**Preprints of the 18th IFAC World Congress**
**Milano (Italy) August 28 - September 2, 2011**

# Controlling Depth Estimation for Robust Robotic Perception

**Sorin M. Grigorescu and Florin Moldoveanu**

*Department of Automation, Transilvania University of Braşov,*
*Romania, (Tel: +40-268-418-836; e-mail: s.grigorescu@unitbv.ro,*
*moldof@unitbv.ro)*

**Abstract:** In this paper, a closed-loop control approach to the robust depth estimation problem in stereo vision is presented. The idea employed in this work is to introduce feedback control techniques at image processing level in order to improve the robustness of a robotic vision system with respect to external influences, such as cluttered scenes and variable illumination conditions. The control strategy detailed in this paper is based on the traditional open-loop mathematical model of the depth estimation process. Block matching has been considered as the technique to be used in the proposed depth calculation system. The suggested control law is derived from the known extremum seeking method which aims at finding optimal actuator values based on the minimum, or maximum, values of a feedback variable. The benefits of using feedback control techniques in machine vision, and particularly in stereo depth estimation, are demonstrated through performance evaluation results.

*Keywords:* Robot vision, Stereo vision, Image segmentation, Feedback control, Closed-loops.

## 1. INTRODUCTION

In many vision based applications, such as mobile robotics and visual guided object grasping, the reliability and robustness of 3D visual perception of the robot's surrounding plays a crucial role in the success or failure of the autonomous system [Kragic and Christensen (2005)].

The most common approach to 3D perception, or depth sensation, is through stereo vision. Mainly, stereo vision exploits the geometry between several perspective cameras imaging a scene. By analysing the perspective views between the acquired images, 3D visual information can be extracted. Traditionally, stereo vision is implemented using a pair of calibrated cameras with a known baseline between their optical points. Having in mind that the geometrical relations between the two cameras are known, by calculating the relative perspective projection of object points in both images, their 3D world coordinates can be reconstructed until a certain accuracy.

In this paper, the authors propose a feedback control approach of a depth estimation system, aiming at compensating the problem of using constant image processing parameters in complex environments. When feedback control techniques are discussed in connection to robot vision, they are usually put in the context of controlling a certain system using visual information. Such devices are typically named *Active Vision* or *Visual Servoing Systems* [Chaumette and Hutchinson (2006)]. There are relatively few publications dealing with control techniques applied directly on the image processing chain.

The idea of feedback image processing has been tackled previously in the computer vision community in papers such as [Mirmehdi et al. (1999)] or [Zhou et al. (2006)]. One of the first comprehensive papers on the usage of feedback information at image processing level can be found in [Peng and Bahnu (1998)], where reinforcement learning was used as a way to map input images to corresponding optimal segmentation parameters. Mirmehdi et al. (1999) developed a hypothesis generation and verification method in order to calculate interest operators which can be used to locate target objects, such as bridges, in noisy data. Also, Zhou et al. (2006) employed a feedback strategy in the self-adaptation of a learning-based object recognition system that has to perform in variable illumination conditions. Marchant and Onyango (2003) used dynamic closed-loop systems to automatically adapt camera parameters at the image acquisition stage. In the area of stereo vision, Gutierrez and Marroquin (2004) adopted probabilistic methods for the robust analysis of depth estimation.

Although the mentioned literature is focused on closed-loop processing, it does not provide a suitable control framework from both the image, as well as from the control point of view. Techniques for image processing inspired from control engineering were used by Ristic (2007) for adapting a character recognition system, as well as for a quality control one. In the field of robot vision, the authors successfully used feedback control concepts to tune region [Grigorescu et al. (2008)] and boundary [Grigorescu et al. (2010)] based segmentation operations in order to improve the visual perceptual capabilities of a service robot. In this paper, feedback machine vision is further investigated by proposing a closed-loop model of depth sensing based on the extremum seeking control paradigm set forth by Ariyur and Krstic (2003).

The paper is organized as follows. In Section 2, the proposed theory behind feedback modelling of image processing systems is presented, followed in Section 3 by a detailed description of the process to be controlled, that is, the depth estimation system. In Section 4, the machine vision extremum seeking control paradigm detailed in Section 2 is applied to the depth sensing process. A performance evaluation of the proposed approach is given in Section 5. Finally, conclusions are stated in Section 6.

## 2. FEEDBACK CONTROL IN IMAGE PROCESSING

In a robotic application, the purpose of the image processing system is to understand the surrounding environment of the robot through visual information. Usually, an object recognition and 3D reconstruction chain for robot vision consists of *low* and *high levels* of processing operations. Low level image processing deals with pixel wise operations aiming to improve the input images and also separate objects of interest from background. Both the inputs and outputs of the low level processing blocks are images. The second type of modules, which deal with high level visual information, are connected to low level operations through a feature extraction component which converts the input images to abstract data describing the imaged objects. The importance of the quality of results coming from low level stages is related to the requirements of high level image processing. Namely, in order to obtain a proper 3D virtual reconstruction of the imaged environment at a high level stage, the inputs coming from low level have to be reliable.

Traditionally, vision systems are open-loop sequential operations, which function with constant predefined parameters and have no interconnections between them. This approach has impact on the final 3D reconstruction result, since each operation in the chain is applied sequentially, with no information between the different levels of processing. In other words, low level image processing is performed regardless of the requirements of high level processing. In such a system, for example, if the segmentation module fails to provide a good output, all the subsequent steps will fail.

The basic diagram from which feedback mechanisms for machine vision are derived can be seen in Fig. 1. In such a control system, the control signal $u$, or *actuator variable*, is a parameter of an image processing operation, whereas the *controlled*, or *state, variable* $y$ is a measure of processing quality.

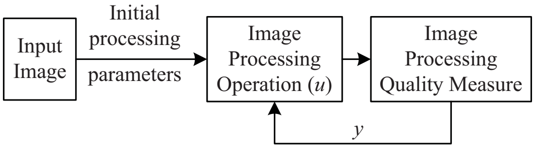

Input Image
Initial processing parameters
Image Processing Operation (u)
Image Processing Quality Measure
y

Fig. 1. Feedback control of an image processing operation.

The design and implementation of feedback structures in machine vision is significantly different from conventional industrial control applications, especially in the selection of the pair *actuator variable - controlled / state variable.* The choice of this pair has to be appropriate from the control, as well as from the image processing point of view.

In order to derive a control strategy for a machine vision system, the following discrete nonlinear state-space representation model of the vision apparatus is suggested:

$$\begin{cases} \dot{\boldsymbol{x}}(k) = f[\boldsymbol{x}(k), \boldsymbol{u}(k)], \\ \boldsymbol{y}(k) = g[\boldsymbol{x}(k)], \end{cases} \tag{1}$$

where $\boldsymbol{x} \in \Re^n$ is the state, $\boldsymbol{u} \in \Re$ is the actuator (input), $\boldsymbol{y} \in \Re$ is the output, $f : \Re^n \times \Re \rightarrow \Re^n$ is the state transition function and $g : \Re^n \rightarrow \Re$ is the output function. $k$ represents the discrete time. Suppose that we have a control law:

$$\boldsymbol{u}(k) = \alpha[\boldsymbol{x}(k), \theta], \tag{2}$$

the control problem is to find the optimal parameter $\theta^*$ which provides an output of desired, or reference, quality. Following the above reasoning, the closed-loop system:

$$\dot{\boldsymbol{x}} = f[\boldsymbol{x}, \alpha(\boldsymbol{x}, \theta)] \tag{3}$$

has its equilibrium point parameterized by $\theta$. Having in mind the high non-linearity of an image processing system, a control strategy based on extremum seeking [Ariyur and Krstic (2003)] is suggested. Thus, the goal of the feedback control system is to determine the optimal parameter $\theta^*$ as the minimum, or maximum, value of the state vector $\boldsymbol{x}$:

$$\theta^* = \text{arg min } \boldsymbol{x}(k) \text{ or } \theta^* = \text{arg max } \boldsymbol{x}(k). \tag{4}$$

The choice of this particular type of control method lies in the fact that, taking into account the non-linearity of an image processing system, it is difficult to determine reference values that could be applied to classical feedback structures. Hence, in the image processing control approach, the desired state of a vision system is given by the extremal values of the state vector. In the following, the proposed model is applied to the depth estimation processed detailed in the next section.

## 3. BLOCK MATCHING AND DEPTH SENSATION

The geometry of stereo vision is known as *epipolar geometry* [Hartley and Zisserman (2004)]. The principle behind it relies on the fact that between an imaged point in real 3D world coordinates and its projection onto 2D images exists a number of geometrical relations. These relations are valid if the cameras are approximated using the *pinhole camera model.* The model refers to an ideal camera with its aperture described as a point and no lenses are used to focus light. Knowing the relative position of two cameras with respect to each other, the imaged 3D point can be reconstructed in a 3D virtual environment using triangulation. In the following, in order to differentiate between 2D and 3D point coordinates, we will refer to the first ones as *pixels*, whereas to the last ones as *voxels*.

When estimating distances through stereo vision, there are mainly five steps that have to be performed:

1. *Calibrate* the stereo camera by calculating its intrinsic (e.g. focal length, optical centre, etc.) and extrinsic (e.g. baseline between the cameras, relative positions and orientations to each other) parameters;
2. Mathematically *remove radial and tangential distortions* introduced by errors in the geometry of the cameras lenses;
3. *Rectify* the angles and distances between the acquired images so that the rows in the left and right images are parallel;

(4) Calculate the *correspondence features* from the left and right images in order to obtain a depth, or disparity, map of the scene;
(5) Knowing the epipolar geometry of the cameras, *re-project* the calculated disparity to a virtual 3D environment.

Since steps 1, 2 and 3 from the above list are out of the scope of this paper, we will take them as granted, that is, we consider the input images to be undistorted, rectified and acquired from calibrated cameras. For more information on these steps the reader may refer to Hartley and Zisserman (2004).

### *3.1 Stereo Camera Setup*

The mechanics of depth estimation are illustrated in Fig. 2. A real world point represented in homogeneous coordinates $P = [V\ W\ Z\ 1]^T$ is projected in the image planes of a stereo camera as the homogeneous 2D image points:

$$\begin{cases} p_L = \left[v_L\ w_L\ 1\right]^T, \\ p_R = \left[v_R\ w_R\ 1\right]^T, \end{cases} \tag{5}$$

where $p_L$ and $p_R$ have the 2D coordinates $(v_L, w_L)$ and $(v_R, w_R)$ projected onto the left $I_L$ and right $I_R$ images, respectively. The $p_L$ and $p_R$ 2D image positions are given by the intersection with the image plane of the line connecting point $P$ in world coordinates with the optical centres $O_L$ and $O_R$ of both cameras, as shown in Fig. 2. The image, or *principal plane*, is located at a distance $f$ from the optical centre of a camera. $f$ is commonly known as the *focal length*. The $z$ axis of the coordinate system attached to the optical centre is referred to as the *principal ray*, or *optical axis*. The principal ray intersects the image plane at image centre $(c_v, c_w)$, also known as the *principal point*. The origin of the image coordinate system is defined as the image top-left corner $(v_0, w_0)$.

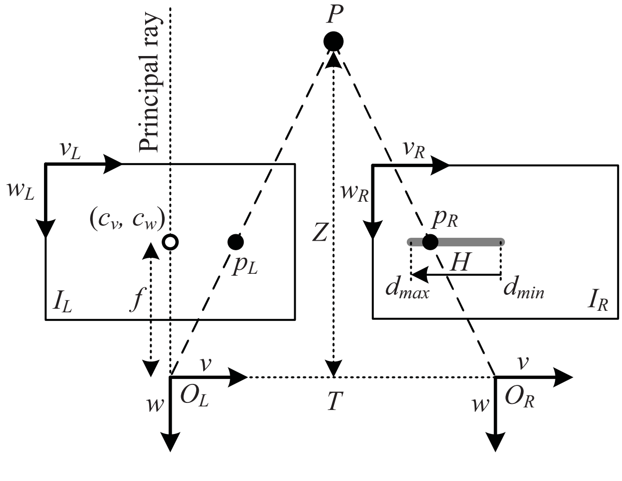


Fig. 2. Principle of depth estimation of a point $P$ on a pair of rectified and undistorted stereo images.

Knowing $p_L$, $p_R$ and the distance $T$ between the optical centres of the two cameras, the goal of depth estimation is to calculate the distance $Z$ between the baseline $T$ and the 3D position of $P$. Having in mind the perspective projection of $P$ onto the image planes, given by $p_L$ and $p_R$, the distance $Z$ can be obtained as:

$$Z = f \cdot \frac{T}{d}, \tag{6}$$

where $d$ is the disparity of the projected point $P$:

$$d = v_L - v_R. \tag{7}$$

From (6) it can be observed that the distance is inversely proportional to the disparity. Since we have considered rectified images as inputs, that is, images with parallel rows, the disparity $d$ is given only by the difference between the point coordinates on the $v$ image axis.

### *3.2 Block Matching Correspondence*

In order to properly compute $Z$, it is needed to establish the location of point $P$ in each camera image, namely, the 2D image points $p_L$ and $p_R$. The correspondence problem is currently one of the most investigated issues in stereo vision. In literature, there are a number of correspondence calculation methods, a comprehensive classification being made by Brown et al. (2003). In this paper, we have chosen to control the so-called *Block Matching* (BM) algorithm in order to obtain reliable 3D scene information.

BM is one of the most popular correspondence matching algorithm used in robotics, its main advantage being fast computation rates, thus making it a good candidate for real-time autonomous systems. BM matches points by calculating a *Sum of Absolute Differences* (SAD) over small sliding windows. The BM method is commonly performed in three steps. In our implementation we have considered the following operations:

- *Pre-filter* the input images with a 7x7 sliding window, containing a moving average filter, to reduce lighting differences and enhance texture;
- *Compute SAD* over a sliding window;
- Eliminate bad correspondence matches through *post-filtering*.

As shown in Fig. 2, the SAD values are calculated using a window shifted in the right images along the interval:

$$H = [d_{min}, d_{max}], \tag{8}$$

where $H$ is referred to as the *horopter*, defined as the 3D volume covered by the search range of BM. The goal of computing SAD is to find the best matching candidate of point $p_L$ in the right image, that is $p_R$, as:

$$m = \sum_{v,w} [I_L(v, w) - I_R(v + d, w)], \tag{9}$$

where $m$ is the SAD, or match, value and $d \in H$. By calculating SAD over $H$, we obtain a characteristic in which its maximum represents the best match candidate of $p_L$ in the right image.

Post-filtering aims at preventing false matches, hence false disparity maps. For filtering bad matches, a uniqueness ratio function is used, defined as:

$$q_r = \frac{(m - m_{min})}{m_{min}}, \tag{10}$$

where $m_{min}$ is the minimum SAD, or match, value. A feature is considered a match if:

$$q_r > T_q, \tag{11}$$

where $T_q$ is a predefined uniqueness threshold value. Bradski and Kaehler (2008) suggested the value of the uniqueness threshold to be $T_q = 12$. As it will be shown in Section 4, a predefined constant value of $T_q$ poses problems in 3D reconstruction since, depending on the imaged scene, it can introduce a large number of outliers in the reconstructed 3D model, or a too few number of voxels. To overcome this problem, we propose in Section 4 a feedback control method for the uniqueness threshold $T_q$. The output of BM is a grey level image $I_d$, also referred to as the disparity map, where the levels of grey represent different distances. In Fig. 3, the disparity map calculated for a typical cluttered service robotics scene is presented. The pixel values with a lower brightness from Fig. 3(b) are considered to be closer to the stereo camera system.

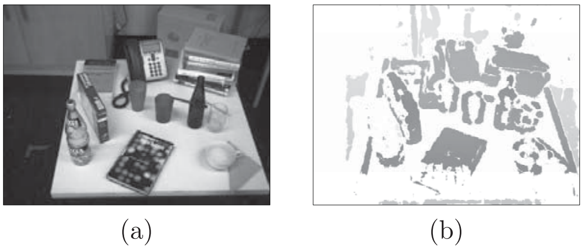


Fig. 3. Depth estimation via block matching. (a) Input left image. (b) Disparity map obtained with $q_r = 16$.

### 3.3 Back-Reprojection into 3D Virtual Environments

As can be seen from Fig. 3, the so far calculated disparity map provides information in pixel metrics, that is, levels of greys in an image, and not real 3D coordinates. The disparity points can be mapped back to a virtual 3D scene using the so-called *reprojection matrix*. The reprojection matrix is derived from the intrinsic stereo camera parameters and is defined as:

$$R_e = \begin{bmatrix} 1 & 0 & 0 & -c_v \\ 0 & 1 & 0 & -c_w \\ 0 & 0 & 0 & f \\ 0 & 0 & -1/T & c_T/T \end{bmatrix}, \tag{12}$$

where, in case the left and right cameras have different optical centers, $c_T = c_v^L - c_v^R$. $c_v^L$ and $c_v^R$ are the left and right centres of the cameras images along the $v$ axis. In this paper, we have considered $c_v^L = c_v^R$, hence $R_{e_{4,4}} = 0$.

Having in mind (12), a real world 3D point $P$ can be reconstructed in a virtual 3D environment through the following homogeneous transformation:

$$\begin{bmatrix} V \\ W \\ Z \\ B \end{bmatrix} = R_e \cdot \begin{bmatrix} v_L \\ w_L \\ d \\ 1 \end{bmatrix}, \tag{13}$$

where $B$ is a scaling factor defining the size of the reprojected voxels. In this work, we have considered a 1-to-1 mapping, that is, a scaling factor $B = 1$. Using (13), the reconstructed virtual environment of the scene shown in Fig. 3 is illustrated in Fig. 4.

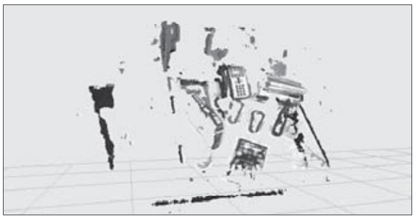


Fig. 4. Reconstructed 3D model of the scene from Fig. 3.

## 4. CLOSED-LOOP CONTROL OF DEPTH ESTIMATION

The main problem with the open-loop depth estimation system described in the previous section is its low performance with respect to variation in the scene structure, such as variable illumination conditions or clutter. An example of using constant parameters of depth estimation is illustrated throughout Fig. 3, 4 and 5. As said before, one of the main factors that influences the depth estimation process is the threshold value $T_q$ of the uniqueness ration $q_r$. If $q_r$ has an optimal predetermined value, as in Fig. 3, the reconstruction from Fig. 4 is fairly reliable, having in mind that we operate only with a pair of images. On the other hand, if the scene parameters change, or $q_r$ has a suboptimal value, 3D reconstruction might fail, as shown in Fig. 5. For a large value of $q_r$, as in Fig. 5(a,c), the 3D results have a low number of object voxels, whereas for a high $q_r$ the number of obtained outliers is too large, as in Fig. 5(b,d).

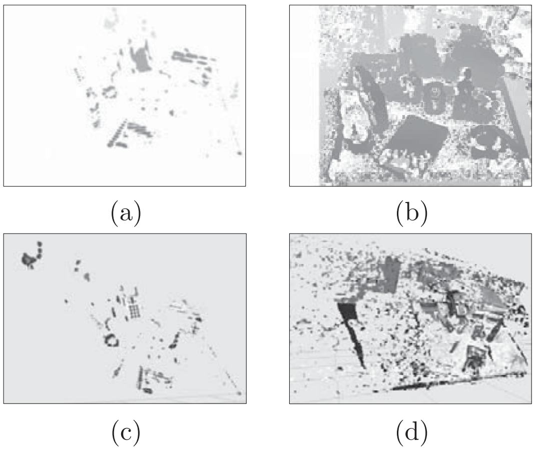


Fig. 5. 3D reconstruction results from suboptimal values of $q_r$. (a) $q_r = 68$. (b) $q_r = 4$. (c,d) Reprojected scenes.

Although, as inferred from Section 3, there are a number of parameters that could be controlled, we have considered the depth estimation process, for simplicity, as a *Single Input Single Output* (SISO) model.

The depth sensing process has been modelled as the nonlinear system from (1). For the sake of clarity, the state vector $\boldsymbol{x}$ is considered to have only one element which describes the behaviour of the modelled process. Since, depending on the chosen uniqueness threshold $T_q$, we obtain a different disparity map $I_d$, as shown in Fig. 5, a straightforward way to derive a state variable for the system is to quantify $I_d$. In this paper, we suggest as quantification of $I_d$ the distance segmentation $I_{th}$ of the obtained 3D model.

Depth segmentation can be implemented by specifying a range interval of interest $H$, also entitled horopter, where the desired objects reside. Using the inverse of (12), the horopter can be translated from real world metric units to pixel values that map depth in the disparity image $I_d$. In Fig. 6, three segmentation examples using a horopter $H = [0.7m, 1.5m]$ and different uniqueness thresholds are illustrated. As can be seen, only the segmentation result from Fig. 6(a) corresponds to optimal segmentation, the other two being either over- or under-segmented.

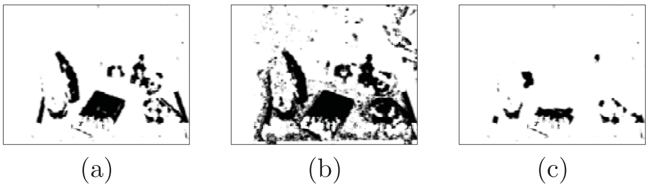


Fig. 6. Depth segmentation using $H = [0.7m, 1.5m]$ and different uniqueness thresholds. (a) Optimal $q_r = 16$. (b) Over-segmented $q_r = 68$. (c) Under-segmented $q_r = 4$.

Using the above described depth segmentation principle based on region segmentation, the problem of controlling the quality of the disparity map $I_d$ is converted into the problem of controlling the quality of the segmented image $I_{th}$. A region segmented image is said to be of good quality if it contains all pixels of the objects of interest forming a "full" (unbroken) and well shaped segmented object region. Bearing in mind the qualitative definition of a segmented image of good quality, the quantitative measure of segmented quality in (14) has been used:

$$i_m = -log_2 p_8,\ i_m(0) = 0, \tag{14}$$

where $p_8$ is the relative frequency, that is, the estimate of the probability of a segmented pixel to be surrounded with 8 segmented pixels in its 8-pixel neighbourhood:

$$p_8 = \frac{no.\ of\ seg.\ px.\ surrounded\ with\ 8\ seg.\ px.}{total\ no.\ of\ seg.\ px.\ in\ the\ image}. \tag{15}$$

Keeping in mind that a well segmented image contains a "full" (without holes) segmented object region, it is evident from (15) that a small probability $p_8$ corresponds to a large disorder in a binary segmented image. In this case, a large uncertainty $i_m$ is assigned to the segmented image. Therefore, the goal is to achieve a binary image having an uncertainty measure $i_m$ as small as possible in order to get a reliable depth segmentation result.

The depth estimation system was modelled according to (1), where the involved variables are:

$$\boldsymbol{x} = [i_m\ q_r]^T, \tag{16}$$

$$\boldsymbol{y} = I_d(v, w), \tag{17}$$

$$\boldsymbol{u} = q_r(i_m, T_q). \tag{18}$$

In Fig. 7, the input-output (I/O) relation between the state variable $i_m$ and the actuator parameter $T_q$ is displayed for the case of the scene from Fig. 3. The goal of the proposed extremum seeking control system is to determine the optimal value $T_q^*$ which corresponds to the minimum of the curve in Fig. 7. $T_q^*$ represents the desired value of the uniqueness threshold. The shape of the obtained I/O curves, as can also be seen from Fig. 7, preserve the controllability of the system, since the value of the actuator converges to the global minimum representing the equilibrium set-point of the considered system.

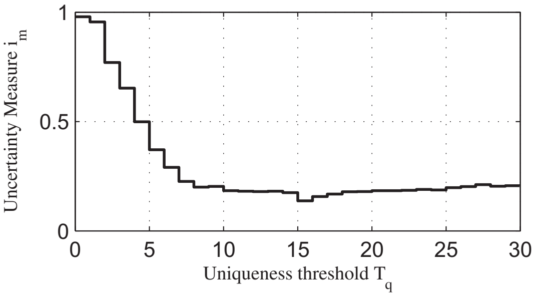


Fig. 7. The uncertainty measure $i_m$ of segmented pixels vs. uniqueness threshold $T_q$.

Following the above presented discussion, the block diagram of the proposed depth sensing system is illustrated in Fig. 8. Firstly, left and right images are processed in order to establish an initial depth map. The core of the method is represented by the state feedback loop which is used to automatically adapt the actuator parameter $T_q$ in order to obtain consistent depth estimation. Once the equilibrium set-point has been achieved, the calculated $I_d$ is used to reconstruct the viewed scene in a 3D environment by reprojecting the voxels using (13).

## 5. PERFORMANCE EVALUATION

The evaluation of the proposed system has been performed on a set of images acquired from service and autonomous robotics systems. A first set of 30 pairs of images of cluttered scenes, such as the one in Fig. 3, were acquired. The images contain common household objects which can be visually grasped and handled by a redundant manipulator. The scene in front of the robot manipulator was imaged using a pre-calibrated Bumblebee® stereo camera. The second series of test images consists of a set of 30 pairs of images acquired from a stereo high speed driver assistance system. In Fig. 9, an example of such a scene is illustrated.

Compared with open-loop depth estimation, which uses constant parameters of $T_q$, the obtain results show an increase in depth estimation accuracy and consistency. As a consequence of the regulation process, the number of outliers in the scene has been drastically reduces, in the same time keeping a sufficient amount of voxels to be used for scene understanding and labeling. The 3D volumes were evaluated based on the amount of obtained object voxels and outliers rate. As future evaluation work, the development of an evaluation procedure for the considered process in taken into account.

## 6. CONCLUSIONS

In this paper, a feedback control approach for machine vision systems based on the extremum seeking control paradigm has been proposed. The suggested approach has

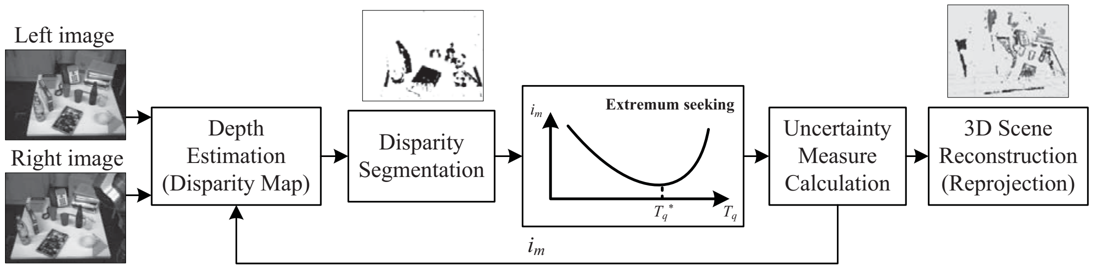
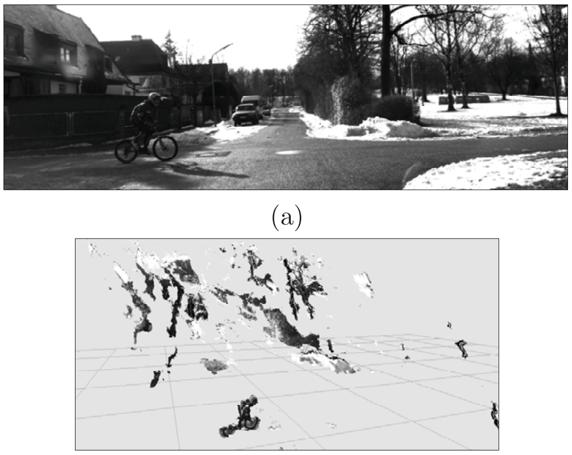

Left image
Right image
Depth Estimation (Disparity Map)
Disparity Segmentation
Extremum seeking
$i_m$
$T_q^*$
$T_q$
Uncertainty Measure Calculation
3D Scene Reconstruction (Reprojection)
$i_m$

Fig. 8. Block diagram of the proposed feedback control system for robust depth estimation.



(a)

(b)

Fig. 9. (a) Outdoor scene from stereo images acquired using a high-speed driver assistance system (Courtesy of Signum GmbH®). (b) Reconstructed scene using the proposed algorithm.

been successfully applied to a state-of-the-art depth sensing system used in autonomous robots. For performance evaluation, indoor cluttered service robotics scenes and outdoor images acquired from a driver assistance component were considered. As future work, the authors plan to enhance the theory behind feedback control in image processing, as well as to increase the number of controlled parameters of the depth sensing system, in order to improve the stability and robustness of the component. Also, the developed framework is planned to be evaluated no only on two-view stereo reconstruction, but on multi-view scene analysis with the goal of optimally reconstructing 3D volumes of indoor and outdoor scenes.

## ACKNOWLEDGEMENTS

This paper is supported by the Sectoral Operational Program Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the project number POSDRU/89/1.5/S/59323.

## REFERENCES

Ariyur, K. and Krstic, M. (2003). *Real-Time Optimization by Extremum Seeking Control*. John Wiley and Sons, New York, USA.

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV, Computer Vision with the OpenCV Library*. O'Reilly Media, USA.

Brown, M., Burschka, D., and Hager, G. (2003). Advances in computational stereo. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 25(8), 993–1008.

Chaumette, F. and Hutchinson, S. (2006). Visual servo control part i: Basic approaches. *IEEE Robotics and Automation Magazine*, 13(4), 82–90.

Grigorescu, S., Natarajan, S., Mronga, D., and Graeser, A. (2010). Robust feature extraction for 3d reconstruction of boundary segmented objects in a robotic library scenario. In *Proc. Of the 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. Taipei, Taiwan.

Grigorescu, S., Ristic-Durrant, D., Vuppala, S., and Graeser, A. (2008). Closed-loop control in image processing for improvement of object recognition. In *Proc. Of the 17th Int. Federation of Automatic Control IFAC World Congress*. Seoul, South Korea.

Gutierrez, S. and Marroquin, J. (2004). Robust approach for disparity estimation in stereo vision. *Image and Vision Computing*, 22, 183–195.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, UK.

Kragic, D. and Christensen, H. (2005). Advances in robot vision. *Robotics and Autonomous Systems*, 52, 1–3.

Marchant, J. and Onyango, C. (2003). Model-based control of image acquisition. *Image and Vision Computing*, 21, 161–170.

Mirmehdi, M., Palmer, P., Kittler, J., and Dabis, H. (1999). Feedback control strategies for object recognition. *IEEE Trans. on Image Processing*, 8(4), 1084–1101.

Peng, J. and Bahnu, B. (1998). Closed-loop object recognition using reinforcement learning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(2), 139–154.

Ristic, D. (2007). *Feedback Control in Image Processing*. Shaker Verlag, Aachen, Germany.

Zhou, Q., Ma, L., and Chelberg, D. (2006). Adaptive object detection and recognition based on a feedback strategy. *Image and Vision Computing*, 24, 80–93.