ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

Object Detection with Gross Size Estimation Using Advanced Segmentation Algorithm in Multimedia Mining

Mudita Bhadania[1], Dr. Bhupesh Gour[2]

Research Scholar, Department of Computer Science and Engineering, Technocrate Institute of Technology, Bhopal, India [1]

Professor, Department of Computer Science and Engineering, Technocrate Institute of Technology, Bhopal, India [2]

Abstract: Segmentation of objects plays an important role in images and real-time for Multimedia applications, Machine vision, Medical imaging and Recognition Tasks. In the recent researches a great challenge is the object detection in a surveillance security system and traffic management there was a great need to find the gross size of the detected object in system to better analysis in different aspects. So In this paper we are trying to detect objects in multimedia files using advance segmentation algorithm. This is a hybrid form of Joint Segmentation Algorithm (JSEG) and edge based algorithms with gross size estimation of objects in Images and Videos.

Keywords: Joint Segmentation, Edge based Algorithm, Gross Size Estimation, and Image Segmentation

I. INTRODUCTION

Video surveillance and security systems are crucial for crime prevention, and to protect critical infrastructure. Current generation surveillance systems unlike closed circuit TV (CCTV) systems have an important advantage that they can use the existing internet infrastructure, thus making it much easier to deploy video surveillance equipment in a cost effective way. On the other hand the emergence of visual sensor networks as a promising way to deal with multiple nodes poses many research challenges. One of the key challenges among others is the ability to manage the huge information flow coming from each imaging node (visual sensor). Transmitting image/video data to a central station for storage/analysis is not efficient when considering a large number of these camera nodes. In order to make effective use of the visual surveillance system it is of key importance to equip each sensing node with autonomous sensing capabilities i.e. push intelligence into the device. One major drawback with most existing surveillance systems is the lack of the ability to process information locally and only send important information to the central station. The amount of data generated by visual sensor networks is huge and the fact that this data can only be used for later analysis and cannot be used to detect events on-line and in real time restricts these systems from realizing their full potential. One way to make these systems autonomous is to use some of the computer vision and image understanding techniques developed over the last few years to detect and analyse events in real-time. Another motivation is the fact that the amount of data transmitted across the network would make human inspection and assessment of events in the monitored area very difficult.

The video-object segmentation problem: The task of video-object segmentation is to identify and separate the important objects in a video scene from the scene background. Clearly, to approach this problem, it is necessary to define what is exactly meant with important objects and how the correct object masks should look like. However, in practice, it turns out that even an unambiguous definition of video objects is a fundamental problem. In the following, the involved definition problems are addressed and grouped into physical problems, being a consequence of the image formation, and semantic problems. The physical problems are as follows:

• *Reflections*: The problem of handling reflections is actually similar to object shadows. However, reflections are more difficult, because the appearance of the reflected images depends on the physical properties of the reflecting surface and because the reflection is not necessarily attached to the object. • Occlusions: The object shape can also change

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

because of occlusions. It depends on the application whether the masks of occluded objects should be extended to their original shape.

• *Translucent objects*: Objects can appear partially translucent since they are made of translucent materials, or because thin structures like hair or cloth appear translucent. Moreover, pixels along object boundaries are always a mixture of foreground colour and background colour. To model the translucency, the segmentation algorithm has to compute a 6 alpha-channel mask which identifies the translucency factor for each pixel instead of only computing a binary object mask. Accurate alpha-channel information cannot be obtained from a single image, but algorithms using heuristic approaches have been proposed.

• *Objects of interest (foreground objects):* The first and obvious question of video segmentation is what parts of an image constitute the foreground object. This issue is already surprisingly difficult, since the intuitive human understanding of foreground objects is strongly depending on the scene context. Mostly, human intuition expects that this should be the main acting objects. For example, in sports broadcast, the players are usually considered foreground and the audience is considered background, even if the audience is moving. This distinction is on a very high semantic level, since it assumes knowledge about the meaning of the scene. Note that the object definition can also vary with the application. A surveillance system in a sports stadium will be interested in other objects than a system for automatic analysis of the sports game.

• *Small background movements:* When taking a more detailed view on the last point, it can be observed that the distinction between foreground and background is in fact gradual. The question is to what extent a background should change such that it is considered part of the foreground. For example, trees may occur in the background with leaves moving slightly in the wind, or there may be a clock on a wall at the back of the room.

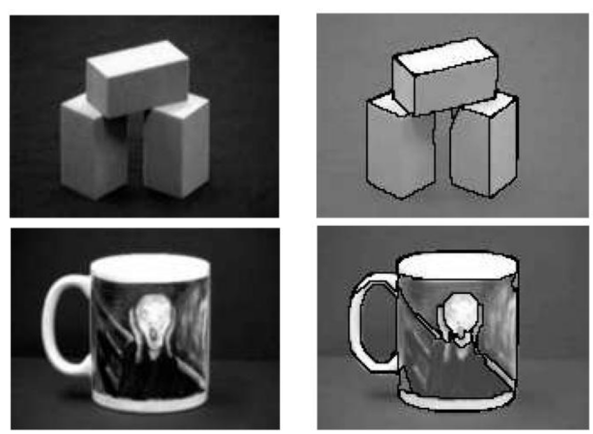
• *Object-status change:* Objects can also change their classification over time. For example, most people would consider a car that drives along a street as an important object. But how to define the object status when the car stops and parks at the side of the street? Alternatively, the opposite case may occur that a car that was parked for a long time suddenly drives away.
Note that it is practically impossible to separate all objects, including the static ones, into independent objects, since this would imply that all future actions would have to be predicted.

• *Multi-body objects:* Objects may be separated into several non connected regions in the image. One reason for this can be that an occluding object cuts the foreground object into pieces. Another complex example are objects that are really composed of several parts but still belonging together like flocking birds.

• *Hierarchical objects:* Additional to multi-body objects, there can also exist a hierarchical relationship between objects. One example is a car object that contains a driver object. When considering all of these problems simultaneously, it can only be concluded that a general-purpose segmentation of video objects is virtually impossible, since the definition of the expected output from the algorithm depends largely on the scene context and the application that we have in mind. However, despite all the mentioned problems, it is still possible to design algorithms that cover a multitude of specific applications and that work well in many practical cases.

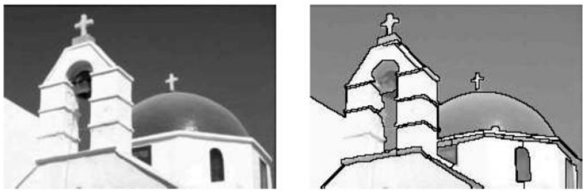
II. IMAGE SEGMENTATION

In designing automated systems for the interpretation or manipulation of image data, system developers often need to perform software imaging operations, called segmentation that extract information about the structure of objects and to separate and discern various parameters of interest within the data. Measurements or attributes of these objects, known as features, can then be calculated and used for defect inspection, quality control, or clinical qualitative analysis. Accordingly, common vision processes deal with the identification of discrete objects within an image. Such processes transform single-pixel representations of the image data into geometric descriptors representing groups of pixel elements.



ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013



(a) (b)

Figure 2.1 (a) Original Image, (b) Segmented Image

These descriptors, known as objects, take the form of points, lines, regions, polygons, or other unique representations.

Figure 2.2 Joint Segmentation

Segmentation techniques are divided into two basic categories: edge-based and region-based. Edge-based segmentation is primarily used to look for image discontinuities. The technique is generally applied where changes of gray-level intensity occur in the image. The assumption is that changes occur in the data at the boundary between objects of interest. The output of edge-segmentation schemes can be x and y gradient two images are used to represent the edges found, one in the x direction and one in the y direction

1. Gradient strength and direction

2. Binary edge map

3. Edge representation.

In contrast, region-based segmentation is used to look for similarities between adjacent pixels. That is, pixels that possess similar attributes are grouped into unique regions. The assumption is made that each region represents one object of interest. Using gray-level intensity is the most common means of assigning similarity, but many other possibilities exist, such as variance, colour, and multispectral features.

Most commercial vision systems use region-based segmentation schemes based on pixel-intensity values. These segmentation techniques assume that the objects of interest possess uniform shading and that a significant and constant gray-level change occurs between the objects of interest and the background. However, in many vision applications, these assumptions have proven erroneous. Therefore, these techniques are considered fragile and commonly require controlled conditions or human supervision.

Effects of uneven sample illumination, shadowing, partial occlusion, clutter, noise, and subtle object-to-background changes can all contribute to errors in basic segmentation processes. They generally result in false segmentations of the background, partial segmentations of the objects of interest, clumping of objects, or inadequate segmentations. Errors in the segmentation of the data can also result in the calculation of erroneous features. Therefore, it is essential that the segmentation method chosen support the final processing goals of the vision system.

Various types of Image Segmentation Techniques

• *Clustering methods:* The K-means algorithm is an iterative technique that is used to partition an image into *K* clusters. The basic algorithm is:

1. Pick *K* cluster centers, either randomly or based on some heuristic.
2. Assign each pixel in the image to the cluster that minimizes the variance between the pixel and the cluster center.
3. Re-compute the cluster centers by averaging all of the pixels in the cluster.
4. Repeat steps 2 and 3 until convergence is attained (i.e. no pixels change clusters).

• *Histogram-based methods:* Histogram-based [14] methods are very efficient when compared to other image segmentation methods because they typically require only one pass through the pixels. In this technique, a histogram is computed from all of the pixels in the image, and the peaks and valleys in the histogram are used to locate the clusters in the image.

• *Edge detection methods:* Edge detection is a well-developed field on its own within image processing. Region boundaries and edges are closely related, since there is often a sharp adjustment in intensity at the region boundaries. Edge detection techniques have therefore been used as the

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

base of another segmentation technique. The edges identified by edge detection are often disconnected. To segment an object from an image however, one needs closed region boundaries. Discontinuities are bridged if the distance between the two edges is within some predetermined threshold.

• *Region growing methods:* The first region growing method was the seeded region growing method. This method takes a set of seeds as input along with the image. The seeds mark each of the objects to be segmented. The regions are iteratively grown by comparing all unallocated neighbouring pixels to the regions. The difference between a pixel's intensity value and the region's mean, δ, is used as a measure of similarity. The pixel with the smallest difference measured this way is allocated to the respective region. This process continues until all pixels are allocated to a region.

III. Previous Work

In the literature, object detection modules are typically designed for specific classes of objects, such as faces and person/pedestrians. Face detection is performed using simple features based on Haar-wavelets [8]. Histogram of Oriented Gradients (HoG) descriptors were designed and effectively used for the task of pedestrian detection [9], which was later applied to generic object detection [10], [11].

In case of videos, the detection time is an important criterion. There are two common approaches to speeding up object detection: the first is a cascade based approach [8], while the other is by using random forests [12]. Zhu et al. [13] integrated classifier-cascades with HoG [9], while Vedaldi et al. [7] uses a cascade of classifiers with progressively increasing complexity, for generic object detection. The major drawback of cascade-based approaches is that one cannot recover from the mistakes committed early in the cascade. The second approach of random forests [12], [14], uses multiple decision trees which typically use simple decision functions at each node. However, they have the disadvantages of over-fitting and lack of a principled method to tune and improve their classification performance.

IV. Proposed Approach

In this paper the proposed approach works on images as well as video frames for object detection and gross size estimation. The frames extracted from the video or images are segmented first, features of each object in the segmented image are extracted on the image and consecutive frames having the desired features in the hand accordingly.

Segmentation-Based Modelling: Each segmented group should be further processed to build an appropriate geometric representation as an object.

• If a group is representing a regular geometric object, for instances, a plane, a polyhedron, or a cylinder, it is straightforward to use some standard methods of fitting these well-defined geometric models to the given data.

• If a group is representing a smooth surface like a human head or a compact object, we could use a level set approach that integrates all joint points, image information and the object boundaries to build an implicit surface model (Lhuillier and Quan, 2005). Alternatively, we used a graph-cut approach that builds a surface model with more details but higher computational cost.

• If a group is representing the specific hair of a given person, a combination of synthesis and analysis method could be used to reconstruct each hair fiber as a curve represented as a set of connected line segments by following the edge orientation in the images.

• If a group is representing an individual leaf of a plant, then we can build a generic leaf model for each plant, and we have developed a method of fitting a generic deformable model to the data in Quan et al. (2006).

For each segments, let r(s) denote the number of times this segment has been generated, across all randomized segmentations for the shape *Wi*. A natural approach would be to rank the segments by their repetition count r. However, this measure can overlook useful segments that are not salient on *Wi* and thus have low repetition count. Instead, we define a more global measure w_s of each segment's potential usefulness. The measure w_s takes into account not only the repetition count of s over the segmentations of *Wi*, but also the repetition counts of similar segments on other database shapes.

In order to define the measure w_s, we consider a shape distance measure *d(s, s')* that evaluates the geometric similarity of two segments. The distance d factors out anisotropic scale variations and is described in detail in Appendix A. Using this distance, we define w_s by considering the repetition counts of the most similar segments to s on all database shapes *{W1, . . . , Wn}:*

$$Ws = \sum_{j=1}^{n} W_{\left(s,s_j^*\right)} r\left(s_j^*\right),$$

Where $s_j^* = \arg min_{s' \epsilon \bar{X}_j} d(s, s')$ is the most similar segment to s in the set, and

$$W_{\left(s,s_j^*\right)} = \exp\left(-\frac{d^2(s,s')}{2\sigma^2}\right)$$

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

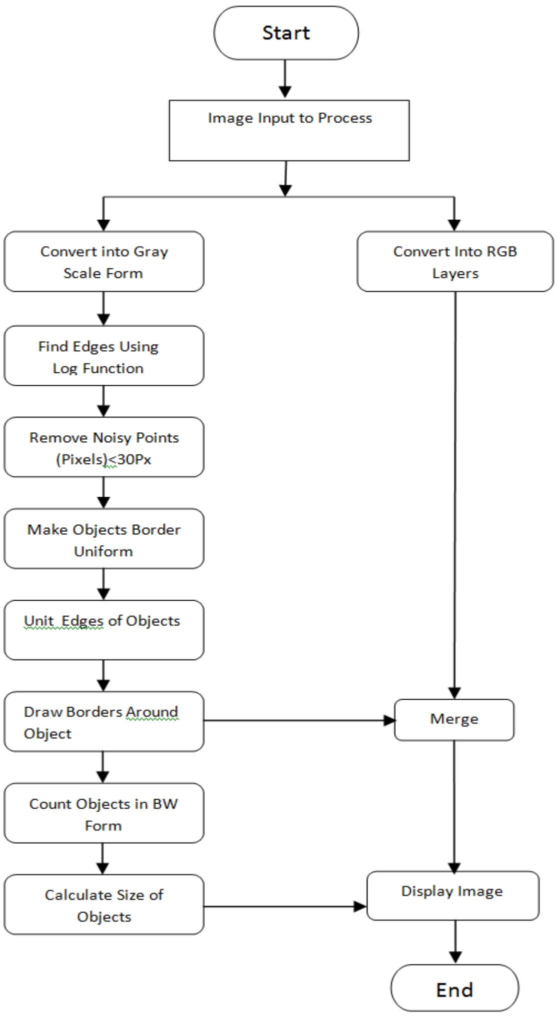


Figure 4.1 Flow chart for Image Object Detection and Gross Size Estimation

where σ is chosen as the median of distances between all pairs of most similar segments from all pairs of input shapes.

Object counting: The detected binary image of the object forms the input image for counting. This image is scanned from top to bottom for detecting the presence of an object. a variables is maintained i.e., count that keeps track of the number of objects.

The Flow chart of the proposed approach is shown in the Figure. 4.1 & 4.2

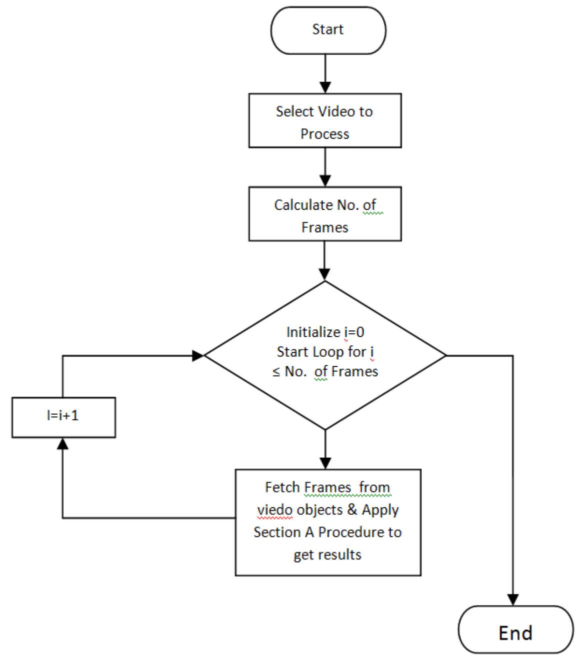
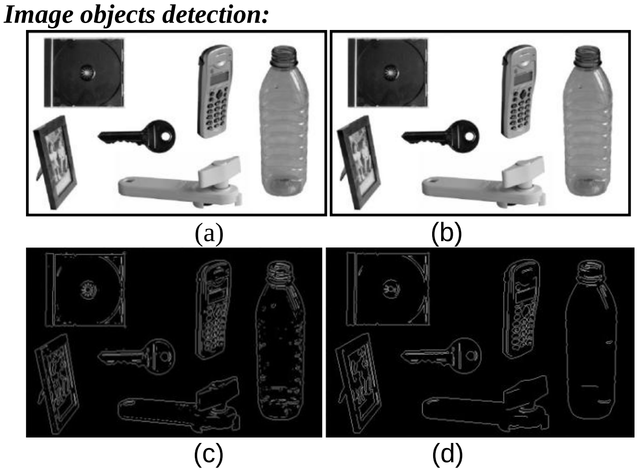


Figure 4.2 Flow chart for Video Object Detection and Gross Size estimation for each frame in the video.

V. SIMULATIONS & RESULTS

This system was implemented on an Intel Core i5 2.5 GHz PC. We have tested the system on image sequences on different scenarios like traffic junction intersection, highways etc. The entire processing requires approximately about 100 frames. Real life traffic video sequence is used to demonstrate the knowledge discovery process. All the videos chosen for vehicle tracking have same light intensity and have been taken during day time. We convert the colour video frames to gray scale images. Multimedia data mining techniques are used to count the number of vehicles passing through the road intersection in a given time duration.

Image objects detection:



(a) (b) (c) (d)

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

(e) (f)

Figure 5.1 The Detection and Gross Size Estimation Process, (a) Original Color Image (b) Greyscale Image (c) edges of the object in the image with small noisy spots due to light illumination (d) image (c) without noise, (e) Inverted Object Image (f) Show Objects and respective gross sizes with green edges and size in yellow.

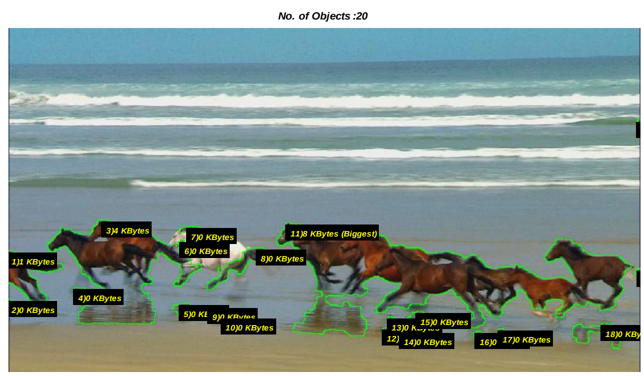
Video object detection:

(a) Original Color Image

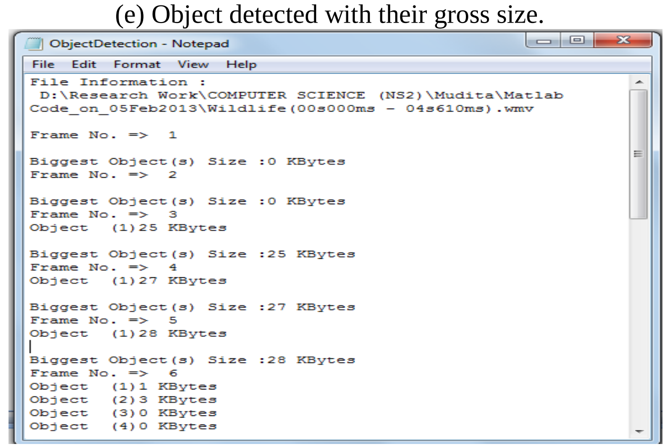
(b) Output of edge based algorithm with Noise

(c) Output of edge based algorithm without Noise

(d) Object detected using Joint Segmentation algorithm



(e) Object detected with their gross size.



```
File Information :
 D:\Research Work\COMPUTER SCIENCE (NS2)\Mudita\Matlab
Code_on_05Feb2013\Wildlife(00s000ms - 04s610ms).wmv

Frame No. =>  1

Biggest Object(s) Size :0 KBytes
Frame No. =>  2

Biggest Object(s) Size :0 KBytes
Frame No. =>  3
Object   (1)25 KBytes

Biggest Object(s) Size :25 KBytes
Frame No. =>  4
Object   (1)27 KBytes

Biggest Object(s) Size :27 KBytes
Frame No. =>  5
Object   (1)28 KBytes

Biggest Object(s) Size :28 KBytes
Frame No. =>  6
Object   (1)1 KBytes
Object   (2)3 KBytes
Object   (3)0 KBytes
Object   (4)0 KBytes
```

(f) Frame by frame details of the object detected

Figure 5.2 The Detection and Gross Size Estimation Process, (a) Original Colour Image (b) Greyscale Image (c) edges of the object in the image with small noisy spots due to light illumination (d) image (c) without noise, (e) Inverted Object Image (f) Show Objects and respective gross sizes with green edges and size in yellow.

VI. CONCLUSIONS

We have proposed an object detection algorithm for images and videos, based on image segmentation and gross size estimation of the segmented objects in images and between frames of a video in a simple feature space. Simulation

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

results for Image and frame sequences with edgebased approach and joint segmentation (JSEG) with gross size estimation, show that our proposed approach will suitable for security surveillance system, criminology, and for parking system. In order to extract colour features of segmented objects, we used the gray value at the center pixel of an object. Gray value turns out to sufficiently represent the object‘s colour features for the tracking purpose. A multi-colour object would be segmented into several parts by the joint segmentation algorithm.

ACKNOWLEDGMENT

I sincerely thank Dr. Bhupesh Gaur (Professor, CSE, Technocrat Institute of Technology, Bhopal) for his mentorship, support and guidance. Without his able guidance this research would not have taken the shape in its present form.

REFERENCES

[1] Badri Narayan Subudhi, Pradipta Kumar Nanda, *Member, IEEE,* and Ashish Ghosh, *Member, IEEE,* “Change Information Based Fast Algorithm for Video Object Detection and Tracking”, *IEEE Transaction On Circuit and System for Video Technology*,Vol.21,no.7,July 2011.

[2] Qingming haung, wen gao, wenjian cai, “Thresholding technique with adaptive window selection for uneven lighting image”, *Elsevier, Pattern recognition letters*, vol 26, page no 801-808, 2005.

[3] Junqiu wang [Member IEEE], yasushi yagi [Member IEEE],”Integrating colour and shape texture features for adaptive real time object tracking”, *IEEE transactions on image processing*, vol no 17,no 2, page no 235-240,2007.

[4] Oleg Michailovich, Member, IEEE, And Allen Tannenbaum, Member, IEEE,*”*Segmentation Of Tracking Sequences Using Dynamically Updated Adaptive Learning*”, IEEE Transactions on Image Processing*, Vol. 17, No. 12, Page No 2403-2412, December 2008.

[5] T. Cour, S. Yu, and J. Shi.” Normalized cuts matlab code”. Computer and Information Science, Penn State University.

[6] F.J. Estrada, A.D. Jepson, and C. Chennubhotla. “Spectral embedding and min-cut for image segmentation”. In *British Machine Vision Conference*, 2004.

[7] P.F. Felzenszwalb and D.P. Huttenlocher. “Efficient graph-based image segmentation”. *Int. J. of Comp. Vis.*, 59(2):167–181, 2004.

[8] J.Malik, S. Belongie, T. Leung, and J. Shi. “Contour and texture analysis for image segmentation”. *Int. J. of Computer Vision*, 43(1):7–27, 2001.

[9] D. Martin and C. Fowlkes. *The Berkeley Segmentation Dataset and Benchmark*.

[10] D. Martin, C. Fowlkes, and J. Malik.” Learning to detect natural image boundaries using local brightness, color, and texture cues”. *IEEE Trans. Pattern Anal. and Machine Intell.*, 26(5):530-549, 2004.

[11] D. Martin, C. Fowlkes, D. Tal, and J. Malik. “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics”. In *Proc. 8th Int’l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[12] J. Shi, C. Fowlkes, D. Martin, and E. Sharon. “Graph based image segmentation tutorial”. CVPR 2004. http://www.cis.upenn.edu/˜jshi/ Graph Tutorial/.

[13] J. Shi and J. Malik. “Normalized cuts and image segmentation”. *IEEE Trans. Pattern Anal. And Machine Intell.*, 22(8):888–905, 2000.

[14] S. Yu and J. Shi.” Multiclass spectral clustering”. In *Proc. Int’l Conf. Computer Vision*, 2003

BIOGRAPHY

Mudita Bhadania is a student of Master of Engineering in Computer Technology and Application, Technocrat Institute of Technology, Bhopal, Madhya Pradesh, India. Area of interest is multimedia mining.