METHOD OF 3D-OBJECT DETECTION BASED ON ORTHOPHOTO DIFFERENCE ANALYSIS

Sergey ZHELTOV,   Alexander SIBIRYAKOV
State Research Institute of Aviation Systems (GosNIIAS), Moscow, Russia
zhl@gosniias.msk.ru,   sibiryakov@mtu-net.ru

Working Group IC-13

ABSTRACT

In the paper we propose the method of object detection based on analysis of orthophoto images difference. Object is registered stereoscopically on smooth surface of arbitrary shape. Orthophoto images for some basic plane are generated from left and right images of the stereopair. Orthophoto generation is based on a priori surface model knowledge. In case of object presence in the scene the orthophoto difference image contains a characteristic structure, which is detected by the matched filtration.

1.   INTRODUCTION

Problem of detection of small 3D-objects included in scene of observation often occurs in various machine vision applications. Typical examples of this problem are obstacle detection for mobile robots or vision-guided navigation [M.Bertozzi and A.Broggi, 1997].

In the presented work the problem is considered when probable detecting object is located on relatively smooth curved surface like car road or way of mobile robot tracking. Analytical surface model might be known a priori with some accuracy. Another way of background surface model acquisition is least square method of fitting to the data set of points, which are exactly known to be not belonging to the object under detection.

In the case under considering object detection should be done stereoscopically, since in the opposite case it is impossible to recognize situations when the object stands on the surface or is drawn on the surface. For stereoscopic system its relative and interior orientation should be known.

The basic idea of the detection method consists of the following. If an analytical model of a surface is known, orthogonal projections of the given surface to some convenient plane with use of left and right images of the stereopair are created. The calculation of a difference of such projections results in appearance of characteristic geometric structures in neighborhoods of 3D-objects not belonging to the given surface.

As orthogonal projections eliminate the distortions caused by surface irregularity the significant brightness variance is appeared in difference image only in 3D-object neighborhood. In case of simple objects with straight-line edges the precise 2D-corner shaped structures are appeared. Angle value and shape of corner are known functions of object position do not depending on the object form. In case of objects with the more complex form (e.g. with plenty of edges) a combination of angular structures dependent only on the object position is appeared on the difference image. Therefore detection algorithm based on angular structures optimal matched filtration is the best in the given problem.

Thus, the work shows that the problem of 3D–object detection can be reduced to a problem of detection of 2D-structures with the predicted properties on a synthetic image.

In the work presented the utmost accuracy of the suggested detection algorithm is analyzed depending on both object elevation under surface and accuracy of relative and interior orientation. It is shown that using of standard video cameras with the basis of the stereo system about 1m provides the possibility of small object detection down to 5cm in height on distance up to 80m. The results of using the method in the problem of road obstacle detection are shown.

2. PHOTOGRAMMETRIC FOUNDATION OF THE METHOD

This section is devoted to exploring the stereophotogrammetric possibilities of the method. To estimate accuracy of object coordinates determination let us consider the case, when the camera orientation parameters are known. Let us introduce an external coordinate system X, Y, Z. The system is selected so that the surface has 2.5D-view $Z(X,Y)$ in some neighborhood of the object. By (X_s, Y_s, Z_s) denote the projection center of the left camera. Designate by a_i, b_i, c_i elements of left camera rotation matrix, which is calculated from three rotation angles α, ω, κ. By (x, y) denote a projection of a 3D-point to the left image. All variables corresponding to the right camera denote by ('). Thus the relation between 3D-scene and image points is given by the following equations:

$$F_x = x - f\frac{a_1(X - X_s) + b_1(Y - Y_s) + c_1(Z - Z_s)}{a_2(X - X_s) + b_2(Y - Y_s) + c_2(Z - Z_s)} = 0 \qquad F_y = y - f\frac{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)}{a_2(X - X_s) + b_2(Y - Y_s) + c_2(Z - Z_s)} = 0$$

$$F_{x'} = x' - f'\frac{a_1'(X - X_{s'}) + b_1'(Y - Y_{s'}) + c_1'(Z - Z_{s'})}{a_2'(X - X_{s'}) + b_2'(Y - Y_{s'}) + c_2'(Z - Z_{s'})} = 0 \quad F_{y'} = y' - f'\frac{a_3'(X - X_{s'}) + b_3'(Y - Y_{s'}) + c_3'(Z - Z_{s'})}{a_2'(X - X_{s'}) + b_2'(Y - Y_{s'}) + c_2'(Z - Z_{s'})} = 0 \quad (1)$$

By B denote a basis of the stereosystem. The following additional condition also used:

$$F_B = B^2 - (X_{s'} - X_s)^2 - (Y_{s'} - Y_s)^2 - (Z_{s'} - Z_s)^2 = 0 \tag{2}$$

3D-coordinates of the object are obtained by least square method:

$$\Delta x = -\left[\mathbf{A^T (BKB^T)^{-1} A}\right]^{-1} \mathbf{A^T (BKB^T)^{-1} F_0} \tag{3}$$

where Δx - vector of corrections to some initial approximation of 3D-coordinates $\mathbf{x} = (X, Y, Z)^T$, $\mathbf{A}$ – matrix of partial derivatives of functions (1),(2) on object coordinates X, Y, Z; $\mathbf{B}$ - partial derivatives matrix of functions (1),(2) on elements of observation vector $\mathbf{\bar{z}} = (x, y, x', y', X_S, Y_S, Z_S, \alpha, \omega, \kappa, X_S', Y_S', Z_S', \alpha', \omega', \kappa', B)^T$, $\mathbf{K}$ is the covariance matrix of observation vector $\mathbf{\bar{z}}$, $\mathbf{F_0} = (F_x, F_y, F_{x'}, F_{y'}, F_B)^T$ is discrepancy vector of functions (1),(2).

Accuracies of X, Y, Z are determined with the help of covariance matrix of spatial coordinates K_x. This matrix is defined by following equation:

$$\mathbf{K_x} = [\mathbf{A^T (BKB^T)^{-1} A}]^{-1} \tag{4}$$

It diagonal elements are the covariances of defined parameters.

To simulate the process of spatial coordinates determination the following values of elements of exterior orientation of images in the S_0XYZ coordinate system (S_0 - projection of the left origin projection to the horizontal plane) were used: $X_S = 0$, $Y_S = 0$, $Z_S = 1m$, $\alpha = \omega = \kappa = 0$, $X_{S'} = 1.5m$, $Y_{S'} = 0$, $Z_{S'} = 1m$, $\alpha' = \omega' = \kappa' = 0$. The following values of coordinates of three points of an object lying on the surface were used: $X_1 = -4.25$ m, $X_2 = 0.75$ m, $X_3 = 5.75$ m, $Y_1 = 50$ m, $Y_2 = 50$ m, $Y_3 = 50$ m, $Z_1 = 0$, $Z_2 = 0$, $Z_3 = 0$. The following values of interior orientation elements were used: $f = f' = 25$ mm.

Covariance matrix $\mathbf{K}$ which is included in (3) and (4), is accepted as diagonal with zero covariance between measurements. Mean square errors are accepted as:

$$\sigma_x = \sigma_z = \sigma_{x'} = \sigma_{z'} = \sigma_0 = \text{const}, \quad \sigma_{xs} = \sigma_{ys} = \sigma_{zs} = \sigma_{x's} = \sigma_{y's} = \sigma_{z's} = \sigma_s = \text{const},$$

$$\sigma_\alpha = \sigma_\omega = \sigma_\kappa = \sigma_{\alpha'} = \sigma_{\omega'} = \sigma_{\kappa'} = \sigma_{angle} = \text{const},$$

In simulation based on typical Sony CCD camera parameters, the following combinations of errors were used:

$$\sigma_{01} = 0.008 \text{ mm}, \; \sigma_{02} = 0.016 \text{ mm}, \; \sigma_{03} = 0.032 \text{ mm}, \; \sigma_{04} = 0.064 \text{ mm},$$

$$\sigma_{B1} = 0.001 \text{ m}, \; \sigma_{B2} = 0.002 \text{ m}, \; \sigma_{\text{angle 1}} = 0.015°, \; \sigma_{\text{angle 2}} = 0.03°, \; \sigma_{\text{angle 3}} = 0.1°$$

The results of mathematical simulation have shown, that:
- convergence of an iterative process is rather high (2-3 iterations);
- initial approximations for determination X, Y, Z could be rather far from a true value;
- symmetry of errors is observed, when the points are lying on the equal distances from a central part of the surface;
- more rough of all coordinates the distance Y is determined, its errors are practically not depend on distribution of points on the line, perpendicular to axial lines of the surface;
- measurement errors of basis are in the range from $\sigma_{B1} = 0.001$m to $\sigma_{B2} = 0.002$m causing practically identical influence on accuracy of spatial coordinates determination.

For various accuracy of images points coordinates σ_0 and various accuracy of exterior orientation angles σ_{yr}, the variance curves of X (Fig.1(a,b)), Y (Fig.1(c)), and Z (Fig.1(d)) are built.

The variance curves of X determination are given both for object points located on one of the axial lines of the surface (Fig.1(a)), and for object points located on an average part of the surface (Fig.1(b)).

For example, from the figures follows if the accuracy of pixel coordinates is 2 pixels ($\sigma_0 = 0.016$mm) and $\sigma_{\text{angle}} = 2'$, then the distance is determined with the error $\sigma_Y = 2$m, and object height is determined with the error $\sigma_Z = 0.04$m.

(a)

(b)

(c)

(d)
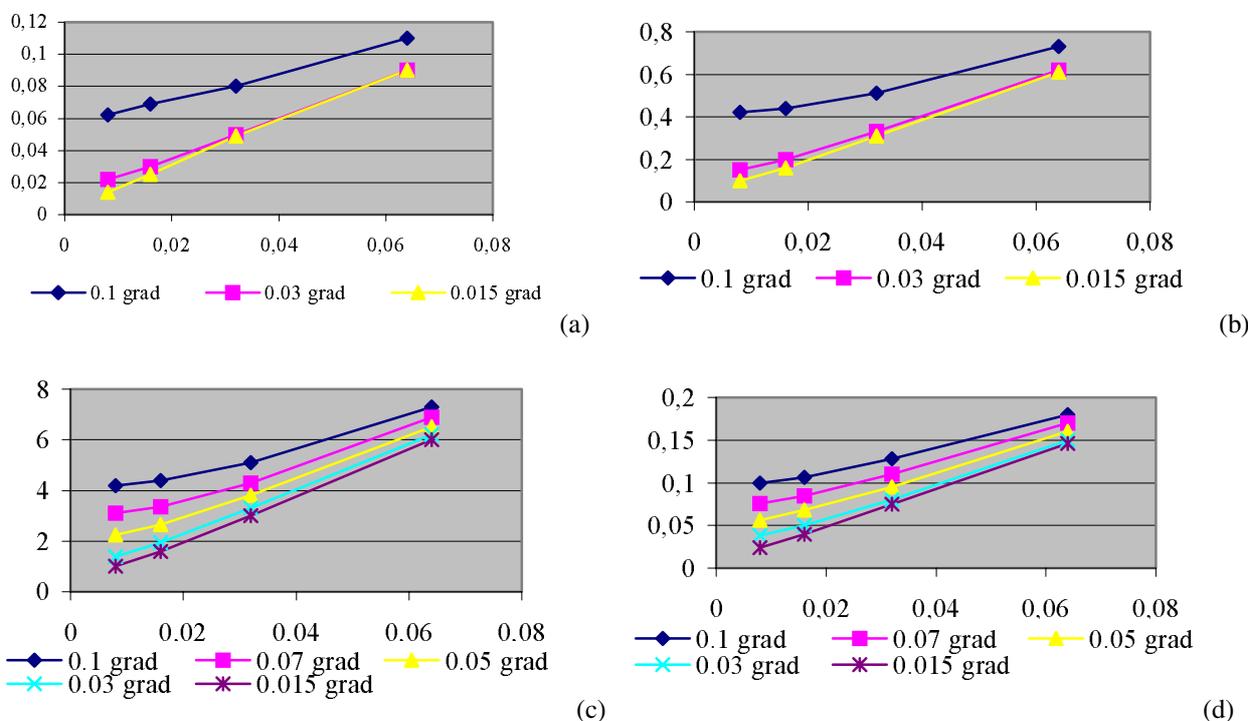
Figure 1. Accuracy of the object 3D-coordinates determination. The horizontal axis on all graphs shows accuracy of image coordinates σ_0 (mm)

(a) Variance curves of X determination in average part of the surface for various accuracy of angular elements of images exterior orientation ($\sigma_{yr} = \{0.1°, 0.03°, 0.015°\}$). The vertical axis shows σ_X (m).

(b) Variance curves of X determination in one of the axial lines of the surface for various accuracy of angular elements of images exterior orientation ($\sigma_{yr} = \{0.1°, 0.03°, 0.015°\}$); The vertical axis shows σ_X (m).

(c) Variance curves of Y determination for various accuracy of angular elements of images exterior orientation ($\sigma_{yr} = \{0.1°, 0.07°, 0.05°, 0.03°, 0.015°\}$). The vertical axis shows σ_Y (m).

(d) Variance curves of Z determination for various accuracy of angular elements of images exterior orientation ($\sigma_{yr} = \{0.1°, 0.07°, 0.05°, 0.03°, 0.015°\}$). The vertical axis shows σ_Z (m).

Thus, the results of simulation show that detection ability of the method is 4-8cm on a distance up to 80m, at use Sony CCD cameras with a digital format of 640x480 pixels.

3. OBTAINING A MODEL OF AN UNDERLYING SURFACE

In the presented work the situation is supposed when detecting object can be present at relatively smooth curved surface. Usual way of background surface model acquisition is least square method of fitting to the data set of points, which are exactly known to be not belonging to the object under detection. To find such data different methods of machine vision can be used.

Various methods of surface points obtaining based on machine vision algorithms are beyond the scope of this paper. As a result of such methods application an irregular set of points is acquired. Different interpolation methods can be used for obtaining a regular grid of surface model. The well-known inverse distance weighting (5) can produce reliable surface model.

$$Z(X,Y) = \sum w_i Z_i \Big/ \sum w_i , \qquad R_i^2 = (X - X_i)^2 + (Y - Y_i)^2 , \qquad w_i = \frac{1}{R_i^2} \tag{5}$$

Here (X_i, Y_i, Z_i) is a set of known points; $Z(X,Y)$ is a surface model.

Another model, which can be effectively used especially in road scenes, is [S.M.Smith, 1995]

$$Z(X,Y) = a_0 + a_1 X + a_2 Y + a_3 Y^2, \tag{6}$$

where coefficients a_i can be found by fitting the model to the known set of points using least squares method.

As an example we show a road model that comes from road signs and markings detection (see Fig.2).
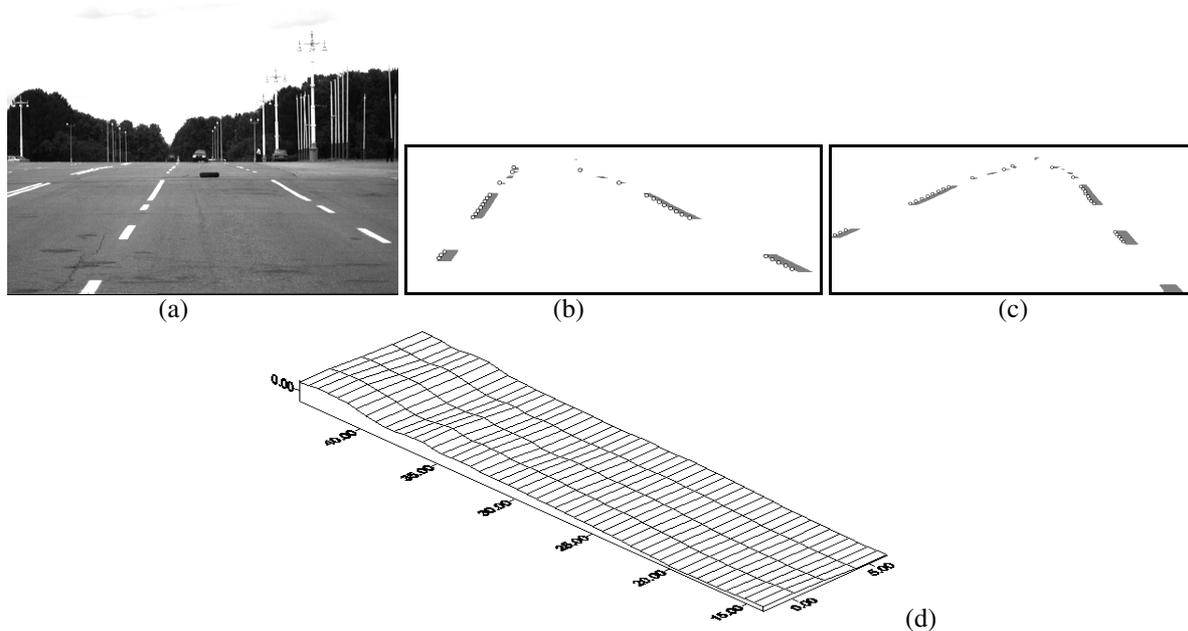
Figure 2. (a) Test image (left image of the stereopair); (b),(c) result of road marking detection for left and right images of the stereopair; (d) road model obtained from the result of road marking dete ction by inverse distance weighting.

4. ORTHOPHOTO DIFFERENCE GENERATION

Orthophoto is orthogonal projection of 3D scene in which all distortions caused by camera orientation and ones caused by 3D-shape of scene object are eliminated. Once orientation parameters of the cameras and reconstructed surface are known it is possible to generate orthophotos of the scene using left and right images (Fig.3(a,b)). The orthophotos are generated in the interest area $X_1 \le X \le X_2$, $Y_1 \le Y \le Y_2$ of the base plane. Denote orthophotos generated using left and

right images by *L(X,Y)*, *R(X,Y)* respectively. Subtraction of the orthophotos (Fig.3(c)) is performed by the following way: *D(X,Y) = L(X,Y) - R(X,Y)*.

| (a) | (b) | (c) |
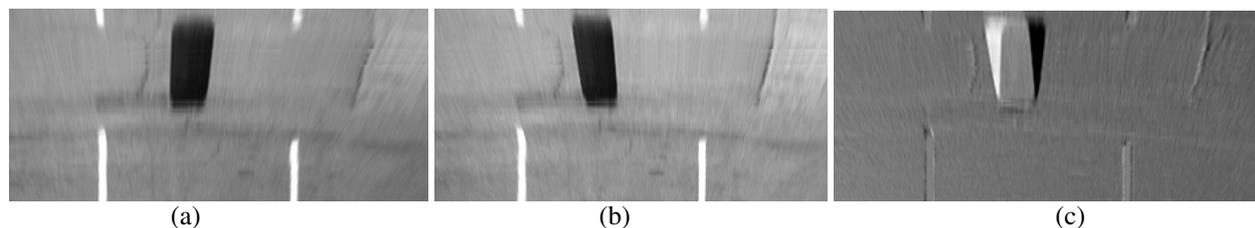
Figure 3. Example of orthophotos and their difference. (a) Orthophoto from left image. (b) Orthophoto from right image. (c) Difference: (a) minus (b)
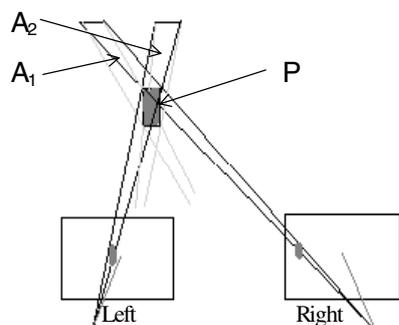
Figure 4. Invisible surface areas for left and right images

Pixel coordinates of *(i,j)*-th orthophoto point are calculated as follows: *(X,Y) = (X₁ + i * Sₓ, Y₁ + j * Sᵧ)*, where S_x, S_y - grid sample distances along X and Y axes respectively. The height *Z(X,Y)* is reconstructed from the surface using the bilinear interpolation of four nearest surface values. To assign a gray value to the *(i,j)*-th orthophoto pixel the point *(X,Y,Z)* is projected to the image using collinearity equations. The gray value is obtained by bilinear interpolation of four image gray values in pixels with nearest to the projected point integer coordinates.

The basic principle of the developed detection method is the significant difference in orthophoto due to different view positions of the cameras (Fig.4). In the left image the surface area A_2 behind the object P is invisible. The same is for the surface area A_1 in the right image. Thus the gray values in invisible areas are taking from object's gray values. This results in «projecting» the object to the invisible surface area.

5. DETECTION BY THE MATCHED FILTER

We assume that the object to be detected has rather contrast edges that are near to perpendicular to the surface. If the stereosystem with respect to the surface is under an angle of view close to zero, then edges of the object become elongated and enough rectilinear.

Let the object be a rectangle of height *h* and width *w*. Let the left-bottom corner of rectangle be located in surface point *(X,Y,Z)* perpendicular to the surface and parallel to the image plane. Then the invisible area behind the rectangle is a trapeze (Fig.5(a)) with dimensions satisfying the following equations (assuming the flat surface near the object (Fig.5(b))):

$$\frac{b}{b+L} = \frac{h}{|Z - Z_S|} = \frac{w}{c}, \quad L = |Y - Y_S| \tag{7}$$

In these formula L is a distance from the camera to the object.
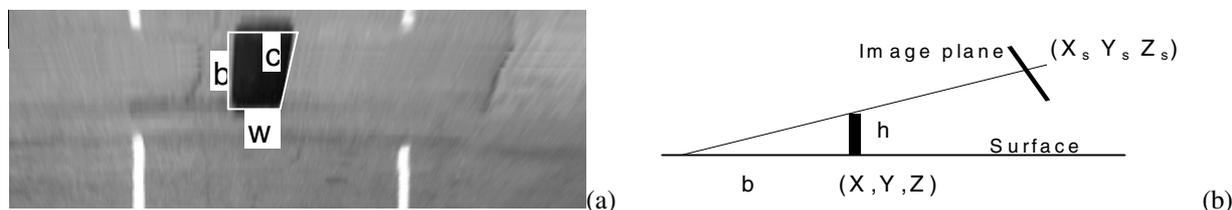
Figure 5. Object dimensions in orthophoto

For the following model parameters: *h* = 10cm., |*Z*- Z_s| = 1 m., *L* = 50 m the parameter *b* is equal to 12.5 m. This shows that Y-dimension of the invisible surface area dominates on other dimensions. So this fact actively used in the detection algorithm, because the orthophoto difference contains clear triangles formed by object edges.

Due to knowledge of relative position of cameras, detection of the objects can be implemented as a matched filter. Fig.6(a) shows the main idea of the method for two arbitrary orthophoto points. In each point (X, Y) a special mask M_{XY} consisting of three regions (S_+, S_-, S_+) is generated. The regions have characteristic dimensions a, b, which can be selected so that to ensure robust detection. The parameter b is selected from (7) when the minimum admissible height of an object is given. If the object height is unknown then the parameter b can be selected by iterations. The angle φ is determined by geometric constraints. Mask pixels in S_- region have values equal to -1, pixels in S_+ regions have values equal to 1, and all other values are 0. The result of the matched filter denoted by $P(X,Y)$ is obtained by a convolution of the orthophoto difference image with mask M_{XY}:

$$P(X,Y) = M_{XY} * D(X,Y) \qquad (8)$$

Fig. 6(d,e) show two examples of matched filter applying.
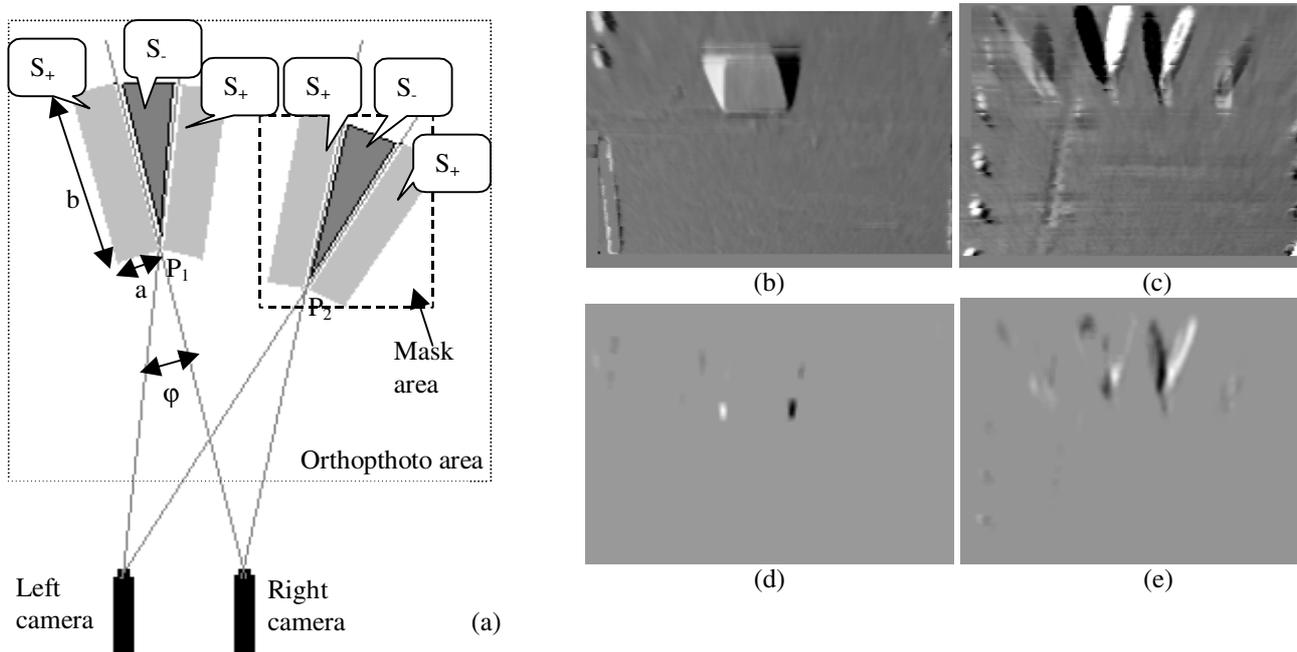
Figure 6. Detection of 3D-objects by the matched filtering.
(a) Configuration of the masks for each orthophoto points (on an example of two arbitrary points P_1, P_2).
(b),(c) Two examples of orthophoto difference images, (b) – one object, (c) – four objects.
(d),(e) Result of matched filtering for (b),(c)

6. RESULTS

This section shows the detection method application in more complex cases. As described in the previous section for the primary detection of the object the matched filter is used. Object selection can be performed by the filtering result $P(X,Y)$ analysis. Different algorithms such as local maximum searching can be used for this purpose. For validation of detected objects the criterion based on statistical decision theory was developed. It uses local features of the filter output to decide about object presence. Then the information from the original images is used for final confirmation of object presence.

Fig.7 shows the examples of the method applying. Detection results are shown on the left image of a stereopair. Each detected object edge is shown as a rectangle. Below the rectangle two parameters of the object are shown. The parameters are distance to the object in meters and object height in centimeters.

Figure 7. Examples of detection. (a) A car tire. (b) A pedestrian.

7. CONCLUSION

Method of using the difference in orthophoto images is well known for refinement of Digital Terrain Models [F. Raye Norvell, 1996]. However, this paper shows that similar idea can be used for object detection in a scene.

Our method has been tested on real data under different conditions. The problem with contrast objects belonging to the underlying surface (for example, drawn objects and signs) is eliminated by preliminary surface model computation. So the surface is excluded from the analysis by subtraction of orthophotos created by using the surface model and original images. The method can detect a large variety of objects since it is robust to the object shape and is sensitive to the abrupt deviation of the object from the surface only. The probability of detection is higher if the obstacle has sharp and contrast edges with respect to the surface. The photogrammetric validation shows high possibilities of the detection on the base of usual CCD cameras.

Further development of the method above is connected with improving various techniques for underlying surface reconstruction.

REFERENCES

F. Raye Norvell, 1996, Using Iterative Orthophoto Refinement to generate and correct Digital Elevation Models (DEM's). Digital photogrammetry: an addendum to the manual photogrammetry, ASPRS,1996

M.Bertozzi and A.Broggi, 1997, Vision-based vehicle guidance, IEEE Computer, vol.30, pp.49-55, July 1997

S.M.Smith, 1995, ALTRUISM: Interpretation of three-dimensional information for autonomous vehicle control, Technical report, http://www.fmrib.ox.ac.uk/~steve