Stereo image visualization for a VISROBOT system

Zdzisław Kowalczuk
Decision Systems Department
ETI Faculty Gdańsk University of Technology
Gdańsk, Poland
kova@pg.gda.pl

Tomasz Merta
Decision Systems Department
ETI Faculty Gdańsk University of Technology
Gdańsk, Poland
tommerta@gmail.com

Abstract—The article describes a novel approach to robotic vision in mobile robot systems. A proposed Visrobot system implements the generic idea of using mobile robots for exploring static indoor environments. The task of such a robot is to visualize a stereo image properly for a human operator. The system uses different stereo baseline values what results in increasing depth resolution for distant objects. Assuming that the robot works in a static environment and that images can be taken from different robot's placements, the idea of the variable baseline stereo imaging can be put into practice. A disparity map can also be determined with the use of this approach. Suitable visualization of the disparity extent allows us to build a stereoscopic picture that ensures a proper perception of the depth and is convenient for displaying a 3D image to a system operator.

I. INTRODUCTION

The main task of autonomous units vision systems is to explore an unknown environment in a similar manner to the human sense of sight, which allows us to recognize static and mobile objects as well as to perceive relationships between them. Eyesight connected with other senses enables us to learn about the surrounding environment.

Vision systems used in mobile robots are highly limited, and usually are adapted to specific tasks. Contemporary possibilities of image processing algorithms still do not allow us to make analysis of complex situation. Such algorithms, compared to human capabilities still do not give satisfactory results, even if the accuracy of the camera and the eye are similar. Implementations of mobile robots provided with a vision system make possible execution of diverse functions (e.g. a mobile vacuum cleaner, underwater vehicle, humanoid robot, soccer player, scout robot, etc.).

In the analyzed context the main tasks of the system are stereo image acquisition as well as visualization for a robot operator. A more specific task is to transfer stereoscopic images of distant objects without approaching them (e.g. in the case of exploring a contaminated environment). Moreover, using a variable stereo baseline, a stereoscopic image of distant objects with accordingly high image depth can be suitably constructed [1].

II. SLAM

In addition to performing the principal task, a mobile robot has to move safely in its environment. Translocation and avoiding obstacles should be performed without any damages, which is achievable in the case of a known and limited environment. However, working in an unknown environment requires that the robot simultaneously determines its own position and updates a corresponding map (Simultaneous Localization and Mapping – SLAM). Knowing the positions of the robot and its obstacles is a basis for a safe operation.

A typical SLAM functioning is based on sensors that collect data from a vicinity of the robot. Current solutions often use a variety of optical sensors, such as lidar, sonar, laser rangefinder and video camera [2]. The sensors are selected depending on a robot’s application, its movement speed and working environment. Implementation of the SLAM function is essential for the automated process of observing environments without a complete supervision of an operator.

The SLAM function is considered in the two- or three-dimensional space (depending on the requirements). 2D maps are sufficient for navigation purposes when the robot moves on a flat ground, where the obstacles are high solids (such as walls). In more complex cases, it may be necessary to take into consideration a detailed 3D map.

It is worth mentioning that the video cameras have special properties. The cameras are passive sensors. They do not send any active signals to study the environment and provide data that express the description of the object located in both near and distant vicinity. Using two or more cameras a stereovision image is obtained (similar as human does). Also the image of a single camera located on an object in motion allows us to create a 3D map using an SfM technique (called Structure from Motion).

The SfM uses the difference in the position of the characteristic points in successive frames [4] and is based on one slowly moving camera, what results in the possibility of obtaining only short stereo bases. The movement itself calls for some involved solutions. On the other hand, our model introduces a strong conceptual simplification and considers arbitrary changes of the camera positions. The stereovision
allows us to determine the depth of an image and the detection of specific objects and it facilitates further analysis of the image.

III. STEREOVISION

The stereovision imitates the human vision system, therefore the visualization of the observed vicinity based on stereovision data is very useful for the operator. Two cameras (a stereo pair) analogous to a pair of eyes provide images that represent objects from two perspectives. The geometric relationship between a given point in space and its projection in the two cameras planes are described by epipolar geometry.

The left and right centers of projection $(O_l, O_r)$ and the image planes $(\pi_l, \pi_r)$ of the left and right camera, respectively, can be seen in Fig. 1. The plane $\pi'$ spanned by the point $p$ and both centers of projection $O_l, O_r$ is called an epipolar plane. The intersection of $\pi'$ with the image planes $\pi_l, \pi_r$ results in conjugated epipolar lines ($l_l$ and $r_r$) within camera matrices. Moreover, the projection $r_l$ of the projection center $O_l$ in the right camera plane $\pi_r$, and the projection $l_r$ of the projection center $O_r$ in the left camera plane $\pi_l$ compose epipolar centers. The segment that connects the centers of projections $O_l, O_r$ is a stereo baseline.

Although the location and position of the camera in space is unknown, a given point $p$ in space is described by vectors $(P_l, P_r)$ anchored in the centers of projection $O_l$ and $O_r$ in each camera. One can determine these vectors after a calibration process, that is, after removing the geometric distortions and determining extrinsic camera parameters. These parameters describe the transformation of space through the translation vector $T$ and rotation vector $R$ [5]. The relationship between vectors $P_l$ and $P_r$ can be then obtained using the formula:

$$P_r = R(R_l - T)$$  \hspace{1cm} (1)

Knowing the vectors $T$ and $R$ is necessary for further analysis, because, in general, searching the projections of a 3D point on the image planes $\pi_l, \pi_r$ is a difficult task. A given real section $P (p, p_1)$ parallel to the camera system (baseline $O_lO_r$) can be projected on planes $\pi_l, \pi_r$ and resulting in sections $L (l_l, l_r)$ and $R (r_l, r_r)$, respectively, as shown in Fig. 2.

In general, these sections are not parallel to each other. Furthermore, they do not coincide with the horizontal lines on the planes $\pi_l, \pi_r$. Therefore the image should be rectified – i.e. all points on planes $\pi_l, \pi_r$ should be converted. It causes that the projections of a point $p$ are located on the same horizontal line in the matrices $\pi_l, \pi_r$. After a proper rectification process the conjugated sections $L$ and $R$ obtain the form of sections $L'$ and $R'$, which are located on the same (e.g. $3^\text{rd}$ from the top) horizontal line on the planes $\pi_l, \pi_r$ as shown in Fig. 3.

Rectified images are needed to determine disparity which specifies in pixels $(d_l - d_r)$ a displacement/shift of a given 3D point projection between the planes $\pi_l, \pi_r$ (Fig. 4). This shift for all 3D points (projected on both camera planes) is represented by means of a disparity map. Taking into account also the intrinsic camera parameters, the points on the disparity map can be attributed to a spatial distance, which, in turn, determines a depth map. The depth in the image that is the distance of 3D points from the camera system, can be computed as follows [9]:

$$z = \frac{Bf}{d_l - d_r} = \frac{Bf}{d} \hspace{1cm} (2)$$

$z$ - image depth [m]
$B$ - stereo baseline width [m]
$f$ - focal length [pixels]
$d$ - disparity [pixels]

whereas the depth error $\Delta z$ [m] is additionally conditioned by the disparity error $\Delta d$ [pixels] according to the equations:

$$\Delta z = \frac{z^2}{Bf} \Delta d \hspace{1cm} (3)$$

$$f = \rho \cdot \Theta \hspace{1cm} (4)$$

$\Theta$ - single pixel size [m/focal in meters]
\( \rho \) - camera resolution [pixels/m].

According to (2) and (4) the value \( z \) is dependent on focal length setting \( f \) and requires knowledge of a single pixel size in meters \( \Theta \) (focal length in pixels). The greater disparity, the closer is a given element to the camera system. The distance measurement error increases with the square of the distance. Therefore, to ensure constant linear measurement when linearly increasing \( z \), a linear increase of both parameters \( B \) and \( f \) is required [6].

\[
\begin{align*}
\text{Figure 4. A segment projection on a rectified camera planes}
\end{align*}
\]

The value of focal \( f \) can be influenced directly by a real focal change, which results in zoom-in or zoom-out. Optics with a large zoom range requires an expensive and accurate lens for a large range of focal length \( f \). Moreover, a large zoom restricts the camera field of view to the central viewing area (a central region of the camera matrix).

The change of the range \( f \) can be carried out also by changing the image resolution. This causes a change in the single pixel size \( \Theta \) (\( f \) calculated in pixels) and in \( f \) (in meters). As the camera resolution \( \rho \) is fixed, the image resolution can only be reduced for small distances \( z \). Reducing the image resolution and increasing the error, while determining the depth is essential, if the priority is to ensure the ongoing demand for computing power for different values of \( z \) [6]. An attempt to modify the accuracy of disparity \( d \) by using subpixels while computing \( d \), is ineffective due to aliasing [7].

Increasing the width of the baseline \( B \) can achieved by moving cameras away from each other. For a mobile robot with sliding cameras the available range of changes of \( B \) may be insufficient. In a static environment, the effect of baseline change can be achieved through robot’s displacement. In this case a pair of stereo images is formed with the image from the previous and current positions of the robot. Such a variable base can be obtained using a group of robots (the stereo pair is distributed between two selected robots).

IV. ROBOTIC IMAGE VISUALIZATION

If the image from the robotic camera system satisfies some parameters of the human vision system, it can be presented to the operator (by suitably displaying the left and right images, respectively). Special eyeglasses with anaglyph, polarization or shutter technique can then be used. An alternative is to use autostereoscopic displays that do not require glasses. However, due to the nature of the image generated by the human eye and limitations of 3D displays, the image from the stereo cameras must be adjusted to ensure proper fusion of the two images (a stereo pair) in the brain of the observer [8].

A. Displaying a stereoscopic image

Images from specifically - configured cameras differ from the images generated by the eyes of the observer, therefore different methods of determining the depth map are suitable in both cases. Eyes are staring at a one chosen point, so that the optical axes of the eyes are crossed at this point (Fig. 5a). In a stereovision system such a ‘toad-in’ (staring) configuration is impractical. Typically, a canonical camera system (Fig. 5b) is applied, where the optical axes of the cameras are parallel, what greatly simplifies the analysis of the images.

In addition, if the priority for the stereovision system is to make an exact 3D view, as if the operator was at the robot’s location, it is necessary to remove stereoscopic distortions. This operation is to ensure that the cameras’ stereo baseline and the eye baseline are equal, the field of view of the observer was the same as the camera’s view, and that there was no depth resolution change (no magnification of depth) [9]. Usually, however, the priority is not to create the impression that the man is at robot’s location, but to observe the environment and to zoom-in essential details.

\[
\begin{align*}
\text{Figure 5. Camera system: (a) toad-in; (b) parallel}
\end{align*}
\]

A proper stereoscopic image synthesis requires a proper description of the relationships between the image depth, the display and displaying conditions. Disparity \( d \) describes the differences between the images (the camera matrices). Parallax \( P \) describes the differences between the images on the 3D display. Parallax occurs while watching a stereo image on a 3D screen and it is realized using a toad-in system (Fig. 6). Therefore, in special conditions these two concepts can be identified. It appears that if the images from the matrices \((L,R)\) in Fig. 2 are rectified and scaled appropriately to the size of the display, the parallax \( P \) can be described as the disparity \( d \):

\[
\begin{align*}
P &= \frac{W_d}{W_S} d \quad (5)
\end{align*}
\]

\( W_d \) - screen width
\( W_S \) - camera matrix width
\( d \) - disparity.

According to Fig. 6 the parallax is positive, when the observer has the impression that a 3D point is located behind the screen. The parallax occurs negative when the eyesight is
focused on a point, which is located in front of the screen. The zero parallax gives natural images.

Knowing the observer’s distance from the screen (Fig. 6), the observer stereo baseline width and the parallax, in general, a depth reconstruction can be done (using Thales theorem):

\[ z = \frac{z_p B_b}{B_b - P} \]

where:
- \( z \) - image depth [m]
- \( z_p \) - distance between screen and observer [m]
- \( B_b \) - eyes stereo baseline (≈ 0.064 m)
- \( P \) - parallax [m].

Therefore, using (5) and (6), the depth perceived by the human observer can be expressed as a function of disparity:

\[ z = \frac{z_p B_b}{B_b - \frac{W_b d}{W_p}} \]

**B. Comfortable viewing range**

In the toad-in eye system the shape of the lens of the eye changes depending on the object on which the eyesight is directed and focused. For any real 3D point and time moment the sight organ adapts its focal length and angle of view. Eye accommodation adjusts the focal to make the image of objects sharp. Eye convergence sets the angle of vision of each eye to the object. The range of convergence angle \( \alpha \) is adjusted to the distance \( z_v \), if the observed object is behind the screen or in front of the screen (Fig. 6). Due to a conflict between accommodation and convergence, the difference between \( z_p \) and \( z_v \) cannot be large.

The range of convergence angle \( \alpha \) variation was determined empirically. This range does not disturb the proper fusion of stereoscopic images in human brain [10]. The difference between the maximum and the minimum value is \( \Delta \alpha = 0.02 \) [rad]. Knowing \( \Delta \alpha \) we can calculate the maximum range of parallax and, according to (5), the corresponding disparity range. Using a single value of \( \Delta \alpha \) for the entire image is not appropriate. What is more, the tolerance for higher \( \Delta \alpha \) is due to the individual observer abilities.

Determining the range of convergence \( \Delta \alpha \) and the distance \( z_p \) from the screen, an acceptable parallax range can be specified as:

\[ \Delta P = P_{\text{max}} - P_{\text{min}} = z_p \Delta \alpha \]

Referring to the range of parallax to the screen size, a relative value can be obtained:

\[ \Delta P_{\text{rel}} = \frac{\Delta P}{W_p} \frac{z_p}{W_s} \Delta \alpha \]

The relative parallax range depends thus on the ratio of the observer distance (to the screen) to the screen width. Sources [8] provide different values of the optimum observer distance from the screen, most of which are in the range \(<1.2, 2.3>\). Assuming that \( z_p, W_p \approx 1.67 \) [8], the relative value of the parallax range \( \Delta P_{\text{rel}} \approx 1/30 \). This is the “1/30 rule”, according to which the parallax value (positive or negative) should be less than 1/30 part of the width of the screen. Assuming \( P_{\text{max}} = W_s / 30 \), according to (5), we obtain the acceptable disparity range \( d_{\text{max}} = W_s / 30 \) (in pixels).

Without taking into account other effects that also contribute to the deterioration of stereoscopy [9], we can conclude that the changes of camera parameters do not deteriorate significantly the stereoscopic images, provided that the disparity is in the allowed range.

**V. VISROBOT SYSTEM WITH VARIABLE BASELINE**

A robot with a stereovision system using a known width of the baseline can judge the distance from the object with greater accuracy. Moreover, with the observance of the restrictions on the permissible scope of disparity (Section IV), using a 3D display or suitable eyeglasses, the operator can be supplied with a correct stereoscopic image. The baseline width is limited by the maximum disparity value \( d_{\text{max}} \) and by the configuration and settings of the camera optics. Maintaining the described relations, larger baseline allows for a higher image depth resolution. For objects located far from the camera system there is the impression of ‘spatial enhancement’ (spatial zoom). This allows for a better distinction between the observed distance objects.

A similar approach was used in [1], where the variable baseline approach was applied solely to calculate a precise depth map, without displaying the resulting image to a robot operator or considering any particular mobile robot system.

**A. Stereo baseline tuning**

Based on (2), and assuming a disparity limit, we can make the \( B \) dependent on the (unknown) distance from the camera:

\[ B = \frac{d_{\text{max}}}{f} z \]

---

**Image 6. Perception of depth in a displayed stereoscopic image**

The range of convergence angle \( \alpha \) variation was determined empirically. This range does not disturb the proper fusion of stereoscopic images in human brain [10]. The difference between the maximum and the minimum value is \( \Delta \alpha = 0.02 \) [rad]. Knowing \( \Delta \alpha \) we can calculate the maximum range of parallax and, according to (5), the corresponding disparity range. Using a single value of \( \Delta \alpha \) for the entire image is not appropriate. What is more, the tolerance for higher \( \Delta \alpha \) is due to the individual observer abilities.

Determining the range of convergence \( \Delta \alpha \) and the distance \( z_p \) from the screen, an acceptable parallax range can be specified as:

\[ \Delta P = P_{\text{max}} - P_{\text{min}} = z_p \Delta \alpha \]

Referring to the range of parallax to the screen size, a relative value can be obtained:

\[ \Delta P_{\text{rel}} = \frac{\Delta P}{W_p} \frac{z_p}{W_s} \Delta \alpha \]

The relative parallax range depends thus on the ratio of the observer distance (to the screen) to the screen width. Sources [8] provide different values of the optimum observer distance from the screen, most of which are in the range \(<1.2, 2.3>\). Assuming that \( z_p, W_p \approx 1.67 \) [8], the relative value of the parallax range \( \Delta P_{\text{rel}} \approx 1/30 \). This is the “1/30 rule”, according to which the parallax value (positive or negative) should be less than 1/30 part of the width of the screen. Assuming \( P_{\text{max}} = W_s / 30 \), according to (5), we obtain the acceptable disparity range \( d_{\text{max}} = W_s / 30 \) (in pixels).

Without taking into account other effects that also contribute to the deterioration of stereoscopy [9], we can conclude that the changes of camera parameters do not deteriorate significantly the stereoscopic images, provided that the disparity is in the allowed range.

**V. VISROBOT SYSTEM WITH VARIABLE BASELINE**

A robot with a stereovision system using a known width of the baseline can judge the distance from the object with greater accuracy. Moreover, with the observance of the restrictions on the permissible scope of disparity (Section IV), using a 3D display or suitable eyeglasses, the operator can be supplied with a correct stereoscopic image. The baseline width is limited by the maximum disparity value \( d_{\text{max}} \) and by the configuration and settings of the camera optics. Maintaining the described relations, larger baseline allows for a higher image depth resolution. For objects located far from the camera system there is the impression of ‘spatial enhancement’ (spatial zoom). This allows for a better distinction between the observed distance objects.

A similar approach was used in [1], where the variable baseline approach was applied solely to calculate a precise depth map, without displaying the resulting image to a robot operator or considering any particular mobile robot system.

**A. Stereo baseline tuning**

Based on (2), and assuming a disparity limit, we can make the \( B \) dependent on the (unknown) distance from the camera:

\[ B = \frac{d_{\text{max}}}{f} z \]
With a constant focal length \( f \) the baseline width \( B \) can be increased while the distance \( z \) increases. To determine an acceptable \( B \) we need to know an approximated distance from the observed objects. For this purpose, initially we obtain \( z \) according to (2) with a small baseline \( B_0 \), and then expand the baseline to improve the resolution depth for distant objects. It should be kept in mind that the expanded baseline cannot contribute to exceed the predetermined disparity value \( d_{\text{max}} \). Objects with constant disparity \( d \) (distance \( z \) from the camera system) can be recognized as a group of pixels (in an image).

Problems may arise from areas of unknown \( z \) value in the image that are caused by errors of the stereovision algorithm. Moreover, obstacles may prevent to move the robot to a required place. It can also occur that close objects occlude objects that are far away. Thus the areas in the image, for which disparity could not be determined, should be rejected. Furthermore, occlusions should be detected using an increased baseline width. For supervised systems, a human operator can resolve such problems.

**B. Environment stereoscopic visualization algorithm ESV in the VISROBOT system**

We assume that the robot’s environment is stationary. A stereo pair is visualized to the operator as a 3D image with the possibility of ‘spatial enhancing’ (by increasing the depth resolution). Different \( B \)'s are realized through relocating cameras and binding images from the previous and the current robot positions (as described in Section I). Thus an additional sensor is necessary to determine the current position of the robot. Such a sensor needs to provide not only current camera position data, but also the elevation and azimuth angles of the camera. An integrated sensor can also be applied that includes 3-axis accelerometer, gravimeter and magnetometer. Another problem that needs to be resolved concerns the issue of the precision of the sensor.

A laboratory mobile robot was constructed and used for testing the performance of the Visrobot system. It includes wheels, PandaboardES platform and a Polulu sensor MiniMU-9 (a 9 dof sensor) composed of a 3-axis accelerometer, magnetometer and a gyroscope. The vision system in this implementation is based on 2 IDS video USB cameras (model 16-40LE) with Mega Pixel lens (2.6mm focal).

At the beginning the ESV algorithm generates the depth map for a certain small baseline \( B_0 \) (10 cm was applied). The stereo image is analyzed to determine its contours, which contain pixels of the same disparity value (the same distance from the camera). For such a pre-processed image the robot’s operator selects an area of interest (a contour with \( d = d_0 \)), which should be observed with higher depth resolution. After selecting the contour the stereo baseline should be increased.

This raises the need to establish adequate conditions for observation. In addition to the cases with reasonable disparity (\( 1 < d < d_{\text{max}} \)) of the selected contours, a borderline case is possible (black points on the disparity map), where \( z = \infty \) and \( d = 0 \) (since \( zd = Bf = \text{const} \)). If the disparity is less than 1 pixel (\( d_0 \approx 0 \)), nor the distance \( z \) can be obtained, neither a new baseline \( B \) that does not exceed the value \( d_{\text{max}} \) can be calculated. Therefore, assuming that a "black point" has a certain distance \( z_{\text{th}} \) which was hypothetically obtained with the previous base \( B \), we obtain an approximate relationship:

\[
B_0 \approx \frac{1}{f} z_{\text{th}}, \quad z_0 = fB_0
\]

Using above and (10), under the assumption that the actual \( d \) cannot exceed \( d_{\text{max}} \) we obtain

\[
B_{\text{max}} = \frac{d_{\text{max}}}{f} z_0 = d_{\text{max}} \cdot B_0
\]

Therefore the maximum baseline \( B_{\text{max}} \) is calculated as:

\[
B_{\text{max}} = \begin{cases} \frac{d_{\text{max}}}{f} z \Leftrightarrow d_0 > 1 \\ B \cdot d_{\text{max}} \Leftrightarrow d_0 \leq 1 \end{cases}
\]

Due to errors of the algorithm that determines the disparity map the baseline should not be increased in a single step from \( B = B_0 \) to \( B = B_{\text{max}} \), but increased in the successive iterations of the algorithm (e.g. \( \Delta B = B_0 \)).

For the actually enlarged stereo baseline the robot moves by the designated value into a previously set (left or right) direction. Next, the extrinsic camera parameters are set that are based on the current data from the sensor that expresses the position and orientation of the robot. Using the current image and the image obtained from the previous step, the ESV algorithm calculates the current \( z \) value of the contour.

The algorithm is terminated if the contour selected by the operator has a maximum disparity \( d = d_{\text{max}} \). Moreover, before the stereoscopic image is displayed, the algorithm verifies if the image does not contain contours with \( d > d_{\text{max}} \). These areas with too large disparity are not displayed to the operator.

**Algorithm ESV Procedure**

Set the initial baseline \( B = B_0 \)

generate a disparity map for an actual \( B \)

obtain contours for the same disparity value

calculate \( z \) for all contours

wait to select contour (by the operator)

determine \( B_{\text{max}} \)

enlarge baseline by \( \Delta B \)

obtain disparity map (and \( d \) for a selected area)

verify if \( d = d_{\text{max}} \)

if yes, hide contours that \( d > d_{\text{max}} \) and terminate

move robot transversally by a given step

read actual position and orientation from a sensor

update extrinsic camera parameters

The ESV algorithm shown in Procedure 1, with the restrictions of section IV, can be evaluated by the Reader observing with anaglyph glasses the 3D full-color images displayed for this purpose on our site [11]. The elements in the image that are close are perceived properly in the case of a small baseline (Fig. 7c) and incorrectly in the case of a very
large baseline (Fig. 8c). It can also be noticed that the distances between the background and the foreground are greater with a larger stereo baseline.

![Image](a) (b) (c) (d)

Figure 7. Visualization of an object with baseline $B=7\text{cm}$: (a) left image, (b) right image, (c) disparity – anaglyph image (d) ‘close’ depth map

![Image](a) (b) (c) (d)

Figure 8. Visualization of an object with baseline $B=20\text{cm}$: (a) left image, (b) right image, (c) disparity – anaglyph image (d) ‘far’ depth map

For a better visualization of the results, the disparity maps for both cases are presented. Without the use of glasses it can be observed how the depth resolution changes with different baselines. Black color indicates that the distance of the point of the camera is unknown. The lighter the color (higher luminance) the closer is a given point to the camera system. Fig. 7 except a real images of an object – left (a) and right (b), anaglyph disparity (c), shows the ‘close’ depth map (d) with the stereo baseline $B=7\text{cm}$. Similarly, in Fig. 8 the same object is shown with $B=20\text{cm}$. Comparing Fig. 7d and 8d it can be seen that with low stereo baseline the resolution depth is small (to improve readability the brightness and contrast of the image has been modified, so the depth in Fig. 7d has a different scale than the depth in Fig. 8d).

Comparing the luminance in Fig. 7d and 8d it can be observed that the near objects are visible with a small baseline, whereas large baseline reproduces better the distant objects (which indicates a higher depth resolution). It should be noted that elements in Fig. 8d with a very high disparity were removed (images 8c and 8d for close objects are not correct). This issue is included in the ESV algorithm procedure, which allows a human operator to interpose to ensure a proper stereoscopic image perception.

VI. CONCLUSION

This paper presents a stereovision system that uses a variable stereo baseline, suitable for a mobile robot. Due to stereovision technology the robot is capable to passively measure the distance of objects in its near and distant vicinity. By a proper change of stereo baseline the position of objects in the background can be determined systematically: to evaluate large distances the depth resolution is increased (‘spatial enhancement’). A stereoscopic image useful for the operator can be obtained at run time. In a further research we plan to automate the spatial mapping process using variable stereo baseline with the possibility of a convenient visualization. In a next stage of the study we may consider a system composed of robots associated in a stereovision system, whose task is to study elements of a static environment.

REFERENCES