International Journal of Computer Engineering Science (IJCES)
Volume 2 Issue 5 (May 2012) ISSN : 2250:3439
https://sites.google.com/site/ijcesjournal http://www.ijces.com/

Depth-Wise Segmentation of 3D Images Using Entropy

S. S. Mirkamali [1] and P. Nagabhushan [1],

[1] Department of Studies in Computer Science, University of Mysore, Mysore, India,
{s.s.mirkamali, p.nagabhushan}@hotmail.com

Abstract. Recent advances in 3D modeling and depth estimation of objects have created many opportunities for multimedia computing. Using depth information of a scene enables us to propose a brand new segmentation method called Depth-Wise segmentation. Unlike the conventional image segmentation problems which deal with surface-wise decomposition, the depth-wise segmentation is a problem of slicing an image containing 3D objects in a depth-wise sequence. The proposed method uses entropy of a depth image to characterize the edges of objects in a scene. Later, obtained edges are used to find Line-Segments. By linking the line-segments based on their object and layer numbers, Objects-Layers are achieved. To test the proposed segmentation algorithm, we use syntactic images of some 3D scenes and their depth maps. The experiment results show that our method gives good performance.

Keywords: Segmentation, Layering, 3D image, Depth image, Entropy.

1 Introduction

Recent advances in structure-from-motion [1] and stereoscopic vision [2, 3] have made it possible to create depth maps with a handheld camera. The increasing prevalence of depth cameras also implies that achieving high-quality depth images is becoming more and more convenient and flexible. For instance, Microsoft Kinect [23] sensors that provide real-time range estimates have been available at commodity prices. Therefore, it is not difficult to imagine that every captured image will have depth data in the future. These advances in depth estimation have created many opportunities for multimedia computing. Using depth information of a scene enables us to propose a brand new segmentation method called Depth-Wise segmentation. Unlike the conventional image segmentation problems which deal with surface-wise decomposition, the depth-wise segmentation is a problem of slicing an image containing 3D objects in a depth-wise sequence. Image and video segmentation has long been a fundamental problem in computer vision which would be useful in many applications, such as object recognition, image/video annotation, video stylization, and video editing. However, unsupervised segmentation is an inherently ill-posed problem due to the large number of unknown factors and the possible geometric and motion ambiguities in the computation. How to appropriately take advantage of the depth information for segmentation is becoming an important issue. To the best of our knowledge, it has not yet been thoroughly discussed in literatures, especially for depth-wise segmentation of a scene.

In this paper, we propose a novel depth-wise image segmentation method. The goal of this paper is to present a method to extract layers of an image preserving both the object's boundaries and their depth order using information available in the depth map of the input 3D image. The entropy measures the energy contribution of a line and it has been used to solve many image processing problems [4]. In this paper we use entropy of a depth image to characterize the edges of objects in a scene.

We assume that there is a depth map for each 3D image and objects are naturally and linearly separable. The depth data could be achieved by using a depth camera or multi-view stereo techniques [2, 5]. In our experiments, we made a 3D virtual scene with the 3D Studio Max software[1], and compute the depth maps by its built-in function. Our method contributes the following two aspects. First, we introduce a novel segmentation method which uses the concept of entropy to separate each row of the depth map into line-segments. Second, we improve the link perception method introduced in [24] to connect parts of the objects and make a complete object more accurately. The link perception method is also used to make object-layers based on the concept of the set theory. Experiment results demonstrate that our segmentation method, which uses the depth map of a 3D scene, can reliably extract objects and their corresponding layers.

2 Related Work

During the past decades, many state-of-the-art image segmentation methods have been proposed, such as thresholding [4, 6], mean shift [7], normalized cut [8], watershed algorithm [9]. Thresholding is undoubtedly one of the most popular segmentation approaches for the sake of its simplicity. It is based on the assumption that the objects can be distinguished by their gray levels. It is an important issue to find a correct gray level threshold that can separate different objects or separate objects from background.

However, the automatic selection of a robust, optimum threshold has remained a challenge in image segmentation. An early review of thresholding methods was reported in [4, 10]. Pun in [11] described a method that maximizes the upper bound of the posteriori entropy derived from the histogram. Wong and Sahoo's method [12] determines the optimum threshold by maximizing the posteriori entropy subject to certain inequality constraints that characterize the uniformity and shape of the segmented regions. Pal [13] developed another entropy-based method by considering the joint probability distribution of the neighboring pixels, which they further modified with a new definition of entropy. Tao et al's., method [14] selects the optimum three-level threshold based on probability partition, fuzzy partition and entropy theory. In their method, the image is segmented to three parts, including dark, gray and white

[1] 3D Studio Max is a graphical software that can virtually generate 3D objects using pre-captured multiple views of those objects with some user interactions.

parts and the fuzzy region is found by genetic algorithm based on the maximum fuzzy entropy principle.

A fast multilevel thresholding technique has been proposed by Yin [15]. The thresholds optimizing the Otsu's [16] or the Kapur's [6] functions are searched by using an iterative scheme. This technique starts from random initial thresholds. Then, these thresholds are iteratively adjusted to improve the value of the objective function. This improvement process stops when the value of the objective function does not increase between two consecutive iterations. The implementation of this method is similar to the one presented by Luo and Tian, where the Kapur's function is maximized by using the Iterated Conditional Modes (ICM) algorithm [17].

By incorporating the additional depth information and layer separation, robust segmentation is achieved with user interaction. For multi-object segmentation from the multi-view video of a dynamic scene, Reid et al., [17] proposed an algorithm that uses Maximum A Posterior (MAP) estimation to compute the parameters of a layered representation of the scene, where each layer is modeled by its motion, appearance and occupancy. The MAP estimation of all layer parameters is equivalent to tracking multiple objects in a given number of views. Expectation-Maximization (EM) is employed to establish the layer occupancy and visibility posterior probabilities. Some other methods formulate the problem of multi-view object segmentation based on layer separation using Epipolar Plane Image (EPI) [18], a volume constructed by collecting multi-view images taken from equidistant locations along a line. Unfortunately, these methods suffer from an important limitation: all of them classify moving objects as the first layer of a scene and other objects are left as background layer.

In summary, many approaches have been proposed to segment a gray scale image into multiple levels based on objects gray levels. However, the problem of how to properly extract the layers based on objects and their depth value remains untouched. In this paper, we propose using depth map in finding the multiple layers of a 3D scene by incorporating a technique for segmentation based on entropy and a link perception algorithm to make object-layers.

3 Proposed Depth-Wise Segmentation Method

Suppose that a 3D image I having a depth map D with L gray levels $L=\{0, 1, \ldots, L\text{-}1\}$, containing the *depth* value of an image I in every pixel coordinate, is to be partitioned into m object-layers $OL=\{ol_1, ol_2, \ldots, Ol_m\}$.

The proposed segmentation method is based on Shannon entropy [19] and a simple link perseption algorithm. It allows the determination of the number of layers as well as appropriate object-layer values.

3.1 Edge Detection

Edge detection is a method to identify points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. In a depth map, a sharp change shows a change between two objects in two different layers. We use this phenomenon to identify the objects in different layers of a 3D image.

International Journal of Computer Engineering Science (IJCES)
Volume 2 Issue 5 (May 2012) ISSN : 2250:3439
https://sites.google.com/site/ijcesjournal http://www.ijces.com/

To generate the edges of the objects Canny edge detection technique [20] is used because of it is a robust and qualified technique [21]. We feed a depth map *D* as an input into the edge detector and the result is a binary image *BI*, which is shown in Fig 1. b.

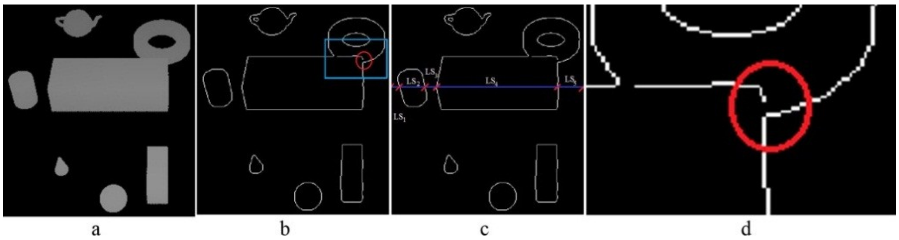

Fig. 1. (a) depth map of a 3D image (b) result of canny edge detection (c) showing a row with 5 line-segments ($LS_1 ... LS_5$) (d) zoomed into blue box of image b.

3.2 Entropy Based Line Division

Edges computed by the edge detection method are fed into the line division section. This section characterizes the number of objects and the layers independently. In this method we use an entropy-based algorithm to extract the line-segments.

The entropy of a system was first defined by Shannon [19] and used to solve many image processing and pattern recognition problems. He defined the entropy of an *n*-state system as

$$H = -\sum_{i=1}^{n} p_i \log(p_i) \tag{1}$$

where p_i is the probability of occurrence of the event *i* and

$$\sum_{i=1}^{n} p_i = 1, \quad 0 \le p_i \le 1. \tag{2}$$

Conventional Entropy Quantifier measures the energy contribution of each row by considering the problem occurrence of *+ve* and *-ve* transitions among the total number of pixels in each row in horizontal direction and quantifies the entropy as

$$E(t) = p \log\left(\frac{1}{p}\right) + (1-p) \log\frac{1}{1-p} \tag{3}$$

where *t* is the *0-1* and *1-0* transitions, *p* represents number of times *0-1* and *1-0* transitions occur in each row and *1-p* represents the non-probable occurrence of transitions.

Depending upon the transition (*0-1* or *1-0*) as described above, *E(t)* could be $E^-(t)$ or $E^+(t)$ so that we could define a total entropy along a row i as

$$E(i)=E^-(t) + E^+(t) \tag{4}$$

where $E^-(t)$ is the entropy due to the transitions from *1* to *0* and *E+(t)* is the entropy due to the transitions from 0 to 1 [25].

A line-segment, $LS_{i,j}$, is defined as a gap between every *-ev* transition followed by a *+ve* transition. In which, every *-ve* transition is considered as the starting point of a line-segment and a *+ve* transition is considered as the ending point of the line-segment $LS_{i,j}$. Fig.1(c) shows an example of *5* extracted line-segments for a row *i*.

We assigned two attributes to a line-segment. The first is object number. We define an object-number of a line-segment as

$$LS_{i,j}.\mathrm{ON} = \begin{cases} LS_{i-1,k}.\mathrm{ON} & if\ \mathrm{LS}_{i,j}.\mathrm{ON} \cap \mathrm{LS}_{i-1,k}.\mathrm{ON} \neq \emptyset \\ & and\quad |V_{i,j} - V_{i-1,k}| \leq T \quad k = 1 \dots p \\ oc + 1 & otherwise \end{cases} \tag{5}$$

where *i* is the row number of the depth map *D* in which a line-segment belongs, *j* is a number of a line-segment in i^{th} row, *k* is a number of a line-segment in $(i\text{-}1)^{th}$ row, *p* is the number of line-segments in $(i\text{-}1)^{th}$ row, *T* is object connectivity value which is calculated manually, *oc* is the number of available object values and

$$V_{i,j} = mode\big(\,D(i,q)\big) \quad q = s_{i,j}\ to\ e_{i,j}. \tag{6}$$

where *s* and *e* are starting and ending points of a line-segment respectively.

The second feature of a line-segment is its layer-number. A layer-number of a line-segment is defined as

$$LS_{i,j}.LN = \begin{cases} \mathrm{k} & if\ V_{i,j} = V_k \quad \text{for k} = 1 \dots \mathrm{lc} \\ V_{i,j}\ \text{and}\ \mathrm{lc} + 1 & otherwise \end{cases} \tag{7}$$

where *lc* is denoting the number of available layers and V_k is the depth value of the k^{th} layer (L_k).

3.3 Link Perception

The idea behind the proposed linking method is based on two assumptions. First, every object should appear only in one object-layer completely. Second, all the objects in a layer are in one object-layer.

The line segmentation algorithm may divide an object into a few parts. In the worst case, there is a possibility of associating an object with different layers. To overcome these problems and also to satisfy assumptions, we drive a straightforward algorithm based on set theories.

The first step in this algorithm is to link all the parts of an object as well as all the layers to which the object belongs and make a compound object *CO* using conditions (8).

$$\begin{cases} CO_l.\mathrm{LN} = CO_l.\mathrm{LN} \cup LS_{i,j}.\mathrm{LN} & if\ LS_{i,j}.\mathrm{ON} = CO_l.ON \\ \begin{cases} CO_{l+1}.\mathrm{LN} = LS_{i,j}.\mathrm{LN} \\ CO_{l+1}.\mathrm{ON} = LS_{i,j}.\mathrm{ON} \end{cases} & otherwise \end{cases} \tag{8}$$

where *CO.LN* and *CO.ON* denote the layer number and object number of a compound object *CO* respectively.

The second step is to link all the compound objects which are in the same layer to construct an object-layer using the following condition.

$$OL_m = OL_m \cup CO_l \quad if\ CO_l.\mathrm{LN} \cap OL_m.\mathrm{LN} \neq \emptyset \tag{9}$$

where l is the compound object number and m is an object-layer number. The novelty of this algorithm compare to [24] lays on consideration of line-segments as blocks instead of pixel-wise calculation of object-layer numbers. In this way the speed and accuracy of the algorithm increases.

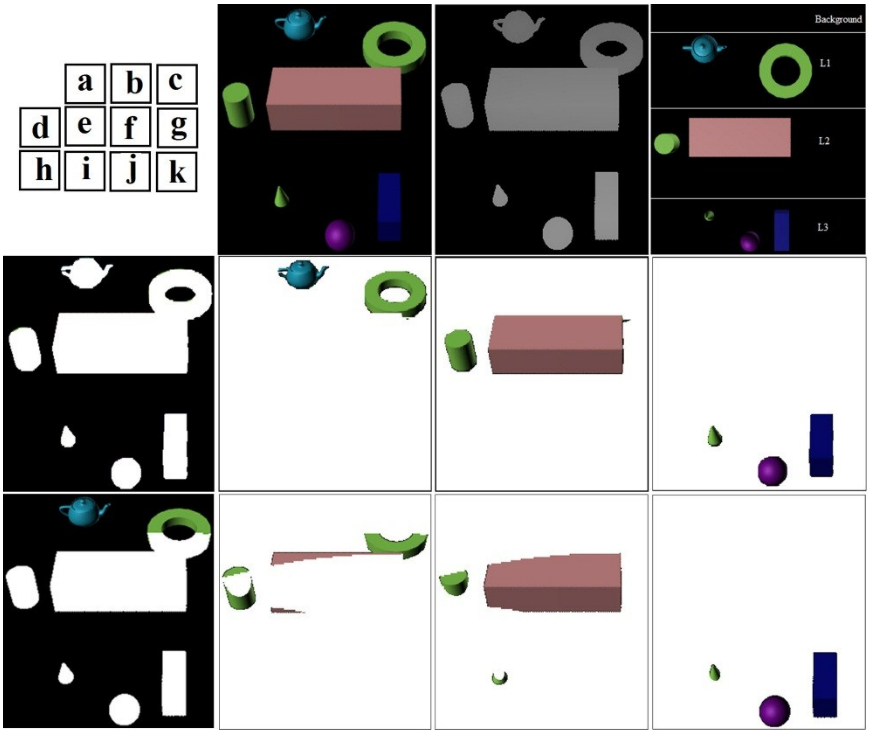

Fig. 2. Well assembled clutter of a scene. (a) Original 3D image; (b) depth map of the given 3D image; (c) top view of the 3D scene; (d-g.) result layers using the proposed method; (h-k) result layers using Multi Otsu method.

4 Results and Discussion

To verify the efficiency of the method, experiments have been carried on many 3D images along with their depth map. We used artificial 3D images and their depth map which have been created using 3D Studio Max. In all the experiments, we fixed the size of all the 3D images as well as their depth maps as 640x480 pixels with the resolution of 96 pixels/inch.

In this paper, we use two critical images to illustrate the efficiency of the method. Figs. 2 and 3 show the results, with (a) showing the original 3D image, (b) showing grayscale image maps (darker colors representing larger z values), (c) showing the top view of the 3D scene to compare the extracted layers visually, (d-h) showing the result using the proposed method with automatically calculated connectivity value of T on the basis of the number of changes in depth ranges of a depth image. Figures 2 and 3 (h-k) showing the results using Multi Otsu method proposed in [22].

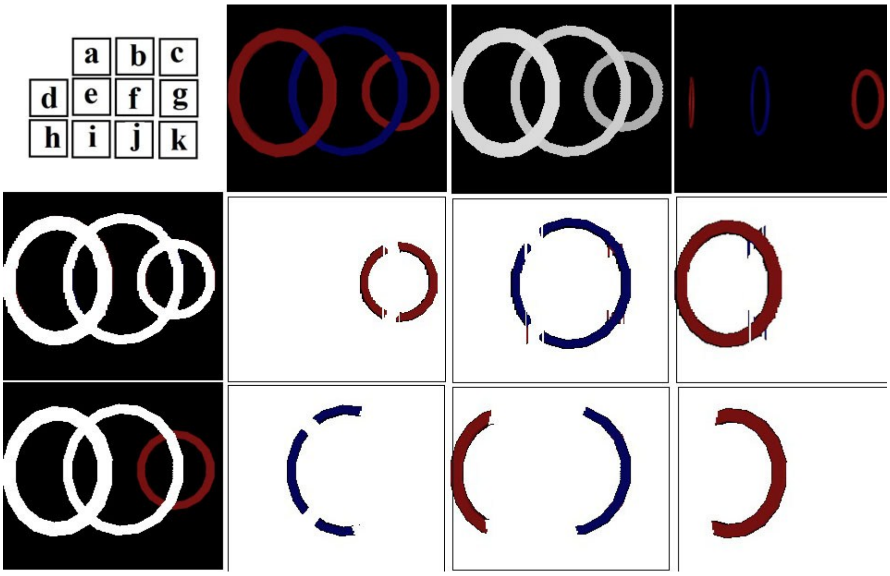

Fig. 3. Three separate rings looks connected. (a) Original 3D image; (b) depth map of the given 3D image; (c) top view of the 3D scene; (d-g.) result layers using the proposed method; (h-k) result layers using Multi Otsu method.

Fig. 2(a) is a 3D image of a well assembled clutter of objects in which some of them are occluded and positioned in different layers. Fig. 2(b) is the depth map of the given 3D image and Fig. 2(c) is the top view of the scene used to identify the exact number of layers. Figs. 2(d-g) are extracted layers of the 3D image. It is shown in these images that the layers are correctly extracted and objects are almost complete with just an small error in boundaries of the cube in the second layer using our method (Fig. 2(e)) which is just because of the broken edges caused by week edge detection approach (see Fig. 1(d)). Figs. 2(h-k) are results using Multi Otsu method. It is clearly visible that the method could not extract the layers truly and objects are broken apart.

Fig. 3(a) is showing another 3D image containing three rings every one of which is occluded by one or two other rings. This image has been created to show how robust the proposed method is against the occlusion. The depth map, top view of the scene is given in Figs. 3(c). In Figs. 3(d-g), we can see that the main features of all the three rings are well preserved and all of them are well classified into their corresponding layers using the proposed method. There are also some small parts of objects which are miss-classified in incorrect layers. In the last row, Figs 3(h-k) show that the Multi Otsu method could not extract the layers and mixed up all the three rings in different layers. More results can be found in Fig. 4.

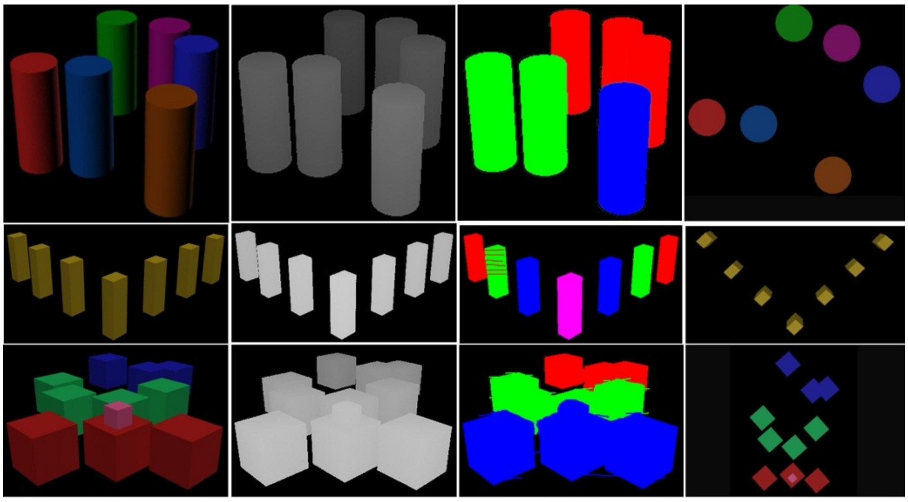

Fig. 4. More results of the proposed method. From left to right in every column, Original 3D images, depth map of the given 3D image, Results of the proposed segmentation method and top view of the 3D scenes.

5 Conclusion

In this paper, we define a new direction of image segmentation by using a depth map of a 3D image. The method is capable of layering a 3D image into multiple layers based on the position of objects in the scene with respect to the camera. An entropy based algorithm was used to divide a depth map into line-segments. After assigning an object number and a layer number to every line-segment, we employ a linking algorithm to merge divided objects and make the completed objects. Later, all the objects in a layer were linked to make object layers. We have conducted experiments on challenging examples in which the scene contained occluded objects. Results show that our proposed method gave good performance.

Future work will focus on finding appropriate methods detecting edges taking joints of connected objects into account. Furthermore, ongoing work is focused on finding a method capable of estimating the object connectivity value automatically.

6 References

1. Bolles, R.C., Baker, H.H., Marimont, D.H., Koenderink, J., Van Doorn, A., Shashua, A., Beardsley, P.A., Zisserman, A., Murray, D.W., Faugeras, O.: structure from motion. International Journal of Computer Vision 1, 7-55 (1987)
2. Guofeng, Z.: Consistent Depth Maps Recovery from a Video Sequence. IEEE Transactions on pattern analysis and machine intelligence 31, 974-988 (2009)
3. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47, 7-42 (2002)

4. Pal, N.R., Pal, S.K.: A review on image segmentation techniques. Pattern recognition 26, 1277-1294 (1993)
5. Kang, S.B., Szeliski, R.: Extracting view-dependent depth maps from a collection of images. International Journal of Computer Vision 58, 139-163 (2004)
6. Kapur, J., Sahoo, P., Wong, A.: A new method for gray-level picture thresholding using the entropy of the histogram. Computer vision, graphics, and image processing 29, 273-285 (1985)
7. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on pattern analysis and machine intelligence 603-619 (2002)
8. Shi, J., Malik, J.: Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22, 888-905 (2000)
9. Haris, K., Efstratiadis, S.N., Maglaveras, N., Katsaggelos, A.K.: Hybrid image segmentation using watersheds and fast region merging. Image Processing, IEEE Transactions on 7, 1684-1699 (1998)
10. Sahoo, P.K., Soltani, S., Wong, A.: A survey of thresholding techniques* 1. Computer vision, graphics, and image processing 41, 233-260 (1988)
11. Pun, T.: A new method for grey-level picture thresholding using the entropy of the histogram. Signal processing 2, 223-237 (1980)
12. Wong, A.K.C., Sahoo, P.: A gray-level threshold selection method based on maximum entropy principle. Systems, Man and Cybernetics, IEEE Transactions on 19, 866-871 (1989)
13. Pal, N.R., Pal, S.K.: Entropy: a new definition and its applications. Systems, Man and Cybernetics, IEEE Transactions on 21, 1260-1270 (1991)
14. Tao, W.B., Tian, J.W., Liu, J.: Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm. Pattern Recognition Letters 24, 3069-3078 (2003)
15. Yin, P.Y., Chen, L.H.: A fast iterative scheme for multilevel thresholding methods. Signal processing 60, 305-313 (1997)
16. Otsu, N.: A threshold selection method for grey level histograms. IEEE Trans. Syst. Man Cybern 9, 62-66 (1979)
17. Xiping, L., Jie, T.: Multi-level thresholding: maximum entropy approach using ICM. Proceedings 15th International Conference on Pattern Recognition., vol. 3, pp. 778-781 vol.773 (2000)
18. Criminisi, A., Kang, S.B., Swaminathan, R., Szeliski, R., Anandan, P.: Extracting layers and analyzing their specular properties using epipolar-plane-image analysis. Computer vision and image understanding 97, 51-85 (2005)
19. Shannon, C.E.: A mathematical theory of communication. Bell Syst.Tech. J. 27, 370-423 (1948)
20. John, C.: A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence 8, 679-698 (1986)
21. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital image processing using MATLAB. Pearson Education India (2004)
22. Liao, P.S., Chen, T.S., Chung, P.C.: A fast algorithm for multilevel thresholding. Journal of information science and engineering 17, 713-728 (2001)
23. Wenjun Zeng,.: Microsoft Kinect Sensor and Its Effect. IEEE Multimedia 2, vol.19, pp. 4-10
24. S. S. Mirkamali and P. Nagabhushan, "*Depth-Wise Segmentation of a 3D Image Using its Depth Map: An Entropy-Based Approach*", 5th Indian International Conference on Artificial Intelligence (IICAI-11), Dec 2011, Tumkur, India .
25. D.Gowda, S. 2009. Equivalence Amongst Document Images at Multiple levels Through Hierarchial Structural Attributes. PhD, University of Mysore.