Large-Scale Scene Classification Using Gist Feature

Reza Fuad Rachmadi, I Ketut Eddy Purnama
Telematics Laboratory
Department of Multimedia and Networking Engineering
Institut Teknologi Sepuluh Nopember Surabaya Indonesia 60111
Email: fuad@its.ac.id, ketut@te.its.ac.id

Abstract *- Scene classification is one of the most challenging research problem in computer vision and image understanding areas. Vividness and viewing effect are several factors that make problem very ambiguous. In this paper, we investigate gist feature performance using several state-of-the art classifier in large-scale scene classification task. Gist feature itself is a collection of gabor filter response from image and its can represented as a region boundary of the object or shape of the scene in the image. In our experiment, we use two state-of-the art classifier, L2-regularized L2-loss SVC and SVM with RBF kernel, and SUN database to evaluating the discriminant aspects of gist feature in large-scale scene classification task. From our experiment we found that the best gist feature extraction parameters for scene classification are 18 orientations, 5 scale, and 4x4 block configuration with 16,46 % accuracy in SUN database with 50 example per class for training task and SVM with RBF kernel as classifier.*

Keywords*: gist feature, large-scale scene classification task, gabor filter, image understanding, SVM classifier, L2-regularized L2-loss SVC*

1. INTRODUCTION

Scene classification is yet one of the most difficult research in computer vision and image understanding area. Over a decade, researcher had found a lot of image descriptors and classification method's with different approach to deal with it. Some approach based on high representation of the image, some object and/or texture that appears in the image, and its correlation with image class [2,4]. For example, if sand, people, and tree is appears in the image, there is a high probability that the image had taken in the beach. Another approach, as in [1,7,9], try to classify with low representation of the image, such as edge, parallel lines, corner, or other local features including SIFT and HOG. In this approach, configuration of the image, like respond of some filter or descriptor, has more contribution to classifier than either there is object or not in the image. Other way to categories scene classification algorithms is algorithm inspired by image statistics and inspired by visual system in living animal. Algorithm inspired by image statistics is a scene classification algorithm that make use of image statistics theory, such as line, corners, descriptor, and object. The second category, algorithm inspired by biological process, is a scene classification algorithm that approximate how visual system works in living animal. A couple algorithm that inspired by biological process are gist feature, FREAK, saliency, and deep-learning. As explained in [9,10], gist feature is an approximation of visual cortex respond signal. Another feature, FREAK, is approximate the LGN area function in mammals visual system and take it as features of the image.

In other area, machine learning researcher had produce a lot of new way to classifying data. In [3], they develop linear and non-linear SVM to classifying either binary class or multi class data. Other way described in [6,8] using linear approach with several constraint. In [8], they use feature sign search algorithm and Lagrange dual to solve quadratic optimization problem (QP) between two classes. Those method claimed very efficient with large-scale sparse data.

In this paper, we try to evaluating gist feature in large-scale dataset for scene classification task. We use two state-of-the art classifier, SVM with RBF kernel and L2-regularized L2-loss SVC, to evaluating the distinctness of gist feature in large-scale scene classification task. In our experiment, we try to find the best configuration for gist feature extraction task.

The paper is organized into 4 section, introduction, gist feature, result and discussion, and conclusions. Section 2 talk about gist feature, and how to extract gist feature from an image as explained in [9] followed by briefly introduction to classifier we use in the experiment. Section 3 talk about dataset used in our experiment, result of our experiment, and discuss it in detail. Section 4 concludes the paper.

2. GIST FEATURE

Biologically-inspired scene classification algorithm had developed since neuroscientist discovered how brain works, especially for vision task. The early research about how brain work in vision task reported in [10]. In those research, they monitor cat brain activity while showing some different picture to the cat. The result from those research show that visual cortex, an area in brain that processed vision task, is more sensitive to orientation and spatial frequency than

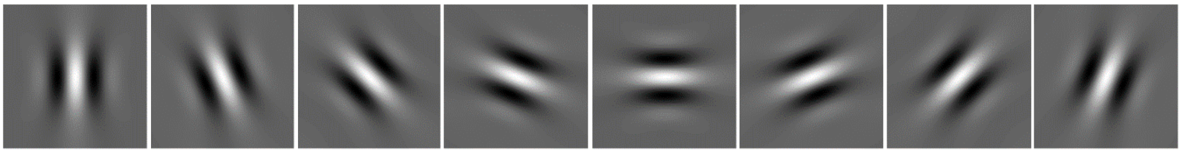

Figure 1. Visualization of gabor filter with different orientation, from left 0º, 22.5º, 45º, 67.5º, 90º, 112.5º, 135º, and 157.5º. This filter used to extracting the gist feature of the image. In implementation, gabor filter will generated in several spatial frequency.

other feature in the image and its can approximate by gabor filter respond from image with different spatial frequency and orientation.

2.1 2D Spatial Gabor Filter

For image processing purpose, gabor filter can approximated by multiplying sinusoidal complex function with 2D Gaussian function and described as follows

$$g(x,y) = s_c(x,y)w_r(x,y) \tag{1}$$

where $s_c(x,y)$ is a carrier function, which is complex sinusoidal function, and $w_r(x,y)$ is envelope function, which is a 2-D Gaussian shape function. Complex sinusoidal function and 2-D Gaussian shape function described in equation (2) and (3).

$$s_c(x,y) = e^{j(2\pi(u_0x+v_0y)+P)} \tag{2}$$

$$w_r(x,y) = Ke^{\left(-\pi\left(a^2(x-x_0)_r^2+b^2(y-y_0)_r^2\right)\right)} \tag{3}$$

In equation 2, (u_0, v_0) and P define spatial frequency in cartesian coordinate and phase of the sinusoidal function. Envelope function define by K; scale magnitude of gaussian function; a, b; scale of two axis of gaussian envelope; and x_0, y_0; center of the gaussian envelope. Subscript r indicate rotation of gaussian envelope with some angle. Visualization of gabor filter can be view in figure 1. Furthermore, gabor filter can used with several orientation and spatial frequency to provided image configuration to the classifier.

2.2. Gist Feature

Gist feature extracted using convolution proses and mean per block calculation. Thus process used gabor filter with different spatial frequency and orientation for convolution process and each mean calculation done by split the image into several block configuration, like 8 x 8 block or 4 x 4 block. Figure 2 illustrated the gist feature extracting process. The convolution process done in fourier domain for computing efficiency and switch back to time domain for mean average per block calculation.

2.3. SVM and L2-Regularized L2-Loss SVC

Support vector machine (SVM) is one of state-of-the art classifier that being used in several years for wide area of applications, including image processing, computer vision, biomedical problem, and other problem that need decision support on it. Naturally, SVM will solve binary classification problem by searching the most distinction function for splitting data in each class. The problem in SVM method can described as follows

$$\min_{\boldsymbol{w},b,\xi} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\xi_i \tag{4}$$

with linear constraint

$$\begin{gathered} y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \\ \xi_i \geq 0, i = 1, \ldots, l \end{gathered} \tag{5}$$

Thus optimization problem done using quadratic solver, like SMO (Sequential Minimal Optimization) [3]. The kernel, $\phi(x_i)$, can be any function to transform data from low dimensional spaces to high dimensional

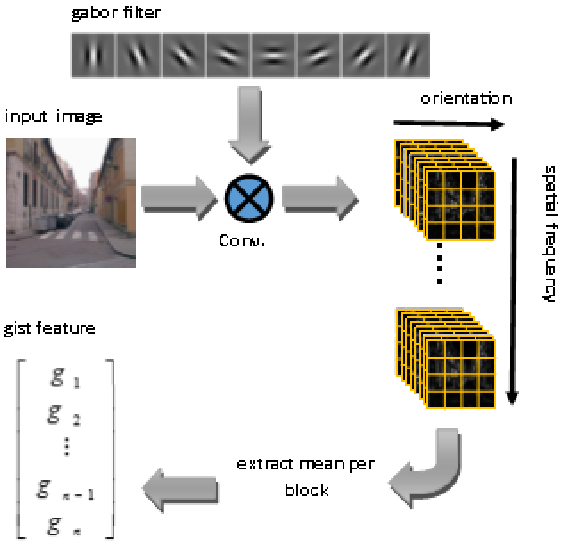

Figure 2. Gisr feature extracting process from input image and some predefine gabor filter.

Table 1. Result from our experiments with several parameters for gist feature extraction process and two state-of-the-art classifier, SVM with RBF kernel and L2-Regularized L2-Loss SVC.

Configuration	Example Per Class	SVM with RBF Kernel		L2-Regularized L2-Loss SVC	
		Accuracy (%)	Raw Data	Accuracy (%)	Raw Data
8 orientations, 4 scales, 4x4 block	1	2.290681	455 / 19850	2.295214	456 / 19850
	5	5.462973	1085 / 19850	5.056423	1004 / 19850
	10	7.880100	1565 / 19850	7.213098	1432 / 19850
	20	10.750120	2134 / 19850	9.561210	1898 / 19850
	50	15.172280	3012 / 19850	12.886140	2558 / 19850
8 orientations, 4 scales, 8x8 block	1	2.071536	412 / 19850	2.233248	444 / 19850
	5	5.408564	1074 / 19850	4.700252	933 / 19850
	10	8.069018	1602 / 19850	6.403526	1272 / 19850
	20	10.762730	2137 / 19850	8.494203	1687 / 19850
	50	14.991430	2976 / 19850	11.532490	2290 / 19850
12 orientations, 5 scales, 4x4 block	1	2.433854	484 / 19850	2.578843	512 / 19850
	5	6.036273	1199 / 19850	5.436776	1080 / 19850
	10	8.686146	1725 / 19850	7.529974	1495 / 19850
	20	11.647350	2312 / 19850	9.957184	1977 / 19850
	50	16.458920	3268 / 19850	13.282620	2637 / 19850
18 orientations, 5 scales, 4x4 block	1	2.343074	466 / 19850	2.585390	514 / 19850
	5	5.732494	1138 / 19850	5.561209	1104 / 19850
	10	8,242820	1637 / 19850	7.832243	1554 / 19850
	20	11.653400	2314 / 19850	10.321423	2049 / 19850
	50	16.459450	3268 / 19850	13.872040	2754 / 19850

spaces. The reason SVM use high dimensional spaces to solve the problem, because in high dimensional spaces there more probabilities that data can be divided linearly as defined in original SVM problem. In this paper, we use implementation of SVM with RBF kernel described by [3].

Other state-of-the art classifier is L2-Regularized L2-Loss SVC [6]. Thus classifier is linear SVM but with different loss and regularized function. The problem of L2-Regularized L2-Loss SVC described as follows

$$\min_{\boldsymbol{w}} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}(\max(0,1-y_i\boldsymbol{w}^T\boldsymbol{x_i}))^2 \qquad (6)$$

Equation 6 is relatively same as equation 4 with different ξ_i. As we mention before, L2-Regularized L2-Loss SVC is same as linear SVM and the only different is how to calculate loss function and regularized function but the solver is very different. L2-Regularized L2-Loss use dual coordinate descent to solve binary classification problem and standard SVM usually use SMO or quadratic programming [6]. There are another type of SVC classifier, L1-Regularized L2-Loss SVC and L1-Regularized L1-Loss SVC. In this paper, we focus on L2-Regularized L2-Loss SVC because in [6] those classifier recommended for regular classification task. Another advantages of L2-Regularized L2-Loss SVC is very fast training time and less memory footprints even with very large dataset.

3. RESULTS AND DISCUSSIONS

As explained in [9], gist feature work very well in image retrieval task and promising to use in large-scale scene classification task. In order to test discriminate aspects of gist feature in large-scale scene classification task, we use SUN database [5] that consist 397 different classes and 130,519 images. Some classes in SUN database may very ambiguous, such as apple orchard and grassland, and very challenging for computer vision system. We split the experiment into 10 part, each part has different training and testing data, and each part has incremental data on training example. Incremental data means that in training process we use 1, 5, 10, 20, and 50 example per class. To provide enough information about the distinctness of gist feature, we use several different parameters for gist feature extraction in each part. For orientation, we use 8, 12, and 18 orientation and for spatial frequency we use 0.32, 0.16, 0.08, 0.04, and 0.02 cycles/pixel. Example of gabor filter respond with 8 orientations and 0.32 spatial frequency is shown in figure 3. All our experiment is summaries in table 1 and the best performance of gist feature is 16.459450 % with extraction parameter: 18 orientations, 5 scales, and 4x4 block configuration. Raw data and acurracy in tabel 1 shown mean result, as percentage or right prediction in testing process, of all part in training/testing split and for raw data columns its use ceiling operation to round decimal value. To extract all gist feature in SUN database with some configuration extraction parameters required about a week or about 5-6 second for each image. To reduce training time for SVM with RBF kernel, we use CUDA based LIBSVM implemented by [11] and run it in our Tesla hardware.

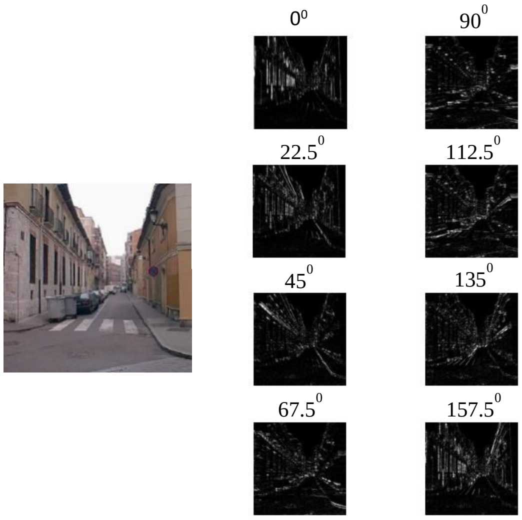

Figure 3. Example of gabor filter respond from input image in the left and with spatial frequency 0.32 cycles/pixel and 8 orientations. Image is part of [9] and freely downloaded in author website

In first experiment, we use same orientation and scale parameters but change the block configuration scheme, 4x4 and 8x8. From the experiments, it was found that the accuracy of the classifier decreases around 0.2 % so it can be concluded that block configuration scheme is not take significant impact to feature distinctness and accuracy of the classifier. Figure 4 shown classifier accuracy for each incremental part of training data for different block configuration scheme.

To test the effect of increasing orientations and scales, we use 3 different orientations (8, 12, and 18) and 2 scales (4 scales and 5 scales). For 4 scales we use 0.32, 0.16, 0.08, and 0.04 cycles/pixel. For 5 scales we add one more scale, 0.02 cycles/pixel. As show in figure 5, the accuracy of classifier increases along with increased orientations. Even the raise is small, about 1 %, for SVM-RBF and L2-regularized L2-loss SVC, we concluded that more orientations can be increasing the accuracy of the classifier and the distinctness of the features.

In our training process, we take a look for memory footprint of each classifier with fix gist extraction parameter: 18 orientations, 5 scales, and 4x4 block configuration and with 50 example per class. We found that for L2-regularized L2-loss SVC, memory usage for training process is about 442 MB. Thus memory usage relatively same with SVM-RBF classifier without GPU. For GPU, memory used for SVM-RBF classifier is about 3.5 GB and it about 8 times larger than we run SVM-RBF without GPU. Those phenomena occur because all matrix calculation stores two matrix, one as input matrix and another matrix as output matrix from GPU process. Training time for SVM-RBF classifier using GPU is about 5-8 times faster than running SVM-RBF in CPU.

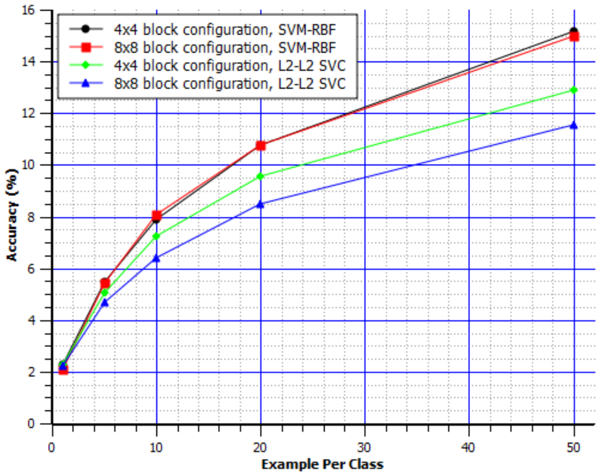

Figure 4. Classifier accuracy for 4x4 and 8x8 block configuration with 8 orientations and 5 scales gist extraction parameters.

4. CONCLUSIONS

In this paper, we shown that gist feature can used for large-scale scene classification task. The best accuracy is achieve using gist feature extraction parameters: 18 orientations, 5 scales, and 4x4 block configuration. We also showed that the increase in block configuration scheme does not improve the accuracy of the classifier even reduce the accuracy of the classifier as shown in table 1 and figure 4. From table 1 and figure 5, we showed that adding scales and orientations had increasing the accuracy of the classifier also increased the distinctness of the features. SVM classifier with RBF kernel has higher accuracy than L2-regularized L2-loss SVC with

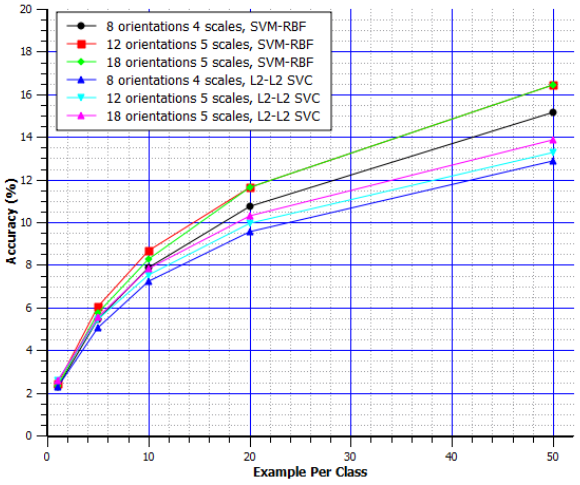

Figure 5. Classifier accuracy for three different orientations and two different scales.

same data, but the training time for L2-regularized L2-loss SVC is faster than SVM with RBF kernel.

For future works, the gist feature can combined with other features, such as saliency or salient region. Saliency or salient region is a computational model of human attentions of some interest things in the image. The result of salient region is a weight of human attentions for all region in the image. By combined between gist feature and saliency, the features will be more focus to some interesting region in the image and may more distinct than before. Another method that can increasing the distinctness of the feature is try use another block configuration scheme. In this paper we use rectangle to represent the block and its can be changed with another shape like rhombus and circle. The overlapping region of the shape or change the method of calculating value per block may use to increasing the distinctness of the features.

REFERENCES

[1] Jorge Sánchez, Florent Perronnin, Thomas Mensink, and Jakob J. Verbeek. “Image Classification with the Fisher Vector: Theory and Practice”. International Journal of Computer Vision 105(3):222-245 (2013).

[2] Guang-Tong Zhou, Tian Lan, Weilong Yang, and Greg Mori. “Learning Class-to-Image Distance with Object Matchings” Proc of CVPR, page 795-802. IEEE, (2013).

[3] Chih-Chung Chang, and Chih-Jen Lin, “LIBSVM: a library for support vector machines”. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.

[4] L.-J. Li, H. Su, E.P. Xing and L. Fei-Fei. “Object Bank: A High-Level Image Representation for Scene Classification and Semantic Feature Sparsification”. Proceedings of the Neural Information Processing Systems (NIPS). 2010.

[5] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba “SUN Database: Large-scale Scene Recognition from Abbey to Zoo”. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2010).

[6] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. “LIBLINEAR: A library for large linear classification”. Journal of Machine Learning Research 9 (2008), 1871-1874.

[7] C. Siagian, L. Itti, “Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 2, pp. 300-312, Feb 2007.

[8] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng. “Efficient sparse coding algorithms”. Proc of NIPS, page 801-808. MIT Press, (2006).

[9] Aude Oliva and Antonio Torralba. “Modeling the shape of the scene: a holistic representation of the spatial envelope”. International Journal of Computer Vision, Vol. 42(3): 145-175, 2001.

[10] Daugman, John G., “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters”. Journal of the Optical Society of America A, 2(7):1160–1169, July 1985.

[11] A. Athanasopoulos, A. Dimou, V. Mezaris, I. Kompatsiaris, "GPU Acceleration for Support Vector Machines", Proc. 12th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2011), Delft, The Netherlands, April 2011.

[12] S. Filipe and Luis A. Alexandre. “From the human visual system to the computational models of visual attention: A survey”. Articial Intelligence Review, pages 1-47, January 2013.