Multi-Class Labeling Improved by Random Forest for Automatic Image Annotation

Motofumi Fukui, Noriji Kato, and Wenyuan Qi
Fuji Xerox Research & Technology Group
6-1 Minatomirai, Nishi-ku, Yokohama-city, Kanagawa 220-8668 Japan
{Motofumi.Fukui, Noriji.Kato, Wenyuan.Qi}@fujixerox.co.jp

Abstract

Recently automatic image annotation (AIA) has been arising as a key technology to support image retrieval. The representative algorithm is Semantic Multiclass Labeling (SML [1]), which constructs a parametric generative model of a distribution of local image features in a class with a gaussian mixture model. Although SML shows good accuracy, SML has not been used widely because of its long training time and annotation time. In this paper we propose a method of improving SML by dealing with Random Forest instead of the gaussian mixture model. We evaluate our proposal by using the standard corpus, Corel5K and IAPRTC-12. The experimental results demonstrate that our method can train very fast and annotate multiple labels very fast, with keeping comparable performance as existing methods.

1. Introduction

As the number of images that are taken with a digital camera is increased, image search technologies have become important to retrieve desired images from an immense amount of image repositories. Most existing image search engines are constructed based on semantic labels attached on images in advance. Then only images with labels matched to user's queries are retrieved. However, the semantic labels which are assigned by human hands sometimes do not contain important labels and include semantic ambiguities among annotators. Therefore, we consider that assigning semantic labels automatically, that is to say, automatic image annotation (AIA), is a key technology of realizing a brilliant image retrieval system.

Recently many leading ideas about AIA have been proposed over this decade [1-10]. In existing AIA techniques, models based on nearest neighbor methods [2,3] attract great deal of interest because of their high performance. In the nearest neighbor methods, the training image with the smallest distance from an input image is found in the appropriate feature space. Then the labels of the training image are assigned to the input image. TagProp [2] is the representative example of the nearest neighbor model, where the distance is measured on a metric space learned using training samples. Although TagProp has excellent results on some AIA tasks with a small data set, it would be hard to apply for realistic applications. It is because millions of the training images are necessary to cover various situations, and the system has to maintain all training data. We believe that parametric models, which represent a probabilistic relation between image features and labels, should be suitable for realistic applications, because only model parameters instead of all training data need to be maintained. SML [1] is such a parametric model based on multiple Naïve Bayes classifiers with local features, where a conditional probability of each local feature to a label is modeled by a hierarchical gaussian mixture model. Though SML shows good annotation accuracy, it has not been used widely, because of its long training time and annotation time. The main reason of the former is its model complexity, and the main reason of the latter is the large number of local features for which probabilities have to be calculated. In this paper, to deal with this problem, we propose to introduce a Random Forest classifier instead of the gaussian mixture model.

Random Forest [11] is an ensemble of decision trees and has become a popular method in many computer vision applications, for example, object segmentation [12,13], image classification [14,15], object classification [16,17], food recognition [18], object detection [19,20], video segmentation [21], and so on. In spite of its success in many applications, Random Forest has not been applied to the AIA task yet. Our main contribution is that we first utilize Random Forest in the AIA task to model a probabilistic relation between a local feature and a class label effectively. There are two advantages of utilizing Random Forest. The first advantage is short training time and annotation time, because it is not necessary to calculate distances in large dimensional feature space. And the second advantage is simplicity of its model to discriminate multiple labels by one classifier. In addition, we try to demonstrate that Random Forest can be applicable for large number of classes by an AIA experiment with more than 100 kinds of labels, though a Random Forest classifier is often used to the small number of classes. In the following section, we explain our proposal in detail.

2. Methodlogy

2.1. Image Annotation with modified SML

SML consists of multiple Naïves Bayes classifiers, each of which predicts a posterior probability of a label given local features, $f_1, f_2, \ldots, f_n$, sampled at the n locations in an image, as Figure 1. Each child node of a classifier has the same probability model $P(f|c)$ of a local feature f given a class label c, which is represented by a hierarchical gaussian mixture model in the original paper.

In this paper, we use a standard gaussian mixture model, where a weight π_j^c , an averaged feature vector m_j^c , and a diagonal covariance matrix $\sum_j^c$ of the jth gaussian distribution are estimated for each label c in a training stage, in order to compare the straightforward performance based on two kinds of classifiers. Furthermore, we use a different type of features from SML as shown in 2.3. We call this model SML*. In a testing stage, a posterior distribution *P(c|F)*of a label c given a set of feature vectors $F = \{f_i\}_1^n$ is estimated by Bayes' Theorem as the following.

$$P(c \mid F) = P(c \mid f_1,\ldots,f_n) = \frac{P(f_1,\ldots,f_n \mid c)P(c)}{P(f_1,\ldots,f_n)} = \frac{P(c)}{P(f_1,\ldots,f_n)}\prod_i P(f_i \mid c) \quad (1)$$

, where $P(f_1,\ldots,f_n \mid c)$can be factorized into $\prod P(f_i \mid c)$ in the Naïves Bayes classifiers. In Eq. (1), *P(c)* denotes a prior of a label *c*. Since *P(F)* is independent of the label, we assign a label *c(I)* on a given image *I* by the following equality.

$$c(I) = \arg\max_c \log P(c \mid F) = \arg\max_c \left\{\log P(c) + \sum_i \log P(f_i \mid c)\right\} \quad (2)$$

In SML, we have only to hold model parameters of *O(mKd)*, where *K* is the total number of labels, *d* is the feature dimension, and *m* is the number of gaussian distributions. While, the nearest neighbor models have to hold model parameters of *O(Nd)*, where *N* is the number of training images. Therefore, SML has large advantage to the nearest neighbor models for an immense volume of image repositories. However, in spite of small demand to the storage, SML has such a shortcoming as its long training time and annotation time.

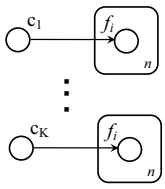

Figure 1. Multiple Naïves Bayes Model.

2.2. Probability Model with Random Forest

Now instead of the gaussian mixture model, we introduce Random Forest as a probability model [22] into an AIA task. Random Forest consists of T randomized decision trees with the D depth. A probability table *P(c|l)* of a label *c* is attached on a leaf node *l* and is calculated in a training stage (Figure 2). It has two kinds of randomness, bootstrap random sampling of a training set and random feature selection to split training data into two groups at each node of the decision tree. The random sampling means that a different subset with the *SN* number of vectors is selected from a whole training set to train each decision tree. Each node of a decision tree has two splitting parameters, a feature index and a threshold. On the node, data whose value of feature specified by the index is higher than the threshold are sent to the right child node, and the others are sent to the left. We utilize the same method as used in Extremely Randomized Clustering Forests [15] to determine the splitting parameters. In [15], the normalized Shannon entropy, as suggested in [23], is calculated on each node, and the parameters are determined such that the entropy is maximal.

A probability table *P(c|l)* can be calculated by using a whole training set, after establishing decision trees. In contrast to classification problems, in the most AIA task, an image has multiple labels. In our implementation, we assume that all labels given on an image are assigned on each feature vector extracted from the image. We estimate *P(c|l)* as the following.

$$P(c \mid l) = \frac{n_l^c + \alpha}{n_l + K\alpha} \quad (3)$$

, where n_l^c is the number of feature vectors with a label c, and n_l is the total number of feature vectors, which reaches on a leaf node *l*. α is a constant parameter (=0.01), which is set up so that *P(c|l)* is not zero. In Eq.(3), note that *ΣP(c|l)* is more than 1.0. In a testing stage, Random Forest estimates a posterior probability of a label from a set of the leaf node on which each evaluated feature vector reaches. Then the probability of a label c to a feature vector *f*, *P(c|f)* is counted by averaging *P(c|l)* for all *T* trees.

$$P(c \mid f) = \frac{1}{T}\sum_{t=1}^{T} P(c \mid l_t) \quad (4)$$

In Eq. (4), l_t denotes a leaf node in the tth tree, on which *f* reaches. *P(f|c)* is calculated by using Bayes' rule, and finally *P(c|F)* is obtained by substituting Eq.(4) into Eq.(1).

$$P(c \mid F) = \frac{P(c)}{P(f_1,\ldots,f_n)}\prod_i P(f_i \mid c) = \frac{P(c)\prod_{i=1}^{n} P(f_i)}{P(f_1,\ldots,f_n)}\prod_{i=1}^{n}\frac{P(c \mid f_i)}{P(c)} = \frac{\prod_{i=1}^{n} P(f_i)}{P(f_1,\ldots,f_n)T^n}P(c)\prod_{i=1}^{n}\sum_{t=1}^{T}\frac{P(c|l_t)}{P(c)} \quad (5)$$

, where we don't have to calculate the first term, because it is independent of the label. Usually with a Random Forest classifier, a probabilistic estimation of a label can be realized as *P(c|l)*, while in our proposal, as *P(c|l)* divided by a prior distribution *P(c)*, as seen in Eq.(5). Dividing by a prior label distribution leads to rescue of a minority label and inhibition of a majority one. This idea is similar to WRF [24] and PRAGMA [25].

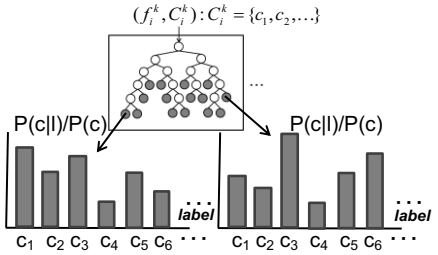

Figure 2. Training Stage by our Proposal.

2.3. Feature Extraction

In general, it is difficult to model *P(f|c)* with a large number of local features. Therefore, SML models *P(f|c)* with 8 feature vectors which consist of gaussian parameters extracted from each image. Although SML reduce training samples drastically, this reduction causes degradation of performance. In contrast, we can utilize the

large number ($n \fallingdotseq 1000$) of local features per image directly, owing to small training time of Random Forest. Our features are region features which are average and deviation of local features over a small region. They are generated as follows. First, images are divided into square regions with fixed size. Next, local features, such as pixel intensity or texture, are extracted at each pixel in the square region. At last, the average and the standard deviation of the local features are calculated.

3. Experiments and Discussion

3.1. Dataset

Firstly we compare our method with the previous approaches, using Corel5K [4]. The Corel5K benchmark set is composed of 4500 training images and 499 testing images. Each image has from 1 to 5 semantic labels, and there are 371 labels in a training set and 260 labels in a testing set. Secondly we evaluate our proposal using IAPRTC-12 [26]. IAPRTC-12 is composed of 17665 training images and 1962 test images. We have the same 291 labels as [2] and [3] both in a training set and in a testing one. In both benchmark sets, there is large imbalance among the number of images with each label.

Each image with an aspect ratio *ar* ($\geqq 1.0$) is shrinked that the length of the short side is changed into 320 pixels. We separate the short side into 24 parts of the same length and the long side into 24**ar* parts. As the results, we acquire the 576**ar* (=*n*) square local regions.

Image annotation performance is evaluated by comparing ground-truth with the labels automatically annotated. Each image is annotated with the 5 most relevance labels. The average recall (R) and the average precision (P) over all the labels, the f-measure (F), which is the harmonic of R and P, and the number of labels with recall>0 (N+) are estimated respectively.

3.2. Image Feature

After a preliminary experiment, we have selected two kinds of region features, color+gabor and cDCT. Color+gabor feature consists of three kinds of color features (RGB/Normalized-RG/CIELAB) and gabor features with 6 orientations and 3 scales. And the resulting region features constitute 52-dimensional feature vectors. While cDCT feature is the same feature as used in [1], and constitutes 126-dimensional feature vectors. All feature values are normalized such that an averaged feature value of each dimension is equal to 0.0 and the standard deviation is equal to 1.0.

3.3. Experiments and Discussion

At first we examine variations of the performance relying on the parameters of Random Forest. Figure 3 shows the results of R and P of averaging 10 trials by using Corel5K with color+gabor features, when one of *T*, *D*, and *SN* is changed from the case of *T*=2, *D*=24, and *SN*=2^{16}. Initially we notice that multiple trees need to be aggregated, because only one tree can recognize specific labels due to overfitting (Figure 3(b)). Next we note that Figure 3(a) appears similar to Figure 3(c). To acquire high performance, enough large *SN* and *D* need to be selected. Since we don't prune the branch in building Random Forest, the number of leaves seems to be proportional to *SN* and *D*. So Figure 3 shows that a lot of leaves are necessary to classify each label. When the number of leaves is set up as a large value in Figure 3(a), R is slightly decreasing, because of the deterioration of randomness among trees. To gain higher performance, in future, we suppose that we should build Random Forest, which can rescue a minority label additionally to acquire the higher recall rate.

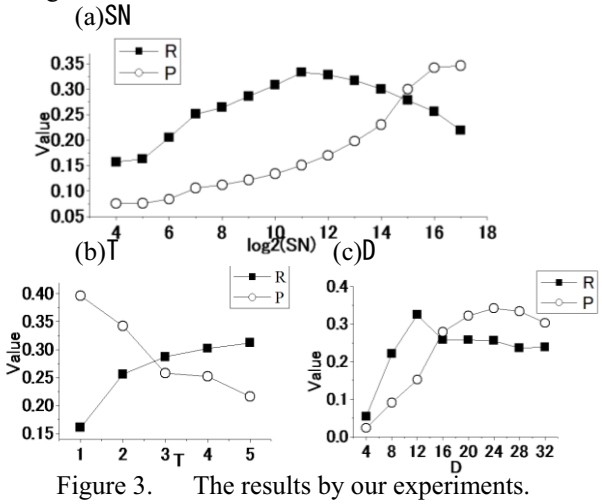

Figure 3. The results by our experiments.

Secondly by using two benchmark sets, we evaluate our model. Table 1 shows overview of performance by our models and some existing works. We have tuned our parameters so as to acquire the largest value of F. Our proposal has the comparable performance as some existing methods except Tagprop. Especially note that SML* and our model(RF) exceed SML [1], because the enough number of feature vectors is extracted from an image. Figure 4 shows the example of annotation results with the best performance by our model.

(a) Corel5K

method	R	P	F	N+
MTL[4]	0.04	0.06	0.05	49
Corr-LDA[10]	0.09	0.06	0.07	59
CRM[9]	0.19	0.16	0.17	107
MBRM[8]	0.25	0.24	0.24	122
SML[1]	0.29	0.23	0.26	137
JEC[3]	0.32	0.27	0.29	139
GS[7]	0.33	0.30	0.31	146
TagProp[2]	0.42	0.33	0.37	160
RF(color+gabor)	**0.26**	**0.34**	**0.29**	**108**
RF(cDCT)	**0.28**	**0.29**	**0.29**	**123**
SML*(color+gabor)	**0.33**	**0.36**	**0.34**	**135**
SML*(cDCT)	**0.31**	**0.28**	**0.30**	**132**

(b)IAPRTC-12

method	R	P	F	N+
MBRM[8]	0.23	0.24	0.24	223
JEC[3]	0.29	0.28	0.29	250
GS[7]	0.29	0.32	0.30	252
TagProp[2]	0.35	0.46	0.40	266
RF(color+gabor)	**0.24**	**0.30**	**0.27**	**226**
SML*(color+gabor)	**0.30**	**0.27**	**0.28**	**266**

Table 1. Overview of performance of our models and some existing works. In SML* *m* is set up as 2^8 in (a) and 2^9 in (b). In our model(RF), *SN* is set up as 2^{16} in (a) and 2^{17} in (b). The other parameters are tuned so that F is the maximum value.

Finally we compare our model(RF) with SML* in terms of training time and testing time. We show the results in Table 2, with the parameters adopted in Table 1. Random Forest has large advantage of shorter computational times, in terms of both training time and testing time. Our proposal can annotate a given image very fast, and more to the point, the annotation time is independent

of the corpus size, different from the nearest neighbor models. The above demonstrations show that our model based on Random Forest is effective on an AIA task.

Table 2. Comparison in terms of computational time.

(a)Corel5K

method	train[s]	test[s] (per image)
RF(color+gabor)	$9.6x10^2$	$3.4x10^{-1}$
RF(cDCT)	$8.4x10^2$	$8.2x10^{-1}$
SML*(color+gabor)	$1.8x10^5$	$1.4x10^1$
SML*(cDCT)	$3.6x10^5$	$3.4x10^1$

(b)IAPRTC-12

method	train[s]	test[s] (per image)
RF(color+gabor)	$2.8x10^3$	$6.7x10^{-1}$
SML*(color+gabor)	$1.5x10^6$	$2.2x10^1$

*Estimate with Core2Duo/3GHz

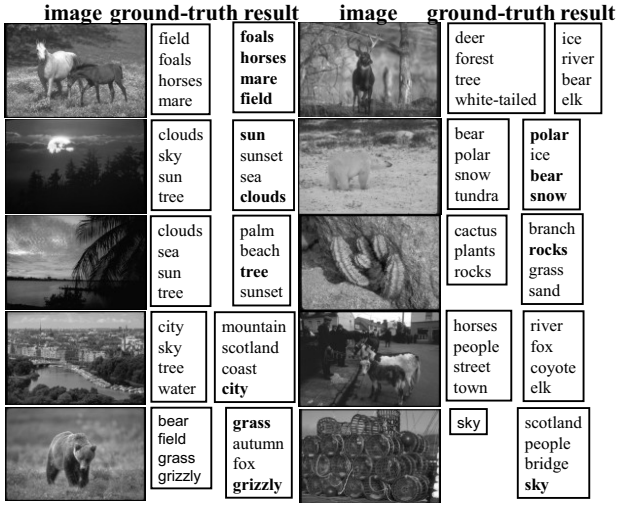

Figure 4. The example of annotation results by our model with Corel5K. A label with bold characters is included in the ground-truth. We show the only 4 most relevance labels.

4. Conclusion

In this paper we have proposed a novel method of applying Random Forest to an automatic image annotation task. Our main contribution is that we have introduced Random Forest into the task by building a model of a probabilistic relation between a local feature and a semantic label. Our model has shown the drastic improvement of training time and annotation time to an existing famous algorithm, SML. Additionally we have demonstrated that our proposal has the comparable performance to the other models. In future we want to challenge to annotate semantic labels on the local region as well as on the whole image, while developing our idea to acquire better performance.

References

[1] G.Carneiro, AB.Chen, PJ.Moreno, and N.Vasconcelos, "Supervised Learning of Semantic Classes for Image Annotation and Retrieval.", *TPAMI*, vol.29, no.3, 2007.

[2] M.Guillaumin, T.Mensink, J.Verbeek, and C.Schmid, "TagProp: Discriminative Metric Learning in Nearest Neighbor Models for Image Auto-Annotation.", *ICCV*, 2009.

[3] A.Makadia, V.Pavlovic, and S.Kumar, "A New Baseline for Image Annotation.", *ECCV*, 2008.

[4] P.Duygulu, K.Barnard, JFG. de Freitas and DA.Forsyth, "Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary.", *ECCV*, 2002.

[5] H.Nakayama, T.Harada, Y.Kuniyoshi, and N.Otsu, "High-Performance Image Annotation and Retrieval for Weakly Labeled Images Using Latent Space Learning.", *PCM*, 2008.

[6] T.Bailloeul, C.Zhu, and Y.Xu, "Automatic Image Tagging as a Random Walk with Priors on the Canonical Correlation Subspace.", *MIR*, 2008.

[7] S.Zhang, J.Huang, Y.Haung, Y.Yu, H.Li, and DN.Metaxas, "Automatic Image Annotation Using Group Sparsity.", *CVPR*, 2010.

[8] A.Feng, R.Manmatha, and V.Lavrenko, "Multiple Bernoulli Relevance Models for Image and Video Annotation.", *CVPR*, 2004.

[9] V.Lavrenko, R.Manmatha, and J.Jeon, "A Model for Learning the Semantics of Pictures.", *NIPS*, 2003.

[10] D.Blei and M.Jordan, "Modeling Annotated Data.", *SIGIR*, 2003.

[11] L.Breiman, "Random Forests.", *Machine Learning*, vol.45, 2001.

[12] F.Schroff, A.Criminisi, and A.Zisserman, "Object Class Segmentation using Random Forests.", *BMVC*, 2008.

[13] J.Shotton, M.Johnson, and R.Cipolla, "Semantic Texton Forests for Image Categorization and Segmentation.", *CVPR*, 2008.

[14] A.Bosch, A.Zisserman, and X.Munoz, "Image Classification using Random Forests and Ferns.", *ICCV*, 2007.

[15] F.Moosmann, E.Nowak, and F.Jurie, "Randomized Clustering Forests for Image Classification.", *TPAMI*, vol.30, no.9, 2008.

[16] R.Lefort, R.Fablet, and JM.Boucher, "Weakly Supervised Classification of Objects in Images using Soft Random Forests.", *ECCV*, 2010.

[17] N.Larios, B.Soran, LG.Shapiro, G.Martínez-Muñoz, J.Lin, and TG.Dietterich, "Haar Random Forest Features and SVM Spatial Matching Kernel for Stonefly Species Identification.", *ICPR*, 2010.

[18] S.Yang, M.Chen, D.Pomerleau, and R.Sukthankar, "Food Recognition using Statistics Local Features.", *CVPR*, 2010.

[19] YW.Chu and TL.Liu, "Co-occurrence Random Forests for Object Localization and Classification.", *ACCV*, 2009.

[20] J.Gall and V.Lempitsky, "Class-Specific Hough Forests for Object Detection.", *CVPR*, 2009.

[21] P.Yin, A.Criminisi, J.Winn, and I.Essa, "Tree-based Classifiers for Bilayer Video Segmentation.", *CVPR*, 2007.

[22] H.Boström, "Estimating Class Probabilities in Random Forests.", *ICMLA*, 2007.

[23] L.Wehenkel, "On Uncertainty Measures Used for Decision Tree Inductions.", *IPMU*, 1996.

[24] C.Chen, A.Law, and L.Breiman, "Using Random Forest to Learn Imbalanced Data.", *Technical Report*, University of California, 2004.

[25] J.Thomas, PE.Jouve, and N.Nicoloyannis, "Optimisation and Evaluation of Random Forests for Imbalanced Datasets.", *LNAI* 4203, 2006.

[26] M.Grubinger, DC.Paul, H.Müller, and T.Deselaers, "The IAPR Benchmark: A New Evaluation Resource for Visual Information Systems.", *Intl. Conf. on Language Resources and Evaluation*, Genoa, Italy, 2006.