Precise 3D Reconstruction of Cultural Objects using Combined Multi-Component Image Matching and Active Contours Segmentation

Christos Stentoumis[1], Georgios Livanos[2], Anastasios Doulamis[2], Eftychios Protopapadakis[2], Lazaros Grammatikopoulos[3] and Michael Zervakis[2]

[1] National Technical University of Athens, Photogrammatric Lab, Zografou, Athens.
[2] Technical University of Crete, Image processing and Computer Vision Lab, Chania,
[3] Technological Education Institute of Athens, Depart. of Topography, Aegaleo, Athens.

Abstract. Cultural and creative industries constitute a large range of economic activities. Towards this expansion we need to state the inclusion of ICT technologies, as such of 3D reconstruction methods. However, precise 3D reconstruction under a computationally affordable manner is a research challenge. One way to precisely reconstruct a cultural object is through the use of photogrammetry with the main goal of finding the correspondences between two or more images to reconstruct 3D surfaces. A cultural object is often surrounded by visual background data that should be excluded to improve 3D reconstruction accuracy. Background conditions dynamically change, especially if the object is captured under outdoor conditions, where many occlusions occur and the shadows effects are not negligible. In this paper, we propose a combine image segmentation and matching method to yield an affordable 3D reconstruction of cultural objects. Image segmentation is performed on the use of active contours while image matching through novel multi-cost criteria optimization functions. Experimental results on real-life ancient column capitals indicate the efficiency of the proposed scheme both in terms of performance efficiency and cost.

1. Introduction

Surveys indicate that cultural and creative industry represents 4.5% of total European GDP and for 3.8% of the workforce [1]; so, this sector of economy can constitute one of the main European engines for growth and job creation. A crucial driving force for the development and economic growth of the creative industries is ICT technologies, which can open new frontiers in growing areas of the creative sector. 3D reconstruction and modeling of tangible cultural objects constitutes a significant task in digitalization area. Our world is a 3D world, and we perceive most of the events occurring in this world by depth information. One way to precisely reconstruct a cultural object is through the use of photogrammetry with the main goal of finding the correspondences between two or more images to reconstruct 3D surfaces. Image matching remains an active research field since finding a unique match or no match at all (occlusions, light variations, geometry distortion) is in fact an ill-posed problem, which can be solved only if suitable constraints are set.

However, a cultural object is often surrounded by visual background data that should be excluded to improve 3D reconstruction accuracy. Background conditions dynamically change, especially if the object is captured under outdoor conditions, many occlusions occur and the shadows effects are not negligible. Thus, new sustainable and innovative computer vision tools should be investigated for automating the capturing technology, able to maximize performance, computational complexity while maintaining the precision in 3D reconstruction.

1.1 Previous Works

The collection of data for 3D modeling can be done by using *passive* methods, e.g. *shape from X* (X: stereo, shade, focus etc.), and /or *active* methods as time-of-flight (ToF), phase shift technology, or structured light scanners. Although, 3D point clouds can be successfully created by using range cameras, the approaches based on this technology are not able to capture the texture of the scene [2]. In [3] a system based on Kinect$^{\mathrm{TM}}$ is proposed to bridge this gap.

On the other hand, matching techniques have been proposed in the literature for describing the dissimilarity between potentially corresponding pixels. [4] evaluates the cost function itself under different optimization schemes in a thorough survey of stereo-methods. Non-parametric image transformations, such as *rank* and *census* [5], produce robust results based on relationships of pixels with their neighborhood. In [6] a dissimilarity measure was proposed to cope with differences in image sampling. Recently, the *mutual information* approach has been proposed for effectively handling radiometric differences [7]. Here, a combination of methods is proposed; the benefits of this approach are thoroughly discussed in [8]. Local approaches of stereo-matching are based on the definition of pixel *neighborhood*. [9] discusses the *support region* formation, where the costs are aggregated to enhance the cost of **p**. A variety of methods exist to enhance this *cost aggregation* step; in [10] weights are attributed in a constant-sized neighborhood around each pixel in accordance with color similarity and geometric proximity; shiftable windows change the position of the central pixel of a support region [11]; and shape-adaptive windows can be based on separate circular sectors across multiple directions around a pixel [12].

To improve 3D reconstruction efficiency image *segmentation* techniques are expoited. Image segmentation extracts attributes of interest considering common properties, such as discontinuities and similarities, within different object classes. Several approaches have been introduced in literature; *point and line detection* techniques where the detected edges are linked in order to accurately represent the shape of each object, *thresholding methods* (histogram, adaptive, multi-level) that divide the image into segments according to distinct bands of pixel intensities and *region growing/splitting methodologies* which iteratively classify neighboring pixels of "seed-points" into a region through appropriately selected similarity criteria [13].

1.2 Our Contribution

In this paper, we propose an innovative 3D reconstruction methodology for cultural heritage objects, by combining state of the art image matching algorithms with novel image segmentation techniques. Initially the image segmentation algorithm is applied on the 2D data to precisely localize object boundaries in color domain. This way, we remove noise in the 3D reconstruction space since the matching algorithm focuses only on the cultural object of interest instead of exploiting the entire imagery plain. Simultaneously, computational cost is significantly reduced since the disparity space is limited on the foreground object. For image segmentation, we use active contours that evolve curves on color domain to obtain the segmentation. Selection of this segmentation technique among many others existing in the literature is due to its efficiency as regards precise localization of the object contours and its robustness in illumination variations since our approach faces outdoor image content. As far as 3D reconstruction is concerned, in this paper, we adopt a methodology that combines state of art techniques and novel considerations regarding a multi-component cost function, an adaptive support region and geometrically constrained 3D smoothing over the cost volume. Experimental results on real-life cultural heritage objects like ancient column capitals of Acropolis reveal the effectiveness of the proposed combined methodology in automatically and simultaneously precisely reconstructed tangible 3D cultural objects from high resolution images.

2 Active Contour based Curve Evolution

In our case, the input images constitute of a column capital bedded on a platform placed on an outdoor environment. Such images usually contain textures, noise, shading, abrupt variations of lighting, background of numerous and overlapping items, thus introducing additional limitations to segmentation techniques and preventing them from converging to accurate object boundaries [14]. As a first step, a transformation to a different color space is required. The RGB model appears adequate for digital representation but unproved for color segmentation. HSV (Hue-Saturation-Value) and Lab (Luminance - a-b (color-opponent dimensions)) models decorrelate the pixel intensity from the pure color components, facilitating the detection of specific color bands, while in the RGB space color information originates from the combination of all the channels (Red, Green, Brown). In addition, the application of adaptive thresholding to the S-V and a-b channels is capable of isolating the foreground object from the background scene. Another limitation arisen through the segmentation procedure is the shading of the objects of interest. To handle this, histogram equalization of the pure color bands is used.

Once the image defects have been subdued through the preprocessing procedure, an appropriate segmentation algorithm must be selected. Contour-based techniques are well established in international bibliography, providing accurate and robust results even in noisy environment, having the drawback of suffering from initialization, local minima and stopping criteria problems. For this reason, we select these approaches in this paper. The principle of these techniques lies on the linking of edge points extracted via an edge detection scheme, attempting to exploit curvilinear continuity to iteratively approximate the borders starting from a closed curve [15].

Then, active contours object detection is applied, by combining curve evolution techniques, level sets and the Mumford-Shah functional, accomplishing to detect corners and any topological change as in [16]. The model begins with a contour in the image plane defining an initial segmentation and then this contour gradually evolves according to a level set method until it meets the boundaries of the foreground region. According to the model, a curve S is represented via a function φ (the level-set function) as $S=\{(x, y)|\varphi(x, y)=0\}$, where (x, y) are coordinates in the image plane while the evolution of the curve is given by the zero level curve at time t of function $\varphi(x, y, t)$. Negative values denote points outside the curve while positive values originate from points belonging to the internal area of the curve, as depicted in Fig.1.

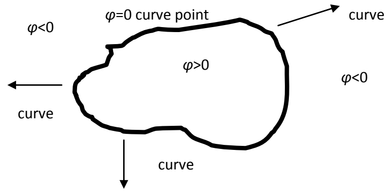


Fig. 1. An example of the proposed method used for curve evolution

At any given time, the level set function simultaneously defines an edge contour and a segment being evolved according to the partial differential Eq. (1), iteratively converging to a meaningful segmentation of the image.

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| F, \phi(x, y, 0) = \phi_O(x, y) \tag{1}$$

where F denotes the speed of the curve evolution.

3. Multi-Component Cost Function Image Matching

The matching process discussed in this paper is wrapped in a hierarchical scheme. Processing of high resolution images is necessarily based on scaled representations of the stereo-pair. The aim is to limit the disparity search space to a computationally feasible range, and also guide matched disparities in a coarse-to-fine context through scale-space. This approach also reveals structures in different layers of image pyramids, which lead from a rough, yet close to reality, 3D surface to finer detail as one proceeds through the image pyramid.

The three matching costs that form the complete matching function C are: the census transformation on image gradients, C_c' (expressed through the Hamming distance), the absolute difference in colour values, C_{ADc}, and the absolute difference on principal image gradients, C_{ADg}. A robust exponential function [17] which resembles a Laplacian kernel [see Eq. (2)], has been preferred to model these costs

$$C(x,y,d) = 1-\exp\left(-\frac{C_c}{\lambda_c}\right) + 1-\exp\left(-\frac{C_{ADc}}{\lambda_{ADc}}\right) + 1-\exp\left(-\frac{C_{ADg}}{\lambda_{ADg}}\right) \tag{2}$$

The terms of C have the quality of truncating costs that are too large, thus preventing outliers from pervading the matching cost. The values of C_c, C_{ADc} and C_{ADg} are also scaled in the same value field, assuming suitable selection of respective regularization factors λ, since each term of C takes values in the field [0, 1), and thus C is always positive. The impact of each dissimilarity measure on the overall cost can be tuned by adjusting the values of each λ.

Census on intensity principal derivatives: To implement census cost C_c we first evaluate the census transformation T_C, being a non-parametric image transformation [5]. For a support neighborhood $N_{m\times n}$ of a pixel **p**, a binary vector forms a map of neighbouring pixels with intensities $I(\mathbf{p})$ less than that of **p**. Unlike usual approaches, in the present implementation the transformation is performed not on grey-scale image intensity function I, but on its principal derivatives $\partial I/\partial x$, $\partial I/\partial y$ [8] providing an extended binary vector

$$T_c(p) = \underset{p\in\left\{\frac{\partial I}{\partial x},\frac{\partial I}{\partial y}\right\}}{\otimes} \; \underset{q\in N_p}{\otimes} c(p,q) \tag{3}$$

where ⊗ denotes the act of concatenation, following the original definition of T_C. In Eq. (3), $c(p,q)$ yields zeros values if $I(\mathbf{p}) \leq I(\mathbf{q})$, otherwise is one.

Census transformation T_C depends on how a pixel relates to its surroundings within the image patch, thus T_C is robust against individual outliers around discontinuities and noisy pixels. The direct introduction of the gradients in two image directions into the binary vector doubles the size of the produced vector T_c, thus exploiting the representational potential of image gradients. Finally, the matching cost C_c between a pixel **p** of the reference image and its corresponding pixel **p**' in the matching image is calculated as the Hamming distance, which represents the number of unequal elements in the two binary vectors:

$$C_{census} = \sum_{x=1}^{n}\left(T_c^{ref}(p) \oplus T_c^{mat}(p')\right) \tag{4}$$

Absolute difference on image color: The absolute difference on color channels (ADc), or on intensity, is a simple and easily implementable measure, widely used in matching in the sense of L_1 norm. Though sensitive to radiometric differences, it has been proven as an effective measure when combined with flexible aggregation areas and referring to combination of all color layers. The cost term C_{ADc} is defined as the average absolute value over all three color channels. This turns out to improve results compared to matching on separate channels or grey-scale.

Absolute difference on image principal gradients: Here, the derivatives of image intensity in the two principal directions are extracted, and the sum of absolute differences of each derivative value in the x and y directions is used as a cost measure. The use of directional derivatives separately, i.e., before summing them up to a single measure ADg [see Eq. (5)], introduces into the cost measure the directional information for each derivative:

$$C_{ADg}(\mathbf{p},d)=\sum_{x,y}\left(\nabla I^{ref}\left(\mathbf{p}(x,y)\right)-\nabla I^{mat}\left(\mathbf{p}(x,y),d\right)\right) \tag{5}$$

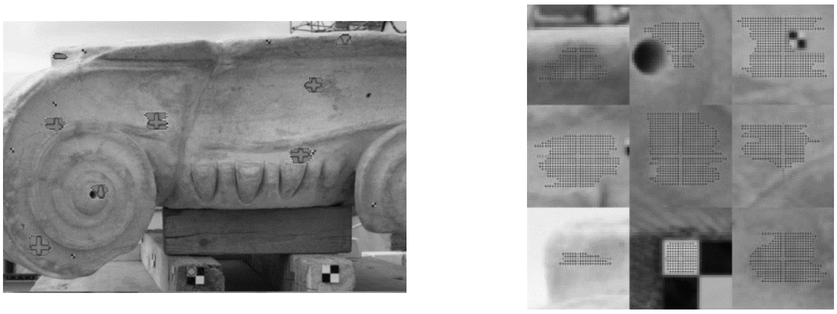
A Gaussian filter (size 3x3, $\sigma = 0.5$) is applied on the grey-scale images before calculating partial derivatives for reducing noise and smoothing around image edges.

3.1. Support Region for Cost Function

In our approach a modification of the cross-based support region approach introduced by [18] is used. *Adaptive* approaches are based on the fact that pixels of a support region ought to have similar colors and are expected to decrease in coherence with their distance from the reference pixel in image space. The construction of the aforementioned cross-based support regions is achieved by expanding around each pixel **p** a cross-shaped skeleton; the support region of **p** is defined by the combination of cross skeletons belonging to pixels in the neighborhood. In [19] a linear threshold is imposed on skeleton expansion based on color similarity $\tau(\cdot)$ of neighboring pixels:

$$\tau(l_q)=-\frac{\tau_{max}}{L_{max}}\times l_q+\tau_{max} \tag{6}$$

Variables in Eq. (6) express: a) the maximum semi-dimension $L_{\max}$ of the window size, b) the maximum color dissimilarity $\tau_{\max}$ between pixels **p** and **q** and c) l_q is the spatial distance between pixel **q** and **p**. The accepted difference τ between successive pixels is also restrained after [20]. Support regions generated according to the above considerations are presented in Fig. 2 for the data-set tested here.



(a) (b)

Fig. 2. Examples of the adaptive support regions formed with the linearly expanded cross-skeletons. (a)The overview of the processed image. (b) Image patches of the adaptation these windows have to image texture.

Aggregation is applied on cost using the *combined* support region W, which are the intersection of support regions of the reference pixel and its corresponding pixel on the matching image. As a result the support region is variable according to each possible disparity value. Aggregated pixel costs C_{aggr} are normalized by the number of pixels in the support region to ensure that costs per pixel have the same scale:

$$C(\mathbf{p},d) = \frac{C_{aggr}}{\left\|W(\mathbf{p},d)\right\|} \tag{7}$$

Cost aggregation is implemented through *integral* images [21], in order to achieve feasible computational load and real-time performance (for low resolution images).

3.2 Geometrically Constrained Smoothing of Cost Volume

The cost values of each pixel per each potential disparity value are stored in *Disparity Space Image* (DSI) representation, thus a *cost volume* is formed. This cost volume is smoothed through a 3D filter based on Gaussian distribution and geometric constraints regarding matching. Cost filtering satisfies the need for 3D support of the aggregation step, since usual cost aggregation is defined on 2D and has the inherent limitation of assuming that all pixels in a neighborhood share the same depth (fronto-parallel assumption). Here, we propose the weighted aggregation of 2D aggregated costs $C_0(x,y,d)$ belonging to geometrically possible disparities around a pixel through the convolution of cost volume with a 3D Gaussian filter:

$$C(x,y,d) = k * C_0(x,y,d) \tag{8}$$

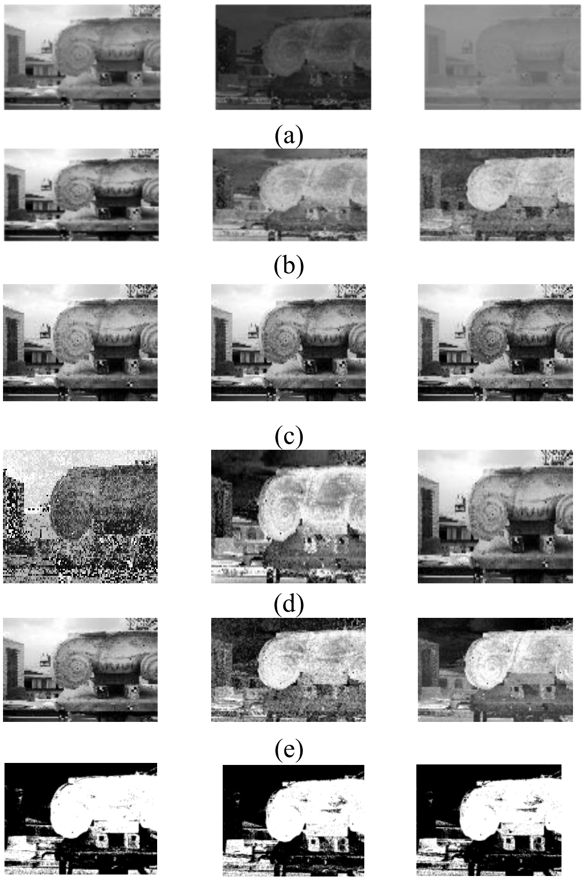
The Gaussian kernel k is adapted in order to serve the *ordering* and *uniqueness constraints* [22]. This kernel has the properties of attributing weights to neighboring costs inversely proportional to their spatial distance in the DSI. The advantage of this approach for 3D local support is that it avoids the need for explicit identification of slanted surfaces in 3D world space.

The estimation of disparity is carried out in the 'winner-takes-all' mode, as in most local and semi-global approaches, i.e., the disparity label with the lowest cost is selected. The estimated disparity map is refined through a robust post-processing procedure which includes left-right consistency check, outlier median smoothing via cross-based regions, occlusion/ mismatch labeling, sub-pixel estimation and edge-preserving smoothing on the disparity map [8, 19].

4. Experimental Results

4.1 Segmentation Results

In this section, we present segmentation results of the proposed active contour methodology. Fig. 3 presents the segmentation results. In this figure, we initially depict the original images along with all the segmentation steps as being described in Section 2.


(a)

(b)

(c)

(d)

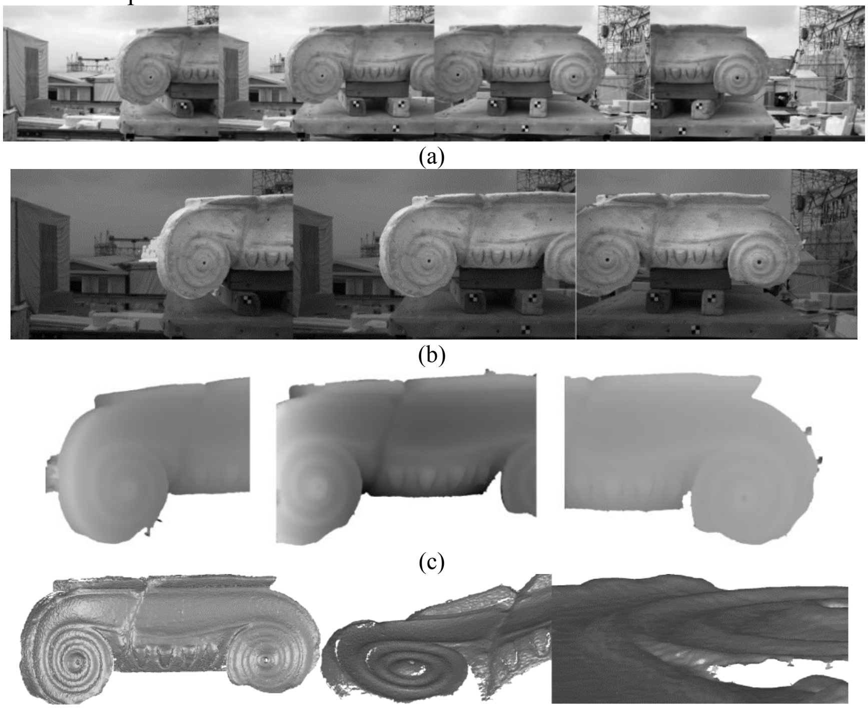
(e)

(f)

Fig. 3. (a) Original images in the RGB, HSV and LAB color space. (b) Equalized images in each color space. (c) Equalized color channels of the RGB space. (d) Equalized color channels of the HSV space. (e) Equalized color channels of the Lab space. (f) mask containing the white objects (HSV space, Lab space and fusion of them).

4.2 Stereo Matching Results

The presented algorithm has been evaluated on an object of high archaeological interest; the column capital in Fig. 4 belongs to the temple of Athena Nike on the Acropolis of Athens (http://ysma.gr/en/athena-nike) and it is a typical part of ancient monuments with complex architectures. Column capitals are structural parts of ancient Greek temples. Thus, as the whole temple was decomposed during restoration, the visual and geometrical documentation should be thorough to ensure that the structural and the aesthetic restoration would be complete. The images seen in the top row of Fig. 4 (camera: 12 Mp Canon EOS5; pixel size: 8.24 μm) had originally been taken for creating an orthomosaic of the capital with conventional photogrammetric techniques and are a part of the Acropolis Restoration Service photographic archive. The intrinsic and extrinsic orientation parameters were determined with our automatic bundle adjustment software. The control points seen in the images (Fig. 4, 1st row) have simply served scaling purposes. Three pairs were used for complete reconstruction, but matching was based on stereo. The main object of interest was detected during the foreground step and it was described through a binary “mask”. These masks define the boundaries of the object and are transformed through the epipolar geometry of each stereo-pair.



(a)

(b)

(c)

(d)

Fig. 4. A highly detailed 3D reconstruction of an important cultural heritage object has been created from multiple-base stereo-matching. (a) Original high resolution images, (b) the recognized foreground in three of the images, that present the reference for the combined stereo-pairs, (c) disparity maps corresponding to each stereo-pair used for 3D reconstruction, (d) illustration the complete face of the capital and some surface details

Acknowledgment

This paper is supported by the project E-Park, "Exploitation of new Technological Trends for payment and handling public parking" approved under the Interreg III Programme, Greek-Cypriot cooperation and funded from European Union and Greek National funds.

References

1. Hesmondhalgh, D: The Cultural Industries. (2002) SAGE.
2. Yan Cui, S. Schuon, D. Chan, S. Thrun, C. Theobalt, "3D shape scanning with a time-of-flight camera," Computer Vision and Pattern Recognition (CVPR), pp.1173-1180, 2010.
3. S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P.t Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, A. Fitzgibbon, “KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera ” in UIST’ (2011), pp. 559-568.
4. Hirschmüller H., Scharstein D., 2009. Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans. on PAMI, 31(9), (2009) 1582-1599.
5. Zabih R., Woodfill J., 1994. Non-parametric local transforms for computing visual correspondence. In: Proc. European Conference on Computer Vision, pp. 151-158.
6. Birchfield S., Tomasi C., 1998. A pixel dissimilarity measure that is insensitive to image sampling. IEEE Trans. on Pattern Analysis and Machine Intelligence, 20(4), pp.401-406.
7. Hirschmüller H., 2008. Stereo processing by semi-global matching and mutual information. IEEE Trans. on Pattern Analysis and Machine Intelligence, 30(2), pp. 328-341.
8. Stentoumis C., Grammatikopoulos L., Kalisperakis I., Petsa E., Karras G., 2013. A local adaptive approach for dense stereo matching in architectural scene reconstruction. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XL–5/W1, pp. 219–226.
9. Tombari F., Mattoccia S.,Di Stefano L., Addimanda E., 2008. Classification and evaluation of cost aggregation methods for stereo correspondence. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8.
10. Yoon K.J., Kweon I.S., 2006. Adaptive support-weight approach for correspondence search. IEEE Trans. on Pattern Analysis and Machine Intelligence, 28(4), pp. 650-656.
11. Bobick A.F., Intille S.S., 1999. Large occlusion stereo. IJCV, 33(3), pp.181–200.
12. Foi A., Katkovnik V., Egiazarian K., 2007. Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images. IEEE Transactions on Image Processing, 16(5), pp. 1395-1411.
13. R. C. Gonzalez and R.E. Woods, Digital Image Processing 2nd Edition, Prentice Hall, New Jersey, 2002.Florida, Richard (2002), The Rise of the Creative Class and How It's Transforming Work, Leisure and Everyday Life, Basic Books.
14. D. Markovic and M. Gelautz, "Experimental Combination of Intensity and Stereo Edges for Improved Snake Segmentation", Pattern Recogn. and Image Analysis, Vol 17, No 1, pp. 131-135, 2007
15. Jitendra Malik,Serge Belongie, Thomas Leung, Jianbo Shi, "Contour and texture analysis for image segmentation", Inter. Journal on Computer Vision, Issue 43, pp 7-27, 2001
16. Tony F. Chan, Luminita A. Vese, "Active Contours Without Edges", IEEE Transactions on Image Processing, Vol 10, No 2, 2001
17. Yoon K.J., Kweon I.S., 2006. Adaptive support-weight approach for correspondence search. IEEE Trans. on Pattern Analysis and Machine Intelligence, 28(4), pp. 650-656.
18. Zhang K., Lu J., Lafruit G., 2009. Cross-based local stereo matching using orthogonal integral images. IEEE Trans. on CSVT, 19(7), pp. 1073-1079.
19. Stentoumis C., Grammatikopoulos L., Kalisperakis I., Karras G., 2012. Implementing an adaptive approach for dense stereo-matching. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVIII/5, pp. 309-314
20. Mei X., Sun X., Zhou M., Jiao S., Wang H., Zhang X., 2011. On building an accurate stereo matching system on graphics hardware. In: Proc. ICCV Workshop on GPU in Computer Vision Applications, pp. 467-474.
21. Viola P., Jones M., 2001. Rapid object detection using a boosted cascade of simple features. In: Proc. IEEE Conf. on CVPR, pp. I-511-518.
22. Yuille A.L., Poggio T., 1984. A generalized ordering constraint for stereo correspondence, MIT, AI Lab., Memo 777.