IPSJ Transactions on Computer Vision and Applications Vol.5 50–54 (July 2013)

[DOI: 10.2197/ipsjtcva.5.50]

Express Paper

Fisher Vector based on Full-covariance Gaussian Mixture Model

Masayuki Tanaka[1,a)] Akihiko Torii[1,b)] Masatoshi Okutomi[1,c)]

Received: March 11, 2013, Accepted: April 24, 2013, Released: July 29, 2013

Abstract: In image retrieval applications, the Fisher vector of the Gaussian mixture model (GMM) with a diagonal-covariance structure is known as a powerful tool to describe an image by aggregating local descriptors extracted from the image. In this paper, we propose the Fisher vector of the GMM with a full-covariance structure. The closed-form approximation of the GMM with a full-covariance structure is derived. Our observation is that the Fisher vector of a higher dimensional GMM yields higher image retrieval performance. The Fisher vector for the GMM with a block-diagonal-covariance structure is also introduced to provide moderate dimensionality for the GMM. Experimental comparisons performed using two major datasets demonstrate that the proposed Fisher vector outperforms state-of-the-art algorithms.

Keywords: image retrieval, image descriptor, Fisher vector, and full-covariance structured Gaussian mixture model

1. Introduction

Image retrieval plays an important role on many computer vision based applications such as image/video copy detection, keyframe indexing of video, and visual robot localization. According to the great success of the bag-of-words (BOW) model, A basic approach of image retrieval consists of three steps: (1) detecting and describing features (keypoints); (2) encoding them as image descriptors using pre-computed codebooks; (3) ranking (making a shortlist) by matching image descriptors of query and database images. Since the efficiency and performance of retrieval essentially depend on how the images are described, the image description has been a widely studied topic [6], [7], [10], [12], [13], [15], [16].

The most popular image descriptor in a decade is the bag-of-visual-words (BoVW) representation [15]. The BoVW representation can be considered as a simple and effective aggregation of powerful local descriptors such as SIFT descriptor [8] or others [17]. More recent successful image descriptor is to encode as Fisher vectors [12], [13]. Fisher vectors associated to parameters of generative model are computed by aggregating local descriptors of images. Computation Fisher vectors generally require a huge computational cost. A closed-form approximation of Fisher vectors for Gaussian Mixture Model (GMM) with a diagonal-covariance is proposed under reasonable assumptions [12]. Another recent image descripor is Vector of Locally Aggregated Descriptors (VLAD) [6], [7] which efficiently aggregate local descriptors with a small visual vocabulary. VLAD encodes local descriptors by accumulating the difference between the local descriptors and their nearest visual words [6]. Although VLAD was originally proposed as an extension of BoVW representation [6], it is shown that VLAD can be interpreted as a simplified Fisher vector associated to the means of the GMM with a scaled-identity-covariance structure [7].

If we focus on the generative models of Fisher vector and VLAD, the difference is only the covariance structure of each GMM component. The diagonal-covariance is assumed for the Fisher vector, while the scaled-identity-covariance is assumed for the VLAD. Since the diagonal-covariance structures are more informative compared to the scaled-identity-covariance structures, Fisher vectors generally give better image retrieval performance than VLAD [7]. Motivated by this fact, we aim at improving the performance of image retrieval by representing more informative Fisher vectors arising from the *richness* of GMM.

In this paper, we refer the diagonal-covariance GMM to the GMM with the diagonal-covariance structure. The diagonal-covariance Fisher vector is referred to the Fisher vector based on the diagonal-covariance GMM.

In the sense of the richness of the covariance structure, the best choice is the full-covariance structure which has full elements of the covariance matrix. However, the full-covariance Fisher vector has two main issues. First, to the best of our knowledge, a closed-form approximation of the full-covariance Fisher vector is not in the literature. Second, the number of the training parameters of the full-covariance GMM is very large. Estimating parameters of the GMM which has a very large number of components by Expectation and Maximization (EM) algorithm [1] is not feasible.

The main contribution of this paper is to derive a closed-form approximation of the full-covariance Fisher vector with the same assumptions as Ref. [12]. By diagonalizing the covariance matrix, the closed-form approximation can be derived in the similar manner to Ref. [12]. We also introduce a block-diagonal-

[1] Tokyo Institute of Technology, Meguro, Tokyo 152–8550, Japan
[a)] mtanaka@ctrl.titech.ac.jp
[b)] torii@ctrl.titech.ac.jp
[c)] mxo@ctrl.titech.ac.jp

covariance Fisher vector whose generative model is the block-diagonal-covariance GMM. This provides a moderate number of parameters to be learned. Also, since the block-diagonal-covariance structure can express both the diagonal-covariance structure and the full-covariance structure as special cases, one can design suitable covariance structure.

2. Fisher Kernel and Diagonal-covariance Fisher Vector

2.1 Fisher Kernel

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be samples of a generative model expressed by a probability density function p with a parameter $\boldsymbol{\lambda}$. Note that a sample implies a set of local descriptors in our case. The Fisher kernel is defined by the inner product between the gradient vectors in feature space [3]:

$$K(\boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{g}_{\boldsymbol{\lambda}}^{\top}(\boldsymbol{X}) \boldsymbol{F}_{\boldsymbol{\lambda}}^{-1} \boldsymbol{g}_{\boldsymbol{\lambda}}(\boldsymbol{Y}), \tag{1}$$

where

$$\boldsymbol{g}_{\boldsymbol{\lambda}}(\boldsymbol{X}) = \nabla_{\boldsymbol{\lambda}} \log p(\boldsymbol{X}; \boldsymbol{\lambda}), \tag{2}$$

$$\boldsymbol{F}_{\boldsymbol{\lambda}} = E_{\boldsymbol{X} \sim p}[\boldsymbol{g}_{\boldsymbol{\lambda}}(\boldsymbol{X}) \boldsymbol{g}_{\boldsymbol{\lambda}}^{\top}(\boldsymbol{X})]. \tag{3}$$

$\top$ represents the transpose operator, $\boldsymbol{g}_{\boldsymbol{\lambda}}$ is the gradient vector of the log-likelihood, and $\boldsymbol{F}_{\boldsymbol{\lambda}}$ is the Fisher information matrix. Since the Fisher information matrix is symmetric and positive definite. it can be decomposed as $\boldsymbol{F}_{\boldsymbol{\lambda}} = \boldsymbol{L}_{\boldsymbol{\lambda}}^{\top} \boldsymbol{L}_{\boldsymbol{\lambda}}$. Using this matrix $\boldsymbol{L}_{\boldsymbol{\lambda}}$, the *Fisher vector* of the sample $\boldsymbol{X}$ is defined as

$$\mathfrak{g}(\boldsymbol{X}) = \boldsymbol{L}_{\boldsymbol{\lambda}} \boldsymbol{g}_{\boldsymbol{\lambda}}(\boldsymbol{X}). \tag{4}$$

2.2 Diagonal-covariance Fisher Vector

Assuming that each local descriptor is independently genarated from a generative model, GMM is a natural choice since it can approximate sufficient classes of probability density distributions. The GMM is defined as

$$p(\boldsymbol{x}) = \sum_{i=1}^{K} w_i \, q(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{S}_i), \tag{5}$$

where w_i is the mixture weight for i-th GMM component, K is the number of the GMM components, and $q(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{S})$ represents the Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{S}$.

In Ref. [12], they derived the closed-form approximation of the diagonal-covariance Fisher vector with two assumptions: (1) the number of the extracted local descriptors is constant, and (2) each GMM component is far from each other[*1]. The diagonal-covariance matrix is expressed by

$$\boldsymbol{S} = \mathrm{diag}(\boldsymbol{\sigma}^2), \tag{6}$$

where $\mathrm{diag}(\boldsymbol{\sigma}^2)$ represents the diagonal matrix whose diagonal components are the elements of $\boldsymbol{\sigma}^2$. The soft assignment factor of the local descriptor $\boldsymbol{x}_t$ to the i-th GMM component is introduced:

$$\gamma_i(\boldsymbol{x}_t) = \frac{w_i \, q(\boldsymbol{x}_t; \boldsymbol{\mu}_i, \boldsymbol{S}_i)}{\sum_{j=1}^{K} w_j \, q(\boldsymbol{x}_t; \boldsymbol{\mu}_j, \boldsymbol{S}_j)}. \tag{7}$$

Using soft assignment factors, the diagonal-covariance Fisher vectors associated to the means and the covariance of i-th GMM component are derived [12]:

$$\mathfrak{g}_i^{\boldsymbol{\mu}}(\boldsymbol{X}) = \frac{1}{\sqrt{T w_i}} \sum_{t=1}^{T} \gamma_i(\boldsymbol{x}_t) \left(\frac{\boldsymbol{x}_t - \boldsymbol{\mu}_i}{\boldsymbol{\sigma}_i} \right), \tag{8}$$

$$\mathfrak{g}_i^{\boldsymbol{S}}(\boldsymbol{X}) = \frac{1}{\sqrt{T w_i}} \sum_{t=1}^{T} \gamma_i(\boldsymbol{x}_t) \left[\frac{(\boldsymbol{x}_t - \boldsymbol{\mu}_i)^2}{\boldsymbol{\sigma}_i^2} - \boldsymbol{1} \right], \tag{9}$$

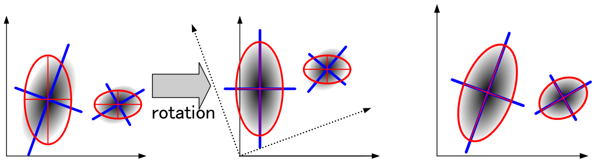
where $\boldsymbol{1}$ is the column vector of which every elenment is one and the dimension is the same as that of $\boldsymbol{x}_t$, and the component-wise vector divisions and square operations are performed. The common normalization factor, $\sqrt{T}$, in Eqs. (8) and (9) are different from that expressed in Ref. [13]. However, the common normalization factor will be canceled in the power normalization process [13]. In this paper, we follow the normalization factor in the original paper [12].

The Fisher vector associated to the means, $\mathfrak{g}^{\boldsymbol{\mu}}(\boldsymbol{X})$, is the concatenation of the vectors $\mathfrak{g}_i^{\boldsymbol{\mu}}(\boldsymbol{X})$ for $i = 1, 2, \cdots, K$. The Fisher vector associated to the covariance, $\mathfrak{g}^{\boldsymbol{S}}(\boldsymbol{X})$, can be obtained by the same manner. The dimensions of the $\mathfrak{g}^{\boldsymbol{\mu}}(\boldsymbol{X})$ and the $\mathfrak{g}^{\boldsymbol{S}}(\boldsymbol{X})$ are $D \times K$, where D is the dimension of the local descriptor. Although the Fisher vector associated to the mixture weight can be obtained, the Fisher vectors associated to the mean and/or the covariance are used in general [7], [13] because the Fisher vectors associated to the mixture weights weights do not significantly contribute to the image retrieval. Therefore, we focus on the Fisher vectors associated to the means and the covariance.

3. Proposed Full-covariance Fisher Vector

The difference between the diagonal-covariance and the full-covariance GMM is shown in **Fig. 1**. The diagonal-covariance GMM can not precisely model the distribution of the local descriptors, the orientation of the axes between them does not coincide. Although the coordinate rotation, for example by the principal component analysis (PCA), can be applied before modeling by the diagonal-covariance GMM, only several components could be fitted, but the others are not. In contrast, as shown in Fig. 1 (b), the full-covariance GMM can precisely model the distribution. Note that although the mixture weights are also the parameters of the GMM, we focus on the Fisher vectors associated to the means and the covariance as discussed in Section 2.2. It is expected that this difference also appears in the image retrieval performance.

The number of parameters of the full-covariance matrix is $D(D+1)/2$ because the covariance matrix is symmetric, while



(a) the diagonal-covariance GMM (b) the full-covariance GMM

Fig. 1 Schematics of the difference between the diagonal-covariance and the full-covariance GMMs, where dark gray region represents high-density region, the blue lines represent dominant directions of the density, and the red ellipses represent the modeled Gaussian component.

[*1] In the original paper Ref. [12], the second assumption is mentioned in a different manner. However, the mathematical meaning is same.

that of the diagonal-covariance matrix is D, where D is the dimension of the local descriptor. The full-covariance matrix, S, can be rewritten by the eigen decomposition:

$$\boldsymbol{S} = \boldsymbol{U}\,\mathrm{diag}(\boldsymbol{\xi}^{2})\,\boldsymbol{U}^{\top}, \tag{10}$$

where $\boldsymbol{U}$ is the matrix whose column vectors are the eigenvectors of $\boldsymbol{S}$, and $\boldsymbol{\xi}^{2}$ is the vector whose elements are the eigenvalues of $\boldsymbol{S}$. It can be considered that the eigenvalues, $\boldsymbol{\xi}^{2}$, of the full-covariance matrix correspond to the variances, $\boldsymbol{\sigma}^{2}$, of the diagonal-covariance matrix.

The matrix $\boldsymbol{U}$ represents the dominant direction of the distribution and the vector $\boldsymbol{\xi}^{2}$ represents the variance along the associated dominant direction. We can consider two different types of the Fisher vectors with respect to the covariance; the Fisher vectors associated to the dominant direction parameters and the variance parameters. Because the derivation of the Fisher vector associated to the dominant direction parameters are intractable, we only derive the Fisher vector associated to the variance parameters or the eigenvalues. The derivation of the Fisher vector associated to the dominant direction parameters is our future work. If we focus on the full-covariance Fisher vector associated the eigenvalues, its closed-form approximation can be derived with the same assumption as in Ref. [12]. The dimension of the full-covariance Fisher vector associated to the covariance is the same as that of the diagonal-covariance Fisher vector associated to covariance.

In order to derive the closed-form approximation, first, we rotate the coordinate, so that the covariance matrix of the current component is the diagonal matrix as shown in **Fig. 2**. The integral is identical for the coordinate rotation. Then, we can derive the Fisher vector associated to the current component by following the same manner in Ref. [12]. The closed-form approximation of the full-covariance Fisher vectors can be derived by rotating for each component as shown in Fig. 2.

Here, we show the closed-form approximation of full-covariance Fisher vector as follow*2:

$$\mathfrak{h}_i^{\boldsymbol{\mu}}(\boldsymbol{X}) = \frac{1}{\sqrt{T w_i}} \sum_{t=1}^{T} \gamma_i(\boldsymbol{x}_t)\boldsymbol{S}_i^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu}_i), \tag{11}$$

$$\mathfrak{h}_i^{\boldsymbol{S}}(\boldsymbol{X}) = \frac{1}{\sqrt{T w_i}} \sum_{t=1}^{T} \gamma_i(\boldsymbol{x}_t)\left[\left\{\boldsymbol{S}_i^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu}_i)\right\}^2 - \boldsymbol{1}\right], \tag{12}$$

where the inverse square root of the matrix $\boldsymbol{S}_i$ is defined by

$$\boldsymbol{S}_i^{-1/2} = \mathrm{diag}(\boldsymbol{\xi}_i^{-1})\boldsymbol{U}_i^{\top}, \tag{13}$$

where $\boldsymbol{U}_i$ is the matrix whose column vectors are the eigenvectors of $\boldsymbol{S}_i$, and $\boldsymbol{\xi}_i^{2}$ is the vector whose elements are the eigenvalues of $\boldsymbol{S}_i$. For the derivation of the closed-form approximation, the number of the local descriptors is assumed to be constant, T, as same as in Ref. [12]. In practice, the number of the local descriptors depends on the image. However, this dependency and the common normalization factor, $\sqrt{T}$, does not affect to the retrieval performance, since we apply the power normalization [13].

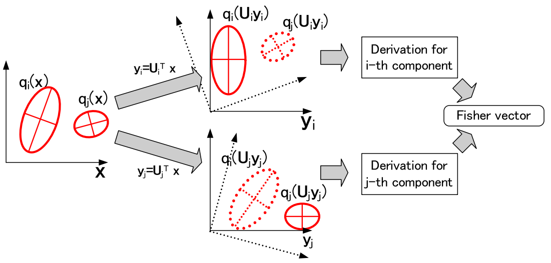


Fig. 2 The schematic flow of the derivation of the full-covariance Fisher vector.

These closed-form approximation of the full-covariance Fisher vectors in Eqs. (11) and (12) represents the diagonal-covariance Fisher vectors in Eqs. (11) and (12) as special case of $\boldsymbol{S} = \mathrm{diag}(\boldsymbol{\sigma}^2)$. The power normalization [13] can be applied to the full-covariance Fisher vector as well as the diagonal-covariance Fisher vector.

Even though the full-covariance GMM can precisely represent the distribution of the local descriptors, the parameter estimation of the full-covariance GMM with large number of components is not an easy task. Accordingly, we also use the block-diagonal-covariance GMM instead of the full-covariance GMM. The block-diagonal-covariance matrix and its inverse can be represented as Ref. [14]

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{B}_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{B}_m \end{pmatrix}, \quad \boldsymbol{S}^{-1} = \begin{pmatrix} \boldsymbol{B}_1^{-1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{B}_m^{-1} \end{pmatrix}, \tag{14}$$

where $\boldsymbol{B}_i$ is the square matrix and m is the number of covariance blocks. Note that the inverse operation of the block-diagonal matrix preserves the block-diagonal structure. This property makes easy to the parameter estimation of GMM with the block-diagonal-covariance by the EM algorithm. In addition, the block-diagonal matrix can express the diagonal and the full matrix as the extreme cases where $m = D$ and $m = 1$, respectively. We can choose the dimensionality of the covariance structure with the block-diagonal-covariance by setting the number of the blocks.

4. Experimental Validation

We use the 128-dimensional SIFT [8] to extract the local descriptors. For the parameter learning of the GMM, we apply the EM algorithm. Note that we can apply the same code for them because the block-diagonal-covariance GMM can represent the diagonal-covariance and the full-covariance GMMs. The power-normalization with $\alpha = 0.5$ is applied as the post-processing. A ranking of a query image is performed according to the inner products between the Fisher vectors of the query image and the database images. Then, an image retrieval performance is evaluated based on this ranking. Although several efficient approximate search algorithms such as a FLANN [9] and a product quantization [5] have been proposed, we use the exact nearest neighbor search to purely compare the performances of the Fisher vectors. The evaluation is performed on two major datasets:

- The University of Kentucky Benchmark (UKB) [11]. This set consists of 2,550 groups of four images each. The most commonly used performance measure of this set is to count how many of the four images which are top-four (including

[*2] Although the diagonal-covariance Fisher vector is expressed with the full-covariance matrix in Ref. [2], they clearly mentioned that the covariance matrix is assumed to be diagonal. So, their expression is only for the diagonal-covariance Fisher vector.

Table 1 Image retrieval performance comparisons on the UKB dataset with the block-diagonal Fisher vector of different number of covariance blocks, where K is the number of the GMM component and D is the dimension of the Fisher vector. The bold indicate the best performance.

			(Full)		number of covariance blocks					(Diagonal)
	K	dim.	1	2	4	8	16	32	64	128
means & cov. (eigenvalues)	16	4096	3.091	3.087	3.097	3.077	3.040	3.040	3.054	3.030
means only	32	4096	**3.273**	3.253	3.242	3.208	3.195	3.141	3.154	3.098
cov. (eigenvalues) only	32	4096	2.782	2.786	2.888	2.932	2.936	2.992	3.020	2.991

Table 2 Image retrieval performance comparisons with state-of-the-art algorithms, where the component number $K = 16$ and there is no data with $K = 16$ on the UKB in Ref. [7]. The PCA reduces the dimension to 512. The bold font represents the best performance.

	Ref. [6]		Ref. [7]			Proposed
	VLAD	Fisher	VLAD	Fisher	Fisher with PCA	Fisher
UKB	3.07	3.07	-	-	-	**3.19**
Holidays	0.496	0.497	0.520	0.540	0.546	**0.591**

the query itself) in the searching result of the 10,200 images. The best value of this performance measure is four and the higher value represents better performance.

- The INRIA Holidays dataset [4]. This set consists of 1,491 images, 500 of them used for queries. For this set, the mean Average Precision (mAP) is used as the performance measure. The best value of the mAP is one and the higher value represents better performance.

Recently, the Fisher vector associated to the means is mainly used [6], [7]. We experimentally confirm the impact of using Fisher vector associated to the means only. **Table 1** summarizes the image retrieval performances on the UKB dataset with three type of the Fisher vectors associated to the means & the covariance (eigenvalues), the means only, and the covariance (eigenvalues) only. The number of the GMM components is set 16 for those of the means & the covariance (eigenvalues) and 32 for those of the means only and the covariance (eigenvalues) only, respectively, so that the dimension of the each aggregated image descriptor equals 4,096. That the covariance block number is one means the full-covariance Fisher vector. That the covariance block number is 128 means the diagonal-covariance Fisher vector. The diagonal-covariance Fisher vector, or the case that the number of covariance blocks is 128, represents existing Fisher vectors and other cases represent the proposed Fisher vectors.

For all number of covariance blocks, the Fisher vector associated to the means only outperforms those of the means & covariance (eigenvalues) and the covariance (eigenvalues) only. These comparisons experimentally validate the superiority of using the Fisher vector associated to the means only. Then, we focus on the performances of the Fisher vector associated to the means only. The image retrieval performance almost monotonically degrades as the number of covariance blocks increases. It shows that the Fisher vector based on the block-diagonal-covariance GMM with the smaller number of the covariance blocks, or the richer generative model, can yield better performance. It also indicates that we can improve the performance by increasing the dimensionality of the GMM while the dimension of the aggregated image descriptors is fixed. It has a positive impact on memory efficiency.

We compare the proposed full-covariance Fisher vector with the state-of-the-art algorithms [6], [7]. For the fair comparisons, we compare with the same number of the GMM components with $K = 16$. The evaluated performances are summarized in **Table 2**. These image retrieval performance comparisons demonstrate that the proposed full-covariance Fisher vector outperforms the existing diagonal-covariance Fisher vector and the VLAD.

5. Conclusion

In this paper, we propose a novel full-covariance Fisher vector for the image retrieval. The closed-form approximation of the full-covariance Fisher vector is derived by using the eigen decomposition of the covariance matrix. The block-diagonal-covariance Fisher vector is also introduced, so that the parameters of the generative GMM are easily learned by the EM algorithm. The experimental comparisons demonstrate that the proposed full-covariance Fisher vector outperforms state-of-the-art algorithms in the image retrieval. Our future works include efficient GMM parameter learning, and the comparisons with a very large dataset.

Acknowledgments A part of this paper is supported by JSPS KAKENHI Grant Number 24700161.

References

[1] Bishop, C. et al.: *Pattern recognition and machine learning*, Springer New York (2006).

[2] Chatfield, K., Lempitsky, V., Vedaldi, A. and Zisserman, A.: The devil is in the details: An evaluation of recent feature encoding methods, *British Machine Vision Conference* (*BMVC*) (2011).

[3] Jaakkola, T., Haussler, D., et al.: Exploiting generative models in discriminative classifiers, *Advances in Neural Information Processing Systems* (*NIPS*), pp.487–493 (1999).

[4] Jégou, H., Douze, M. and Schmid, C.: Hamming embedding and weak geometric consistency for large scale image search, *European Conference on Computer Vision* (*ECCV*), pp.304–317 (2008).

[5] Jégou, H., Douze, M. and Schmid, C.: Product quantization for nearest neighbor search, *IEEE Trans. Pattern Analysis and Machine Intelligence* (*PAMI*), Vol.33, No.1, pp.117–128 (2011).

[6] Jégou, H., Douze, M., Schmid, C. and Pérez, P.: Aggregating local descriptors into a compact image representation, *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp.3304–3311 (2010).

[7] Jégou, H., Perronnin, F., Douze, M. and Schmid, C., et al.: Aggregating local image descriptors into compact codes, *IEEE Trans. on Pattern Analysis and Machine Intelligence* (*PAMI*), Vol.34, No.9, pp.1704–1716 (2012).

[8] Lowe, D.: Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91–110 (2004).

[9] Muja, M. and Lowe, D.: Fast approximate nearest neighbors with automatic algorithm configuration, *International Conference on Computer Vision Theory and Applications* (*VISSAPP*), pp.331–340 (2009).

[10] Nakayama, H., Harada, T. and Kuniyoshi, Y.: Global gaussian ap-

proach for scene categorization using information geometry, *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp.2336–2343 (2010).

[11] Nistér, D. and Stewénius, H.: Scalable Recognition with a Vocabulary Tree, *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vol.2, pp.2161–2168 (2006).

[12] Perronnin, F. and Dance, C.: Fisher kernels on visual vocabularies for image categorization, *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp.1–8 (2007).

[13] Perronnin, F., Sánchez, J. and Mensink, T.: Improving the fisher kernel for large-scale image classification, *European Conference on Computer Vision* (*ECCV*), pp.143–156 (2010).

[14] Petersen, K.B. and Pedersen, M.S.: *The Matrix Cookbook*, Technical University of Denmark (2008).

[15] Sivic, J. and Zisserman, A.: Video Google: A text retrieval approach to object matching in videos, *International Conference on Computer Vision* (*ICCV*), pp.1470–1477 (2003).

[16] van Gemert, J., Veenman, C., Smeulders, A. and Geusebroek, J.: Visual word ambiguity, *IEEE Trans. Pattern Analysis and Machine Intelligence* (*PAMI*), Vol.32, No.7, pp.1271–1283 (2010).

[17] Winder, S. and Brown, M.: Learning local image descriptors, *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp.1–8 (2007).

(Communicated by *Eisaku Maeda*)