The method of image feature selection for integration of image classification by Bag-of-Keypoints

Kyohei MATSUMOTO *, Ryotaro OKADA *, Takafumi NAKANISHI † and Takashi KITAGAWA ‡

* *Graduate School of Systems and Information Engineering,University of Tsukuba*
Tsukuba, Ibaraki, Japan
{kyohei,ryotaro}@mma.cs.tsukuba.ac.jp

† *Center for Global Communications(GLOCOM),International University of Japan,*
Roppongi, Minato-ku, Tokyo, Japan
takafumi@glocom.ac.jp

‡ *Division of Information Engineering,University of Tsukuba*
Tsukuba, Ibaraki, Japan
takashi@cs.tsukuba.ac.jp

***Abstract*—In this paper, we propose a method of image feature selection for integration of image classification by Bag-of-Keypoints method and efficient search method. Our method is integration of image classification, which provides high-precision classification for various images. Therefore, in order to select a type of image feature in accordance with various purposes, it is necessary to select method easily. In addition, our method integrates various kinds of image feature by Bag-of-Keypoints method. Classification with a combination of some detectors and descriptors is more effective than with single detector and descriptor. To realize combination of some detectors and descriptors, we propose an integrating method.**

This paper is LATE BREAKING PAPERS for CSCI-ISAI.

***Keywords*-Bag-of-Keypoints, Image classification, Method integration, Feature Selection**

I. Introduction

Recently, anyone take photos easily and post them on the Internet since mobile phone and smartphone become popular. Therefore, huge amount of images are always accumulated and scattered on the Web. The research that is using their images as a huge image database has developed. Constructing that image database requiers indexing which give contexts to image. When a computer understands contents of images, it is possible to give indexes to each images automatically. Automatic Indexing is realized using image classification method.

Extratcting image features is popularly applied to image classification. Image feature is a vector generally, which is roughly classfied into global features or local features. Our method focus on local features because that are devised many types of feature. Local descriptor has been found that description of value or vector, which represents image feature from keypoints detected by keypoint detector. It is able to match similar images to calculate similarity of between local descriptors. Many local descriptors are proposed such as SIFT [1], SURF [2] and ORB [3], they have a different metric, dimension and usage. A structure and an usage of local descriptors are indicated. Still, an effectiveness for what kind of material images or material purposes is not indicated. Therefore, users construct a prototype system including various local descriptors and validate an effectiveness of local decriptors. Namely, it is necessary to select image features easily.

We proposed a method of image feature selection for integration of image classification by Bag-of-Keypoints method. Our method provides effective detectors and descriptors as output from classified image sets as input. This integration method makes better precision for classification. Our system has an evaluation measure, that is calculated to varidate which detectors and descriptors are better for image sets. Our system searches and suggests the best detector and the best descriptor. It is able to select automatically detector and descriptor without implementing image features experimentally.

In addition, our method provides a combination of some detectors and descriptors. There is a possibility classification with a combination of some detectors and descriptors is more effective than with single detector and descriptor. In order to realize a combination of some detectors and descriptors, we proposed an integrating method by Bag-of-Keypoints method [4].

In this paper, we propose efficient search on our method without Brute-force search. Our previous method has a problem that cause a combinational explosion following increase the number of detectors and descriptors, there is a possibility not to finish a search for the best detectors and descriptors. In order to solve this problem, we propose

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.72

Table I
FORMAT OF LOCAL DESCRIPTORS : A DIFFERENT DESCRIPTOR HAS A DIFFERENT METRIC, DIMENSION AND USAGE.

Algorithm	Format	Distance(general)
SIFT	128 dimensional vector	Euclidian-distance
SURF	64 dimensional vector	Euclidian-distance
ORB	256 bit binary code	Hamming-distance

two prunings. The first pruning set a limit to the number of detectors and descriptors that included in a combination. The second pruning filter detectors and descriptors that include in a combination. The efficient search is realized by their pruning apply to our method.

II. RELATED WORK

A. Keypoint detection

Keypoint is a point which represents a feature of an image. Keypoint detection has been found that detection of keypoints on an image. The number of keypoints and locations of keypoints are changed according to kind of detector. For examples of keypoint detector, Difference of Gaussian(DoG) [1],Hessian matrix approximation(which is adopted by SURF [2]), Harris Corner detector [5], FAST [6],and so on.

B. Local descriptor

Local descriptor has been found that a description of value or vector, which represents image feature from keypoints detected by keypoint detector. Local descriptor consists of elements such as luminance , RGB color or etc. around the keypoints. Table I shows the format of local descriptors. For examples of local descriptor, SIFT [1], SURF [2], ORB [3], and so on.

C. Feature selection in a field of image

Feature selection is especially developed in a field of machine learning. I.Guyon *et al.* [7] explain feature selection method, There are manifold methods, creating ranking of feature, space dimentionality reduction, using filters, clustering, AdaBoost [8], and so on.

D. Bag-of-keypoints

Bag-of-Keypoints method is proposed by G.Csurka *et al.* [4] ,that is a simple classfier for visual categorization. This method is based on vector quantization of affine invariant descriptors of image patches. This method constructs histogram which represents frequency of descriptors, using local descriptors descripted single image. This method converts multiple descriptors to simple representation. Bag-of-Keypoints method is also called Bag-of-Features and Bag-of-Visual-Words. In this paper, it is be named Bag-of-Keypoints following the original research by G.Csurka [4].

Bag-of-Keypoints approach is motivated by an analogy to learning methods using the bag-of-words [9] reperesentation for text categorization. The bag-of-words method constructs a histogram which represents frequency of word groups. Bag-of-Keypoints method is image classfier adapted to the bag-of-words method.

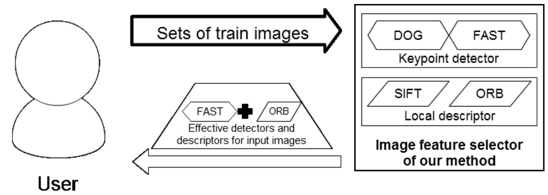

Figure 2. Overview of our method : the system by our method suggests effective detectors and descriptors.

The visual-words is a vocabulary of image, which corresponds to word groups on the bag-of-words. The visual-words consists of the features, which is a local descriptor mainly, chosen from huge amount features extracted on train images prepared for creating visual-words.

Moreover,that histogram in Bag-of-Keypoints method is a vector, is not a histogram in the strict sense of the word. In this paper, the vector constructed by Bag-of-Keypoints is called histogram observe the established procedures. Therefore, similarity of histogram written in this paper is the same as similarity of vector.

III. IMAGE FEATURE SELECTION FOR INTEGRATION OF IMAGE CLASSIFICATION

A. Image feature selection using evaluation measure

Figure 2 shows overview of our method. The user pick up several train images from large image sets that are wanted to be classified. The user classify manually that train images. The user gives classified train images as input data to our system.

Our system searches and suggests detectors and descriptors that get the best value of evaluation measure depending on train images as input. Constructing a classification system using suggested detectors and descriptors, the large image sets are expected to be classified as classification of train images.

B. Creation of Visual-words in our method

We use Bag-of-Keypoints method for image feature selection, which requires the visual-words. The visual-words' are created on each local descriptor. This process is following procedure of G.Csurka's research, thus some details are omitted. Refer to Bag-of-Keypoints paper [4] for details of creating the visual-words.

1) Keypoint detection
   We use a Grid-sampling , that is also called Dense-sampling, extracts keypoints from lattice point that divited a regular intervals. Erik Nowak *et al.* [10] show Grid-sampling is more effective sampling method on creation of the visual-words.

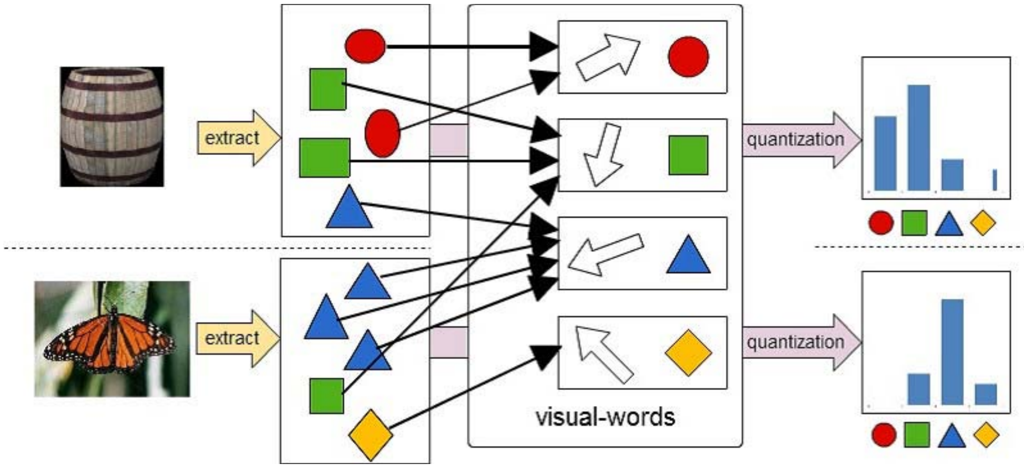

Figure 1. Bag-of-Keypoints : Multiple descriptors are converted to simple representation.

2) Local descriptor extraction
   Each kind of local descriptor is extracted from the same keypoints extraced Keypoint detection above.
3) Clustering local descriptors
   Clusters are constructed from extracted local descriptors by k-means, and output cluster centroids as the visual-words.
   The number of clusters is equal to the number of bins of histogram. Therefore, using the same number of clusters regardless of kind of local descriptor, different dimension of vector is integrated depending on local descriptor.
4) creation of visual-words correspond to local descriptor

   The visual-words' are created corresponding to each kind of local descriptor. For example, when the system is implemented include three local descriptors such as SIFT, SURF and ORB, requires three visual-words' such as visual-words for SIFT, visual-words for SURF and visual-words for ORB.

C. Integration method of local descriptors with Bag-of-keypoints

We propose integration method of local descriptors to classify for better precision. Our method integrates some detectors and some descriptors. Therefore, the system by our method suggests a combination of some detectors and some descriptors.

1) Integration of keypoint detectors: To realize the integration of keypoint detectors, each set of keypoints by each detector are unified into single set of keypoints. Figure 3 shows this method.

2) Integration of local descriptors: To realize the integration of local descriptors, local descriptors from the same keypoints are concatenated into single vector. However,feature vectors extracted by different descriptor has

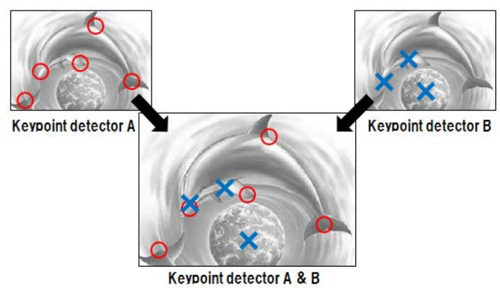

Figure 3. Integration of keypoint detectors : "Keypoint detector A&B" is realized using both keypoints from "Keypoint detector A" and "Keypoint detector B"

independent metric, dimension and usage. Therefore, vectors are concatenated after integration of local descriptors by simplification with Bag-of-Keypoints.

Figure 4 shows integration of local descriptors. In our method, local descriptors are converted to meta-local-descriptors by Bag-of-Keypoints method. Local descriptors that have different metric, dimension and usage are integrated into the same metric, dimension and usage of Bag-of-Keypoints. Therefore, a similarity of image is computed due to calculate a distance between of vectors as concatenated meta-local-descriptors. We adopt Euclidean distance for our method refering to the research by G.Csurka *et al.* [4] .

D. Pruning for search the best combinations of detectors and descriptors

When calculate a evaluation measure to varidate combinations of detectors and descriptors, our previous method has a problem that cause a combinational explosion following increase the number of detectors and descriptors. The number of detectors α and the number of descriptors β are defined, our previous method makes process in $\mathcal{O}(2^{\alpha+\beta})$. There is

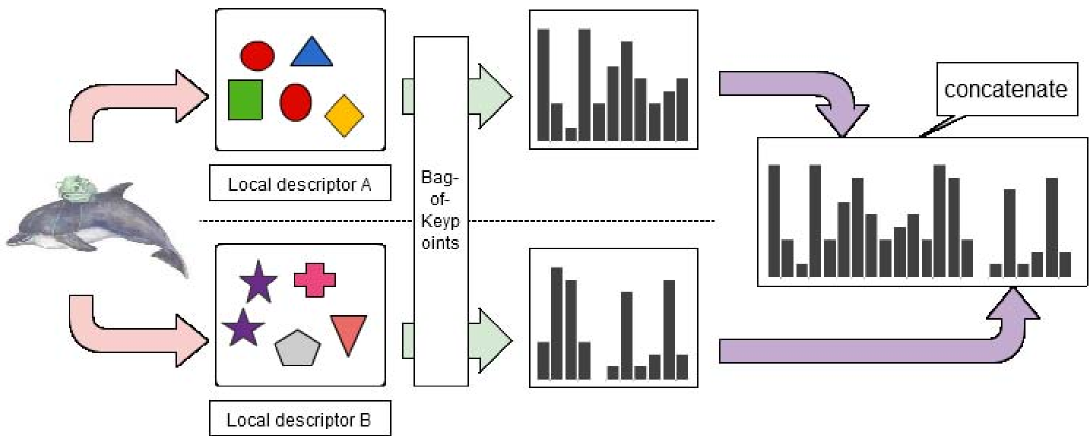

Figure 4. Integration of local descriptors : “Local descriptor A&B” is realized to concatenate meta-local-descriptor A and meta-local-descriptor B created by bag-of-keypoints

a possibility not to finish a search of many detectors and descriptors.

In order to solve this problem, we propose two prunings below. These prunings are efficient searches on our method without Brute-force search.

1) Set a limit to the number of detectors and descriptors that included in a combination: The first pruning is to set a limit to the number of detectors and descriptors that included in a combination.

The maximum number of detectors and descriptors that include in a combination n is defined. Our method makes process in $\mathcal{O}((\alpha+\beta)^n)$ at most.

In addition,our method has a tendency to increase evaluation time following to increase the number of detectors and descriptors include in a combination. This pruning reduces evaluation time to except a combination include many detectors and descriptors preferentially.

2) Filtering detectors and descriptors that include in combination by ranking: The second pruning is to compose a ranking by extraction with single detector and single descriptor before combining detectors and descriptors. Detectors and descriptors that ranked higher use for evaluation.

The number of detector and descriptors extracted from the top of ranking m is defined. Our method makes process in $\mathcal{O}(2^m)$ at most.

IV. System flow

This section shows our system flow. Figure 5 shows processess.

A. Compose and input train images

The user picks up several train images from a large image sets that are wanted to be classified. The user classifies manually that train images. Classified train images are used later as a correct image sets. The user gives classified train images as input data to our system.

B. Prunings

1) Set a limit to the number of detectors and descriptors that included in a combination: The user determines the maximum number of detectors and descriptors that included in a combination n. Combinations that consisted of more than n features are ignored in following process.

2) Filtering detectors and descriptors that include in combination by ranking: The user determines the number of detector and descriptors extracted from the top of ranking m. The system finds out features beforehand that cause a bad condition.

The system composes a ranking by extraction with single detector and single descriptor. The system considers bad features that located less than m -th, which combinations included that features are ignored in following process.

C. Keypoint detection

Each keypoint detector extracts set of keypoints from input data. The number of sets of keypoints is equal to the number of implemented keypoint detector. Various kind of sets of keypoints are extracted correspond to keypoint detector.

D. Integration of keypoint detector

More than 2 sets of keypoints are unified into single set of keypoints, an unified set of keypoints is viewed as keypoints extracted by integrated keypoint detectors, following III-C1. This operation is equal to extract from the same image by several keypoint detectors. Therefore,the system creates unified set of keypoints that reperesents property of several keypoint detectors. α keypoint detectors are implemented, at most $(2^\alpha - 1)$ sets of keypoints are obtained.

E. Extraction of Local descripter

Each local descriptors are extracted from $(2^\alpha - 1)$ sets of keypoints. β method of extrating local descriptor are

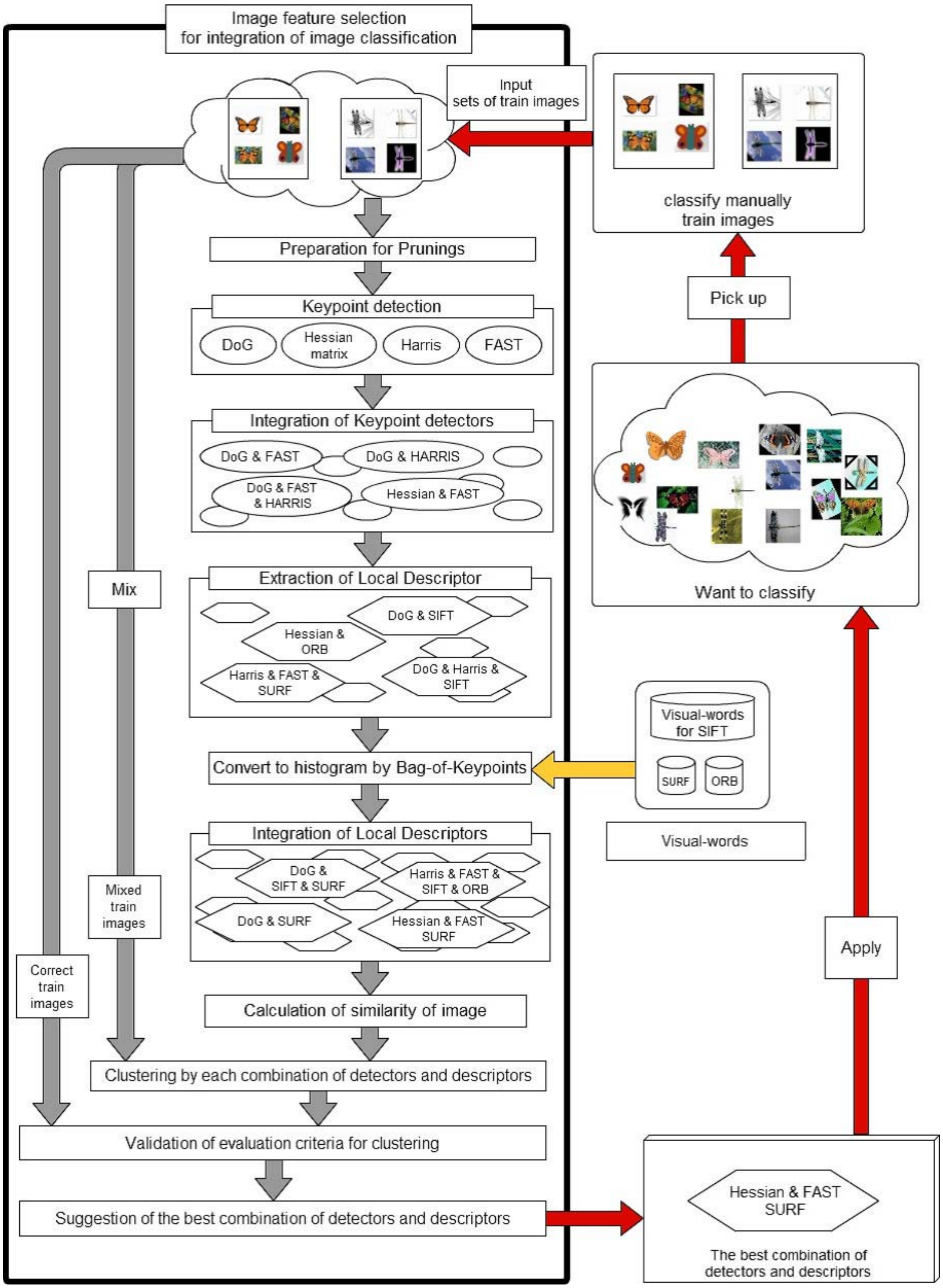

Figure 5. System Flow

implemented, at most $(2^{\alpha} - 1) \cdot \beta$ sets of local descriptors are obtained.

F. Convert to histogram by Bag-of-Keypoints

Histograms are composed with Bag-of-Keypoints. Above sets of local descriptors converted to sets of histogram. At most $(2^{\alpha} - 1) \cdot \beta$ histogram per a image are obtained.

G. Integration of Local descriptors

Histograms that converted on section IV-F are concatenated into single histogram corresponding to integrated local descriptors, following III-C2. A concatenated histogram is viewed as integrated local descriptors. Bins of histogram and a range of value are integrated in this process, it is able to concatenate without protruding certain local descriptor. Histograms are concatenated in at most $(2^{\beta}-1)$ ways.

H. Calculation of similarity of image

A similarity of between images is calculated with concatenated histograms that converted by the same combination of detectors and descriptors. A similarity is calculated in the Euclidean distance.

I. Clustering train images

The number of detectors α and the number of descriptors β are implemented, The system makes $(2^{\alpha}-1)\cdot(2^{\beta}-1)$ combinations of detectors and descriptors via above processes. Accordingly, the system mixes train images and performs clustering with similarities of train images on every combinations. The number of clusters is equal to the number of groups of train images.

J. Validation of evaluation measure

Precisions of clusters are calculated using evaluation measure to compare result of clustering and correct train images. The best clustering is a cluster that shows the best value calculated by evaluation measure.

K. Suggestion of the best combination of detectors and descriptors

The system suggests a combination of detector and descriptor that shows the best result of clustering. If several combinations show the same value, it suggests a combination that performs on the shortest time. The reason why there is purpose to extract features from video in real time, and extracting features such as SURF [2], FAST [6] or etc. is recommended processing time rather than precision.

V. Conclusion

In this paper,we propose a method of image feature selection for integration of image classification by Bag-of-Keypoints method. This integration method provides effective image classification for images given by user. This method is able to suggest the best combination of detectors and descriptors. In addition, this method suggests by efficient search using proposed two pruning without Brute-force search. The suggested feature by system validates the evaluation measure.

This integration method provides only an effective combination of detectors and descriptors. We will design the meta-system that provides the system that implemented including an effective combination of detectors and descriptors. It is necessary to not implement a suggested combination. This meta-system is more familier with users.

There are future works in this paper. The first is to plan the adjustment method of the parameters. Inappropriate parameters cause the best combination to be pruned possibly. The second is implementation other method of pruning. Feature selection is especially developed in a field of machine learning. There is a possibility of reducing search costs in our method that applied to various kind of feature selection.

In addition, we focus on Image Classification in this paper. There are many purpose of Image Features such as *Object Recognition*, *Image Retrieval* and so on. We will propose integration methods for purpose other than Image Classification.

References

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *International Conference on Computer Vision*, Barcelona, 11/2011 2011.

[4] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *In Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, pp. 1–22.

[5] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[6] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*, ser. ECCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 430–443.

[7] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[8] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," 1995.

[9] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.

[10] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Proceedings of the 9th European Conference on Computer Vision - Volume Part IV*, ser. ECCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 490–503.