Hierarchical System for Content based Categorization and Orientation of Consumer Images

Gaurav Sharma[1], Abhinav Dhall[1], Santanu Chaudhury[2], and Rajen Bhatt[1]

[1] Samsung Delhi R&D, D5 Sec. 59, Noida
grvsharma@gmail.com, {abhinav.d, rajen.bhatt}@samsung.com
[2] Multimedia Lab, Dept. EE, Indian Institute of Technology Delhi
santanuc@ee.iitd.ac.in

Abstract. A hierarchical framework to perform automatic categorization and reorientation of consumer images based on their content is presented. Sometimes the consumer rotates the camera while taking the photographs but the user has to later correct the orientation manually. The present system works in such cases; it first categorizes consumer images in a rotation invariant fashion and then detects their correct orientation. It is designed to be fast, using only low level color and edge features. A recently proposed information theoretic feature selection method is used to find most discriminant subset of features and also to reduce the dimension of feature space. Learning methods are used to categorize and detect the correct orientation of consumer images. Results are presented on a collection of about 7000 consumer images, collected by an independent testing team, from the internet and personal image collections.

1 Introduction

In this paper we present a digital content management (DCM) solution which (a) automatically categorizes consumer images into four broad categories and (b) detects their correct orientation, based on their content. We first categorize the images into four categories namely Mountains, Monuments, Water bodies and Portraits. As the rotation of the input image is unknown (among multiples of 90 degrees), we do the categorization in a rotation invariant way. Then within each category we detect the correct orientation of the image by methods tuned to the statistics of that category.

We are interested in fast solutions suitable for implementation in a resource limited target. Hence, we use simple and inexpensive features based on color and edge information. To further speed up we use a recently proposed feature selection method [11], based on information theory concepts, to reduce the dimension of the feature space by mining out a small subset of most discriminant features. Support vector machine (SVM), Gaussian mixture models and variant of the boosting classifiers are used as the learning methods for the various tasks.

Fig. 1. Example images from the database; Mountains, Monuments, Water bodies and Portraits

1.1 Related work

Image categorization is an area of much recent research. However, most of the work uses high time complexity point detection e.g. Scale invariant feature transform (SIFT) [7]. A recent representation of images, bag of words [14], has been shown to be very good for categorization tasks e.g. [14, 6, 4]. However, resource constraints prevent us from using point detection.

Orientation detection is also a well researched field. A Bayesian learning framework was presented in [10] for estimating the orientation of the images. [12, 13] did content based image orientation detection using SVMs with spatial color moments (CM) and edge direction histograms (EDH) features. [15] used AdaBoost algorithm with the CM and EDH features. They trained a indoors versus outdoors classifier using similar AdaBoost algorithm. [1] proposed a scalable boosting approach for image reorientation. The features used were statistics of different sized image blocks from RGB and YUV channels and vertical and horizontal edge images with comparisons as weak classifiers.

2 Proposed framework

2.1 Categorization method

We categorize the images into four classes namely Mountains, Monuments, Water bodies and Portraits. Fig. 1 shows examples of the images from the database. We solve the problem with a learning based framework. First we extract color based rotation invariant features (color correlograms) to represent the image and then use a statistical learning method (support vector machines, SVM) to do the categorization.

Low level features. The color correlogram (CC) [5] feature is used for the categorization task. For any pixel, the CC gives the probability of a pixel at a distance k away to be of certain color, and is defined as $\gamma^{(k)}_{c_i,c_j} = \Pr_{p_1 \in I_{c_i}, p_2 \in I}[p_2 \in I_{c_j} || p_1 - p_2| = k]$. The choice of feature was motivated by two factors: (a) the input images may be rotated by multiples of 90 degrees and (b) images are expected to be color images and color distribution is a good discriminant for the different classes. CC [5] feature captures the spatial correlation of colors. It has been proposed and used for image indexing and retrieval. We show that the feature along with a suitable classifier can be used for image categorization giving good results.

Support Vector Machines. Once the images are represented with CC vectors we use support vector machines (SVM) [9] for categorization of the vectors

into 4 classes. SVM finds a separating hyperplane, in the $\phi(.)$ induced high dimensional space, having the maximum margin [9] and has been found useful in many machine learning applications. We train one-vs-one SVM classifier on the training data, the type of SVM used was C-SVC with an radial basis function (RBF) kernel, given by $K(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \exp(-\gamma||x_i - x_j||^2), \quad \gamma > 0$. The cost parameter C and the kernel parameter γ were optimized using cross validation. We used libsvm [3] for the experiments.

2.2   Image orientation detection

Once the images are categorized into four categories, we proceed to detect the correct orientation of the images. Each category is expected to have (a) different distributions of the low level features for the different orientations and (b) different features which are more discriminant for the task. Keeping these two points in mind we design feature extraction, feature selection and orientation detection modules which are tuned to the particular category.

Low level features. We use low level color and edge features for the current task due to the resource limitations on embedded platform on which the system is expected to run. We extract the color moments (mean and variance) of the normalized R and G planes (which lends some robustness against illumination differences) i.e. $R_{norm} = \frac{R}{R+G+B}, \quad G_{norm} = \frac{G}{R+G+B}$ with the planes divided into $k \times k$ blocks. We use this feature for estimating the orientation of the Mountains and Monuments class. Further, for Water bodies class color is not enough because of similar colored sky/water. We use edge direction histograms as the texture of the water surface is a good cue for the orientation of the image. The features are calculated for $k \times k$ image blocks. The horizontal and vertical edges images are calculated using Sobel operators. The pixels with small edge responses are discarded and edge direction θ at each edge pixel (x, y) is calculated as, $\theta(x, y) = \tan^{-1} \frac{G_y(x,y)}{G_x(x,y)}$. The edge directions obtained are then used to construct b bin histograms for each block of the image, which are concatenated to form the feature for the full image.

Information theory based feature selection. To reduce the dimensionality and to mine out the most discriminant features we use a recently proposed information theory based feature selection method [11]. The method maximizes the mutual information (MI) among the features and the class labels given by $I(X;Y) = \sum_i \int_\chi p_{X,Y}(x, i) \log \frac{p_{X,Y}(x,i)}{p_X(x)p_Y(i)} dx$ where X is the random process generating features x and Y is the random process generating labels i. [11] decompose the mutual information (MI) into two components, $I(X;Y) = M(X;Y) + C(X;Y)$, marginal mutual information (MMI) given by $M(X;Y) = \sum_{k=1}^{b} I(X_k;Y)$ and the conjunctive component of mutual information (CCMI) given by $C(X;Y) = \sum_{k=2}^{b} [I(X_k; X_{1,k-1}, Y) - I(X_k; X_{1,k-1})]$ where $X_{i,j}$ is set of features from index i through j and X_k is the kth feature. The MMI measures the discrimination power of individual feature while CCMI measures discrimination power of interdependence of features. They use theoretical tools and empirical validation to arrive at the important conclusion that,

for common features, only pairwise dependence of features results in practical gains and modeling dependence of features beyond pairs complicates the model and reduces performance. We use the approximate Infomax algorithm, given in [11], to mine out the most discriminant features.

Gaussian mixture model classifier. We use gaussian mixture models (GMM) to capture the distribution of the features (after feature selection) for the 4 rotation classes of the images. GMM pdf given by,

$$f(x; M) = \sum_{i=1}^{k} \pi_i \frac{1}{(2\pi|\Sigma_i|^d)^{\frac{d}{2}}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right) \qquad (1)$$

where $M = (\pi, \mu, \Sigma)$ are the GMM model parameters for the present GMM, is a light weight generative model for capturing the distribution of the feature vectors extracted from images with different rotations. We learn the GMM parameters $\{M_l = (\pi_l, \mu_l, \Sigma_l)|l = 1 \ldots k\}$ using the expectation maximization (EM) algorithm on the reduced feature space after feature selection.

We fit four GMMs, one for each rotation of 0, 90, 180 and 270 degrees and assign orientation to a new image based on the likelihood of the four models of having generated that vector i.e. $l^* = \arg\max_l f_l(x; M_l)$.

Scalable boosting. The Portrait class is more varied because of the different backgrounds. We use a recently proposed scalable boosting algorithm [1] for orientation detection of the Portrait class. First basic images with RGB, HSV normalized versions of RGB, HSV and horizontal and vertical edge maps are generated. Then the features are extracted from those images i.e. the mean and variances of multiple sized blocks and strips (both horizontal and vertical). Finally a modified boosting algorithm [1] is used to learn a strong classifier with the weak classifiers being simple comparison etc. between the feature values as used in [1]. We train the strong classifier for 0 degree vs other rotations. When a new image is presented, we extract only the required parts of the basic images and calculate the features which are required for evaluation of the strong classifier. We calculate the response of all four rotated version of the new image and decide the final orientation of the image by the maximum response value.

3 Experimental results

We tested our system on a database of about 7000 images collected over the internet and from personal image collections, by an independent testing team. To test our system, we split the dataset randomly into 100 images per category for testing and the rest for training. We did 10 such random splits and report the average performance here.

3.1 Categorization results

Table 1 gives the confusion matrix for the categorization task, where the column is the true class. The overall performance for the method was 83.6 %, with the most confusion classes being Monuments and Mountains with 10.5 % confusion.

Table 1. Confusion matrix for categorization task (column is the true class)

	Mountains	Monuments	Waterbodies	Portraits
Mountains	83.75	10.50	9.25	2.75
Monuments	5.25	82.50	3.75	9.75
Waterbodies	9.00	2.75	83.50	3.00
Portraits	2.00	4.25	3.50	84.50

3.2 Orientation detection result

Mountains and Monuments classes. We use color moment features with Infomax feature selection for the Mountains and Monuments classes. Table 2 shows the confusion matrix for the orientation detection task. The mirror image classes show relatively more confusion among themselves e.g. for Mountains 0 degree images are 8.5 % confused with 180 degrees class. The performance for all the rotations are almost same for both the classes showing that the classifier is not biased towards one rotation. The average accuracy for orientation detection achieved for Mountains and Monuments class was 86.1 % and 79.4 % respectively.

Water bodies class. Table 2 shows the confusion matrix for the water bodies class. Again there is high confusion between mirror image rotations and very less confusion between images rotated with complementary angles. The average performance reached is 74.0 % for water bodies class with uniform performance for all rotations.

Portrait class. Portrait class is the most varied class in terms of appearance due to the varied backgrounds against which the pictures have been taken. Table 2 shows the confusion matrix for the orientation detection of the Portrait classes. Again there is much confusion between the classes with opposite rotations. The performance reached for Portrait class is 80.4 %.

Table 2. Confusion matrices for orientation detection

Mount	0 deg.	90 deg.	180 deg.	270 deg.	Monu	0 deg.	90 deg.	180 deg.	270 deg.
0 deg.	86.2	3.0	9.4	2.9	0 deg.	81.6	6.2	11.0	4.5
90 deg.	3.4	85.2	1.8	8.4	90 deg.	5.0	79.2	5.3	10.4
180 deg.	8.5	2.8	86.6	2.2	180 deg.	8.0	4.8	78.9	7.1
270 deg.	1.9	9.0	2.2	86.5	270 deg.	5.4	9.8	4.8	78.0

Wbod	0 deg.	90 deg.	180 deg.	270 deg.	Port	0 deg.	90 deg.	180 deg.	270 deg.
0 deg.	76.8	1.6	24.7	0.8	0 deg.	81.4	1.1	17.9	1.6
90 deg.	0.6	72.8	0.7	25.3	90 deg.	1.9	79.5	1.1	15.8
180 deg.	21.9	1.1	73.5	1.1	180 deg.	15.2	2.1	79.1	1.1
270 deg.	0.7	24.5	1.1	72.8	270 deg.	1.5	17.3	1.9	81.5

3.3 Discussion and future work

When compared to scene categorization method of [8] we achieve a performance of 83.6 % on 4 categories while they achieve 89 % on 4 scene categories. However, we use simple features for reducing time complexity are able to achieve comparable results. Compared to the large scale performance evaluation in [2] we achieve better accuracies c.f. their 79.1 % (except in water bodies) by first categorizing the images into more homogeneous categories and then training the algorithm independently on each category.

To sum up, we have presented a system for categorization and orientation detection of consumer images using low level features and learning methods. The system is meant to be run on a low resource device and so only allows for the computation of simple features. This doesn't allow us to compute higher complexity sift features [7], which could have potentially resulted in better results. As a future task, we would like to try out more complicated features as much allowed by our constraints.

References

1. S. Baluja. Automated image-orientation detection: a scalable boosting approach. *Pattern Analysis and Applications*, 10:247–263, 2007.
2. S. Baluja and H. A. Rowley. Large scale performance measurement of content-based automated image-orientation detection. *ICIP*, 2003.
3. C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
4. L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. *CVPR*, 2005.
5. J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih. Image indexing using color correlograms. *CVPR*, 1997.
6. S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *CVPR*, 2006.
7. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
8. A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV*, 42, 2001.
9. B. Scholkopf and A. J. Smola. *Learning with kernels. Support Vector Machines, Regularization, Optimization, and Beyond.* The MIT Press, 2001.
10. A. Vailaya, H.-J. Zhang, C. Yang, F.-I Liu, and A. K. Jain. Automatic image orientation detection. *IEEE Trans. IP*, 11(7):746–755, 2002.
11. M. Vasconcelos and N. Vasconcelos. Natural image statistics and low-complexity feature selection. *PAMI*, 31(2):228–244, 2009.
12. Y. Wang and H. Zhang. Content-based image orientation detection with support vector machines. *CBAIVL*, 2001.
13. Y.M. Wang and H. Zhang. Detecting image orientation based on low-level visual content. *CVIU*, 93:328–346, 2004.
14. J. Willamowski, D. Arregui, G. Csurka, C. R. Dance, and L. Fan. Categorizing nine visual classes using local appearance descriptors. *IWLAVS*, 2004.
15. L. Zhang, M. Li, and H.-J. Zhang. Boosting image orientation detection with indoor vs. outdoor classification. *WACV*, 2002.