Feature Selection to Relate Words and Images

Wei-Chao Lin[1] and Chih-Fong Tsai*[,2]

[1]*Department of Computing, Engineering and Technology, University of Sunderland, Sunderland, SR6 0DD, UK*

[2]*Department of Information Management, National Central University, 300 Jhongda Rd., Jhongli, Taiwan*

Abstract: Image annotation, i.e. mapping words into images, is currently a major research problem in image retrieval. In particular, images are usually segmented into a number of regions, and then low-level image features are extracted from the segmented regions for annotation. As the extracted image features may contain some noisy features, which could degrade the recognition performance when the number of keywords assigned to images is very large, image feature selection needs to be considered. In this paper, a Pixel Density filter (PDfilter) and Information Gain (IG) are used as the feature selection techniques. By using Corel as the dataset, 10, 50, 100, 150 and 190 keywords annotation are setup for comparisons. The experimental result shows that PDfilter and IG can increase the precision of image annotation by colour or texture features. However, they do not enhance the annotation performance by the combined colour and texture features.

1. INTRODUCTION

Multimedia databases have become very popular for various applications, such as digital libraries, medical images, and news photos. Content-based Image Retrieval (CBIR) systems provide an effective and efficient retrieval technique to access multimedia databases for users to query relevant images through some perceptual (or low-level) features, such as colour, texture, shape and spatial object layouts. This technique requires complete indexing facilities and adequate data structures, to filter out images unrelated to a query, then visualise the final retrieval results in an accurate way [1]. The most successful applications are based on supervised machine-learning classifiers to match images with relevant keywords for general keyword-based queries. Therefore, noisy features are the primary problem for image indexing that cause current systems still not to be very accurate with large numbers of vocabularies.

This paper focuses on noisy feature filtering, to investigate one possible reason why current systems are unlikely to scale to larger vocabularies. A Pixel Density filter (PDfilter) and Information Gain (IG) [2, 3] are suggested to solve this problem. PDfilter is used for image feature representation, to make the image feature signature allow similar discriminating power among images to a detailed pixel by pixel comparison. IG focuses on image feature selection that holds most information about the analysis category, in order to identify the most useful data in the system training stage. Finally, the similarity measure tool is setup by a k-Nearest Neighbour (k-NN) classifier that provides a supervised machine-learning technique to allow the system to assign relevant keywords to the training and testing images.

A short review of CBIR and its challenges are given in Section 2. In Section 3, PDfilter and IG are described to solve the image feature selection problem. Section 4 presents the experiments and compares the performance of (1) k-NN, (2) IG (3) Colour histogram (4) PDfilter, and (5) PDfilter+IG for the image classification task over 190 vocabularies. Conclusion and future work are provided in Section 5.

2. RELATED WORK

Content-based Image Retrieval (CBIR) was proposed in the early 1990s and has been an active research area for over ten years. It is a technique based on the extracting image contents to retrieve relevant images, such as colour, texture, shape, etc. In particular, CBIR systems extract and index images' low-level features automatically, aiming to provide the capability to support visual queries, an intuitive query approach, and automatic description content features of images [4].

To extract image contents, image segmentation is one major task. The result of image segmentation provides more detailed information that describes an analysed image in terms of different areas or regions before low-level feature extraction [5]. In literature, complex segmentation schemes have been applied in current image retrieval systems. To segment an image, the optimal shape and size of sub-segments can be determined by different resolutions, which is an active area of research. Then low-level features can be extracted automatically from each of the sub-segments, including colour, texture, shape content, spatial object layout recognition, and so on.

Based on a combination of different low-level features to identify high-level concepts, image semantics can be found automatically using suitable rules and previous knowledge [6]. Additionally, images may be given descriptive metadata off-line manually, typically catalogue information, which is about the object's creator, event activity, time of creation, and so on.

The retrieval stage is followed by the image indexing stage. Various retrieval environments have been developed that enable users to query image databases through different kinds of retrieval strategy including Query by Keyword (QBK), Query by Example (QBE), and Query by Feature

*Address correspondence to this author at the Department of Information Management, National Central University, 300 Jhongda Rd., Jhongli, Taiwan; Tel: +886-3-422-7151; Fax: +886-3-4254604; E-mail: cftsai@mgt.ncu.edu.tw
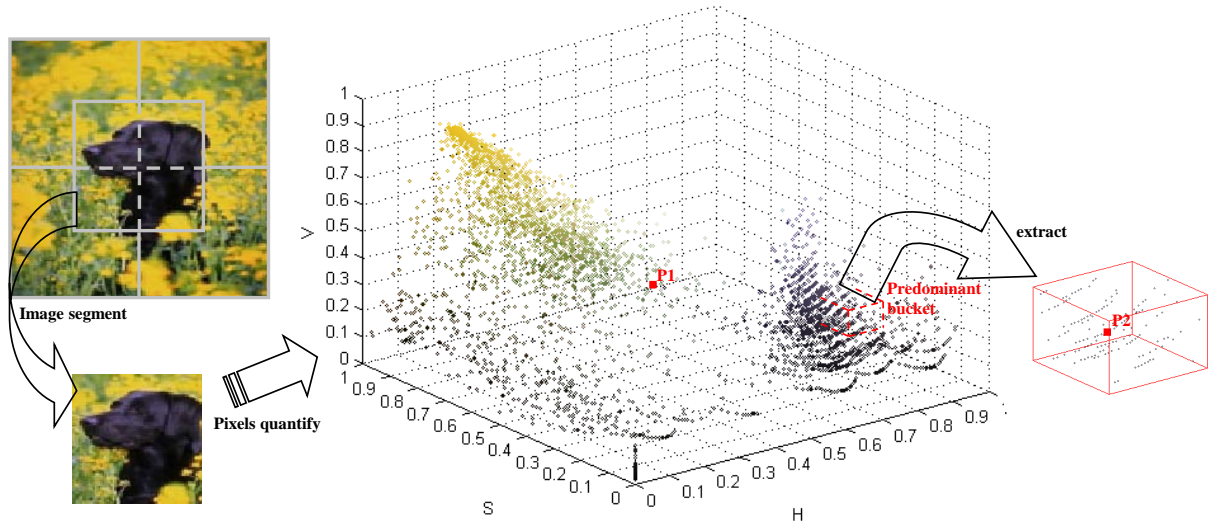
Fig. (1). Example of PDfilter operation[1].

(QBF) [1]. Visualisation tools inspect the initial set of images retrieved in response to the query, and improve the effectiveness of the system by filtering out the images which are not suitable for the query [7]. Relevance feedback is the process of analysing responses to retrieved information that permits query refinement to adjust retrieval repeatedly in order to match the user's original needs. User judgements of search results are fed back into the system to refine the query so that new results are more fitting.

The semantic gap problem is the main challenge for CBIR systems. It occurs in the translation of low-level features into high-level concepts. This problem causes some keywords never to be assigned to their correct images, and thus users are very likely un-satisfied by the retrieval results of current systems [8]. A number of investigations have worked towards deciphering this problem. Tsai *et al.* [6] have shown the possibility of using machine learning technology to automatically assign 150 controlled keywords to unknown and unseen images. Although the process is far from perfectly accurate it appears sufficiently effective for practical use, if combined with state-of-the-art browsing technology. However, ideally powerful image searching environments would like to operate with much larger vocabularies than at present, in order to enable current CBIR systems to be incorporated into commercial search engines, like Google and Yahoo, for general users to approach.

3. METHODOLOGIES

3.1. Pixel Density Filter (PDfilter)

Image feature values are derived from individual pixels initially. However, image's representative features cannot simply use every single characteristic, since every image is composed with very large numbers of individual pixels that will introduce more noise than considering an overall value for each region. In most related studies, such as Barnard *et*

al. [8] and Tsai [6], the representative feature comes from the average of all pixel values within a region or tiling area. This courses image's representative features can be too similar to each other. For example, regions composed of a mixture of red and green will be both represented in the same way, as brown colour. Particularly, the colour histogram [9] is the most extracted feature to solve this problem. It provides more discrimination power when more colour buckets are comprised [10]. Nevertheless, the number of buckets is usually limited by colour look-up on human vision. Therefore, there are only up to 256 colour buckets in current approach, and the majority is less than 128 buckets.

PDfilter is used for pixel value selection which can make representative features more similar to their values in the original image. In general, it may increase colour buckets. It supplies a better clustering model to perform with the individual feature space in each region. Every single pixel can be addressed into a bucket area with formula (1), where D means the dimensions out of N dimensions; ^{x}D is the value of each dimension and $^{x}D_m$ 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 (we set $S = 10$ to have 1,000 buckets for analyses). Finally, the representative feature values come from the average of the pixel values in the predominant bucket. An example with *hue* (H), *saturation* (S) and *value* (V) colour space [11] is shown in Fig. (**1**). The representative feature to the analysed region after PDfilter selection will be the symbol at point P2 (0.736, 0.25, 0.249), with black, in bucket 227, in contrast to the average of all pixel values at location P1 (0.389, 0.441, 0.418), with green.

$$A(p) = \sum_{D=1}^{N} \left(\left\lfloor \frac{^{x}D \times S}{^{x}D_m} \right\rfloor \times S^{D-1} \right) + 1 \tag{1}$$

3.2. Information Gain (IG)

Information Gain (IG) [3] is applied for training data selection, which allows each category to be represented by

---

[1]In order to show clearly the distribution of the pixel values, this example uses the tiling segmentation. However, the focus of this paper is to segment an image into regions with colour similarity.

its most important feature vectors with noise and uncertainty reduction. In addition, it is one of the most used feature selection method in text categorisation. Based on formula (2), IG has weighted sum of gain value to predict the presence or absence of a term in the analysed document.

$$G(t) = -\sum_{i=1}^{m} P(Ci) \log P(Ci)$$

$$+ P(t) \sum_{i=1}^{m} P(Ci \mid t) \log P(Ci \mid t)$$

$$+ P(\bar{t}) \sum_{i=1}^{m} P(Ci \mid \bar{t}) \log P(Ci \mid \bar{t}) \qquad (2)$$

In our approach, *k*-means clustering is used before performing IG, in order to sort original training feature vectors of the analysis category into a number of *k* clusters. We set *k* = 10 in the experiments since the feature vectors are too similar between each other. That is, without using PDfilter some documents cannot be grouped into more than 10 clusters. In formula (2), let *m* be the total number of regions in an image and $\{Ci\}_{i=1}^{m}$ is the set of images within the analysis category, then the gain value $G(t)$ to every individual cluster can be obtained. In addition, IG is intuitively appealing for uncertainty data reduction. There are no algorithms for the optimal splitting value identified. In practice, the threshold is based on heuristics to find a near-optimal value [12]. Only the documents with an above threshold gain value are re-

tained as new training data for future classification. Finally, the new training data to our experiments are taken from the single cluster with the highest gain value.

4. EXPERIMENTS

4.1. Task

We focus on feature value selection by PDfilter and IG, in order to recognise useful training data for the system. Our research questions are (1) can the PDfilter make a colour and/or texture feature signature to allow similar discriminating power among images as a detailed pixel by pixel comparison? (2) does IG provide a suitable model for colour and/or texture feature vector selection?

4.2. Experimental Set Up

For experiments, five different approaches are evaluated, which are (1) feature extractor, (2) PDfilter, (3) IG, (4) colour histogram and (5) similarity measure, as shown in Fig. (**2**). As a result, five independent systems (shown in Table **1**) were implemented, with different combinations of feature value calculation and training region selection. Other experimental systems are divided into three experiments by different image feature representations: (1) colour (2) texture and (3) colour+texture. For the colour feature, particularly, the colour histogram system is used for the comparisons, which is based on 125 colour buckets [13].
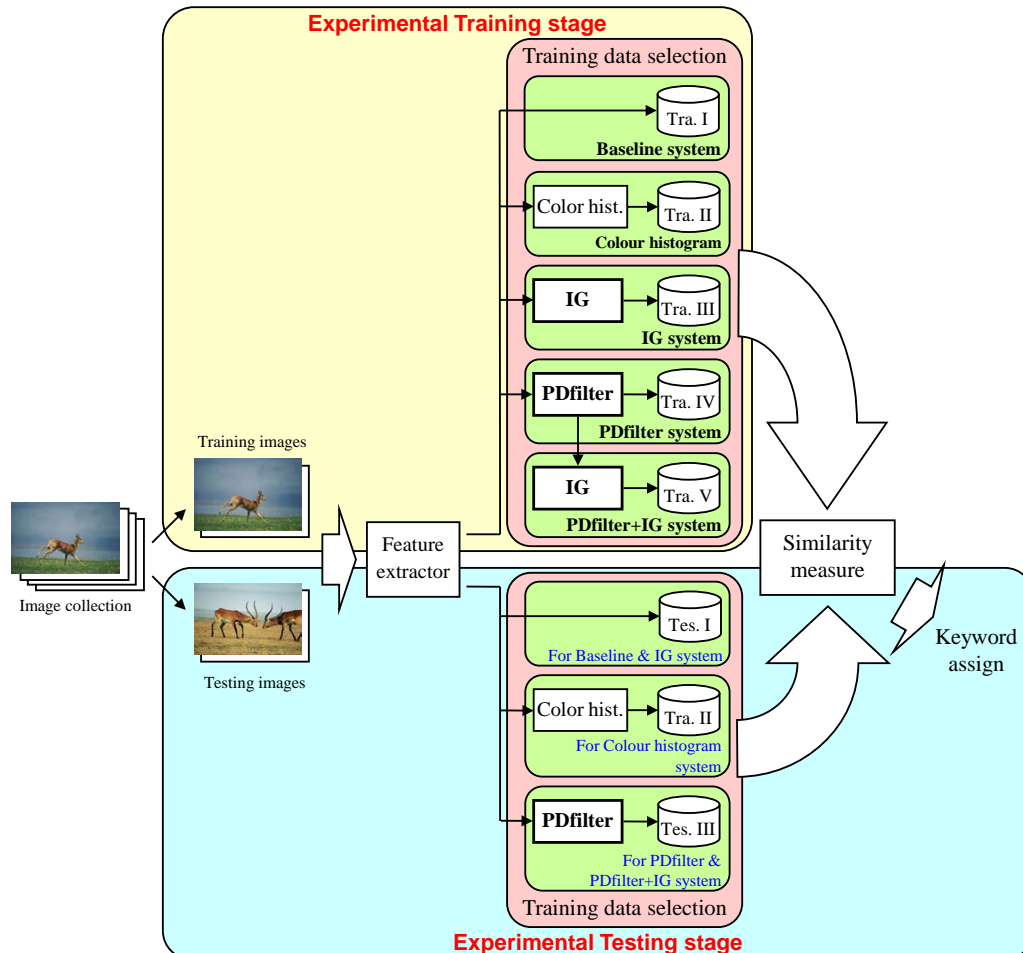

Fig. (2). Experimental framework.

Table 1.   Experimental Systems

	Feature Val. Calc.			Training Reg. Sel.	
	I	**II**	**III**	**A**	**B**
Baseline	×			×	
IG	×				×
Colour histogram		×		×	
PDfilter			×	×	
PDfilter+IG			×		×

I: average of all pixels' value within the region; **II**: average most colour in the region; **III**: average most predominant values in the region.
A: the region holding the central pixel (64, 64); **B**: the region selection by IG.

The image collection comes from the Corel Stock Photo Libraries[2], which is the most commonly used dataset for CBIR experiments. Corel use 100 images to describe a topic (category) and some of the descriptions are spread over more then one category. In this paper, we only consider one category to such descriptions, as we use one keyword to represent a single category. At the same time, the WordNet[3] lexical reference system is used to determine whether a keyword is concrete or abstract. The experiments only use concrete concepts and up to 190 keywords. Each individual keyword category provides 20 and 40 images for the training and test data respectively. Additionally, experiments were performed over 10, 50, 100, 150 and 190 keywords. The 10 to 150 keyword experiments consist of ten sub-experiments using random subsets of keywords selected from the 190 keyword set.

Feature extractor is used in automatic low-level feature extraction. There are three steps in this component. Firstly, images are resized into 128x128 pixels. Then each image is segmented by the Normalized Cuts (*N*cut) algorithm [14] into five regions based on their colour similarly. Finally, colour and texture features are extracted from these five sub-images (regions) respectively. The experiments described here map image features into 3 dimensions in HSV colour space and 16 dimensions in four levels of Daubechies' wavelet texture feature decomposition [15].

Different components were implemented in every independent system to have different combinations of the training and testing data. However, the *k*-NN classifier supplies a similarity measure between the training and testing data. According to Jain *et al.* [16], we set $k = 1$ (1NN), which can provide reasonable classification results for most pattern recognition problems. It assigns relevant keywords into new instances, in order to provide an effective query method through keyword-based retrieval.

4.3. Evaluation and Discussion

F-measure, as weighted harmonic mean of recall and precision, is used for system centred evaluation in this paper.

Fig. (**3**) shows the experimental results using various feature selection techniques for the colour feature. The results are improved after IG selection. After PDfilter, both

systems are better then the Baseline and Colour histogram. In particular, for PDfilter+IG, the F-measure increases to 6.33% from 4.13% provided by the baseline over 190 categories. That is, the PDfilter and IG increase the precision of assigning keywords to their related images for colour features.
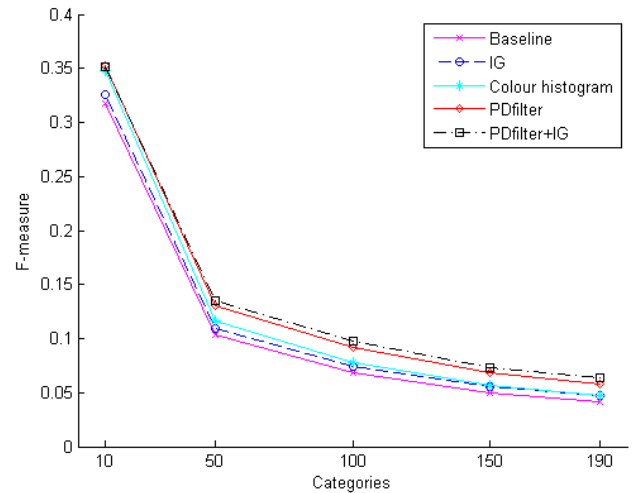

Fig. (3). Experimental results for the colour feature approach, over 10, 50, 100, 150 and 190 categories.

Table **2** shows that only IG improves the annotation performance with texture features (very slightly). Additionally, the results show that PDfilter and IG do not enhance the annotation performance by the combined colour and texture features.

Table 2.   Experimental Results of Texture and Colour+Texture Approach, in 190 Categories Only

F-measure	Baseline	IG	PDfilter	PDfilter+IG
texture	4.70%	4.83%	3.76%	3.65%
colour+texture	7.10%	6.88%	5.65%	5.33%

5. CONCLUSION AND FUTURE WORK

PDfilter and IG are two promising approaches to image annotation using colour features. For the texture features, every pixel's feature value is computed by the relationship with its neighbouring pixels, a relationship obscured by PDfilter and IG. This is the main reason why the systems cannot improve the performance by texture and colour+texture features.

In future we will expand the investigation by directly extracting texture features, attempting to apply IG between each pixel and its neighbours, to improve precision. We will also work with larger vocabularies, including abstract keywords. In addition, other image data sets, such as the IAPR TC-12 Benchmark [17], the TRECVid [18], and the University of Washington image collection[4], will be used for system evaluation.

REFERENCES

[1]   A. del Bimbo, *Visual Information Retrieval*. San Francisco: Morgan Kaufmann, 1999.

[2]   L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. California: Wadsworth & Brooks, 1984.

[3]   Y. Yang and J. O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," in *14th International Conference on Machine Learning (ICML '97)*, 1997, pp. 412-420.

[4]   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, pp. 1349-1380, Dec. 2000.

[5]   F. Idris and S. Panchanathan, "Review of Image and Video Indexing Techniques," *Visual Communication and Image Representation*, vol. 8, pp. 146-166, Jun. 1997.

[6]   C.-F. Tsai, K. McGarry, and J. Tait, "CLAIRE: A modular support vector image indexing and classification system," *ACM Transactions on Information Systems (TOIS)*, vol. 24, pp. 353-379, Jul. 2006.

[7]   J. K. Wu, M. S. Kankanhalli, J. Lim, and D. Hong, *Perspectives on Content-based Multimedia Systems*. London, Boston, Dordrecht: Kluwer Academic Publishers, 2000.

[8]   K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *Journal of Machine Learning Research*, vol. 3, pp. 1107-1135, Feb. 2003.

[9]   M. J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, pp. 11-32, Nov. 1991.

[10]   F. Long, H. Zhang, and D. D. Feng, "Fundamentals of Content-Based Image Retrieval," in *In Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, D. D. Feng, W. C. Siu, and H. Zhang, Eds. Berlin, Heidelberg, New York: Springer-Verlag, 2003.

[11]   F. van der Heijden, *Image Based Measurement Systems: Object Recognition and Parameter Estimation*. Chichester: John Wiley & Sons, 1994.

[12]   C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. London, Cambridge: MIT, 1999.

[13]   T. Lai, *CHROMA: A Photographic Image Retrieval System*, UK: University of Sunderland, 2000.

[14]   J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888-905, 2000.

[15]   I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics, 1992.

[16]   A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transitions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4-37, Jan. 2000.

[17]   M. Grubinger, P. Clough, H. Müller, and T. Deselears, "The IAPR TC-12 Benchmark: A New Evaluation Resource for Visual Information System," in *International Workshop OntoImage '06 Language Resources for Content-Based Image Retrieval*, Genoa, Italy, 2006, pp. 13-23.

[18]   A. F. Smeaton, W. Kraaij, and P. Over, "TREC Video Retrieval Evaluation: A Case Study and Status Report," in *RIAO 2004: Coupling Approaches, Coupling Media and Coupling Languages for Information Retrieval*, Avignon, France, 2004, pp. 25-37.