

## 3D-OBJECT DETECTION METHOD BASED ON THE STEREO IMAGE TRANSFORMATION TO THE COMMON OBSERVATION POINT

V. M. Lisitsyn<sup>\*</sup>, S. V. Tikhonova<sup>\*\*</sup>

State Research Institute of Aviation Systems, Moscow, Russia

<sup>\*</sup>lvn@gosniias.msk.ru

<sup>\*\*</sup>svetlana-t@mtu-net.ru

**KEY WORDS:** algorithms, stereoisimages processing, 3D object detection, computer vision.

### ABSTRACT

The method of 3D-object detection by stereo image processing is considered where object is situated on the smooth surface. Stereo image is obtained by passive sensor such as TV-camera, IR-camera and so on. It is supposed that some long landmarks are present on the underlying surface and position of those landmarks is determined for both left and right images of the stereo-pair.

To detect object both images of a stereo-pair are transformed to a common observation point. For images' tying, equations are used that describe position of long landmarks on the left and right images. After the considered transformations, the fragments of the stereoscopic picture are subtracted. On the difference image, the patterns corresponding to figures on the underlying surface are close to zero, while contrast points related to three-dimensional objects are marked as face-down triangles. The presence of such patterns becomes a sign that 3D-object is detected. To search for such patterns, the contour selection on the difference picture and matched filtering of contour preparation is performed.

### 1 INTRODUCTION

The method of 3D-object detection by stereo image processing is considered where object is situated on the smooth surface. Stereo image is obtained by passive sensor such as TV-camera, IR-camera and so on.

As a smooth surface the plane or cylindrical surface, i.e. surface obtained by transition of a straight line parallel to itself along some curve, can be considered. Here the underlying surface does not shade itself and is oriented along an optical axis of cameras. It is supposed that some long landmarks are present on the underlying surface and position of those landmarks is determined for both left and right images of the stereo-pair. The landmarks position is used for images tying on both left and right images of the stereoisimage. Landmarks should be oriented along an optical axis of cameras. Landmarks could be given by lines or curves equations in a camera plane or as a array of points for both left and right images of the stereo-pair.

### 2 DESCRIPTION OF THE METHOD

To detect objects both images of a stereo-pair are transformed to a common observation point, i.e. the stereoscopic images are transformed in such a way as if the TV-cameras have been displaced in a plane-parallel manner to the common observation point. For example it may be the mid-position between the cameras. Such transformation corresponds to rotation of the underlying surface around an infinitely remote point.

In the image plane this leads to such an image transformation when the pixels on the line of horizon (that correspond to an infinitely remote point) remains unaltered and pixels below the line of horizon are shifted. And in this case, the lower the pixel is (the less is the distance to the observed surface) the greater is the shift.

Lets define a co-ordinates with the center in the focal plane of a camera and with the axis  $Y$  directed upwards, axis  $X$  along an optical axis of the camera and axis  $Z$  to the right. In a such a co-ordinates the shift  $S$  in pixels is determined as

$$S = \frac{\Delta \cdot F}{D_{np} m_z}$$

where  $D_{np}$  is the projection of distance to the point on an axis  $X$ ;  $F$  is the focal length of camera lenses;  $m_z$  is the scale along the axis  $Z$ ;  $\Delta$  is virtual shift of the camera along the axis  $Z$ .

If the plane where the landmarks are situated is parallel to the defined  $XOZ$  plane and  $H$  is the distance between planes then

$$S = \frac{\Delta \cdot m_y}{H \cdot m_z} (n_{line} - N)$$

where  $m_y$  is the scale along the axis  $Y$ ,  $n_{line}$  is the line number from the bottom of the image, and  $N$  is the line number corresponding to the camera optical axis position. Thus, for a planar model the shift value is a linear function and such transformation is an affinity one.

After these transformations, all lines and figures located on the underlying surface are reduced to a form identical on both left and right images of the stereo pair. At the same time, three-dimensional objects acquire a certain inclination: to the right - on the left image, and to the left - on the right image.

After the considered transformations, the fragments of the stereoscopic picture are subtracted: the right image is subtracted from the left. On the difference image, the intensity of patterns corresponding to figures on the underlying surface are close to zero, while contrast points related to three-dimensional objects are marked as face-down triangles. The presence of such patterns becomes a sign that 3D-object is detected. The above considered transformations are analogous, to a certain extent, to methods considered in (Zheltov, 2000), but without modeling of the underlying surface.

Suppose there is a vertical plane perpendicular to the camera optical axis and there is a straight line in this plane that form angle  $\gamma$  with the vertical axis. Then after above considered transformation this line will be transformed into the straight line and angle  $\gamma$  will be transformed into the angle  $\gamma_n$ . Here

$$tg(\gamma_n) = tg(\gamma) + \frac{\Delta \cdot m_y}{H \cdot m_z}$$

Thus if two contours on a difference image were generate by the same contour of some 3D object, then the following is true:

$$|tg(\gamma_{left}) - tg(\gamma_{right})| = \frac{B \cdot m_y}{H \cdot m_z} \quad (1)$$

where  $B$  is the base of stereo pair,  $\gamma_{right}$  is the angle of the contour on difference image that was generated by the contour of 3D object on the right image, and  $\gamma_{left}$  is the angle of the contour on difference image that was generated by the contour of 3D object on the left image.

Thus if some 3D object has nonzero contrast with the background, then after above considered transformations patterns looking like face-down triangles will be present on difference image in positions defined by the position of the 3D object edges. The angle between the sides of such a triangular pattern is determined by (1).

Different methods could be used to search for these patterns. Here we consider the method based on the matched filtration. This method allows us to search for contours with beforehand determined incline (in this case – vertical contours that are typical for artificial objects).

For convenience we form two difference images: one from positive pixels and another from negative pixels. To search for characteristic patterns (i.e. face-down triangles) contour selection of difference images and matched filtration of

contours are performed (we consider matched filter features later). We will be searching for pairs of contours – one from “positive” part of difference image and other from it’s “negative” part – that touch each other in the bottom point and have beforehand determined angels of incline difference. In order to increase operational stability of algorithms threshold processing and binarization of difference images and suppression of isolated points are performed after contour selection.

We say that object is detected if output response of the matched filter has the value exceeding some threshold value. In case when the matched filter gives a multiple response, these multiple responses are combined into connected regions. In each region the maximum is identified. Position of the object and the distance to it could be determined by the pixel corresponding to this maximum.

Block diagram of the algorithm is presented on Fig.1.

### 3 MATCHED FILTER DESCRIPTION

The matched filter consists of a doubled rectangular mask. One part of the mask is designed to detect positive contours inclined in the pre-determined range, and the other part is designed to detect negative contours inclined in the pre-

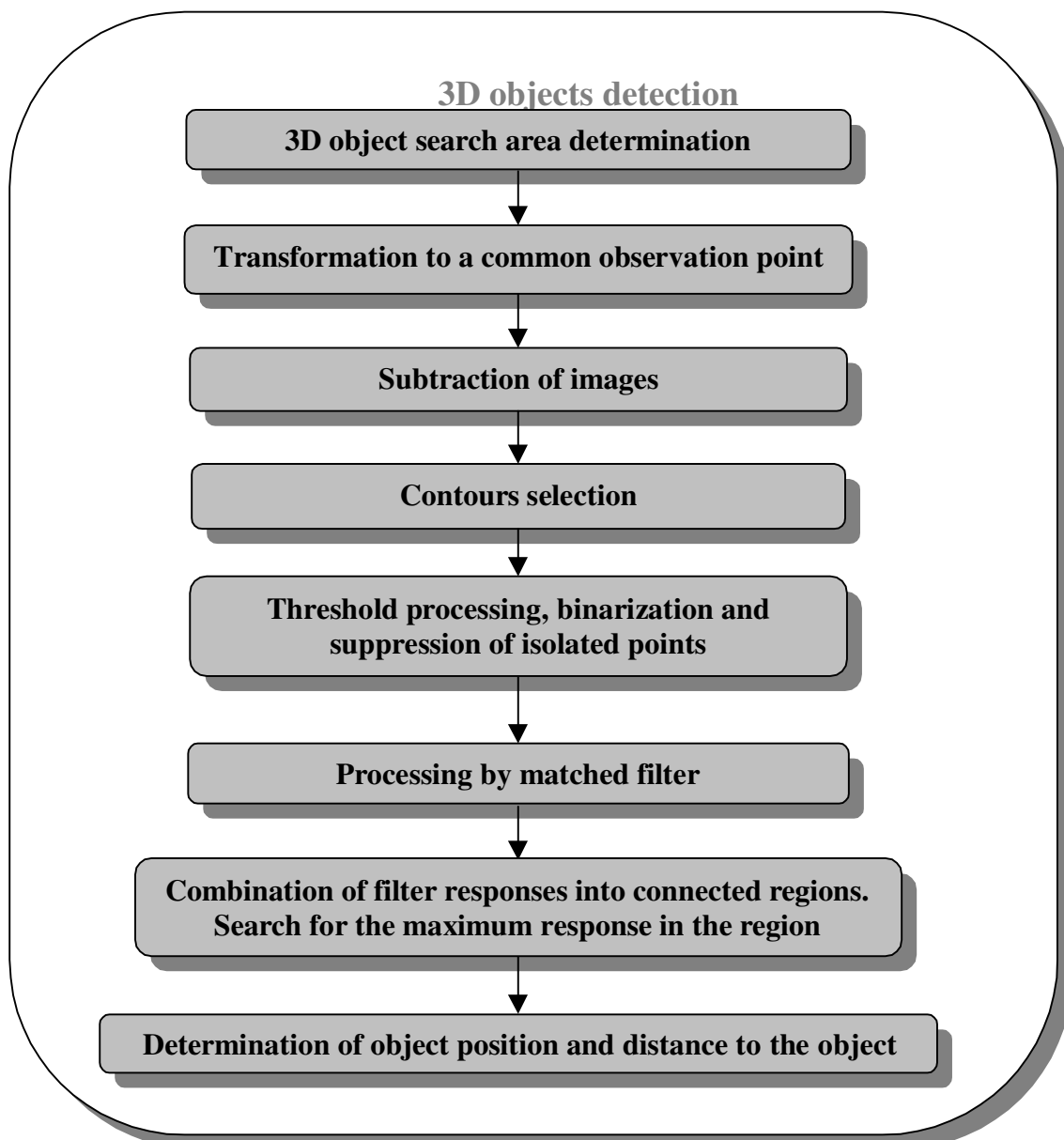


Figure 1.

determined range. In this case, both binary images of positive and negative contours are being processed simultaneously.

The filter operates as follows. The image of positive contours is scanned using the semi-mask detecting positive contours of the required inclination. In points where output response of "positive" semi-mask exceeds the threshold, the calculation of "negative" semi-mask response on the negative image is performed. Horizontal displacement of this semi-mask in some minor neighborhood area relative to the position set by the "positive" semi-mask is performed. The maximal response value within the above neighborhood area is found, and this value is summed with the "positive" semi-mask response value. If the total filter response exceeds the pre-determined threshold, then it is assumed that a contour belonging to an 3D object is present in the given point. If not, then scanning of the positive contour images is continued.

The displacement of the "negative" semi-mask in some region is intended to diminish the effects of TV-cameras' pitch misalignment and inaccuracies of images' transformation to the common observation point on the operation of the algorithm. In such cases, on the difference image there appear not the typical facedown triangles, but facedown trapeziums, which are successfully detected by this procedure.

#### 4 MODELING RESULTS

Figs. 2-6 illustrate the algorithm operation.

Fig. 2 shows the initial stereoscopic image with long landmarks (road marking lines in this case) that were determined using some procedure. This procedure is not considered in this paper. Landmarks positions are defined in an image coordinates by equations of straight lines approximating road marking inner contours. Object detection is performed within the rectangular area.

Fig. 3 illustrates the results of image transformations to the common observation point. This transformation is performed for the object search areas only. The image of a 3D-object has got certain inclination: to the right on the left image, and to the left on the right image.

Fig. 4 represents the results of subtraction of the image fragments corresponding to the object search areas. Here, pixels that have positive values on the difference image form the right image and pixels that have negative values on the difference image form the left image. The images incorporate four types of pixels: values of pixels derived by subtracting background from background, object from object, object from background and background from object. The values of pixels obtained by subtracting background from background and object from object are close to zero, while the values of pixels obtained by subtracting background from object and object from background, are significant and proportional to the object contrast relative to the background. The difference picture incorporates road marking contour lines resulting from the imprecise matching of images. In points of the object boundaries typical facedown triangle patterns are present on the difference picture.

Fig. 5 shows the results of the inclined contours selection on the difference images (positive and negative), threshold processing and suppression of isolated points. Contours of corresponding inclination have appeared in place of triangles. Each triangle is transformed into a pair of contours: one positive, and the other negative. The presence of such a pair is the sign that an 3D-object is detected in the given point.

Fig. 6 represents the results of the matched filter operation. The detected object is marked by bright dots on its edges.

Investigations of the algorithm sensitivity to the TV-cameras misalignment were performed. To do it, one image of a stereo pair (left or right) was displaced by software. The modeling has shown that the algorithm operation is stable (i.e. the algorithm detects 3D-objects on images where they are present and does not give false alarms in case of images with no 3D-objects) for images with vertical misalignment up to 5-6 pixels. The algorithm's sensitivity to a horizontal camera misalignment is even lower, as it affects primarily the preciseness of determination of the distance to the object.

As a whole, the performed modeling has proved stable operation of the algorithm.

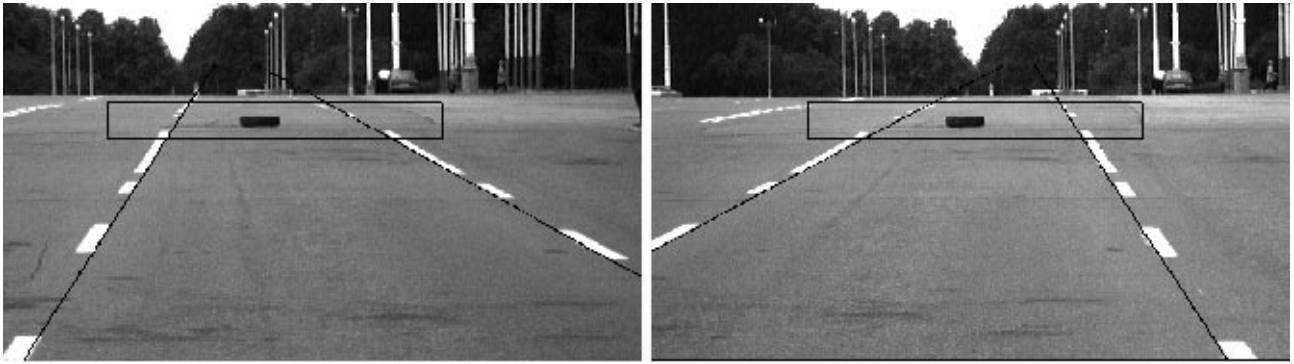


Figure 2.

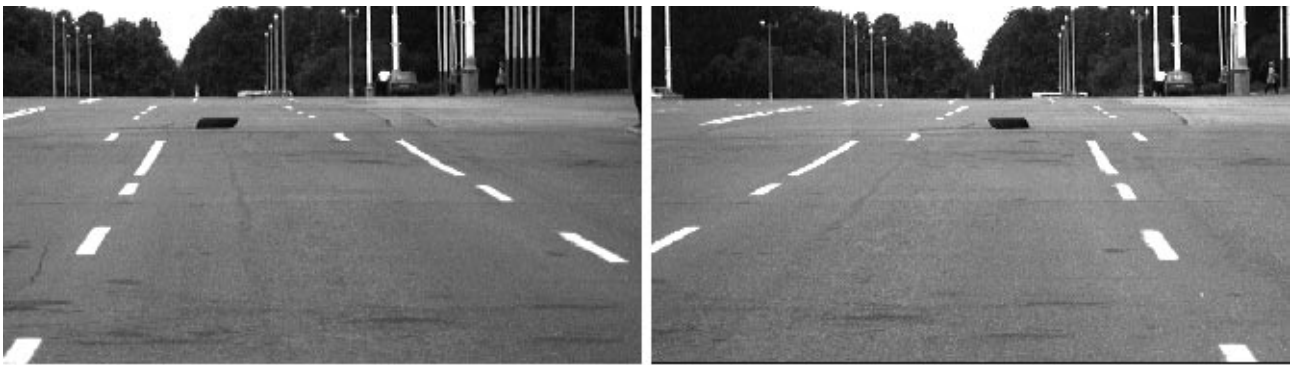


Figure 3.



Figure 4.



Figure 5.

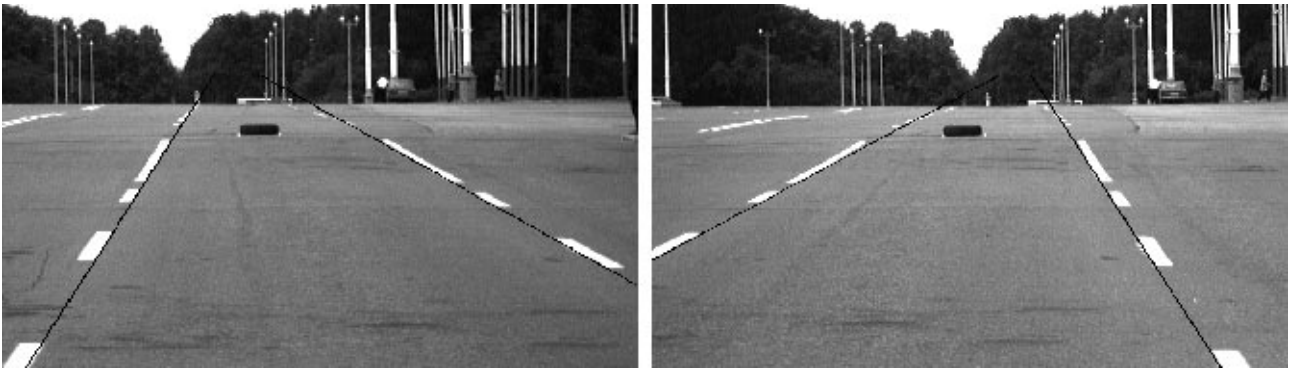


Figure 6.

## REFERENCES

Zheltoy, S. and Sibiryakov, A., 2000. Method of 3D-Object Detection Based on Orthophoto Difference Analysis. In *Archives of ISPRS*, vol. XXXIII, part 3, Amsterdam.