Boosting Image Classification with LDA-based Feature Combination for Digital Photograph Management

Xuezheng Liu[1,†], Lei Zhang[2*], Mingjing Li[2], Hongjiang Zhang[2], Dingxing Wang[1]

[1] Department of Computer Science and Technology, Tsinghua University, Beijing 100080, China

[2]Microsoft Research Asia, Beijing 100080, China

*Corresponding author. Email:leizhang@microsoft.com

†This work was performed while the author was visiting Microsoft Research Asia.

Abstract

Image classification is of great importance for digital photograph management. In this paper we propose a general statistical learning method based on boosting algorithm to perform image classification for photograph annotation and management. The proposed method employs both features extracted from image content (i.e., color moment and edge direction histogram) and features from the EXIF metadata recorded by digital cameras. To fully utilize potential feature correlations and improve the classification accuracy, feature combination is needed. We incorporate linear discriminant analysis (LDA) algorithm to implement linear combinations between selected features and generate new combined features. The combined features are used along with the original features in boosting algorithm for improving classification performance. To make the proposed learning algorithm more efficient, we present two heuristics for selective feature combinations, which can significantly reduce training computation without losing performance. The proposed image classification method has several advantages: small model size, computational efficiency and improved classification performance based on LDA feature combination.

Keywords

Image classification, Boosting algorithm, LDA-based feature combination, Training acceleration

1. Introduction

With the rapid development of digital cameras and scanners, digital photographs are becoming a commodity. As a result, digital photographs are being accumulated rapidly and automated tools for organizing these photographs become more and more desirable. Unfortunately, though there are many commercial products available, annotating semantic content of photographs required in organizing photographs, is still left to users. Therefore, automated image annotation is of great importance as it can significantly reduce the manual labeling efforts and thus facilitate the photograph management.

In this paper we consider a specific kind of tasks in image annotation, i.e., labeling images with “indoor” vs. “outdoor”, “city” vs. “landscape” or detecting image’s proper orientation, according to image content. The reason that we focus on these specific cases is because they are the most common processes for digital photograph management. Indoor/outdoor and city/landscape are the basic genres of digital photographs, generally representing very different kinds of human activities. Therefore, if we can accurately annotate photographs with these labels, it will be of great benefit to indexing and searching in digital albums. Orientation detection is another important task for digital photographs. Usually, an image uploaded from either digital cameras or scanners may be in wrong orientation (rotated by 90º, 180º or 270º) while users expect each image should be upright when displayed. Though orientation detection does not belong to the image annotation problem, algorithm wise, it shares the similar properties as the annotation problem.

All these special annotation problems can be treated as image classification: two-class classification (indoor/outdoor, city/landscape) or four-class classification (orientation in 0º, 90º, 180º, 270º). However, these classifications remain challenging. Humans classify images and identify the correct orientation based on object recognition and other contextual information, but state-of-the-art of image content analysis is currently far from this type of image understanding [14]. In this paper, we present our works on image classification aiming at the above problems, which utilize features extracted from both image content (i.e., color, texture and spatial structure) and the camera-recorded EXIF (Exchangeable Image File, [10]) metadata associated with images.

Researchers have made considerable works on indoor/outdoor as well as city/landscape classifications. Most of their attempts have been to classify images by mapping low-level features to high-level semantics. Szummer et al [5] proposed an algorithm for indoor/outdoor classification based on the K-NN classifiers and three types of features, i.e., histograms in Ohta color space (color feature), multiresolution autoregressive model parameters (texture feature) and coefficients of a shift-invariant DCT (frequency feature). Vailaya et al [6] formalized the classification problem to the Bayesian framework using vector quantization (VQ), in which the optimized size of codebook is selected based on MDL criterion. They proposed a hierarchical classification scheme to classify images into indoor and outdoor classes at the highest level, and further classify outdoor images into city and landscape classes.

Earlier works in automated image orientation detection regard it as a four-class classification problem.

Vailaya, et al [1] utilized Bayesian learning framework similar to image classification [6] to classify the image orientations. In their experiment, local regional features are used to obtain rotation sensitive features for orientation detection. They examined a variety of features including color moments [3], color histograms, edge direction histograms and MRSAR texture features, and found that color moments features are much more effective than the others. Based on the same features (i.e., color moments), Hwang et al [2] employed the Hierarchical Discriminant Regression and obtained a slightly lower error rate than LVQ based method in [1]. Wang et al [3] pointed out that the color itself would not be discriminative enough for general image orientation detection. By adopting both the luminance (structural) and chrominance (color) low-level content features, they used Support Vector Machine (SVM) to learn from extracted features and achieved a better result than the LVQ method.

Despite the improved accuracy in comparison with LVQ method, the main drawback of the work proposed by Wang [3] is that the SVM models are too large to be used in a practical system with limited memory space. As a result, the speed of the classification is also slow when using SVM models with many support vectors. To address this drawback, we exploit boosting algorithm instead of SVM and VQ method, from which we can obtain an extremely small training model with very fast speed in classification, simultaneously achieving comparable classification accuracy.

In addition, since the extended boosting algorithm in [9] is capable of dealing with missing features, we can adapt it to exploit EXIF metadata associated with images, which are nowadays commonly available in most digital cameras. EXIF metadata [10] contain important parameters of shooting environment and camera settings, e.g. shutter speed, aperture value, white balance, subject distance and flash utilization. This information is valuable for classification, especially for indoor vs. outdoor because camera settings are usually quite different for indoor or outdoor photographs. Unfortunately, EXIF metadata are usually not completely recorded due to different implementations and incomplete support of early digital cameras. Therefore, if we use EXIF metadata for classification, some of the features will be occasionally missing. For example, the "subject distance" in EXIF metadata is a good feature for the indoor vs. outdoor classification, but it is only supported by some of the digital cameras. Since boosting algorithm can perfectly cope with missing features, we can benefit from boosting algorithm and utilize EXIF features for classification. With the help of EXIF features, we are able to further improve the indoor/outdoor classification accuracy about 3%~5% in comparison with only image features.

Boosting algorithm with weak classifiers, each of which only depends on only a single feature, can be regarded as a feature selection process [3]. However, it does not fully utilize feature correlations. To exploit correlations between features and improve the classification accuracy, we further combine the features linearly to generate new features. Instead of globally linear analysis on all dimensions of feature vectors, we only exploit localized correlations within a small number of features (we use only two-feature-combination in the learning algorithm) and adapt LDA [12] to generate new "combined features". These combined features as well as the original single features are jointly used as weak classifiers in the boosting algorithm. In comparison with conventional LDA techniques, our localized-LDA-with-boosting method has much lower computational requirements, while can exploit feature correlations and improve accuracy of classification. (See Section 3 for the detailed algorithm with discussions, and Section 4 for experimental results).

This paper is organized as follows. In section 2, we will provide our basic algorithm in detail. In Section 3, we will describe feature combination based on localized LDA and some accelerating techniques for training model in our implementation. In Section 4, we will present the experimental environment and results of the proposed algorithm. Thereafter, we will give concluding remarks in Section 5.

2. Boosting image classification scheme

In this section we present our basic algorithm used for indoor/outdoor, city/landscape and four-orientation classification, i.e., the boosting algorithm on image features and EXIF features, without feature combination. Due to similarity of the three classification problems and in order to reduce computation in feature extraction, we use same image features in all our classifications.

2.1 Image feature extraction

For indoor/outdoor classification, we need features that effectively represent the image content, especially the difference between indoor and outdoor photographs. As we know, image objects in indoor and outdoor photographs are quite different, e.g. indoor photographs often contain furniture, food, walls and curtains, while outdoor photographs are mainly sceneries with sky, grass, mountains and appearance of buildings. In order to represent the difference of objects, we utilize local image features that are capable of describing image details. Therefore, we first divide image into $N \times N$ blocks and then extract the image features from these local regions.

Considering the requirement of processing speed and our classification problem, only low-level features are used, including both color features and texture features. It is shown in [3, 13] that color moments (CM) in the LUV color space are simple yet effective in color image analysis, therefore we adopt CM as features to represent image's color information. To represent textures and other spatial structures, we adopt edge direction histogram (EDH) features as in [3]. Using both CM and EDH features, we achieve a simple yet effective image representation containing both color and texture information for indoor/outdoor classification.

Comparing the overall classification performance according to different block numbers N, we finally chose $N = 5$ for CM features which has almost the same accuracy as $N = 8$ and $N = 10$, but has much fewer dimensions. For EDH features, we compared the performance according to the quantization number of edge directions and found that 12 directions are good enough to yield a good result. Consequently, for each image, we extract the following CM features and the EDH features: The CM vector size is: 5×5 blocks × 6 = 150, where six features (3 mean and 3 variance values of L, U, V components) are extracted from each block. The EDH vector size is: 5×5 blocks × (12 + 1) = 325, where (12+1) features (12 directions for edge pixels and 1 total number of non-edge pixels) are extracted from each block and the normalization method for (12+1) features is the same as in [3]. Totally, we have 475 dimensional image features that represent local information of image.

2.2 Boosting learning algorithm for classification

First we briefly introduce the learning algorithm used in our classification, i.e., the *AdaBoost.MH* with abstaining [9]. Recent researches on boosting algorithms [9, 15, 16] show that using linear combination on a number of weak classifiers, one can finally get a considerably strong classifier with remarkable performance. The basic boosting algorithm (AdaBoost, [16]) uses +1 and -1 as output of weak classifiers, while an extended algorithm [9] makes weak classifiers output real values as the confidence-rated predictions. Furthermore, it is capable of dealing with samples with missing values in some dimensions.

Consider we have N dimension features numbered as $1 \ldots N$. Each sample $\boldsymbol{x}$ is denoted by an N-dimensional vector in feature space, i.e., $\boldsymbol{x}=(x_1, x_2, \ldots, x_N)^{\mathrm{T}}$ where x_j is the j-th feature of sample $\boldsymbol{x}$. A weak classifier $h_j(\mathbf{x}, \theta_j)$ depends on only the j-th feature and threshold θ_j, defined as:

$$h_j(\boldsymbol{x}, \theta_j) = \begin{cases} c_l & \text{if } x_j < \theta_j \\ c_g & \text{Otherwise} \end{cases} \tag{1}$$

where j is the feature index, x_j is a non-missing feature of $\mathbf{x}$, θ_j is the threshold which best separates positive and negative samples in dimension j, and c_l and c_g are the confidence-rated output in case of x_j less than or greater than θ_j, respectively. The output of $h_j(\mathbf{x}, \theta_j)$ gives a predication of $\boldsymbol{x}$'s class label (positive or negative), and when x_j is a missing feature, $h_j(\mathbf{x}, \theta_j)$ outputs 0 as an unbiased predication.

In each stage of boosting algorithm, each sample is associated with a weight indicating its importance for training. Thus, for j-th feature, threshold θ_j partitions the set of all training samples (denoted by $\boldsymbol{S}$) into the following five subsets:

$$\begin{aligned} \mathbf{PL} &= \{\boldsymbol{x} \in \boldsymbol{S} \mid \boldsymbol{x} \text{ is positive, } x_j < \theta_j\} \\ \mathbf{NL} &= \{\boldsymbol{x} \in \boldsymbol{S} \mid \boldsymbol{x} \text{ is negative, } x_j < \theta_j\} \\ \mathbf{PG} &= \{\boldsymbol{x} \in \boldsymbol{S} \mid \boldsymbol{x} \text{ is positive, } x_j \geq \theta_j\} \\ \mathbf{NG} &= \{\boldsymbol{x} \in \boldsymbol{S} \mid \boldsymbol{x} \text{ is negative, } x_j \geq \theta_j\} \\ \mathbf{Miss} &= \{\boldsymbol{x} \in \boldsymbol{S} \mid x_j \text{ is a missing feature}\} \end{aligned} \tag{2}$$

where positive and negative are the labels of samples (e.g. we define indoor images as positive while outdoor as negative in indoor/outdoor classification). Thus, the confidence-rated predication of $h_j(\mathbf{x}, \theta_j)$ (i.e., c_l and c_g) are calculated as:

$$c_l = \frac{1}{2}\ln(\frac{\|\mathbf{PL}\|}{\|\mathbf{NL}\|}), \quad c_g = \frac{1}{2}\ln(\frac{\|\mathbf{PG}\|}{\|\mathbf{NG}\|}) \tag{3}$$

where || · || denotes the sum of weights associated to elements in the set. Therefore, weak classifiers with larger confidence outputs (i.e., larger $|c_l|$ and $|c_g|$) and better applicability (i.e., smaller ||**Miss**||) are more valuable in classification and should be picked up in early stages. Therefore, in every stage t, boosting algorithm selects the best threshold θ for each feature j, and then picks up the best classifier (denoted by feature j_t and threshold θ_t) from all weak classifiers in all feature dimensions, which has the smallest normalization factor Z calculated by

$$Z = \|\mathbf{Miss}\| + 2(\sqrt{\|\mathbf{PL}\| \cdot \|\mathbf{NL}\|} + \sqrt{\|\mathbf{PG}\| \cdot \|\mathbf{NG}\|}) \tag{4}$$

And the final strong classifier is:

$$H(\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t h_{j_t}(\boldsymbol{x}, \theta_t) \tag{5}$$

where α_t (set as 1 here) is the weight for the weak classifier h_{j_t}. The input sample $\boldsymbol{x}$ can be classified based on the sign of $H(\boldsymbol{x})$. For more details, please refer to [9].

We can see that boosting algorithm has the extremely small model size and fast classification speed in comparison with SVM. For each chosen weak classifier in boosting model, we only need to store its feature index j, threshold θ and two predications c_l and c_g, which need no more than 16 bytes. Furthermore, boosting-based classification is very efficient with only comparison and addition operations. All these properties make it practical to perform boosting algorithm on thousands of features. In contrast, when feature set is large, SVM frequently suffers from both large model size due to the large number of support vectors and high computation cost in classification. In addition, boosting has the capability of dealing with missing features. This property is crucial for using EXIF metadata in indoor/outdoor classification, since some of EXIF features are occasionally missing due to incomplete support of digital cameras. Taking into account all these considerations, we adopt boosting algorithm as the learning algorithm for image classification problem.

The boosting algorithm can be directly used in indoor/outdoor and city/landscape classification. For orientation detection, for simplicity of implementation, we transform the four-class classification problem into the two-class classification problem based on image's proper orientation. If an image is in its correct orientation (i.e., the up right), we categorize it to the positive class, otherwise to the negative class. Thus, we can train a model to detect whether an image is in the correct orientation with boosting algorithm. In order to detect orientation of an image, we first rotate the image to 0º, 90º, 180º and 270º to generate four images, and use the model to classify the four images one by one. After that, the image with the largest classification output is considered to be in correct orientation. And then, from the rotation we know the correct orientation of the original image. Obviously, this simplification is equivalent to training four models, one for each orientation.

2.3 EXIF features for indoor/outdoor classification

As aforementioned, to improve the indoor/outdoor classification accuracy, we also utilize EXIF metadata associated with images, which are nowadays supported by most digital cameras. An image file created by a digital camera usually contains EXIF metadata in the file header that includes the most important parameters of camera settings when the photograph was taken. According to our knowledge, there is no previous work explicitly utilizing this information for the classification problem we are addressing. Considering the different indoor and outdoor environments for digital cameras, EXIF metadata are supposed to be effective features for indoor/outdoor classification. Since EXIF contains only global information, it has no benefits for orientation detection and little benefits for city/landscape classification. Therefore, we use EXIF only in indoor/outdoor classification. EXIF metadata can be easily extracted from image file [10]. We use the following EXIF features generated from EXIF metadata, as shown in Table 1.

We discard items in EXIF metadata which have clearly no correlations with the classification (e.g. width, height and compression of image). In Table 1, all features except the last one are directly extracted from EXIF metadata. Note that some digital cameras do not support complete EXIF format, therefore the EXIF features may be missing in some dimension.

Table 1: EXIF features used in indoor/outdoor classification

EXIF features	**Meaning or values**
Aperture value	the F-number of lens aperture
CCD width	metric of image resolution
Distance	the subject distance
Date time	Date and time when the photograph is taken.
Exposure bias	value for adjusting exposure value
Exposure program	auto, aperture priority or shutter priority
Exposure time	shutter speed, recorded by the time for exposure
File data time	the time when the image file is created
Flash used	whether camera flash is used or not
Focal length	the length of focal
ISO equivalent	the equivalent film's ISO value
Metering model	center weighted, spot or matrix
White balance	sunny, fluorescent, incandescent or cloudy
Exposure Value (*EV*)	$\log_2(Aperture^2 \cdot \frac{1}{Exposure\ Time} \cdot \frac{ISO}{100})$

The last feature in Table 1 is calculated from other EXIF features as follows. From information of aperture number and shutter speed of camera, we can obtain the exposure value (EV), a professional term in photography, which refers to the amount of light for a given exposure. In digital cameras, EV is used to calculate the correct combination of aperture and shutter speed in order that the photograph is correctly exposed. If the photograph is correctly exposed, EV can be calculated from the combination of sensitive of the CCD/CMOS (i.e., ISO equivalence), aperture and shutter speed as in Table 1. In general, EV value is larger in bright environment than in dark environment, which is a critical feature to discriminate between indoor and outdoor photographs, because indoor photographs are usually taken with low EV values while outdoor photographs are usually taken with high EV values.

However, if the photograph is not appropriately exposed, only EV may not be able to describe the absolute environment lighting condition. For example, if the photograph is over-exposed (i.e., it looks too bright), we will obtain a lower EV than actual measure of environment light, because the photograph should have been taken with a lower exposure time (or larger aperture number) to consist with environment light. To deal with such case of inappropriate exposure, we further take the average brightness of photograph into consideration and use it to complement EV. Therefore, in addition to the 14 EXIF features in Table 1, we also employ global brightness features. We calculate the CDF (cumulative distribution function) of histogram of image luminance in 256 grey levels, and use the grey levels corresponding to 5, 15, 25,…,95 percentile of CDF, as well as the mean luminance value, as our 11 brightness features.

Combining 14 features from EXIF metadata and 11 brightness features, we have 25 global features which have only relationship with environment and global statistic, regardless of local structures. In this paper we also call these 25 features as "EXIF features", since brightness features only provide minor supplementary helps.[2]

From the boosting training results based on EXIF features (see details in Section 4.3), we can see some simple and reasonable rules which are consistent with intuitions for classifying images. For example, the hypothesis (i.e., selected weak classifier) in the first boosting stage is EV feature, which indicates that the most discriminative feature between indoor and outdoor photographs is the environment light. In the first 10 hypotheses of training model, we find that the EV feature, White Balance, Exposure bias and mean value of brightness are selected. This means that from the exposure parameters recorded by cameras, we can obtain valuable information for indoor/outdoor classification. On the other hand, EXIF features with no correlation to exposure may also helpful, e.g. the subject distance of photograph can provide some preference of indoor or outdoor class, because outdoor images usually have longer subject distances than indoor images.

2.4 Boosting on image features and EXIF features for indoor/outdoor classification

Now we present how we use EXIF features as well as image features in indoor/outdoor classification (for city/landscape and orientation only image features are used). We have two feature sets, i.e., 475 dimensional image features and 25 dimensional EXIF features. Rather than merging them into (475+25) dimensional features for indoor/outdoor classification, we utilize the two sets separately. In fact, we trained three models for

[2] Only 14 EXIF features without brightness features are much better in classification than only brightness features.

indoor/outdoor classification on the two feature sets, i.e., the “image-only” model on only image features, the “EXIF-only” model on only EXIF features, and “EXIF-with-rank” model on EXIF features with one additional “rank feature”, which is generated by the boosting classifier $H(x)$ in (5) based on image features and image-only model. These models are listed in Table 2.

Table 2: Three models based on two feature sets for indoor/outdoor classification

Model name	Feature utilization
Image-only	475 image features
Exif-only	25 EXIF features
Exif-with-rank	25 EXIF features + 1 rank feature, where rank feature equals to the boosting’s final output $H(x)$ performed on the 475 image features, based on *image-only* model.

The designing of above models is derived from both research and practical considerations. Intuitively, the image features properly describe images’ local structures, while EXIF features are some global features of images. Thus they might have essential complementarities in indoor/outdoor classification. Therefore, we first use the two feature sets separately by training and testing image-only and EXIF-only model, so as to investigate their respective capabilities. Then, we regard 475 image features as one combined strong feature (rank feature), and add it to EXIF features. This operation can examine the complementarities of the two feature sets. Additionally, it also provides instance in using strong feature (the rank feature) as boosting algorithm’s weak classifier, which may have interesting behaviors for research.

For practical considerations, we also need the above three models. EXIF features are quite different from image features: they are very easily extracted, but sometimes totally unavailable. Thus, we need image-only model to classify images without EXIF metadata, e.g. images obtained from a scanner. The EXIF-only model is suitable for cases that all images contain EXIF metadata, e.g. photographs imported from digital cameras. Since EXIF-only model has much faster speed in classification, it is valuable for fast photograph processing. Finally, EXIF-with-rank model is capable of dealing with either EXIF or non-EXIF images, and performs better than the other two models (see Section 4), thus we can use it for indoor/outdoor classification when there is no severe requirement of process time.

3. LDA-based feature combination

Each of the weak classifiers in above algorithm depends on only a single feature, thus the boosting process, which selects a new weak classifier in each stage, can be viewed as a feature selection process [4]. For instance, during the training of image-only model, in every boosting loop a best feature with currently minimum error is selected from 475 dimensional image features and added to the final classifier $H(x)$. Therefore, the $H(x)$ or the training model is essentially a set of selected features from original features.

From this analysis we can see that boosting algorithm performed on single-feature-dependent weak classifiers as in Section 2 does not fully exploit feature correlations. There are definitely significant correlations between single features (e.g. CM features extracted in neighboring blocks), and utilizing these correlations may further improve the performance. It is worthwhile to use not only single features but *combined features* in boosting algorithm, since correlations between features might lead to a more effective approach to accurate classification.

In this section, we propose the LDA-based approach to combining features, which can exploit correlations between features and also has very low computational requirements.

3.1 Linear feature combination and LDA

As aforementioned, single-feature-dependent weak classifiers cannot exploit feature correlations. In order to take advantage of correlations between features, the weak classifiers used in boosting algorithm should be not only single features but also combined features, i.e., classifiers that depend on more than one feature. Based on principles of simplicity in machine learning, we use linear feature combination on single features, where some features are combined together based on linear operations to form new features. All these combined features, as well as original single features, are selected by boosting algorithm in each stage so as to choose best features and form the final classifier.

Consider we have totally N original features defined the same as in Section 2.3. A combined feature f is defined by the linear projection with vector $\boldsymbol{w}=(w_1,w_2,\ldots w_N)^{\mathrm{T}}$ that maps N-dimensional feature space to a scalar. For a sample $\boldsymbol{x}$, the "f-feature" in $\boldsymbol{x}$ is denoted by x_f and calculated as:

$$x_f = \sum_{i=1}^{N} w_i \cdot x_i = \boldsymbol{w}^{\mathrm{T}} \cdot \boldsymbol{x} \qquad (5)$$

Thus, similar to weak classifier based on single feature, a weak classifier depends on combined feature f is defined as:

$$h_f = h_{\mathbf{w}}(\boldsymbol{x},\theta_{\mathbf{w}}) = \begin{cases} c_l & \text{if } \boldsymbol{x}_f < \theta_{\mathbf{w}} \\ c_g & \text{Otherwise} \end{cases} \qquad (6)$$

where $\theta_{\boldsymbol{w}}$ is the threshold for making predication on combined feature f. If $\boldsymbol{x}$ has some missing features, the x_f is also missing unless we can calculate x_f appropriately without the missing features in $\boldsymbol{x}$, in other words, all factors in $\boldsymbol{w}$ corresponding to the missing features in $\boldsymbol{x}$ are zeros.

Based on (5), (6), a combined feature f and threshold $\theta_{\boldsymbol{w}}$ also partition the whole sample set $\boldsymbol{S}$ into **PL**, **NL**, **PG**, **NG** and **Miss** sets. The confidence-rated predication c_l and c_g, as well as normalization factor Z of weak classifier h_f can also be calculated with (2), (3), (4), by using x_f instead of x_j, as if x_f is a real and independent feature. In this way we can employ boosting algorithm for feature selection among both single features and combined features.

The remaining work is to generate a set of combined features of value. We need to choose the linear combination vector $\boldsymbol{w}$ which defines a discriminative projection in feature space for our classification. Therefore, we adapt LDA (linear discriminant analysis) [12] to find useful $\boldsymbol{w}$, which is an effective algorithm for linear feature combination and classification. LDA can find the best linear projection based on Fisher's discriminants such that the two classes of samples are well separated.

Given a sample set S containing M samples $\{\boldsymbol{x}^1,\boldsymbol{x}^2, \ldots \boldsymbol{x}^M\}$ and their associated weights $(W^1, W^2, \ldots, W^M)$ in a certain boosting stage, we calculate the weighted centers of positive samples and negative samples in feature space, as follows:

$$m_p = \frac{1}{\|\mathbf{Pos}\|} \sum_{\boldsymbol{x}\in\mathbf{Pos}} W(\boldsymbol{x}) \cdot \boldsymbol{x}, \quad m_n = \frac{1}{\|\mathbf{Neg}\|} \sum_{\boldsymbol{x}\in\mathbf{Neg}} W(\boldsymbol{x}) \cdot \boldsymbol{x} \qquad (7)$$

where **Pos** and **Neg** denote positive sample set and negative sample set, respectively. We also use $W(\boldsymbol{x})$ to denote the associated weight of sample $\boldsymbol{x}$ for clarity. Therefore, the between- and within- class scatter variance S_B and S_W are the following formulas:

$$S_B = (m_p - m_n) \cdot (m_p - m_n)^{\mathrm{T}} \tag{8}$$

$$\begin{aligned} S_W = & \sum_{\boldsymbol{x} \in \mathbf{Pos}} W(\boldsymbol{x}) \cdot (\boldsymbol{x} - m_p)(\boldsymbol{x} - m_p)^{\mathrm{T}} \\ & + \sum_{\boldsymbol{x} \in \mathbf{Neg}} W(\boldsymbol{x}) \cdot (\boldsymbol{x} - m_n)(\boldsymbol{x} - m_n)^{\mathrm{T}} \end{aligned} \tag{9}$$

Based on Fisher's discriminants, the projection $\boldsymbol{w}$ that best separates the two classes should satisfy:

$$\boldsymbol{w} = \arg\max_{\boldsymbol{w}} J(\boldsymbol{w}) = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^{\mathrm{T}} S_B \boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}} S_W \boldsymbol{w}} \tag{10}$$

By solving (10) we can obtain the best projection $\boldsymbol{w}$ during each stage, i.e., the most discriminatively combined feature f based on current weights.

3.2 LDA-based feature combination on small feature subset

Instead of calculating the best projection $\boldsymbol{w}$ on all dimensions in feature space, each time we only choose a few features and seek projection $\boldsymbol{w}$ within the subspace spanned by these features. In other words, we use LDA to find best feature combination over a small number of features (typically over two features), rather than the whole feature set. This is derived from the following considerations.

First, performing LDA in a large feature space are computationally expensive. LDA on M training samples in an N-dimensional feature space needs $O(MN^2)$ computation for calculating S_W, which is very time consuming for large M and N. If we plan to carry out LDA operation in each boosting stage, such training program with LDA over the entire feature set is usually impractical, and may also enlarge model size and slow down classification speed. Second, from intuition the feature combination should be carried out on correlated features. However, many features in the feature sets have none of intuitive inter-correlations, especially the features with different types or meanings (e.g. one CM feature and one EDH feature). We consider linear combination on features with different meanings to be of no benefits. Therefore, we need to perform LDA on small and correlated feature subsets, instead of on the entire set. In addition, because the boosting is effective in combining weak classifiers to strong classifiers, we don't have to use very strong features. The effectiveness of boosting is derived from extension and diversity of weak classifiers, not the high accuracy of them. As well acknowledged, boosting-like algorithms are not suitable to combine stable base learners, because they are insensitive to variations of samples' weights and tend to output homogeneous class boundaries in various boosting stages. However, LDA performed on the entire feature space is exactly such kind of stable learner. Consequently, we utilize weakened LDA features generated from multiple correlated feature subsets, instead of one global LDA-based combined feature.

In the implementation, we perform LDA on feature subset that contains only two correlated features. We consider this pairwise feature combination to be sufficient for the image classification, and simultaneously with low computation cost in LDA operations.

3.3 Heuristics for selective feature combination

The subsequent problem is to select a series of feature pairs for LDA-based combination. If we have N original features, there are totally C_N^2 different feature pairs. This is usually a "feature explosion", meaning that there are too many feature subsets for us to calculate feature combination. Therefore, we need some heuristics for selective feature combination, which select and combine the feature pairs with potential benefits in boosting algorithm, so that we can obtain effective combined features within acceptable computation.

One way to selective feature combination is to use rigid restriction. Since we need features to be correlated, we can require them to have exactly the same meaning. We call this heuristic method as "meaning-based" selective method. To explain this, we use image features in Section 2 as an example. If we allow combination of any two CM features or two EDH features, we meet with $C_{325}^2 + C_{150}^2 = 63828$ different feature pairs to be combined with LDA and selected with boosting. To reduce the large number of potential combinations, we restrict features to exact same meanings, i.e., only the equivalent CM features in two blocks, or EDH features in the same edge direction in two blocks can be combined. Thus, we only need to calculate $C_{25}^{2} \cdot 6 + C_{25}^{2} \cdot 13 = 5700$ combined features. Intuitively, most valuable combined features belong to the latter case, thus it is more efficient and also effective to use such restriction for feature combination.

Another way to selective feature combination is to combine only “good” original features, i.e., original features that are rather powerful in separating the two classes in their respective dimensions. We call this heuristic method as “greedy” selective method. Here, instead of using Fisher’s discriminant in one dimension, we utilize the criterion of boosting algorithm in selecting best classifier, i.e., the normalization factor Z in (4). The Z factor is a metric indicating how much the boosting algorithm can take advantage of class separation in the corresponding feature, so it is as well a measure of class separation for single feature. In addition, no matter how we choose potential feature pairs for combination, we always need to calculate all Z factors for each single-feature-dependent weak classifier. Thus we utilize Z factors of each original feature for selecting “good” features and make LDA-based combination on them, i.e., we can pick up the best K features (with the K smallest Z values) and perform LDA-based combination on each pair of these K features. The number K is subject to computational capability used in training classifier. If K equals N, we thus employ no selective feature combination but the entire pairwise feature combination.

In Section 4, we evaluate both meaning-based and greedy method for selective feature combination, in comparison with no heuristics as well as single features.

4. Performance evaluation

In this section we present performance evaluation of proposed algorithm. We begin with explanation of the evaluation methodologies, including image sets and our experiments. Then we give and discuss our experimental results.

4.1 Experiments and image sets

In the experiments, we evaluated the following aspects of our approach: (a) the performance of boosting algorithm on image classification; (b) the effectiveness of EXIF features, and (c) the improvement achieved by using LDA-based feature combination, as well as the heuristics approach to training acceleration. To evaluate boosting image classification, we compare accuracy achieved by boosting algorithm with those by SVM method used in [3]. We use the same image set (both training and testing set) as those used in [3] and extract the same image features, so that we can make a fair comparison between the previously proposed results and ours (see Table.3 for details of the image sets in our experiments). To see the effectiveness of EXIF features, we examine performance of the three models mentioned in Section 2.4, i.e., the image-only, EXIF-only and EXIF-with-rank model in indoor/outdoor classification. The results based on these EXIF models are also compared with CM and EDH features. Finally, we compare LDA-based feature combination with non-combined single features in boosting algorithm for all the three classification tasks (i.e., indoor/outdoor, city/landscape and four orientation classification), in order to evaluate LDA-based feature combination algorithm. When training boosting classifiers, we always run boosting algorithm for 2000 stages (i.e., selecting 2000 weak classifiers), and choose the best number of weak classifiers in testing set.

For image sets, in order to make evaluation more objective, we collect the training images and testing images separately, rather than divide one image set into training and testing data. For training set, we use Corel photo gallery and personal photographs, which contains about 30,000 images. For testing set, we collect another 3000 images from personal photographs taken by different individuals with those in the training set. These images are categorized into 1000 indoor images and 2000 outdoor images for indoor/outdoor classification, and outdoor images are further categorized into 1000 city and 1000 landscape images for city/landscape classification. We perform the three kinds of classification tasks on these images, in order to evaluate LDA-based feature combination algorithm. Besides, as abovementioned, we also use the same data set in [3] for boosting vs. SVM experiments. To evaluate EXIF features, we collect other two separate digital photograph sets, in which all images contain EXIF metadata. Our image sets and experiments are listed in Table.3, and Section 4.3 further explains the experiment for EXIF evaluation. In the following sections, we describe and discuss the experimental results in detail.

Table 3: Image sets used in the experiments

Experiments	Classification	Training set	Testing set
Boosting vs. SVM	*Orientation detection*	5416 (same with [3])	5422 (same with [3])
EXIF features vs. image features	*Indoor/outdoor*	9022/2662	3098/881(correlated) 375/1145(non-correlated)
LDA-based feature combination vs. non-combined single features	*Indoor/outdoor*	4994/5000	1000/2000
	City/landscape	4998/4997	1000/1000
	Orientation detection	10644	3000

4.2 Comparison between Boosting image classification and SVM

In order to evaluate our boosting image classification approach, we also test the SVM-based algorithm proposed in [3], so that we can compare the performances between our algorithm and SVM. We run boosting algorithm and SVM on the same training data as in [3], with the same CM and EDH features. The experimental results are shown in Table.4.

Table 4: Performance comparison between boosting method and SVM in orientation detection

Classifier architecture	Training set	Accuracy with Rejection			
		0%	**10%**	**20%**	**50%**
SVM (single layer)	5416	78.4	82.1	89.9	96.5
SVM (Two layers)	5416 (1^{st} set) 3619 (2^{nd} set)	79.8	83.2	90.9	97.5
Boosting with LDA-based feature combination	5416	81.0	85.2	88.5	96.7
Boosting on single feature	5416	76.7	80.6	84.4	95.4

From Table.4 we can see that boosting algorithm on single feature has comparable performance with that of SVM. If we use LDA-based feature combination, it significantly improves the accuracy and achieves better results than SVM. In particular, in orientation detection, the combined features can greatly improve performance, and LDA-based boosting method achieves 4% more accuracy increase than previous approaches.

In the experiment, the number of selected weak classifiers by boosting method is less than 2,000, and thus the training model is only 24K bytes in size. This is much smaller than SVM's training model, e.g. four SVM models are trained for four orientations in [3], and the total size of four SVM models is about 47M bytes. Additionally, the classification speed of boosting algorithm is extremely faster than that of SVM. In our experiment, the time spent on the pure classification for 5,422 images is 487 seconds in a workstation with Pentium IV CPU for the SVM classifier with single layer, whereas it is only 9 seconds for the boosting classifier. All these results illustrate the advantages of our boosting-based approach.

4.3 Ability of EXIF features for indoor/outdoor classification

Second, we test the ability of EXIF features in comparison with image features (i.e., CM and EDH) for indoor/outdoor classification. For this evaluation, we collect two non-correlated digital photograph sets (all photographs have EXIF metadata): a large set with 12120 indoor photographs and 3543 outdoor photographs, and a small set with 375 indoor photographs and 1145 outdoor photographs. Data in these two sets have very different distributions. The large set is further divided into a training set and a correlated testing set, while the small set is only used for testing (see Table.3). This makes us evaluate classification accuracy on both correlated and non-correlated testing set, so that we can examine the sensibility and the generalization capability of these features to training data distribution.

As mentioned in Section 2.4, We test the accuracy of three models on our training set listed in Table 2, i.e., Image-only model on 475 dimensional single image features, EXIF-only model on 25 dimensional EXIF features and EXIF-with-rank on 25 EXIF features and 1 "rank" features based on the output of classifier on Image-only feature. Fig.1 and Fig.2 show the classification accuracy with different rejection on the correlated and non-correlated testing set.

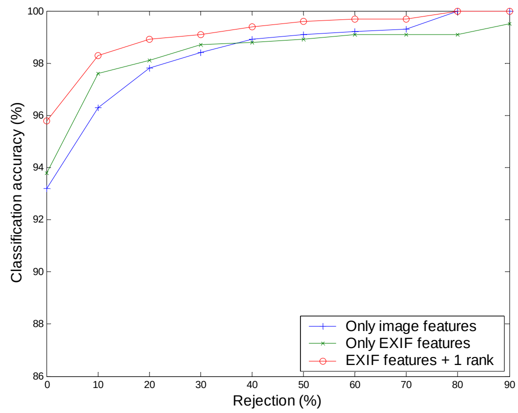

Figure.1. Comparison of EXIF and image features on correlated testing set

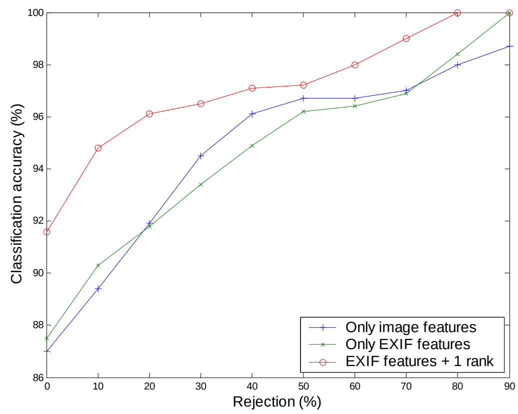

Figure.2. Comparison of EXIF and image features on non-correlated testing set

From Fig.1 and Fig.2 we can see that EXIF-only model achieves about 90% accuracy, which is comparable or slightly better than that achieved by low level image features for indoor/outdoor classification. With only 25 EXIF features (i.e., 13 features from EXIF metadata, 1 exposure degree feature calculated from EXIF metadata and 11 simple features from image histogram), the classifier can achieve very high classification accuracy, performing at least as same better as 150 CM and 325 EDH features. This indicates that information stored in EXIF metadata is very powerful for indoor/outdoor classification, because we know only image histogram features are very weak to classify indoor/outdoor images, and it should be the camera settings and parameters stored in EXIF metadata that facilitate the classification. As extraction of EXIF features is much easier than image feature extraction, in future we can greatly benefit from EXIF features for similar classification tasks.

In addition, EXIF features associated one more “rank” feature derived from image features significantly outperforms only EXIF features or image features, especially in the severe non-correlated testing set. This illustrates complementarities of image features and EXIF features, which consists with our intuition. In Fig.2, the accuracies of all these training models descend, because we use a difficult testing set with very different data distribution (e.g. outdoor samples are about 3 times more than indoor samples in this testing set, while it is exactly the inverse situation in training set). Comparing accuracy in Fig.2 and Fig.1 with no rejection, classifier with only image features or only EXIF features has 6%~7% accuracy degradation, while EXIF-with-rank model degrade the accuracy at only 4%. This means that the combination of image and EXIF features is much more insensitive to training data, while EXIF features and image features have similar sensibility to training data. To further understand the complementarities between image and EXIF features, we examined classification errors of the Image-only and EXIF-only model, as shown in Fig.3 and Fig.4. We present two indoor and two outdoor photographs in both figures. Photographs in Fig.3 are incorrectly classified based on their image features (Image-only model), but can be correctly classified based on EXIF features (EXIF-only model). Fig.4 shows the contrary case, where photographs are correctly classified by their image features but are incorrectly classified by EXIF features. Finally, all photographs in Fig.3 and Fig.4 are correctly classified by EXIF-with-rank model.

3.(a). Upper area is similar to sky and clouds

3.(b). Bright upper area

3.(c). Night place with black sky

3.(d). Night place with black sky

Figure.3. Images that are correctly classified by EXIF features, but not by image features

4.(a). Long subject distance

4.(b). Long subject distance

4.(c). Wrong camera time setting
(due to time zone change in travel)

4.(d).indoor-like dim scene

Figure.4. Images that are correctly classified by image features, but not by EXIF features

In Fig.3, the Image-model classifier seems to be misled by the extracted low level image features. Based on CM and EDH features, the image-model may “find out” that outdoor images usually have blue or white sky in upper area, while indoor images have other colors on the walls, ceiling and furnishings. However, image 3.(a) to 3.(d) conflict with this property. Image 3.(a) has an upper area similar to sky and clouds in color, and image 3.(b) also has a very bright upper area. In contrary, image 3.(c) is a restaurant in the open air, in which the color of sky is in not blue but black, Image 3.(d) shows another night place with very weak environment light. Based on CM and EDH features the image-only model cannot obtain the correct class of images in Fig.3. With EXIF features, the correct class of image 3.(a), 3.(b) and 3.(d) might be easily discovered from their subject distance (short distance in 3.(a), 3.(b) and long distance in 3.(d)). For image 3.(c) (and also 3.(d)) EXIF features contain date and time information, from which the classifier may “know” that the photograph is taken in the evening. Based on training data where outdoor images taken in the evening are always dark in histogram, the classifier can find the correct class of image 3.(c) (and image 3.(d)).

The four images in Fig.4 demonstrate the contrary case where EXIF features are misleading. The boosting classification based on EXIF features is essentially a (probably compromised) decision between many simple rules learned from training data. We find that the EXIF features corresponding to the brightness and lighting conditions of image (e.g., the average brightness feature, the EV feature and white balance feature) are selected in the earlier steps in the boosting training process, and obtain larger weight α_t in Eq (5) in the trained EXIF model. In addition, the subject distance is also valuable for indoor/outdoor classification. As observed in training data that outdoor images usually have better lighting conditions and bright scenes, and contain more distant objects than indoor images, it is reasonable that the classifier learns and utilizes such rules. Image 4.(a) and 4.(b) have very large subject distances in their camera settings (65m in 4.(a) and 24.5m in 4.(b)). Therefore, EXIF-only model might draw the wrong conclusions based on the fact that average outdoor images have larger subject distances. Image 4.(d) looks unlike most outdoor scenery photographs in that it is dark and clouded. Thus, without detailed image features, EXIF-only model tend to classify it into indoor class, because of the indoor-like dim lighting. Image 4.(c) has a wrong date time stamp which may be caused by the time zone change during the travel. The date time records that the shooting time is 21:00. Thus the classifier may realize that it is a photograph taken in the evening, yet with bright scene. Therefore, only if it is an indoor image can the wrongly recorded EXIF features be reasonable. In contrast, with image features we can easily classify image 4.(a) and 4.(b) into indoor class based on the color of walls and furnishings, and classify 4.(c), 4.(d) into outdoor class based on the sky in upper area. The images in Fig.3 and Fig.4 have clearly illustrated the complementarities between image and EXIF features. Using both of the two kinds of features, the EXIF-with-rank model becomes more robust and can correctly classify all above images.

4.4 LDA-based feature combination vs. single feature

Finally, we evaluate LDA-based feature combination. As described in Section 3.4, we use feature combination between two features, selected from all feature pairs based on some heuristics. We have proposed two heuristic methods to reduce potential feature pairs, i.e., the meaning-based method and the greedy method, so that we don't suffer from the extremely large computation. There are two problems to answer in the evaluation. First, we should examine whether LDA-based feature combination is valuable for classification. Although it is reasonable in intuition, we need to test the final accuracy in real classification. Second, because the two heuristics method can significantly speed up feature combination by rejecting most feature pairs in combination, we should investigate whether and how this rejection influence the performance and depress final accuracy of classification by non-heuristic algorithm with pairwise feature combination.

Therefore, in our experiments we trained the following four boosting models on image features: (a). Single model, i.e., the boosting model on only single-feature-dependent weak classifiers, without feature combination. (b). Meaning-based combination model, i.e., using LDA-based feature combination and only combing features with the same “meaning”. (c). Greedy combination model, i.e., using greedy method to select features and only combine the 100 most powerful single features. (d). Pairwise combination model, i.e., using LDA to combine all feature pairs, and no heuristics for feature combination. Comparing the performance of above four models in our classification tasks, we can get to know the effectiveness of our algorithm. The details of these models are listed in Table.5.

Table 5: Models for evaluating LDA-based feature combination

Model name	Training method	Number of weak classifiers in each boosting stage	Time for training one model on a large training set with 10,000 samples
Single	No LDA	475	20 minutes
Meaning-based combination	With LDA	6,100	2~3 days
Greedy combination	With LDA	10,000	4~5 days
Pairwise combination	With LDA	110,000	More than one month

In each row of Table 5 we also list the referenced time for training one model, i.e., finishing 2000 boosting stages on 475 dimensional features and 10,000 training samples in workstation with Pentium IV CPU. We see that Single model has very fast training speed, with no LDA operation and only hundreds of weak classifiers in each boosting stage. Heuristic models have moderate training time, with thousands of LDA operations and weak classifiers in boosting stage. The Pairwise model has extremely large number of combined features, so it must spend months for training models. Since the number of our features (475) and training samples (tens of thousands) are the typical cases in most actual classification tasks, we consider the first three algorithms to be much more practical than the last one.

In our experiments, we finished training the Single model, Meaning-based and Greedy combination model for all training sets in the last experiment listed in Table 3. For the Pairwise combination model, due to its extremely long training time, we can only partially finish it by resampling. In each boosting stage, we ignore the most unimportant samples whose weights totally take up 10% of the entire weights. This resampling strategy can speed up boosting algorithm with only negligible decreasing in classification accuracy. By means of resampling we can finish training Pairwise model for indoor/outdoor and city/landscape classification. For orientation detection, due to the large training set (more than 10,000 images with four orientations, i.e., more than 40,000 training samples), we cannot accomplish training Pairwise model within limited time, and only evaluate the other three models.

Now we present our experimental results on the four models. Fig.5. shows the descending of training error with number of hypothesis increasing for indoor/outdoor classification. In all other training sets the models have very similar behaviors, so we omit figures of training error descending in training sets for city/landscape and orientation classification. From Fig.5 we can compare the convergence speed of different algorithm (the algorithm with smaller training error with a fixed hypothesis number has a more rapid convergence speed). We see boosting algorithm on single features converges slowly, while all LDA-based methods have a much faster descending in training errors. This illustrates the effectiveness of LDA-based feature combination: the combined features are “powerful” than single features, since training error can become much smaller with the same number of hypotheses.

In Fig.5 we also notice that Pairwise combination has a faster convergence speed than Greedy combination method, and Greedy is faster than Meaning-based combination. The reason is very simple and natural: because Pairwise combination model traverses the entire feature pair set and utilizes non heuristics, it might find more powerful combined features than features found by the other two heuristic models. With these features, Pairwise model makes training error descend more quickly. Similarly, the Greedy model also has less training errors than Meaning-based model, because it searches more combined features (10000 combined features) than Meaning-based model (5700 combined features) in each boosting stage, with the greedy strategy.

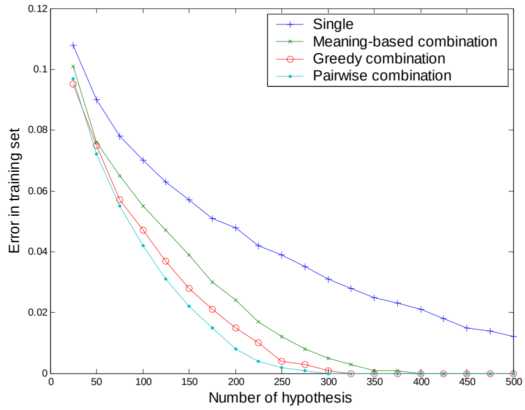

Figure.5. Convergence in training set: the descending of training error

Fig.6, Fig.7 and Fig.8 are the performance comparison among our models. For easy comparison, we also list the value of accuracy in Table 6 (in Table 6 the "hypo num" means the number of hypotheses in the final classifier). From the comparison in all three classifications, we can see that LDA-based feature combination does further improve classification. How much improvement the LDA-based feature combination can achieve depends on the problem, e.g. the improvement by LDA in city/landscape is more than that in indoor/outdoor classification. And in orientation detection, we notice that LDA can greatly improve classification performance. We guess the reason why LDA does not improve indoor/outdoor classification that much is because that, in indoor/outdoor classification only single features has already achieved a good performance, i.e., there remains less "space" for improvement than that in city/landscape or orientation detection.

In [11] we have proposed a very simple feature combination method, i.e., the subtraction combination between features for orientation detection. We compare the results of LDA-based combination, and see that LDA-based method also outperforms subtraction combination method in orientation detection. In addition, subtraction combination method is designed for solving specific problem, derived from the need of rotation sensitive features in orientation detection. In contrast, the LDA-based algorithm is a universal method for general classification problems, and it can always find the best way for linear feature combination, no matter it should be subtraction or addition combination between features.

From the experimental results we also find that Meaning-based combination and greedy combination have very similar performance to enumeration-based Pairwise combination. In the case of orientation detection, greedy combination has even slightly better performance than pairwise combination in some rejection point. We believe this is mainly because we use resampling strategy in pairwise combination (otherwise we cannot finish training the model), while did not resample in greedy and meaning-based model. As experienced in previous works, resampling can cause a slight degradation in model's performance. Therefore, we ascribe abnormal less performance of Pairwise model to resampling strategy, not the pairwise combination algorithm. Our experimental results show that the proposed two heuristics are valuable in practice, as they can significantly reduce the impracticably long training time in pairwise LDA-based feature combination, while does not lose performance.

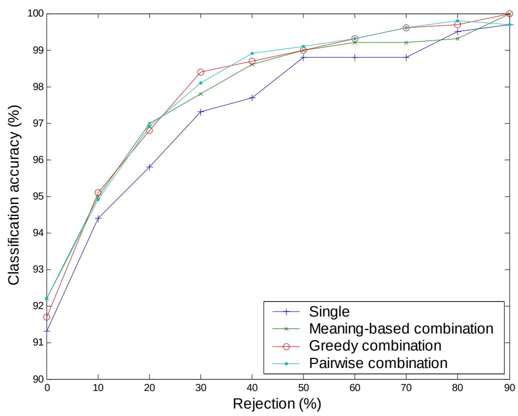

Figure.6. Accuracy comparison in indoor/outdoor classification

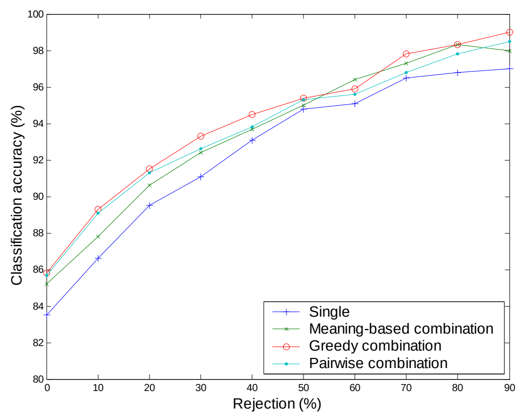

Figure.7. Accuracy comparison in city/landscape classification

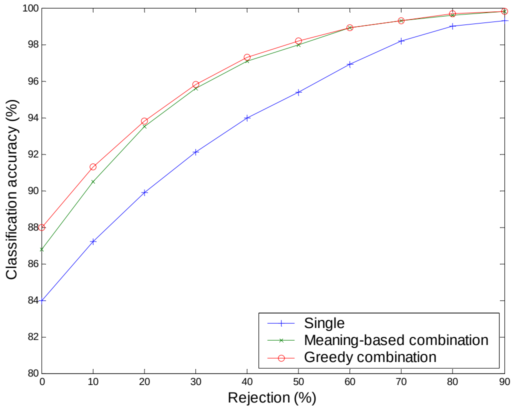

Figure.8. Accuracy comparison in orientation detection

Table 6: Comparison of accuracy in all three classification tasks

Model name	**Indoor/outdoor**					**City/landscape**					**Orientation**				
	Hypo num	Accuracy with Rejection				Hypo num	Accuracy with Rejection				Hypo num	Accuracy with Rejection			
		0%	10%	20%	50%		0%	10%	20%	50%		0%	10%	20%	50%
Single	1500	91.3	94.4	95.8	98.8	700	83.5	86.6	89.5	94.8	2000	84	87.2	89.9	95.4
Meaning-based	1550	92.2	95	97	99	1350	85.2	87.8	90.6	95	1600	86.5	90.5	93.5	98
Greedy	1550	91.7	95.1	96.8	99	1500	85.8	89.3	91.5	95.4	1800	88	91.3	93.8	98.2
Pairwise	1400	92.2	94.9	96.9	99.1	1500	85.7	89.1	91.3	95.3	-	-	-	-	-

5. Conclusions

In this paper we have proposed a general method for image classification and used it in three important tasks: indoor/outdoor classification, city/landscape classification and image orientation detection. We utilized boosting algorithm as the classification architecture, in order to obtain small training model and fast classification speed. We extracted EXIF metadata to generate EXIF features, and our experiments showed that EXIF features are very powerful for indoor/outdoor classification, providing a slightly better accuracy to common high-dimensional image features. In addition, we took advantage of the complementarities between EXIF features and image features, and merged them into a new feature for indoor/outdoor classification, which was more accurate than previous features, and is very stable and insensitive to training set.

In order to further improve classification accuracy, we exploited correlations between different features by using linear feature combination. We employed LDA to find best linear combination for selected features, and used boosting algorithm to choose powerful combined features for the final classifier. Our LDA-based feature combination is a general method which improves performance in all our classification tasks, especially in orientation detection. Finally, we proposed two heuristics for speed up model training, which were very efficient without losing the final performance. Overall, the proposed method in this paper has the following advantages: small model size, fast classification speed, improved classification performance based on feature combination and computational efficiency.

References

[1] A. Vailaya, H.J Zhang and A. K. Jain, "Automatic image orientation detection", in Proc. *IEEE Intl. Conf. on Image Processing*, October 1999. pp. 600-604.

[2] W. Hwang and J. Weng, "An Online Training and Online Testing Algorithm for OCR and Image Orientation Classification Using Hierarchical Discriminant Regression", in Proc. *Fourth IAPR Intl. Workshop on Document Analysis Systems,* December, 2000.

[3] Y. Wang and H.J Zhang, "Content-Based Image Orientation Detection with Support Vector Machines", in Proc. *IEEE Workshop on Content-based Access of Image and Video Libraries*, December 2001. pp.17-23.

[4] P. A. Viola and M. J. Jones, "Robust real-time object detection", *Technical report*, COMPAQ Cambridge Research Laboratory, Cambridge, MA, Feb. 2001.

[5] M. Szummer and R.W. Picard, "Indoor-Outdoor Image Classification", in Proc. *IEEE International Workshop on Content-based Access of Image and Video Databases*, 1998. pp. 42-51.

[6] A. Vailaya, M. Figueiredo, A.K. Jain and H.J. Zhang, "Content-Based Hierarchical Classification of Vacation Images", in Proc. *IEEE Multimedia Computing and Systems*, June 1999. pp 518-523.

[7] R. E. Schapire, "Theoretical views of boosting and applications", in Proc. *Tenth International Conference on Algorithmic Learning Theory*, 1999.

[8] A. Vailaya and A. K. Jain, "Rejection option for VQ-based Bayesian classification", in Proc. *15th Intl. Conf.*

on Pattern Recognition, October 2000. pp. 48-51.

[9] R. E. Schapire and Y. Singer. "Improved boosting algorithms using confidence-rated predictions", In Proc. *11th Annual Conf. on Computational Learning Theory*, 1998. pp 80-91.

[10] Specification of Exchangeable Image File Format: http://www.exif.org/specifications.html

[11] L. Zhang, M. Li, H.J. Zhang, Boosting image orientation detection with indoor vs. outdoor classification, *Proc. IEEE Workshop on Applications of Computer Vision* (*WACV*), 2002

[12] Duda R.O., Hart P.E. and Stork H.G., Pattern Classification, Wiley, New York, NY 2000.

[13] M.Stricker, M. Orengo. Similarity of color images, in Proc. *SPIE Storage and Retrieval for Image and Video Databases*. 1995

[14] Y. Rui, T. S. Huang, S. F. Chang. Image retrieval: current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*. 1999, 10:39-62.

[15] Y. Freund, and R. E. Schapire. Experiments with a new boosting algorithm. In Machine Learning: Proceedings of the Thirteenth International Conference, pp. 148–156. 1996

[16] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139. 1997