Program

Preface

3D is expected to be the next big thing in audio-video entertainment. This is a logical continuation of a longer historical development in which video content representation has been continuously evolving to offer a greater realism and a more pleasant experience. First, it was colour, then it was high-definition. Now, it seems the right time to introduce the 3rd dimension.

A 3DTV system would include the usual main functional components: 3D scenery capture, processing the captured data for storage and transmission, and displaying the result for creating 3D visual sensation. The development of each of these components is a challenging task and their seamless integration is even more challenging as the ultimate goal is to deliver full-scale, fascinating and very realistic 3D TV service to the consumers.

3DTV-CON series of conferences has the ambition to bring together top researchers and developers from academia and industry with diverse experience and activity in distinct, yet complementary, areas to discuss the development of next generation 3DTV technologies, applications and services. The conference originated from the FP6 3DTV Network of Excellence and has gotten continuing support from FP6/FP7 projects working in the area of 3D Media. The first three conferences were held in Kos Island, Greece (2007), Istanbul, Turkey (2008), and Potsdam, Germany (2009). This year we have the 4th edition, the 3DTV Conference 2010, to be held on June 7-9 2010, in Tampere, Finland. The conference is organized by the Department of Signal Processing, Tampere University of Technology and Tampere International Center for Signal Processing (TICSP) and sponsored by the Federation of Finnish Learned Societies, Nokia, and the IEEE SP/CAS Chapter of Finland.

We have selected Tampere as the conference venue as the city is world-wide known as a research and innovation center in the field of digital multimedia. It has been named the most desirable place to work, live and study in Finland. The conference venue offers a peaceful experience of Finnish forests and lakes yet walking distance away from the effervescent city center thus being the ideal place for a scientific event.

As with the previous conferences, we aim at offering a reach technical program comprising plenary talks, tutorials, special and regular sessions so to stimulate discussions and interaction between conference participants. A wide range of disciplines including imaging and computer graphics, signal processing, telecommunications, electronics, optics and physics have been addressed in the technical contributions. We truly hope that the technical program accompanied by pleasant social gatherings will be of full benefit of the participants and will stimulate their future research.

Tampere, June 2010

Atanas Gotchev

Levent Onural

Tuesday, 8 June 2010

9:00-17:00 — **Registration**

9:00-10:15 — **Tutorial II: Challenges and Solutions for Immersive 3D Video Communication**
Chair: Gozde Bozdagi Akar
Oliver Schreer
Fraunhofer Institut for Telecommunications, Heinrich-Hertz-Institut, Berlin, Germany

10:15-10:30 — **Coffee break**

10:30-12:00 — **Tutorial II: Challenges and Solutions for Immersive 3D Video Communication**
Chair: Gozde Bozdagi Akar
Oliver Schreer
Fraunhofer Institut for Telecommunications, Heinrich-Hertz-Institut, Berlin, Germany

12:00-13:00 — **Lunch**

13:00-14:00 — **Keynote: Recent R&D Activities on 3DTV Broadcasting in Korea**
Chair: Levent Onural
Jinwoong Kim
Electronics and Telecommunications Research Institute (ETRI), Korea

14:00-14:20 — **Coffee break**

14:20-16:00

3D Quality of Experience
Chair: Moncef Gabbouj

- *Evaluation Of Depth Compression And View Synthesis Distortions In Multiview-video-plus-depth Coding Systems*
  Noha El-Yamany; Kemal Ugur; Miska Hannuksela; Moncef Gabbouj
- *Reduced-reference Quality Metric For 3d Depth Map Transmission*
  Chaminda T. E.R. Hewage; Maria G. Martini
- *Perceptual Video Quality Metric For 3d Video Quality Assessment*
  Prince Joveluro; Hossein Malekmohamadi; Anil Fernando; Ahmet Kondoz
- *Impact Of Depth Map Spatial Resolution On 3d Video Quality And Depth Perception*
  Gokce Nur; Safak Dogan; Hemantha Kodikara Arachchi; Ahmet Kondoz
- *Simulator Sickness - Five Experiments Using Autostereoscopic Mid-sized Or Small Mobile Screens*
  Satu Jumisko-Pyykkö; Timo Utriainen; Dominik Strohmeier; Atanas Boev; Kristina Kunze

View synthesis/rendering
Chair: Yo-Sung Ho

- *Depth Image Based Rendering: A Faithful Approa*
  Michael Schmeing; Xiaoyi Jiang
- *Biological-aware Stereoscopic Rendering In Free V*
  Sammy Rogmans; Maarten Dumont; Gauthier L
- *Alpha Model Based Mixed Pixel Processing For Vie*
  Hariprasad Kannan; Kiran N Iyer; Kausik Maiti;
- *Hybrid Motion/depth Oriented Inpainting For Virtu*
  Kuan-Yu Chen; Pei-Kuei Tsung; Pin-Chih Lin; Hs
- *User Directed View Synthesis On Omap Processo*
  Mursel Yildiz; Gozde Bozdagi Akar

16:00-17:30

Poster session: 3D Applications

- *An Experiment Study Of Presence Perception in 3DTV Program*
  Sang-Hee Kweon; Eun-Joung Cho; Byung-Chul Cho; Kyung-Ho Whang
- *Feasible Mid-air Virtual Reality With The Immaterial Projection Screen Technology*
  Ismo Rakkolainen
- *A Comparison Between Two 3d Free-viewpoint Generation Methods - Player-billboard And 3d Reconstruction -*
  Tetsuya Shin; Nozomu Kasuya; Itaru Kitahara; Yoshinari Kameda; Yuichi Ohta
- *Seam Carving For Stereo Images*
  Kei Utsugi; Takuma Shibahara; Takafumi Koike; Keita Takahashi; Takeshi Naemura
- *Adaptive Parallax For 3d Television*
  Kai Ide; Thomas Sikora
- *Experienced Audiovisual Quality For Mobile 3d Television*
  Timo Utriainen; Satu Jumisko-Pyykkö
- *A Novel Method For Automatic 2d-to-3d Video Conversion*
  Youwei Yan; Qionghai Dai; Feng Xu; Xiaodong Liu

Poster session: 3D processing

- *Multi-view Image Denoising Based On Graphical Model Of Surface Patch*
  Zhou Xue; Jingyu Yang; Qionghai Dai; Naiyao Zhang
- *3d Shape Measurement System Based On Structer Light And Polarization Analysis*
  Piotr Garbat; Marek Sutkowski
- *Real-time Depth Map Generation Architecture For 3d Videoconferencing*
  John Congote; Iñigo Barandiaran; Javier Barandiaran; Tomas Montserrat; Julien Quelen
- *Optimizing The Disparity Map By The Integration Of Hvs Binocular Properties For Efficient Coding Of Stereoscopic Color Images*
  Rafik Bensalma; Mohamed-Chaker Larabi
- *Impact Of Downsampling Ratio In Mixed-resolution Stereoscopic Video*
  Payman Aflaki ; Miska M. Hannuksela; Jukka Häkkinen; Paul Lindroos; Moncef Gabbouj

Poster session: 3D scene reconstruction

- *Real-time Free Viewpoint Image Rendering By Using Fast Multi-path Dynamic Programming*
  Norishige Fukushima; Toshiaki Fujii; Yutaka Ishibashi; Tomohiro Yendo; Masayuki Tanimoto
- *Rapid Radiometric Enhancement Of Colored 3d Point Clouds Using Color Balancing*
  Ulas Yilmaz; Olaf Hellwich
- *Improved 3d Video Synthesis Combining Graph Cuts And Chroma Key Technology*
  Kosmas Dimitropoulos; Theodoros Semertzidis; Nikos Grammalidis
- *Coherent Grouping Of Pixels For Faster Shadow Cache In 3d Holoscopic Computer Graphics*
  Amar Aggoun

Poster session: Depth estimation

- *Improvement Of Segment-based Depth Estimation Using A Novel Segment Extraction*
  Gi-Mun Um; Gun Bang; Won-Sik Cheong; Namho Hur; Soo In Lee
- *Fast Iterative Motion And Disparity Estimation Algorithm For Multiview Video Coding*
  Zhi-Pin Deng; Yui-Lam Chan; Ke-Bin Jia; Chang-Hong Fu; Wan-Chi Siu
- *A Simple And Efficient Way To Compute Depth Maps For Multi-view Videos*
  Gael Sourimant
- *Depth Map Estimation From Single-view Image Using Object Classification Based On Bayesian Learning*
  Jae-Il Jung; Yo-Sung Ho

DEPTH MAP ESTIMATION FROM SINGLE-VIEW IMAGE USING OBJECT CLASSIFICATION BASED ON BAYESIAN LEARNING

Jae-Il Jung and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro, Buk-gu, Gwangju, 500-712, Korea
Telephone: +82-62-715-2258, Fax: +82-62-715-3164
E-mail: {jijung, hoyo}@gist.ac.kr

ABSTRACT

Generation of three-dimensional (3D) scenes from two-dimensional (2D) images is an important step for a successful introduction to 3D multimedia services. Among the relevant problems, depth estimation from a single-view image is probably the most difficult and challenging task. In this paper, we propose a new depth estimation method using object classification based on the Bayesian learning algorithm. Using training data of six attributes, we categorize objects in the single-view image into four different types. According to the type, we assign a relative depth value to each object and generate a simple 3D model. Experimental results show that the proposed method estimates depth information properly and generates a good 3D model.

Index Terms **—** 2D-to-3D conversion, Depth estimation, Monocular depth cues, 3D scene generation, Single-view image

1. INTRODUCTION

Although two-dimensional (2D) images are successfully exploited in various multimedia services nowadays, interest on three-dimensional (3D) images is increasing rapidly and 3D image processing techniques are attracting more attention. The 3D image processing technology includes a wide range of different operations from 3D scene acquisition to 3D display. Among them, 3D contents generation is one of the most essential parts for the 3D image service.

In order to capture a 3D scene, we need special equipments, such as stereo or multi-view cameras and a depth camera [1]. Even if 3D image contents have been produced and become available, the amount of 3D contents is not enough to satisfy the user demand yet. On the other hand, there are abundant 2D image contents captured by conventional single-view cameras. Hence, generation of 3D scenes from 2D contents can be an alternative solution to overcome the current discrepancy and fill up the lack of 3D image contents.

However, it is not straightforward to generate a 3D scene from a single-view image since we lost some 3D information when capturing a real scene with a single-view camera. The 3D information includes the distance information between objects in the 2D image and the camera. The distance information of each pixel in the 2D image from the camera is called as the depth value, and the matrix of depth values for all the pixels in the 2D image is called as the depth map of the image.

Since accuracy of the depth map strongly affects the quality of the generated 3D scene, depth estimation plays an important role in 2D-to-3D conversion. In general, it is very challenging to obtain an accurate depth map from a single-view image. If we have multi-view images captured by two or more cameras, we can estimate the depth map using stereo matching algorithms. However, it is much more difficult to estimate a depth map from a single-view image because there is no additional information, such as camera parameters and disparity information. Therefore, we can only estimate relative depth values by analyzing monocular depth cues in the single-view image.

Recently, there are several proposals to estimate the depth map from the single-view image. S. Batiato *et al.* generated a depth map in the following steps: generation of gradient planes, depth gradient assignment, consistency verification of detected region, and final depth map generation [2]. J. Ko *et al.* proposed an automatic conversion method based on the degree of focus of segmented regions and generated a stereoscopic image [3]. They utilized higher-order statistics to check the degree of focus. S. A. Valencia *et al.* presented a depth estimation method by measuring focus cues, which consists of a local spatial frequency measurement using multi-resolution wavelet analysis and Lipschitz regularity estimation of significant edges [4]. Tam *et al.* found that the most critical depth information tends to be concentrated at object boundaries and image edges [5]. They generated the depth map in a single-view image using the Sobel edge detector. Chang *et al.* explored the motion by a frame difference method, and used the K-means algorithm to realize color segmentation; thus, the depth map was acquired from both time and spatial information [6]. Derek Hoiem *et al.* proposed the learning method to generate 3D models [11]. Their model is made up of texture-mapped planar billboards based on several labels.

However, previous works simply assigned depth values to all the pixels in the image using the same algorithm without considering different object types. Since images contain various types of objects, we propose a new depth estimation method considering object types in the single-view image. Our main contribution is that we define four different object types and six effective attributes to describe object units, and classify each object using the Bayesian classifier based on the training data. According to the object type, we assign relative depth values in different ways.

2. MONOCULAR DEPTH CUES

Depth perception arises from a variety of depth cues, and the depth cues are typically classified into two types according to the number of required eyes. Binocular depth cues that require input from both eyes include stereopsis and convergence. Monocular depth cues require an input from one eye. Only monocular depth cues exist in a single-view image and they make people perceive depth in 2D images.

There are various types of monocular depth cues. When there are two objects of the same size, we can measure the relative distance from their relative sizes. The object which subtends the larger visual angle appears closer. When an observer moves, the apparent relative motion of several stationary objects against the background gives hints about their

978-1-4244-6378-7/10/$26.00 ©2010 IEEE

relative distance. If information about the direction and velocity of movement is known, motion parallax can provide absolute depth information. Occlusion of objects by others is also a clue which provides information about relative distance. This information only allows the observer to know a ranking of relative nearness.

Among the monocular depth cues, linear perspective is one of the very powerful cues. Lines that are parallel in the 3D world appear to get closer together as they recede in the distance. The fact helps us figure out the distance between two objects. It also induces relative size, motion parallax, and text gradient. In this paper, we focus on linear perspective to estimate a depth map from a single-view image.

3. OBJECT CLASSIFICATION

In our approach, the types of input images are limited, because we estimate a depth map based on the linear perspective depth cue. The input constraint is that an input image should contain the vanishing point and be an outdoor scene. Before object classification, the vanishing point of an input image is detected by extracting edge components and finding the most overlapped point of their extended lines [10]. Only straight lines extracted by Hough transform are considered as candidates.

Then, the input image is divided into segments by the mean shift algorithm [9] as shown in Fig. 1(b). This process induces an effect that the boundaries of each object become distinctive. Segments are merged into object units with the grow-cut algorithm which is the manual image segmentation algorithm [7] as shown in Fig. 1(c). These objects will be used as basic units for the proposed algorithm.


(a) Original input image


(b) Segmented image

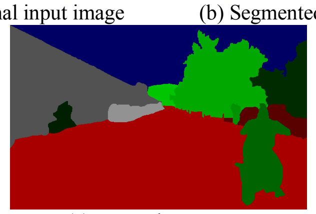

(c) Merged segments

Figure 1. Input image and segmentation results

3.1 Object Type

A real photograph can contain various objects, such as a building, a human, a car, and so on. Conventional depth estimation algorithms do not consider object types and handle them with the same method. However, it is unsuitable because it disregards object's own properties. In this paper, we divide objects into four types: SKY, GROUND, PLANE, and CUBIC. As you easily know through the names, the SKY and the GROUND types actually mean the sky and the ground in the world. The PLANE type stands for the object facing the perpendicular direction to the camera ray, and has a constant depth value. Examples of the PLANE type are a human, a tree, and so on. The CUBIC type is regarded as the object having the different depth values according to the distance from the vanishing point, and includes a building, a wall, and so on. Examples of each type are shown in Fig. 2.

Figure 2. Examples of object types:
SKY, GROUND, PLANE, and CUBIC

3.2 Attributes

Before object classification, we describe features of the object types. Six attributes are defined and will be used as the criteria of classification. Table 1 lists the attributes and their elements. The proposed algorithm automatically describes whole objects in an image with these six attributes. In this section, we introduce each attribute and how each element is selected in detail.

Table 1. Attributes and their elements

notation	attribute	elements
a_1	Horizon	<contact, include, none>
a_2	Vanishing Point	<include, none>
a_3	Vertical Line	<include, none>
a_4	Boundary	<top, bottom, left, right, none>
a_5	Complexity	<HH, HL, LH, LL>
a_6	Object Size	<HH, HL, LH, LL>

The horizon attribute a_1 describes the relation between the horizontal line and an object. The horizon acts an important role to distinguish the SKY and GROUND types from other types. The object including the horizon has a very low probability that it is the SKY or GROUND types. This attribute consists of three elements, "none", "contact", and "include". Figure 3 illustrates each element.

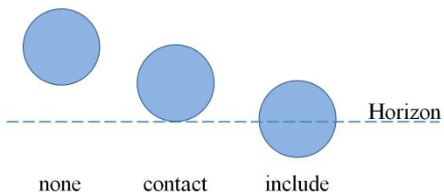


Figure 3. Form of intersection and selected element
for Horizon attribute

The next attribute a_2 describes the relation between the line passing through the vanishing point and an object. The attribute helps to classify CUBIC objects from PLANE objects. CUBIC objects have a high probability that the extended lines of their edges pass through the vanishing point, but PLANE objects do not. If at least one extended line passes through the vanishing point, the "include" element is selected, else the "none" element is selected.

The third attribute a_3 is the vertical line. Because CUBIC objects frequently include the line passing the vanishing point and the vertical line at the same time, the inclusion relationship of the vertical line provides excellent cues to distinguish the CUBIC object from other objects. In order to select a proper element, we check whether the edges include any vertical lines or not. If at least one vertical line is included in the object the "include" element is selected, else the "none" element is selected.

The information, which objects contact with the image border, gives the classifier important cues. For example, the SKY object has a high probability that the object contacts with the top border of the image, and may not contact with the bottom border. Although there can be more than one border contacted with the object, we discard other borders except one border contacting with the object in the largest area. The "none" element is for the object which does not contacts with any image border.

Texture complexity a_5 represents how much high frequency textures are included in the object. Generally the sky contains the low frequency texture. For calculating texture complexity, the average of difference between the original texture and the low-pass filtered texture is calculated by Eq. (1).

$$Texture\ \ complexity = \frac{1}{N_O}\sum_{x,y\in O} p(x,y)\text{-}p(x,y)*g(x,y) \quad (1)$$

where $p(x,y)$ is the original objects, N_o is the number of the pixels in the object, and $g(x,y)$ is the two-dimensional Gaussian filer. In Eq. (1), the operator, *, is convolution, and the summation is calculated for only pixels in the object. According to the degree of complexity, we divide elements into four levels, High-High (HH), High-Low (HL), Low-High (LH), and Low-Low (LL).

The final attribute a_6 relates to object size. The percentage of object size can be cues for classification. It has same elements of level with the texture complexity attribute a_5.

In order to generate training data, we manually classify object types in several images. The training data for fifty objects were gathered, and it is used to classify new objects in an input image.

3.3 Bayesian Classification

For a new input image, we classify each object into the pre-defined object type by analyzing its attributes. Although the approach using intuitive classification is possible, we use the approach based on the probability leaning method because images can contain various exceptions [8]. The Bayes theorem provides a way to calculate the probability of an object type based on its prior probability from the training data set. The Bayes theorem is the basis of Bayesian learning methods because it provides a way to calculate the posterior probability $P(h|T)$, from the prior probability $P(h)$, together with $P(T)$ and the likelihood probability $P(T|h)$, where T and h represent the object type and hypothesis, respectively.

$$P(h|T) = \frac{P(T|h)P(h)}{P(T)} \quad (2)$$

The Bayesian approach to classifying the new object is to assign the most probable target value, t_{MAP}, given the attribute values $<a_1, a_2, \ldots, a_6>$ describing the object.

$$t_{MAP} = \arg\max_{t_j \in T} P(t_j|a_1,a_2,\ldots,a_6) \quad (3)$$

The expression can be re-written using the Bayes theorem as Eq. (4).

$$t_{MAP} = \arg\max_{t_j \in T} \frac{P(a_1,a_2,\ldots,a_6|t_j)P(t_j)}{P(a_1,a_2,\ldots,a_6)} \quad (4)$$

Two terms in Eq. (4) can be estimated on the basis of the training data set. It is easy to estimate each of the $P(t_j)$ simply by counting the frequency with which each target value t_j occurs in the training data. However, It is not possible to estimate the different $P(a_1,\ a_2,\ldots,a_6|t_j)$ terms in this fashion. Therefore, we need to see every instance in the instance space many times in order to obtain reliable estimates. The simplified assumption that the attribute values are conditionally independent given the target value is adopted in our approach. Under the assumption, the probability of observing the conjunction can be simplified to the product of the probabilities for the individual attributes as Eq. (5). It is called the naive Bayes classifier.

$$t_{NB} = \arg\max_{v_j \in V} P(t_i)\prod_i P(a_i \mid t_j) \quad (5)$$

where t_{NB} denotes the target value by the naive Bayes classifier. We select the object type having the highest probability with the attributes of the input object.

4. DEPTH ASSIGNMENT

4.1 Fundamental Depth Map

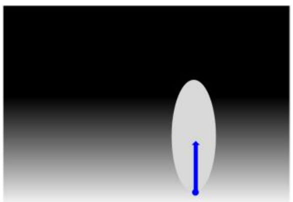
After classification, we make a fundamental depth map used as a reference depth map during depth assignment. The fundamental depth map reflects the properties of the ground and the sky. Zero value is assigned to the upper area than the vanishing point, because the sky has infinite distance from the camera. In order to generate the fundamental depth map, depth values are assigned by Eq. (6).

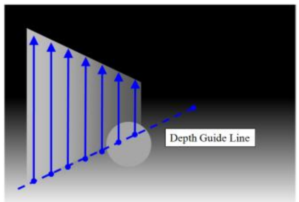
$$depth_y = \begin{cases} \dfrac{255(y - VP_y)}{height - VP_y} & if\ \ y > VP_y \\ 0 & otherwise \end{cases} \quad (6)$$

In Eq. (6), y and VP_y represent the current row position and the row position of the vanishing point, respectively. The height stands for the input image's height. By applying this formula to whole rows, we can obtain the fundamental depth map.

4.2 Depth Assignment for Objects

According to the object type, we assign the proper depth values to classified objects by different ways. For PLANE objects, a constant depth value located at the bottom position of the object from the fundamental depth map is copied, and the object is filled with it, as shown in Fig. 4(a).





(a) PLANE (b) CUBIC

Figure 4. Depth assignment for PLANE and CUBIC objects

Contrary to PLANE objects, the assignment for the CUBIC object is different, because the depth values of the CUBIC object have to become different according to the distance from the vanishing point. We copy the depth value from the fundamental depth map and fill the one column with it. By repeating this process per each column, we obtain the distance-dependent depth values from the vanishing point.

However, when a CUBIC object is hidden by other objects, wrong depth values are assigned due to the wrong information of the bottom boundary of the CUBIC object. In order to overcome this problem, we define a depth guide line which acts as a guide line when we copy the depth value from the fundamental depth map as shown in Fig. 4(b).

5. EXPEIMENTAL RESULTS

In order to show the performance of our proposed algorithm, we took tests with two outdoor photographs having the linear perspective cue. Figure 5(a) and Fig. 6(a) are the input images, and the images of (b) are the estimated depth map with the proposed algorithm. Whole objects in the images are reasonably classified and are filled with appropriate depth values. With the input images and the depth maps, the 3D scenes are generated using 3D warping techniques as shown in Fig. 5(c) and Fig. 6(c). From the results, we can know that the proposed algorithm estimates the similar depth from single-view images with our perception.


(a) Input image


(b) Depth map

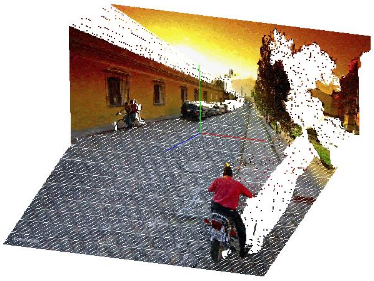

(c) 3D scene

Figure 5. Experimental results for *Rider* image


(a) Input image


(b) Depth map


(c) 3D scene

Figure 6. Experimental results for *Red Building* image

6. CONCLUSIONS

Because interest on 3D contents is increasing gradually, 3D image processing techniques are attracting more attention. The 3D scene generation from a single-view image is an essential technology for the 3D contents. Among its relevant problems, the depth estimation is the most significant and complicated task. In this paper, we proposed the depth estimation algorithm from a single-view image using object classification based on the Bayesian learning. On the basis of the training data set about six attributes, objects in a single-view image were categorized into four types: SKY, GROUND, CUBIC, and PLANE. According to their types, relative depth values can be assigned with our algorithm. Experimental results show that the proposed method estimates the depth maps which is similar to our perception, and successfully generates the 3D scene of the input images.

7. ACKNOWLEDGEMENT

This research was supported in part by MKE, Korea, under the ITRC support program supervised by NIPA (NIPA-2010-(C1090-1011-0003))

8. REFERENCES

[1] Y. Kang, E. Lee, and Y. Ho, "Multi-Depth Camera System for 3D Video Generation," Proceedings of International Workshop on Advanced Image Technology, pp. 44(1-6), Jan. 2010.

[2] S. Battiato, A. Capra, S. Curti, and M. La Cascia, “3D Stereoscopic Image Pairs by Depth-Map Generation,” Proceedings of International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 124-131, Oct. 2004.

[3] J. Ko, M. Kim, and C. Kim, “2D-To-3D Stereoscopic Conversion:Depth-Map Estimation in a 2D Single-View Image,” Proceedings of the SPIE, vol. 6696, pp. 66962A, Sept. 2007.

[4] S. A. Valencia and R. M. Rodriguez-Dagnino, “Synthesizing Stereo 3D Views from Focus Cues in Monoscopic 2D mages,” Proceedings of the SPIE, vol. 5006, pp.377-388, Oct. 2003.

[5] W. J. Tam, F. Speranza, L. Zhang, R. Renaud, J. Chan, and C.Vazques, “Depth Image Based Rendering for Multivew Stereoscopic Display: Role of Information at Object Boundaries,” Proceedings of the SPIE, vol. 6016, pp. 601609, Oct. 2005.

[6] Y. Chang, C. Fang, L. Ding, S. Chen, and L. Chen, “Depth Map Generation for 2D-to-3D Conversion by Short-Term Motion Assisted Color Segmentation”, Proceedings of IEEE International Conference on Multimedia and Expo, pp. 1958-1961, July 2007.

[7] V. Vezhnevets and V. Konouchine, “GrowCut - Interactive Multi-Label N-D Image Segmentation By Cellular,” Proceeding of Graphicon, pp. 150-156, June 2005.

[8] T. M. Mitchell, “Machine Learning,” Mc Grow Hill, 1997.

[9] D. Comanicu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, May 2002.

[10] V. Cantoni, L. Lombardi, M. Porta, and N. Sicari, ″Vanishing Point Detection: Representation Analysis and New Approaches”, Proceedings of International Conference on Image Analysis and Processing, pp. 90-94, Sept. 2001.

[11] D. Hoiem, A. Efros, and M. Hebert, “Automatic Photo Pop-up”, ACM SIGGRAPH, Aug. 2005.