Automatic Image Annotation Using SURF Descriptors

Muhammed Anees V
Department of Computer Science
Cochin University of
Science and Technology,
Cochin, India 682022
Email: mohdaneesv@gmail.com

G. Santhosh Kumar
Department of Computer Science
Cochin University of
Science and Technology,
Cochin, India 682022
Email: san@cusat.ac.in

Sreeraj. M
Department of Computer Science
Cochin University of
Science and Technology,
Cochin, India 682022
Email: sreerajtkzy@gmail.com

***Abstract*—In recent years there is an apparent shift in research from content based image retrieval (CBIR) to automatic image annotation in order to bridge the gap between low level features and high level semantics of images. Automatic Image Annotation (AIA) techniques facilitate extraction of high level semantic concepts from images by machine learning techniques. Many AIA techniques use feature analysis as the first step to identify the objects in the image. However, the high dimensional image features make the performance of the system worse. This paper describes and evaluates an automatic image annotation framework which uses SURF descriptors to select right number of features and right features for annotation. The proposed framework uses a hybrid approach in which k-means clustering is used in the training phase and fuzzy K-NN classification in the annotation phase. The performance of the system is evaluated using standard metrics.**

keywords- Automatic Image Annotation, SURF feature extraction, Image classification, K-means clustering, Fuzzy K-NN.

I. Introduction

The availability of modern and economical image capturing devices increase the use of digital images in recent years. Despite of the extensive research in this field, correct retrieval of image from large collection remain a challenging task. This is due to the difficulty in mapping of semantic content of the image as perceived by humans. The process of assigning meaningful textual descriptions to an image based on its content is known as automatic image annotation (AIA). Due to the higher quantity of visual digital content, the modeling of multimedia and especially the semantic gap between the low level visual features and high level semantic concepts become an important domain [3]. The problem of AIA and its applications become more relevant in image databases and social networking websites. In traditional method each image is tagged manually with suitable keywords to search and retrieve images efficiently. It is a tedious work and needs huge amount of man power and time.

Content Based Image Retrieval (CBIR) system was proposed in early $1990'$s to organize and search these images efficiently to overcome the difficulties of traditional image retrieval methods by matching the low level features of images [1]. User queries are used to retrieve images in traditional CBIR systems. In this study, An AIA system is proposed which assigns suitable key words to digital images depending on the information containing it. Automatic image annotation can be extended to an image retrieval system called annotation based image retrieval(ABIR).

Speed-up Robust Feature Extractor (SURF) [2] is used to extract fixed number of key points from the training images. SURF generates the descriptors of each key points which is extracted from the training images. Each key point is represented by SURF using its coordinates, scaling, orientation and 64 by 1 descriptors. These 64 by 1 descriptor matrix is used to annotate the image. Fixed size feature matrix is converted in to a feature vector using probability density function and Rayleigh estimation [24]. N number of training images are clustered in to K number of clusters using k-means [4] clustering. Class names are assigned to each clusters to create an annotation database. A model is created from the feature vectors, and cluster labels and this model is used to annotate a test image in the annotating phase. Features are extracted from test images using the same SURF extraction method used in the training phase and test images are classified using the fuzzy K-nearest neighbour (Fuzzy K-NN) algorithm. According to the class label assigned by the fuzzy K-NN algorithm, class names are retrieved from annotation database and displayed on the image.

The rest of this paper is structured as follows. In section 2, related work is reviewed. The proposed system framework and each component in detail is discussed in Section 3. The result and performance analysis is described in Section 4. Finally, Conclusions and future works are discussed in Section 5.

II. Related Work

Many techniques are proposed for automatic image annotation system. Mori et al. [6] developed a co-occurrence model, in which the co-occurrence tags corresponding to the images are computed. However, the model tends to require large numbers of training samples to estimate the correct probability. Moreover, it tends to map frequent words to every possible images. P. Duygulu et al. [7] proposed a machine translation system which improves the performance of co-occurrence model which uses a vocabulary of blobs to annotate

an image. Blie and Jordan [8] proposed the correspondence latent Dirichlet allocation (Corr-LDA) model to find a conditional relationship between image features and textual features. Monay and Gatica-Perez [9] used Latent Semantic Analysis (LSA) and Probabilistic Latent Semantic Analysis (PLSA) for image annotation. Cross media relevance model (CMRM) was proposed by Jeon et al. [10] which uses the joint distribution of image regions and set of keywords. Cross media relevance model was later improved by Lavrenko et al. [11], who introduced Continues space Relevance Model (CRM). Cross media relevance model was also improved by S. L. Feng et al. [12] by introducing Multiple Bernoulli Relevance Model (MBRM).

There exists many approaches to implement an AIA system. Based on the portion of the image used to extract the annotation process is classified in to segmental approach and holistic approach. Segmental approach considers the image as a combination of semantically meaningful parts. Images are segmented or parts are taken from the image and features are extracted from these parts as described in [7]. Holistic approach considers the image as a whole. Features are extracted from the whole image and a relation is explored directly between the image and the annotation words [13].

It is also possible to classify the annotation process based on the features extracted. Color Features, Texture Features, Scene Features and Scale and rotation invariant Features being the features used in general. Color features have been widely used feature for image annotation. Image annotation based on colour features are described in [14]. Texture features are another important feature used in image annotation systems. The term $texture$ generally refers to the presence of a spatial pattern that has some properties of homogeneity. Gabor filter [15] is a method used for extracting texture features. Scale and rotation invariant features are recently used for implementing image annotation systems. Scale-Invariant Feature Transform (SIFT) [16] and Speeded Up Robust Feature (SURF) [2] are most commonly used Scale and rotation invariant feature extraction method. All related works on automatical image annotation is explained in detail by Dengsheng Zhang [25]

III. Proposed System

The proposed architecture of the automatic image annotation system is shown in Figure 1 .

The proposed framework is broadly divided into two phases. They are training phase and annotation phase. Training phase contains 3 stages. First step in training phase is SURF [2] feature extraction from training images, which is used to extract the features from it. clustering the training images is done by K-Means [4] clustering in step 2, which clustered the training images in to k number of clusters. Suitable labels are assigned to each clusters depend upon the general behaviour of the clusters. These extracted features along with their cluster label is used to generate a the model for training. This generated model is used to annotate the test images in the annotation phase

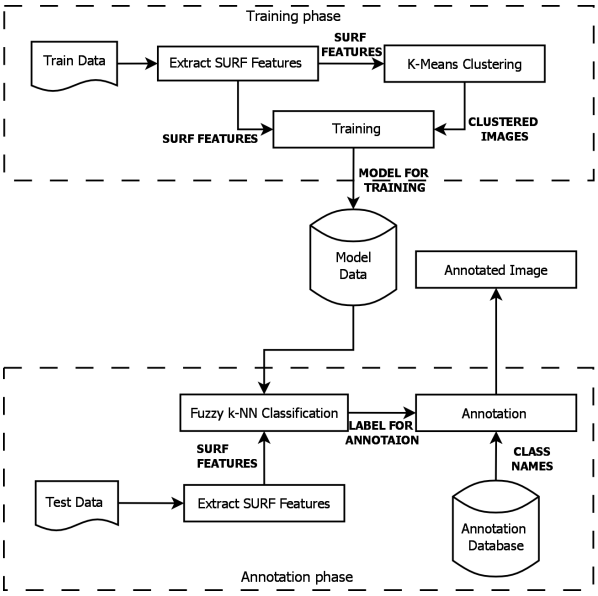

Fig. 1. Architecture of the proposed system

SURF feature extraction from test images is the first step in annotation phase. In second step classify each test images using Fuzzy K-NN algorithm based on the model created in the training phase. Annotation of the test images is done in third step using the class label assigned by the fuzzy K-NN algorithm. Cluster names corresponding to cluster labels are retrieved form the annotation database and displayed on the image. The architecture of the proposed automatic image annotation system is described in detail below.

A. Training Phase

1) SURF Feature Extraction: Feature extraction is the first and important stage of any classification and annotation problem. The proposed architecture uses speeded up robust feature extraction method (SURF) to extract the features of both training and testing images. SURF extraction method is a scale and rotation invariant feature extraction method, which is faster than widely used feature extracting method scale invariant feature transform (SIFT) [16] [2]. SURF is used in this proposed automatic image annotation system due to this higher performance over SIFT [2]. SURF focuses on scale and in-plane rotation invariant detectors and descriptors of an image. Integral image is calculated from the image and calculate sum of pixel intensities in the integral image[17] using the equation 1

$$I_{\Sigma}(x,y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i,j) \tag{1}$$

The locations in the image where the determinant of Hessian matrix [18] is maximum are detected and the matrix is calculated by the equation 2. Pixel intensities are high where the determinant of Hessian matrix is maximum, so determinant of Hessian matrix gives the maximum intensity points in an image. The features of these maximum intensity points

are extracted to implement the proposed automatic image annotation system.

$$H(x,\sigma) = \begin{bmatrix} L_{x,x}(x,\sigma) & L_{x,y}(x,\sigma) \\ L_{x,y}(x,\sigma) & L_{y,y}(x,\sigma) \end{bmatrix} \quad (2)$$

where $L_{x,x}(x,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\delta^2}{\delta x^2}g(\sigma)$ with the image I in point X. Scale space is a continuous function which is used to find extremes across all possible scales. This step is known as scale space representation in which images is scaled in to all possible scales and find the high intensity points in all scales. Interest points below the threshold are eliminated and remains only the strongest points in interest point localization step. The selected points show high intensity across all possible scales.

Now SURF describes the features of the selected points. SURF selects N number of high intensity points in an image. SURF describes the distribution of the intensity content within the interest point neighbourhood. These descriptors are build on distribution of first order Haar wavelet responses in x and y direction and create 64 by 1 descriptor matrix is created for each interest points detected. A feature matrix is created for whole image from these individual matrices. The feature matrix is then converted in to a feature vector using probability distribution function and Rayleigh estimation [24].

2) K-Means Clustering: Now N number of rows are used to represent N number of training images. N feature vectors are clustered in to k-clusters using K-means clustering. k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. Given a set of observations $(x_1, x_2, ..., x_n)$, where each observation is a d-dimensional real vector, k-means clustering aims at partitioning the n observations into k sets $(k \leq n)S = S_1, S_2, ..., S_k$ so as to minimize the within cluster sum of squares. The basic idea of this interactive algorithm is to assign the feature vector to the cluster such that the sum of squared error E is minimum.

$$E = \sum_{i=1}^{k} \sum_{j=1}^{N_j} \|x_{ij} - \mu_i\|^2 \quad (3)$$

where x_{ij} is the j^{th} point in the i^{th} cluster, μ_i is the mean vector of i^{th} cluster and N_j is the number of patterns in the i^{th} cluster.

3) Model Data Generation: The feature vectors corresponding to the N number of images are extracted using SURF feature extraction, and k number of clusters. N number of images are clustered in to k clusters using k-means algorithm. A model for image annotation system is created using this image feature vectors and corresponding clusters. This image annotation model is used to classify the test images with the help of fuzzy K-NN classification algorithm in annotation phase.

B. Annotation Phase

Features are extracted from test images using SURF feature extraction method. The process of feature extraction is the same as the feature extraction method described in the training phase. Test image features are extracted and converted in to a row vector using the same techniques described in the training phase.

1) Fuzzy K-NN classification: Keller [5] extended the K-NN [19] algorithm using the fuzzy concept. The theory of Fuzzy set and fuzzy membership functions are introduced in KNN algorithm to implement fuzzy K-NN classification algorithm. Fuzzy K-NN algorithm assigns class membership to a pattern rather than assigning the pattern to a particular class. The membership values for the pattern should provide a level of assurance to accompany the resultant classification. The basis of the algorithm is to assign membership as a function of the pattern distance from its k-nearest neighbors and those neighbors memberships in the possible classes.
Fuzzy K-NN algorithm first finds the distance from unknown vector to all classes using Euclidian distance as given in the equation 4. This equation finds the distance between q and p.

$$d(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \quad (4)$$

The fuzzy K-NN algorithm assigns class membership values to the test image. The class membership values are calculated using equation 5

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij}(1/\|x - x_j\|^{2/(m-1)})}{\sum_{j=1}^{k}(1/\|x - x_j\|^{2/(m-1)})} \quad (5)$$

u_{ij} be the membership in i^{th} class of j^{th} vector of labelled sample.x is an unknown, unlabelled data. x_i be a member of $\{x_1, x_2....., x_n\}$, set of labelled data and m defines how heavily the distance is weighted. Fuzzy k-nn algorithm assigns class numbers to the test images based on the membership value given by the member ship function. Fuzzy K-NN shows membership value of test images in all classes and select the class with highest membership value as the assigned class.

2) Image Annotation: Test images are classified using Fuzzy-KNN algorithm and we use this classification result to annotate an image. Annotation database consists of class numbers and corresponding class names. According to the class numbers assigned by the fuzzy K-NN algorithm, class names are retrieved from annotation database and displayed on the image.

MATLAB [22] is used to implement the proposed system.

IV. Result and Analysis

Experiment is conducted with standard datasets Caltech 101 dataset [20] and corel 1000 data set [21].The data set contains 10 classes and these classes are taken from Caltech 101 dataset [20] and corel 1000 data set [21]. The images are clustered into 10 classes using k-means clustering algorithm. The data set is a combination of small objects and sceneries. A total number of 300 images for training and 500 images for testing are used for the experimental study.

TABLE I
CONFUSION MATRIX OF THE EXPERIMENTAL STUDY

Actual Class	Predicted class										
	Airplane	Building	Headphone	Car	Sunflower	Mountain	Butterfly	Sea	Human	Tree	Unclassified
Airplane	48	0	0	2	0	0	0	0	0	0	0
Building	0	37	4	0	2	1	2	2	0	1	1
Headphone	1	0	39	0	0	2	3	2	1	0	2
Car	0	1	0	49	0	0	0	0	0	0	0
Sunflower	0	1	0	0	48	0	0	0	1	0	0
Mountain	0	5	0	0	2	38	0	3	1	1	0
Butterfly	0	0	0	0	0	0	38	3	0	8	1
Sea	3	0	0	0	1	5	2	35	3	0	1
Human	0	3	1	1	1	1	1	3	39	0	0
Tree	0	0	0	0	0	0	1	0	1	48	0

Table I describes the experimental result of our automatic image annotation system. From table I we found that most of the test images are annotated accurately. The images are annotated based on the classification label assigned by the fuzzy k-nn algorithm. During the experimental study, we have seen that some test images in the remains unclassified due to the lower strength of membership value assigned by the membership function used in the fuzzy K-NN classification step.

A. Evaluation

The performance of an automatic image annotation system can be measured using the standard statistical measures like precision, recall, F-score and accuracy [23]. These parameters can be calculated using the standard measures True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN). The performance matrix of the proposed system is illustrated in Table II.

TABLE II
PERFORMANCE MATRIX

CLASS	TP	FP	FN	TN	PRECISION	RECALL	F-SCORE	ACCURACY
AIRPLANE	48	4	2	446	0.92	0.96	0.94	0.99
BUILDING	37	10	13	440	0.79	0.74	0.76	0.95
HEADPHONE	39	5	11	445	0.89	0.78	0.83	0.97
CAR	49	3	1	447	0.94	0.98	0.96	0.99
SUNFLOWER	48	6	2	444	0.89	0.96	0.92	0.98
MOUNTAIN	38	9	12	441	0.81	0.76	0.78	0.96
BUTTERFLY	38	9	12	441	0.81	0.76	0.78	0.96
SEA	35	13	15	437	0.73	0.70	0.71	0.94
HUMAN	39	7	11	443	0.85	0.78	0.81	0.96
TREE	48	10	2	440	0.83	0.96	0.89	0.98

Average values of precision, recall, F-score and accuracy of the system were obtained as 0.85, 0.84, 0.84, and 0.96, respectively.

V. CONCLUSION

In this work, the problem of automatic image annotation system is investigated through SURF feature extraction algorithm combined with well known algorithms like k-means clustering and fuzzy K-NN classifier. From the performance evaluation, we can conclude that this system shows an accuracy of 0.96.

The proposed automatic image annotation system can be improved by extracting features from a particular region in the image instead of extracting features from the whole image. Segmentation before extracting the features can be applied to improve the performance of the annotation system. This help us to assign multi labels to an image. Another modification can be done using any other classification method, which performs well than fuzzy k-NN classifier.

REFERENCES

[1] V.N.Gudivada and J.V. Raghvan, *Special issues on content based image retrieval system,* IEEE Com. Magzine, 1985.

[2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, *"SURF: Speeded Up Robust Features",* Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346–359, 2008.

[3] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain., *Content-based image retrieval at the end of the early years,* TPAMI, 22(12):13491380, 2000.

[4] J. B. MacQueen, *"Some Methods for classification and Analysis of Multivariate Observations",* Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297,1967

[5] J. M. Keller, M. R. Gray, and J. A. Givens, Jr. *"A Fuzzy K-Nearest Neighbor Algorithm",* IEEE Transactions on Systems, Man, and Cybernetics, Vol. 15, No. 4, pp. 580-585,1985.

[6] Y.Mori, H.Takahashi,and R.Oka. 1999, *"Image-to-word transformation based on dividing and vector quantizing images with words",* MISRM'99 First International workshop on Multimedia Intelligent Storage and Retrieval Management, 1999.

[7] P.Duygulu,K.Barnard,J.F.G.de Freitas,and D.A.Forsyth, *"Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary",* ECCV 02:Proceedings of the 7th European Conference on Computer Vision-Part IV, (London, UK), pp. 97112, Springer-Verlag, 2002.

[8] D. M. Blei and M. I. Jordan, *"Modeling annotated data",* In Proceedings of ACM SIGIR International Conference on Research and Development in Information Retrieval, 2003, pp. 127134.

[9] F. Monay and d. Gatica-perez, *"plsa-based image auto annotation: constraining the latent space",* In proceedings of the 12th annual acm international conference on multimedia, 2004, pp. 348351.

[10] J. Jeon, V. Lavrenko, and R. Manmatha, *"Automatic image annotation and retrieval using cross-media relevance models",* Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, 2003.

[11] V. Lavrenko, R. Manmatha, and J. Jeon, *"A model for learning the semantics of pictures",* In Proceedings of Advances in Neural Information Processing Systems, 2003.

[12] S. L. Feng, R. Manmatha and C. Lavrenko, *"Mulitple Bernouli Relevance Models for Image and Video Annotation",* Proceedings of Of CVPR, Washington, DC, June, 2004.

[13] E. Akbas, *"Automatic image annotation by ensemble of visual descriptors",* Masters thesis, Middle East Technical University, Ankara, Turkey, 2006

[14] Supheakmungkol Sarin, Michael Fahrmair, Matthias Wanger, Wataru Kameyama , *"Holistic Feature Extraction for Automatic Image Annotation",* Fifth FTRA International Conference on Multimedia and Ubiquitous Engineering,2011

[15] M. Idrissa and M. Acheroy, *"Texture classication using gabor lters",* Pattern Recognition Lett, vol. 23, pp. 10951102,2002.

[16] Lowe, David G., *"Object recognition from local scale-invariant features",* Proceedings of the International Conference on Computer Vision. 2. pp. 11501157,1999.

[17] Paul Viola and Michael Jones, *"Rapid object detection using a boosted cascade of simple features",* In CVPR , pages 511-518, 2001.

[18] T. Lindeberg, *"Feature detection with automatic scale selection",* IJCV, 30(2):79 -116, 1998.

[19] P.Hart, *"The condensed nearest neighbour rule",* IEEE trans, Information Theory, Vol IT-14,pp 515-516,1968.

[20] L. Fei-Fei, R. Fergus and P. Perona., *"Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories",* IEEE. CVPR 2004, Workshop on Generative-Model Based Vision, 2004.

[21] James Z. Wang, Jia Li, Gio Wiederhold, *"'SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries",* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 23, no.9, pp. 947-963, 2001.

[22] MATLAB 7.8, ., Natick, *"The MathWorks Inc",* MA, 2009.

[23] Olson, David L.; and Delen, Dursun, *"Advanced Data Mining Techniques",* Springer, 1st edition , page 138,2008

[24] Papoulis, A, *". Probability, Random Variables, and Stochastic Processes, 2nd ed.",* New York: McGraw-Hill, pp. 104 and 148, 1984.

[25] Dengsheng Zhang n, Md. Monirul Islam, Guojun Lu, *".A review on automatic image annotation techniques.",* Pattern Recognition, Volume 45, Issue 1, January 2012, Pages 346-362