A multi-class image classification system using salient features and support vector machines

Wenbin Shao
University of Wollongong, wenbin@uow.edu.au

Son Lam Phung
University of Wollongong, phung@uow.edu.au

G. Naghdy
University of Wollongong, golshah@uow.edu.au

A multi-class image classification system using salient features and support vector machines

Abstract

This paper addresses the problem of automatic image annotation for semantic retrieval of images. We propose an image classification system that is capable of recognizing several image categories. The system is based on the support vector machine and a set of image features that includes MPEG-7 visual descriptors and a custom feature. The system is evaluated on a large dataset consisting of 14400 images in four categories - landscape, cityscape, vehicle and portrait. We find that the proposed edge direction histogram and the MPEG-7 edge histogram perform better than other features in this application. Experiment results indicate that the pair- wise SVM approach performs better than the one-versus-all SVM approach. The pair-wise method with confidence score voting has better classification rates compared to the pair-wise method with majority voting.

Keywords

image classification, MPEG-7, SVM, image annotation, image retrieval

Disciplines

Physical Sciences and Mathematics

A Multi-class Image Classification System Using Salient Features and Support Vector Machines

Wenbin Shao, Son Lam Phung, and Golshah Naghdy
School of Electrical, Computer and Telecommunications Engineering
University of Wollongong
Northfields Av, Wollongong, NSW 2522, Australia
E-mails: {*ws909, phung, golshah*}@*uow.edu.au*

Abstract

This paper addresses the problem of automatic image annotation for semantic retrieval of images. We propose an image classification system that is capable of recognizing several image categories. The system is based on the support vector machine and a set of image features that includes MPEG-7 visual descriptors and a custom feature. The system is evaluated on a large dataset consisting of 14400 images in four categories - landscape, cityscape, vehicle and portrait. We find that the proposed edge direction histogram and the MPEG-7 edge histogram perform better than other features in this application. Experiment results indicate that the pair-wise SVM approach performs better than the one-versus-all SVM approach. The pair-wise method with confidence score voting has better classification rates compared to the pair-wise method with majority voting.

1. Introduction

The rapid increase in the availability of digital cameras has greatly facilitated the acquisition of images and brought millions of images to this world. A challenging problem, resulted from the creation of huge images databases, is the effective search, navigation and management of large image archives. There are roughly two major techniques for image retrieval: text-based and content-based. Text-based approaches utilise textual keywords or descriptions generated by human annotators [1]. Given the rapid growth in the number of digital images, manual image annotation is time-consuming and annotator dependent.

Content-based image retrieval was developed to address some of the shortcomings of the manual annotation. It is based on the idea of retrieving images using information automatically extracted from pixels. CBIR substantiates demanding multimedia retrieval needs such as sketch search, example search and similarity search [2], [3]. Although noteworthy progress has been made in the research, very few CBIR systems are capable of widespread commercial use [2], [4].

The main challenge in content-based image retrieval is to bridge the semantic gap between low-level features and conceptual contents [2], [3], [5]. For image retrieval, image representation by keywords is more suitable than by low-level features, because people often describe an image with keywords, not low-level features [6]. Our research will focus primarily on extracting automatically the keywords that describe an input image. This research problem is also known as *automatic annotation.*

In this paper, we propose a multi-class image classification system that is based on support vector machines and a set of salient features. Given an input image, our system will automatically generate keywords that describe the image contents. We propose a new feature called gradient direction histogram that is suitable for automatic semantic annotation task. We also analyse several MPEG-7 visual descriptors. The proposed system is applied to annotate landscape, cityscape, vehicle and portrait images. This paper is organised as follows. In Section 2, we review existing techniques for classifying images. In Section 3, we describe the proposed system that combines salient features and multi-class support vector machines. In Section 4, we present and analyse the experiment results. Finally in Section 5, we give the concluding remarks.

2. Background

To date, many works that attempt to bridge the semantic gap have been undertaken. Li and Wang [7] propose a statistical modelling approach to automatic linguistic indexing. In their approach, they employ a statistical model known as the two-dimensional multi-resolution hidden Markov model. This model is applied separately to each category. Therefore, when a new category of images is added to the database, relearning of the existing categories is avoided. Their CBIR system indexes images with linguistic descriptions based on statistical model comparison. Li and Wang use a subset of the COREL database to evaluate their system. There is no overlap in the descriptive words among different categories in the images subset. An advantage of Li and Wang's approach is that it does not require image segmentation.

Ozcanl and Yarman-Vural [8] propose a CBIR system based on region classification. The system is aimed at users who want to query certain objects in the database. In their approach, the input image is first segmented into regions. Each region is represented by a 239-dimension feature vector that consists of all visual descriptors in the MPEG-7 standard set. The feature vectors are then processed by a fuzzy ARTMAP classifier. ARTMAP is a class of neural network architectures based on adaptive resonance theory. In the experiments, the training set

contains a few background regions. However, their system is evaluated on object regions only. The authors use 938 images of ten classes in the experiments and choose a baseline system without using the fuzzy ARTMAP classification. Their system performs better than the baseline system across most of the classes.

Lu et al. [9] propose an approach for automatic annotation based on model space. First, they divide training images into K categories and manually annotate each image with four to five keywords. Then they transform low-level visual features into high-level semantic model-vectors. The annotation is propagated from training examples to the remaining images. In feature extraction, they select six types of low-level features including colours and textures, and reduce the dimension of the feature from 369 to 114. Lu et al. use pair-wise coupling (PWC) method to construct support vector machines. The SVMs are then employed to train the images based on summation of negative probability (SNP). Lu et al. compare the PWC-SNP model-vector with the low-level feature vector, using the minimum distance classifier. They show that their approach performs better. However, their approach requires training to be repeated when a new category is added.

3. PROPOSED APPROACH

We propose a multi-class image annotation system that combines salient visual descriptors and support vector machines. The block diagram of the system is shown in Fig. 1. MPEG-7 and custom features are first extracted from the image. Multi-class support vector machines are then used to classify the visual descriptors into different image categories such as *landscape, cityscape, vehicle,* and *portrait.*

A. Visual descriptors

In this section, we propose a new image feature called gradient direction histogram and describe several MPEG-7 visual descriptors for image classification task.

1) Gradient direction histogram (GDH): The proposed feature is a normalized histogram of gradient directions, computed across all edge pixels in the image. To compute the feature, we first apply an edge operator to calculate the edge magnitude along the horizontal and vertical direction. In this paper, the Prewitt operators [10] are used:

$$\mathbf{h}_x = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad \mathbf{h}_y = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad (1)$$

Next, edge pixels are found by thresholding the edge magnitude. For each edge pixel, the gradient angle is calculated as

$$\theta = \arctan \frac{G_y}{G_x} \quad (2)$$

where G_y and G_x are the edge magnitude along the vertical and horizontal direction, respectively. Gradient angles for all edge pixels are represented as a histogram of 36 bins. To provide invariance against image rotation, the histogram is normalized to 35 bins.

2) MPEG-7 visual descriptors: Multimedia Content Description Interface, generally known as MPEG-7, is a standard for multimedia content description. It is aimed at a wide range of applications involving image, video and audio search. For still images, there are three major categories of visual descriptors defined in the MPEG-7: colour, texture and shape.

- **Colour descriptors.** There are four main descriptors: dominant colour, scalable colour, colour structure and colour layout. All descriptors can All colour descriptors can be extracted from an image or an image region. *Dominant colour* describes the representative colours in an image or an image region. *Scalable colour* is a pixel histogram in the HSV colour space; the histogram is encoded using the Haar transform. *Colour structure* is a histogram of colour distribution and spatial colour structure. *Colour layout* represents images by spatial colour structure in the YCbCr colour space; it is invariant to image resizing.

- **Texture descriptors.** Texture descriptors can be extracted from an image or an image region. There are three common texture descriptors: edge histogram, homogeneous texture and texture browsing. *Edge histogram* describes the local spatial distribution of edges in an image. For this descriptor, an input image is divided into 16 sub-images (4 rows and 4 columns). Then, each sub-image is further divided into non-overlapping square blocks. For each block, the edge pattern is determined from five possible edge patterns: four directional and one non-directional. Next, a five-bin histogram is extracted from every sub-image. Finally, an edge histogram descriptor of 80 bins in total is formed for the entire input image. *Homogeneous texture* describes region texture using the mean energy and the energy deviation whereas *texture browsing* specifies region texture in terms of regularity, coarseness and directionality.

- **Shape descriptors.** For each 2-D region, there are two types of shape descriptors: *region-based shape* represents the shape of a region whereas *contour-based shape* reveals the properties of the object contour.

Further details about the above MPEG-7 visual descriptors can be found in [11]–[14].

B. Constructing classifiers

In this work, we use support vector machines (SVMs) as the basic tool for multi-class classification. SVMs are a machine learning approach with a solid theoretical foundation [15]–[17]. They have been demonstrated to perform well in numerous practical applications.

1) Two-class SVMs: Using SVM for a problem involving two classes, the decision boundary is obtained directly from the training data by finding an optimal separating hyperplane that maximizes the margins between the two classes; a constrained quadratic programming problem can be constructed
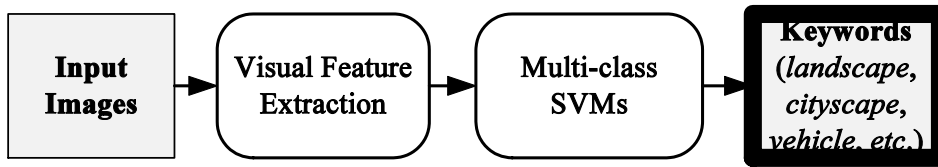
Fig. 1: Proposed image annotation system.

to solve the problem. This learning strategy is based on the Vapnik-Chervonenkis theory and shown to boost the generalization capability of the classifier. By projecting data from the input space to a higher-dimensional feature space via a mapping function, we can apply SVMs to non-linear problems. Compared to other classifiers, the SVM approach works well when the number of training samples is small.

Suppose we have a training set given by $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_L, y_L)\}$, where $\mathbf{x}_i$ is a vector in n-dimensional space and $y_i \in \{1, -1\}$. The input samples are projected to a higher-dimensional feature space using a mapping function $\phi(\mathbf{x})$. In the new space, a linear hyperplane that produces the maximum separation between the two classes can be found by solving the following quadratic programming problem. Minimize

$$Q(\mathbf{w}, b, \epsilon) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + c \cdot \sum_{i=1}^{L} \epsilon_i \qquad (3)$$

subject to constraints

$$\{\mathbf{w}^T\phi(\mathbf{x}_i) + b\} \cdot y_i \geq 1 - \epsilon_i \text{ and } \epsilon_i \geq 0$$

where $\mathbf{w}$ is a vector perpendicular to the optimal separating hyperplane, b is a bias term, ϵ_i is a non-negative slack variable and C is the learning cost. The learning cost represents a compromise between maximizing the margin and minimize the classification error. Note that $H(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle$ is called the kernel function, where $\langle \cdot \rangle$ represents dot product. The kernel method is applied to reduce the computational complexity. In this paper, we use the radial basis function kernel:

$$H(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}, \gamma > 0 \qquad (4)$$

To use the SVM effectively for classification, we need to find the appropriate features, the kernel function, and the training parameters.

2) Multi-class SVMs: SVMs are originally designed for two-class problems. To apply the SVMs for our classification problem involving multiple image categories, it is necessary to extend the two-class SVMs. In this paper, we investigate two strategies for multi-class SVMs: *one-versus-all* and *pair-wise* [17]–[20].

- **one-versus-all SVMs.** In one-versus-all SVMs, a k-class problem is decomposed to k problems, each involving only two classes [17], [18], [21]. This strategy requires k SVMs, each is trained using the entire training set. For the i-th SVM, the samples in the i-th class are labelled as positive, and the rest of the training set are labelled as negative. In other words, the i-th SVM classifier separates class i from all other classes.
- **Pair-wise SVMs.** For a k-class problem, this strategy constructs $k(k-1)/2$ two-class SVM classifiers. Each SVM classifier is trained with samples from two classes. Let C_{ij}, where $i < j$, be the SVM classifier that is trained on data from the i-th class and j-th class. The final classification results are obtained by combing all classifiers with a voting scheme.

Methods for polychotomous classification can be adapted to pair-wise SVMs. For example, Friedman [22] proposed a simple *majority voting* approach in which the final class label is the most common label found by all pair-wise classifiers. When two classes have identical voting scores, the class with a small index can be chosen [18].

We propose another method, called *confidence score voting*, which takes into account the confidence score produced by each pair-wise SWM. For each input sample, the confidence score is related to the distance from the sample to the optimal separating hyperplane. A larger distance means a higher confidence score. The final label for the input sample is the class with the highest accumulated confidence score.

3) Finding SVM training parameters: A difficult task in classifier design and evaluation is finding the training parameters to produce systems that generalize well. In the proposed system, k-fold cross validation is used for this task [23]. We divide the training set into k partitions. In each training round, $(k-1)$ partitions are used to train, and the remaining partition is used to validate the classifier. This step is repeated k times until all partitions have been evaluated. Finally, the average classification rate across k folds is computed. The above process is repeated for different sets of training parameters and the parameters producing the highest classification rate are selected.

Once the training parameters are estimated, they are used to train the final SVM classifier on the entire training set; this classifier is evaluated on the test set.

4. RESULTS AND ANALYSIS

This section describes an application of the proposed multi-class image annotation system. Firstly, we describe the data collection and experimental procedure. Secondly, we present
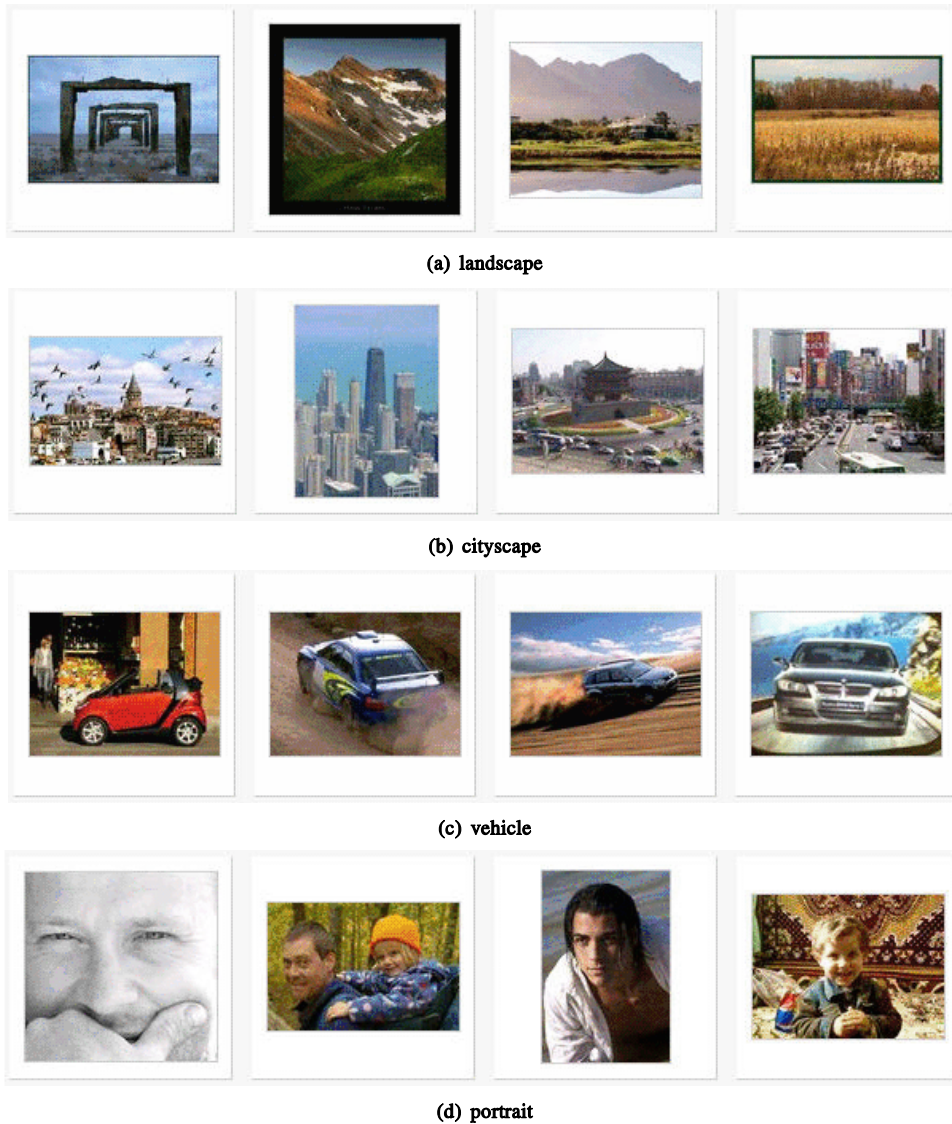
433

(a) landscape

(b) cityscape

(c) vehicle

(d) portrait

Fig. 2: Example images in the dataset of 14400 images.

the results of selecting salient visual features. Thirdly, we compare the image classification performance of three methods for multi-class SVMs. Finally, we compare the propose system with an existing image classification system.

A. Data preparation and experimental steps

We use the same database presented in [24] except that a few ambiguous images are replaced. This database consists of over 14400 images in four categories - landscape, cityscape, portrait and vehicle. These images vary widely in size, quality, and contents; some images even have blurred or monochrome appearance. Examples of the images are shown in Fig. 2.

In this paper, we use 8400 images for training and 6000 images for testing; the training set and test set are separated. For each image category, 2100 images are used for training and 1500 images are used for testing.

In the experiments, the MPEG-7 reference software called *eXperimentation Model* (XM) [25] is used to extract most MPEG-7 visual descriptors. MATLAB is used to extract the dominant colour descriptor. An software library called *LIBSVM* [23] is chosen to train the SVMs. The SVMs use the *radial basis function* kernel. We experiment with different SVM parameters: training cost c and kernel radius γ. The range of c values is from 2^{-5} to 2^{15}, and γ values from 2^{-15} to 8. Based on our earlier study [24], the MPEG-7 descriptors such as the texture browsing descriptor, the region-based shape descriptor and the contour-based shape descriptor are excluded from our analysis because of their poor performance.

B. Selection of salient features

There are four image classes considered in this study. For each pair of classes, we train SVM classifiers that use one of the following seven visual features:

- the proposed gradient direction histogram (35 elements),
- the MPEG-7 edge histogram (80 elements),
- the MPEG-7 homogeneous texture (62 elements),

434

TABLE 1: CLASSIFICATION RATES OF PAIR-WISE SVMs FOR DIFFERENT VISUAL FEATURES, ESTIMATED USING FIVE-FOLD CROSS VALIDATION ON TRAINING SETS.

Feature Classifier	Scalable colour	Colour layout	Edge histogram	Homogeneous texture	Colour structure	Dominant colour	GDH
landscape - cityscape	82.0	71.1	87.4	74.6	84.8	66.3	91.8
landscape - vehicle	86.6	80.4	92.3	76.2	89.3	69.5	92.7
landscape - portrait	88.9	86.9	92.0	81.7	89.4	77.4	84.6
cityscape - vehicle	87.6	77.5	94.5	75.7	88.9	68.9	91.6
cityscape - portrait	90.2	84.0	90.4	76.8	90.8	77.9	90.9
vehicle - portrait	90.9	82.3	97.5	80.7	91.3	75.6	93.8

- the MPEG-7 colour structure (256 elements),
- the MPEG-7 scalable colour (256 elements),
- the MPEG-7 colour layout (12 elements) and
- the MPEG-7 dominant colour (32 elements).

For each visual feature, we apply the five-fold cross validation technique (see Section 3-B.3) to estimate its corresponding classification rate on the validation set. The results in Table 1 show that across six pair-wise classification tasks, the proposed gradient direction histogram and the MPEG-7 edge histogram outperform all other visual features. Among the visual features, the MPEG-7 dominant colour performs the worst: this proves that there are no clear dominant colours in the four image classes of landscape, cityscape, vehicle and portrait. Based on the cross-validation results on the training set, we choose the proposed GDH and the MPEG-7 edge histogram for subsequent analysis.

C. Comparison of three methods for multi-class SVMs

We compare three different methods for multi-class SVMs: one-versus-all, pair-wise with majority voting and pair-wise with confidence score voting. The SVMs use either the proposed GDH feature or the MPEG-7 edge histogram: the exact feature to use is automatically determined based on the cross-validation results in the training set.

- **One-versus-all SVM method.** This method requires only four SVMs. Table 2 shows the classification results obtained by one-versus-all SVM method. Only 68.4% landscape images and 70.1% cityscape images are classified correctly. The classification rates for vehicle and portrait categories are better, at 85.7% and 80.1% respectively.

- **Pair-wise SVM with majority voting method.** This method requires six pair-wise SVM classifiers. We assign indices 1, 2, 3, 4 to landscape, cityscape, vehicle, portrait respectively. When two classes have identical voting scores, we choose the class with a smaller index. Table 3 shows the classification performance of this method, in the form of a confusion matrix.

- **Pair-wise SVM with confidence score voting method.** This method requires six pair-wise SVM classifiers. The classification rates of this method are shown in Table 4.

- **Comments.** Results in Tables 2, 3 and 4 show that both pair-wise SVM methods clearly outperform the one-

versus-all SVM method. The overall classification rates for one-versus-all SVM, pair-wise SVM with majority voting, and pair-wise SVM with confidence score voting are 76.1%, 81.6% and 83.4%, respectively. The confidence score voting method performs better than the majority voting method.

TABLE 2: CLASSIFICATION RATES FOR THE ONE-VERSUS-ALL SVM METHOD, ON THE TEST SET OF FOUR CLASSES.

Category	Classification rate (%)
Landscape	68.4
Cityscape	70.1
Vehicle	85.7
Portrait	80.1

TABLE 3: CONFUSION MATRIX OF PAIR-WISE SVM WITH MAJORITY VOTING, ON THE TEST SET OF FOUR CLASSES. ENTRY AT (ROW i, COLUMN j) IS THE PERCENTAGE OF CLASS i SAMPLES THAT ARE CLASSIFIED AS CLASS j SAMPLES.

(%)	Landscape	Cityscape	Vehicle	Portrait
Landscape	83.9	5.3	3.5	7.3
Cityscape	12.7	78.8	3.3	5.2
Vehicle	13.9	1.9	83.3	0.9
Portrait	8.1	10.4	1.2	80.3

TABLE 4: CONFUSION MATRIX OF PAIR-WISE SVM WITH CONFIDENCE SCORE VOTING, ON THE TEST SET OF FOUR CLASSES.

(%)	Landscape	Cityscape	Vehicle	Portrait
Landscape	84.9	5.1	3.2	6.8
Cityscape	8.9	80.7	4.0	6.4
Vehicle	12.0	2.1	84.1	1.8
Portrait	6.4	8.9	0.9	83.8

D. Comparison with other techniques

For comparison purpose, we apply the k-NN classifier on the same dataset. We experiment with the k-NN classifier where k varies from 1 to 15. Tables 5 and 6 show the classification rates of the k-NN classifiers using the proposed GDH feature and the MPEF-7 edge histogram. The overall classification rates of the k-NN using these two features are 72.6% and 74.5%, respectively. These results show that on average, the k-NN method performs worse than the three SVM methods presented above. We are conducting further comparison study with more complex classifiers including neural networks.

TABLE 5: CONFUSION MATRIX OF THE k-NN CLASSIFIER USING THE PROPOSED GDH FEATURE AND $k = 11$, ON THE TEST SET OF FOUR CLASSES.

(%)	Landscape	Cityscape	Vehicle	Portrait
Landscape	68.1	10.7	6.6	14.6
Cityscape	13.5	72.0	7.5	7.0
Vehicle	10.7	6.0	80.2	3.1
Portrait	15.0	11.1	3.9	70.0

TABLE 6: CONFUSION MATRIX OF THE k-NN CLASSIFIER USING MPEG-7 EDGE HISTOGRAM AND $k = 13$, ON THE TEST SET OF FOUR CLASSES.

(%)	Landscape	Cityscape	Vehicle	Portrait
Landscape	63.9	15.1	16.1	4.9
Cityscape	14.3	73.5	6.2	6.0
Vehicle	9.5	3.7	86.3	0.5
Portrait	10.6	13.5	1.5	74.4

5. CONCLUSION

In this paper, we have presented a multi-class image classification system that combines salient visual features and support vector machines. We propose a visual feature called edge direction histogram and investigate a wide range of MPEG-7 visual descriptor in an application that involves four image categories - landscape, cityscape, vehicle and portrait. We find that the proposed edge direction histogram and the MPEG-7 edge histogram perform better than other features in this application. The pair-wise SVM approach is found to perform better than the one-versus-all SVM approach. The pair-wise method with confidence score voting has better classification rates compared to the pair-wise method with majority voting.

REFERENCES

[1] Y. Chen, J. Li, and J. Z. Wang, *Machine Learning and Statistical Modeling Approaches to Image Retrieval*. New York: Kluwer Academic Publishers, 2004.

[2] F. Long, H. Zhang, and D. D. Feng, "Fundamentals of content-based image retrieval," in *Multimedia Information Retrieval and Management - Technological Fundamentals and Applications*, D. Feng, W. Siu, and H.J.Zhang., Eds. Berlin: Springer, 2002.

[3] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[4] O. Marques, L. M. Mayron, G. B. Borba, and H. R. Gamba, "Using visual attention to extract regions of interest in the context of image retrieval," in *Proceedings of the 44th annual Southeast regional conference*. Melbourne, Florida: ACM Press, 2006.

[5] J. P. Eakins, "Retrieval of still images by content," in *Lecture Notes in Computer Science: Lectures on Information Retrieval*. Berlin: Springer, 2001, vol. 1980/2001, pp. 111–138.

[6] L. Wang, L. Liu, and L. Khan, "Automatic image annotation and retrieval using subspace clustering algorithm," in *2nd ACM International Workshop on Multimedia databases*. Washington, DC, USA: ACM Press, 2004.

[7] J. Li and W. J. Z., "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.

[8] O. C. Ozcanl and F. Yarman-Vural, "An image retrieval system based on region classification," in *Lecture Notes in Computer Science: Computer and Information Sciences*. Berlin: Springer, 2004, vol. 3280/2004, pp. 449–458.

[9] J. Lu, S.-p. Ma, and M. Zhang, "Automatic image annotation based-on model space," in *Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering*, 2005, pp. 455–460.

[10] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Prentice Hall, 2002.

[11] B. S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG-7: multimedia content description interface*. Chichester: Wiley, 2002.

[12] M.-. V. Group, "Text of iso/iec 15938-3/fdis information technology - multimedia content description interface - part 3 visual," in *ISO/IEC JTC1/SC29/WG11/N4358*, Sydney, 2001.

[13] F. Nack and A. T. Lindsay, "Everything you wanted to know about mpeg-7, part 1," *IEEE Multimedia*, vol. 6, no. 3, pp. 65–77, 1999.

[14] ——, "Everything you wanted to know about mpeg-7, part 2," *IEEE Multimedia*, vol. 6, no. 4, pp. 64–73, 1999.

[15] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[16] N. Cristianini and J. Shawe-Taylor, *An Introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press, 2001.

[17] S. Abe, *Support vector machines for pattern classification*. New York: Springer, 2005.

[18] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[19] L. Wang and X. Fu, *Data Mining with Computational Intelligence*. Berlin: Springer, 2005.

[20] B. Schlkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, 1st ed. Cambridge, MA: The MIT Press, 2001.

[21] V. N. Vapnik, *The nature of statistical learning theory*, 2nd ed. New York: Springer, 2000.

[22] J. H. Friedman, *Another Approach to Polychotomous Classification*, 1996. [Online]. Available: http://www-stat.stanford.edu/~jhf/ftp/poly.ps.Z.

[23] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2007. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[24] W. Shao, G. Naghdy, and S. L. Phung, "Automatic image annotation for semantic image retrieval," in *Lecture Notes in Computer Science: VISUAL2007*, G. Qiu, Ed. Berlin Heidelberg: Springer-Verlag, 2007, vol. 4781, pp. 372–381.

[25] Institute for Integrated Systems, *MPEG-7 eXperimentation Model (XM)*, 2005. [Online]. Available: http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/mpeg7.html.