SEGMENTATION OF IMAGE PAIRS FOR 3D RECONSTRUCTION

Hani. M. Mohammed *, Naser El-Sheimy

Department of Geomatics Engineering, University of Calgary, 2500 University Dr. N. W. Calgary, AB, Canada T2N 1N4 - (hmmohamm, elsheimy)@ucalgary.ca

ICWG II/III: Pattern Analysis in Remote Sensing

KEY WORDS: Segmentation, Disparity map, Homography, Image pair, 3D reconstruction, Camera geometry

ABSTRACT:

Image segmentation is an essential task in many computer vision applications such as object detection and recognition, object tracking, image classification, 3D reconstruction. Most of the current techniques utilise the colour or grayscale information of an image without considering the camera geometry. In this paper, a method is proposed to utilise the camera relative orientation of a pair of images to find a reliable object segmentation. The inputs to the method are a rectified image pair and a disparity map which could be computed from the rectified image pair, the disparity map is used to determine a set of local homographies between planar surfaces in the two images. The planar surfaces are corresponding to image segments despite the inconsistency of the RGB information. Homography based segmentation alone is not reliable due to possible noise in the disparity map and existence of non-planar objects in the scene. Therefore, an RGB technique is used as a complementary approach to enhance the segmentation result. Two colour-based segmentation techniques are used here, the first is the colour edge detector, and the second is Grabcut. Experimental results show the although the colour edge detector is a simpler algorithm than Grabcut, it does not include noisy data in the segmentation results. the This useful for 3D reconstruction, as it is preferable to exclude noisy areas like the sky and window glass. The outcome of the proposed segmentation algorithm is an object-based segmentation of the pair of images as well as a segmented disparity map.

1. INTRODUCTION

Image segmentation is a fundamental task in many applications, including 3D reconstruction, classification, object recognition, and motion detection. Most of the current segmentation algorithms are based solely on the radiometric properties of the image. Furthermore; in most cases, only a single image is used at a time. However; there are some scenarios in which multiple images for the same scene exist. For instance, in the problem of 3D reconstruction, the scene is reconstructed from multiple images, and therefore, segmentation can be performed over a sequence of images to enhance the segmentation quality. In addition, one can take into consideration the geometrical aspects of image pairs or triplets.

In this research, a new method of image segmentation is proposed in which both the radiometric and geometric properties of a sequence of images are utilised. The proposed method is focusing on the segmentation for 3D reconstruction, in which image segmentation could be used prior to the reconstruction process (e.g. to enhance the disparity map) and could be used for post-processing of the constructed point cloud (e.g. point cloud classification). There are other possible applications of the proposed method, such as object detection and recognition, especially when a depth camera is available onboard.

In the proposed method, pairs and triplets of images are considered. Images can be taken by stereo cameras (stereo-rectified) or can be taken as a sequence. The images in each pair or triplet are related by geometrical constraints, such as the fundamental matrix and local homographies between corresponding planes in those images. We utilise the information from the geometrical entities to build an initial segmentation (homography based segmentation), then enhance it with the RGB information associated with the images.

The proposed method consists of five main steps. First, image rectification is performed, if necessary, and therefore, a disparity map can be computed using any of the available algorithms, such as the Semi-Global Dense Matching (SGM) (Hirschmüller 2008). Since the dense image matching is an essential step towards the 3D reconstruction and assuming the disparity is already computed, it could be used without worrying about adding extra processing time. Furthermore, in many applications, a depth camera is used along with the RGB cameras, and therefore, the depth map already exists. In the second step, a region growing algorithm with a small threshold is applied to the disparity map to segment it based on the changes in depth. The threshold is selected such that the allowed change of disparity values is within one or two pixels when the object is far from the camera (i.e. having small disparity values) and the threshold increases when the object is close to the camera (i.e. having larger disparity values).

The disparity map will be over segmented with different segments representing different depths or distances from the camera. This segmentation is not very accurate due to several factors, such as the amount of noise or errors in the computed disparity, and the errors resulting from the region growing algorithm. However; we can consider the over-segmentation as just an initial step and that it could be enhanced over time.

Since each segment is locally uniform (i.e. disparity values are changing smoothly), then we can consider that each segment represents an approximated planar surface in the scene. Therefore, we can fuse the disparity with the pair of images to approximate a local homography for each segment in the disparity, which is the third step in the proposed method. In the next step, we use colour segmentation algorithms to enhance the initial segmentation. We first the colour edge detector to specify a boundary for the homography based segmentation. Then, in the fourth step, we extend the algorithm to work with three images

* Corresponding author

instead of just two. Finally, Grabcut algorithm is used to find an accurate and automatic segmentation of the scene. Grabcut requires three trimaps, one for the background, another for the foreground and the third trimap is for the unknowns. We can construct these trimaps from the initial segmentation performed in the second step.
It is well known that Grabcut and several other algorithms require user intervention to feed the algorithm with the trimaps. Therefore, our contribution is to automate the creation of the trimaps from the initial segmentation.
The next section provides an overview of the related work and paper contribution. This section is followed by a section of a detailed discussion of the proposed method. Section (4) provides the results and discussion of the experimental tests, followed by the conclusion section.

2. RELATED WORK AND PAPER CONTRIBUTION

Image segmentation has been used in a wide range of applications. Image segmentation algorithms can be categorized based on how they perform and on their applications. Classical segmentation algorithms could be either edge-based segmentation algorithms or region-based segmentation algorithms. One of the most famous examples of edge-based segmentation is Canny edge-detector (CANNY 1987). Region-based segmentation includes the flood fill algorithm (Heckbert 1990). Recently other smart algorithms and methods were implemented to deal with more complex scenarios. Watershed (Meyer 1992) and Grabcut (Rother, Kolmogorov, and Blake 2004) are considered region growing algorithms for the segmentation of colour images. Grabcut is one of the most reliable algorithms for segmenting an object in an image. However; it requires human intervention. On the other hand, there are several machine learning and deep learning algorithms to segment and classify objects in the scene, but such algorithms require the training of a massive dataset in order to perform accurately and efficiently.
There are many other variants of those algorithms, but all of them depend only upon the RGB or grayscale information of the image without considering the geometrical characteristics of the scene or the camera.
In this research, we propose a new method in which both the geometry and RGB contents are utilised to obtain an automatic and reliable object segmentation. The method starts with a pair of images and a disparity map, the output is a labelling matrix that can be used to overlay the original image pairs and the disparity map. So, the proposed method segments not only the image but also the disparity map. Furthermore; the method is being enhanced once by using a colour edge detector to eliminate the disparity noise, and another time by combining the proposed method with the Grabcut algorithm. The overall method proved to be automatic and reliable, especially when applied to images of variable textures.

3. METHODOLOGY

3.1 Overview of the proposed method

Figure (1) summarizes the main steps in the proposed method. We start with two image pairs, then compute the rectification transformation, so that the images can be stereo rectified, and a disparity map could be computed. This is followed by the initial homography based segmentation, the colour edge detection and Grabcut. The method is discussed in details in the following subsections.

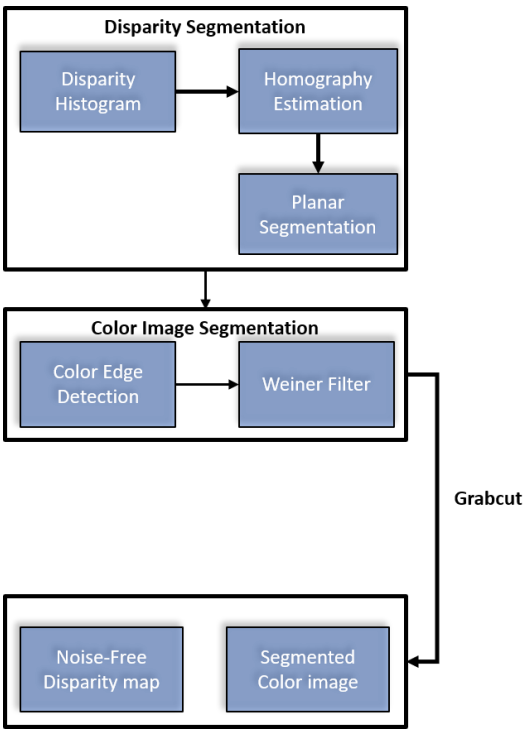


Figure 1. workflow of the proposed method

3.2 Image rectification

In the proposed method, we start with a pair of images that were taken by a single camera; therefore; the images are not initially rectified. The two images need to be stereo-rectified such that the epipolar lines intersect at infinity, and a scanline matching is then performed. There are two famous rectification algorithms, Hartley's uncalibrated stereo rectification algorithm (Hartley and Zisserman 2003) and Bouguet's calibrated stereo rectification algorithm (Kaehler and Bradski 2016).
In this paper, we consider the uncalibrated stereo rectification scenario, in which the relative orientation between the two images is unknown. However; it might be convenient for some readers to assume calibrated stereo rectification if the relative orientation of the image pair is known beforehand.
Hartley's algorithm is straight forward, and it is summarized in the following steps:

1. Perform image feature detection and matching between the image pair.
2. Compute the fundamental matrix.
3. Find the two epipole of the two images using the fundamental matrix from:

$$\boldsymbol{F}.\boldsymbol{e_l} = \mathbf{0} \quad , \quad \boldsymbol{e_r}.\boldsymbol{F} = \mathbf{0} \tag{1}$$

where $\boldsymbol{F}$ is the fundamental matrix and $\boldsymbol{e_l}$ and $\boldsymbol{e_r}$ are the epipoles of the left and right images respectively.

4. Find a homography $\boldsymbol{H_r}$ that maps the right epipole to infinity:

$$\boldsymbol{e_r} \longrightarrow (1,0,0)^T \tag{2}$$

5. Find a homography $\boldsymbol{H_l}$ that maps the left epipole to infinity as well:

$$\boldsymbol{e_l} \longrightarrow (1,0,0)^T \tag{3}$$

6. Apply the homography transformations to stereo rectify the left and right images.

The detailed description of Hartley's algorithm could be found in (Hartley and Zisserman 2003).
After the image stereo rectification, the disparity map is computed using any scanline matching method. In this paper, the SGM is used, since it is considered as the state-of-the-art method in terms of accuracy and time efficiency.

3.3 Initial Segmentation of the disparity map

The disparity associated with the image pair is generally smooth except in areas in which noise exists and in areas of edges and blobs. Therefore; we can assume that the disparity is smooth or having continuous values over planar surfaces. Furthermore; if an object is far enough from the camera, it will have small changes in the disparity from one point to another. Thus, we can assume that objects that are far from the camera are having smooth disparities as well, even if they are not representing planar surfaces.
The segmentation of the disparity map is performed based on the histogram of its values. Figure 2) shows an example of the histogram of the disparity of one of the test image pairs. Some peaks appear with a continuous change in the disparity around them. Therefore; it is possible to take each peak and perform region growing around it to obtain an initial segmentation of the disparity map. This initial segmentation results in an inaccurate over-segmented disparity map, which needs to be refined. Example of the initial region growing segmentation of disparity map is shown in Figure (3-b)

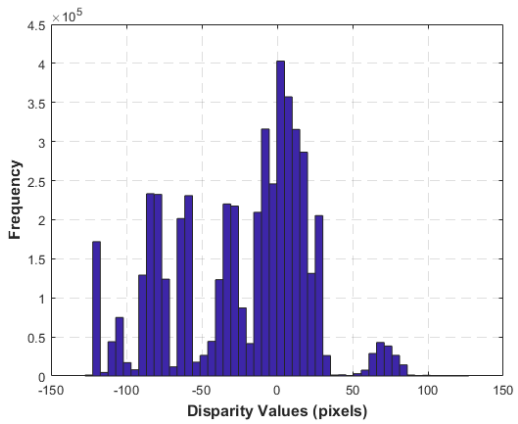
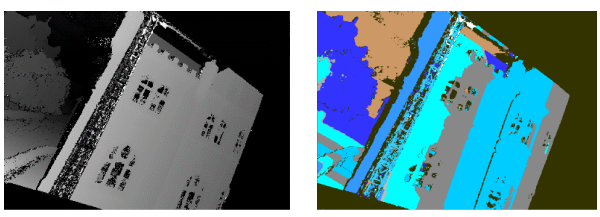


Figure 2. Example of the histogram of the disparity values in a disparity map.



(a) (b)

Figure 3. Example of the initial segmentation of the disparity map using the histogram and region growing.

3.4 Homography Based Segmentation

Homography based transformation is used to enhance the segmentation of disparity and to determine planar surfaces and approximated planar surfaces in the image pairs.
Each planar surface in the left image is connected to a planar surface in the right image via a local homography transformation. The homography transformation from one image to another requires the knowledge of the corresponding points between the two images. The disparity map can be used to find the correspondence between pixels in the left and right images. The correspondence relationship can be written as:

$$\boldsymbol{q} = \boldsymbol{p} + \boldsymbol{d}(\boldsymbol{q}) \tag{4}$$

where $\boldsymbol{q}$ and $\boldsymbol{p}$ are the pixel locations in the left and right images respectively, and $\boldsymbol{d}(\boldsymbol{q})$ is the disparity at the pixel position $\boldsymbol{q}$.
It should be noted here that equation (4) holds for stereo rectified image pair, and an inverse transformation back to original images is required to find the correspondence between planar surfaces in the original images.
The initial segmentation of the disparity map is now used to obtain the correspondence relation in equation (4). First, the largest segment is considered. This segment is used in equation (4) to find some pixels positions $\boldsymbol{q_i}$ in the left image and their corresponding pixel positions in $\boldsymbol{p_i}$ in the right image. An inverse transformation $\boldsymbol{H_l^{-1}}$ is applied to the points $\boldsymbol{q_i}$ and similarly $\boldsymbol{H_r^{-1}}$ is applied to the points $\boldsymbol{p_i}$:

$$\boldsymbol{Q_i} = \boldsymbol{H_l^{-1}}\boldsymbol{q_i} \tag{5}$$

and

$$\boldsymbol{P_i} = \boldsymbol{H_r^{-1}}\boldsymbol{p_i} \tag{6}$$

Then it is required to estimate the local homography $\boldsymbol{H_j}$ between the points $\boldsymbol{Q_i}$ and the points $\boldsymbol{P_i}$:

$$\boldsymbol{Q_i} = \boldsymbol{H_j}\boldsymbol{P_i} \tag{7}$$

It is then expected that the homography relation in equation (7) is valid for all the points on the same planar surface. Therefore, all the points in the pair of images are tested using the inequality:

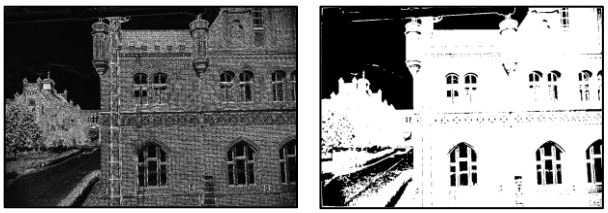
$$\boldsymbol{Q}_i - \boldsymbol{H_j}\boldsymbol{P_i} < \boldsymbol{\varepsilon} \tag{8}$$

where $\boldsymbol{\varepsilon}$ is a variable threshold that is selected based on the distance of the objects from the camera (or simply their disparity values) and the disparity range in a disparity map.
If the inequality in (8) holds for the points $\boldsymbol{Q_i}$ and $\boldsymbol{P_i}$ then these points belong to the same planar surface which is defined by $\boldsymbol{H_j}$.
As a result of this test, the over segmented disparity map in figure (3-b) will be refined, and points on the same planar surface will be connected, even if their disparity values are non-uniform.
The problem facing this method at this stage is the existence of noise in the disparity map. Noise in the disparity map leads to incorrect estimation of local homographies, and in some cases, the failure of points to pass the test in (8). In some other cases, the noise in disparity might result in inaccurate segmentation. Therefore; it is useful to use the images' RGB information at this stage of the method to enhance the overall segmentation quality.

3.5 Segmentation enhancement using RGB information

To eliminate the noise in the disparity map and the segmented image, we constraint the homography based segmentation using a colour edge detector. The colour edge detector we use here is based on the work by Silvano Di Zenzo (Di Zenzo 1986) and is implemented by Joao Henriques. The problem with any edge

detector is that it usually generates noise in the image. Therefore, we use the adaptive low-pass Wiener filter to reduce the amount of noise introduced by the edge detector (Rangayyan 2004). Figure (4) shows an example of the application of the colour image detector and Wiener filter to a test image.



(a) (b)

Figure 4. (a) Colour edge detector with noise, (b) color edge detector after applying Wiener filter.

To impose the constraints of the colour edges on the homography based segmentation, a binary image is created from the edge image, then the label matrix $\boldsymbol{L}$ is element-wise multiplied by the edge image $\boldsymbol{E}$. The resultant label matrix $\boldsymbol{L'}$ is given by:

$$\boldsymbol{L'} = \boldsymbol{L} \odot \boldsymbol{E} \tag{9}$$

3.6 Extension to a triplet of images

The combination of homography based segmentation and the colour information is reliable for the segmentation of objects that have the same geometrical characteristics. However; the proposed method is limited by the amount of noise in the disparity map and the occlusion in the scene caused by the camera geometry. In order to overcome these issues, a third image is added such that two image pairs are now available instead of one image pair. The first image pair consists of the first and second images in the sequence, and the second image pair contains the second and the third images. The same method steps will be applied to the new image pair to find an acceptable segmentation.

Now, we would like to fuse the segmentation results from the two image pairs. We can assume that the first image in the triplet is related to the second image by a projective transformation. Therefore; the first image can be mapped to the second image and vice versa.

$$\boldsymbol{I_1} = \boldsymbol{P}\, \boldsymbol{I_2} \tag{10}$$

where $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$ are the first and the second image respectively. We can then fuse the two segmentation by registering the two images and finding the discrepancy between each common segment.

3.7 Automation of Grabcut

One of the main drawbacks of Grabcut is the lack of automation and the requirement of successive human intervention to obtain good segmentation results. Grabcut requires three trimaps as input, one for the foreground, another for the background and the third for the unknown pixels.

The proposed method can provide the three trimaps to the Grabcut algorithm to further enhance the segmentation. So, the overall segmentation process becomes automated and more reliable.

Consider the result of the segmentation of one of the objects in the scene is a logical mask $\boldsymbol{B}$, in which the foreground is 1 and the background is 0. Then, the background and the foreground masks can be considered as:

$$\boldsymbol{B}, \qquad \boldsymbol{F} = NOT(\boldsymbol{B}) \tag{11}$$

The unknown map (mask) is chosen to be an area surrounding the foreground map.

4. RESULTS AND DISCUSSION

4.1 Dataset

The dataset was provided by the International Society of Photogrammetry and Remote Sensing (ISPRS) through the ISPRS and EuroSDR benchmark on multi-platform photogrammetry (Nex et al. 2015). The images were taken using an Unmanned Aerial Vehicle (UAV) in a close range mode over the area of Zollern Colliery (Industrial Museum) in Dortmund, Germany. A subset of three images was selected from the huge dataset based on the structure of the scene. The selected images contain several textures and non-planar objects, which give a good indication of the validity of the proposed method to perform under challenging conditions.

(a) (b)

(c)

Figure 5. Selected images from the Zollern Colliery.

4.2 Experimental Results

Figure (6) shows the disparity map of the first pair of images after performing the required stereo rectification. The noise in the disparity map is obvious and could affect the segmentation quality as well as the quality of the 3D reconstruction.

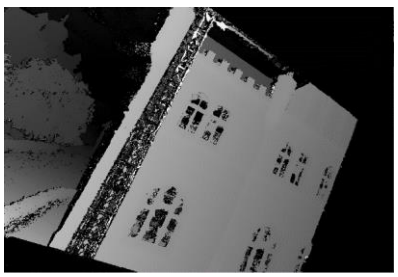
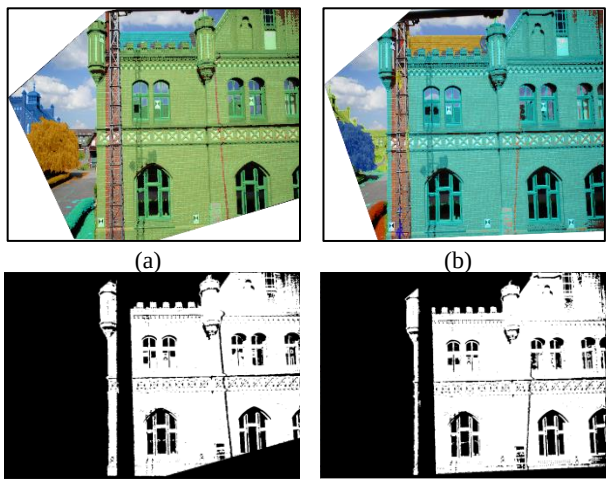


Figure 6. Disparity map of the first image pair computed using SGM.

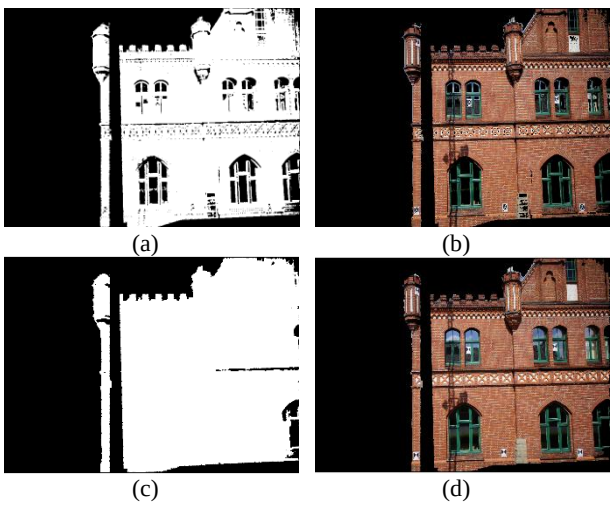
Figure (7) shows the results after applying the homography based segmentation and the color edge detector. The results in (7-a) and (7-b) are obtained using the first image pair, while the results in (7-c) and (7-d) are obtained using the second image pair. Areas of the segmented images were cropped as a result of the stereo rectification and remapping. These parts of the images could be retrieved if more images of the scene were added to the selected dataset.

It can be visually noticed that most of the objects in the scene are segmented, except parts of the road and the sky, which is a result of applying the colour edges and Wiener filter. It is preferable to exclude the sky from the segmentation, especially if the data are to be used for 3D construction, as the sky is a huge source of noise in the generated point cloud. On the other hand, roads could be retrieved if more images were to be added.
Although different structures, like trees, buildings and roads appear in the scene, the proposed method was able to segment most of those objects. In general, the Homography based segmentation is not reliable when dealing with objects of a complex structure like trees. However; when the objects are far from the camera, the change over disparity is small, and objects could be approximated by planar surfaces. But if the object is close to the camera, a relaxed threshold for the homography test should be used.



(a) (b) (c) (d)

Figure 7. The segmentation result before Grabcut. (a) segmented first image using the first image pair, (b) segmented second image using the second image pair, (c) binary map for the segmentation of the first image, (d) binary map for the segmentation of the second image.



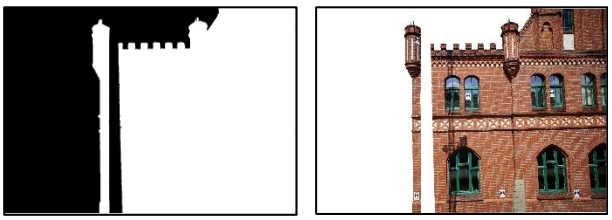
(a) (b) (c) (d)

Figure 8. The segmentation result after fusing a triplet of images and using Grabcut. (a) Binary map for the segmentation using three images, (b) Segmented object using three images, (c) binary map for the segmentation after Grabcut (d) Segmentation after Grabcut.

To judge the quality of the segmentation, only the main object in the scene is selected. In figure (8), the results after fusing the two image pairs and after using Grabcut are depicted. The quality of the segmentation has improved in both cases. Although it seems as if the segmentation with Grabcut is better than that with color edge detector, Homography based segmentation removes the points at which the disparity is incorrect. Thus, it has the benefit of cleaning the disparity map and therefore, reduce the noise in the generated point cloud at the stage of 3D reconstruction.
For example, windows are always a source of noise, as the reflection on the window's glass results in incorrect disparity values. Therefore; it is better to remove the window glass before the 3D reconstruction process.

4.3 Segmentation Quality Assessment

The confusion matrix and Cohen's kappa coefficient (Cohen 1960) are used here as a measure of the quality of the segmentation. Two classes are considered here: foreground and background. A ground truth image and mask were created using the GIMP software (https://www.gimp.org/) with the pencil selection tool.



(a) (b)

Figure 9. Ground truth for the segmented object.

Classified	Ground Truth	
	Foreground	Background
Foreground	2941416	583734
Background	8945	2465905

Table 1. The confusion matrix for the result of segmentation without Grabcut.

The kappa coefficient is given by the following equation:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{11}$$

where p_0 is the sum of the diagonal elements of the confusion matrix, and p_e is the sum of the product of the off-diagonal elements.

Classified	Ground Truth	
	Foreground	Background
Foreground	3257740	267410
Background	10562	2464288

Table 2. The confusion matrix for the result of segmentation with Grabcut.

The values p_0, p_e and the kappa coefficient for the segmentation without Grabcut are:

$$p_0 = 0.90$$
$$pe = 0.498$$
$$\kappa = 0.803$$

and for the segmentation with Grabcut are:

$$p_0 = 0.95$$
$$p_e = 0.529$$

$$\kappa = 0.90$$

Larger Cohen coefficient κ indicates better quality of the classification. It implies that the proposed segmentation method has an excellent segmentation result.

5. CONCLUSION

In this paper, a new segmentation method was proposed. The method utilises the geometrical as well as the radiometric characteristics of image pairs. Furthermore; development was made to extend the method to work over image triplets. Also, the method was combined with the Grabcut algorithm to provide an automated, reliable segmentation method. The method was tested using a challenging dataset with variable textures, and the method's accuracy and robustness were proved both visually and statistically in the

ACKNOWLEDGEMENTS

This project was funded by research grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair funds of El-Sheimy.

The authors would like to acknowledge the provision of the datasets by ISPRS and EuroSDR, released in conjunction with the ISPRS scientific initiative 2014 and 2015, led by ISPRS ICWG I/II.

REFERENCES

Canny, J., 1987. A Computational Approach to Edge Detection. *Readings in Computer Vision*, January, 184–203. doi.org/10.1016/B978-0-08-051581-6.50024-6.

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. doi.org/10.1177/001316446002000104.

Hartley, R., Zisserman, A., 2003. *Multiple View G eometry in Computer Vision*. *Cambridge University Press*. 1 . doi.org/10.1017/CBO9781107415324.004.

Heckbert, P.S., 1990. "Graphics Gems." In , edited by Andrew S Glassner, 275–77. San Diego, CA, USA: Academic Press Professional,Inc.http://dl.acm.org/citation.cfm?id=90767.90829.

Hirschmüller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328–41. doi.org/10.1109/TPAMI.2007.1166.

Kaehler, A., Bradski, G., 2016. *Learning OpenCV 3:Computer Vision in C++ with the OpenCV Library*. O'ReillyMedia, Inc.

Meyer, F., 1992. Color Image Segmentation. In *1 992 International Conference on Image Processing and Its Applications*, 303–6.

Nex, F., Gerke, M., Remondino, F., Przybilla, H.-J., Bäumker, M., Zurhorst, A., 2015. ISPRS Benchmark For Multi-Platform Photogrammetry. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W4, 135–42. doi.org/10.5194/isprsannals-II-3-W4-135-2015.

Rangayyan, Rangaraj, M., 2004. *Biomedial Image Analysis*. CRCPress.

Carsten, R., Kolmogorov, V., Blake, A., 2004. "'GrabCut': Interactive Foreground Extraction UsingIterated Graph Cuts." *ACM Transactions on Graphics*, 23(3), 309. doi.org/10.1145/1015706.1015720.

Silvano Di, Z.. 1986. A Note on the Gradient of a Multi-Image. *Computer Vision, Graphics and Image Processing*, 33 (1): 116–25. doi.org/10.1016/0734-189X(86)90223-9.