# Using geometrical rules and a priori knowledge for the understanding of indoor scenes

C. Coelho, M. Straforini, M. Campani

Dipartimento di Fisica dell' Università di Genova,
Via Dodecaneso 33, 16146 Genova, Italy

*The main goals of Computer Vision are the recovery of the three dimensional structure and the recognition of viewed objects from T.V. images. The understanding of images of indoor scenes, such as corridors and offices, is greatly simplified by the exploration of geometrical properties of the viewed scene. In these scenes there are many long straight edges which are either parallel or orthogonal and the viewed objects can be modeled as elements of Legoland. Legoland is a block world where physical edges have at most three directions, which are mutually orthogonal.*

*In this paper a general procedure is proposed for the understanding of indoor scenes which is likely to be useful in Computer Vision. The proposed algorithm makes use of an extensive 2D image processing which is able to provide a faithful line drawing of the viewed scene. It also uses procedures adequate for both the understanding of the 3D structure of the viewed scene and for the recognition of several items in the scene.*

The proposed vision system has two main components: an extensive processing of the 2D image followed by a reasoning system, which uses geometrical rules and a priori knowledge [2],[5]. The aim of the first component is to provide a reliable line drawing of the image where vertexes and polygons have been located and identified. The second component provides rules for the understanding of the 3D structure of the viewed scene and is able to recognize several classes of objects. The two components of this vision system are described in part I and II of this paper. The proposed scheme for the understanding of indoor scenes performs remarkably well by making use of the fact that these scenes can be described as belonging to a polyhedral world, called *Legoland*, where straight edges have at most three orthogonal directions [2]. The implementation of this system runs on a Sun SparcStation 1 in less then 10 seconds.

# PART I

## VANISHING POINTS

Under perspective projection, segments in the image plane, which correspond to parallel lines in 3D space, converge to a common point of intersection called *vanishing point*. An new approach for the detection and localization of vanishing points has recently been proposed [8], which detects vanishing points with a linear computational complexity and good accuracy also in the presence of objects not belonging to the Legoland world. Vanishing points are very important because they allow the recovery of the normal vectors of surfaces belonging to a polyhedral world such as *Legoland* [2].

## THE RECOVERY OF VERTEXES AND OF THE LINE DRAWING

In this section we describe procedures able to recover junctions from a list of segments (see Fig. 2). These junctions or vertexes are usually poorly detected by standard edge detection schemes [3], [6] which very often break trihedrical vertexes and round off $L$ junctions. The procedure for the recovery of vertexes can be divided into two steps:

i) merging of collinear and adjacent segments;

ii) detecting vertexes;

### Merging Segments

For the reconstruction of the polygons it is useful to reduce the number of segments extracted from the image by a procedure of elimination and merging. The procedure starts first by selecting only those segments with a length longer than a threshold, and that converge to at least one vanishing point . The algorithm for merging segments is divided into three steps:

i) Two cellinear and adjacent segments converging to the same vanishing point are merged together. Fig.1 illustrates the role of two thresholds, $\epsilon_c$ (10 pixels) to establish collinearity and $\epsilon_a$ (100 pixels) to establish adjacency.

ii) As a result of the previous step, two merged segments may cross another segment. This intersection is not a real vertex in the line drawing and therefore it must be eliminated by breaking the merged segment. This test allows us to use such a high value for the parameter $\epsilon_a$, without creating spurious junctions.



BMVC 1990 doi:10.5244/C.4.41

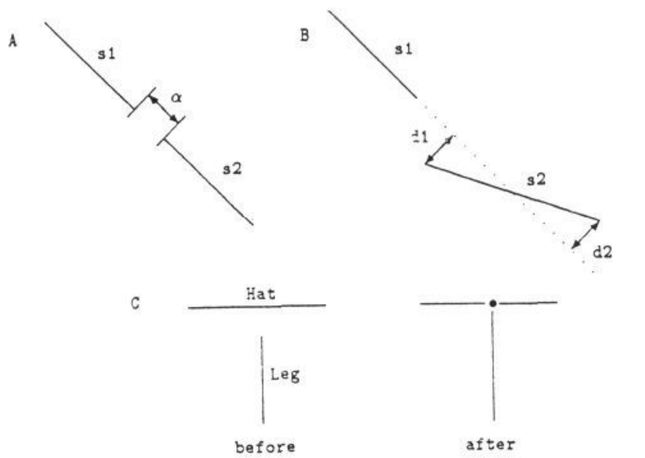


Figure 1: *A: two collinear segments s1 and s2 are adjacent if the distance $\alpha$ between the two nearest end points is smaller than $\epsilon_\alpha$. B: two segments s1 and s2 are collinear if $d_1 < \epsilon_c$ and $d_2 < \epsilon_c$. C: hat and leg before and after the formation of a T junction*

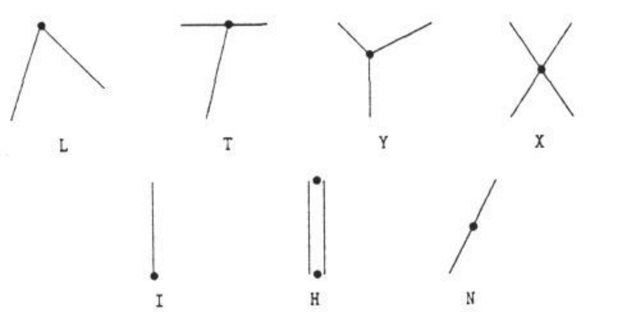


Figure 2: *Vertices and junctions which are likely to occur in real images.*

iii) $T$ junctions (see Fig. 1C) are rather important for the understanding of line drawings [7]. A $T$ junction is usually formed by 3 segments, two of which are collinear and can be named as *hat* and the third segment can be referred to as *leg*.

## Vertexes

Figure 2 illustrates the different kinds of vertexes which are usually found in real images. More complicated junctions are not considered here. The first four vertexes, indicated as $L, T, Y$ and $X$, play a crucial role in the proposed approach, while the remaining three vertexes, indicated as $I, H$ and $N$, are less important and for this reason they are deleted. The algorithm for the detection of a vertex explores all end points in the segments list. Every end point, which has not yet been identified as a vertex, is connected to the nearest end point provided that the distance between end points belonging to the same vertex is smaller than a given threshold (about 10 pixels).

## Simple Polygons

A polygon is *simple* when its sides (with the exception of the segments constituting the frame of the image) are formed by segments that converge to only two vanishing points [2]. Other types of polygons are considered *not simple*. The *colour* of the simple polygon is a coding of the pair of vanishing points on which its sides converge. A simple polygon can be interpreted as the perspective projection of a not occluded planar panel [1]. This default rule, which can be referred to as *the rule of maximal visibility*, is usually true and provides a good clue for the understanding of indoor scenes.

In many cases, because of the presence of shadows or reflexes, a panel is seen in the image plane as being composed of several not simple polygons, and so it is useful to merge them into polygons which are simple. In other cases it is possible that a region is not simple because in its perimeter there are short segments converging to a third vanishing point. *Almost simple polygons* are those where at least 90 % of their perimeter is formed by segments converging to only two vanishing points. Almost simple polygons are coloured as simple polygons.

In order to obtain a useful segmentation of the 2D image and to understand the structure of the 3D scene, it is useful to group adjacent simple polygons of the same colour into *maximal polygons.*

## Experimental Results

The images grabbed by a CCD camera are stored in a matrix of $512 \times 512$ pixels with 256 gray levels. Edges are detected from the images by the Canny operator [3] followed by a chain follower and a polygonal approximation [4].

Figure 3 illustrates an image of a corridor at the Department of Physics (A) and the segments obtained by the polygonal approximation of the edge chains (B). The output of the algorithm described in section 3.1 is reproduced in (C). The three vanishing points are first detected and located and only long segments (longer than 10 pixels) converging to one of the three vanishing points were retained. Finally, collinear and adjacent segments are merged and $T$ junctions were recovered. The final set of segments after the deletion of spurious segments is shown in (D). Figure 4 reproduces the final line drawing with vertexes labeled as $L, Y, T$ and $X$ junctions. The obtained line drawings seem to capture basic geometrical features of viewed scenes correctly. The results of the detection and fusion of simple polygons (and of almost simple polygons) from the line drawings of Fig. 4 are shown in Fig. 5.

# PART II

## LABELING EDGES AND GEOMETRICAL RULES

In this section, it is shown how the labeling of edges as *convex*, *concave* or *occluded* and the use of some simple geometrical rules allows a partial recovery of the 3D structure of the viewed scene from the 2D line drawing of the image. The algorithm for labeling edges is described in [2].

Let us consider the maximal polygons in Fig. 6 obtained from the original image shown in Fig.3A. The four $Y$ junctions connecting polygons with three different colours, indicated by a circle, can be used for the

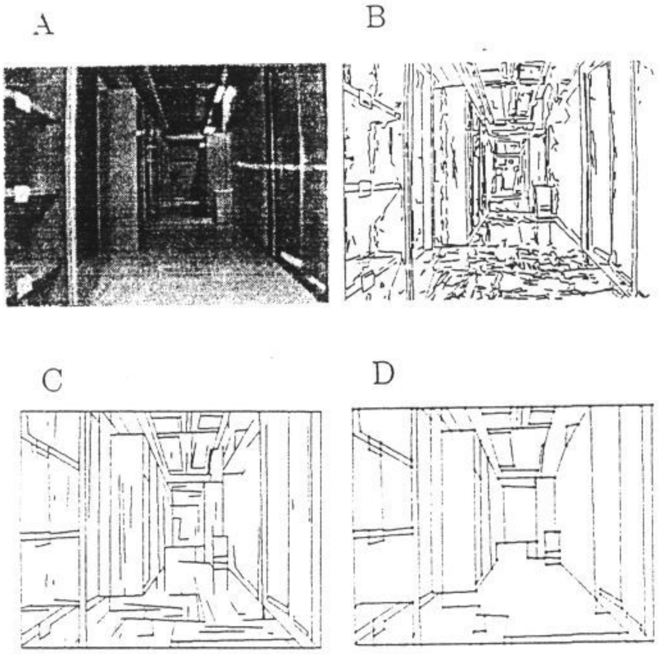


Figure 3: *A: an image of a corridor at the Department of Physics. B: segments obtained by the polygonal approximation of edge chains. C: Segments after the merging procedure described in section 3.1 D: The planar graph after the elimination of spurious edges and vertexes.*

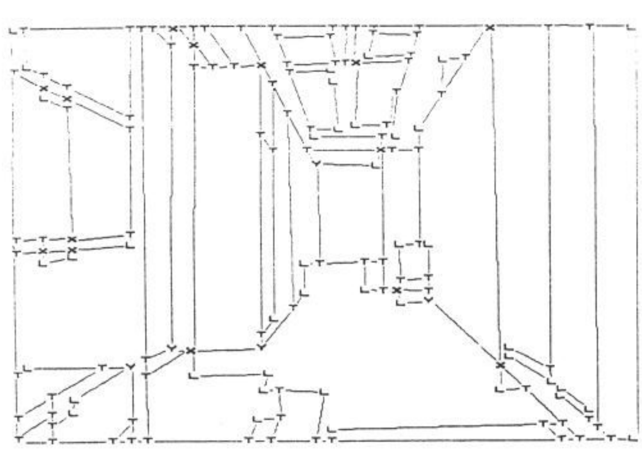

Figure 4: *The final graph of detected edges (from the original image in Fig. 3A) and of vertexes, labeled as $L, Y, T$ and $X$.*

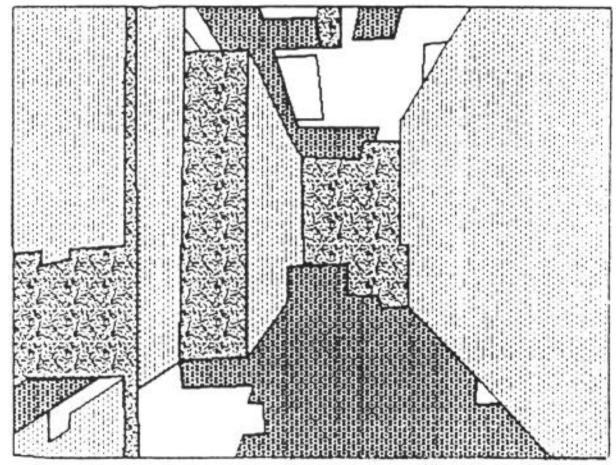

Figure 5: *Maximal polygons from image 3A. The different texture of polygons represents a different colour.*

understanding of the 3D structure of the viewed scene. The proposed procedure is divided into three steps:

i) the horizon lines are drawn and they divide the image plane into sectors where visible planar panel can have three different normal vectors at most (see [1],[2]).

ii) $Y$ and $T$ junctions are extracted from the graph of labeled vertices. By using the heuristic that edges between $Y$ junctions are either concave or convex (i.e. they cannot be occluding edges) and the labeling algorithm described in [2], several edges are labeled. After the completion of this procedure, the $T$ junctions are considered. The hat (see Fig. 1C) of a $T$ junctions is considered an occluding edge [7], provided that the same edge has not been previously labeled as convex or concave. In the use of conflict, the heuristic based on the $Y$ junction is assumed to be stronger than the heuristic based on the $T$ junction.

Fig. 6 illustrates edges labeled by the proposed procedure and it is evident that the basic $3D$ features of the scene have been captured. The planar panels corresponding to the two walls and the floor are correctly identified and located. The bottom end of the corridor is correctly identified, although it is merged with other simple polygons of the same colour. The third step explicitly uses simple geometrical rules:

iii) simple polygons with the same colour separated by simple polygons with a different colour through a convex and a concave edge are the perspective projection on the image plane of not coplanar 3D planar panels.

These geometrical rules give rise to the conclusion that in the 3D world the planar panel 1 is nearer the observer than the planar panel 2.

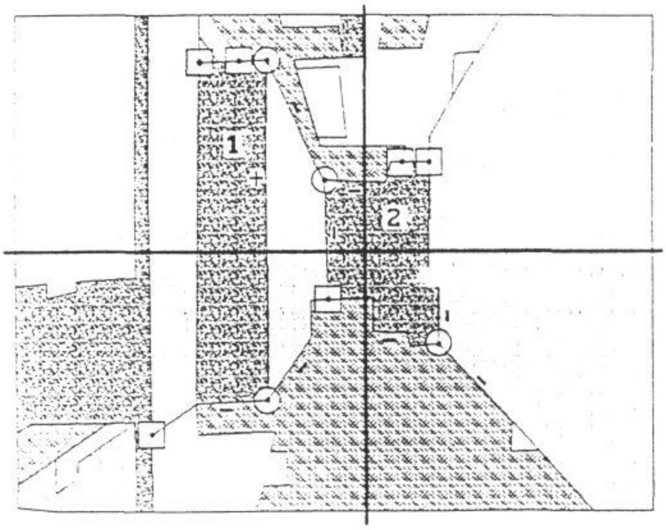


Figure 6: *Understanding of maximal polygons shown in Fig. 5D. Horizon lines divide the image into four sectors. Edges labeled as + are convex edges and edges labeled as - are concave edges. Y and T junctions have been marked by circles and squares respectively. The planar panel 1 is interpreted as nearer to the viewing camera than the planar panel 2.*

## THE ANALYSIS OF SEQUENCES OF IMAGES

It is of some relevance to enquire about the stability and robustness of the proposed algorithm. Fig.7 illustrates 9 out of 16 frames obtained in the same corridor, shown in Fig. 3A, taken while the T.V. camera was translating. During the acquisition of the images a door was opened, causing clear changes in the illumination of the scene. Fig. 8 reproduces the maximal polygons obtained from the sequence of the 9 images. The detected polygons exhibits some degree of stability, because the values of parameters were not adjusted in each frame.

## THE USE OF A PRIORI KNOWLEDGE

Experience in the analysis of real images, suggests that a robust identification and recognition of doors, floor, ceiling, and of similar objects requires the use of a priori knowledge of the viewed environment. A priori knowledge leads to visual routines specifically for the detection in the scene of elements such as *the bottom end of the corridor*, the em left and right wall, the *floor* and *vertical obstacles*.

In the case of central projection (when a vanishing point is close to the optical center) these routines are simplified. For example:

- The routine for the detection of the bottom of the corridor searches for the maximal polygon (not necessary simple) inside which a vanishing point is located.
- The routine for the detection of walls searches for simple polygons (with the colour corresponding to vertical panels oriented perpendicularly to the im-

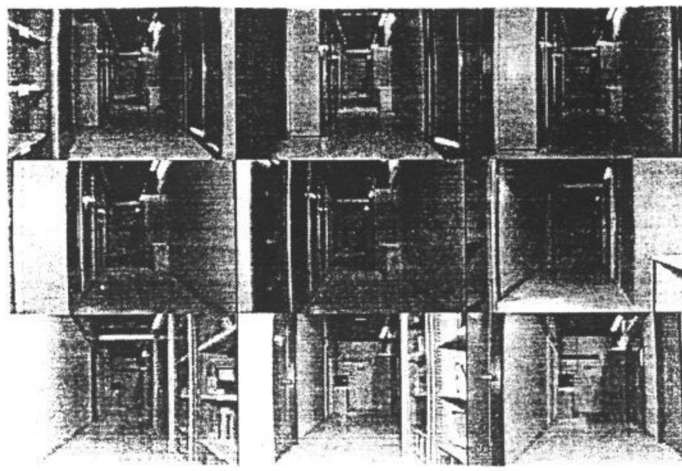

Figure 7: *A sequence of 9 images taken in a corridor at the Department of Physics. Notice how the illumination changes from frame to frame, and how the images are rich in reflexes and shadows.*

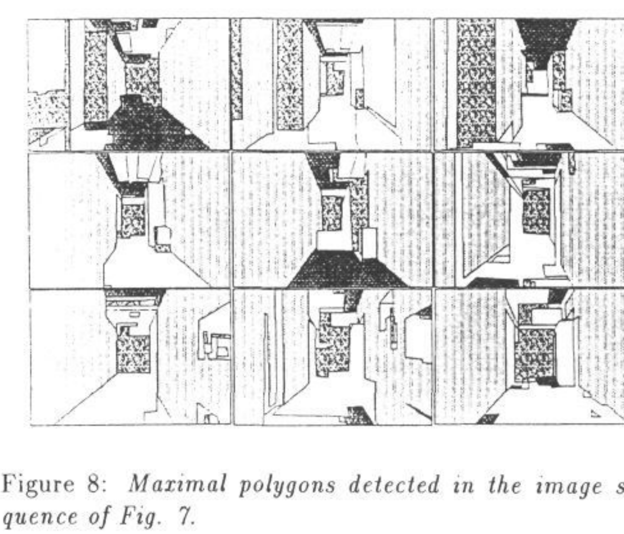

Figure 8: *Maximal polygons detected in the image sequence of Fig. 7.*

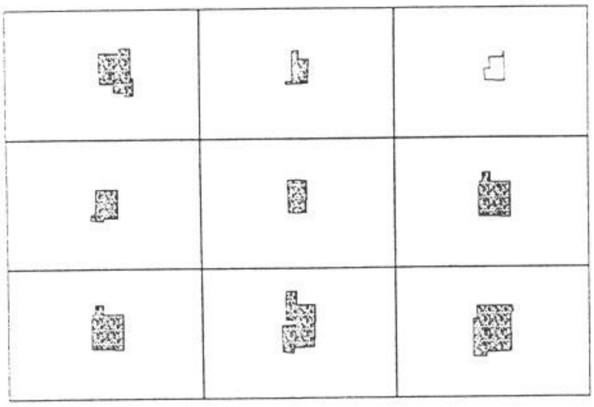

Figure 9: *The detection of the bottom end of the corridor in the image sequence of Fig. 7.*

age plane) with highest and lowest vertices above and below the horizontal horizon line respectively.

- The routine for the detection of the floor searches for the region of the image plane between polygons identified as right wall, left wall and bottom of the corridor.

When we are not in central projection, these routines are slightly different and appear to work successfully in a variety of different offices.
Fig. 9 illustrates the detection of the bottom end of the corridors from the images sequence of Fig. 7. The end of the corridor is always correctly identified although the detected polygon is sometimes merged with neighbouring regions which are not the projection of the end wall. It is possible, however, by an appropriate reasoning procedure to refine the detection of the end wall by eliminating regions erroneously merged to the polygon identified as the projection of the bottom end of the corridor. The routine is able to detect as the bottom end also a polygon not classified as simple (see the third frame). Fig. 10 reproduces the detection of the right and left walls and of vertical obstacles, which proves to be rather reliable. In the eighth frame part of the floor has been merged to a polygon identified as the right wall, because of the presence of a strong highlight erroneously confused as a physical edge. Fig. 11 illustrates the output of the algorithm for the detection of the floor. It is now evident that the recognition of these elements is more stable and robust. From the results shown in fig. 10 is clear that the floor is usually detected, although its borders are not exactly identified.

## DISCUSSION

The aim of this paper is two fold: firstly to present a 2D image processing, able to provide a faithful line drawing of the viewed scene, and secondly to show procedures which are adequate for both the understanding of the 3D structure of the viewed scene, and for the recognition of several items in the scene. Experience so far on the analysis of images suggests that successful use of

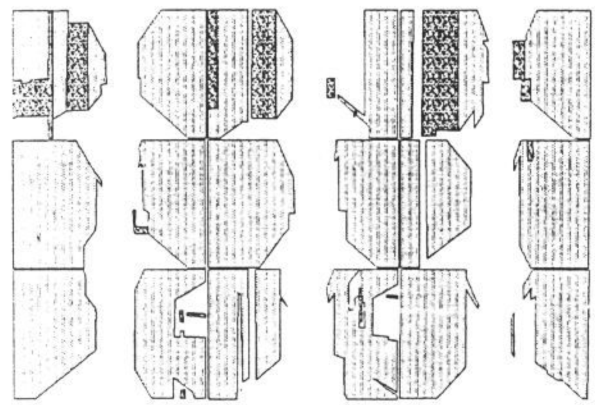

Figure 10: *The detection of walls and vertical obstacles in the image sequence of Fig. 7.*

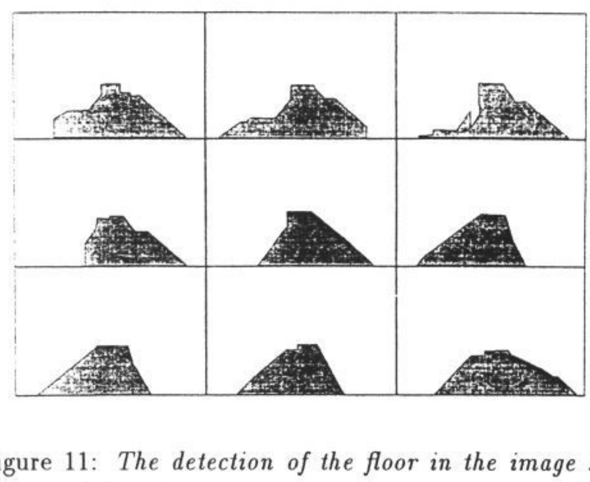

Figure 11: *The detection of the floor in the image sequence of Fig. 7.*

*high level vision* procedures needs an efficient *low level* processing and that high level vision by itself is not sufficient to overcome the poor low level processing. Here several techniques for understanding images have been analyzed: labeling edges, reasoning about geometry and using a priori knowledge [2],[5]. Consequently, it is suggested that machine vision has to investigate the properties and the best strategies of these vision routines in order to go beyond the limits of low level vision.

## ACKNOWLEDGEMENT

This work has been supported by the ESPRIT project 2502 VOILA, PROMETEUS project and Progetto Finalizzato ROBOTICA (C.N.R.). Dr. Christopher Coelho was supported by a fellowship from ELSAG (Elettronica San Giorgio s.p.a.). Dr. Marco Straforini is supported by Genova Ricerche. Clive Prestt checked the English.

## REFERENCES

1. **Fiumicelli,A. and Torre,V.** "On the Understanding of Line Drawings", *SPIE* vol. 726 Intelligent Robots and Computer Vision: Fifth in a Series 1986.

2. **Bellutta, P., Collini, G., Verri, A. & V. Torre.** "Recovering $3D$ information from vanishing points". *IEEE workshop on interpretation of 3D scenes*, Austin, Texas,U.S.A., 1989.

3. **Canny,J.** "A Computational Approach to Edge Detection", *IEEE Transaction on PAMI*, vol. 8, no. 6, 1986, pp. 679-698.

4. **Pavlidis,T.** "Algorithms for Graphics and Image Processing", *Computer Science Press*, Rockville, MD, 1982.

5. **Kapur, D. and Mundy, J.L.** "Geometric Reasoning and Artificial Intelligence: Introduction to the Special Volume", *Artificial Intelligence*, Vol. 37, No. 1, 1988, pp. 1-11.

6. **De Micheli E.,Caprile B.,Ottonello P. and Torre V.** "Localization and Noise in Edge Detection", *IEEE Transaction on PAMI*, vol. 11, no. 10, 1989, pp. 1106-1116.

7. **Huffman D.** "Impossible objects as nonsense sentence ", *Artificial Intelligence*, (2), 1971.

8. **Coelho C., Straforini M., Campani M.** "A fast and precise method to extract vanishing points", *SPIE's International Symposia on Applications in Optical Science and Engineering*, Boston 1990.