

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

A New Approach to Boundary Matting for 3DTV System

¹Lingshan Liu, ^{1,2}Ming Xi, ^{1,2}Dong-Xiao Li and ^{1,2}Ming Zhang

¹Institute of Information and Communication Engineering, Zhejiang University,
Hangzhou 310027, China

²Zhejiang Provincial Key Laboratory of Information Network Technology,
Hangzhou 310027, China

Abstract: Precision of depth has a significant impact on quality of virtual view synthesis in 3DTV system. Artifacts exist in stereo correspondence. Object boundary artifact is one of the most difficult problems. The object boundary pixels receive colors from both foreground and background thus leads to wrong mapping. The occlusion fill-up process causes object boundary expansion which makes boundary pixels unreliable. We propose a boundary matting technique based on natural image matting algorithms to calculate the alpha value of the boundary pixels. Depth values which are suggested by alpha values can be recalculated. The limitation of matting algorithms is that they define only boundaries between two layers and cannot solve the multiple-layer definition problem. Our contribution is the localized methods of boundary matting which can solve the multiple layer boundaries in complex depth map. Our goal is superior view synthesis from stereo, thus the quality of virtual views are valued in experiments. The result shows that our technique reduces the boundary artifacts effectively.

Key words: Boundary matting, natural image matting, localization, stereo, 3D

INTRODUCTION

3D technology is becoming more and more popular in recent years (Ruzinor *et al.*, 2011). The 3D online Marketing, 3D model construction, 3DTV are all popular applications (Sharma *et al.*, 2011; Kies *et al.*, 2005). Natural 3DTV system applies stereo algorithm to generate depth map (Yu *et al.*, 2011). With the main-view image and its corresponding depth map, DIBR algorithm generates multiple virtual view images. By delivering stereo image pairs into the left and right eyes individually, Natural 3DTV is able to give the observer an extraordinary feeling of 3D real scene. Quality of stereo depth map plays an important role in Natural 3DTV system (Yao *et al.*, 2012). Although stereo is developing, problems still remain (Zhixiang *et al.*, 2011; Shuchun *et al.*, 2011). One of the problems of stereo is the object boundaries, where the pixel colors are mixtures of foreground and background colors (Hasinoff *et al.*, 2006). If the mixed pixel colors are used during rendering, visible artifacts would result in virtual view images. Another drawback of depth is that the boundaries expand after hole-filling or disocclusion process. The boundary refinement technique is one of the most difficult problems in 3DTV system. Most current

methods do not perform matting to solve mixed boundary pixels. Hasinoff *et al.* (2006) proposed a boundary matting technique which seems to similar to ours (Chuang *et al.*, 2001).

Matting refers to the problem of accurate foreground estimation in images and video (Wang and Cohen, 2007). Early in 1984, researchers established the problem mathematically and introduced the alpha channel as the means to control the linear interpolation of foreground and background (Mitsunaga *et al.*, 1995). Recently, natural image matting has developed. Based on methodologies, natural image matting can be classified into two categories (Wang and Cohen, 2007): color sampling matting methods and defining affinities matting methods. Color sampling matting methods can be classified into parametric sampling algorithm and nonparametric sampling algorithm. Ruzon and Tomasi proposed an early parametric sampling algorithm (Ruzon and Tomasi, 2000). Based on Ruzon and Tomasi's methods, Chuang *et al.* (2001, 2002) proposed a Bayesian matting approach, this approach models foreground and background colors as mixtures of Gaussians and uses a maximum-likelihood criterion to estimate the optimal opacity, foreground and background

(Chuang *et al.*, 2001, 2002; Wexler *et al.*, 2002). Parametric sampling methods would generate large fitting errors when color distributions are non-Gaussian. Nonparametric methods can avoid this problem. Knockout system calculates the foreground value of an unknown pixel as a weighted sum of nearby known foreground colors. Instead of directly estimating the alpha value at each pixel, another way of using local image statistics is defining affinities between neighboring pixels and modeling the matte gradient across the image lattice. Poisson matting assumes that intensity changes in the foreground and background are locally smooth and formulate the matting problem as solving Poisson equations (Sun *et al.*, 2004). Closed form matting derives a cost function from local smoothness assumptions on foreground and background colors to obtain a quadratic cost function in alpha (Levin *et al.*, 2008a). In recent years, matting algorithm is widely used. Nsib studied the matting of polymeric pigmented coating and analyzed changes in color and rheology of coating caused by adding a matting agent (Nsib *et al.*, 2008).

Hasinoff *et al.* (2006) proposed boundary matting which is the most similar to our work. They exploit multiple views to perform alpha matting and to simultaneously refine stereo depth at the boundaries (Hasinoff *et al.*, 2006). The approach starts from an initial estimate derived from stereo and optimizes the boundary curve parameters and foreground colors near the boundaries. However their methods rely on multiple inputs, we propose a simplified boundary matting algorithm which implements on only one color image and several constrains.

In this study we focus on the boundary artifacts problem in virtual synthesis and attempt to solve the boundary artifacts in complex scenes. The key feature of our approach is that boundaries can be defined locally. Compared to natural image matting algorithm, we extend the matting application to 3DTV system. The compositing equation shows that matting only solve the boundary pixels between definite foreground and background, however scenes in 3DTV application may be complex thus boundaries between sub-foreground and sub-background need to be estimated. Our distinguish contribution is invention of a localized method of matting algorithm to solve the multiple-layer identification problem.

DIGITAL MATTING

Matting equation: The compositing equation of matting is (Chuang *et al.*, 2001; Wang and Cohen, 2007):

$$I = \alpha F + (1-\alpha) B \tag{1}$$

where, C, F and B are the composite, foreground and background colors, α is the opacity component used to linearly blend foreground and background.

In RGB color space the compositing equation can be described as:

$$\begin{cases} I_R = \alpha F_R + (1-\alpha)B_R \\ I_G = \alpha F_G + (1-\alpha)B_G \\ I_B = \alpha F_B + (1-\alpha)B_B \end{cases} \tag{2}$$

Given a single input image, composite color C is known while F, B and α are unknown. Thus 7 unknown values need to be solved out of 3 known values, this makes matting an ill-posed problem. Additional constraints or user-interface is needed to solve the matting problem (Wang and Cohen, 2007). Most matting algorithms use a trimap to provide additional conditions. A trimap segments the input image into three regions: definitely foreground F, definitely background B and unknown region U. The accuracy of trimap affects the performance of a matting algorithm. Different matting algorithms make the demand for the trimap different. We have tried several matting algorithms and the results show that Bayesian matting and spectral matting perform better than others.

Bayesian matting: Bayesian matting models both foreground and background color distributions with spatially-varying sets of Gaussians and assumes a fractional blending of the foreground and background colors to produce the final output. Bayesian matting is based on Bayesian frame work which uses a maximum-likelihood criterion to estimate the optimal opacity, foreground and background simultaneously (Wang *et al.*, 2010). The maximization over a probability distribution P can be expressed as:

$$\begin{aligned} \arg \max_{F,B,a} P(F,B,a|C) &= \arg \max_{F,B,a} \frac{P(C|F,B,a)P(F)P(B)P(a)}{P(C)} \\ &= \arg \max_{F,B,a} L(C|F,B,a) + L(F) + L(B) + L(a) \end{aligned} \tag{3}$$

Through a continuously sliding window for neighborhood definitions, it marches inward from the foreground and background regions, using nearby computed F, B and alpha value in constructing oriented Gaussian distributions. They estimate the color probability distribution with the known foreground or background colors. Considering the weights and clusters, it calculates mean F and covariance matrix:

$$\bar{F} = \frac{1}{w} \sum_{i \in N} w_i F_i \quad (4)$$

$$\sum_{F} = \frac{1}{w} \sum_{i \in N} w_i (F_i - \bar{F})(F_i - \bar{F})^T \quad (5)$$

The solution of α becomes:

$$a = \frac{(C-B)(F-B)}{\|F-B\|^2} \quad (6)$$

The Bayesian matting algorithm effectively handles objects with intricate boundaries. In comparison of Ruzon and Tomasi's method and Knockout, it is relatively easy to implement. Thus we choose Bayesian matting as one of our methods.

Spectral matting: Spectral matting (Levin *et al.*, 2008b) automatically computes a set of fundamental fuzzy matting components from the smallest eigenvectors of a suitably defined Laplacian matrix. Spectral matting is based on spectral segmentation methods which perform unsupervised image segmentation by examining the smallest eigenvectors of the image's graph Laplacian matrix (Ng *et al.*, 2001). As spectral matting is an ill-posed problem, extensions of user-interface is needed. It generalizes the matting compositing equation by assuming that each pixel is a convex combination of K image layers:

$$I_z = \sum a_z^k F_z^k \quad (7)$$

where, the K vectors are the matting components of the image, they specify the fractional contribution of each layer to the final color observed at each pixel. This approach shows that fuzzy matting components may be

extracted from the smallest eigenvectors of matting Laplacian. Spectral matting can be implemented by unsupervised or interactive mode. User guidance provide foreground and background constraints which can make spectral matting more effective. Spectral matting is effective in automatable matte extraction; however it is challenging to perform on highly cluttered images.

Drawback of depth map: Stereo correspondence is the most common method to generate depth map in 3DTV system, however, problems exist. One of the problems is the boundary pixels which receive the impact from both foreground and background. If the mixed pixel colors are used during rendering, visible artifacts will result. In our 3DTV system, dynamic programming algorithm is applied to generate the depth map. In order to fill the occlusion regions, object boundaries are expanded. Thus the boundary pixel colors become unreliable. In conclusion, object boundaries of depth map are not reliable and should be improved.

PROPOSED METHOD

We apply matting algorithm to recalculate the depth values of boundary pixels. Trimap which provides additional information is initialized during the rendering process. The matting algorithm takes the color image and trimap as the input. The alpha value which is the result of matting algorithm suggests the depth value of the boundary pixel. A threshold value of alpha is needed to determine whether the new depth value belong to foreground or background. The flow chart is illustrated in Fig. 1.

Trimap: As is described in previous section, matting is an ill-posed problem and additional constraints or user

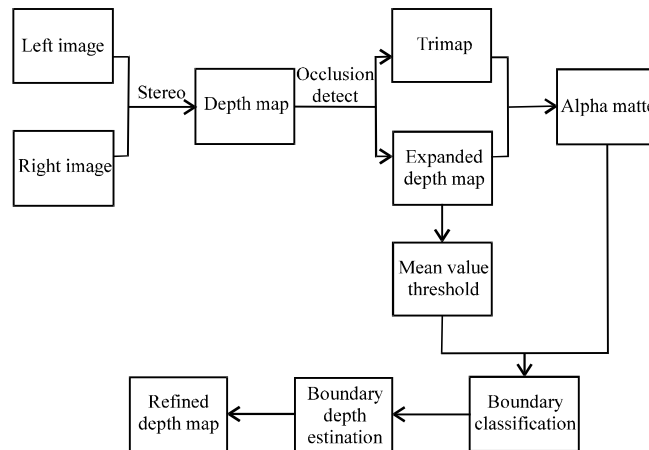


Fig. 1: Flow chart of general boundary matting

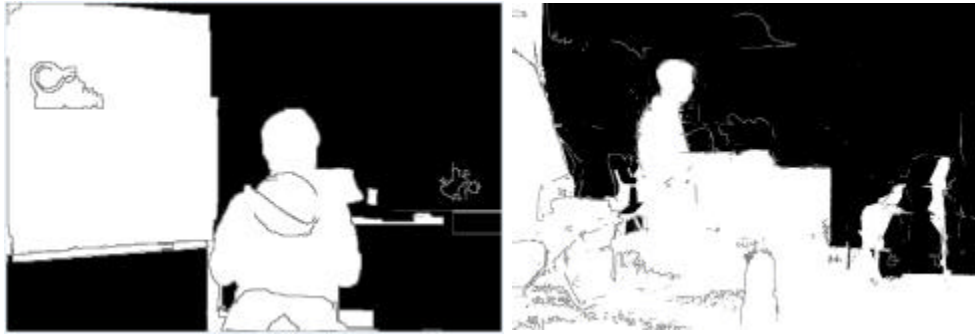


Fig. 2: Examples of Trimap, Left: first frame of basketball sequence; right: first frame of book_arrival sequence

guidance is needed for solving the matting equations. A trimap segments the input image into three regions: definite foreground F (with $\alpha = 0$), definite background B (with $\alpha = 1$) and the unknown regions U where α values need to be estimated. Thus the problem reduces to estimating F , B and α values in the unknown regions based on the definite known regions.

In our 3DTV system, virtual view synthesis process uses the stereo depth to generate virtual views. In order to fill the occlusion regions, borderlines are expanded. The pixels along the borderlines thus become unreliable. Suppose the expanded boundaries are unknown regions marked with U , the remaining pixels are separated into two regions (background and foreground) by dichotomy. The average value of these rest pixels are threshold. Foreground pixels are given α value of 255, while background pixels 0 and the unknown regions 128. Although dichotomy is not perfect, it is simple and effective. The segment of trimap should be improved later. Examples of trimap are illustrated in Fig. 2.

General boundary matting algorithm: With the color map and the trimap, the matting algorithm calculates the α value which shows the opacity of the boundary pixels. We classify the boundary pixels into foreground or background according to these α values. The threshold T of foreground or background is defined according to the mean value of the non-boundary pixels. If the α value of a boundary pixel is larger than T , it is figured as foreground pixel and vice versa. New depth value of the deduced foreground or background pixel is defined through the neighborhoods. Suppose a former unknown pixel p_i is figured as foreground pixel, we search in the neighborhoods of former definite foreground for the maximum depth value which is then defined as the new depth value of p_i . Similar process is implemented on the deduced background pixels. After this process, all boundary pixels have updated depth values. The depth of

former definite foreground and background pixels remain the same. Although the maximum value of square neighborhood is not precise, it provides the credible foreground or background value. The experimental results shows that the object boundaries shrink and the virtual view images get more legible. The precision of the new depth values relies on the following two factors: (1) precision of the matting algorithm. The appropriate segmentation of trimap has a great impact on matting result. Accurate and effective segmentation methods should be developed in future work, (2) the assignment of the new depth value. As the former depth value of a boundary pixel is totally unreliable, the new depth value should be estimated through an appropriate method after it is sorted. We have used the extreme value in the neighborhoods. Other feasible methods could be developed.

We have implemented several matting algorithm on boundary refinement, Bayesian matting performs best. Although it relies on the accurate of trimap, the performance is better than others. Spectral matting also gives suitable estimation, however it is a risk to implement on images that do not have visible components. With the trimap generated in virtual view synthesis process, Bayesian matting calculates the opacity of unknown regions (Fig. 3). The unreliable pixels along the object boundaries in depth map thus get their α values. Note that Bayesian matting algorithm does not estimate depth, only the α values. Depths are calculated by using the extreme value of definite nearby background or foreground according to the sort results. With the same trimap as user guidance, matting result of spectral matting is illustrated in lower row of Fig. 3.

Localized algorithm: In matting algorithm, images are arbitrarily defined into two layers. Only object boundaries between the global foreground and background could be updated. This is not generally applicable because scenes



Fig. 3: Alpha matte results of matting. Upper row is the result of Bayesian matting Second row is the result of spectral matting with the same trimap as constraints

in depth images are usually complex. Objects boundaries may consist in local part of image which leads to a problem that more than two layers of depth should be identified. We call this multiple-layer identification. None of the previous work has mentioned this problem. We propose a technique named localization of matting to deal with the multiple-layer identification problem. In our approach, matting and depth refinement are done locally, different object boundaries are identified in different blocks of image. The trimap and color image are divided into several blocks, matting algorithm and depth refine process are implemented on each divided block pairs. Similar approach to general algorithm is implemented on each local image pairs. First, we calculate the mean value of each block and judge whether it is uniform. For uniform blocks, we compare the uniform depth value with the mean value of the whole image to decide whether the block totally belongs to foreground or background. For non-uniform blocks, we mark the boundary pixels with gray-scale value of 128 and separate the remaining block pixels locally. Threshold is determined in each block. Thus foreground and background can be defined locally. By the same block size, color image and the corresponding trimap are divided into several blocks. Matting algorithm then is implemented on each block pair. In this way, the

boundaries are distinguished and classified locally; multiple-layer identification problem is solved. The whole process is illustrated in Fig. 4.

Although, boundaries between multiple layers can be identified by the methods mentioned above, problems still remain: (1) initialization of trimap. Block size should be determined beforehand and the technique of estimating block size should be improved. In this approach, we define the block size according to the complexity of input-files, (2) same with the general methods, the definition of threshold in each block should be improved. In our approach, mean value of non-boundary pixels in each block is regarded as the local threshold, (3) blocking effect. Pixels in the same layer of depth may be identified to different layers in different blocks. Thus the alpha values are estimated based on different clusters in different blocks. It causes the blocking effect, (4) time cost. Matting algorithm is not real-time. The process of localization takes quadruple time longer than general method does.

RESULTS

We test our technique on several datasets, using both general and localization techniques. The result of

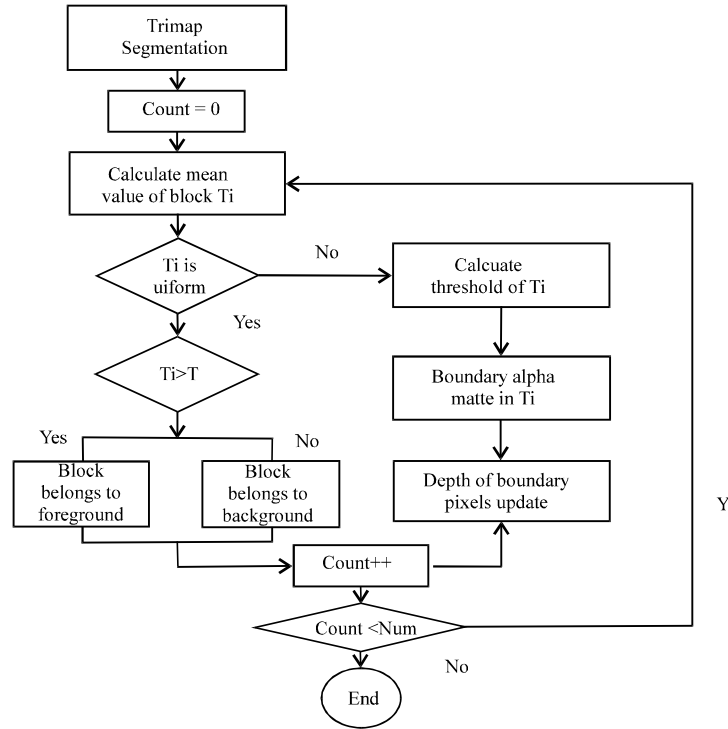


Fig. 4: Flowchart of localized algorithm

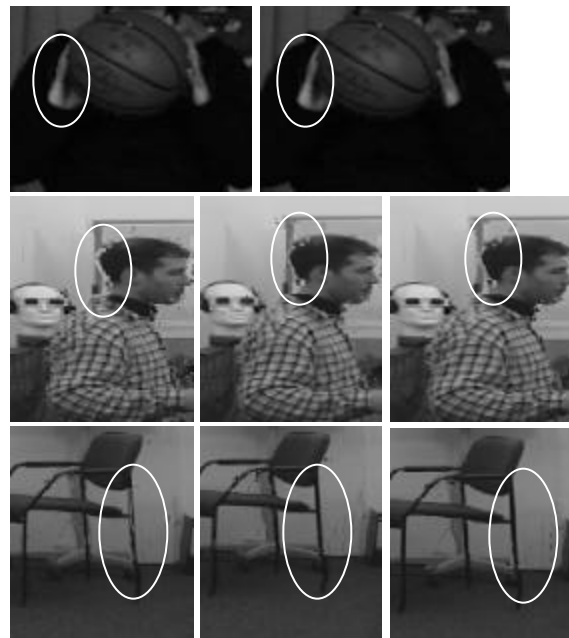


Fig. 5: Result of general algorithm, (a) Basket ball sequence, left top is the original virtual view image without matting refinement, while right top is the output after Bayesian matting refinement, Spectral matting is not appropriate to this sequence as these images are not visible components, (b) and (c) Cut out from the output of book_arrival sequence, Left ones are original virtual views, middle ones are refined by general Bayesian matting and right ones are refined by spectral matting, Regions of interest are highlighted

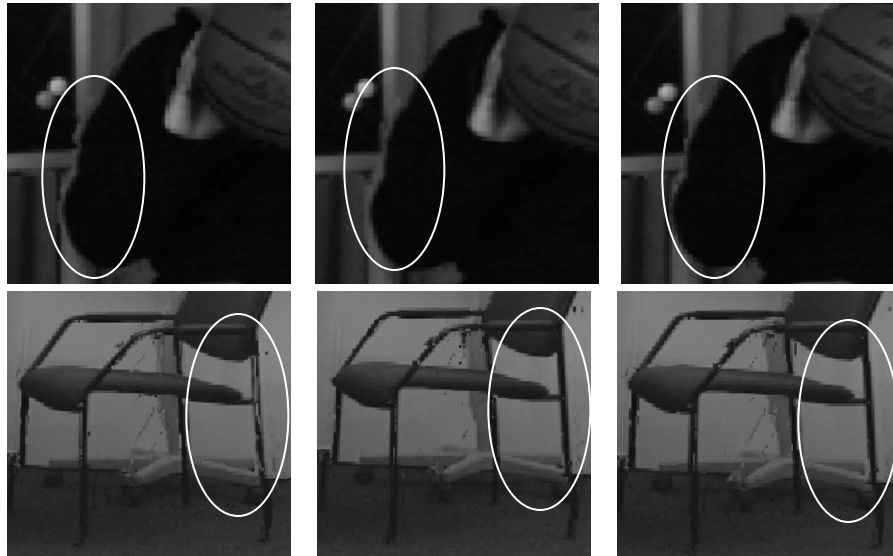


Fig. 6: Result of localization of Bayesian matting, (a) Basketball sequence; left is the original virtual view images, middle ones are outputs of the general Bayesian matting refinement and right ones are outputs of localization of Bayesian matting. Regions around the left arm are shrunk, (b) Output of book_arrival sequence with the same order, Regions of interest are highlighting

Book_arrival of HHI and basketball sequence made by our lab are shown in this study. Figure 5 shows the result of first frame of basketball sequence and first frame of book_arrival sequence, views of camera 8 and camera 10 are chosen as the input of our virtual view synthesis project. Figure 6 shows the refined virtual view image of localization methods by using Bayesian matting.

CONCLUSIONS

Mixed boundary pixels must be correctly sorted in virtual view synthesis in 3DTV system. Depth values of unreliable boundaries pixels should be recalculated. Our boundary matting methods and localized methods can effectively achieve this goal. We have tested this approach on several databases the results show that boundary artifacts are less than before. The unique point in this paper is that multiple-layer definition problem is mentioned and rightly solved by localization of matting algorithm. Limitations still remain: (1) matting algorithm should be improved. Although Bayesian matting provides accurate matting result, it is time consuming and initialization of trimap often causes problem. Spectral matting pay less reliable on constrains however it is not applicable for non-visible segmented images, (2) constrains or user guidance. Trimap is the general method to guide matting process. The precision of matting relies

on the segmentation of trimap. The generation of trimap should be further developed, (3) Estimation of new depth value of boundary pixels. In this study, we choose the extreme value in neighborhoods for simplicity. This process should be more effective and reliable. In future work, we will focus on these three limitations and improve our boundary refinement technique.

ACKNOWLEDGMENTS

This study was supported in part by the National Natural Science Foundation of China (Grant No. 60802013, 60971061, 61072081), the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2009ZX01033-001-007), Key Science and Technology Innovation Team of Zhejiang Province, China (Grant No. 2009R50003) and China Postdoctoral Science Foundation (Grant No. 20110491804).

REFERENCES

- Chuang, Y.Y., A. Agarwala, B. Curless, D.H. Salesin and R. Szeliski, 2002. Video matting of complex scenes. Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, July 21-26, 2002 ACM, New York, pp: 243-248.

- Chuang, Y.Y., B. Curless, D. H. SalesIn and R. Szeliski, 2001. A Bayesian approach to digital matting. *Comput. Vision Pattern Recognit.*, 2: 264-271.
- Hasinoff, S.W., S.B. Kang and R. Szeliski, 2006. Boundary matting for view synthesis. *Comput. Vision Image Understanding*, 103: 22-32.
- Kies, K., N. Benamrane and A. Benyettou, 2005. 3D medical image segmentation and surface modeling using the power Crust. *Inform. Technol. J.*, 4: 377-381.
- Levin, A., A. Pav-Acha and D. Lischinski, 2008a. Spectral matting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30: 1-14.
- Levin, A., D. Lischinski and Y. Weiss, 2008b. A closed form solution to natural image matting. A closed form solution to natural image matting. *Trans. Pattern Anal. Machine Intell.*, 30: 228-242.
- Mitsunaga, T., T. Yokoyama and T. Totsuka, 1995. AutoKey: Human assisted key extraction. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive techniques*. August 06-11, 1995, Los Angeles, CA, USA, pp: 265-272.
- Ng, A. M. Jordan and Y. Weiss, 2001. On Spectral clustering: Analysis and an algorithm. *Adv. Neural Inform. Proc. Sys.*, 14: 849-856.
- Nsib, F., N. Ayed and Y. Chevalier, 2008. Matting agent concentration and its effect on the colour and the rheology of matted coatings. *J. Applied Sci.*, 8: 1527-1533.
- Ruzinor, C.M., A.R.M. Shariff, A.R. Mahmud and B. Pradhan, 2011. Online 3D terrain visualization: Implementation and testing. *J. Applied Sci.*, 11: 3247-3257.
- Ruzon, M. and C. Tomasi, 2000. Alpha estimation in natural image. *Comput. Vision Pattern Recognit.*, 1: 18-25.
- Sharma, G., L. Baoku and W. Lijuan, 2011. Online marketing in second life virtual world. *Asian J. Mark.*, <http://scialert.net/abstract/?doi=ajm.0000.35544.35544>
- Shuchun, Y., Y. Xiaoyang, S. lina, Z. Yuping and S. Yongbin *et al.*, 2011. A reconstruction method for disparity image based on region segmentation and RBF neural network. *Inform. Technol. J.*, 10: 1050-1055.
- Sun, J., J. Jia, C.K. Tang and H.Y. Shum, 2004. Poisson matting. *Proc. ACM SIGGRAPH.*, 23: 315-321.
- Wang, J. and M.F. Cohen, 2007. Image and video matting: A survey. *Found. Trends Comput. Graphics Vision*, 3: 97-175.
- Wang, W., X.H. Huang and M. Wang, 2010. Out-of-sequence measurement algorithm based on Gaussian particle filter. *Inform. Technol. J.*, 9: 942-948.
- Wexler, Y., A. Fitzgibbon and A. Zisserman, 2002. Bayesian estimation of layers from multiple images. *Eur. Conf. Comput. Vision*, 3: 487-501.
- Yao, L., D.X. Li and M. Zhang, 2012. Temporally consistent depth maps recovery from stereo videos. *Inform. Technol. J.*, 11: 30-39.
- Yu, S., D. Yan, Y. Dong, H. Tian, Y. Wang and X. Yu, 2011. Stereo matching algorithm based on aligning genomic. *Inform. Technol. J.*, 10: 675-680.
- Zhixiang, T., W. Hongtao and F. Chun, 2011. Target capture for free-floating space robot based on binocular stereo vision. *Inform. Technol. J.*, 10: 1222-1227.