Boosting Image Orientation Detection with Indoor vs. Outdoor Classification

Lei Zhang, Mingjing Li, Hong-Jiang Zhang
Microsoft Research Asia
49 Zhichun Road, Beijing 100080, China
{i-lzhang, mjli, hjzhang}@microsoft.com

Abstract

Automatic detection of image orientation is a very important operation in photo image management. In this paper, we propose an automated method based on the boosting algorithm to estimate image orientations. The proposed method has the capability of rejecting images based on the confidence score of the orientation detection. Also, images are classified into indoor and outdoor, and this classification result is used to further refine the orientation detection. To select features more sensitive to the rotation, we combine the features by subtraction operation and select the most useful features by boosting algorithm. The proposed method has several advantages: small model size, fast classification speed, and effective rejection scheme.

1. Introduction

With advances in the multimedia technologies and the advent of the Internet, more and more users are very likely to create digital photo albums. However, an image uploaded from either digital cameras or scanners is often in a wrong orientation while users expect each image should be up right when displayed. Therefore, automated image orientation detection is desirable in these applications.

However, automatic detection of image orientation is a very difficult task. Humans identify the correct orientation of an image through the contextual information or object recognition, which is difficult to achieve with present computer vision technologies. In this paper, we present our work that attempt to automatically detecting image orientations with texture and color features as well as spatial structures of images.

Earlier works in automated image orientation detection have treated it as a four-class classification problem. Vailaya, *et al* [1] proposed a Bayesian framework to classify the image orientations according to the color features extracted from the images. Based on the same feature, Hwang et al [2] employed the Hierarchical Discriminant Regression and obtained a slightly lower error rate than the 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) as the classifiers and achieved a better result than the LVQ method.

Compared to the 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 due to many support vectors in the SVM models. In addition, all the works above have a drawback that the rejection scheme is based only on the confidence score outputted by the classifiers. However, by analyzing the detection result in detail, we find that the accuracy of indoor images is much lower than that of outdoor images. An effective rejection scheme should reject more indoor images than outdoor images at the same level of confidence score.

To address the drawbacks above and improve the accuracy as well, we propose an approach based on the boosting algorithm to estimating the orientation of an image and rejecting images based not only on the confidence score of the orientation detection but also on the indoor and outdoor classification result. Benefiting from the boosting algorithm, we can train a model which is small enough in size and fast enough in classification [4]. Meanwhile the accuracy is comparable to that of SVM based method in [3].

Indoor/outdoor classification is another difficult problem. A number of attempts have been made to classify the 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, one each for color, texture and frequency information. Vailaya et al [6] formalized the classification problem to the Bayesian framework using vector quantization (VQ) and proposed the hierarchical classification to classify the images into indoor/outdoor classes at the highest level.

We find that the orientation detection of indoor images is more difficult than that of outdoor images. To detect the orientation of an indoor image, we usually need to recognize the object in the image first. However, the task of detecting orientations of outdoor images is relatively easier because there are lots of useful information which can be mapped to low-level features, such as sky, grass, building and water. Naturally, we propose a new rejection scheme which rejects more indoor images than outdoor images based on the indoor/outdoor classification result.

This paper is organized as follows. In section 2, we will provide a brief definition of the problem, and then present the proposed algorithm in detail. In Section 3, we will describe the experimental environment and provide the result of the proposed algorithm. Thereafter, we will give concluding remarks in Section 4.

2. Image orientation detection scheme

As in [1, 3], we represent the orientation detection problem as a four-class classification problem, i.e. given an image from a scanner or a digital camera, determine its correct orientation from among the four possible ones: $\omega_1 \Leftrightarrow 0°$, $\omega_2 \Leftrightarrow 90°$, $\omega_3 \Leftrightarrow 180°$, and $\omega_4 \Leftrightarrow 270°$. Note that an image in an arbitrary orientation can easily be rotated into one of the above four orientations by aligning the image boundaries horizontally and vertically.

2.1 Feature extraction

There are many kinds of features to represent the content of an image in content based image retrieval. Rotation invariant is usually an important issue in image retrieval. However, the features for image orientation detection must be sensitive to rotation. Therefore, instead of global features, local regional features are used to capture the spatial content for classification [1, 3]. An image is represented in terms of $N \times N$ blocks and the features are extracted from these local regions.

Similar to the method in [3], we adopt the color moment (CM) feature and the edge direction histogram (EDH) feature. But the dimensionality of the feature used in both [1] (600) and [3] (288+925) is so large that it is difficult to train the classifier and leads to a large size of model in [3]. For the CM feature, we compared the classification performance according to different block numbers from N = 4 to N = 10. Preliminary results show that N = 5 is good enough and has almost the same accuracy as N = 8 and N =10. For EDH feature, we compared the performance according to the quantization number of edge directions and found that 12 directions also yielded a good result compared to 36 directions in [3].

Consequently, for each image, we extract the CM feature and the EDH feature. The CM vector size is: 5x5 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: 5x5 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].

Notice that once the CM and EDH feature in 0° degree are extracted, the features in other three orientations can be simply transformed from the feature in 0°. Thus the feature extraction can be apparently sped up.

2.2 Learning classification functions

The classifier we adopted here is based on the AdaBoost [7]. Recent developments have shown that by simply combining weak learners, boosting based method may have a surprisingly effective combined performance. Moreover, if the weak classifier depends only on a single feature, the boosting process, which selects a new weak classifier in each stage, can be viewed as a feature selection process [4]. A weak classifier $h_j(x)$ is thus defined as the following:

$$h_j(x) = \begin{cases} +1 & \text{if } p_j x_j < p_j t_j \\ -1 & \text{otherwise} \end{cases} \qquad (1)$$

where j is the feature index, x_j is the j-th dimensional feature, t_j is the threshold which best separates positive and negative examples in dimension j, p_j is the parity indicating the direction of the inequality sign.

And the final strong classifier is:

$$H(x) = \sum_{t=1}^{T} \alpha_t h_{j_t}(x) \qquad (2)$$

where α_t is the weight for the weak classifier h_{j_t}. The input sample x can be classified based on the sign of $H(x)$.

In order to generate a probability output, we use the following method as in [4]:

$$\Pr_H(y = +1 \mid x) = \frac{e^{H(x)}}{e^{H(x)} + e^{-H(x)}} \qquad (3)$$

In this way, the final model of the classifier will be extremely small compared to that of SVM because for each weak classifier, we only need to store an index of the selected feature, a weight, a threshold, and a parity sign for the classifier, which need only 12 bytes to store in our system. For example, the size of the final model with 2,000 features is only 24K bytes. Moreover, the classification speed is very fast since it only needs the addition and comparison operations, whereas in SVM, it needs the addition, multiplication and exponential (as Gaussian kernel is used) operations.

So far, the AdaBoost algorithm can be carried out on the CM feature and EDH feature. In every boosting loop,

a best feature which leads to the minimum error is selected among the 150+325=475 dimensional features. However, based only on these features, we can not train a model which is better than SVM according to our experimental results.

As AdaBoost can be treated as a process of feature selection, we can design a large set of features by means of feature combination. Motivated by the fact that the positive and negative samples are extracted from images in different orientations, e.g. 0° is for positive samples, 90°, 180° and 270° are for negative samples, we use the subtraction operation to combine two features together and form a new feature. Thus, if an image is rotated, the values of combined features must change. In order to avoid the combination explosion, we only combine two features which have the same meaning. For instance, means in L channel in two blocks can be combined, or same edge direction in two blocks can be combined.

The total number of combined features is $(6+13)\binom{25}{2}+150+325=6{,}175$, where 6 is the number of CM features in a block, 13 is the number of EDH features in a block, 150 and 325 are the number of the original CM feature and EDH feature respectively. Obviously, this set of features is very sensitive to rotation.

In both [1] and [3], image orientation detection is considered as a four-class classification problem. In [1], four class-conditional probability densities are estimated under a VQ framework. In [3], four SVM classifiers are trained based on one-against-all scheme.

However, orientation detection is a special case of four-class classification problem in that for each image we have four features corresponding to four orientations. Instead of classifying an image by four classifiers using the feature in its original 0° degree, we can classify four features corresponding to four orientations with only one classifier and output the orientation that the corresponding feature has the maximum classification result. Therefore, we only need to train one classifier, 0° against 90°, 180° and 270°. The experimental result verifies this analysis.

Interestingly, Wang et al proposed a re-enforced rejection scheme by considering the results for both 0° and 180° rotated images. But this re-enforced rejection scheme is only a slightly better than the regular rejection scheme, because the classification result of 0° and 180° rotated image are always consistent. The experimental result in [3] also verifies our analysis above.

2.3 Rejection scheme

The rejection scheme is always useful in the orientation detection problem in that lots of images are difficult to be detected unless the objects in the images can be recognized. On the other hand, some images are already in the correct orientation. Vailaya et al first proposed a simple rejection scheme to reject those images whose maximum *a posteriori* probabilities are less than a threshold T [1], then proposed a little more complex rejection scheme [8]. Wang et al proposed a rejection scheme by rejecting images for which the classifier has a low confidence [3].

However, all the rejection schemes proposed in [1] and [3] are only based on the result of image orientation detection. When we tested the performance of orientation detection separately on indoor images and outdoor images, we found that the accuracy of indoor images is much lower than that of outdoor images. This observation motivates us to design a rejection scheme to reject more indoor images than outdoor images at the same level of confidence score.

Fortunately, the features we used for orientation detection can be well used for indoor/outdoor classification. Using the same approach as training for orientation detection, we can train a model for indoor/outdoor classification. Moreover, because the classification speed of AdaBoost is very fast and the features need not to be extracted again for indoor/outdoor classification, the using of indoor/outdoor classification will not slow down the detection speed.

Different from the orientation detection, we extend the feature set by combining any two features with addition operation and thus we get a very large feature set of $\binom{475}{2}+150+325=113{,}050$ dimensions. This feature set yields to a model with good performance for indoor/outdoor classification.

Let $\Pr_{IO}(x)$ denote the confidence score that the image x should be classified into indoor images. For simplicity, let f_i denote $\Pr_H(x_i)$. Based on both $\Pr_H(x_i)$ and $\Pr_{IO}(x)$, we can define our rejection scheme as the following:

If the image is an indoor image, i.e. $\Pr_{IO}(x) > t_{IO}$, then the image is classified into class ω_i (i = 1, 2, 3, 4) if the following three conditions are met: (i) $f_i \geq f_j$, $\forall j \neq i$; (ii) $f_i \geq ti_1$; (iii) $f_i - f_j \geq ti_2$, $\forall j \neq i$; If the image is not an indoor image, i.e. $\Pr_{IO}(x) \leq t_{IO}$, then the image is classified into class ω_i (i = 1, 2, 3, 4) if another three conditions are met: (i) $f_i \geq f_j$, $\forall j \neq i$; (ii) $f_i \geq to_1$; (iii) $f_i - f_j \geq to_2$, $\forall j \neq i$; Otherwise, the image is rejected. That is, it will hold its original orientation.

In the above scheme, t_{IO} represents the threshold for indoor/outdoor classification. ti_1 and ti_2 represent the threshold for ambiguity rejection for indoor images. to_1 and to_2 represent the threshold for ambiguity rejection

for indoor images.

In practical, t_{IO} is easy to be determined. ti_1 can be set equal to ti_2 and to_1 can be set equal to to_2. If ti_1 and ti_2 are greater than to_1 and to_2, we can implement the rejection scheme to reject more indoor images.

If we set $ti_1 = to_1$ and $ti_2 = to_2$, the rejection scheme will reduce to the regular rejection scheme as in [3].

2.4 Digital camera mode

The reason that we especially describe the digital camera mode is that the digital camera is becoming a dominant way of image acquisition for most of the consumers. When taking photos, peoples often hold the cameras in the normal way, i.e. they hold the cameras in 0° to take horizontal photos, or rotate the camera by 90° or 270° to take vertical photos. But they seldom rotate the camera by 180°. This phenomenon results in that the images uploaded from digital cameras are seldom in the wrong orientation of 180°.

Hence we can classify the images with the detection result of 180° into the orientation of 0°. This trick can reduce the error rate by half, because almost half of the errors are generated by the images wrongly classified into 180° in our experiment.

3. Experimental results

First, we carried out the experiment to compare the performance of the proposed method and the method in [3]. We use the same training data as in [3]. That is, the image data are from the Corel photo gallery and the number of training examples is 5,416, and the size of test set is 5,422. The regular rejection scheme [3] is used to reject images.

From the result shown in Table 1, we can see that with the same training data, the proposed AdaBoost based method by combining features with subtraction operation performs better than the SVM methods with single layer and two layers. Notice that SVM with two layers also use another training set of 3619 images for the second layer.

We also list the result based on the model trained by AdaBoost on the original single feature. By combining features with subtraction operation, the accuracy with rejection rate 0% increases from 76.7% to 81.0%. It shows that the feature combination is very effective.

In the experiment, the number of selected features by boosting method is 2,000 and the model is only 24K bytes in size, which is much smaller than the model of SVM. If we train four SVM models for four orientations as in [3], the total size of four SVM models is about 47M bytes. In addition, the classification speed of AdaBoost is extremely faster than that of SVM. In the experiment, the time spent on the pure classification for 5,422 images is 487 seconds for the SVM classifier with single layer, whereas it is only 9 seconds for the AdaBoost classifier.

Table 1. **Performance comparison of different methods with regular rejection**

Classifier architecture	Training Set Size	Accuracy with Rejection			
		0%	10%	20%	50%
SVM Single Layer	5416	78.4	82.1	89.9	96.5
SVM Two Layers	5416 (1st set) 3619 (2nd set)	79.8	83.2	90.9	97.5
AdaBoost Sub-combination	5416	81.0	85.2	88.5	96.7
AdaBoost Single feature	5416	76.7	80.6	84.4	95.4

However, the test set is from the same database with the training data and thus is highly correlated to the training data. Moreover, most of the images in the test set are outdoor images taken by professional photographers. In order to test the algorithm objectively, we use another set of test data, which consists of 1,000 indoor images and 2,000 outdoor images. All the images are either from the clipart of Microsoft Office XP or from some personal photos. All the experiments below will use this database as the test set.

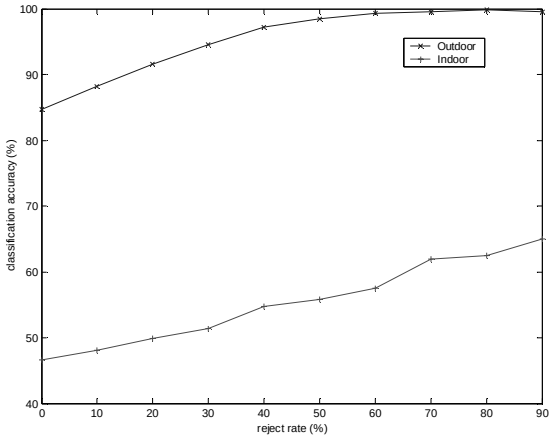

Figure 1. **Comparison of indoor images and outdoor images**

To test the effect of the proposed rejection scheme based on the indoor/outdoor classification, we first tested the accuracy of orientation detection for indoor images and outdoor images separately. From Figure 1, we can see that the accuracy of indoor images is much lower than that of outdoor images. Even if we reject 90% of indoor images, we can only obtain the accuracy of 65%.

Based on this observation, we collected 5,000 indoor images and 5,000 outdoor images as training data, and trained a model of 1,400 features to classify

indoor/outdoor images. The classification accuracy for indoor/outdoor on the test set of 3,000 images is 92.1% with no image rejected, and 95.5% with 10% images rejected.

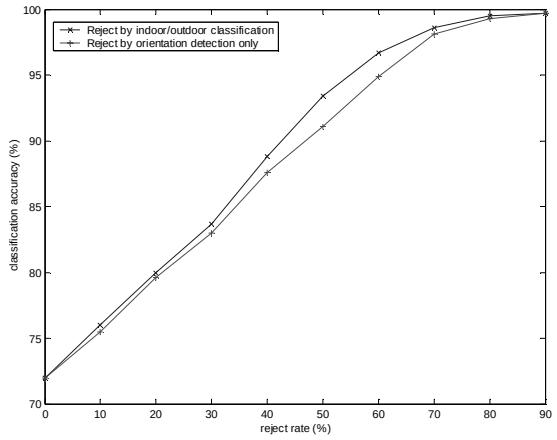

Figure 2. **Comparison of different rejection schemes**

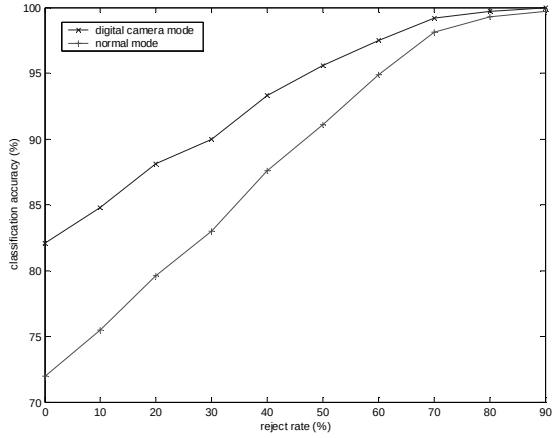

Figure 3. **Comparison of the normal mode and the digital camera mode**

Then we implemented the proposed rejection scheme based on this trained model. We select t_{IO} to achieve an accuracy of 95% for indoor/outdoor classification. We set $ti_1 = ti_2$ and $to_1 = to_2$ for simplicity, and let $ti_1 > to_1$ to reject more indoor images. The result shown in Figure 2 illustrates that the rejection scheme by indoor/outdoor classification performs better than the previous rejection scheme.

Note that the computation complexity does not increase much in this rejection scheme. We need to afford only one additional cost of indoor /outdoor classification, which is extremely fast compared to the feature extraction.

Finally, we have tested the accuracy of orientation detection in digital camera mode, which we described in section 2. We can see clearly from Figure 3 that by simply adopting digital camera mode, the accuracy increases from 72.0% to 82.1%. The result is promising since the digital camera is becoming an increasingly dominant way of image acquisition.

4. Conclusions and future works

We have proposed a new method for image orientation detection. To select orientation sensitive features, we combine the features by subtraction operation and select the most useful features by boosting algorithm. In addition, we present a new rejection scheme based on the indoor/outdoor classification. The proposed method has several advantages: small model size, fast classification speed, and effective rejection scheme. All these factors make it possible to be used in a practical system

However, the general image orientation detection is still a challenging problem. It is especially difficult to detect the orientations of indoor images because we lack the discriminative features for indoor images. The directions of our future work will concentrate on indoor images along the following ways: feature extraction, face detection, and the use of the meta-data in the digital camera.

5. 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.