ISM TRECVID2008
High-level Feature Extraction

Tomoko Matsui[1], Jean-Philippe Vert[2,3], Shin'ichi Satoh[4], Yuji Uchiyama[5]

[1]Institute of Statistical Mathematics, Tokyo, Japan
[2]Centre for Computational Biology, Mines ParisTech, Fontainebleau, France
[2]Institut Curie, Inserm U900, Paris, France
[4]National Institute of Informatics, Tokyo, Japan
[5]Picolab Co., Ltd, Tokyo, Japan

ABSTRACT

We studied a method using support vector machines (SVMs) with walk-based graph kernels for the high-level feature extraction (HLF) task. In this method, each image is first segmented into a finite set of homogeneous segments and then represented as a segmentation graph where each vertex is a segment and edges connect adjacent segments. Given a set of features associated with each segment, we then obtain a positive definite kernel between images by comparing walks in the respective segmentation graphs, and image classification is carried out with an SVM based on this kernel. We submitted six runs using this method with several combinations of the values of the kernel and SVM parameters.

1. INTRODUCTION

The HLF task can be regarded as a set of supervised binary classification tasks, where each image must be assigned a set of binary labels to indicate whether or not it belongs to each concept class. Unlike more specific tasks such as face or character recognition, the emphasis in HLF is on obtaining generic and versatile automatic tools that can learn any concept from a set of examples belonging to the concept class.

For the HLF task, we investigate a strategy where each image is first automatically segmented into a finite set of "homogeneous" segments and then represented as a segmentation graph, where each vertex is a segment and edges connect adjacent segments. A set of features such as size, color, and texture are associated with each segment. Using this graph-based representation, we apply a graph classification method to classify the images. More precisely, we investigate the use of graph kernels in combination with support vector machine (SVM) classification.

2. METHOD

Our method for HLF contains three steps, as shown in Figure 1: (i) image segmentation, (ii) kernel calculation, and (iii) SVM classification. In (i), each input image is automatically segmented and represented as a segmentation graph, as explained in Section 2.1. In (ii), a walk-based positive definite kernel between segmentation graphs is computed, as explained in Section 2.2. Finally, HLF treated as a set of binary classification problems is performed with an SVM using the walk-based kernel between segmentation graphs to classify images.

2.1. Graph-based representation of images

The first step of our approach is to automatically split each image into a variable number of homogeneous regions, using an unsupervised segmentation method[2], as in Figure 2. The image is then represented as a *segmentation graph*, i.e., a simple graph $G = (V, E)$, whose vertices V are the segments obtained by automatic segmentation and whose edges E connect vertices corresponding to adjacent segments of the image. The number of vertices (i.e., of segments) depends on the image. Furthermore, each segment is characterized by a set I of 23 features presented in Table 1. The 12 texture features (nos. 12–23) are the responses to a small filter bank of orientation and spatial-frequency selective linear filters[3]. For each segment $v \in V$ of a segmentation graph, we denote by $F(v) = (f_i(v))_{i \in I} \in \mathbf{R}^I$ the vector of the features (23-dimensional in our case).

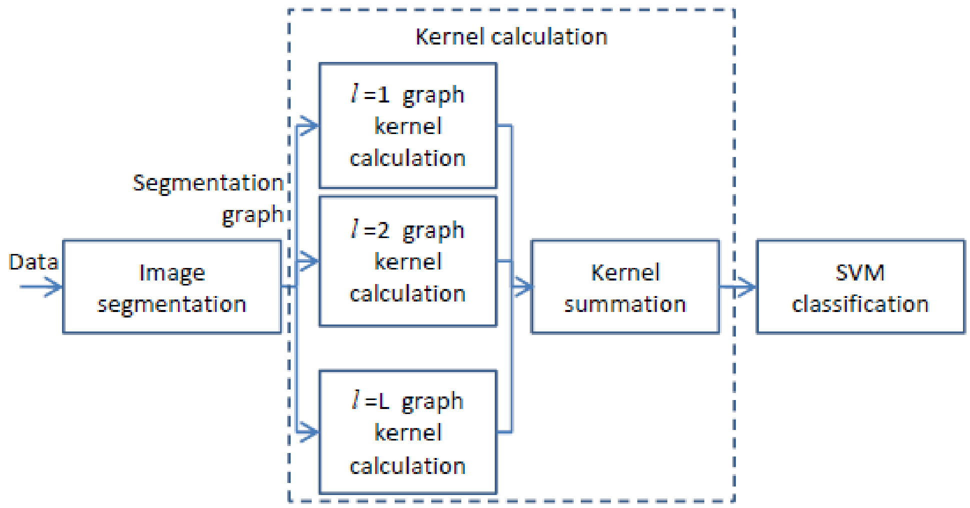

Figure 1. Overall procedure of our method.

Figure 2. Example of segmented image from data set of TRECVID2005.

Table 1. Features characterizing each image segment.

Feature no.	Description
1	Average x
2	Average y
3	Area in pixels
4	Boundary length divided by area
5	Second moment of area
6–8	Average red, green, blue (RGB) intensities
9–11	Standard deviations of RGB intensities
12–23	Texture features

2.2. Walk-based graph kernel

We use the notion of walk-based graph kernels[4,5,6] to define positive definite kernels between segmentation graphs. We note that similar ideas were previously investigated by Harchaoui and Bach[7] using the notion of subtree graph kernels[8,9], and by Aldea et al.[10] using the notion of a marginalized kernel[4], in both cases for more specific image classification problems.

In order to define the walk-based graph kernel, we first define a walk w in a graph $G = (V, E)$ as a finite sequence of connected vertices, i.e., $w = (v_1, ..., v_l)$ with $v_i \in V$ for $i = 1, \ldots, l$ and $(v_i, v_{i+1}) \in E$ for $i = 1, \ldots, l$. Here, l is called the length of walk w. Furthermore, we impose the constraint that the walk does not totter in the sense of [6], i.e., that $v_i \neq v_i + 2$ for $i = 1, \ldots, (l-2)$. We denote by $W_l(G)$ the set of walks of length l in G.

We now define positive definite kernels between vertices. For any vertices in two graphs $v_1 \in V(G_1)$ and $v_2 \in V(G_2)$, we define a kernel between v_1 and v_2 as a kernel between their respective features, e.g., a Gaussian kernel:

$$K_V(v_1, v_2) = \exp\left(-\gamma \left\| f_I(v_1) - f_I(v_2) \right\|^2\right) = \exp\left(-\gamma \sum_{i \in I} (f_i(v_1) - f_i(v_2))^2\right). \quad (1)$$

Given two walks of length l in two graphs $w = (v_1, ..., v_l) \in W_l(G)$ and $w' = (v'_1, ..., v'_l) \in W_l(G')$, we now define a walk kernel between w and w' as the function:

$$K_W(w, w') = \prod_{i=1}^{l} K_V(v_i, v'_i). \quad (2)$$

Then we define the walk-based graph kernel of depth l between two graphs G and G' as

$$K_l(G, G') = \sum_{w \in W_l(G)} \sum_{w' \in W_l(G')} K_W(w, w'). \quad (3)$$

It should be noted that if $l = 1$, no adjacency information is taken into account in the kernel. An image is then considered to be a "bag-of-segments", and the kernel between two images is simply the sum of the vertex kernels between all possible pairs of segments. When $l > 1$, the adjacency information is taken into account.

Finally, we define the walk-based kernel as the sum for multiple depths $l = 1, \ldots, L$ between two graphs G and G' as

$$K_{L-SUM}(G, G') = \sum_{l=1}^{L} K_l(G, G'). \quad (4)$$

We implemented the walk-based graph kernel using a recursive process, as explained in [6]. Since this kernel is positive definite, we can perform image classification with an SVM using the kernel on the segmentation graph representation of the images.

3. EXPERIMENTS

3.1. Description of our submitted runs

We submitted six runs using our method with six combinations of the values of $\gamma = 8, 16$ in eq. (1) and the penalty parameter of the error term $C = 1, 10, 100$ in SVM. These values were selected through the benchmark experiments as explained in Section 3.2. We set $I = [1, 23]$ and $L = 5$ for the walk-based graph kernel.

For the training data, we used 28,835 key frames (13,835 positive and 15,000 negative frames). The positive frames consist of 8,029 positive frames in the TRECVID 2008 development data set and 5,806 positive frames in the TRECVID 2005 development and test data sets. The negative frames were randomly selected from 28,233 negative frames in the TRECVID 2008 development data set.

Unfortunately there was a mistake in our script used in the experiments for the runs, and our results submitted were incorrect. We revised the script and reevaluated our method using the ground truth data ("feature.qrels.tv08"). Figure 3 shows the inferred average precisions (AP) for our method with $\gamma = 16$ and $C = 10$ and the median score of the TRECVID 2008 HLF results ("feature.stats.tv08") for each feature. The performance of our method was close to the median and the averaged inferred AP was 0.0406.

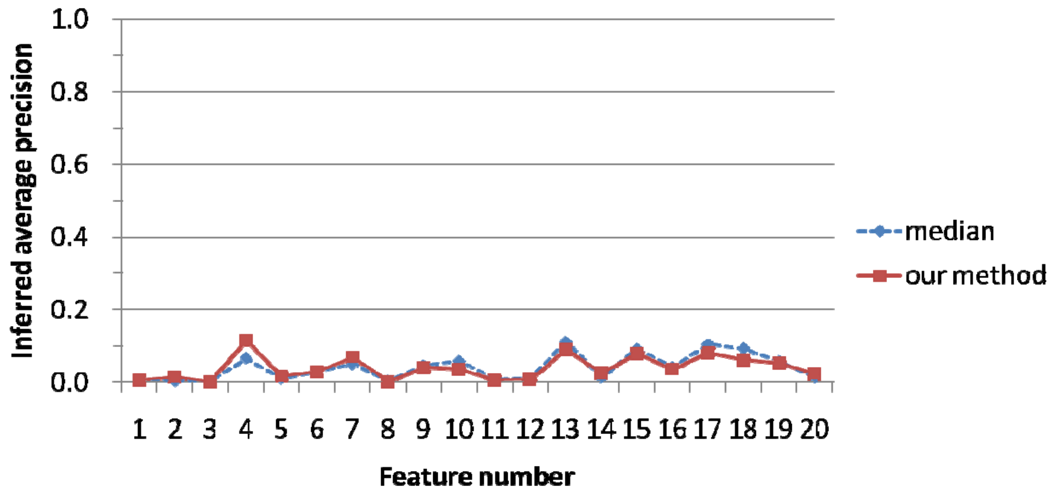

Figure 3. Inferred APs of our method with $\gamma = 16$ and $C = 10$ and the median scores of the TRECVID 2008 HLF results.

3.2. Benchmark experiment for the graph kernel and SVM parameters

To find appropriate values of the graph kernel and SVM parameters, γ and C, we tested our method in the benchmark experiment (called "Experiment 1") of the MediaMill challenge problem[1], which is often used as a benchmark for HLF systems. This problem contains data from the HLF track of the TREC Video Retrieval Evaluation (TRECVID) 2005/2006 benchmark[11]. The goal is to assign one or several of 101 concepts to individual images extracted from videos. The dataset contains 30,993 images in the training set and 12,914 in the test set, both with human annotation. We set $I = [1, 23]$ and $L = 5$ for the walk-based graph kernel.

Figure 4 shows the mean average precisions (MAPs) for different parameters of the graph kernel in eq. (1), $\gamma = 5$, 8, or 16, and for different penalty parameters of the error term of the SVM, $C = 1$, 10, or 100. The MAP with $\gamma = 16$ and $C = 10$ was the best and 0.341. Therefore, these parameters were centrally selected for our runs. It is noted that the MAP of the baseline method, called "Experiment 1", of the MediaMill challenge problem was 0.261, and our method outperformed the baseline method (relative increase of 58%).

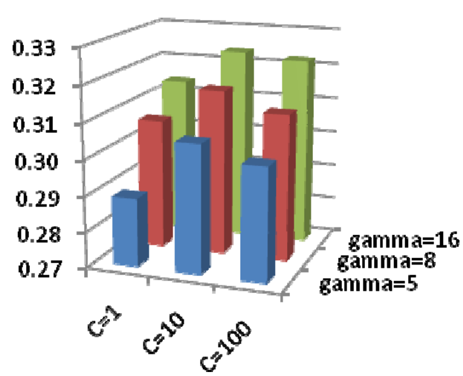

Figure 4. MAPs with several values of the graph kernel and SVM parameters.

4. CONCLUSIONS

In this paper, we described our HLF method using the walk-based graph kernel. For the TRECVID 2008 HLF task, the inferred APs of our method were close to the median scores of the TRECVID 2008 HLF results. In the benchmark experiment on the MediaMill challenge problem, we obtained a relative increase of 58% compared with the baseline performance.

5. REFERENCES

[1] G. M. Snoek, M. Worring, J. C. van Gemert, J. M. Geusebroek, and A. Smeudlders. The challenge problem for automated detection of 101 semantic concepts in multimedia. Proc. ACM Multimedia, 2006.

[2] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. IEEE Trans. Pattern Anal. Mach. Intell., 23(8):800–810, Aug 2001.

[3] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. Int. J. Comput. Vision, 43(1):29–44, 2001.

[4] T. Gärtner. Exponential and Geometric Kernels for Graphs. In NIPS Workshop on Unreal Data: Principles of Modeling Nonvectorial Data, 2002.

[5] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized Kernels between Labeled Graphs. In T. Faucett and N. Mishra, editors, Proceedings of the Twentieth International Conference on
Machine Learning, pp. 321–328. AAAI Press, 2003.

[6] P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert. Graph kernels for molecular structureactivity relationship analysis with support vector machines. J. Chem. Inf. Model., 45(4):939–51,
2005.

[7] Z. Harchaoui and F. Bach. Image classification with segmentation graph kernels. In 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), pp. 1–8. IEEE Computer Society, 2007.

[8] J. Ramon and T. Gärtner. Expressivity versus efficiency of graph kernels. In T. Washio and L. De Raedt, editors, Proceedings of the First International Workshop on Mining Graphs, Trees and Sequences, pp. 65–74, 2003.

[9] P. Mahé and J.-P. Vert. Graph kernels based on tree patterns for molecules. Technical Report ccsd-00095488, HAL, September 2006.

[10] E. Aldea, J. Atif, and I. Bloch. Image classification using marginalized kernels for graphs. In Graph-Based Representations in Pattern Recognition, volume 4538/2007 of Lecture Notes in Computer Science, pp. 103–113. Springer Berlin/Heidelberg, 2007.

[11] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVid. In Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pp. 321-330. ACM Press, New York, NY, 2006.