

LETTER

Tree Fusion Method for Semantic Concept Detection in Images

Jafar MANSOURI^{†a)}, Student Member and Morteza KHADEMI[†], Nonmember

SUMMARY A novel fusion method for semantic concept detection in images, called tree fusion, is proposed. Various kinds of features are given to different classifiers. Then, according to the importance of features and effectiveness of classifiers, the results of feature-classifier pairs are ranked and fused using C4.5 algorithm. Experimental results conducted on the MSRC and PASCAL VOC 2007 datasets have demonstrated the effectiveness of the proposed method over the traditional fusion methods.

key words: semantic concept detection in images, fusion, C4.5 algorithm

1. Introduction

With the ever increasing development of communication systems, Internet and network technology, and storage and compression technology, a great number of images are available to customers. Automatic semantic concept detection in images is a tool for managing and retrieving these data, and it is essential for many applications like content-based image search. Image concept detection systems are mainly based on classification models in which concepts are assigned to the image by learning the relationship between concepts and low-level visual features extracted from the image [1], [2].

It is recommended that for a better detection, usually a large number of features and various classifiers with different characteristics should be used. Every classifier has its own advantages and weaknesses relative to features with different characteristics (for a feature, one classifier gives better result and for another feature, another classifier gives better result). Since usually many low-level features like color, texture, edge and various classification algorithms like SVM and K-nearest neighbor are used, a fusion must be used. Three traditional fusion methods are early, late, and multiple kernel learning (MKL) fusions. In the early fusion, all low-level features are first put into a feature vector, and classification is then performed on the feature vector. In the late fusion, different types of features are separately given to classifier(s). Then, results of classifications are combined. The combination of the above methods can also be used. MKL is another method for the fusion of different features and special classifiers like SVM, using the kernel trick [3].

In this Letter, a new fusion method, named tree fusion, is presented. Different types of features and classifiers are

used and the results of feature-classifier pairs are fused using the well-known C4.5 algorithm. Experimental results from MSRC [4] and PASCAL VOC 2007 [5] datasets show the superiority of the proposed fusion method over the traditional fusion methods for the semantic concept detection in images.

2. C4.5 Tree Fusion

2.1 The Conventional Decision Tree Algorithm

The C4.5 is a famous decision tree classifier which has the capability to induce classification rules in the form of decision trees from a set of examples [6]. C4.5 grows a tree using the divide-and-conquer algorithm for a set S of cases as follows [7]–[9]:

- If all the cases in S belong to the same class or S is small, the tree is a leaf labeled with the most frequent class in S .
- Otherwise, choose a test based on a single attribute with two or more outcomes. Make this test the root of the tree with one branch for each outcome of the test, partition S into corresponding subsets S_1, S_2, \dots according to the outcome for each case, and apply the same method recursively to each subset. C4.5 usually uses the gain ratio criterion to rank possible tests.

The derived model summarizes all given information from training data set but states it in a more concise manner. The test process is in the form of passing through a path in the built tree from the top of the tree to a certain leaf node and the related class label is assigned into the case. For more details about the C4.5, please refer to [7]–[9].

2.2 The Proposed Tree Fusion Method

In the proposed method, the C4.5 algorithm is used to learn a fusion, and the induced fusion rules are expressed in the form of a decision tree. In the tree, attributes are feature-classifier pairs and classes are concepts. In the tree fusion, for each concept the tree determines which feature-classifier pairs are useful and ranks them in accordance with their importance as described above. The most important and informative feature-classifier pair is in the top of the tree. To clarify the algorithm, an example illustrating a sample of a tree structure created with feature-classifier pairs is shown in Fig. 1 for a classification of a specific concept. In this

Manuscript received February 5, 2014.

Manuscript revised April 23, 2014.

[†]The authors are with Electrical Engineering Department, Ferdowsi University of Mashhad, Iran.

a) E-mail: jafar.mansouri@gmail.com

DOI: 10.1587/transinf.E97.D.2209

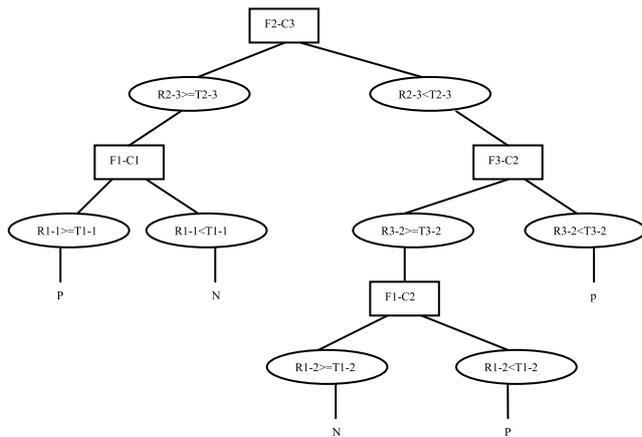


Fig. 1 An example of C4.5 tree fusion with three features and three classifiers.

example, there are 3 different features and 3 types of classifiers. In the figure, the pair of feature “i” and classifier “j” is shown with F_i-C_j . In this example, C4.5 tree determines that feature 2 and classifier 3 is the most important feature-classifier pair and first, this pair is tested. Then according to the result (condition) of classifier 3 on feature 2 (R_{2-3}), the next step is determined. If R_{2-3} is larger than threshold T_{2-3} , the pair of feature 1 and classifier 1 is tested. Otherwise, the feature 3 with classifier 2 is considered. This process continues until reaching a positive (P) or negative (N) leaf node.

This is different from the late and MKL fusions with different weights for feature-classifier pairs. In the late and MKL fusions, the results of classifications are combined (usually by addition with different weights) and the overall result is obtained. In the tree fusion, according to the conditions (results of feature-classifier pairs) that tree checks them, it is possible that some feature-classifier pairs are not used or their order changes. Furthermore, some features may be unnecessary for detection, or some classifiers may not be good classifiers for some features. These issues are intended by C4.5 tree fusion. In these situations, feature-classifier pairs that may not be necessary or suitable for detection (in accordance with the conditions), are not taken into account. For example, for detection of a concept, in the first case the values of features of texture and edge can be such that there is no need to know the values of color feature; while in the second case with different values of features of texture and edge, the values of color feature are important. C4.5 tree fusion can solve this matter and separate cases. But in the late fusion, since results of classifications are added (even with different weights), in the first case, the value of color feature can lead to the wrong detection while there is no need to color feature. This problem exists in the early fusion, too. In the MKL fusion, features and classifiers get different weights, which are obtained by an optimization process. These weights are fixed in the test phase and can not be changed in accordance with the conditions. Hence, the above problem exists in this fusion, too.

Also, in the early fusion, if features have high-dimensionality and the number of features is high, the fused vector has very high dimensionality and this can make distance concentration [10] that results in performance degradation. In the tree fusion, this problem does not exist since features are considered separately.

3. Experimental Results

To evaluate the performance of the proposed method for concept detection in images, we conducted experiments on two widely-used MSRC and PASCAL VOC 2007 image datasets. MSRC contains 591 images with 23 concepts. Two of which (‘horse’ and ‘mountain’) have been removed from the evaluation due to their few positive samples, as suggested in the evaluation protocol proposed in [4]. Thus, there are totally 21 concepts. We randomly divide the dataset into two subsets: 296 images for training and 295 images for test. The PASCAL VOC 2007 dataset consists of 20 classes. The dataset consists of 9963 images divided into a predefined training and test set of, respectively, 5011 and 4952 images. PASCAL VOC 2007 dataset is more challenging than the MSRC dataset due to more significant background clutter, illumination effects and occlusions. Note that we only use the image-level ground truth to train the models.

In this Letter, the extracted features are: 81-d color moments descriptor, 64-d color histogram descriptor, 80-d edge direction histogram descriptor, 62-d homogenous texture descriptor and 128-d SIFT descriptor. We use a set of 128-d SIFT features to describe keypoints in an image. All the SIFT features from the training set are clustered, and centroids of clusters are used to build the visual world vocabulary. We then represent each image with a histogram of visual words.

Four classifiers are used: SVM with radial basis kernel function, SVM with χ^2 kernel function, K-nearest neighbor, and Gaussian mixture model. We use four fusion methods. For Fusion1, all low-level features are concatenated (early fusion) and then the resultant feature vector is given to the classifiers and results of classifications are added (late fusion) with different weights, which are determined based on 2-fold cross-validation. For Fusion2, each feature is separately given to classifiers and then results of all feature-classifier pairs are added (late fusion) with different weights. For Fusion3, The aforementioned features and SVM with χ^2 and radial basis function kernels are used for MKL fusion [11]. Finally, in the tree fusion, each feature is given to different classifiers and fusion is performed using C4.5 algorithm on the feature-classifier pairs.

For performance evaluation, average precision (AP) is used [12]. AP approximates the area under the precision-recall curve. To evaluate the overall performance, we use mean average precision (MAP) which is the mean value of the APs over all concepts. Figures 2 and 3 depict the AP for concepts in MSRC and PASCAL VOC 2007 datasets. For MSRC dataset, the MAP for the tree fusion is 0.641

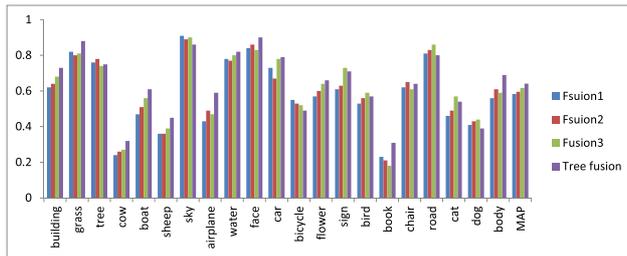


Fig. 2 Comparison of performances of four fusion methods for MSRC dataset.

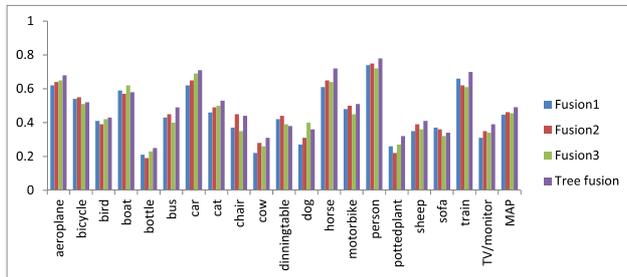


Fig. 3 Comparison of performances of four fusion methods for PASCAL VOC 2007 dataset.

while for Fusion1, Fusion2, and Fusion3 is 0.583, 0.596, and 0.617. For PASCAL VOC 2007 dataset, the tree fusion, Fusion1, Fusion2, and Fusion3 have MAPs 0.491, 0.447, 0.462, and 0.455, respectively. These show the superiority of the proposed fusion over Fusion1, Fusion2, and Fusion3, which use traditional fusions. This superiority is due to the properties of decision tree which were stated above.

4. Conclusion

In this Letter, a new method for fusion of different features and classifiers based on C4.5 decision tree was proposed for concept detection in images. The results from experiments conducted on the MSRC and PASCAL datasets demonstrated the superiority of the proposed tree fusion over the

traditional fusion methods.

References

- [1] G. Nasierding and A.Z. Kouzani, "Empirical study of multi-label classification methods for image annotation and retrieval," Proc. Int. IEEE Conf. on Digital Image Computing: Techniques and Applications, pp.617–622, 2010.
- [2] M. Jian, C. Jung, and Y. Zheng, "Discriminative structure learning for semantic concept detection with graph embedding," IEEE Trans. Multimedia, vol.16, no.2, pp.413–426, 2014.
- [3] M. Varma and D. Ray, "Learning the discriminative power-invariance trade-off," Proc. Int. IEEE Conf. on Computer Vision, pp.1–8, 2007.
- [4] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextronBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," Proc. European Conf. on Computer Vision (ECCV), pp.1–15, 2006.
- [5] M. Everingham, L. VanGool, C.K.I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes challenge 2007 (VOC2007) results," [Online]. Available: <http://www.Pascal-network.org/challenges/VOC/voc2007/workshop/index.html>
- [6] A. Fakhari and A.M.E. Moghadam, "Combination of classification and regression in decision tree for multi-labeling image annotation and retrieval," Applied Soft Computing, vol.13, pp.1292–1302, 2013.
- [7] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.-H. Zhou, M. Steinbach, D.J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," Knowl. and Inf. Syst., vol.14, no.1, pp.1–37, 2008.
- [8] J.R. Quinlan, C4.5. Programs for Machine Learning, Morgan Kaufmann, LosAltos, CA, USA, 1993.
- [9] R. Kohavi and J.R. Quinlan, "Decision tree discovery," in Handbook of Data Mining and Knowledge Discovery, ed. W. Klossan, and M. Zytchow, Chapter 16.1.3, pp.267–276, Oxford University Press, 2002.
- [10] D. François, V. Wertz, and M. Verleysen, "The concentration of fractional distances," IEEE Trans. Knowl. Data Eng., vol.19, no.7, pp.873–886, 2007.
- [11] <http://research.microsoft.com/en-us/um/people/manik/code/GMKL/download.html>
- [12] G. Ding and K. Qin, "Semantic classifier based on compressed sensing for image and video annotation," Electron. Lett., vol.46, no.6, pp.417–418, 2010.