INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK

TECHNIKI INFORMACYJNE
TEORIA I ZASTOSOWANIA

Wybrane problemy
Tom 2 (14)

poprzednio

ANALIZA SYSTEMOWA W FINANSACH I ZARZĄDZANIU

Pod redakcją
Andrzeja MYŚLIŃSKIEGO

Warszawa 2012

INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK

TECHNIKI INFORMACYJNE
TEORIA I ZASTOSOWANIA

Wybrane problemy
Tom 2 (14)

poprzednio

ANALIZA SYSTEMOWA W FINANSACH
I ZARZĄDZANIU

Pod redakcją
Andrzeja Myślińskiego

Warszawa 2012

Wykaz opiniodawców artykułów zamieszczonych w niniejszym tomie:

Dr hab. inż. Andrzej MYŚLIŃSKI, prof. PAN

Dr hab. inż. Ryszard SMARZEWSKI, prof. KUL

Dr hab. Dominik ŚLĘZAK

Prof. dr hab. inż. Andrzej STRASZAK

Prof. dr hab. inż. Stanisław WALUKIEWICZ

Dr hab. Adam WIERZBICKI

ISBN 9788389475442

Disparity Map Detection Methods for Stereo Images for Augmented Reality

Konrad Koniarski

Studia Doktoranckie IBS PAN,
e-mail: *konrad.koniarski@gmail.com*

Streszczenie. In this paper, two popular methods of scene disparity estimation are compared for application in augmented reality. Augmenting computer rendered objects between real objects is challenging problem. Especially when image comes from video sequence or real time video. The proposed approach is base on the methods of disparity estimation for all pixels using stereo cameras. The group of pixels representing one object usually can be aggregated to one planar layer. Given the depth of any layer, the virtual object can be blended between the layers. Experimental results shows that using stereo calibrated cameras for scene depth estimation is effective. Graph cut method gives more accurate results then block matching method at a cost of lower speed.

Key words: Augmented reality, OpenCV, disparity map, stereo vision, image processing

1 Introduction

1.1 Augmented Reality

Computer systems are widely used to process image for enhance of particular details. In most cases image comes from real time video cameras. Computer systems can collect data using many different sensors, but this give engineers a challenging problem to solve. How to present data from many sensors on one screen? There are many approaches for data presentation in modern computer systems. The most widespread computer system interface in computer-human interactions is computer display. Computer displays are still improved and become more and more ergonomic and presents more detailed information. Display can be used for the presentation of image from cameras augmented for virtual information. Augmented reality techniques allow computer system to capture and process visual information and depending on the data provide various context information which are useful for system user. Recently, augmented reality technologies are widely used in education, entertainment, architecture,

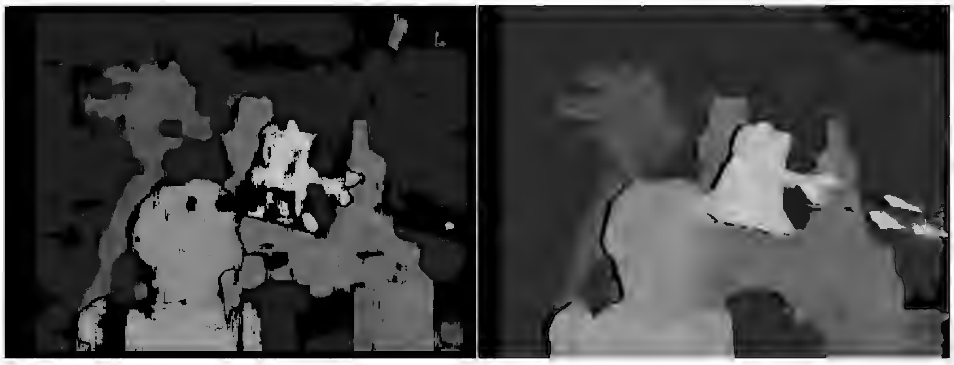
medical, machinery manufacturing and maintenance, advertisement, military, regional planning and many other important fields.
Virtual reality is widely used today, but augmented reality has a few advantages and become more and more popular. Augmented reality has better sense of reality because only artificial elements are added to displayed image. In augmented reality the real word is not processed into virtual model. In other hand, if only part of image is augmented then the virtual part should be composed into real objects in a natural way. There are many aspects that need to be taken under consideration like perspective projection, object obscuring or lightning.
The virtual reality is not giving as good interaction as the augmented reality systems. Augmented reality responds to any changes of real world scene and update presented information in real time. Augmented reality system depends on data that are gathered by various sensors. The most common types of sensors are optical systems like camera, GPS localization or magnetic field sensors. Some augmented reality systems need artificial markers that can be detected by sensor for identification and calibration of virtual objects.
Based on data gathered by sensors which describes real world scene and description of virtual objects augmented reality system can build one realistic scene, which can greatly enhance the interaction of created virtual scene.

Augmented reality systems can be split into four steps

- Gather information from sensors,
- Analyse the real world information and pre-processing,
- Generate virtual model and apply pose and projection transformation,
- Integration of virtual model and real scene.

In this research two frames from stereo cameras video sequence are used for computing the disparity map. Then disparity map is used for building virtual scene in augmented reality system. There are a few methods for creating disparity maps based on stereoscopic image. In this paper block matching and graph cut method are compared. Disparity maps created by block matching method is presented on Figure 1a and disparity created by graph cut method is presented on Figure 1b.

Traditional augmented reality systems use markers to augment virtual object in real scene. Markers are designed to be easy to recognize and the camera matrix can be recovered from detected marker pose. The position of virtual object is computed using camera matrix and the detected marker



(a) Block matching alghoritm (b) Graph cut alghoritm

Rys. 1: Disparity map results using block matching method

pose for correct virtual object projection. This approach usually allow placing virtual object in front of real image. To merge virtual object with real one and allow to obscure virtual object by real one disparity information is requested.

1.2 Related work

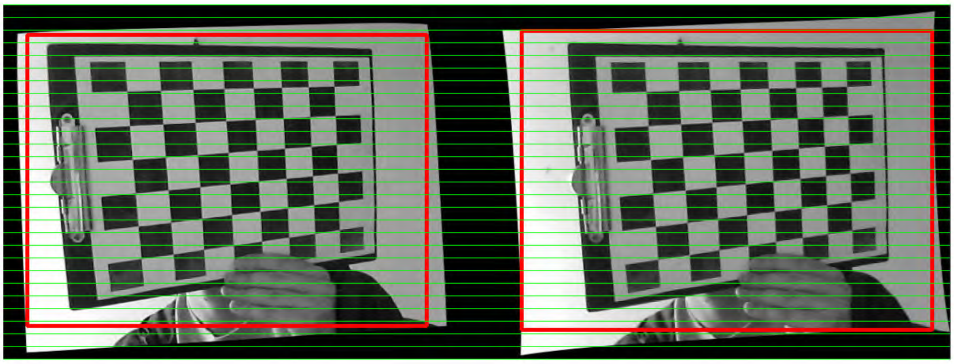
There has been much interest in building augmented reality systems. The majority of related work focuses on maker augmented reality systems. Kato et al [1] designed ARToolkit which is comprehensive system for recognition, tracking and rendering for marker augmented reality system. There are many systems extended ARToolkit like ArUco [2], ARTag [3], NyARToolkit [4]. Marker-less methods are usually based on detection image features that are good for tracking [5]. Klein [6] designed PTAM (Parallel Tracking and Mapping) as marker-less method which revealing image feature using FAST (Features from Accelerated Segment Test) method and Kalman filter for tracking. This method has good performance but virtual object are presented only in front of real scene. Another method proposed by Chari et al [7] use combination of image segmentation and SIFT (Scale-invariant feature transform) features to estimate the approximate depth of layers. In this method front object obscuring problem is solved, but this method is not working in real time. Processing off-line file takes up to 1 hour. Similar method that also uses segmentation was proposed by Bleyer and Gelautz [8][9] where mean-shift based clustering algorithm was used. Bleyer and Gelautz method allows handling of large untextured regions and precision localization of deep boundaries. This method gives very good disparity map quality and performance.Using calibrated stereo camera rig deep map can be extracted using block matching method [10] or graph cut method by [11]. Those two methods are implemented in OpenCV image processing library [12]. Augmented reality system can be build using only disparity estimation computed form stereo image rig using block matching or graph cut method for building disparity map of 3D scene. Disparity map is used for building planer layers of 3D scene which is segmented and all segments have assigned distance from camera. If those segments are presented in 3D it is easy to put virtual images between them. By using this approach virtual objects that are further from camera then real object are obscured by real objects. All steps of this method are presented in Table 1.

Tabela 1: Algorithm steps

Step	Description	Output
Calibration	Estimation of stereo rig intrinsic and extrinsic parameters using Zhang [13] method	Intrinsic, extrinsic matrices, translation and rotation, perspective projection matrix
Stereo Image capture	Acquisition stereo images	Left and right image
Pose estimation and rectification	Rectification of left and right images	Rectificated left and right image
Stereo correspondence error estimation	Rectification can be estimated using epipolar geometry	Reprojection error value
Building disparity map	Building disparity map with block matching or graph cut method	Disparity map
Virtual object and disparity map rendering	Rendering virtual object and disparity map as surface	Augmented reality image

1.3 Calibration

Stereo images must be calibrated and rectified in order to create a disparity map. Stereo cameras parameters can be estimated using calibration proce-



Rys. 2: Stereo image calibration with epipolar lines

dure which is implemented using OpenCV library [12]. That algorithm is constructed by Zhang [13]. A number of images of a chessboard pattern, placed at different positions in front of the stereo rig, is taken. Knowing structure of the pattern Zhang's method can estimate left and right camera intrinsic matrices, left and right camera distortion coefficients matrix, rotation matrix and transaction matrix relating left and right camera positions. Then stereo image can be rectified using cameras parameters. Rectification can be evaluated using epipolar geometry. Results of calibration are presented on Figure 2.

1.4 Block matching method

The block matching stereo deep map method, [10], is very fast one-pass stereo matching algorithm that uses sliding sums of absolute differences between pixels in the left image and the pixels in the right image, shifted by some varying amount of pixels. In order to improve quality and readability of the disparity map, the algorithm includes pre-filtering and post-filtering procedures.

1.5 Graph cut method

The graph cuts stereo correspondence algorithm, described in [11], is non-real time stereo correspondence algorithm that usually gives very accurate depth map with well-defined object boundaries. The algorithm represents stereo problem as a sequence of binary optimization problems, each of those is solved using maximum graph flow algorithm.

2 Experimental Results

2.1 Technology

All of the software created for this paper for proof of concept and testing was written in C/C++ language and compiled for MinGW.
OpenCV 2.3 development library was used for image processing and manipulations. Then OpenGL library is used for rendering 3D objects and composing it into real scene. All tests were run on machine with Intel i7 2640M 2.80GHz processor, 8GB ram. Stereo images resolution was 640x480 pixels.

2.2 Results

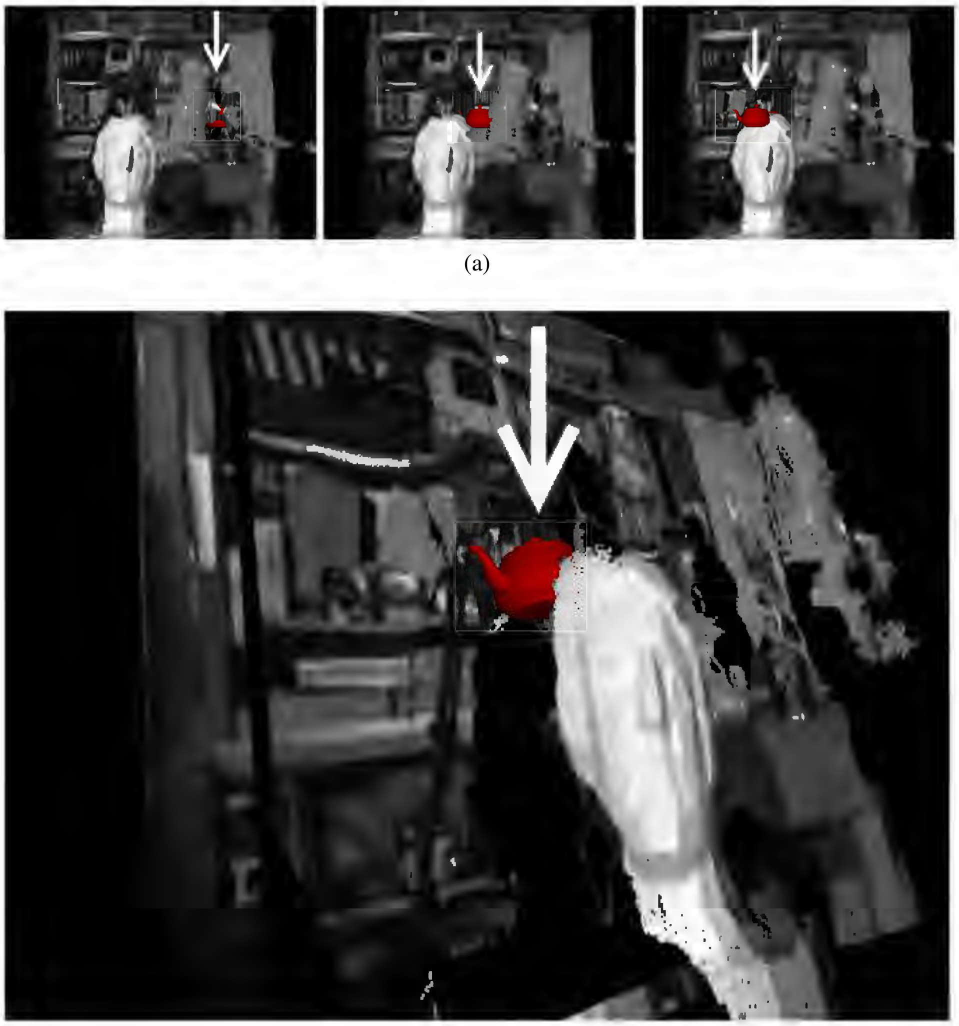
Disparity map methods were tested using images form Middlebury Stereo Datasets [14] and live web-cam images.

The first scenario use static image "Tsukuba" from Middlebury Stereo Datasets as shown on figures 3a and 4a. Three images show key frames form video sequence. On "Tsukuba" image there is a few occluding object. The aim of this test was to put virtual object (red tea pot) between planar layers.
Figure 3b and 4b presents planar layers from different perspective. Average time of disparity map generation was 12.123 milliseconds for block matching method and 98.232 milliseconds for graph cut method.

The second scenario use live web-cam images. Stereo web-cameras rig was calibrated before running disparity map algorithms. For this scenario the same configuration settings were used for all algorithms. The aim of this test was to put virtual object (red tea pot) between planar layers. Average time of disparity map generation was 13.291 milliseconds for block matching method and 121.639 milliseconds for graph cut method. Disparity maps for live web-cam images were quite poor comparing to static image.

2.3 Conclusion

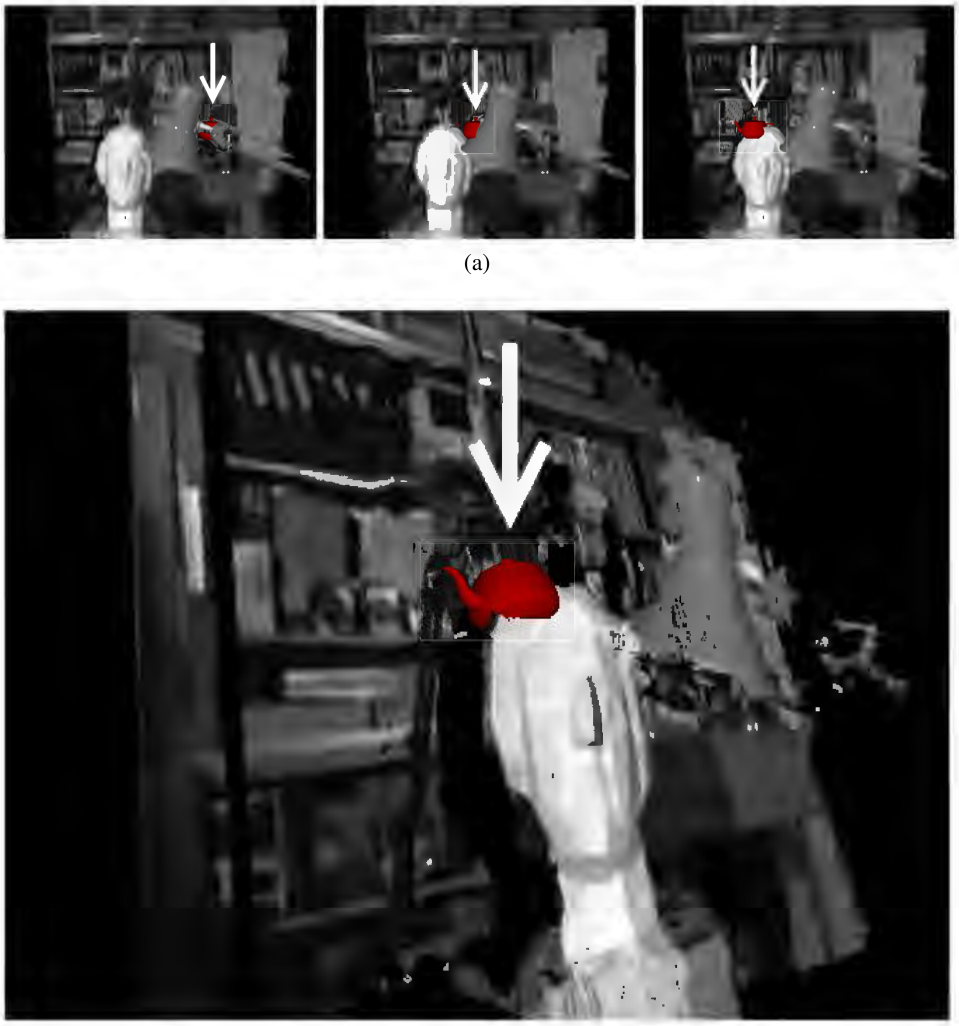
In this paper, an approach to augmented reality system that relies on disparity map of stereo images is presented. Obtained results show that graph cut and block matching method can be used for augmented reality systems. For better image quality need to be supported with more information about 3D scene composition. Both methods give better results when static



(a)

(b)

Rys. 3: Disparity map results using block matching method. Red teapot (pointed out white by arrow) is moving between layers.



(a)

(b)

Rys. 4: Disparity map results using graph cut method. Red teapot (pointed out white by arrow) is moving between layers.

image are used. This is because static image contains a lot of details while web-cam image usually show bigger solid plates. Both methods have problem to assign disparity to all pixels. The further works to use disparity maps approach need improve quality of disparity map with solid plates.

Literatura

1. Kato H., Billinghurst M., (1999) *Marker tracking and HMD calibration for a video-based augmented reality conferencing system*, Augmented Reality, (IWAR '99) Proceedings. 2nd IEEE and ACM International Workshop, 85 - 94.
2. ArUco http://www.uco.es/investiga/grupos/ava/node/26 Augmeterd reality system.
3. ARTag http://www.artag.net/ Augmented reality system.
4. NyArToolkit http://www.artag.net/ Augmented reality programming library.
5. Koniarski K., (2011) *Metody detekcji cech obrazów w analize wielokadrowej*, Techniki informacyjne teoria i zastosowania, 1(13), 68-82.
6. Klein G., Murray D., (2009) *Parallel Tracking and Mapping on a Camera Phone*, Mixed and Augmented Reality, 83-86.
7. Chari V., Singh JM., Narayanan PJ., (2008) *Augmented Reality using Over-Segmentation*, National Conference on Computer Vision Pattern Recognition Image Processing and Graphics.
8. Bleyer M., Gelautz,. (2004) *A Layered Stereo Algorithm Using Image Segmentation and Global Visibility Constraints*, Image Processing, **5**, 2997 - 3000.
9. Malik H,. Roth G,. McDonald C,. (2002) *Robust 2D Tracking for Real-time Augmented Reality*, Proceedings of Vision Interface, 399-406.
10. Konolige K., (2010) *Projected Texture Stereo*, Robotics and Automation (ICRA), 2010 IEEE International Conference, 148 - 155.
11. Kolmogorov V., (2003) *Graph Based Algorithms for Scene Reconstruction from Two or More Views*, PhD thesis, Cornell University.
12. OpenCV http://opencv.willowgarage.com Augmented reality programming library.
13. Zhang Z., (2000) *A Flexible New Technique for Camera Calibration*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **22(11)**,1330-1334.
14. Middlebury Stereo Datasets http://vision.middlebury.edu/stereo/data/.

METODY DETEKCJI GŁĘBI W OBRAZIE STEREOSKOPOWYM W ZASTOSOWANIU ROZSZERZONEJ RZECZYWISTOSCI

Abstract. W niniejszej pracy przedstawiony jest problem wyznaczania różnic pomiędzy obrazami z kamer stereo. Dwie popularne metody są porównane, które można wykorzystać w systemach rozszerzonej rzeczywistości. Wstawianie wirtualnych obiektów pomiędzy rzeczywiste obiekty jest trudnym problemem. Szczególnie, jeśli odbywa się to w czasie rzeczywistym, a obraz pochodzi z kamery wideo. Proponowane metody segmentują obraz rzeczywisty na warstwy, a każdy segment otrzymuje etykietę określającą jego głębię w stosunku do kamery. Wyniki eksperymentów pokazały, że obie metody działają efektywnie. Metoda graph cut daje lepsze wizualne wyniki, natomiast metoda block matching działa znacznie szybciej.

ISBN 9788389475442