Depth Map Up-Sampling Using Random Walk

Gyo-Yoon Lee and Yo-Sung Ho

Gwnagju Institute of Science and Technology (GIST),
261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712 Republic of Korea
{gyoyoon,hoyo}@gist.ac.kr

Abstract. For the high quality three-dimensional broadcasting, depth maps are important data. Although commercially available depth cameras capture high-accuracy depth maps in real time, their resolutions are much smaller than those of the corresponding color images due to technical limitations. In this paper, we propose the depth map up-sampling method using a high-resolution color image and a low-resolution depth map. The proposed method is appropriate to match boundaries between the color image and the depth map. Experimental results show that our method enhances the depth map resolution successfully.

Keywords: 3D broadcasting, multi-view video, FTV, TOF cameras, depth map, interpolation, random walk.

1 Introduction

Three-dimensional (3-D) video currently attracts public attention in a variety of multimedia applications. The current 3-D videos provide 3-D effects using the stereoscopic images which are based on binocular depth cues. In the near future, users will be able choose their viewpoints of themselves in the immersive visual scenes created by 3-D videos.

Since we cannot transmit videos of all viewpoints, we synthesize the viewpoint's video using transmitted video-plus-depth data for 3-D TV [1]. To provide the high quality synthesized view, accurate depth information is important. In general, depth estimation methods are classified into two categories: active depth estimation and passive depth estimation. The active depth estimation method directly obtains the depth map using physical sensors. On the contrary, the passive depth estimation method calculates the depth values using acquired 2-D images.

The passive depth estimation method uses two or more 2-D images. Typical examples are shape from focus [2] and stereo matching [3]. The passive depth estimation method can be performed at a low price because it needs only 2-D images. However it does not guarantee quality of depth map because the performance of passive depth estimation depends on image properties. Active depth estimation uses the physical sensor such as lasers, infrared rays (IR), or light patterns. There are structured light patterns [4] and depth cameras [5]. If we use physical equipment, we can obtain more accurate depth values. However, depth cameras are expensive and they capture low-resolution depth maps only. To get accurate depth maps, the hybrid

Y.-S. Ho (Ed.): PSIVT 2011, Part I, LNCS 7087, pp. 384–394, 2011.

camera system was proposed [6]. To overcome problems of previous depth estimation methods, the hybrid camera system consists of the multi-view color cameras and the depth cameras. Thus, it can perform both active depth estimation and passive depth estimation.

Although depth estimation methods have been researched continually, more accurate depth estimation method remains an unsolved problem. We obtain more accurate depth values when we use the active depth estimation method. However, we need the up-sampling process due to the difference between color images and depth maps. Besides, depth up-sampling can be used for the depth encoding algorithm. We improve coding efficiency by transmitting the down-sampled depth map which is represented much fewer bits than that of original depth information. In the decoder, the transmitted depth map can be used through the up-sampling process. Thus if accuracy of up-sampled depth values is higher, coding efficiency will be improved. In this paper, we propose the efficient depth map up-sampling method.

Section 2 explains the previous depth up-sampling methods. In Section 3, we describe the proposed up-sampling method. Then, Section 4 demonstrates the experimental results. We conclude in Section 5.

2 Related Works

The hybrid camera system also has the problem that resolution of depth maps captured by depth cameras is smaller than that of the corresponding color images due to technical limitations of the depth cameras [7]. Figure 1 shows the resolution difference between the color image and the depth map of the hybrid camera system.

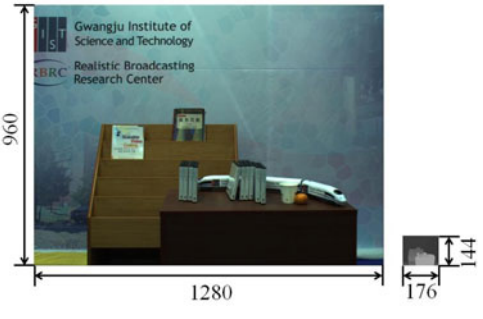


Fig. 1. Resolution of the color image and the depth map

Since the inaccurate depth information deteriorates the quality of synthesized views and 3-D video, the quality of depth maps is very important for an image-based rendering. Thus the accurate enhancement method of the low-resolution depth map is required. To obtain accurate depth information, we need to keep following properties of depth maps.

1. Boundaries of depth maps match corresponding color image boundaries.
2. Depth values of neighboring pixels in the same object are similar.

To solve this problem, various methods have been proposed. In the beginning of the research, general image interpolation methods were used such as bilinear, nearest-neghbor, and bicubic interpolations [8]. However, they do not guarantee the depth map properties. So, the Markov random field probability model and the bilateral filter are proposed.

2.1 Markov Random Field

Diebel *et al.* interpolated depth values using the Markov random field probability model (MRF) and the designed the adaptive weighting function according to the color gradient [9]. The MRF is composed of 5 node types. Figure 2 shows the designed MRF.

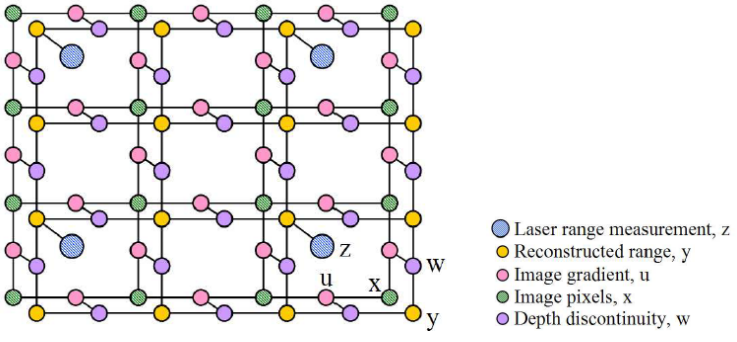


Fig. 2. Node types of MRF

The MRF is defined through thefollowing conditional probability.

$$p(y|x,z) = \frac{1}{Z}\exp(-\frac{1}{2}(\Psi + \Phi)) \tag{1}$$

where Ψ is depth measurement potential and it represents the difference between the laser range measurement and the reconstructed range. Φ is depth smoothness term. As computing the optimization problem of Eq. (1), we obtain the depth values.

2.2 Joint Bilateral Up-Sampling

Kopf *et al.* proposed the post-processing step using the bilateral filter [10]. The bilateral filter is an edge-preserving filter. The idea is to apply a spatial filter to the low resolution S, while similar range filter is jointly applied on the full resolution image. The up-sampled solution $\tilde{S}_p$ is then obtained as

$$\tilde{S}_P = \frac{1}{k_p}\sum_{q_\downarrow \in \Omega} S_{q\downarrow} f(\|p_\downarrow - q_\downarrow\|) g(\|\tilde{I}_p - \tilde{I}_q\|) \tag{2}$$

$g(\|\tilde{I}_p - \tilde{I}_q\|)$ is color distances between pixel p and q in full resolution $\tilde{I}$. And $f(\|p_\downarrow - q_\downarrow\|)$ represents the spatial distance.

After releasing the joint bilateral filter, many up-sampling methods which use modified bilateral filter are proposed. Yang *et al.* proposed the post-processing step using the bilateral filter [11]. This method enhances the low-resolution depth map by refining iteratively initial depth values.

3 Proposed Depth Up-Sampling

We generate the new depth value using the initial depth values. We warp the pixel from low-resolution depth map to color image. We define the initial values as warped depth values.

3.1 Initial Value

Camera Calibration. To match a depth map and a color image of different cameras, it is important to find out relative camera information through camera calibration [12]. We apply a camera calibration algorithm [13] to each camera and obtain projection matrices.

$$P = K[R|t] . \tag{3}$$

where P is the projection matrix of each camera. It is consist of the intrinsic matrix K, the rotation matrix R, and translation vector t.

3-D Warping. The camera parameter represents the relative position of the camera and world coordinates. Since we have position information and depth information of cameras, we can find the any position of the depth map in the world coordinate using Eq. (4) which is consist of camera parameter ***R, K, t***.

$$\mathrm{X_r} = \mathrm{R_r^{-1}} \cdot \mathrm{K_r^{-1}} \cdot \mathrm{x_r} \cdot \mathrm{d_r(x_r)} - \mathrm{R_r^{-1}} \cdot \mathrm{t_r} . \tag{4}$$

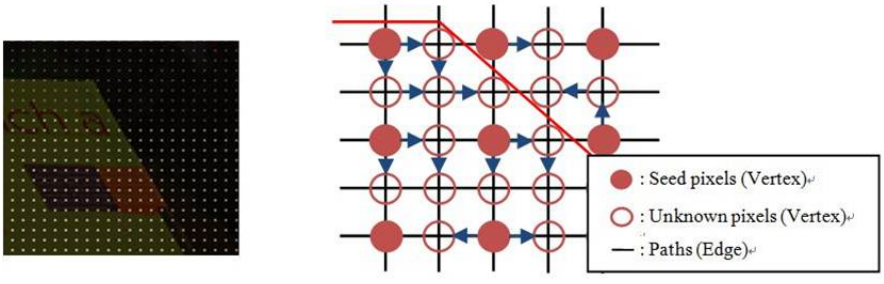
where $\boldsymbol{X_r}$ means the position in the real world coordinates of a pixel $\boldsymbol{x_r}$ in the depth map, and $d_r(\boldsymbol{x}_r)$ is the return value of the corresponding depth value of $\boldsymbol{x}_r$. After finding position in the world coordinates, we then reproject the 3-D points into the color image. Equation (5) represents reporjection equation which is composed of camera parameter of the color camera and the geometric position of the depth map.

$$\mathrm{x_t} = \mathrm{P_t X_r} . \tag{5}$$

where $\boldsymbol{x_t}$ is the corresponding position of $\boldsymbol{x_r}$ in the depth map. All pixels of the depth map have the position corresponded a color image through 3-D warping. However the 3-D warping technique cannot guarantee perfectly matching positions of the depth map and the color image.

3.2 Depth Up-Sampling Using Random Walk

We classify pixels as seed pixels and unknown pixels. If pixels have initial depth values, they are seed pixels and other pixels are unknown pixels. We assume that depth values of neighboring pixels which have similar color values are similar. Thus, we copy the depth values of unknown pixels from the depth values of a seed pixel which has similar color values and the low distance cost. Figure 3 shows the concept of our up-sampling method.



(a) Initial depth value (b) Example of depth interpolation

Fig. 3. Depth map interpolation

We calculate the random walk probability of each seed pixels. Figure 4 is an example of the random walk probability of seed pixels. After calculation of the probability, unknown pixels have the probability values corresponded with each seed pixels. Thus we copy depth values of unknown pixels from that of a seed pixel which has the largest probability values.

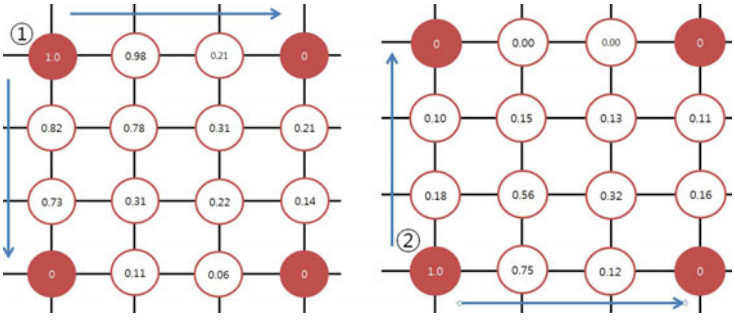


Fig. 4. Example of random walk probability

To calculate random walk probability each path between neighboring pixels has the cost. The cost between pixel *i* and pixel *j* represents

$$w_{ij} = \exp\left(-\frac{\left| z_i - z_j \right|^2}{\sigma} \right). \tag{6}$$

where $\left| z_i - z_j \right|$ represents the Euclidean color distance and σ means the variance. If unknown pixel is far from the seed pixel, random walk probability decreases as cost between two pixels. However since there are many path between the seed pixel and the unknown pixel, random walk probability depends on the path between pixels. Figure 5 shows the random walk probability corresponding paths.

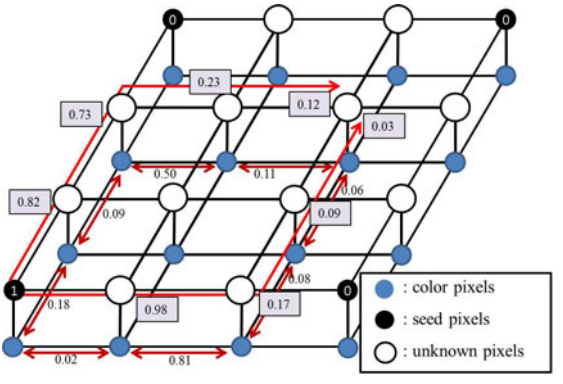


Fig. 5. Random walk probability of each path

We define the random walk probability as the largest probability among each path's probability. The largest probability means that the sum of path cost is the smallest. Thus random walk probability is

$$P(j|seed(i)) = \text{argmin}\left(\sum_{i \to j} w_{ij} \right). \tag{7}$$

i is seed pixel and *j* is unknown pixel.

To find minimum cost there several methods. Graph-Cut is widely used in variety field for optimization. Graph-Cut method minimizes weight of connections between groups. However it only considers external cluster connections. It does not consider internal cluster density.

We solve this minimum cost path problem using spectral graph theory [14]. It solves the problem using simple matrix calculation. The graph is made of edges and vertices. The vertices mean the pixels and the edges mean the connectivity information. As a neighborhood system (4-neighbor or 8-neighbor), the number of edges and the shape of graph are different. We select 4-neighbor system. As shown in Figure 5, each pixel is the vertex and the connected lines are edges. All edges have

cost values as defined in Eq. (6), represented as Gaussian weighting color distribution. The variance of Eq. (6) controls weighting of color and distance. Large variance increases color weighting. If variance is small, the distance weighting is larger than color weighting.

The desired random walk problem has the same solution as combinatorial Dirichlet problem [15], [16]. The Dirichlet integral is defined as

$$D[u] = \frac{1}{2}\int |\nabla u|^2 d\Omega \,. \tag{8}$$

The harmonic function satisfies the Laplace equation. Since the Laplace equation is the Euler-Lagrange equation for the Dirichlet integral, the harmonic function minimizes the Dirichlet integral [17]. Finally, the Dirichlet integral is the same and it is defined as

$$D[x] = \frac{1}{2}(Ax)^T C(Ax) = \frac{1}{2}x^T Lx = \frac{1}{2}\sum_{i,j\in E} w_{ij}(x_i - x_j)^2 \,. \tag{9}$$

Equation (7) is substituted matrix calculation $x^T Lx$ in Eq. (9). L represents the Laplacian matrix. Equation (10) represents the Laplacian matrix. w_{ij} is cost between each neighboring pixel and d_i is sum of cost between pixel i and neighboring pixels.

$$L_{ij} = \begin{cases} d_i & \text{if } i = j, \\ -w_{ij} & \text{if } v_i \text{ and } v_j \text{ are adjancent,} \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

$x^T Lx$ of Eq. (9) is

$$\begin{aligned} D[x] &= \frac{1}{2}[x_{Seed}^T \quad x_{Unknown}^T]\begin{bmatrix} L_{Seed} & B \\ B & L_{Unknown} \end{bmatrix}\begin{bmatrix} x_{Seed} \\ x_{Unknown} \end{bmatrix} \\ &= \frac{1}{2}(x_{Seed}^T L_{Seed} x_{Seed} + 2x_{Unknown}^T B^T x_{Seed} + x_{Unknown}^T L_{Unknonw} x_{Unknown}) \,. \end{aligned} \tag{11}$$

where B is a combination of the seed matrix and the unknown matrix. Since we minimize Eq. (11), we find the critical point. To find critical point, we differentiate the Eq. (11).

$$L_{Unknown} x_{Unknown} = -B^T x_{Label} \,. \tag{12}$$

Because we know B, x_{Seed}, and $L_{Unknown}$, we can find the probability $x_{Unknown}$ from Eq. (12) and fill the unknown pixel with the depth value of the seed pixel which has maximum probability. Because the Laplacian matrix is too large, solving the Eq. (12) is difficult. However since the Laplacian matrix is symmetric and sparse, it is easily solved.

4 Experimental Results

4.1 Depth Map Interpolation

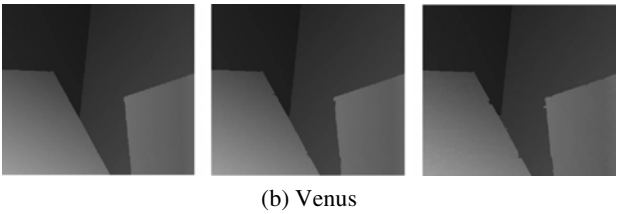
To evaluate the objective performance of the proposed interpolation method, we use the data set of *Middlebury* website [18]. We apply the proposed method with the down-sampled depth maps. The down-sampled depth map consists of pixels which are on positions of multiples of up-sampling rate in the original depth map. To improve accuracy we interpolate the depth value of a block unit. Block based interpolation method is efficient because it considers only near values and positions of initial value is regular. As comparing the interpolated depth values to the original depth values, we find the error percentage which is used by *Middlebury*. In addition, we compare the error percentage with other interpolation method such as MRF refinement [9] and iterative joint bilateral filter [11]. Table 1 shows the comparison of all error percentage.

Table 1. Comparison of all error percentages

	Tsukuba			*Venus*			*Teddy*			*Cone*		
Up-sampling rate	2	4	8	2	4	8	2	4	8	2	4	8
MRF	2.51	5.12	9.68	0.57	1.24	2.69	2.78	8.33	14.5	3.55	7.52	14.4
Bilateral filter	1.16	2.56	6.95	0.25	0.42	1.19	2.43	5.95	11.5	2.39	4.76	11.0
Proposed method	0.69	1.23	2.33	0.18	0.27	0.31	2.92	3.91	5.98	3.01	3.67	5.37

In the low up-sampling rate, the error percentage is similar to the bilateral filter method. However as the up-sampling rate is higher, the proposed algorithm shows the much better performance than that of previous methods. Figure 6 shows the up-sampled depth map.

(a) Teddy



(b) Venus

Fig. 6. Double, quadruple, octuple interpolated images

4.2 Boundary Noise Remove

We obtain the depth map (176×144) using SR-4000 of the hybrid camera system. To interpolate the depth values of TOF cameras, we must consider the wrong initial values which are caused by 3-D warping error and depth map noise. When depth maps of TOF cameras are up-sampled, depth maps have noise. The depth map noise is caused by camera parameter errors, 3-D warping errors, and non-discontinuity depth value on boundary. Figure 7 shows the boundary noise of warped depth values.

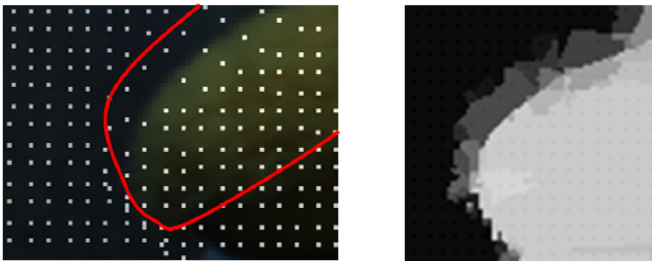
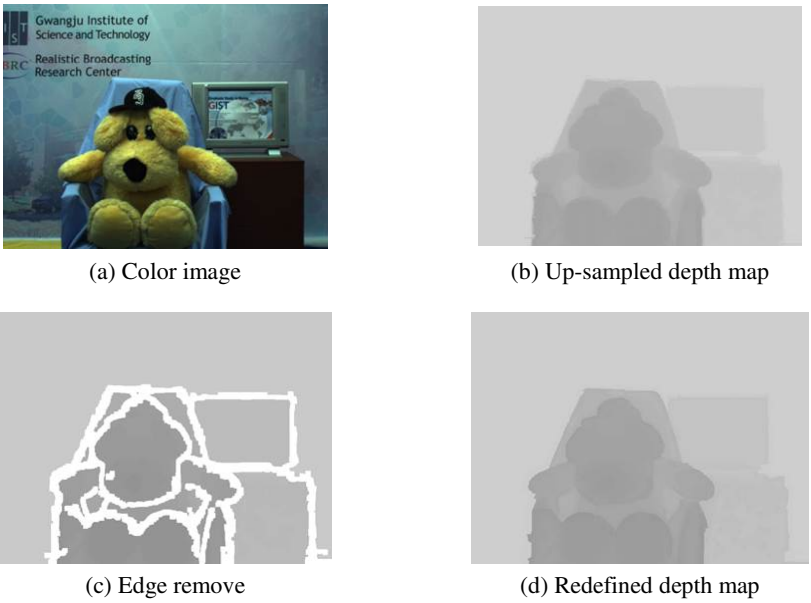


Fig. 7. Initial value noise

To overcome the problem, we redefine the depth values neighboring edges using the proposed depth hole filling method. Figure 8 shows the up-sampled depth map, color image (800×600), and boundary redefined depth map. To improve the clarity we reverse the depth values. Figure 9, 10 represent the up-sampled depth from 176×144 to 1190×950 using proposed method and 3-D rendering result.



(a) Color image (b) Up-sampled depth map

(c) Edge remove (d) Redefined depth map

Fig. 8. Up-sampled depth map

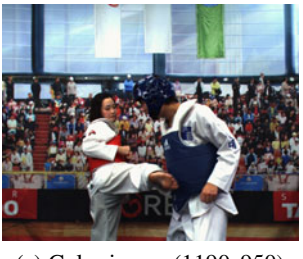
(a) Color image (1190×950)

(b) Depth map

(c) Up-sampled depth map

(d) Rendering result

Fig. 9. Up-sampled depth map and rendering result



(a) Color image (1190×950)

(b) Depth map

(c) Rendering result

Fig. 10. Rendering result using the up-sampled depth map

5 Conclusion

To render the 3-D scene, depth information is essential data. Depth maps which are captured by the depth camera cannot match color images due to resolution difference. In this paper, we propose the random walk probability model for depth up-sampling. We objectively evaluate proposed method by up-sampling *Middlebury* data sets. The proposed method enhances the accuracy of up-sampled depth maps. As the block based up-sampling rate is larger, quality variations of the proposed method is smaller than that of previous methods. Besides we enhance the depth map which is captured by TOF cameras. We interpolate the depth map using global method. And we apply the post processing to overcome problem of 3D-warping and global method. The result of proposed method shows accuracy improvement of discontinuity regions neighboring object boundary.

Acknowledgments. This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2011-(C1090-1111-0003)).

References

1. Kauff, P., Atzpadin, N., Fehn, C., Muller, M., Schreer, O., Smolic, A., Tanger, R.: Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. Signal Processing: Image Communication 22(2), 217–234 (2007)
2. Nayar, S.K., Nakagawa, Y.: Shape from focus. IEEE Transactions on Pattern Analysis and Machine Intelligence 16(8), 824–831 (1994)
3. Zitnick, C., Kang, S., Uyttendaele, M., Winder, S., Szeliski, R.: High-quality video view interpolation using a layered representation. In: Proceedings of ACM SIGGRAPHS, vol. 23(3), pp. 600–608 (2004)
4. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 195–202 (2003)
5. Iddan, G., Yahav, G.: 3D imaging in the studio and elsewhere. In: Proceedings of SPIE Videometrics and Optical Methods for 3D Shape Measurements, vol. 4298, pp. 48–55 (2001)
6. Lee, E., Ho, Y.: Generation of high-quality depth maps using hybrid camera system for 3-D video. Journal of Visual Communication and Image Representation 22, 73–84 (2011)
7. Jung, J., Ho, Y.: MRF-based Depth Map Interpolation using Color Segmentation. In: Asia-Pacific Signal and Information Processing Association, pp. 19–22 (2010)
8. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical Recipes in C: The art of Scientific Computing. Cambridge University Press (1992)
9. Diebel, J., Thrun, S.: An application of Marko random fields to range sensing. In: Advances in Neural Information Processing Systems, vol. 18, pp. 291–298 (2005)
10. Johannes, K., Michael, C., Dani, L., Matt, U.: Joint Bilateral Upsampling. In: Proceedings of ACM SIGGRAPHS, vol. 26(3) (2007)
11. Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: International Conference on Computer Vision and Pattern Recognition (2007)
12. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 1330–1334 (2000)
13. Camera Calibration Toolbox Program for Matlab provided by Caltech, `http://www.vision.caltech.edu/bouguet/calibdoc/`
14. Gardy, L.: Random Walks for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11) (2006)
15. Kakutani, S.: Markov processes and the Dirichlet problem. Proc. Jap. Acad. 21, 227–233 (1945)
16. Biggs, N.: Algebraic potential theory on graphs. Bulletin of London Mathematics Society 29, 641–682 (1997)
17. Courant, R., Hilbert, D.: Methods of Mathematical Physics, vol. 2. John Wiley and Sons (1989)
18. `http://vision.middlebury.edu/stereo/`