Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **6**, November **2018**

Bayesian-Based Probabilistic Architecture for Image Categorization Using Macro- and Micro-Sense Visual Vocabulary

Chaur-Heh Hsieh[1], Chung-Ming Kuo[2*], Yao-Sheng Hsieh[2]

[1]Information Engineering College, Yango University
Mawei, Fuzhou, China

[2]Department of Information Engineering, I-Shou University
No.1, Sec. 1, Syuecheng Road, Dashu Township 840, Kaohsiung, Taiwan, China.
*corresponding author, e-mail: kuocm@isu.edu.tw

Received September, 2018; revised October, 2018

Abstract. *Visual vocabulary representation approach has been successfully applied to many multimedia and vision applications, and we have been developed a novel visual vocabulary with macro-based and micro-based visual words in previous work. In this work, we will present a category-specific visual model for image categorization based on the above-mentioned visual vocabulary. The category-specific visual model is composed of macro and micro visual words description, respectively. Because the image contains macro and micro contents and they are exclusive each other, we can categorize image by considering macro or micro content in a flexible way. In our work, we will propose a Bayesian-based probabilistic method that achieves effective and excellent image categorization. The performance evaluation for the proposed systems indicates that the new categorization scheme achieves promising results.*
Keywords: Visual vocabulary, Category-specific, Categorization, Bayesian-based

1. **Introduction.** The rapid growth of Internet-based services makes many multimedia applications available on the Internet for users to access. The tools for content-based image processing are important for many applications such as image browsing/retrieval [1-3], intelligent vehicle/robot navigation and image/object recognition [4-6], therefore the related studies have received much attention in recent years. For image categorization, it generally requires to automatically classify images into a limited number of categories with semantic label [7-11]. However, images are usually composed of various entities in any possible layout. Therefore, image categorization is a difficult and challenging issue [12]-[20].

The categorization schemes widely use the distribution of low-level features and assumes that the categorization of images always behave similarity in selected features domain [18]. However, due to the variety and complexity of image contents, two images with large similar regions may have very different interpretation and thus fall into different categories. The working principle of human visual system is most likely an integrator. Thus, the whole image interpretation is always with higher priority. Thus, the system perceives a scene from coarse (global) to fine (local). In Fig. 1, these pictures will have very similar interpretation for human, although they are actually not the same. In order to consider this property, the global and fine content should be modeled independently.

In [14], we developed a novel visual vocabulary with macro-based and micro-based visual words. We also presented an effective image description method based on the

FIGURE 1. Similar macro sense (Grass land) images with different local details

new visual vocabulary. According to extensive simulation, the visual vocabulary achieves excellent results for image retrieval, thus it can effectively extract the visual features from images.

For image categorization, in this work we will present a category-specific visual model for each category based on the above-mentioned visual vocabulary. The category-specific visual model is composed of macro and micro visual words description, respectively. Because the image contains macro and micro contents and they are exclusive each other, we can categorize image by considering macro or micro content in a flexible way. In our work, we will propose a Bayesian-based probabilistic method that achieves effective and excellent image categorization.

2. **The fundamental architecture for image categorization.** The structure of the proposed image categorization method is illustrated in Fig. 2. It contains two main components: visual vocabulary construction and image categorization. The scheme of visual vocabulary construction is similar to our previous work [14]. We briefly review the scheme as follows. The developed visual vocabularies are with macro and micro sense of visual words based on the characteristics of homogeneity and completeness. For macro sense, the words represent the whole and global sense of visual perception and with completeness of meaning. On the contrary, the micro sense visual word usually represents the detail of image content and with incomplete meaning. Each vocabulary is homogeneous in words content, so it can effectively represent an image according to its content, and thus it is effective in retrieving and categorization. For more details, please refer to [14].

In this paper, we will focus on the algorithm of image categorization. There are two phases in the proposed categorization algorithm; one is to construct category-specific model for each class, and the other is to design a Bayesian-based probabilistic classifier. In the following section, we will describe the procedures in details.

2.1. **Construction of category-specific model for each category.** We aim at constructing a category-specific model, which effectively represents the images of each class in a compact way. The model contains important features including macro and micro visual words and the proportion of macro and micro content for each category. The

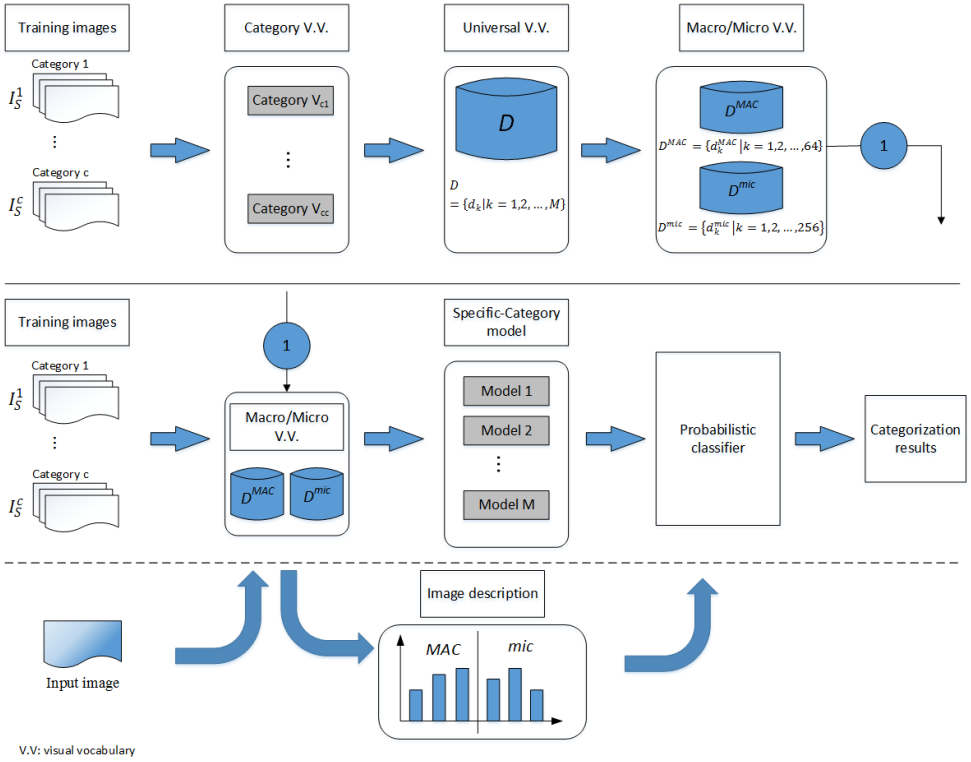

FIGURE 2. The structure of proposed image categorization

category-specific model represented as

$$\psi(i) = \left\{ H_b^{\psi(i)}(m_b), P_b^i\% \;\middle|\; \begin{array}{c} i = 1...L,\ m_0 = 1...N_{MAC}\ or\ m_0 = 1...N_{mic} \\ b = 0,1,\ 0 = Macro,\ 1 = Micro \end{array} \right\} \tag{1}$$

where $\psi(i)$ is the model of category i, and the macro or micro feature description for category i are represented by $H_b^{\psi(i)}(m_b)$, where the subscripts b=0, or 1 represents the macro and micro feature respectively. The proportion of macro and micro content in percentage for category i is represented by $P_b^i\%$. In our work, the macro/micro description is the histogram of visual words defined as

$$H_b^{\psi(i)}(m_b) = \left[(d_b^i(k_b), d_b^i(k_b)\%), k_b = 0,1,...,N_b \right] \tag{2}$$

where $d_b^i(k_b)$ is the kth visual word in macro or micro visual vocabulary and $d_b^i(k_b)\%$ is the probability of the visual word appearance. On the other hand, the overall macro and micro content in each category is also a very important property. We can easily find that the background and details in various classes are usually very different. Thus, the proportion of overall background and details in each class is an important feature, and we use $P_b^i\%$ to represent this feature. The construction of category model is illustrated in Fig. 3. [1]

[1]Note: In Fig. 3, the symbol $L(B_s^c(n))$ is the label of each input block using visual vocabulary, and can be expressed as $L(B_s^c(n)) = k* = \underset{k}{\text{argmin}}(\|B_s^c(n) - d_k\|), k = 1,...C_M, n = 1,...N$, where $B_s^c(n)$ is the input block, i.e., macro or micro, and d_k is the visual word in visual vocabulary. $\sum_{n=1}^{N_{MAC}} \delta(L(B_s^{MAC}(k) - n))$ denotes the label histogram with macro-based class model; and $\sum_{n=1}^{N_{mic}} \delta(L(B_s^{mic}(k) - n))$ is the label histogram with micro-based class model.

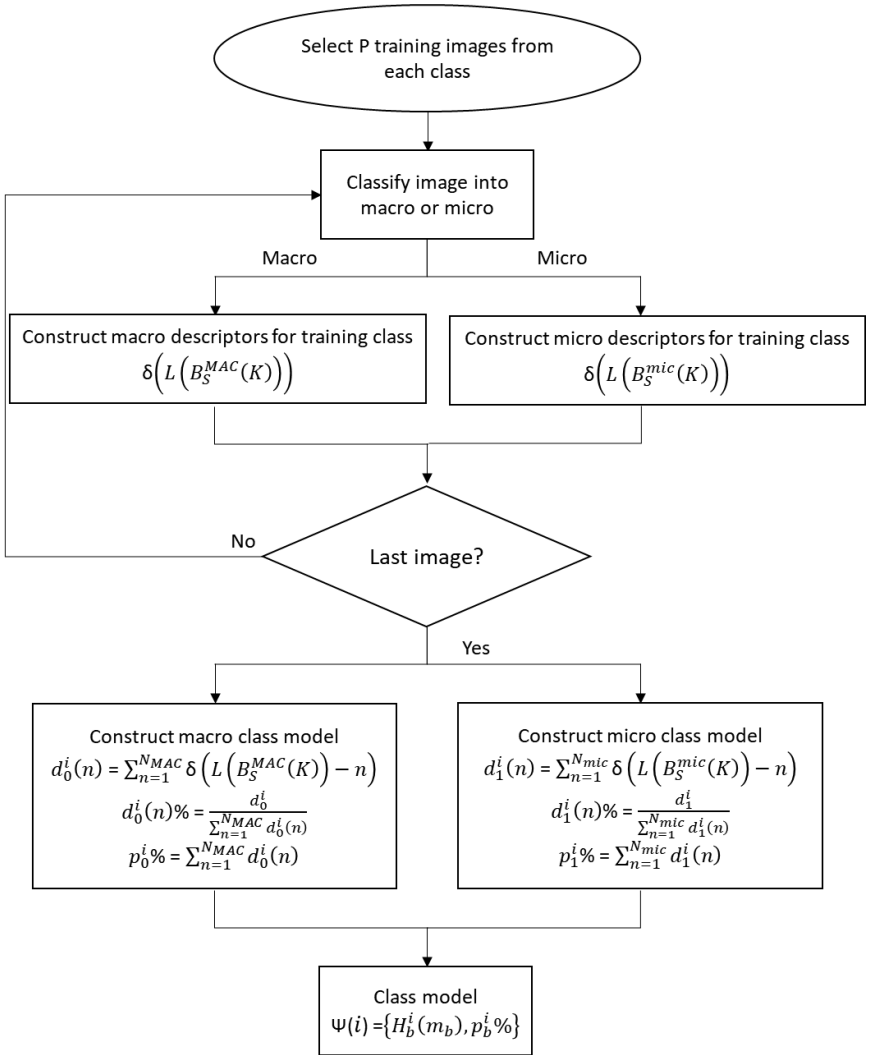

FIGURE 3. [1]The construction of categorization model

In the training procedure, the number of training samples should be carefully considered. Because the size of visual vocabulary is fixed, the probability of visual words will increase when the number of the training samples increases. Consequently, the uniqueness of image class perhaps decreases. This will significantly degrade the representativeness of category model. Therefore, the number of training samples should be restricted to a limited number. In our work, the number of training images for each class is set to 5. On the other hand, for considering the uniqueness and representativeness further, we need to filter out the insignificant visual words to maintain the representativeness of category model. In our work, we will filter out the visual words of low appearance rate to decrease the ambiguity of class model. We define the average occurrence rate T as follows.

$$T_{MAC} = \frac{1}{N_{MAC}} \sum_{k=1}^{N_{MAC}} d_{MAC}^i(k)\% \tag{3}$$

$$T_{mic} = \frac{1}{N_{mic}} \sum_{k=1}^{N_{mic}} d_{mic}^i(k)\% \tag{4}$$

The visual words whose appearance rates are below the threshold will be filtered out in the training phase.

The training procedure is shown in Fig. 4. It is worth mentioning that the training is performed one image at a time. Therefore, the filtering process can not only remain the common characteristics for the image category but also keep the uniqueness of each image in same class.

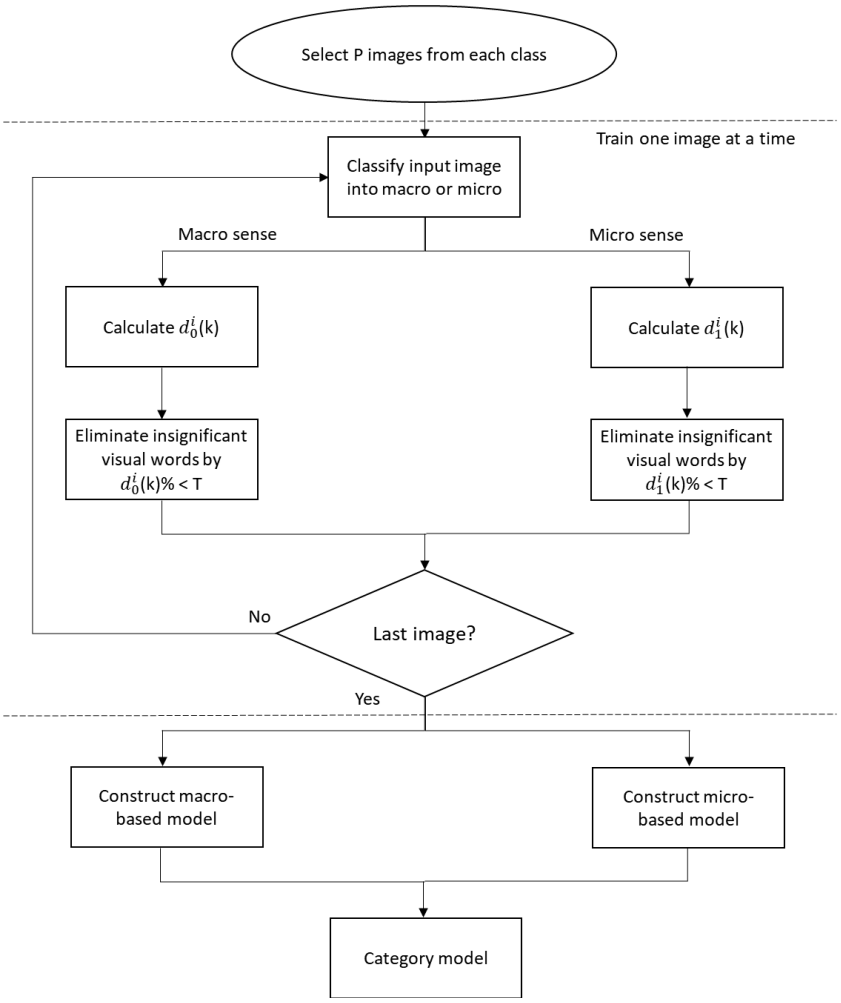

FIGURE 4. The training procedure for image class model

3. **Bayesian-based probabilistic approach for image categorization.** For image categorization, we propose a maximum a posteriori (MAP) approach to achieve accurate categorization. As mentioned above, the image is composed of macro and micro contents, therefore we express the image as $I = \{I_{MAC}, I_{mic}\}$, where I_{MAC}, I_{mic} represent the macro and micro contents in the image I, respectively. Since the two contents are mutually exclusive, i.e., $I_{MAC} \cap I_{mic} = \phi$, the similarity of images for macro content and micro content should be calculated separately. The description for image I is expressed as $H(I) = \left\{H_0^I(m_0), H_1^I(m_1)\right\}$, where $H_b^I(m_b)$, b=0 or 1, are macro and micro description for the image, respectively. The category of image I is obtained by MAP probabilistic

model as

$$\begin{aligned} Cate(I) &= \arg\max_{\Psi(i)} P(\Psi(i)\,|I) \\ &= \arg\max_{\Psi(i)} \left[P\left(H_0^{\Psi(i)}(m_0)\,\middle|H_0^I(m_0)\right) + P\left(H_1^{\Psi(i)}(m_1)\,\middle|H_1^I(m_1)\right)\right] \\ &= \arg\max_{\Psi(i)} \left[\begin{array}{l} P\left(H_0^I(m_0)\,\middle|H_0^{\Psi(i)}(m_0)\right) P\left(H_0^{\Psi(i)}(m_0)\right) + \\ P\left(H_1^I(m_1)\,\middle|H_1^{\Psi(i)}(m_1)\right) P\left(H_1^{\Psi(i)}(m_1)\right) \end{array} \right] \end{aligned} \tag{5}$$

where $P\left(H_0^I(m_0)\,\middle|H_0^{\Psi(i)}(m_0)\right)$ and $P\left(H_1^I(m_1)\,\middle|H_1^{\Psi(i)}(m_1)\right)$ are maximum likelihood functions, which are respectively used to measure the similarity of macro and micro contents between image and category models. The maximum likelihood functions can be calculated by their distances. We adopt the distance measure [29][19] as

$$\begin{aligned} &P\left(H_b^I(m_b)\,\middle|H_b^{\Psi(i)}(m_b)\right) \\ &= \sum_{j=1}^{M}\left[\left(1-\left|H_b^I(j)-H_b^{\Psi(i)}(j)\right|\right)\right] \times \min\left(H_b^I(j), H_b^{\Psi(i)}(j)\right) \end{aligned} \tag{6}$$

For the priori probability, we let $P\left(H_b^{\Psi(i)}(m_b)\right) = P_b^{\Psi(i)}\%$, which is the proportion of macro or micro content in the class i. The proportion is a dominant factor for categorization. Even though the similarity is high for macro or micro content, however the small proportion will degrade the overall similarity. As shown in Fig. 5, to categorize an image into a specific class we shall consider not only the content similarity but also the proportion of macro and micro content for each category. Therefore, the proposed method can accurately measure the similarity between image and category model in a reasonable way.

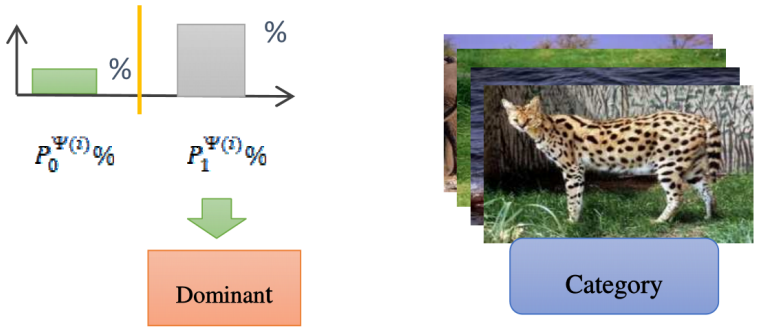

FIGURE 5. The illustration of the proportion of macro and micro content

Finally, the categorization should also consider the weighting of macro and micro content according to the similarity of macro/micro content between image and category model. In Eq. (5), the categorization strategy is out of consideration for one important factor that is the similarity of the proportion between categorized image and class model. To consider this factor, Eq. (5) is modified into

$$Cate(I) = \arg\max_{\Psi(i)} \left[w_{MAC} \times P\left(H_0^{\Psi(i)}(m_0)\,\middle|H_0^I(m_0)\right) + w_{mic} \times P\left(H_1^{\Psi(i)}(m_1)\,\middle|H_1^I(m_1)\right)\right] \tag{7}$$

where the weights wMac and wmic are calculated by $w_{MAC} = \min\left(P_0^{\Psi(i)}\%, I_{MAC}\%\right)$, $w_{mic} = \min\left(P_1^{\Psi(i)}\%, I_{mic}\%\right)$.

Fig. 6 is used to illustrate the modification. The overall similarity between image and category model should consider not only the content details but also their overall macro and micro content occupation.

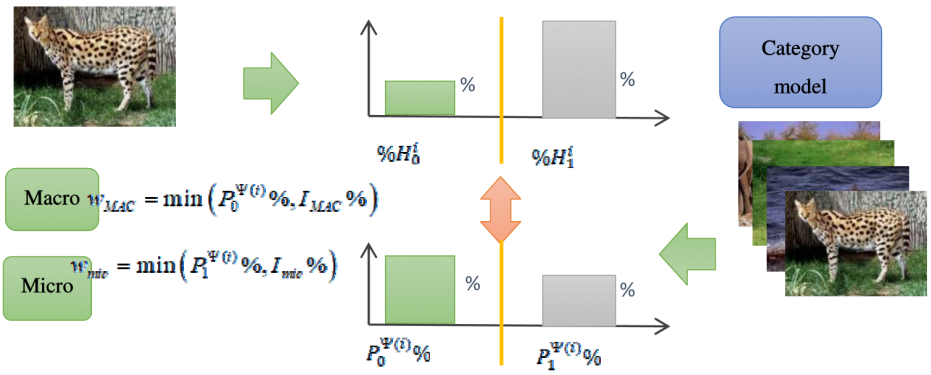

FIGURE 6. The determination of the weighting value

The weighting factor is very important for improving categorization accuracy. We use an example to demonstrate the importance. In Fig. 7, the proportions of macro and micro for category image A are 80% and 20%, respectively. And the proportion of macro and micro for the category B are 20% and 80%, separately. Assume that the macro similarity of between class model A and the image is 90%, and micro similarity is 30%. According to Eq. (5), the total similarity is equal to 0.9*0.2+0.3*0.8= 0.42 (42%). Although the macro proportion is smaller, it dominants the overall similarity. If the weighting factors are included, the overall similarity changes to 0.2*0.42=0.084(8.4%). It is obvious that the categorization accuracy is improved.

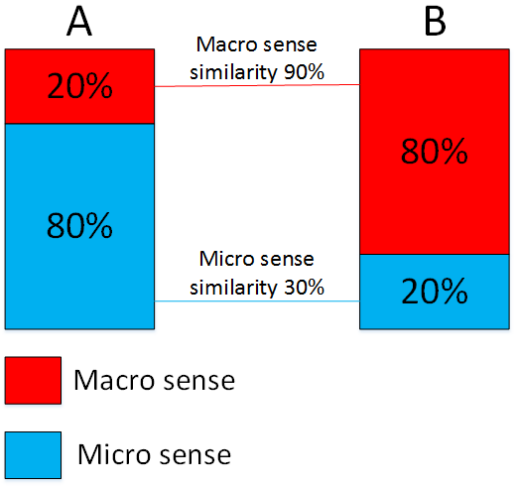

FIGURE 7. The demonstration of weighting value

4. **Experimental Results.** We use a database (31 classes, 3901 images) from Corel's photo to test the performance of the proposed method. The database contains a variety of images; the example images and the number of images for each class are shown in [35][18]. We selected 300 test images, as demonstrated in Fig. 8. There are five main categories, in which global color is the dominant feature for discrimination of the categories. We select

60 images in each main category and uniformly divide each main category into three sub-categories, thus yielding 20 images per sub-category. For fairly evaluation, five images are randomly selected from each sub-category as training images. We use accuracy rate as the objective criterion, which is defined as

$$accuracy = \frac{number\ of\ successfully\ categoried\ images}{total\ umber\ of\ images\ in\ category} \tag{8}$$

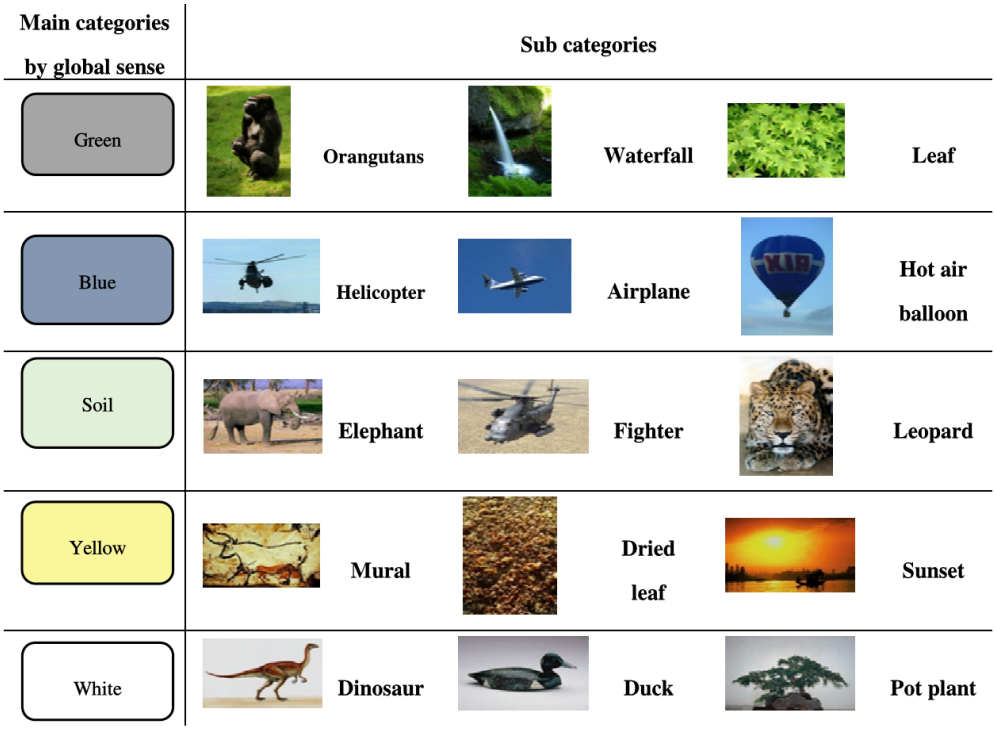

FIGURE 8. The example sample images: five main categories, and each includes three sub-categories

In the following, several experiments will be conducted to evaluate the efficiency and effectiveness of proposed method. We summary the categorization results for main categorizes and sub-categorizes in Table 1 and Table 2. For different quantities of testing samples, the categorization performance is different for the setting of macro and micro content ratio. For examples, in Table 1, the performance is varied from 0.8533 to 1 according to different macro and micro content ratio. In Table 2, the variation is more significant due to the large amount testing samples. However, the proposed method can always achieve the performance to the best one. Obviously, the proposed method detects and determines the correct macro and micro content ratio automatically and thus optimize the performance. Similar results are also observed in categorization of sub-categories, as shown in Tables 3 and 4. The proposed method can precisely describe micro and macro feature, meanwhile it can also appropriately adjust the weighting factor of macro and micro content according to dominant image class. The whole procedure does not need human intervention, thus we can conclude that the proposed method is very effective for image categorization.

Finally, we use some visual examples to explain the limitation of the proposed categorization method. Fig. 9 lists the wrongly categorized images for both macro-based and micro-based categories. We can find that the wrongly categorized images have very similar category features. For the example in Fig. 10 (a), leaf and waterfall images are similar

TABLE 1. Categorization performance for macro-based category (30 Images/category)

30 images/class	Results					
methods / Class*	Proposed method	Macro : Micro (0.1 : 0.9)	Macro : Micro (0.3 : 0.7)	Macro : Micro (0.5 : 0.5)	Macro : Micro (0.7 : 0.3)	Macro : Micro (0.9 : 0.1)
(A)	30	25	30	30	30	30
(B)	30	30	30	30	27	12
(C)	30	23	29	30	29	29
(D)	28	20	26	30	30	27
(E)	30	30	30	30	30	30
Correctness	0.986667	0.853333	0.966667	1	0.973333	0.853333

TABLE 2. Categorization performance for macro-based category (60 Images/category)

60 images/class	Results					
methods / Class*	Proposed method	Macro : Micro (0.1 : 0.9)	Macro : Micro (0.3 : 0.7)	Macro : Micro (0.5 : 0.5)	Macro : Micro (0.7 : 0.3)	Macro : Micro (0.9 : 0.1)
(A)	55	32	47	56	58	58
(B)	56	60	56	56	46	14
(C)	56	27	47	57	58	55
(D)	56	28	40	54	53	38
(E)	54	58	57	54	51	51
Correctness	0.923333	0.683333	0.823333	0.923333	0.886667	0.72

TABLE 3. Categorization performance for micro-based category (10 Images/category)

10 images/class		Results					
Methods / Input		Proposed method	Macro : Micro (0.1 : 0.9)	Macro : Micro (0.3 : 0.7)	Macro : Micro (0.5 : 0.5)	Macro : Micro (0.7 : 0.3)	Macro : Micro (0.9 : 0.1)
Green	(A)	8	4	10	9	8	7
	(B)	10	9	10	10	10	10
	(C)	10	10	10	9	10	8
Blue	(D)	8	10	9	9	9	9
	(E)	10	9	10	7	9	7
	(F)	10	10	9	10	8	8
Soil	(G)	9	10	10	9	10	9
	(H)	10	10	10	10	10	10
	(I)	10	10	10	10	10	9
Yellow	(J)	10	9	10	10	10	10
	(K)	10	10	10	10	10	10
	(L)	9	10	10	9	9	8
White	(M)	10	10	10	10	10	10
	(N)	10	10	10	10	10	10
	(O)	10	10	10	10	10	10
Accuracy		0.96	0.94	0.986667	0.946667	0.953333	0.9
(A)Orangutans, (B)Waterfall, (C)Leaf, (D)Helicopter, (E)Airplane, (F)Balloon, (G)Elephant, (H)Fighter, (I)Leopard, (J)Mural, (K)Dried leaf, (L)Sunset, (M)Dinosaur, (N)Duck, (O)Pot plant.							

TABLE 4. Categorization performance for micro-based category (20 Images/category)

20 images/class		Results					
Methods / Input		Proposed method	Macro : Micro (0.1 : 0.9)	Macro : Micro (0.3 : 0.7)	Macro : Micro (0.5 : 0.5)	Macro : Micro (0.7 : 0.3)	Macro : Micro (0.9 : 0.1)
Green	(A)	15	9	16	18	14	15
	(B)	20	13	16	18	18	19
	(C)	16	14	15	16	14	15
Blue	(D)	14	16	15	15	15	15
	(E)	17	18	17	15	14	16
	(F)	17	14	15	16	14	15
Soil	(G)	16	15	18	17	15	14
	(H)	17	11	15	19	18	17
	(I)	20	16	20	20	19	20
Yellow	(J)	17	12	15	17	17	17
	(K)	20	17	17	20	20	20
	(L)	20	20	20	20	13	19
White	(M)	20	20	20	20	20	20
	(N)	20	20	20	20	20	20
	(O)	13	14	14	14	15	13
Accuracy		0.873333	0.763333	0.843333	0.883333	0.82	0.85
(A)Orangutans, (B)Waterfall, (C)Leaf, (D)Helicopter, (E)Airplane, (F)Balloon, (G)Elephant, (H)Fighter, (I)Leopard, (J)Mural, (K)Dried leaf, (L)Sunset, (M)Dinosaur, (N)Duck, (O)Pot plant.							

no matter what the features of color or details are. The reasons of the categorization error might be the selected features are not comprehensive enough. The issue is worthy to be investigated further in the future.

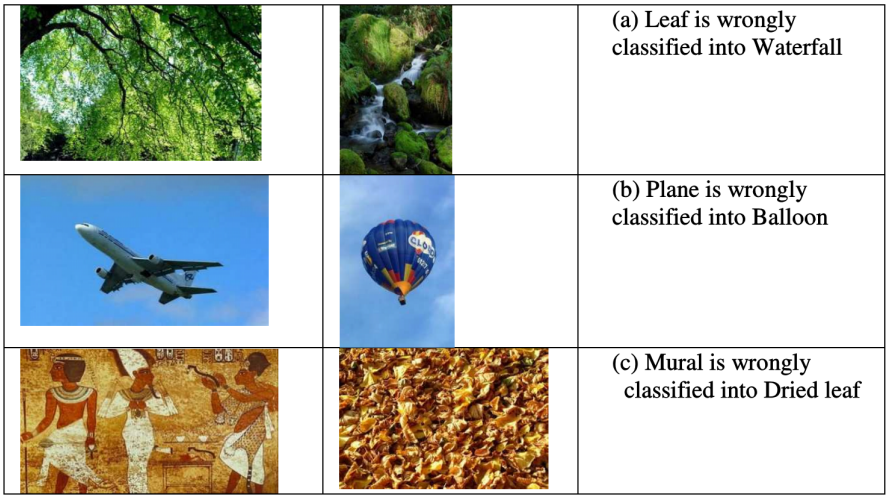

FIGURE 9. The examples of wrong categorization

5. **Conclusion.** In this paper, we have proposed a systematical approach to construct a probabilistic architecture for image categorization using macro- and micro-based visual vocabulary. In order to evaluate the performance of proposed visual vocabulary, extensive simulations on image categorization were performed. The experimental results indicate the visual vocabulary achieves promising results for categorization. Therefore, we can

conclude that the proposed visual vocabulary can effectively extract the visual features from images. The proposed approach is very effective for image categorization because it achieves excellent performance without human intervention.

Acknowledgment. This work was supported by the Ministry of Science and Technology Granted MOST 106-2221-E-214-050.

REFERENCES

[1] A. Yamada, M. Pickering, S. Jeannin and L. C. Jens, MPEG-7 Visual Part of Experimentation Model Version 9.0-Part 3 Dominant Color, *ISO/IEC JTC1/SC29/WG11/N3914*, Pisa, January. 2001.

[2] A. Mojsilovic, J. Hu, E. Soljanin, Extraction of Perceptually Important Colors and Similarity Measurement for Image Matching, Retrieval, and Analysis, *IEEE Trans. on Image Processing*, vol. 11, no. 11, November 2002.

[3] N. C. Yang, W. H. Chang, C. M. Kuo, and T. H Li, A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval, *Journal of Visual Communication and Image Representation*, vol. 19, pp. 92-105, February. 2008.

[4] W. Zhou, H. Li, Y. Lu and Qi Tian, Principal Visual Word Discovery for Automatic License Plate Detection, *IEEE Transactions On Image Processing*, vol. 21, no. 9, pp.4269-4279, September 2012.

[5] T. Li, T. Mei, I.S. Kweon, and X.S Hua, Contextual Bag-of-Words for Visual Categorization, *IEEE Transactions on Circuits and Systems For Video Technology*, vol. 21, no. 4, pp.381-392, April 2011.

[6] A. Bosch, A. Zisserman, and X. Munoz, Scene Classification Using a Hybrid Generative/Discriminative Approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp.712-727, April 2008.

[7] L. Wu, S. C. H. Hoi, and N. Yu, Semantics-Preserving Bag-of-Words Models and Applications, *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp.1908-1920, July 2010.

[8] F. Perronnin, Universal and Adapted Vocabularies for Generic Visual Categorization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, bol. 30, no. 7, pp.1243-1256, July 2008.

[9] J. Qin and N. C. Yung, Scene categorization via contextual visual words, *Pattern Recognition 43 (2010)*, pp.1874-1888, November 2009.

[10] A. Bolovinou, I.Pratikakis and S.Perantonis, Bag of spatio-visual words for context inference in scene classification, *Pattern Recognition 46 (2013)*, pp.1039-1053, September 2012.

[11] C Fredembach, M Schroder, S Susstrunk, Eigenregions for image classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 12, Dec. 2004.

[12] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky, Describing visual scenes using transformed dirichlet processes, *Advances in neural information processing systems*, pp. 1297-1304, 2005.

[13] D. M. Blei, Probabilistic topic models, *Communications of the ACM*, vol. 55, no.4, pp. 77-84, 2012.

[14] C.M. Kuo, C.H. Hsieh, N.C. Yang, C. Kuo, C.K. Chang and Y.M. Chen, Constructing a discriminative visual vocabulary with macro and micro sense of visual words, *Multimedia Tools and Applications*, vol. 75, no. 24, pp. 16983-17017, Dec. 2016.

[15] P. Guerrero, N. J. Mitra, P. Wonka, RAID: A Relation-Augmented Image Descriptor, *ACM Transactions on Graphics (TOG)-Proceedings of ACM SIGGRAPH 2016*, vol. 35, no. 4, July 2016.

[16] Z.M. Lu and Y.P. Feng, Image retrieval based on histograms of EOPs and VQ indices, *Electronics Letters*, vol. 52, pp.1683-1684, 2016.

[17] D. Liu, S. Yan, R.R. Ji, X.S. Hua and H.J. Zhang, Image Retrieval with Query-Adaptive Hashing, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 9, February 2013.

[18] H. Qi, K. Li, Y. Shen and W. Qu, Object-Based Image Retrieval with Kernel on Adjacency Matrix and Local Combined Features, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 8, November 2012 Article, no. 54, November 2012.

[19] S. Antaris and D. Rafilidis, Similarity Search over the Cloud Based on Image Descriptors Dimensions Value Cardinalities, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol 11, April 2015 Article no. 51, April 2015.

[20] X. Li, T. Uricchio, L. Ballan, M. Bertini, C. G. M. SNOEK and A. D. Bimbo, Socializing the Semantic Gap: A Comparative Survey on Image Tag Assignment, Refinement, and Retrieval, *ACM Computing Surveys (CSUR)*, vol. 49, July 2016 Article no. 14, July 2016.