Context Based Image Retrieval using Image Features

T.Bagyammal
Lecturer, Amrita School of Engineering
Amrita Vishwa Vidyapeetham, Ettimadai
India
t_bagyammal@cb.amrita.edu

Latha Parameswaran
Professor, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Ettimadai,
India
p_latha@cb.amrita.edu

Abstract -- Context based image retrieval is useful to effectively retrieve images based on its context from the massive collection available on the web. Content based image retrieval system is effective only to recognize and retrieve specific objects. It fails when the image has multiple objects and the objects are cluttered. In this work an algorithm has been proposed which facilitates the search based on the context,"arrangement of an object with respect to other objects" in a given image. Image containing the objects with its context is chosen as input, images which have the chosen object with the similar context have been retrieved from the dataset. Supervised segmentation and labelling of an image has been done. Features such as Histogram Orientation of Gradients (HOG), Scale Invariant Feature Transform (SIFT), 1-D signature, Gray Level Co-occurrence Matrix (GLCM) texture properties, Statistical properties (mean, standard deviation), Speeded Up Robust Features (SURF) and Colour Histogram are extracted from the image. Given a training set of images, context is learnt using the minimum distance classifier based on the extracted features. Study of this algorithm on the Kitchen Scene in the Scene UNderstanding (SUN 09) data set has been done. Image retrieval is done based on all the set of features extracted from the images. Context based image retrieval system has been widely used in object recognition, person identification within the context of personal photo collection and image annotation.

Key words – Context based image retrieval, image retrieval, context, minimum distance classifier, visual features.

1. INTRODUCTION

With advances in the Internet and new digital image sensor technologies, the volume of digital images produced by scientific, educational, medical, industrial, and other applications available to users has increased dramatically [1]. The difficulties faced by text-based retrieval became more and more severe [2]. The efficient management of the rapidly escalating visual information became an important problem. Content based image retrieval (CBIR) uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. While content based image retrieval is generally effective, it may not achieve a reliable search result in cases where the visual information extracted from the small object region are unable to reliably reveal the search intent of the user. In [15] this uncertainty has been handled by taking into account the fact that objects in real-life images hardly occur in isolation. So the visual information outside the region of interest can be seen as the context in which region of interest (ROI) is specified as a search query.

This work proposes a Context-based image retrieval model that effectively employs the visual context information together with the ROI. In typical context-based image retrieval systems, the context (visual contents) of the images in the database are extracted and described by multidimensional feature vectors. In [1] feature database is constructed from the feature vectors of the images in the database. In [3] to retrieve images, users provide the retrieval system with the query image that contains object with its context or sketched figures. These examples are then represented as a feature vector by the system. The similarities between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed. Recent retrieval systems have integrated users' relevance feedback to modify the retrieval process in order to generate perceptually and semantically more meaningful retrieval results. Figure 1 shows the system diagram of context based image retrieval system.

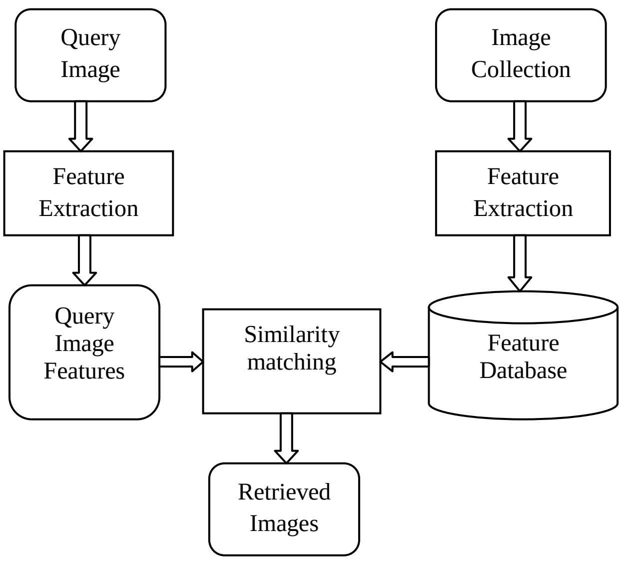

Figure 1 Image Retrieval system

Context based image retrieval is a new area of research. There are many works in literature based on using the context of the object in an image for object identification and object categorization in computer vision domain. Context of an object may be learnt from the visual content of the image or from the text description (Global Positioning System [GPS] information, date and time information of an image) of an image.

2. Background Study

In the task of visual object categorization, semantic context can play very important role of reducing ambiguity in objects visual appearance. In [4] for object categorization, objects category labels has been assigned with respect to other objects in the scene, assuming there is more than one object present. Semantic object context has been incorporated as a post-processing step into any off-the-shelf object categorization model. Conditional random field (CRF) framework has been used to maximize object label agreement according to contextual relevance. Results have revealed that incorporating context into object categorization greatly improves categorization accuracy.

Identifying people in photos is an important need for users of photo management systems. In [6] MediAssist has been used, which facilitates browsing, searching and semi-automatic annotation of personal photos, using analysis of both image content and the context in which the photo has been captured. This semiautomatic annotation includes annotation of the identity of people in photos. In this work, person identification techniques based on a combination of context and content has been used. Language modelling and nearest neighbour approaches to context-based person identification was proposed, in addition to novel face colour and image colour content-based features (used along with face recognition and body patch features).A comprehensive empirical study of these techniques using the real private photo collections of a number of users have been conducted, and the result showed that combining context- and content-based analysis improves performance over content or context alone. Most image annotation systems consider a single photo at a time and label photos individually. In [7] the authors focused on collections of personal photos and utilize the contextual information naturally implied by the associated GPS and time metadata. First a constrained clustering method has been employed to partition a photo collection into event-based sub collections, considering that the GPS records may be partly missing. They then use conditional random field (CRF) models to exploit the correlation between photos based on 1) time-location constraint and 2) the relationship between collection-level annotation and image-level annotation. With the introduction of such a multilevel annotation hierarchy, they have addressed the problem of annotating consumer photo collections that requires a more hierarchical description of the customers' activities than do the simpler image annotation tasks.

Conventional approaches of image indexing and retrieval from digital libraries include content-based, metadata-based, and keyword-based approaches. In [8] the authors addressed a different way of image retrieval motivated by real-life applications for an intelligent system that can automatically select appropriate background images from textual passages. A technique for developing automatic image-retrieval systems based on essential contextual information of a textual passage was explored. A framework that applies semantic role labelling techniques and a commonsense knowledge base, ConceptNet was proposed. The primitive results indicate that the proposed methodology has a potential on applications with textual passages that describe things and events that are regularly seen in everyday life.

In [5] the authors have described the issues in a model based on relevance feedback for retrieval, which is based on the Rocchio's algorithm. In this the user issues a (short, simple) query. The system returns an initial set of retrieval results. The user marks some returned documents as relevant or irrelevant. The system computes a better representation of the information need based on the user feedback. The system displays a revised set of retrieval results. Relevance feedback can go through one or more iterations of this sort.

Features extracted for image retrieval includes SIFT, SURF, 1-D signature, GLCM texture feature, colour histogram and statistical features (mean, standard deviation). SIFT algorithm [10] takes an image and transforms it into a collection of local feature vectors. Each of these feature vectors is supposed to be distinctive and invariant to any scaling, rotation or translation of the image. SIFT can robustly identify objects even among clutter and under partial occlusion, because the SIFT feature descriptor is invariant to uniform scaling, orientation, and partially invariant to affine distortion and illumination changes.

In [11] The Speeded-Up Robust Features algorithm, scale and rotation invariant interest point detector and descriptor which are computationally very fast is discussed.1-D Signature [17] is a 1-D representation of a 2-D object boundary. One of the simplest is to plot the distance from the centroid to the boundary as a function of angle. Feature vector includes the fourier transform of this 1-d representation. The shape of the object is best represented by this feature. GLCM [18] is a tabulation of how often different combinations of pixel grey levels co-occur in an image. GLCM has been used for second order texture calculations.

Second order measures consider the relationship between groups of two (usually neighbouring) pixels in the original image. The features extracted using GLCM are Contrast, Energy, Correlation and Homogeneity.

Colour histogram is supplemented with other features in the literature for image retrieval to increase the efficiency and recall rate of the algorithm. Feature vector comprises of 64 bin colour histogram calculated from the given RGB image.

3. Proposed Method

The proposed algorithm aims at retrieving images from the database based on the context in the given query image. Supervised segmentation and labelling of an image has been done to identify the objects in the image. Given a set of images for training, learning has been done based on the minimum distance classifier. Features that have been considered for classification and image retrieval are

1. Scale Invariant Feature Transform (SIFT)
2. Histogram of Oriented Gradients (HOG).
3. Speeded Up Robust Features (SURF)
4. Color Histogram
5. Statistical features (mean, standard deviation)
6. Gray Level Co-occurrence Matrix (GLCM)
7. 1-D signature

This proposed algorithm has two major steps a. Object Recognition b. Image Retrieval. Object Recognition is done by using the minimum distance classifier which has i) Training Phase ii) Testing Phase.

3.1 Object Recognition

A. Training Phase

In the proposed work training has been done using minimum distance classifier by using all the features.The steps involved are given in the Figure 2.

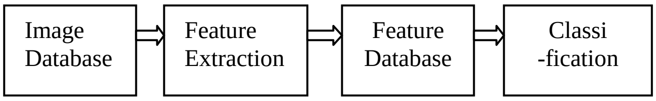

Figure 2 Steps of Training phase

B. Minimum Distance Classifier

The minimum distance classifier [13] is used to classify unknown image data to classes which minimize the distance between the image data and the class in multi-feature space.The Minimum Distance algorithm is a supervised classification method which first analyses the training data and calculates a mean vector for each prototype class, described by the class or cluster centre coordinates in feature space. The Minimum distance algorithm then determines the absolute distance from each unclassified feature to the mean vector for each prototype class and assigns that feature to the closest class. Features are assigned to different classes based on their closeness to the class mean vectors i.e.

$$x \in \omega_i \ if \ d(x, z_i) = \min d\left(x, z_j\right.$$

Where,

$$d(x_i,x_j) = \sum_{j-1}^{n} abs\,|(f_j(x) - f_j(i))|$$

$f_j(x)$ - Feature of unknown object j

$f_j(i)$ - Feature of known i^{th} object

n - Number of Features

All the required parameters for classification are learnt during training phase.

C. Testing Phase

In the testing phase, an image is given as an input. Once the features are extracted, minimum distance classifier has been applied and the classification is done. Based on the class to which the identified context has maximum probability, the corresponding class has been assigned for the image. The steps involved in the Testing phase are as shown in Figure 3.

Figure 3 Steps of Testing phase

3.2 Image Retrieval

Scene image is given as an input with the object of interest marked in it. Objects are selected by marking the top left and bottom right corner on each of the object. Supervised segmentation of the object is done. Classification of the object has been done using the proposed algorithm. Images in the database, which contain the selected objects, have been ranked based on the similarity between the objects in the query image. Top-k images from the dataset which has the chosen objects have been retrieved from the database.

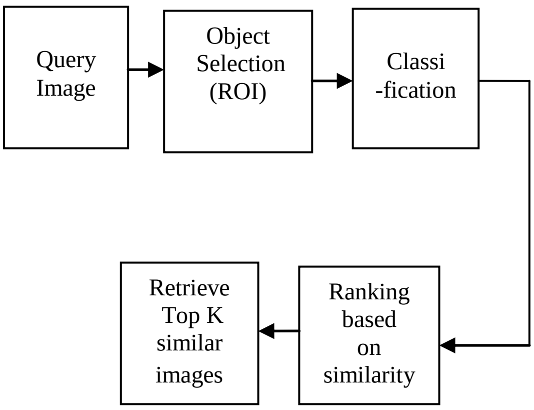

Figure 4 Image Retrieval

4. Dataset

The dataset used in this work is SUN (Scene UNderstanding) 09 dataset containing 908 Scene categories (from Abbay to Zoo) which contains 3819 Object categories. Proposed algorithm is experimented on Kitchen Scene with 5 Object categories consisting of 100 images of kitchen scene from Scene UNderstanding (SUN 09). For training a classifier 16 images from each object category (16*5=80) and for testing 25 images from different class have been considered. Context between the objects has been learnt using the features. Objects that have been considered for classification are Microwave oven, Mug, Shelf, Stove and Tap.

5. Results and Discussion

Images in the database have been trained for five different object categories Microwave oven, Mug, Shelf, Stove and Tap. During testing a kitchen scene image is taken and a supervised segmentation has been done, then object classification is done using the minimum distance classifier by using the features SIFT, HOG, SURF, Color Histogram, GLCM, 1-D Signature, mean and standard deviation.

5.1 Object Recognition

Object classification using high level features (SIFT, SURF, HOG, Color Histogram), texture features (GLCM) and 1-D Signature is given as a confusion matrix in Table 1,2,3,4,5 and 6. From the confusion matrix shown in Table 1 it is understood that misclassification rate for Mug, Shelf, Tap and Stove is 20% and for Microwave oven it is 0% by using 1-D Signature. The estimated misclassification rate for Microwave oven, Mug, Shelf, Tap and Stove is 20% by using the different combination of high level features SIFT, SURF,1-D signature, colour histogram and HOG feature is shown in table 2,3,4 and 5. The estimated misclassification rate using all the features is depicted in table 6 and it is 20% for Microwave oven, Mug and Stove and 0% for Microwave oven and Mug.

Table 1-Confusion matrix using the feature 1-D Signature

1-D Signature					
	Microwave oven	Mug	Shelf	Tap	Stove
Microwave oven	100	0	0	0	0
Mug	0	80	0	0	20
Shelf	0	0	80	0	20
Tap	20	0	0	80	0
Stove	0	0	20	0	80

Table 2-Confusion matrix using SIFT and HOG features

HOG and SIFT features					
	Microwave oven	Mug	Shelf	Tap	Stove
Microwave oven	80	0	20	0	0
Mug	0	80	0	20	0
Shelf	20	0	80	0	0
Tap	20	0	0	80	0
Stove	0	0	20	0	80

Table 3-Confusion matrix using SIFT, HOG and SURF features

HOG,SIFT and SURF features					
	Microwave oven	Mug	Shelf	Tap	Stove
Microwave oven	80	0	20	0	0
Mug	0	80	0	20	0
Shelf	20	0	80	0	0
Tap	20	0	0	80	0
Stove	0	0	20	0	80

Table 4-Confusion matrix using SIFT, HOG, SURF and Colour Histogram

HOG, SIFT, SURF, Colour Histogram features					
	Microwave oven	Mug	Shelf	Tap	Stove
Microwave oven	80	0	20	0	0
Mug	0	80	0	20	0
Shelf	20	0	80	0	0
Tap	20	0	0	80	0
Stove	0	0	20	0	80

Table 5-Confusion matrix using SIFT, HOG, SURF, Colour Histogram and 1-D Signature

HOG,SIFT,SURF,Color Histogran,1-D Signature					
	Microwave oven	Mug	Shelf	Tap	Stove
Microwave oven	80	0	20	0	0
Mug	0	80	0	20	0
Shelf	20	0	80	0	0
Tap	20	0	0	80	0
Stove	0	0	20	0	80

Table 6- Confusion matrix using SIFT, HOG, SURF, Colour Histogram, GLCM and 1-D Signature

HOG,SIFT,SURF,Colour Histogram ,1-D Signature and GLCM					
	Microwave oven	Mug	Shelf	Tap	Stove
Microwave oven	80	0	0	20	0
Mug	0	80	20	0	0
Shelf	0	0	100	0	0
Tap	0	0	0	100	0
Stove	0	0	0	20	80

5.2 Context Based Image Retrieval

The classification results have improved when all the features are considered. Hence all the features are considered for image retrieval. Input image for the image retrieval algorithm is shown in Fig. 5 with object of interest marked in red colour region (object of interest are Shelf and Microwave oven).Top 3 output images is given in Fig 6.a, Fig 6.b and Fig 6.c which contains the similar objects that are all marked in the query image. It is observed that the shape descriptive feature is important for image retrieval.

Figure 5 Image Retrieval- Sample Input Image (chosen objects are highlighted in a red colour region)

Figure 6.a

Figure 6.b

Figure 6.c

Figure 6.a,6.b,6.c Image Retrieval - Top 3 Output Images containing the object or interest

6. CONCLUSION

A technique for context based image retrieval system has been proposed in this work. Supervised segmentation and labelling of images is done. Context has been generated based on the similarity between the features learnt by using Minimum distance classifier. During testing, query image has been considered and the GLCM texture feature, Histogram Orientation of Gradients (HOG), Scale Invariant Feature Transform (SIFT), 1D signature, Gray Level Co-occurrence Matrix (GLCM) texture properties, Statistical properties(mean, standard deviation), Speeded Up Robust Features (SURF) and Colour Histogram features are extracted and then features are then compared for similarity with the classes. The class which gives better matching would be ranked and the top k images are retrieved corresponding to the class of the query.

During testing 1D signature gives good classification result and when it is combined with the high level features and texture features classification results have improved.

This work can be extended to include semantic context (co-occurrence of an object with other object) and spatial context (left, right, top and bottom). The above mentioned Context can be incorporated using Conditional Random Field framework.

ACKNOWLEDGEMENT

We thank Amrita Vishwa Vidyapeetham for having provided the required resources in the Amrita-Cognizant Innovation Lab for carrying out the research work.

REFERENCES

[1] Showkat AhmadDar, Zahid Gulzar Khak, "Content Based Image Retrieval", IOSR Journal of Computer Engineering (IOSR-JCE), Vol 12, Issue 2(May. -Jun. 2013), PP 87-92.

[2]Zakariya S.M, Aligarh, Ali R, Ahmad N, "Combining visual features of an image at different precision value of unsupervised content based image retrieval", Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on 28-29 Dec. 2010, pp. 1 – 4.

[3] Latika Pinjarkar, Manisha Sharma, Kamal Mehta "Relevance Feedback Techniques in Content Based Image Retrieval: A Survey", International Journal of Engineering Research and Development, Vol 1, Issue 9 (June 2012), PP.19-23.

[4] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Be-longie, "Objects in Context", in Proceedings of 11th IEEE International Conference Computer Vision (ICCV '07), Oct. 2007, pp. 1–8.

[5] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review", Multimedia Systems, vol. 8, no. 6, pp. 536–544,Apr. 2003.

[6] Neil O' Hare and Alan F. Smeaton "Context-Aware Person Identification in Personal Photo Collections", IEEE Transactions on Multimedia, Vol, 11, No.2, Feb'09, pp.220-228.

[**7**] Liangliang Cao,Jiebo Luo,Henry Kautz and Thomas S.Huang, "Image Annotation Within the Context of Personal Photo Collections Using Hierarchical Event and Scene Models", IEEE Transactions on Multimedia, Vol, No. 2, Feb'09, pp.208-219.

[**8**] Sheng-Hao Hung, Pai-Hsun Chen,Jen-Shin Hong, and Samuel Cruz-Lara, "Context –based image retrieval: A case study in background image access for multimedia presentations", IADIS International Conference WWW/Internet 2007, Vila Real : Portugal (2007).

[9] Dalal N and Triggs B, "Histograms of Oriented Gradients for Human Detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, San Diego, CA, USA.

[10] David G. Lowe, "Distinctive image feature from scale invariant keypoints", International Journal of Computer Vision, 60, 2 (2004), pp. 91-

110.

[11] Edouard Oyallon, Julien Rabin,"An analysis and implementation of the SURF method, and its comparison to SIFT", IPOL Journal Image Processing On Line, Feb 2013.

[12] Dalal N. France,"Histograms of oriented gradients for human detection ", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 2005 (Vol 1), pp.886 – 893.

[13] Kanika Kalra, Anil kumar Goswami, Rhythm Gupta,"A Comparative Study of Supervised Image Classification Algorithms for Satellite Images", International Journal of Electrical, Electronics and Data Communication, Vol-1, Issue-10, Dec-2013 pp. 10-16.

[14] Myung Jin Choi, Antonio Torralba, and Alan S. Willsky," A Tree-Based Context Model for Object Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, VoL. 34, 2012.

[15] Linjun Yang, Bo Geng, Yang Cai, Alan Hanjalic, Xian-Sheng Hua, "Object Retrieval Using Visual Query Context", IEEE Transactions On Multimedia, Vol. 13, NO. 6, Dec-2011.

[16] Jianxiong Xiao, Hays J, Ehinger K.A, Oliva A, Torralba A," SUN Database: Large-scale Scene Recognition from Abbey to Zoo", Conference on Computer Vision and Pattern Recognition (CVPR) 2010, San Francisco,CA, pp. 3485 – 3492.

[17] Rafael C. Gonzalez, Richard E.Woods, *Digital Image Processings*, India: Pearson Prentice Hall, 2009.

[18] Haralick, R.M., K. Shanmugan, and I. Dinstein, "Textural Features for Image Classification", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-3, 1973, pp. 610-621.

Authors Profile

Bagyammal T. working as a Lecturer at department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham.Received the B.E degree in Information Technology from Bharathiar University in the year 2004, the M.Tech degree in Computer Vision and Image Processing in the year 2014 from Amrita Vishwa vidyapeetham and the research interests include Image Processing, Computer Vision and Image Mining.

Dr. Latha Parameswaran working as a Professor and Chairperson in the Department of Computer Science and Engineering, Amrita School of Engineering,Amrita Vishwa Vidyapeetham. She completed her Master's Degree in Computer Applications from PSG College of Technology. Her Ph. D. is from Bharathiar University. Her areas of research include Image Processing, Image Mining, Information Security and Theoretical Computer Science.