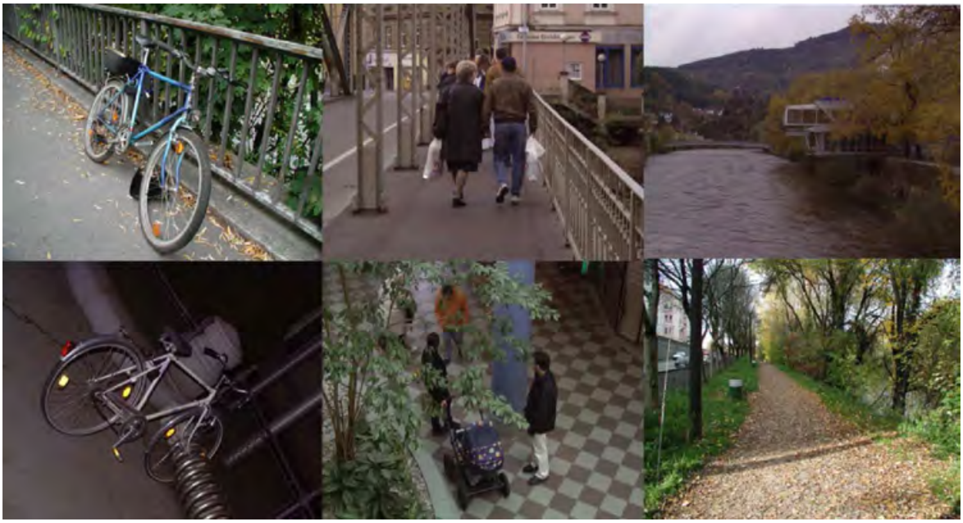


Fig. 7. Sampling images of GRAZ-01. From left to right, the categories are bikes, people and backgrounds.

Table 1. Performance comparison of several approaches on GRAZ-01.

	Bikes		Persion	
Method	EER	AUC	EER	AUC
Moment invariants	73.5	76.5	63.0	68.7
SIFT	78.0	86.5	76.5	80.8
BoW	80.3	87.2	79.6	82.3
SPM	81.1	87.9	77.4	79.5
BIM	82.4	88.5	75.5	83.3
SM	83.5	89.6	56.5	59.1
MBIM	84.3	91.2	76.8	85.3
PBIM	85.5	94.3	86.9	92.7

related approaches (i.e. moment invariants, SIFT, similarity-measure-segmentation (SM), BoW, SPM, BIM, modified biologically inspired model (MBIM)).[1,3,13,21] The experimental GRAZ-01 dataset results are shown in Table 1.

Table 1 shows the ROC curves results: PBIM achieves the best performance in all cases. PBIM by far outperforms the moment invariants, SM, BoW, SIFT and SPM approaches for both bikes and persons. The performance of PBIM are similar to that of the MBIM method at the bike cases; however, PBIM significantly outperforms the MBIM method at the person recognition tasks. In general, our proposed model achieves competitive results.

5. Conclusion

In this paper, we presented an OGHM-based patch selection method (PSGHM), and extended the BIM model with the PSGHM method. The OGHM-based features have properties that are robust to in-image distortions, including rotation. The proposed PBIM model increases the rotation invariance for local feature representation and refine the selected patches. PBIM provides a better balance between selective representation and invariance. Experiments on three different datasets demonstrated significant improvements as compared to the conventional BIM. Our work thus far has focused mainly on the low layers of BIM. Enhancing a deeper hierarchy of features will constitute our future work.

Acknowledgments

The authors Lu and Jia Contributed equally to this work. This work was supported by the National Science Foundation of China (Grant No. 61603389) and partially supported by the National Natural Science Foundation of China (Grant Nos. 61502494 and 61210009) and also by the Strategic Priority Research Program of the CAS (Grant No. XDB02080003) and the Development of Science and Technology of Guangdong Province Special Fund Project (Grant No. 2016B090910001).

References

1. L. Bastian, A. Leonardis and B. Schiele, Combined object categorization and segmentation with an implicit shape model, in *Proc. 8th Workshop on Statistical Learning in Computer Vision*, ECCV (Springer, Prague, 2004), pp. 17–32.
2. H. Bay, A. Ess, T. Tuytelaars and L. Van, Speeded-up robust features (SURF), New method for feature extraction based on fractal behavior, *Comput. Vis. Image Underst.* **110**(3) (2008) 346–359.
3. C. Chang and C. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol.* **2**(3) (2011) 27.
4. N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (IEEE, San Diego, 2005), pp. 886–893.
5. X. Hou and L. Zhang, Saliency detection: A spectral residual approach, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (IEEE, Minneapolis, 2007), pp. 1–8.

6. D. Hubel and T. Wiesel, Receptive fields of single neurons in the cat's striate cortex, *J. Physiol.* **148**(3) (1959) 574–591.
7. L. Itti and C. Koch, Computational modelling of visual attention, *Nat. Rev. Neurosci.* **2**(3) (2001) 194–203.
8. L. Itti, C. Koch and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11) (1998) 1254–1259.
9. T. Joachims, Text categorization with support vector machines, in *Learning with Many Relevant Features* (Springer, Berlin-Heidelberg, 1998), pp. 137–142.
10. T. Kang *et al.*, Enhanced SIFT descriptor based on modified discrete Gaussian–Hermite moment, *ETRI J.* **34**(4) (2012) 572–582.
11. T. Leung and J. Malik, Representing and recognizing the visual appearance of materials using three-dimensional textons, *Int. J. Comput. Vis.* **43**(1) (2001) 29–44.
12. D. Lowe, Distinctive image features from scale-invariant keypoints, New method for feature extraction based on fractal behavior, *Int. J. Comput. Vis.* **60**(2) (2004) 91–110.
13. Y. Lu *et al.*, Extended biologically inspired model for object recognition based on oriented Gaussian–Hermite moment, *Neurocomputing* (2014) 189–201.
14. Y. Lu *et al.*, Enhanced hierarchical model of object recognition based on a novel patch selection method in salient regions, *IET Comput. Vis.* **9**(5) (2015) 663–672.
15. Y. Lu *et al.*, Dominant orientation patch matching for HMAX, *Neurocomputing* **193** (2016) 242–250.
16. B. Ma, X. Chai and T. Wang, A novel feature descriptor based on biologically inspired feature for head pose estimation, *Neurocomputing* **115** (2013) 1–10.
17. R. Manduchi, A. Castano, A. Talukder and L. Matthies, Obstacle detection and terrain classification for autonomous off-road navigation, *Auton. Robots* **18**(1) (2005) 81–102.
18. K. Mikolajczyk and C. Schmid, A performance evaluation of local descriptors, *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(10) (2005) 1615–1630.
19. P. Nagabhushan, D. Guru and B. Shekar, An efficient approach for appearance based object recognition, *Neurocomputing* **69**(7) (2006) 934–940.
20. B. Olshausen, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature* **381**(6583) (1996) 607–609.
21. A. Opelt, A. Pinz, M. Fussenegger and P. Auer, Generic object recognition with boosting, *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(3) (2006) 416–431.
22. H. Qiao *et al.*, Biologically inspired visual model with preliminary cognition and active attention adjustment, *IEEE Trans. Cybern.* **45** (2015) 2612–2624.
23. H. Qiao *et al.*, Human-inspired motion model of upper-limb with fast response and learning ability — A promising direction for robot system and control, *Assem. Autom.* **36**(1) (2016) 97–107.
24. M. Riesenhuber and T. Poggio, Hierarchical models of object recognition in cortex, *Nat. Neurosci.* **2**(11) (1999) 1019–1025.
25. E. Rosten, R. Porter and T. Drummond, Faster and better: A machine learning approach to corner detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(1) (2010) 105–119.
26. T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio, Robust object recognition with cortex-like mechanisms, *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(3) (2007) 411–426.
27. J. Shen, Orthogonal Gaussian–Hermite moments for image characterization, in *Proc. Intelligent Systems & Advanced Manufacturing*, International Society for Optics and Photonics (SPIE, USA, 1997), pp. 224–233.

28. L. Shen, R. Rangayyan and J. Desautels, Detection and classification of mammographic calcifications, *Int. J. Pattern Recognit. Artif. Intell.* **7**(6) (1993) 1403–1416.
29. Y. Tang and X. You, Skeletonization of ribbon-like shapes based on a new wavelet function, *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(9) (2003) 1118–1133.
30. Y. Tang, L. Yang and J. Liu, Characterization of dirac-structure edges with wavelet transform, *IEEE Trans. Syst. Man. Cybern. B Cybern.* **30**(1) (2000) 93–109.
31. Y. Tang, T. Yu and L. Ernest, New method for feature extraction based on fractal behavior, *Pattern Recognit.* **35**(5) (2002) 1071–1081.
32. R. Young, The Gaussian derivative model for spatial vision: I. Retinal mechanisms, *Spatial Vis.* **2**(4) (1987) 273–293.
33. R. Young and R. Lesperance, The Gaussian derivative model for spatial-temporal vision: I. Cortical model, *Spatial Vis.* **14**(3) (2001) 261–319.
34. T. Zhang, B. Fang, Y. Y. Tang, G. He and J. Wen, Topology preserving non-negative matrix factorization for face recognition, *IEEE Trans. Image Process.* **17**(4) (2008) 574–84.
35. T. Zhang, B. Fang, Y. Y. Tang, Z. Shang and B. Xu, Generalized discriminant analysis: A matrix exponential approach, *IEEE Trans. Syst. Man Cybern. B Cybern.* **40**(1) (2010) 186.
36. T. Zhang, B. Fang, Y. Yuan, Y. Y. Tang, Z. Shang, D. Li and F. Lang, Multiscale facial structure representation for face recognition under varying illumination, *Pattern Recognit.* **42**(2) (2009) 251–258.
37. T. Zhang, Y. Y. Tang, B. Fang, Z. Shang and X. Liu, Face recognition under varying illumination using gradientfaces, *IEEE Trans. Image Process.* **18**(11) (2009) 2599–2606.
38. H. Zhang *et al.*, B-HMAX: A fast binary biologically inspired model for object recognition, *Neurocomputing* **218** (2016) 242–250.