International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 3 Issue: 5

ISSN: 2321-8169
3053 - 3057

Object Recognition and Clustering based on Latent Semantic Analysis (LSA)

Mr. Vinaykumar Hebballi
Department of Computer Science
PES Institute of Technology
Bangalore, India
vinaykumar.hebballi@gmail.com

Ms. Vidhu Rojit
Asst. Professor, Department of Computer Science
PES Institute of Technology
Bangalore, India
vidhurojit@pes.edu

Abstract— Object Recognition and clustering are prime techniques in Computer Vision, Pattern Recognition, Artificial Intelligence and Robotics. Conventionally these techniques are implemented in Visual-Feature based methods. However, these methods have drawbacks they do not efficiently deal with the differences in shapes and colours of objects. Another method which uses semantic similarity to solve this kind of problem, i.e. Cosine Similarity method, but this method also has problems. The problems are synonymies and polysemies. In this paper we propose a method in which objects with different shapes and different colours which function similarly can be recognized and clustered. If the text printed on the object the semantic feature of that object is extracted and clustered according to semantic feature. Proposed method is based on semantic information so we conduct an experiment with the dataset of images which contains the packing cases of commercial products (e.g. Mobile, Laptop etc). Semantic information in dataset is retrieved using text extraction module and then the results of text extraction are passed through an Internet search module. Finally objects are described and clustered using the latent semantic analysis (LSA) module. The clustering results are more accurate than the Visual feature based method and cosine similarity based methods.

Keywords- *Object Recognition, Image Processing, Text Mining, Object Clustering*

*****

I. Introduction

Object recognition and clustering are prime techniques in pattern recognition, artificial intelligence, and robotics and computer vision, enabling a robot or an application to differentiate objects or to automatically associate same objects. Many researchers use visual feature based methods to solve this kind of recognition problem. There are many publicly available data processing algorithms which promise high accuracy and performance, such as scale-invariant features transform (SIFT) [1] and Speeded-up robust features (SURF) [2]. These algorithms may not be able to find variation in shapes and colours between objects, the visual feature of similar objects may be small and may not be recognized or clustered correctly by Visual feature based method. In contrast, if the object surface contains any printed text this can be used to recognize and cluster different types of objects i.e. Cosine Similarity method [13] [14], but this method has some problems. The problems with cosine similarity are synonymies and polysemy [15]. In synonymies many ways to refer to the same objects for e.g. car and automobile, which lead to poor recall. Similarly, in polysemy most words have more than one meaning, e.g. model, python and chip, which lead to poor precision.

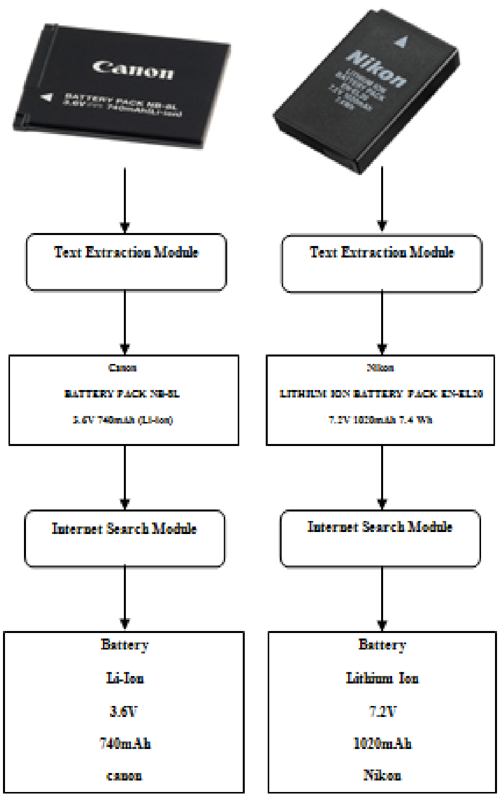

Fig.1. Method for extracting semantic features.

We introduce a method that can describe, recognize and cluster similar objects by measuring the semantic similarities. To get similarity between objects, we must obtain

 ISSN: 2321-8169
3053 - 3057

semantic feature of the object. We give input as scanned image then obtain the text from the image using Optical Character Recognition (OCR) technique. These texts are regarded as the object's semantic feature from which the object is identified, described and clustered.

The text printed on the object surface is not enough to recognize the object and classify the objects. For more accurate classification we need to expand this semantic feature. So we introduce the idea of strong and weak words. The strong words are those words which are retrieved from text image using OCR. The strong words are directly related to the object. Some text is indirectly related to object, but directly related to strong word. This will increase the semantic feature of object the weak words are obtained by passing strong word as keyword to internet search. After that cluster the objects according to latent semantic analysis (LSA) results.

II. RELATED WORK

An effective recent technique for visual feature extraction is SIFT (Scalable Invariant Feature Transform). This introduced by D. G. Lowe [1]. In SIFT an image is transformed into a huge collection of local feature vectors and each is independent with respect to image translation, scaling and rotation, but moderately dependent in terms of illumination changes and even and 3D projection. In SURF (Speeded-Up Robust Features) by H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool [2] is partly inspired by the SIFT descriptor. SURF detects interest points using a basic Hessian matrix approximation, which lowers the computational cost. Features are retrieved by summing the Haar wavelet response around the interest points.

The image segmentation and Thresholding are very well known procedure to select particular segment in an image. The automatic threshold selection for image segmentation is pretend in N. Ostu's Method [3].

Object recognition has been researched for almost half a century by the author S. Ullman and G. J. Power [4], and image-based object categorization has advanced rapidly in recent years by the Authors J. Ponce, M. Hebert, C. Schmid, and A. Zisserman [5]. Object recognition is typically based on geometry, appearance or features. The geometric approach was adopted in early object recognition attempts. It was mainly focused on geometrical primitives such as lines and curves [6]. Appearance-based techniques emerged later and are more widely used. Eigen face methods are computationally efficient and accurate first-face recognition systems. This was introduced by M. Turk and A. Pentland [7], are of this type. The third approach is most commonly used today. The SIFT technique performs well for objects containing rich texture information, but it requires sophisticated indexing and matching algorithms for effective object recognition [1] [5].

III. SYSTEM OVERVIEW

A system that describes and clusters different objects into several logical groups on the basis of their semantic features has been implemented. For example, given two laptops images of different shapes and colours, we extract their respective semantic features by reading text on their and figure out logical relationships from these features.

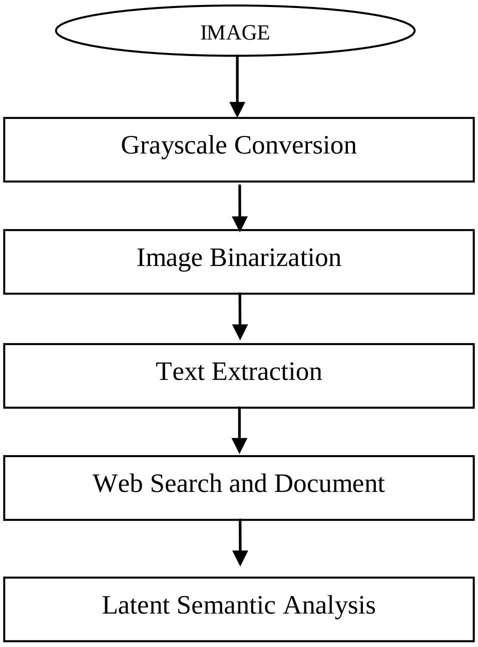

Fig.2. System overview of semantic feature extraction and object description and clustering

The system is divided into 5 modules. The System first converts the original image into grayscale image then that image is converted into binary image using Ostu's method. The image is input to an OCR engine which translated text image into texts. The text is sent to the internet search module using the search results we extract the semantic feature. Finally the Semantic features are calculated using Latent Semantic Analysis and clustered using k- mean algorithms.

IV. MODULES

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

 ISSN: 2321-8169
3053 - 3057

A. GRAYSCALE CONVERSION

The first problem in our research is to read text from image. Failure of getting text from image results in early termination of reading text. Therefore the primary mission is to detect text region.

In this module first we convert the original scanned images into grayscale images by using luminosity Algorithm. The luminosity algorithm is sophisticated algorithm to convert colour image to grayscale image. The luminosity algorithm has many formulae to calculate grayscale level but we use the following formula:

$P = 0.21 \times R + 0.71 \times G + 0.07 \times B$

Using this we convert original images to grayscale image. Then that image is converted to binary image.

B. IMAGE BINARIZATION

To binarize our image, we will use the Otsu's method [3] [8], which is a very clever method created by Nobuyuki Otsu. The algorithm assumes that the image to be threshold contains two classes of pixels (e.g. Foreground and background) then calculates the optimum threshold separating those two classes so that their combined spread is minimal. Its main advantages are speed and the ease of implementation. The method involves iterating through all possible threshold value and calculating a measure of spread for the pixel levels, each side of the threshold, i.e. the pixels that either falls in the foreground or background. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum.
The Ostu's Method is composed of following steps:

STEP 1: Compute histogram and probabilities of each intensity level.

STEP 2: Set up initial ωi (0) and μi (0).

STEP 3: Step through all possible thresholds t=1… maximum intensity.

Update ωi and μi.

Compute $\sigma^2 b$ (t).

STEP 4: Desired threshold corresponds to the maximum $\sigma^2 b$ (t).

STEP 5: You can compute two maxima (and two corresponding thresholds). $\sigma^2 b1$ (t) is the greater max and $\sigma^2 b2$ (t) is greater or equal maximum

STEP 6: Desired threshold = (Threshold1 + Threshold2) / 2 .

Fig.3. original image to grayscale and grayscale to binary image.

C. TEXT EXTRACTION MODULE

In this module text region images have been detected and separated. This module converts these images into text strings using Asprices OCR engine. The OCR engine outputs the corresponding text and their confidence scores, computed from text region images. Texts whose confidence score exceeds 75% are considered as successful OCR results while those with confidence scores below 75% are considered as failures. Text strings of confidence score exceeding 75% are retained, while those of lower confidence scores are filtered out. The above modules retrieve strong words only. Weak words are obtained by applying the Internet Search module to the strong words.

D. WEB SEARCHING AND DOCUMENT

In this module strong words are referred to as keywords for internet search. These keywords are successively searched with Google Search API. Each keyword search returns 10 results, and each result contains keyword-related information, namely, a title, URL and snippet. The HTML code is removed from search results by an HTML special character filter. Meaningless HTML codes returned with the results, such as "", are removed by an HTML special character filter.

For each given keyword the internet search module results 10 titles and 10 snippets. The URL and HTML links are removed from results because they are irrelevant to concept. The title and snippets are joined to form single word or sentence. To avoid grammatical problems in semantic feature extraction we remove stop word like to, are, are etc from sentence. Then we store the sentence in document file to find semantic feature using latent semantic analysis. The implementation of the module is shown in Fig 4.

E. LATENT SEMANTIC ANALYSIS

To extract the semantic feature from document we use latent semantic analysis (LSA) [9] [10]. The LSA is a method to find hidden concept, in document data. In this module each document and term is expressed in terms of vectors with element corresponding to their concept. Each element in vector given the frequency of occurrence in document or term in the concept.

The first step is to represent the text as a matrix in which each row stands for unique keyword and column stands for document. Each cell contains the frequency of occurrences in the document which is denoted by column. We call this matrix as X. $B=X^TX$ is document-document matrix and $C=XX^T$ term-term matrix. Both B and C are square and Symmetric matrix.

Next, LSA applies Singular Value Deposition (SVD) to the matrix. In SVD a rectangular matrix is decomposed into the product of three other matrices. One component matrix describes the original row entities as a vector of orthogonal factor values. Another describes the original column entities in the same way; the third is a diagonal matrix containing scaling values such that when the three components are matrix multiplied the original matrix reconstructed. When fewer than the necessary number of factors is used, the reconstructed matrix is least- square best fit. One can reduce the dimensionality of the solution simply by deleting coefficients in diagonal matrix ordinarily starting with smallest.

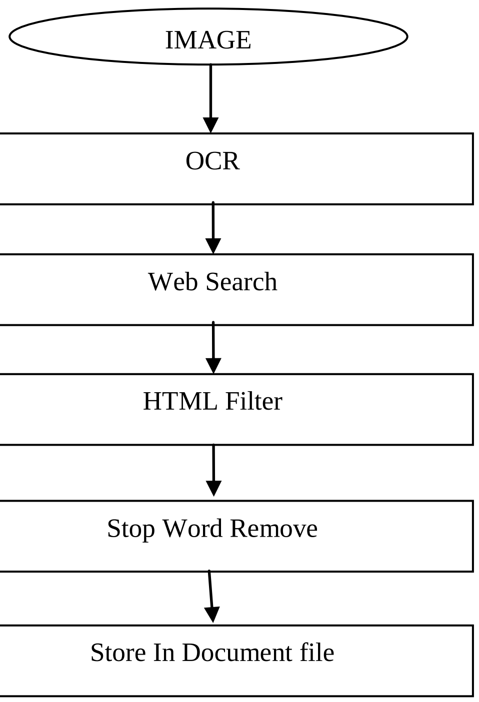

Fig.4. Overview of web search and Document File.

$$\{X\}= \{W\} \{S\} \{P^T\}$$

Where W is matrix of eigenvector of B. P is the matrix of Eigen vector of C. S is the diagonal matrix of singular value observed as square root of Eigen value of B.

In Singular Value Decomposition (SVD) some of singular values are too small so in latent semantic analysis (LSA), we ignore such small singular value and replace it by value 0.we keep only k singular value in S. The S will be containing all zeros except first k value. We can reduce the S into Sk which is kxk matrix which contains only k singular values and it also reduce the W and P^T into Wk and P^Tk to have only k rows and k columns.

$$\mathbf{X_k = W_k\, S_k\, P^T_k}$$

V. EXPERIMENTS AND RESULTS

We tested our system with dataset of 1000 text images of different products like camera, keyboards, lens, mobile etc. To calculate the effectiveness of our system, we cluster them into 10 groups with iterative k-mean algorithm [12] using latent semantic feature set.

TABLE I. Sample of Results for Describing Objects in Semantic Feature Documents.

Objects	DOC1	DOC2	DOC3	DOC4	DOC5
XBOX 360 GoPro60GB HARD DRIVE	1	0	0	0	0
Logitech G940 Joystick	0	1	0	0	0
Nikon SB-910 Flash	0	0	1	0	0
Canon NB-6L Battery	0	0	0	1	0
LG Keyboard	0	0	0	0	1

The experiment took around 2 hours in machine with Intel i5 processor. The extent to which our system describes an object is demonstrated for five objects only and each of 5 features in Table I. the first column contains objects name second to sixth are frequency values for latent semantic analysis.

Clustering performance was assessing from four indexes: ARI, RI, NMI, and ACC. Table II shows the results of clustering images. The results are compared with cosine similarity and visual feature based similarity. By looking at the results TABLE II our proposed system has better performance that visual feature based method and cosine similarity based method.

ISSN: 2321-8169
3053 - 3057

VI. CONCLUSION

The paper presents an algorithm that automatically reads the text printed on the surface of object and subjects them to web search. From web search results the objects described as feature word expressed in human language. Objects with similar feature word description (based on Latent Semantic Analysis) are clustered together. Our Latent Semantic Analysis algorithms demonstrate a clustering performance higher than that of vector space method and visual feature based clustering method.

TABLE II. Evaluation of Performance of Latent Semantic Feature in Clustering Images.

	LSA(Latent Semantic Similarity)	Cosine Similarity Method	Visual Feature based Method
ARI	0.382	0.278	0.245
RI	0.842	0.804	0.714
NMI	0.573	0.541	0.481
ACC	0.982	0.828	0.629

In future research, we will introduce probabilistic Latent Semantic Analysis or NLP algorithms to improve precision of feature word extraction. We can also improve text extraction by considering real word problems. The proposed method only deals with objects related to text printed on surface. We are going to expand the use of the proposed method to deal with objects with text that are not related to objects.

REFERENCES

[1] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[3] N. Ostu, “A Thresholds Selection Method from Gray-Level Histograms,” in systems, man, and cybernetics*, 1979 IEEE Transaction*.

[4] S. Ullman and G. J. Power, “High-level vision: Object-recognition and visual cognition,” *Optical Engineering*, vol. 36, no. 11, pp. 3224–3224, 1997.

[5] J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, *Toward categorylevel object recognition*. Springer, 2007, vol. 4170.

[6] J. L. Mundy, A. Zisserman *et al.*, *Geometric invariance in computer vision*. MIT press Cambridge, 1992, vol. 92.

[7] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[8] Wolf, C.; Jolion, J.; Chassaing, F., "Text localization, enhancement and binarization in multimedia documents," *Pattern Recognition, 2002. Proceedings. 16th International Conference on* , vol.2, no., pp.1037,1040 vol.2, 2002.

[9] Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, **25**, 259-284.

[10] Bellegarda,J.R. , Butzberger, J.W, Yen-Lu Chow, Coccaro, N.B. Naik, D., “A novel word clustering algorithm based on latent semantic analysis ,” Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on , vol. 1, , no., pp.172,175 vol. 1, 7-10 May 1996.

[11] M. MeilPa and J. Shi, “Learning segmentation by random walks,” 2001.

[12] Chang Wen Chen; Jiebo Luo; Parker, K.J., "Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications," Image Processing, IEEE Transactions on , vol.7, no.12, pp.1673,1683, Dec 1998.

[13] Jing Xu; Okada, S.; Nitta, K., "A Semantic-Similarity-Based Method for Object Description and Clustering," Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on , vol., no., pp.3669,3674, 13-16 Oct. 2013.

[14] Soe-Tsyr Yuan; Jerry Sun, "Ontology-based structured cosine similarity in document summarization: with applications to mobile audio-based knowledge management," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol.35, no.5, pp.1028,1040, Oct. 2005.

[15] Lushan Han; Finin, T.; McNamee, P.; Joshi, A.; Yesha, Y., "Improving Word Similarity by Augmenting PMI with Estimates of Word Polysemy," Knowledge and Data Engineering, IEEE Transactions on , vol.25, no.6, pp.1307,1322, June 2013