HIGH-LEVEL FEATURE EXTRACTION USING SVM WITH WALK-BASED GRAPH KERNEL

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

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

ABSTRACT

We investigate a method using support vector machines (SVMs) with walk-based graph kernels for high-level feature extraction from images. 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. In a benchmark experiment on the MediaMill challenge problem, the mean average precision increased from 0.216 (baseline) to 0.341 when our method was utilized.

Index Terms— High-level feature extraction, graph kernel, walk kernel, support vector machine.

1. INTRODUCTION

Our goal is to develop a method of high-level feature extraction (HFE) from images, e.g., detecting whether an image is a landscape or contains an object such as a car or a dog. If a list of concepts (such as "is a landscape" or "contains a dog") is given, this task can also 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 HFE is on obtaining generic and versatile automatic tools that can "learn" any concept from a set of examples belonging to the concept class.

To reach this goal, 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.

We confirmed the relevance and effectiveness of our method by evaluating it in a benchmark experiment on the MediaMill challenge problem[1] and we report promising results.

2. METHOD

Our method for HFE 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, HFE treated as a set of binary classification problems is performed with an SVM using the walk-based kernel between segmentation graphs to classify images.

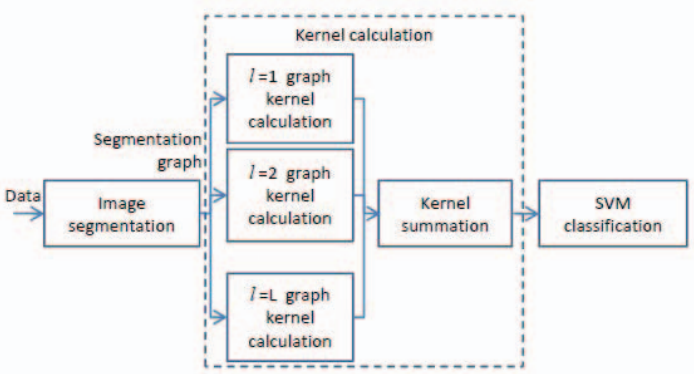

Fig. 1. Overall procedure of our method.

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]. Among them, the first six filters are the first derivatives of the Gaussian filter with $\sigma_x = 1$ (pixel) and $\sigma_y = 3$, with six orientations. The following six filters are the second derivatives of the Gaussian filter with $\sigma_x = \sqrt{2}$ and $\sigma_y = 3\sqrt{2}$, again with six orientations. 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).

Fig. 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. Data specification and benchmark experiment

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 HFE systems. This problem contains data from the HFE 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 compared our method with the baseline method of [1] in which each image is first converted to a 120-dimensional vector of visual features, and classification is performed with an SVM. Visual features express image concepts and each feature corresponds to a bin characterizing one of the pairs of 15 low-level visual

concepts (e.g., road and sky) and 8 concepts characterizing both global and local color-texture information.

To assess the performance of the methods, we measured for each concept the average precision (AP) as the area under the precision/recall curve. The AP for each concept was averaged to produce the final mean average precision (MAP).

In the experiments, we set I = [1,23] and L = 5 for the walk-based graph kernel. For γ in eq. (1) and the penalty parameter of the error term C, we set γ = 16 and C = 10. These values were selected through three-fold cross validation on the training set, as explained in Section 3.3.

3.2. Comparison with baseline performance

The APs for the concepts are shown in Figure 3. For almost all concepts, except for 11 concepts out of 101, ("candle," "drawing," "fireweapon," "hassan_nasrallah," "motorbike," "nightfire," "people_marching," "racing," "religion_leader," "sharon," and "tank"), the APs obtained with our method were higher than the baseline ones. The MAP was 0.341 for our method versus 0.216 for the baseline method. This corresponds to a significant relative increase of 58%. Since both the baseline method and our method are based on an SVM, the difference in performance highlights the importance of choosing the correct kernel.

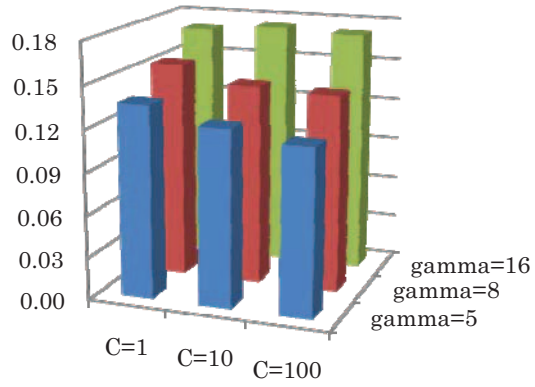

Fig. 4. MAPs for three-fold cross validation for the graph kernel and for SVM parameters γ and C.

3.3. Cross validation for the graph kernel and SVM parameters

Figure 4 shows the MAPs in three-fold cross validation on the training set for different parameters of the graph kernel in eq. (1), γ = 5, 8, or 16, and for different regularization parameters of the SVM, C = 1, 10, or 100. The MAP with γ = 16 and C = 10 was the best, and these parameters, selected using only training data, were therefore selected for evaluating the test data (section 3.2). We note that the choice of parameters can have an important effect on the final performance, so parameter tuning is important. It should be noted that, in an open evaluation of the different parameters on the test data, we also obtained the best MAP with the same parameter values.

3.4. Influence of walk length

An important parameter of walk-based kernels is the length of the walks considered. By default, we summed the kernels corresponding to walks of lengths 1 to 5, which amounts to considering all walks of length 1 to 5 to compare two graphs. To further assess the importance of the length and of combining different lengths together, we investigated the performance of walk-based kernels for a specific length (3) and compared them with the sum kernel (4). To save computation time, here we use only a subset of the benchmark data and focus on only 39 concepts out of 101. The subset contains 1000 images selected randomly from the original training and 1000 similarly selected from the testing data. MAPs obtained when separately using the graph kernels with depth l = 1, …, 5 and the sum (L-SUM) are shown in Figure 5. Recalling that the case l = 1 corresponds to not taking into account the adjacency information of segments on the images, we first observe a significant improvement when this adjacency information is taken into account (l>1). Second, we observe that the sum kernel is better than each individual kernel, suggesting that walks of different lengths may be relevant for classification, and that simply adding together kernels for different lengths is a simple yet effective way to exploit this information.

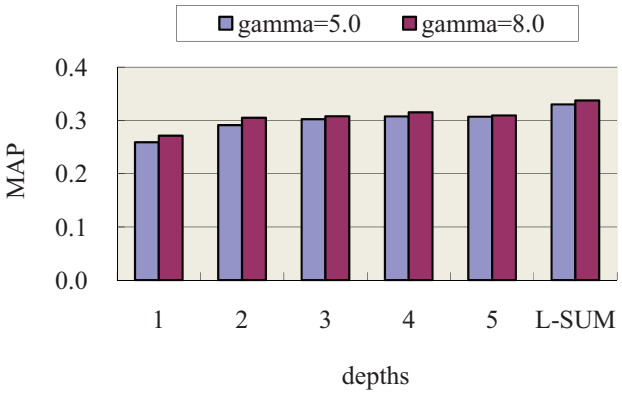

Fig. 5. MAPs for 39 concepts as a function of depths of graph kernel

4. CONCLUSIONS

In this paper, we investigated an HFE method using the walk-based graph kernel. In the benchmark experiment on the MediaMill challenge problem, we obtained a relative increase of 58% compared with the baseline performance. This confirms the relevance of our approach.

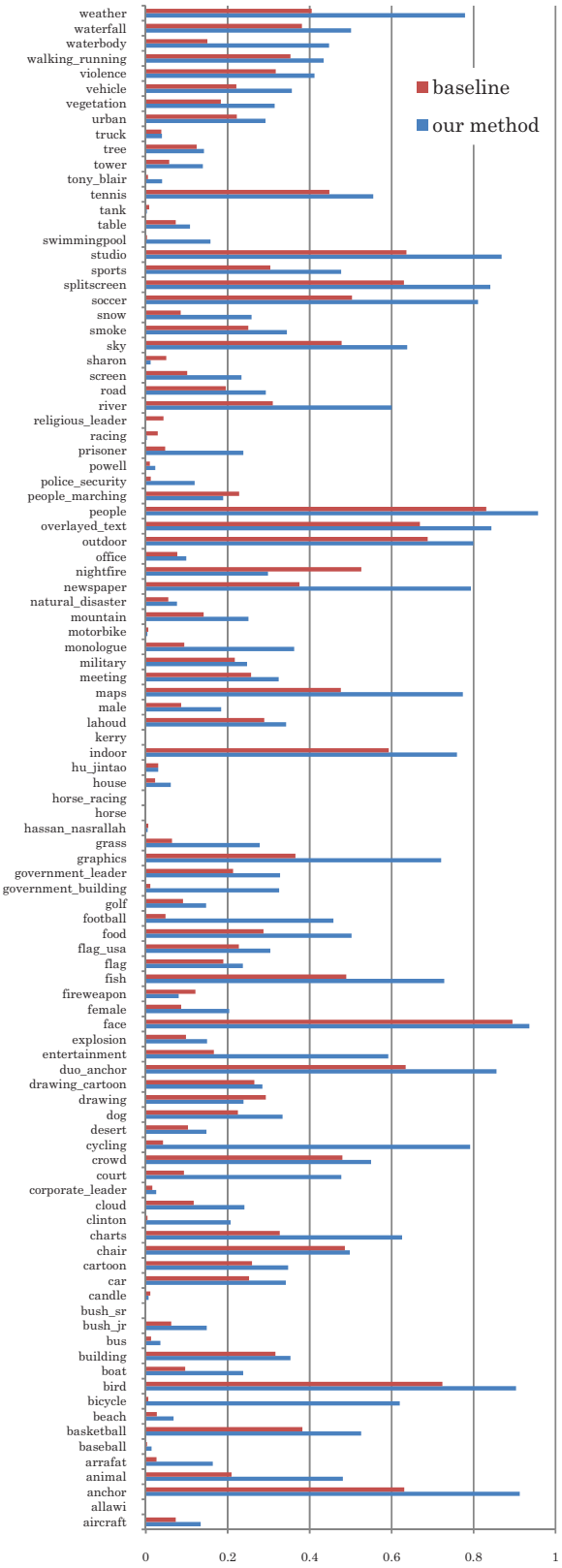

Fig. 3. APs for each concept using $K_{L\text{-}SUM}$.

Our future work will include kernel design for not only vertices but also edges of the segmentation. Moreover, we plan to investigate the optimal subset I selection for features and the weighted sum of the walk-based graph kernels.

5. ACKNOWLEDGMENTS

Part of this work was supported by the Function and Induction Research Project, Transdisciplinary Research Integration Center, Research Organization of Information and Systems.

6. 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.