A Robust and Effective Method for Bidimensional Recognition of 2D and 3D Objects from Intensity Images

Sergiu Nedevschi
Computer Science and Engineering Department,
Technical University of Cluj-Napoca, Romania

Abstract: *This paper presents a robust and effective method for bidimensional recognition of 2D or 3D objects, from intensity images based on: matching of the models with symbolic structures of the scene, using of the inexact matching, intensive using of techniques for search space reduction.*

Keywords: *Object recognition, inexact matching, intermediate symbolic representation, search space reduction.*

1. INTRODUCTION

The most recognition methods try to realise a 1:1 matching between sensorial and model features. The request of 1:1 matching between features reduces the applicability of the methods.

Because of the impossibility to maintain an absolute control over the scene, the successive images of the same object, in the same positions towards the sensor, may give birth to variations which after the segmentation process, generate different representations.

There are two kinds of variations of the representations:

- node and arc attributes variations;
- structural variations, caused by the appearance or disappearance of some nodes and arcs.

To obtain a robust and effective recognition method, the inexact matching is introduced.

Inexact matching allows the pairing of a model feature with more sensorial features or fragments of sensorial features.

The major problem of inexact matching is the combinational explosion that it generates, during the pairing process.

To obtain effective inexact matching methods it is necessary to intensively use techniques for search space reduction.

The search space dimension is determined by:

- number of model and sensorial features;
- the order of feature considering;
- search focus;
- constraints using;
- the use of heuristics for search guiding, abandoning or ending.

2. THE INEXACT MATCHING PROBLEM

Let it be: $S = \{s_i \mid i = 1,...,s\}$ a set of sensorial features and $M = \{m_i \mid i = 1,..., m\}$ a set of model features characterised by model constraints like: $C_M(m_1, m_2,..., m_k) = Z$.

An inexact matching IM is a set of pairing between the model features and combinations of the sensorial features:

$$IM = \{(m_{i_k}, S_{j_k})\} \tag{1}$$

where:

$$S_{j_k} = \{s_r \vee P(s_r) | s_r \in S \text{ and } P(s_r) \text{ is part of } s_r\} \tag{2}$$

An inexact matching is consistent if the ensemble of the paired sensorial data may correspond to the perception of the model features.

Let us consider the following consistence levels:

Local consistency: The inexact matching IM is local consistent if and only if:

$$\forall (m_{i_1}, S_{j_1}),...,(m_{i_k}, S_{j_k}) \in IM \tag{3}$$

$$C_M(m_{i_1},..., m_{i_k}) = Z \Rightarrow C_S(S_{j_1},..., S_{j_k}) = Z \tag{4}$$

where C_M is a model constraint and C_S a sensorial one.

Global consistency: Is the consistency at the object level. The inexact matching IM is global consistent if and only if:

$$\exists \text{ Tr}, \ \forall (m_{i_k}, S_{j_k}) \in IM, \ S_{j_k} \cong Tr(m_{i_k}) \tag{5}$$

where Tr is a rigid transformation.

3. THE METHOD

It is presented a robust and effective method for bidimensional recognition of 2D or 3D objects, from intensity images. The method is based on:

- matching the models with symbolic structures of the scene;
- using of the inexact matching;
- intensive using of techniques for search space reduction.

3.1. The scene and models representation

For scene and models representation an intermediate representation is used as defined in [4] and characterised by:

- local features;
- the representation primitive is the straight line segment;
- 2D index features associated to 3D forms and which correspond to projection invariant properties of 3D forms;
- the association of localisation and identification attributes with each feature;
- the association of significance factor with each feature;
- hierarchical nature of the representation.

The aggregation of the elementary features belonging to the same physical entity in index features, generates compounded features with greater discrimination and indexing power.

The intermediate representation can be viewed as an explicit relational attributed graph in which the straight line segment elementary features represent the nodes and the compound features expressing relational constraints between elementary features represent the arcs.

The hierarchical nature of the representation allows the optimal hierarchical level selection to realise indexing or matching.

The signification factors allow the selection of the privileged features and the features ordering according to their importance.

3.2. Model indexing

Sequential extraction of the models from model library to realise a matching with sensorial features is acceptable in the case of small number 2D object libraries. Considering a great number of 2D objects or the 3D object library, the former being specified by projections corresponding to their different aspects, it is necessary to use indexing techniques[2],[3].

The index features defined in the intermediate representation are used in the implementation of indexing. It is used a technique based on model level accumulation of the votes coming out after a successful indexing of the models by the index features from the scene. The vote number for each model will be normalised.

The models will be sorted depending on to the accumulated vote number. In the recognition process first the models with greater vote number will be used.

3.3. The realisation of inexact matching

To realise the inexact matching it is used the hypothesis generation and verification method. This method allows the focus of the search at the model, feature and algorithm level[1].

The focus at the model level consists of the selection of a set of sensorial features corresponding to the virtual appearance of the model in the scene. It is realised for each generated hypothesis.

Hypothesis generation is realised by aligning the model features with the same kind of sensorial features having similar identification attributes and determining the associated rigid transformations. To get smaller generated hypothesis number, and the hypothesis to be as probable as possible, it is necessary to use some privileged features with high discrimination power. In this respect, the index features, from the adopted intermediate representation, showing relations between elementary features, have a higher discrimination power than the elementary features. At the same time, for the reducing of the hypothesis number, the feature pairs used have to insure only one aligning possibility. We call a topological claim relation, a relation, from a model represented by an attributed relational graph, for which having given its positioning attributes we can specify the positioning attributes of all nodes.

Among the topological claim relation there is the angle relation. The parallelogram index feature has a higher discrimination power but does not have topological claims. To use the discrimination power of parallelogram feature and other of the same type it is incremented the signification factor of the included angle features.

The hypothesis generation will be realised by aligning the angle type features.

The evaluation of the hypothesis is done by predicting the zones in the scene where the virtual model features appear: $m_i^*=T(m_i)$.

This virtual segment is used to realise a new focus at the feature level.

Only the segments intersecting or contented into a rectangle, which has as a median line the m_i^* segment, and has an orientation close to the one of the m_i^* segment are selected from the sensorial segment set associated to the hypothesis.

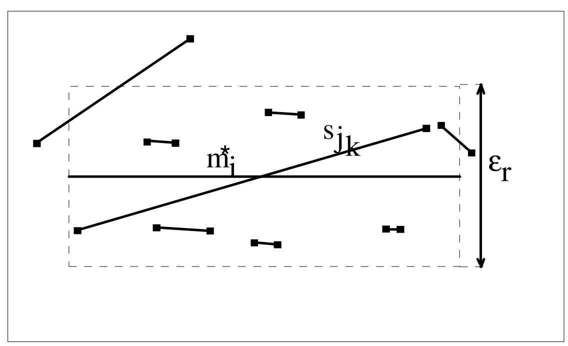

Fig. 1 Focus at the feature level.

A 1:1 pairing is tried. In the case of edge line features, the likeness factor of two features m_i and s_{ij} can be calculated as it follows[1]:

$$\sigma = \begin{cases} 1 - a \cdot \dfrac{\Delta M}{\Delta M_{MAX}} - b \cdot \dfrac{\Delta \theta}{\Delta \theta_{MAX}} - a \cdot \dfrac{\Delta L}{\Delta L_{MAX}}, \text{ if } \begin{cases} \Delta M < \Delta M_{MAX} \\ \Delta \theta < \Delta \theta_{MAX} \\ \Delta L < \Delta L_{MAX} \end{cases} \\ 0, \text{ else} \end{cases}$$

(6)

where:

a, b, c are the positive weighting factors, so that a+b+c=1

ΔM is the distance between the middle of m_i^* and s_{ji} segments;

$\Delta\theta$ Is the absolute value of the difference of orientations:

$$\Delta\theta = \left|\theta(s_{ji}) - \theta(m_i^*)\right| \tag{7}$$

ΔL is the absolute value of the m_i^* and s_{ji} length difference, relatively to the length of the m_i^* segment:

$$\Delta L = \frac{\left|l(m_i^*) - l(s_i)\right|}{l(m_i^*)} \tag{8}$$

The likeness factor has the maximum value of 1 when m_i^* and s_{ji} segments covers each other exactly and decreases when ΔM, $\Delta\theta$, and ΔL grow.

If a 1:1 matching is not possible one has to try an inexact matching.

For each of the sensorial segments associated with the virtual appearance of the model segment on the scene it is calculated the d_m distance between their middle and the support line of the m_i^* segment. Only the segments with the distance $d_m<D_{max}$ having the same gradient orientation are taken. Each of these segments will contribute to the matching in a ratio which is proportional to the length of its projection on the m_i^* segment and an inverse ratio to the d_m.distance. So, the contribution of a s_{jk} segment at the pairing with m_i^* segment will be :

$$\frac{1}{l(m_i^*)} \cdot l(pr_{m_i^*}(s_{jk})) \cdot \frac{1}{1+\frac{d_m}{D_{max}}} \tag{9}$$

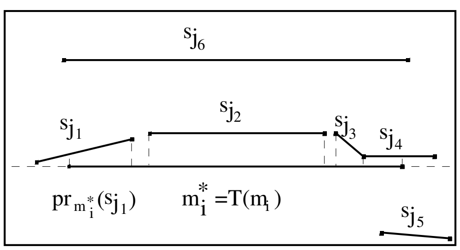

Fig. 2 The inexact matching at the feature level.

The likeness factor between the m_i^* segment and the fragments of sensorial segments which satisfy the above conditions are calculated as it follows:

$$\sigma = \begin{cases} \frac{1}{l(m_i^*)} \cdot \sum_j l(pr_{m_i^*}(s_{jk})) \cdot \frac{1}{1+\frac{d_{mj}}{D_{max}}}, \text{if } \begin{cases} d_{mj} < D_{max} \\ \Delta\theta < \Delta\theta_{max} \end{cases} \\ 0, \text{else} \end{cases} \tag{10}$$

The likeness factor is used to update the quality factor and the covering factor. The quality factor can be calculated with the iterative formula:

$$Fc_i = Fc_{i-1} + \sigma(m_i^*, s_{ji}) \cdot \frac{l(m_i)}{p} \tag{11}$$

with $Fc_0=0$, where p is the model's perimeter.

The covering factor (Fa) is calculated every time when the correspondent of a model segment is covered in the scene:

$$Fa_i = Fa_{i-1} + \frac{l(m_i)}{p}, \tag{12}$$

with $Fa_0=0$, where:

Fa_i corresponds to i covered segments;

p is the model's perimeter.

The quality and covering factors are used for heuristic guiding, abandoning or terminating of the search.

The 1:1 pairings are used for iterative refinement of the transformation as show in [1].

4. EXPERIMENTAL RESULTS

The above presented method was tested on a set of industrial objects in the following conditions: unproper lighting; noises; touching and partial overlapping of the objects.

In fig. 3 and 4 there are presented the results of the recognition of a polygonal object using the presented method, and the HYPER method respectively[1].

The use of the inexact matching, determines an improvement of the recognition accuracy. So the quality factor obtained through the presented method is 0.882 instead of 0.689 with the HYPER method.

The tables 1 and 2 reveal the realised matching through the two methods. For every pairing the identifier and the length of the model feature, the identifier and the length of the sensorial feature, the likeness factor, the weight of the paired model segments and the quality factor of the partial matching are presented. The better result of the presented method is determined by the inexact matching. So the model feature 4 is paired with the sensorial features 14 and 13 contributing with a weight of 0.14 at the quality factor, if the model segment has a weight of 0.15 in the model's perimeter. The model feature 5 it is paired with a fragment from the sensorial feature 16 contributing with a 0.06 weight at the quality factor if the model segment has a weight of 0.11. In the HYPER method the segments 4 and 5 of the model are not found in scene, determining a decrease of the quality factor.

The figures 5 to 7 show the recognition of some objects in noisy scenes with partial overlapping. In all these situations the quality factor of the recognition is close to the weight of the visible perimeter of the objects.

In fig. 8 the recognition of a cub with hard 3D features from a noisy occluded image is presented.

5. CONCLUSIONS

This paper presents a robust and effective method for bidimensional recognition of 2D or 3D objects, from intensity images based on:

- matching of the models with symbolic structures of the scene;
- using of the inexact matching;

- intensive using of techniques for search space reduction.

The inexact matching problem is presented and it is justified it's necessity. The major problem of inexact matching is the combinational explosion which it generates.

To realise the inexact matching the hypothesis generation and verification method is used. This method allows the focus of the search at the model, feature and algorithm level.

A formula for likeness factor computing in inexact pairing is presented.

The recognition accuracy is tested on scenes with noise and occlusions. The obtained accuracy is higher than the one obtained by HYPER methods, and this result is obtained due to using of inexact matching.

References

[1] N. Ayache, O.D. Faugeras, HYPER: A New Approach for the Recognition and Positioning of Two-Dimensional Objects, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 1, January 1986, pp. 44-54.

[2] W.E.L. Grimson, Object Recognition by Computer: The Role of Geometric Constraints, *MIT Press*, Cambridge, Massachusetts, 1990.

[3] W.I. Grosky, R. Mehrotra, Index-Based Object Recognition in Pictorial Data Management, *Computer Vision, Graphics, and Image Processing*, 52, 1990, pp. 416-436.

[4] S. Nedevschi, C. Goina, Intermediate Representation for 3D Model Based Recognition from Intensity Image, *ACAM*, Vol. 2, No. 1, 1993, pp. 26-33.

[5] A.M. Wallace, A Comparison of Approaches To High-Level Image Interpretation, *Pattern Recognition*, Vol.21, No.3, 1988, pp 241-259

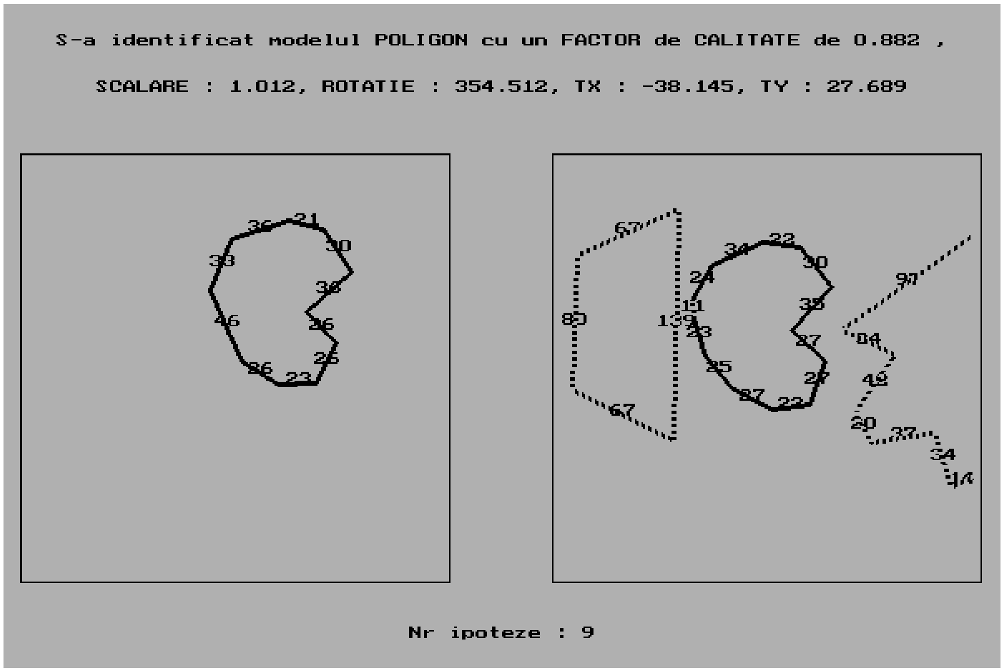

Fig. 3 Polygonal object recognition using inexact matching.

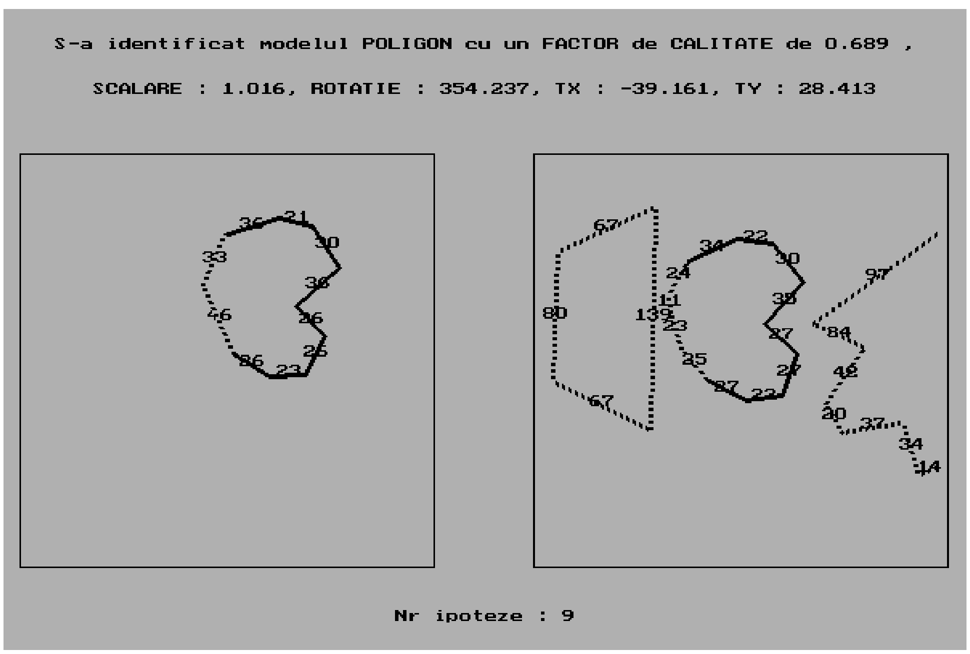

Fig. 4 Polygonal object recognition using HYPER method.

Model segment identif.	Model segment length	Scene segment identif.	Scene segment length	Likeness factor	Model segments weight	Quality factor (Fc)
4	46.60	14	23.60	0.45		
4	46.60	13	25.50	0.45	0.15	0.14
9	36.20	8	35.80	0.95	0.27	0.25
6	36.00	17	34.00	0.93	0.39	0.36
5	33.20	16	24.20	0.54	0.50	0.42
8	30.80	7	30.00	0.96	0.60	0.51
1	26.70	10	27.60	0.86	0.68	0.59
10	26.40	9	27.60	0.89	0.77	0.66
3	26.10	12	27.10	0.92	0.85	0.74
2	23.00	11	23.20	0.95	0.93	0.81
7	21.90	6	22.30	0.96	1.00	0.88

Table 1 Pairing list for inexact matching (object polygon).

Model segment identif.	Model segment length	Scene segment identif.	Scene segment length	Likeness factor	Model segments weight	Quality factor (Fc)
4	46.60	*	*	0	0.15	0
9	36.20	8	35.80	0.95	0.27	0.11
6	36.00	17	34.00	0.93	0.39	0.22
5	33.20	*	*	0	0.50	0
8	30.80	7	30.80	0.96	0.60	0.32
1	26.70	10	27.60	0.86	0.68	0.39
10	26.40	6	27.60	0.89	0.77	0.47
3	26.10	12	27.10	0.92	0.85	0.55
2	23.00	11	23.20	0.96	0.93	0.62
7	21.90	6	22.30	0.97	1.00	0.69

Table 2 Pairing list for exact matching (object polygon).

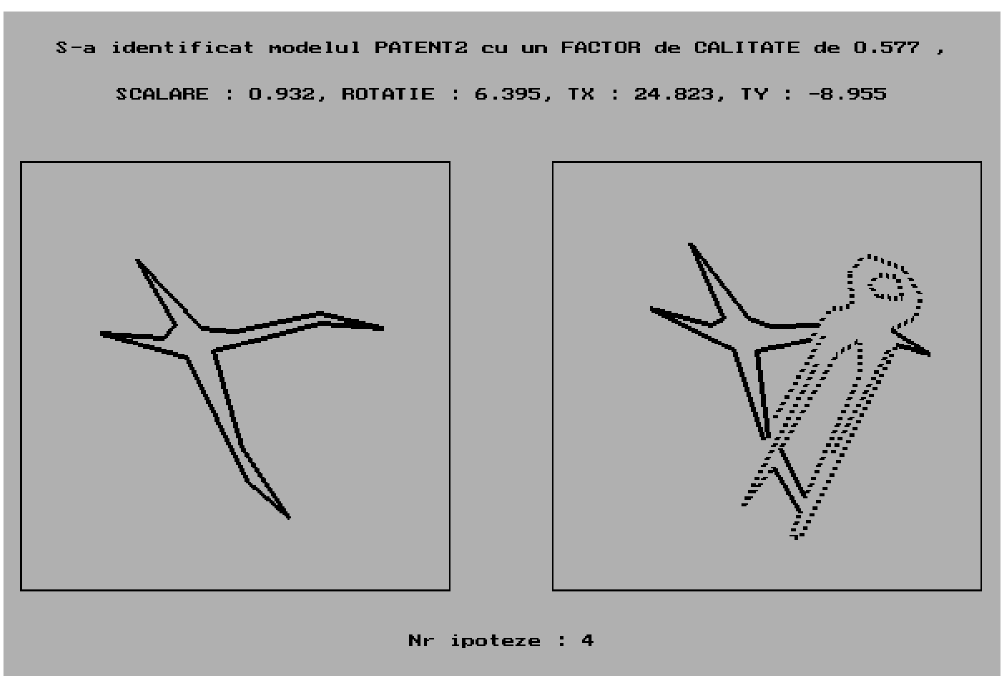

Fig. 5 "patent2" object recognition from a noisy scene with occlusions.

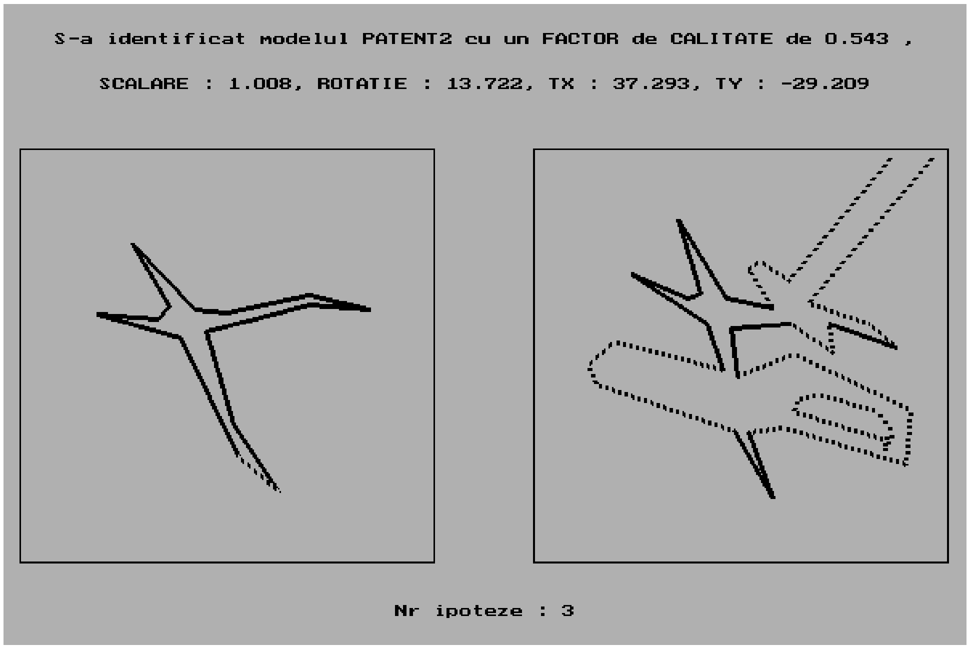

Fig. 6 "patent2" object recognition from a noisy scene with occlusions.

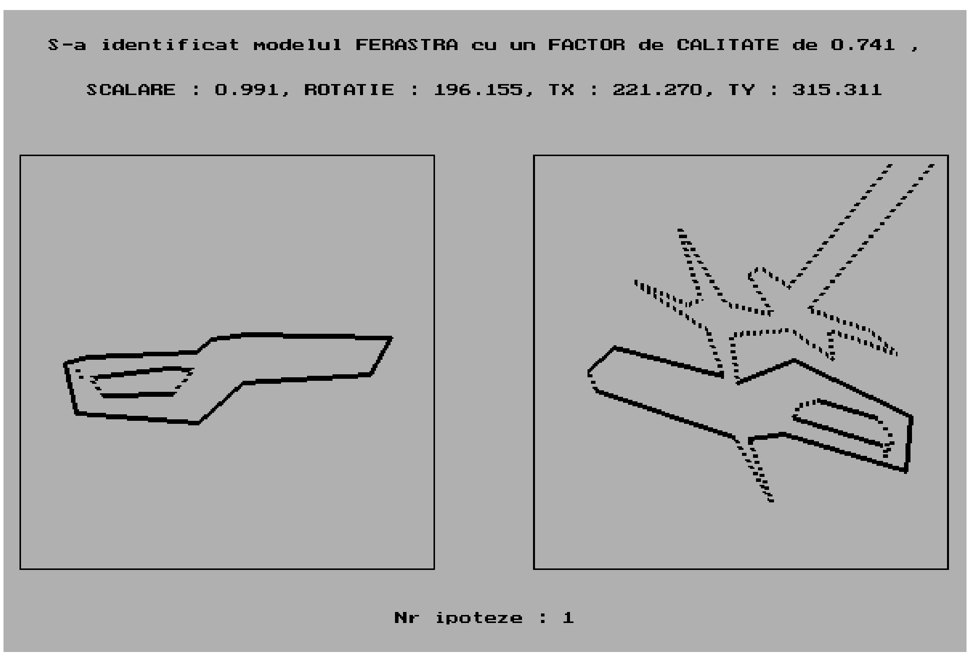

Fig. 7 "ferastrau" object recognition from a noisy scene with occlusions.

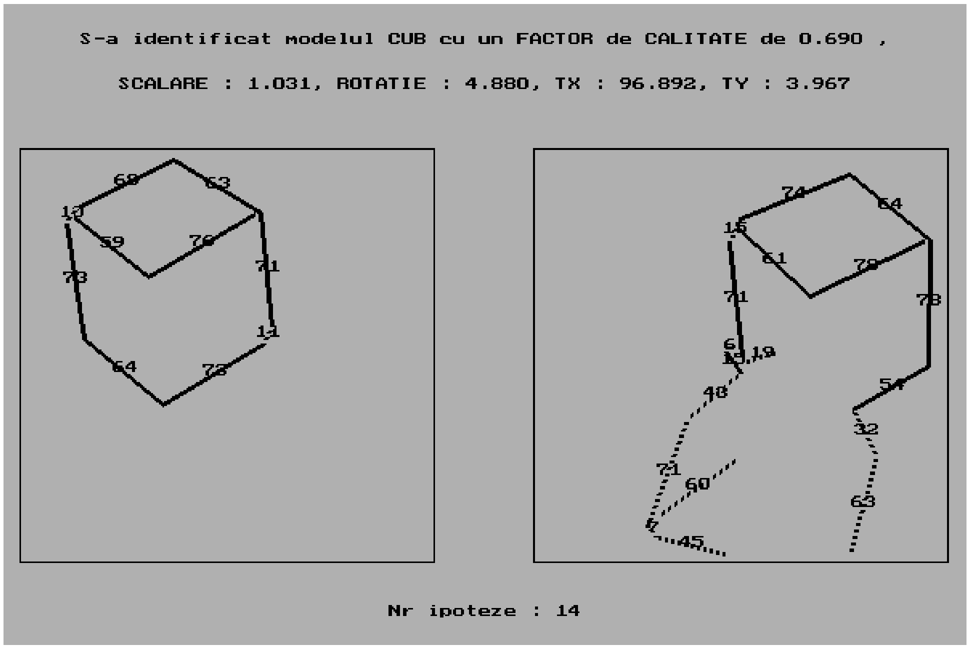

Fig. 8 "cub" object recognition from a noisy scene with occlusions.