Cogn Process
DOI 10.1007/s10339-008-0247-6

RESEARCH REPORT

Improving image annotation via useful representative feature selection

Wei-Chao Lin · Michael Oakes · John Tait · Chih-Fong Tsai

Received: 1 February 2007 / Revised: 16 July 2008 / Accepted: 13 November 2008

Abstract This paper describes the automatic assignment of images into classes described by individual keywords provided with the Corel data set. Automatic image annotation technology aims to provide an efficient and effective searching environment for users to query their images more easily, but current image retrieval systems are still not very accurate when assigning images into a large number of keyword classes. Noisy features are the main problem, causing some keywords never to be assigned to their correct images. This paper focuses on improving image classification, first by selection of features to characterise each image, and then the selection of the most suitable feature vectors as training data. A Pixel Density filter (PDfilter) and Information Gain (IG) are proposed to perform these respective tasks. We filter out the noisy features so that groups of images can be represented by their most important values. The experiments use *hue*, *saturation* and *value* (HSV) colour feature space to categorise images according to one of 190 concrete keywords or subsets of these. The study shows that feature selection through the PDfilter and IG can improve the problem of spurious similarity.

Keywords Image annotation · Image retrieval · Information gain

W.-C. Lin (✉) · M. Oakes
Department of Computing, Engineering and Technology, University of Sunderland, Sunderland SR6 0DD, UK
e-mail: wei-chao.lin@sunderland.ac.uk
URL: http://www.cet.sunderland.ac.uk/IR/

M. Oakes
e-mail: Michael.Oakes@sunderland.ac.uk
URL: http://www.cet.sunderland.ac.uk/IR/

J. Tait
Information Retrieval Facility, Eschenbachgasse 11/3 Stk., 1010 Vienna, Austria
e-mail: John.Tait@ir-facility.org
URL: http://www.johntait.net/home.html

C.-F. Tsai
Department of Information Management, National Central University, Jhongli 32001, Taiwan
e-mail: cftsai@mgt.ncu.edu.tw

Introduction

Nowadays, due to the speedy development of the Internet and computing technologies, the number of visual information collections is increasing day-by-day. Millions of people access digital images and/or multimedia documents from the Internet daily. Therefore, effective and efficient retrieval techniques need to be developed and incorporated into current search engine style systems. Automatic image annotation technology assigns relevant keywords to each image in the data set, in order to provide an efficient and effective searching environment for users to query their image databases more easily (Del Bimbo 1996).

Due to the semantic gap problem (Gupta et al. 1997), current image retrieval systems are still not very accurate when they are required to assign images into one of a large number of keyword classes. An example of the semantic gap problem is when two images, such as a fingerprint and a wave pattern near the sea shore, have similar edge-orientation visual features but represent very different concepts (Vailaya 2000). Lavrenko et al. (2003) and Jeon and Manmatha (2004) classified images into more than 100 categories with low accuracy. In addition, the semantic gap problem also resulted in some keywords being found to be

Published online: 13 December 2008

unreachable or unassignable to any class when Tsai et al. (2006) carried out 150 keyword classification. The most successful current image classification systems are based on supervised machine learning (Carneiro et al. 2007; Del Bimbo 1999). With machine learning, noisy features are the main problem for image classification, which cause some keywords never to be assigned to their correct images. For example, a noisy image of a car might include other image types, such as tree, road or grass. In training, the features typical of a car might be swamped by those of the other image types.

Following Tsai (2005), this work interprets the image annotation task as being the task of identifying that class of images to which an annotation label applies at some level of accuracy.

In this paper, we will address two main tasks in machine learning for image classification. First, we will consider how best to perform feature selection. A feature is any characteristic used to represent an image which is used to assist with its classification, features typically being colours, textures or shapes. The Pixel Density filter (PDfilter) is proposed here to select representative features which are more similar to their values in the original image than is possible with traditional approaches. It works by averaging the feature values in the area of HSV colour space which contains most pixels. The second task we address is that of training data selection. Rather than using all the feature vectors derived from training set images as training data, we select just those vectors which best discriminate between the various keyword categories. For the task of feature vector selection, we propose the use of Information Gain (IG) (Breiman et al. 1984), which is used to select the features which hold most information about each keyword category. IG gives a weight, called gain value, between the feature vectors for each image and each single keyword category, and then extracts the most useful feature vectors (those with highest gain value with respect to a category) to enable training data improvement. A supervised machine-learning technique is then applied for image annotation. The machine learning technique used in this paper is based on the *k*-Nearest Neighbour (*k*-NN) classifier (Bishop 2006), where training and testing data sets enable classification of feature vectors into particular keyword classes.

Related work

Content-based image retrieval (CBIR), first proposed in the early 1990s, has been a lively research area over the past few years (Müller et al. 2003). It not only provides a search methodology which enables images to be retrieved by the contents of the images themselves, but also aims to supply: (1) the ability to handle visual queries, (2) a friendly query model for users to approach, and (3) automatic descriptions of image content features (Wu et al. 2000). In the indexing process, CBIR systems extract image features automatically to facilitate retrieval (Eakins and Graham 1999). Images may be segmented or analysed in terms of their constituent areas or regions, before low-level feature extraction takes place (Idris and Panchanathan 1997). Complex schemes are required to segment an entire image. The size and shape of segments that can be assigned by different resolutions is an active area of research (Tsai 2005). The perceptual features used in automatic low-level feature extraction include colour (Lai 2000; Wu et al. 2000), texture (Howarth and Rüger 2004), shape content (Shanbehzadeh et al. 2000) and spatial object layout recognition (Mandal et al. 1999). Images may be annotated off-line manually, typically with information such as the image's creator, creation date or event activity (Eakins and Graham 1999). Various authors have developed retrieval engines, which allow users to search for images by keywords (Barnard et al. 2003), query by example, where users submit an image similar to the one they are looking for (Jin and French 2003), or query by image feature (Del Bimbo 1999). The users are typically given a searching environment to inspect the initial set of images retrieved in response to the query, and to filter out the images not considered relevant. In systems which allow relevance feedback, user judgements of search results are fed back into the system to refine the query, so that new results are more fitting (Lew 2001). The main challenge for current CBIR systems is overcoming the semantic gap, which is the arduous task of translating low-level features into high-level concepts (Gupta et al. 1997). While CBIR systems store and retrieve images by low-level indexing, most users would like to submit queries for images using semantically meaningful high-level concepts. As a step towards closing the semantic gap, Barnard et al. (2003) and Tsai et al. (2006) have shown the feasibility of using a probabilistic model, such as Latent Dirichlet Allocation (LDA) (Blei et al. 2003), and/or machine learning techniques, such as the *k*-NN classifier (Bishop 2006) and Support Vector Machines (SVMs) (Burges 1998), to automatically assign controlled vocabularies of around 150 words to unseen images. However, a powerful image searching environment would need to operate with much larger vocabularies, to allow users to query for images of interest efficiently and effectively (Tsai 2005). How to filter out noisy features to solve the semantic gap problem is the main challenge before larger vocabulary approaches. The PDfilter and IG are suggested in this paper, which select the most representative features of images, and then select the feature vectors most useful for classifying each image, thus aiming to avoid the influence of noisy features within image classification.

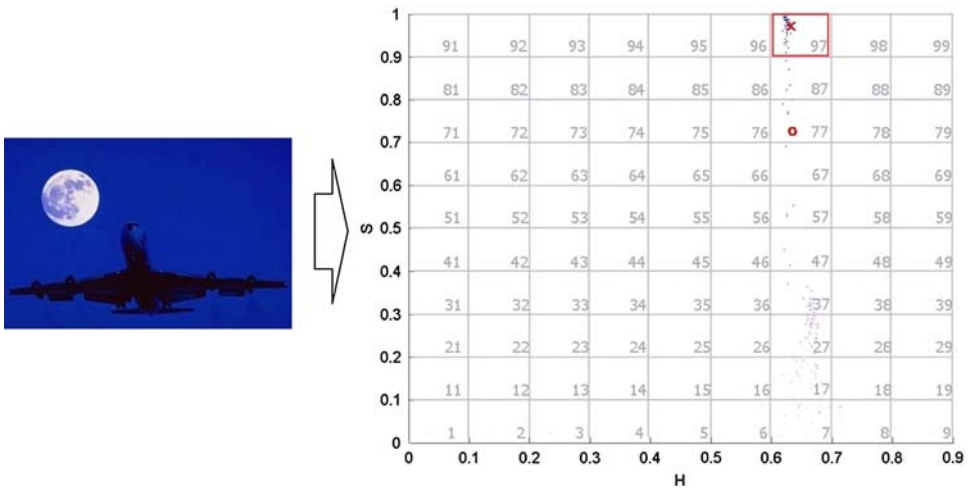

Fig. 1 Example of Pixel Density filter operation

Technology approach

Pixel Density filter

In automatic classification tasks, a set of features which characterise the original objects must be chosen. When selecting such features for images, we cannot simply use every single characteristic of the original image, since images are composed of a very large number of individual pixels. Individual pixels in the original image can be characterised by various values, including values mapped onto the range 0–1 for hue, saturation and value, which is called HSV colour space (van der Heijden 1994). However, using every single pixel's value will introduce more noise to the system than choosing a single representative value for all the pixels in each region into which the image has been segmented. In most related work, such as Barnard et al. (2003) and Tsai (2005), the representative image feature comes from the average of all pixel values within a region or tiling area. This can mean image feature values can be too similar to each other. For example, two regions composed of a mixture of red and green areas will be both represented similarly, as types of brown colour. The PDfilter proposed in this paper, however, will characterise each region by its predominant (modal) pixel value, leading to the possibility that one region would be represented by red, the other by green. Thus PDfilter aims to solve the problem of providing more representative features, which are more similar to their values in the original image. The method of Tsai (2005) is the baseline against which the novel approaches described in this paper are compared.

The inspiration for the PDfilter comes from the colour histogram (Swain and Ballard 1991). Here the colour space is quantised into a number of buckets, and the image is characterised by the number of pixels falling into the range of each bucket. Lai (2000) lists 125 bins, such as "red 0, green 0.55, blue 1.00" to represent sky blue. The colour histogram has greater discrimination power when the number of colour buckets is increased (Long et al. 2003). The PDfilter not only provides a much larger number of predefined buckets than the colour histogram, but also allows the use of different colour spaces, such as HSV and HSI colour spaces (Gonzalez et al. 2004), and a high dimensional texture feature representation.[1]

The experiments described here are based on HSV colour space. The PDfilter works with the feature spaces in individual regions of the analysed images. Based on Eq. 1, where d means the dimension number out of N dimensions; Xd is the value of each dimension and Xd_m is the maximum value of each dimension in the whole image collection, each pixel's feature value is quantified into a bucket $A(p)$ of a coordinate figure in S divisions. Finally, the representative feature values come from the average of the pixel values in the predominant bucket. As shown in Fig. 1, each pixel's value in the image is quantified into a coordinate figure, and then the area that holds most pixels is computed, as the representative feature average of the most predominant colour.

$$A(p) = \sum_{d=1}^{N} \left(\left\lfloor \frac{Xd \times S}{Xd_m} \right\rfloor \times S^{d-1} \right) + 1. \quad (1)$$

In order to show clearly the distribution of the pixel values, we simplify colour into just two dimensions, showing only *hue* (H) and *saturation* (S) in this example. Each pixel's value in the image is quantified into a coordinate figure, as determined by Eq. 1. For example, a pixel with real values (0.641, 0.971) will be allocated to bucket number 97, since $A(p) = \left(\left\lfloor \frac{0.641 \times 10}{1} \right\rfloor \times 10^{1-1} + \left\lfloor \frac{0.971 \times 10}{1} \right\rfloor \times 10^{2-1} \right) + 1 = (6 + 90) + 1 = 97$.

[1] Images can be characterised by texture as well as colour. Texture is typically described by the wavelet transform (Daubechies 1992). However, the emphasis of this paper is on colour.

After PDfilter selection, the representative feature of the whole region will be the symbol at point **x** (0.630, 0.976) in area 97, in contrast to the average of all pixel values at location **o** (0.626, 0.714), as used in the baseline system (Tsai 2005). Thus, the baseline system and the PDfilter system both represent an analysed image by coordinate dimensions, but their values are different.

In the experiments described here, each image was initially segmented into five regions, each of which as a result of the PDfilter becomes represented by a three-valued vector corresponding to H, S and V.

Information Gain application

The concept of information gain comes from Shannon's (1948) information theory and the decision trees of decision theory (Mitchell 1997). It is an information-theoretic criterion in the field of machine learning and is frequently employed in feature selection in text categorisation (Quinlan 1986; Yang and Pedersen 1997). This paper applies IG for training data selection, in order to allow each single category to be represented by its most important feature vectors after noise and uncertainty reduction. IG is the weighted sum of the entropies or gain values of all the feature vectors in each keyword category. It measures the information required to predict the presence or absence of a feature vector in a given keyword category. In the original training data set, there are 20 training images for each keyword category. Each image is automatically segmented into 5 regions, yielding a total of 100 training vectors for each keyword. For each keyword, *K-means* clustering (Manning and Schütze 1999) is used to sort the feature vectors of the analysed images into a number k of clusters; we set $k = 10$. In Eq. (2), m is the total number of regions in an image and $\{Ci\}_{i=1}^{m}$ is the set of images within a single category. The gain value of an individual cluster (t) is then defined (Yang and Pedersen 1997) as follows:

$$\begin{aligned} G(t) = &- \sum_{i=1}^{m} P(Ci) \log P(Ci) \\ &+ P(t) \sum_{i=1}^{m} P(Ci|t) \log P(Ci|t) \\ &+ P(\bar{t}) \sum_{i=1}^{m} P(Ci|\bar{t}) \log P(Ci|\bar{t}). \end{aligned} \tag{2}$$

The overall IG for a cluster is thus the sum of the gain values for each image vector in that cluster. IG reduces uncertainty, since only the documents of the clusters with an above threshold gain value are retained as training data for future classification. However, there are no exact algorithms to find the optimal splitting value. In practice, the threshold is based on heuristics to find a near-optimal value (Manning and Schütze 1999). The training data for our experiments are all taken from the single cluster with the highest gain value.

Main experiment

Research question

In this paper, we explore two main questions: (1) can the PDfilter solve the similarity problem in representative feature selection? and (2) can IG operate effectively in training data selection?

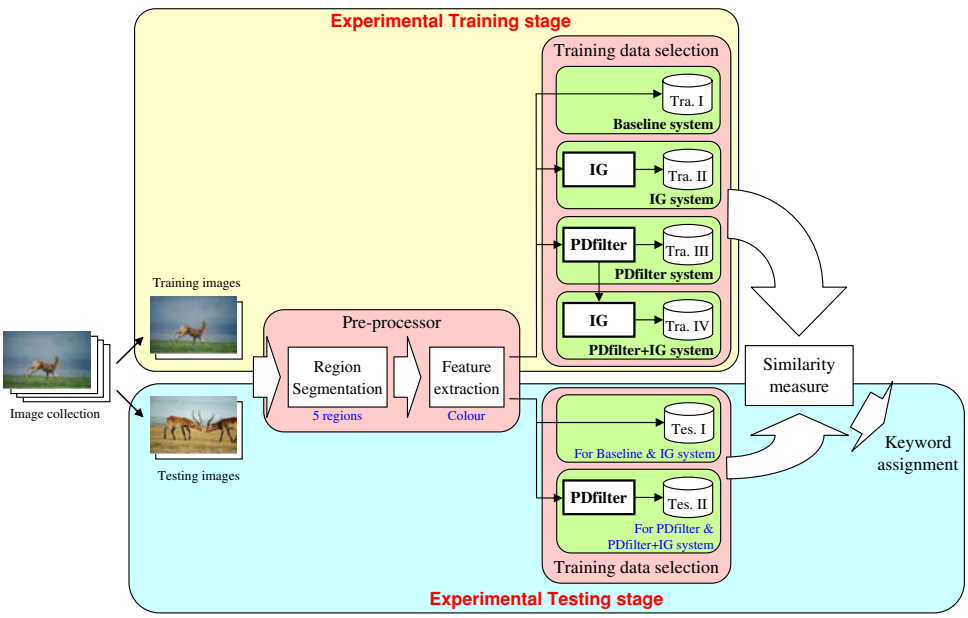

Fig. 2 Framework of the experiment

The PDfilter and IG are the main focuses of experimentation in this paper. This combination has never been applied into any related image retrieval systems before. Simultaneously, this paper also extends the size of the controlled vocabulary to 190 concrete keywords.

Experimental framework

The experimental framework (shown in Fig. 2) describes how we assign related keywords to each analysed image, in order to enable users to submit queries for images of interest through keyword-based retrieval. The experimentation is separated into a training stage and a testing stage, and the image collection is also divided into two groups, one for each stage. Based on different operational models, the experiment consists of four systems, and each system will create its own database for its related feature storage and analysis. The experimental execution is enabled by the following components:

Region segmentation

Since we are going to select useful training data by IG, image segmentation is based on region-based methods in the Normalised Cuts (Ncut) algorithm (Shi and Malik 2000), which segments images by the pixels' colour similarity and proximity cues. According to Barnard's et al. (2003) analysis, segmentation into a large number of regions can provide more exact annotation. However, in this paper, the experiment is restricted to five regions or sub-image segments to reduce the computational cost.

Feature extraction

In order to prove the idea that feature selection by pixel density can improve the problem of similarity between each feature value, the experiments of this paper only use colour feature vectors, which are the positions in HSV colour space (van der Heijden 1994) that represent colour by every pixel's *hue* (H), *saturation* (S) and *value* (V). The advantage of the HSV colour space, in contrast to the *red* (R), *green* (G) *blue* (B) colour model, is that the distances between colours correspond to human perception (Mathias and Conci 1998).

Pixel density filter

This component aims to represent each image by selected features, which are more similar to their values in the original image than would be the case in the baseline system. Each pixel's value is quantified into a coordinate figure, and then the area that holds most pixels is computed, and the representative feature is the average of the pixel values in this predominant area.

Information gain

This component applies IG to select useful representative image feature vectors for training data for each single category. This filters out any noise in the classification of that category since we only use the most important feature vectors.

Similarity measure

The *k*-NN classifier assigns new instances of images into their *k* nearest neighbours (Mitchell 1997), and uses the Euclidean Distance similarity measure between the training and testing data. Analysis by Jain et al. (2000) showed that $k = 1$ (*1NN*) allows reasonable classification for most applications. In our experiments we also apply *1NN*, which allows us to assign a training example's keyword to the nearest testing set simply and easily.

Data set

The image collection for the experiments comes from the four Corel Stock photo libraries[2] and the Corel Gallery 1,300,000.[3] This consists of 68,600 photographs that are sorted into 686 categories published by Corel Corporation, and each category contains 100 images to represent a topic.

In contrast with Li and Wang's (2003) application, our experiment aims to represent each group of images by a single keyword, in order to connect low-level features with their related keywords. This enables the system to understand which kinds of feature can be explained by which kind of keyword, and then to assign relevant keywords into each testing image. We assigned a single keyword to each category according to the original description by the publisher, Corel. Some of the descriptions from Corel were spread over more than one category, such as "children" and "children II" or "classic cars" and "classic automobiles". We combined such categories into single keyword classes.

The WordNet[4] online lexical reference system can be used to discriminate the type of keyword. According to WordNet, a word which is a kind of physical entity is a concrete concept, such as "agate", "car", "dog", and so on. A kind of abstraction and/or human activity is an

[2] A software review is given at: http://www.uottawa.ca/academic/cut/options/Nov_96/CorelCDs.htm.

[3] A software review is given at: http://www.gtpcc.org/gtpcc/corelgallery.htm.

[4] For more information about WordNet please visit http://www.wordnet.princeton.edu/.

Table 1 The 190 concrete keywords

1 agate	39 cougar	77 garden ornament	115 mushroom	153 soldier
2 aircraft	40 cowboy	78 gemstone	116 musical instrument	154 space
3 aircraft illustration	41 cruise ship	79 glass	117 national park	155 space voyage
4 amusement park	42 crystal	80 goat	118 navy SEAL	156 spice & herb
5 animal	43 decorated pumpkin	81 graffito	119 nest	157 port car
6 antelope	44 deer	82 hairstyle	120 object	158 statue
7 ape	45 desert	83 hand-painted	121 ocean life	159 steam engine
8 backyard wildlife	46 dessert	84 harbour	122 office	160 steam train
9 bald eagle	47 dining	85 hawk	123 oil painting	161 steamship
10 balloon	48 dinosaur	86 heavy machinery	124 orbit	162 swimsuit
11 bark texture	49 dish	87 helicopter	125 orchid	163 tall ship
12 beach	50 dog	88 horse	126 owl	164 tennis
13 bead	51 dog sled	89 hotel	127 penguin	165 textile
14 bear	52 doll	90 house & cottage	128 pet	166 tiger
15 beverage	53 dolphin and whale	91 iceberg	129 pill	167 tool
16 bird	54 door	92 insect	130 plant	168 toy
17 bird art	55 drawing	93 isle	131 plant microscopy	169 train
18 boat	56 duck decoy	94 jewellery	132 playing card	170 tram
19 bobsled	57 earth	95 landmark	133 polar bear	171 transport
20 bonsai	58 Easter egg	96 landscape	134 portrait	172 tree & leaf
21 botanical prin	59 elephant	97 leopard	135 predator	173 tulip
22 bridge	60 everglade	98 lighthouse	136 prehistoric world	174 UK royal
23 Buddha	61 fabric	99 lion	137 pub sign	175 university & college
24 building	62 face	100 mammal	138 pyramid	176 vegetable
25 bus	63 firearms	101 marble	139 reef	177 warplane
26 butterfly	64 fireworks	102 mask	140 reflection	178 warship
27 cactus flower	65 fish	103 men	141 religious stained glass	179 waterfall
28 canal and waterway	66 flora	104 merchant ship	142 road	180 waterscape
29 car	67 flower	105 microscopic image	143 road sign	181 whitetail deer
30 castle	68 flowerbed	106 mineral	144 rock formation	182 wildcat
31 cat	69 foliage	107 model	145 rose	183 wilderness
32 cave	70 food	108 molecule	146 sailboard	184 wildflower
33 cavern	71 frost	109 monument	147 sailboat	185 wildlife painting
34 children	72 fruit	110 mosaic	148 sailing ship	186 wolf
35 church and cathedral	73 fur feathers and skin	111 motorcycle	149 sculpture	187 woman
36 cloud	74 furniture	112 mountain	150 seed	188 WWII Planes
37 coast	75 game bird	113 mural	151 shell	189 yellow
38 costume	76 garden	114 museum	152 sky	190 young animal

abstract keyword, like "agriculture", "design" or "summer". Unlike Tsai (2005), we regarded an assemblage of multiple physical or entity objects as a single entity, such as "harbour" or "building", and as a concrete class. In addition, location is also used as an attribute. This work is similar to the assignment of high-level concepts into different levels, as in Jörgensen's et al. (2001) work to categorise levels into generic/specific/abstract levels.

Altogether, there are 446 keywords defined in total, 190 concrete keywords, 138 abstract keywords and 119 location classes. The experiments described in this paper used up to 190 concrete keywords, and the list is shown in Table 1. These experiments are based on concrete keywords, in order to prove the concept of the PDfilter and IG for feature selection. In future we will also examine the performance of these techniques with abstract and location keywords.

Experimental set up

A total of 190 concrete keywords were used in this paper. Separate experiments were performed with 10, 50, 100,

150 and 190 concrete keywords. The experiments of 10, 50, 100 and 150 concrete keywords constituted 10 sub-experiments, each using a different random selection of keywords from the 190 concrete keyword set. With the experiment of 190 concrete keywords only one combination of keywords was used, the full set.

Following Tsai's (2005) work, the images for analysis were resized into 128 × 128 pixels prior to system execution, and then segmented into five regions of sub-images by the pixels' colour similarity and the proximity cues used to determine texture. The experiments worked with 20 images in the training and testing sets, respectively. In this study based on different combinations of training data and testing data, four independent systems are created: (1) the *baseline* system, (2) the *IG* system, (3) the *PDfilter* system, and (4) the *PDfilter* + *IG* system. The *baseline* system is based on Tsai's (2005) work. In these experiments each region is represented by a feature vector, and its values come from one of two different calculations for all training and testing data. The *baseline* and *IG* systems represent a region by the average of all pixels' values within the region, and the remaining systems, which are related to the PDfilter, use the average of the most predominant values selected by the PDfilter. Training data are presented as one feature vector for an image. The *baseline* and *PDfilter* systems consider only the central region, which includes the central pixel of (64, 64), while in the *IG* and *PDfilter* + *IG* systems feature vectors are selected by IG. Table 2 lists the details of every feature value calculation and training data applied in these experiments.

Evaluation model

Recall and *precision* evaluation measures are the most common in many information retrieval applications (Baeza-Yates and Ribeiro-Neto 1999). Recall, shown in Eq. (3), measures the fraction of relevant images retrieved or placed in a desired category relative to the total number of relevant images. Precision, shown in Eq. (4), measures how many of the images retrieved are in fact relevant to the user's interest (Belew 2000; Oakes 1998). They can be defined by the following equations:

$$\text{Recall} = \frac{\text{TP}}{\text{Relevant}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{Retrieved}}. \tag{4}$$

In our experiments, TP (*true positives*) is the number of testing images correctly classified into a category; *Relevant* is number of testing images assigned to this category by the Corel data set; *Retrieved* means how many images are assigned to this category by the automatic system.

Table 2 Details of feature value calculation and training data application

	Feature value calc.		Training data			
	A	B	I	II	III	IV
Baseline	×		×			
IG	×			×		
PDfilter		×			×	
PDfilter + *IG*		×				×

Feature value calculation

A: average of all pixels' value within the region

B: average of the pixel values of the predominant colour

Training data

I and III: feature vector from the central region

II and IV: useful feature vector selection by information gain

This study also evaluates the number of unassigned keywords. In Eq. (5), *U* is the percentage of the keywords that are never assigned to any related images. However, a small number of unassigned keywords do not necessarily prevent the system from producing high recall or precision.

$$U = \frac{\text{Unassigned keywords}}{\text{Total number of keywords}}. \tag{5}$$

Results and discussion

Recall and precision evaluation

Figure 3 shows the results of this study, and summarises the performance of the *baseline*, *IG*, *PDfilter*, and *PDfilter* + *IG* systems by recall and precision. At the same time, error bars of one standard deviation over the 10 sub-experiments accompany the 10, 50, 100, and 150 category results.

Focusing on the recall curve, the experiments which applied IG alone produced disappointing results. However, results were better for the PDfilter application. For 10 categories, the *PDfilter* and *PDfilter* + *IG* systems are both below the *baseline* system, but the curve overtakes the baseline system at 50 and 100 categories. Additionally, both systems are better than the *baseline* system in terms of precision. The experiments performed especially well with the PDfilter. This suggests that the PDfilter model can be applied to discover the distinguishing features between each category, which improves performance.

Unassigned keyword evaluation

After implementation of both systems, some category keywords still could not be assigned to their related

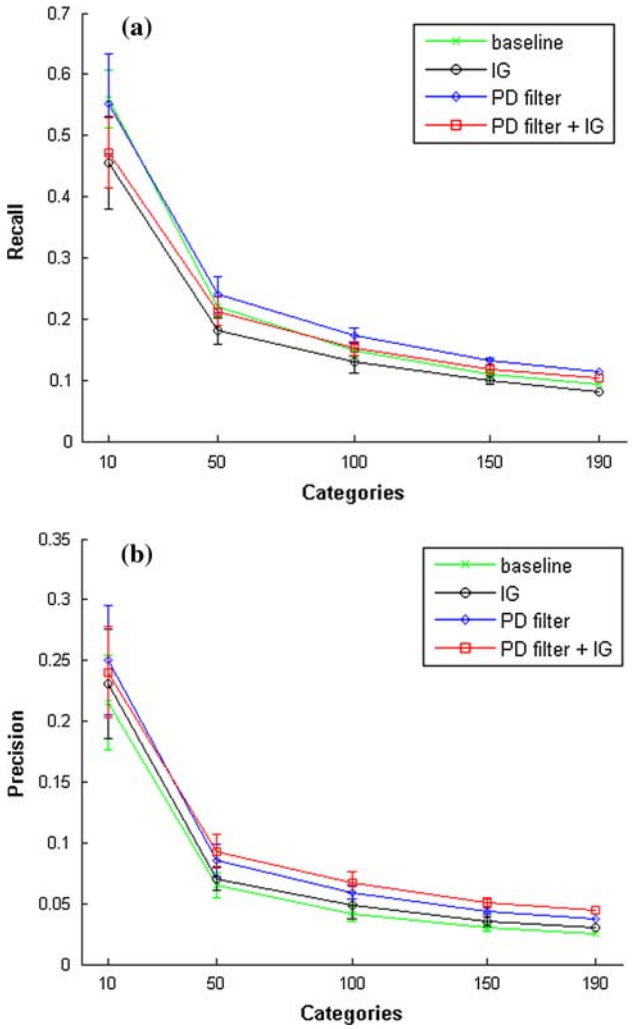

Fig. 3 Experimental results for recall and precision, over 10, 50, 100, 150 and 190 categories with standard deviation for *baseline*, *IG*, *PDfilter*, and *PDfilter* + *IG* systems

images. Figure 4 illustrates this problem using as an example keywords which start with "a" or "b". There are some keywords, such as "bird", which cannot be assigned to their related image by any of the observed systems.

The results for unassigned keywords are shown in Fig. 5. Error bars of one standard deviation are shown to examine the variations of each system for each number of categories. According to this analysis, the *Baseline* system supplies the fewest unassigned keywords for 50, 100 and 150 category classification, but performs badly when images are classified into 190 categories. There are more keywords which cannot be assigned by the *IG* and *PDfilter* + *IG* systems, but both systems perform better than the *baseline* in the precision analysis. In addition, the *PDfilter* system produces the fewest unassignable keywords in 190 categories, and seems to offer a flexible system for image annotation.

Discussion

The recall and precision analyses show that image classification can be improved successfully after colour feature representation obtained from the PDfilter. The reason is that the accuracy of image classification and retrieval will be impaired when image feature values are too similar to each other. When people search for their objects of interest, they usually focus on one specific feature first. For example, the colour of "tiger" is always linked with the orange specific to the tiger itself rather than the dark yellow which would come from the overall average colour value of an image of a tiger hiding in grass. The PDfilter likes to represent each category by its most important value, thus making each feature more related to its original character. This paper has proved that the PDfilter model can work with colour analysis, and in future we will examine its performance in image classification experiments based on texture.

The results for the systems which make use of IG show that when using IG alone without the PDfilter, low recall occurs as a result of noisy selection. Nevertheless, the *PDfilter* + *IG* combined system produces the highest recall and precision of all, showing the benefit of the PDfilter in feature representation. Even though IG can be a good method for filtering out noisy features, when feature values are too similar, the system will also filter out some important data.

Regarding the unassigned keyword evaluation, there were more keywords that cannot be assigned to any image when using the *PDfilter* + *IG* system than for any of the other experiments. Thus PDfilter and IG improved recall and precision, but did not solve the problem of unassignable keywords. In addition, the problem of unassigned keywords may also depend on the image collection of the dataset. Testing data that cannot be assigned may be of a very general nature, such as "animal", "insect", or "ocean life". As a result, this study needs to consider image data sets other than Corel in future experiments.

Conclusion and future work

The results show that the PDfilter is a promising approach for image annotation systems. Moreover, the use of IG further increases recall and precision when applied in conjunction with the PDfilter. The PDfilter enabled us to select representative features with values similar to their values in the original image, resulting in better classification into keyword categories. In future we will extend the investigation by looking at other feature extraction methods. We also plan to use the PDfilter with other image features such as texture and shape, to try and solve the problem of unassignable keyword categories. Experiments

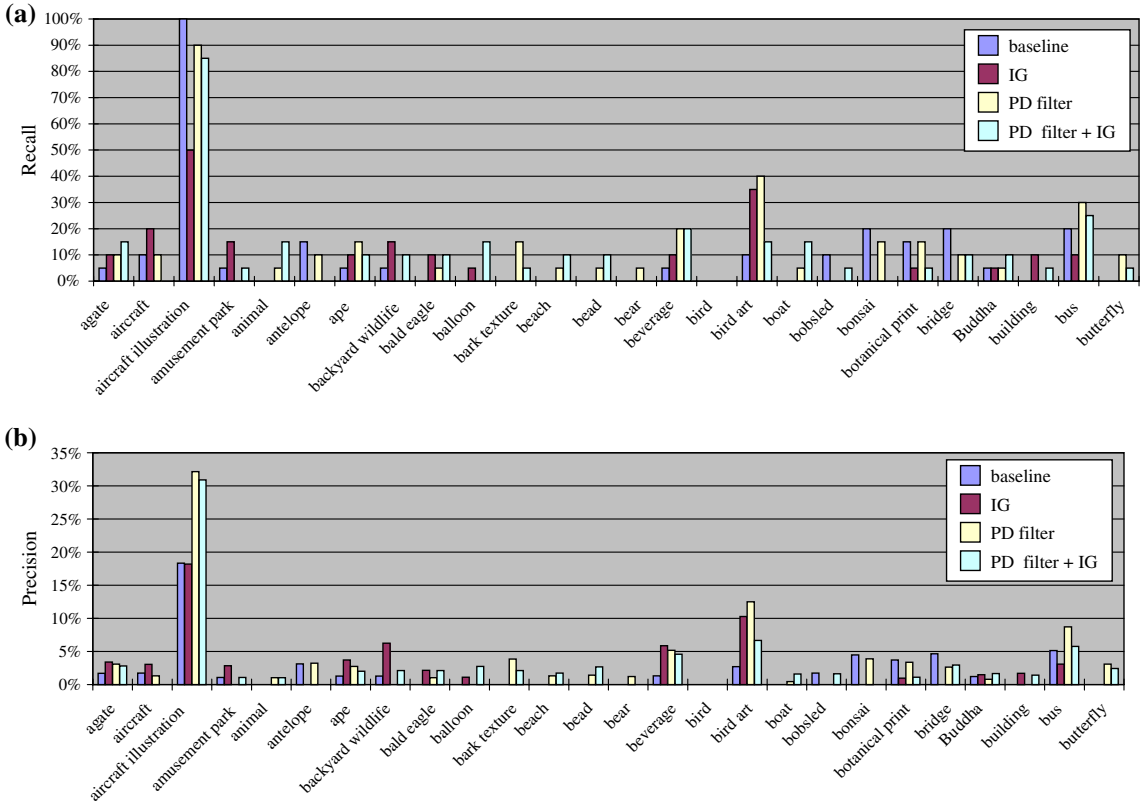

Fig. 4 Recall and precision for example categories from the experiment of 190 categories

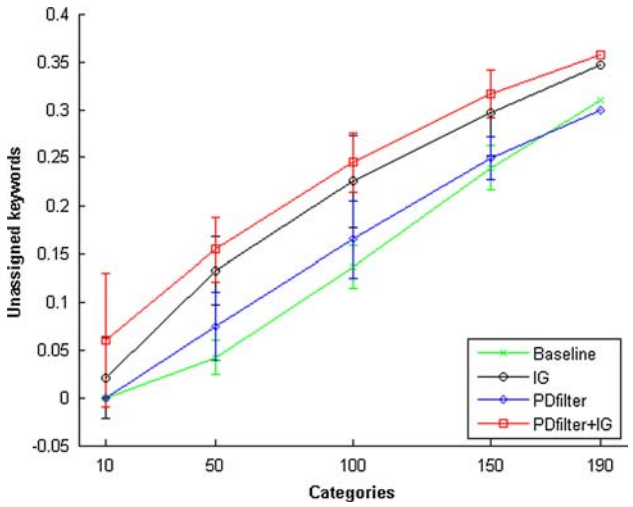

Fig. 5 Unassigned keywords measured, over 10, 50, 100, 150 and 190 categories with standard deviation for *baseline*, *IG*, *PDfilter*, and *PDfilter* + *IG* systems

will be conducted to observe the recall and precision in large vocabulary applications. In addition, The TRECVid data set (Smeaton et al. 2004), the IAPR TC-12 Benchmark (Grubinger et al. 2006) and the image collection built by the University of Washington[5] will be used for system evaluation, so as to show how this approach performs with other data sets.

Machine learning technology and related probabilistic classification methods will also be explored, in order to compare the approach of this study against other techniques. Finally, human centred evaluations will be performed to allow system improvement via a review of system usability.

References

Baeza-Yates R, Ribeiro-Neto B (1999) Modern information retrieval. Addison Wesley, England

Barnard K, Duygulu P, Forsyth D, de Freitas N, Blei DM, Jordan MI (2003) Matching words and pictures. J Mach Learn Res 3:1107–1135

Belew RK (2000) Finding out about: a cognitive perspective on search engine technology and the WWW. Cambridge University Press, Cambridge

Bishop CM (2006) Pattern recognition and machine learning. Springer, New York

[5] Available at: http://www.cs.washington.edu/research/imagedatabase/.

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022
Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. CRC Press, Boca Raton
Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Mining Knowl Discov 2(2):121–167
Carneiro G, Chan AB, Moreno PJ, Vasconcelos N (2007) Supervised learning of semantic classes for image annotation and retrieval. IEEE Trans Pattern Anal Mach Intell 29(3):394–410
Daubechies I (1992) Ten lectures on wavelets. Society for Industrial and Applied Mathematics, Philadelphia
Del Bimbo A (1996) Image and video databases: visual browsing, querying and retrieval. J Vis Lang Comput 7(4):353–359
Del Bimbo A (1999) Visual information retrieval. Morgan Kaufmann, San Francisco
Eakins JP, Graham ME (1999) Content-based image retrieval: a report of the JISC technology applications programme. The Joint Information Systems Committee (JISC). http://www.jisc.ac.uk/uploaded_documents/jtap-039.doc (26 January 2007)
Gonzalez RC, Woods RE, Eddins SL (2004) Digital image processing using MATLAB. Pearson Prentice-Hall, Upper Saddle River
Grubinger M, Clough P, Müller H, Deselaers T (2006) The IAPR TC-12 Benchmark—a new evaluation resource for visual information systems. In: Proceedings of the International Workshop OntoImage'2006 Language Resources for Content-Based Image Retrieval, held in conjunction with LREC'06. Genoa, Italy, 22 May 2006, pp 13–23
Gupta A, Santini S, Jain R (1997) In search of information in visual media. Commun ACM 40(12):35–42
Howarth P, Rüger S (2004) Evaluation of texture features for content-based image retrieval. International Conference on Image and Video Retrieval (CIVR), Dublin, pp 326–334
Idris F, Panchanathan S (1997) Review of image and video indexing techniques. J Vis Commun Image Represent 8(2):146–166
Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 22(1):4–37
Jeon J, Manmatha R (2004) using maximum entropy for automatic image annotation. In: Proceedings of the International Conference on Image and Video Retrieval, Dublin, Ireland, July 21–23 2004: 24–32
Jin X, French JC (2003) Improving image retrieval effectiveness via multiple queries. In: Proceedings of the First ACM International Workshop on Multimedia Database, New Orleans, LA, USA, pp 86–93
Jörgensen C, Jaimes A, Benitez AB, Chang S (2001) A conceptual framework and research for classifying visual descriptors. J Am Soc Inf Sci 52(11):938–947 Special Issue on Image Access: Bridging Multiple Needs and Multiple Perspectives
Lai T (2000) CHROMA: a photographic image retrieval system. PhD Thesis. University of Sunderland, UK
Lavrenko V, Manmatha R, Jeon J (2003) A model for learning the semantics of pictures. In: Proceedings of the International Conference on Neural Information Processing Systems, Vancouver, Canada, 8–13 December 2003
Lew MS (2001) Principles of visual information retrieval. Springer, London
Li J, Wang JZ (2003) Automatic linguistic indexing of pictures by a statistical modelling approach. IEEE Trans Pattern Anal Mach Intell 25(9):1075–1088
Long F, Zhang H, Feng DD (2003) Fundamentals of content-based image retrieval. In: Feng DD, Siu WC, Zhang H (eds) Multimedia information retrieval and management—technological fundamentals and applications. Springer, Germany
Mandal MK, Idris F, Panchanathan S (1999) A critical evaluation of image and video indexing techniques in the compressed domain. Image Vis Comput 17(7):513–529
Manning CD, Schütze H (1999) Foundations of statistical natural language processing. MIT, London
Mathias E, Conci A (1998) Comparing the influence of color spaces and metrics in content based image retrieval. In: Proceedings of the IEEE International Symposium on Computer Graphics, Image Processing, and Vision. Rio de Janeiro, Brazil, 20–23 October 1998, pp 371–378
Mitchell TM (1997) Machine learning. McGraw Hill, New York
Müller H, Müller W, Marchand-Maillet S, Pun T, Squire DM (2003) A framework for benchmarking in CBIR. Multimedia Tools Appl 21(1):55–73
Oakes MP (1998) Statistics for corpus linguistics. Edinburgh University Press, Edinburgh
Quinlan JR (1986) Induction of decision trees. Mach Learn 1(1):81–106
Shanbehzadeh J, Moghadam AME, Mahmoudi F (2000) Image indexing and retrieval techniques: past, present, and next. In: Proceedings of SPIE, The International Society for Optical Engineering, 3972, pp 461–490
Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J. 27, July and October: 379–423 and 623–656
Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 22(8):888–905
Smeaton AF, Kraaij W, OverP (2004) The TREC video retrieval evaluation (TRECVID): a case study and status report. In: Proceedings of the RIAO 2004 Conference. Avignon, France, 26–28 April 2004, pp 25–37
Swain MJ, Ballard DH (1991) Color indexing. Int J Comp Vis 7(1):11–32
Tsai C (2005) Automatically annotating images with keywords. PhD Thesis, University of Sunderland, UK
Tsai C, McGarry K, Tait J (2006) Qualitative evaluation of automatic assignment of keywords to images. Inf Process Manage 42(1):136–154
Vailaya A (2000) Semantic classification in image databases. PhD Thesis. Michigan State University, USA
van der Heijden F (1994) Image based measurement systems: object recognition and parameter estimation. Wiley, Chichester
Wu JK, Kankanhalli MS, Lim J, Hong D (2000) Perspectives on content-based multimedia systems. Kluwer Academic Publishers, London
Yang Y, Pedersen JO (1997) A comparative study on feature selection in text categorization. In: Proceedings of the Fourteenth International Conference on Machine Learning. 8–12 July 1997, pp 412–420