

# Story Board Creation by Using Image Segmentation

Avanti Ralegaonkar, Nitin Pise

**Abstract**— In present world we are standing by such a generation of internet where sharing of photos and videos has become very easy. Thanks to great improvement in technology, different websites and various application they provide. Also credit goes to awareness among people and comfort they find in using these facilities. In today's world people share photos (i.e. images) for any and every occasions and places they visit and while doing so they mainly specify captions as place name or occasion but not much detail is said about the actual photo (i.e. image and its contents). Present methods of searching images over the internet emphasize on file names / captions rather than actual content which leads to inaccurate results. So here, we are proposing a solution for this problem that will generate more accurate results while searching for images which is based on contents rather than additional information and thus giving end user more appropriate results. Entire work is divided broadly in two simple steps; 1. Extracting objects from input image 2. Generating keywords for them and storing them in human understandable language.

**Index Terms**— Content based image retrieval, Image parsing, Image retrieval, Image Segmentation, Object Recognition.

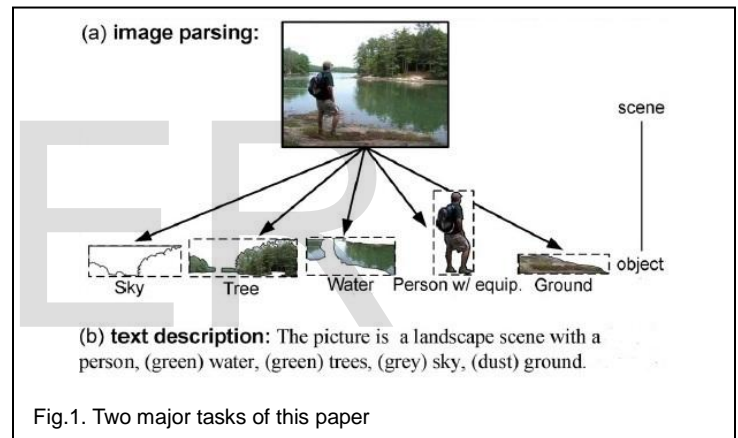
## 1 INTRODUCTION

Fast growth of public photo and video sharing websites, such as Flickr and YouTube provides a huge corpus of unstructured image and video data over the Internet. Searching and retrieving visual information from the Web, however, has been mostly limited to the use of metadata, user-annotated tags, captions, and surrounding text. In this paper we present Story Board Creation By Using Image Segmentation [5][1] framework generates text descriptions in natural language based on understanding of image content.

Fig. 1 illustrates two major tasks are image parsing and text description. By analogy to natural language understanding, image parsing computes a parse graph of the most probable interpretation of an input image. This parse graph includes a tree structured decomposition for the contents of the scene, from scene labels, to objects, to parts and primitives, so that all pixels are explained. The parse graph is similar in spirit to the parsing trees used in speech and natural language understanding for specifying relationships and boundary sharing between different visual patterns.

Segmentation involves partitioning an image into a set of homogeneous and meaningful regions, such that the pixels in each partitioned region possess an identical set of properties. Image segmentation is one of the most challenging tasks in image processing and is a very important pre-processing step in the problems in the area of image analysis, computer vision, and pattern recognition [5, 6].

Large numbers of segmentation algorithms are present, but there is no single algorithm that can be considered good for all images. Algorithms developed for a class of images may not always produce good results for other classes of images.



Normally, when we search images based on a specific keyword. Any search engine takes the keyword and searches it by file name of the given image and not by content. Search engine has a backbone of textual information which is used for indexing. That is why search is based on image name. So, image search results in inaccurate results.

As simple as the Story Board Creation By Using Image Segmentation task in Fig. 1 may seem to be for a human, it is by no means an easy task for any computer vision system today especially when input images are of great diversity in contents (i.e., number and category of objects) and structures (i.e., spatial layout of objects), which is certainly the case for images from the Internet. For this reason, our objective in this paper is twofold. 1) We use a semiautomatic method (interactive) to parse general images from the Internet in order to build a large scale ground truth image data set. Our goal is to make the parsing process more and more automatic using the learned models. 2) We use automatic methods to parse images in specific domains.

The rest of the paper is organized as : Section 2 gives a brief description of related work in the field of Image Parsing, important papers and literature that we have studied or used

- Avanti C. Ralegaonkar is currently pursuing masters degree program in computer engineering in Pune University, India, PH-9561079824. E-mail: [ralegaonkar\\_avanti@yahoo.in](mailto:ralegaonkar_avanti@yahoo.in)
- Nitin N. Pise, my guide is currently working as associate professor in Pune University, India. E-mail: [nitin.pise@mitpune.edu.in](mailto:nitin.pise@mitpune.edu.in)

as a part of our literature survey. Section 3 explains the design of project which includes mathematical model, data flow architecture and turing machine. Section 4 shows the results of work done so far. Section 5 summarizes the conclusion and future work related to this project.

## 2 RELATED WORK

Dmitriy Fradkin, Ilya Muchnik, described the concept of constructing hierarchical classifiers using cluster analysis and suggest a new method and improvement in these approaches. They also suggest a new method for constructing features that improve classification accuracy of Multiclass SVM[8].

The method of simple and efficient implementation of Lloyds k-means clustering algorithm, which we call the filtering algorithm. This algorithm is easy to implement, requiring a kd-tree as the only major data structure is present in an efficient k-means clustering algorithm. It was proposed by T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatco, R. Silverman, and A. Y. Wu[9].

In 2007, Chris Vutsinas described the concept of Image Segmentation: K-means and EM Algorithms. In this method, two algorithms for image segmentation are studied. K-means and Expectation Maximization algorithm are each considered for their speed, complexity, and utility [10].

Faguo Yang and Tianzi Jiang described the concept of pixon-based adaptive scale method for image segmentation. The key idea of the approach is that a pixon-based image model is combined with a Markov random field (MRF) model under a Bayesian framework [11].

H. D. Cheng, X. H. Jiang, Y. Sun, Jingli Wang, described the concept of monochrome image segmentation approaches operating in different color spaces such as histogram thresholding, characteristic feature clustering, edge detection, region-based methods [12].

A method of color image segmentation based on fuzzy homogeneity and data fusion techniques is presented. The general idea of mass function estimation in Dempster-Shafer evidence theory of the histogram is extended to the homogeneity domain. The fuzzy region in each primitive color, whereas, the evidence theory is employed to merge different data sources in order to increase the quality of the information and to obtain an optimal segmented image. Segmentation results from the proposed method are validated and the classification accuracy for the test data available is evaluated, and then a comparative study versus existing techniques is presented. It was described by Salim Ben Chaabane, Mouniri Sayadi, Farhat Fnaiech and Eric Brassart in 2009[13].

The method to split color information is the image to be segmented. Hence, this is a blind color image segmentation method. It consists of four subsystems: pre-processing, cluster detection, cluster fusion and postprocessing. It was proposed by Ezequiel Lopaz-Rubio, Jose Munoz-Perez, Jose Antonio Gomez-Ruiz[14].

The concept of a new color thresholding method for detecting and tracking multiple faces in video sequence. The proposed method calculates the color centroids of image in

RGB color space and segments the centroid region to get ideal binary image at first. Then analyse the facial features structure character of wait-face region to fix face region. The novel contribution of this paper is creating the color triangle from RGB color space and analysing the character of centroids region for color segmenting. It was proposed by Jun Zhang, Qieshi Zhang, and Jinglu Hu in 2009[15].

The concept of Fusion of multispectral image with hyper spectral image generates a composite image which preserve the spatial quality from the high resolution (MS) data and the spectral characteristics from the hyperspectral data, is presented in Performance analysis of high resolution and hyper spectral data fusion for classification and linear feature extraction. It was proposed by Shashi Dobhal in 2008[16].

Cheolha Pedro Lee in 2005, described the concept based on the statistics of image intensity where the statistical information is represented as a mixture of probability density function defined in a multi-dimensional image intensity space. Depending on the method to estimate the mixture density functions, three active contour models are proposed: unsupervised multi-dimensional histogram method, half-supervised multivariate Gaussian mixture density method, and supervised multivariate Gaussian mixture density method is present in Robust Image Segmentation using Active Contours [17].

Fahimeh Salimi, Mohammad T. Sadeghi in 2009, introduced a new histogram based lip segmentation technique is proposed considering local kernel histogram in different illumination invariant color spaces. He histogram is computed in local areas using two Gaussian kernels: one in color space and the other in the spatial domain. Using the estimated histogram, the posterior probability associated to non-lip class is then computed for each pixel. This process is performed considering different color spaces. A weighted averaging method is then used for fusing the posterior probability values. As a result a new score is obtained which is used for labeling pixels as lip or non-lip. The advantage of the proposed method is that the segmentation process is totally unsupervised [18].

Extracting objects from image is of major interest in numerous applications for example image processing, computer vision. Segmentation techniques usually rely on pixel level classification schemes exploiting local image related information. The classification step is achieved either using thresholding technique or Bayesian framework. The computation of a primary layer of spatial region is processed either relying on motion based criteria or on intensity, texture and color information. The second type of technique usually supplies a better localization of boundaries, likely to correspond to intensity, texture or color contours. In fact, starting from this initial spatial partition, a 2D parametric model, generally an affine one, is attached to each spatial region, and spatial regions are merged according to object properties. The major drawback of these approaches is that they prevent from processing a very fine spatial segmentation, since parametric estimators generally require a significantly large estimation support to be reliable. Thus, it may result in the loss of boundaries.

### 3 PROGRAMMER'S DESIGN

This is a proof of concept. I have limited this scope to following- Already pre-defined objects (flower, nature, greenery, water, sky, human being, animal, bird, ball, buildings, vehicle, ) only high level classification (can identify animal but can not say which animal it is).

#### Features

To identify objects, classification of objects and indexing, image retrieval (exact matching keywords).

#### Goals

For this proof of concept (flower, nature, greenery, water, sky, human being, animal, bird, ball, buildings, vehicle) various objects are considered. Improve image quality, Image enhancement, Feature extraction. To achieve this goal, we intend to combine known techniques in image segmentation based algorithms for image processing.

The results of the study will be used both to reduce the complexity of algorithms computing homology groups of image objects and to determine the relevance of homology elements depending on the application context. Our research will be led along the following topics: stability of image operations, classification of images, and efficient computation of image parsing information, and to succeed I will address the following questions:

- a) How stable are parsing algorithms (noise, data distortion, or transformations (fusion of objects))?
- b) Is it possible to extract information of an object from some of its projections or cuts?
- c) Which algorithms are well suited for efficient homology computation?
- d) Is there a notion of adjacency for classes of objects, defined by algorithms?

#### Objectives

From a given parse graph, the task of our objective is twofold. 1) I try to develop an automatic method (interactive) to parse general images from the Internet in order to build a large scale ground truth image data set. My goal is to make the parsing process more and more automatic using the learned models. 2) I also try to develop automatic methods to parse images/videos in specific domains. For example, in the surveillance system, the camera is static, so we only need to parse the background (interactively) once at the beginning, and all other components are done automatically. Although the manual image parsing algorithm may produce some errors, it is fully automatic.

To make this concept actual working on limited number of objects and standard database this can be expanded in future.

#### Constraints

As this is just a proof of concept, so I have limited this to number of objects, limited number of image databases.

Input data for this work would be an image set available at Microsoft. This image set would be a mix of different images having animals, flowers, landscape etc.

Output Data would be a story board with objects identified in layman language viz. an image contains a horse or a cow etc.

So, success definition for this work would be - on input of an image we should get correct objects being identified in layman language which is usually used by humans to search for images. System architecture is explained in Fig 2.

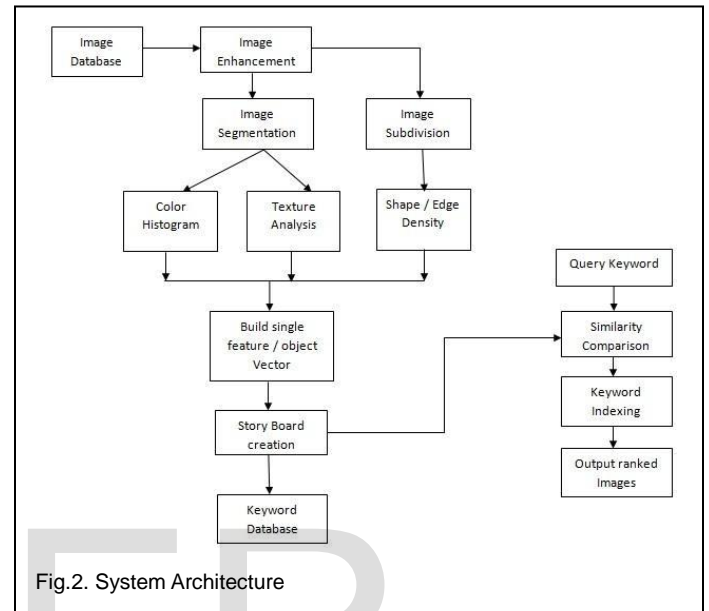


Fig.2. System Architecture

### 3.1 Phases

#### Phase I: Extraction of object from image file

This phase will take image as an input. And produces the story board for the specified image (which contains the keyword i.e. objects) as an output. This phase is further subdivided into four sub phases. The first phase will convert the input image into pre-processed image. Then this pre-processed image is then fed into the second phase which will process that image to extract the object. The output of this phase is given to a natural language processing which will produce the story board.

#### Phase II : Querying keyword

This phase will provide the user interaction to our system. User can provide his desired keyword as a query. We will process the story boards and database for the image files which was produced in the Phase I and produce the output.

#### Steps to carry out proposed work

1. Parsing: Parse the image files to get the story board. For indexing step, input image files must be converted into story board.
2. Keyword Extraction: Extract the keywords (i.e. objects) using different algorithms for color, texture and edge.
3. Indexing: Index the image file. Indexing is for searching.
4. Retrieval: Then based on the keywords , various images that will match to the keyword are extracted.

### 3.2 Experimental Setup

The algorithm presented here was designed to solve the problem of correctly segmenting objects. It works on various types of everyday images: both color and gray scale. We will test our algorithm on Microsoft Segmentation Database (MSD), which is image segmentation bench-mark set. It Contains 3000 real life images of different categories and same size 481x321 pixels. For the processing of images we are working on Intel R core i3 2.77GHz (min), with 4 GB RAM, and we have used C # · NET (GUI and Processing) on Windows 07 operating system.

The algorithm used so far is Fuzzy C means because, particularly fuzzy C-means based clustering and its variants, have been widely used in the task of image segmentation due to their simplicity and fast convergence. It is a soft segmentation method. It produces an idea of overlapping membership function. Each pixel has a degree of belonging to different clusters, as in fuzzy logic rather than completely belonging to just one cluster. It finds a good partition of image by finding suitable prototype that minimizes the objective function. It is an iterative clustering method that produces 'C' optimal partitions by minimizing the weighted within group sum of squared error objective function  $J_m$

$$J_m = \sum_{i=1}^N \sum_{j=1}^c u_{ji}^m d^2(x_i, v_j) \quad (1)$$

A solution of the objective function (1) can be obtained through an iterative process, which is carried as follows.

- 1) Set the values for c, m and  $\epsilon$ .
- 2) Initialize the fuzzy partition matrix  $U_{(0)}$ .
- 3) Set the loop counter  $b=0$ .
- 4) Calculate the 'c' cluster centers  $(b) j v$  with  $U_{(b)}$ .

$$v_j^{(b)} = \frac{\sum_{i=1}^N (u_{ji}^{(b)})^m x_i}{\sum_{i=1}^N (u_{ji}^{(b)})^m} \quad (2)$$

- 5) Calculate the membership matrix  $U_{(b+1)}$ .

$$u_{ji}^{(b+1)} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ji}}{d_{ki}}\right)^{2/m-1}} \quad (3)$$

If  $\max \{U_{(b)} - U_{(b+1)}\} < \epsilon$  then stop else set  $b=b+1$  and go to step 4. Where 'X' is the data set in the m-dimensional vector space, N is the number of data items, c is the number of clusters and its minimum value is two and maximum value is N.  $\epsilon$  is the threshold value for clustering.  $U_{ji}$  is the degree of membership of  $x_i$  in  $j^{th}$  the cluster, m is the weighting exponent on each fuzzy membership,  $v_j$  is the prototype of the center of cluster,  $d^2(x_i, v_j)$  is a distance measure between object and cluster center.

### 3.3 Data Independence and Data Flow Architecture

This section depicts Data Flow Diagram which is a graphical representation of the "flow" of data through a sys-

tem, modeling its process aspects. Fig 3 describes Data Flow Diagram for the proposed system containing different processes.

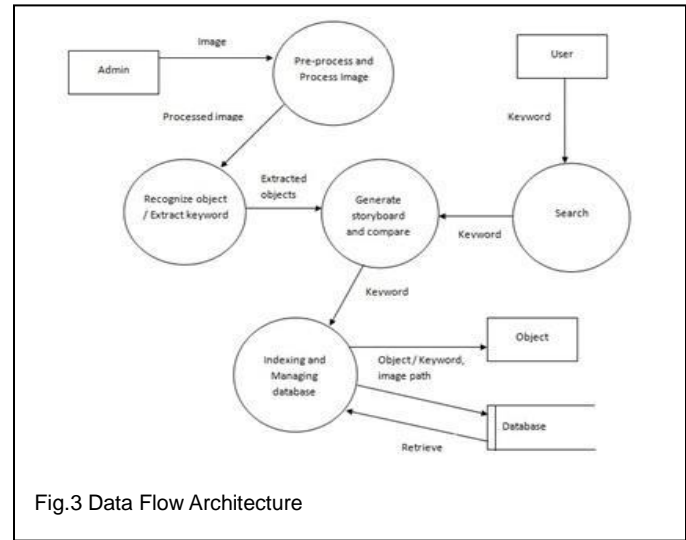


Fig.3 Data Flow Architecture

### 3.4 Turing Machine

This section depicts Turing Machine which is a simple state diagram showing all possible changes in states or the behavior of systems. Fig 4 is showing Turing Ma-chine for my system.

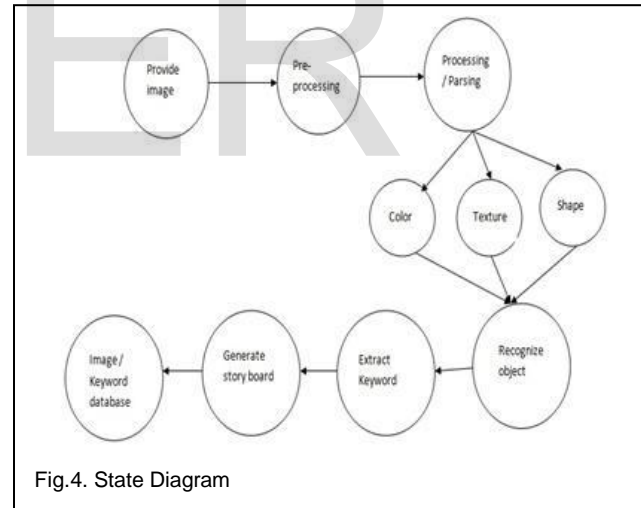


Fig.4. State Diagram

### 4 FUTURE SCOPE

As a part of future work we will incorporate different types of image features. To improve the segmentation and retrieval quality, I will try to develop a methodology to enhance the importance of image central part, which is closer to human judgment. We can also try to develop an image similarity evaluation method between images with multi objects.

### 5 CONCLUSION

This paper proposes model for object extraction, image processing and storing of data in the form of contents as a key-

word to a system. We use combine methods and characteristics of image processing to get the best optimum result.

## REFERENCES

- [1] Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee, and Song-Chun Zhu, "12T: ImageParsing to Text De-scription", No. 8, August 2010 – Proceedings of the IEEE.
- [2] Nida Aslam, Irfanullah, Kok-Keong Loo, Roohul-lah, "Limitations and Challenges: Image/ Video Searching and Retrieval" in doi : 10.4156/jd-cta.vol3.issue1.aslam
- [3] Eva Horster, Rainer Lienhart, Malcolm Slaney, "Retrieval on Large-Scale Image Databases", ACM Trans. CIVR'07, July 9-11.
- [4] S. Nandagopalan, Dr. B. S. Adiga, and N. Deepak, "A Universal Model for Content-Based Image Re-trieval", World Academy of Science, Engineering and Technology 46, 2008.
- [5] A. Zipern, News Watch; A Quick Way to Search for Images on the Web. [Online] Available: <http://www.nytimes.com/2001/07/12/technology/news-watch-a-quick-way-to-search-for-images-on-the-web.html>.
- [6] C. Manning, H. Schutze, and M. Press, Foundations os Statistical Neural Language Processing. Cambridge, MA: MIT Press, 1999.
- [7] L. Li, R. Socher, and L. Fei-Fei, "Towards total scene understanding: Classification, annotation and segmentation in an automatic framework", in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2036-2043.
- [8] Dmitriy Fradkin, Ilya Muchnik, "A Study of K-Means Clustering for Improving Classification Accuracy for Multi-Class SVM ", DIMACS Technical Report 2004-02, April 2004.
- [9] T. Kanungo, D. M. Mount, N. Netanyahu, C. Pi-atko, R. Silverman, and A. Y. Wu, "An Efficient k-means clustering algorithm: analysis and implemen-tation", IEEE Trans. Pattern Analysis and MachineIntelligence, 24(2002), 881-892.
- [10] Chris Vutsinas, "Image Segmentation: K-Means and EM Algorithms", 2007.
- [11] Faguo Yang and Tianzi Jiang, "Pixon-based Image Segmentation, with Markov Random Fields", Vol.12, No. 12, December 2003
- [12] H. D. Cheng, X. H. Jiang, Y. Sun, Jingli Wang, "Color Image Segmentation: Advances and Prospects", Vol. 34, No. 12. (December 2001), pp. 2259-2281.
- [13] Salim Ben Chaabane, Mounir Sayadi, Farhat Fnaiech and Eric Brassart, "Color Image Segmen-tation Using Homogeneity Method and Data Fusion Techniques", from Hindwai Publishing Corporation, May 2009.
- [14] Ezequiel Lopaz-Rubio, Jose Munoz-Perez, Jose An-tonio Gomez-Ruiz, "A four-stage system for blind color image segmentation", Vol. 10, No. 2, 2003, pp. 127-137.
- [15] Jun Zhang, Qieshi Zhang, and Jinglu Hu, "RGB Color Centroids Segmentation (CCS) for Face Detection", ICGST-GVIP Journal, ISSN 1687-398X, Vol. 9, Issue (II), April 2009.
- [16] Shashi Dobhal, "Performance analysis of high-resolution and hyperspectral data fusion for classifi-cation and linear feature extrac-tion", January 2008.
- [17] Cheolha Pedro Lee, "Robust Image Segmentation us-ing Active Con-tours: Level Set Approaches", 2005.
- [18] Fahimeh Salimi, Mohammad T. Sadeghi, "Decision Level Fusion od Color Histogram Based Classifiers for Clustering of Mouth Area Images", International Conf. on Digital Imag Processing, 2009.