# The use of ANN and 3-point matching for objects recognition and localization in occlusion environment

H. Senoussi, H. Temimi and B. Belmekki

University of Science and technology of Oran U.S.T.O, Electrical Engineering Faculty.P.o. Box 1505, El M'naouer, Oran, Algeria. E\_mail: senoussih@yahoo.com, Fax: 041 42 15 81.

Abstract— Applications associating the robotics and the vision are often confronted to problems of objects identification and localization in a multiform scene. Recognition of partially occluded objects has been recently made possible by describing an object boundary with a sequence of local feature segments localized with control or dominant points, and then matching the representation with those of set of known reference shapes. Most the works of localization of objects had been based on the structural approach to make the recognition and the localization at the same time by using a segmented representation of the shape. In this work, the aim is to combine two approaches (structural and global) to make the localization of occluded objects. The global approach using neural networks is adopted to make the selection of the feature. The structural approach is used to make localization in the scene. In essence, Curvature Guided Polygonal approximation is employed for detectiong the dominant points of the boundaries. A 3-point matching is developed for detecting the dominant points of known boundary selected by neural networks which matched against reference objects. The recognition phase is very important because it permits of to reduce the number of the feature to localize. One uses the ANN to inform us to priori of shapes that are present in the scene (the feature that are going to be used thereafter). Then the system of localization goes us to inform a posteriori on shapes presents in the scene. The scheme has been successfully applied for localizing multiple overlapped handtools in different sizes and orientations.

Keywords— Pattern recognition, neural network, AR prediction, Fourier descriptors, principal component analysis, points dominants

#### I. INTRODUCTION

Recognition of partially occluded objects has been an important issus in the fields of industrial automation applications because occlusion cause significant problems when one attemps to identify and locate an object in the workspace of robots, baggage inspection in airports, etc. Occlusion occurs when two or more objects in a given image touch or overlap with one another. In such situations vision techniques using global features to identify and locate an object may fail because descriptors of a part of a shape may not have any resemblance to the descriptors of the entire shape. Typical representations employed in global techniques include Shapes Numbers, Moment, Fouriers descriptors, Regional descriptors and Hough Transformation [16]. To solve this problem a lot of methods developed (structural Methods). The more part of these methods requires the previous creation of models of objects. It probably explains the fact that the proposed solutions are generally

limited to objects polygonal [5] [17]. The polygonal models drive to manipulate segments, whereas some more general models require the management of complex features.

M. H. Han [15] used a conceptually simple technique to solve the occlusion problem by following the order of the matched vertices in the model and the input image. The algorithm used vertices resulting from local maximum curvature approximation. His method then compared every vertice in the model with vertices in the input image using a graph matching. Aedt fter all line vertices had been compared, the graph contained the all compatible nodes. This method is very simple but exhibits a lack of tolerance. Therfore, the information based only on locally focused features may fail to match because there is no relational information from features of remote nodes.

In this paper a new scheme capable of combining the global method by using neural networks as a graph matching technique and a structural method by a 3-point matching technique developed for locating overlapped objects in the input image. This is achieved with the use of dominant points on the boundaries which are extracted with Curvature Guided Polygonal Approximation. Each dominant point on the selected contour by the neural networks is tested with the 3-point matching technique to locate its counterpart in the reference object, forming clusters of contiguous matched point pairs.

The inherent parallelism of neural networks allows rapid pursuit of many hypotheses in parallel with high computation rate [12] [16]. Moreover, it provides a great degree of robustness or fault tolerance compared to conventional computers because of many processing nodes, each of which is responsible for a small portion of the task.

Information that must be presented to the neural network must contain the minimum possible information in order to reduce the time of treatment of the calculator. For that raison a compression of data is necessary. In the case of the present work we used three methods (the linear prediction, the Principal Components Analysis and Fouriers descriptors to put in evidence there influences on the first hand in the length of the training and in the other hand on the recognition. A detailed description on each of the processes is explicated in the following sections.

## II. Shape recognition by neural networks

In this work, we proposes a new method based on the utilization of neural networks. The idea consists of learning to the network a certain number of situations of every isolated object. Finally training should be capable :

- to recognize f shapes not belonging to the set of training.
- to discriminate between superimposed two shapes.

One arranges a basis of examples of freestanding real shapes (hammer, key of pitchfork, gun). shapes belong to three different classes. Every prototype of the example basis exists in different angles of rotation to satisfy the condition of spatial invariance of the classifier.

## .1 Compression by AR prediction

The linear prediction of n order is a particular case of the general problem of the linear medullization in the least square sense [1] [4]. Equation (1) represents an example of auto-regressive linear filter (AR: filter all-pole) of n order, transforming the following of inputs  $\{u_t; t = 0, 1, ...\}$  as the following of exits  $\{y_t; t = 0, 1, ...\}$  according to the rule:

$$y_t + \sum_{i=1}^n a_i^{(n)} y_{t-i} = u_t$$
,  $t = 0, 1, \cdots$  (1)

where  $u_t$  is a white noise of variance  $\sigma_n$  and  $a_i$  coefficients.

For every vector  $y_t$  corresponds a vector valued  $y_t^*$  or corresponds some  $a_i$  coefficients more precisely. For our application  $y_t$  represent the matrix picture transformed in vector of dimension  $(1 \times M^2)$ , and instead of the image is represented by this vector it will only be represented by a vector of dimension  $(1 \times n)$  whose composantes are coefficients of the predictor filter (coefficients of prediction).

#### .2 Principal Components Analysis (PCA)

Although the rate of compression raised by using the linear prediction but it tend to make correlation between vectors which have initially minimum correlation. The utilization of these coefficients will put a problem at the time of the utilization of neuronal classifier. In the goal to decorrelate vectors of AR descripteurs we applied at last the Principal Components Analysis[2][3]. The Fig 1 illustrates the effect of the PCA on vectors of data.

In addition the PCA has a character of data compression.

#### A. Recognition of occluded objects

Tests of recognition have been operated in the first phase on pictures of isolated shapes, to fix the minimal number of training prototypes. In second phase, tests of recognition have been operated on the overlaped shapes.

Knowing that prototypes belong to three classes, all networks that one is going to use have three neurons in the layer of exit. The first neuron is sensible to the key to pitchfork, the second is sensible to the hammer and the third neuron is sensible to the gun. The different tests that we are going to present, will be achieved while using the compression by:

• AR prediction,



Fig. 1. Correlation histogram for: (a) AR coefficients; (b) AR + ACP coefficients.

# • AR+PCA.

We used a multilayer perceptron (MLP), driven by the algorithm of backpropagation of the gradient improven (Adaptation of the training speed, utilization of the moment, the initialization of matrix weight by the Nguyen and Widrow formalism, cross-correlation function [11][10]).

## III. OBJECTS LOCALISATION USING 3-POINT MATCHING

After extracting the outermost object boundary, Curvature Guided Polygonal Approximation is applied to construct a representation of the contour. The representation is in the form of a sequence of dominant points joined by straight lines to form a polygon which best approximates the original boundary.

#### A. Localization of dominant points

The algorithm can be divided into two sections. In the first part, the positions on the boundary corresponding to the extrema of the smoothed curvature function are taken as the initial breakpoints. In the second part of the algorithm dominant points in the object contour are detected from the initial breakpoints through a series of iterative operations.

Let

$$P^{(m)} = \left\{ B_1^{(m)}, B_2^{(m)}, ..., B_i^{(m)}, ..., B_N^{(m)} \right\}$$

be the sequence of breakpoints after the  $m^{th}$  iteration with N(m) elements and  $B_{N(m)+1} = B_1$ .  $B_i$  and  $B_{i+1}$  are successive break points  $\forall B_i \in P^{(m)}$ .  $S_{ij}$  is the  $j^{th}$  point on a segment bounded by  $B_i$  and  $B_{i+1}$  and  $D_{ij}$  is the perpendicular distance from  $S_{ij}$  to the straight line  $\overline{B_i B_{i+1}}$  joining the two breakpoints.  $T_{ij}$  is the  $j^{th}$  point on a segment bounded by  $B_i$  and  $B_{i+2}$  and  $E_{ij}$  is the perpendicular distance from  $T_{ij}$  to the straight line  $\overline{B_i B_{i+2}}$  joining the two breakpointset (figure 2).



Fig. 2. A point dominant insertion or supression

1. m=0. start with initial set of breakpoints  $P_0$ .

2. i = 1. start with first boundary segment joining  $B_1$  and  $B_2$ .

3. Determine  $S_{ij}$  such that  $D_{ij} = \max \{D_{ij}, D_{i2}, ..., D_{ij}\}$ .  $S_p = S_{ij}$ .

4. if  $D_{ij} >$  thershold, then  $P_{m+1} = P_m \cup \{S_p\}$ .

5. i = i+1. progress analysis to the next adjacent segment. 6. repeat (3)-(6) if  $i < N^{(m)} + 1$ 

7. i = 1.start with first boundary segment joining  $B_1$  and  $B_3$ .

8. Determine  $T_{ij}$  such that  $E_{ij} = \max \{E_{ij}, E_{i2}, \dots, E_{ij}\} \cdot T_p = T_{ij}$ .

9. if  $E_{ij} < \text{thershold}$ , then  $P_{m+1} = P_m \setminus \{T_p\}$ .

10. i = i+1 progress analysis to the next adjacent segment. 11. repeat (8)-(11) if  $i < N^{(m)}$ .

12. if  $P_m \neq P_{m+1}$  then m = m + 1. repeat (2)-(12).

13. if  $P_m = P_{m+1}$  then  $P_{m+1}$  is the set of domonant points for the object boundary.

#### B. The matching

The domonant points extracted from the scene and reference boundaries are recorded in the sequences

$$P^m = \{B_1^m, B_2^m, \dots, B_N^m\}$$

and

$$P^{s} = \{B_{1}^{s}, B_{2}^{s}, \dots, B_{M}^{s}\}$$

respectively. For clarity of explanation, the matching algorithm is described with the aid of the boundary of the scene in fig 4 and the reference contour in fig 3. The dominant points on the two boundaies are located and labled with an identity number. the following terminology is adopted in the analysis.

• The length  $||L(B_i, B_j)||$  is defined as the distance between two points  $B_i$  et  $B_j$ .

• The length ratio of  $||L(B_i, B_j)||$  and  $||L(B_j, B_k)||$  is defined as.

$$R(B_i, B_j, B_k) = \frac{\|L(B_i, B_j)\|}{\|L(B_j, B_k)\|}$$
(2)

• The interior angle  $\theta$   $(B_i, B_j, B_k)$  is defined as

$$\theta\left(B_{i}, B_{j}, B_{k}\right) = \arccos \frac{\left\langle L\left(B_{i}, B_{j}\right), L\left(B_{j}, B_{k}\right)\right\rangle}{\left\|L\left(B_{i}, B_{j}\right)\right\| \left\|L\left(B_{j}, B_{k}\right)\right\|} \quad (3)$$

where  $\langle A, B \rangle$  is the inner product of A and B and ||A|| is the norm of A.

Each dominant point  $B_i$  is characterized by the length ratio and the interior angle of its associated line segments formed  $B_{i-1}$  and  $B_{i+1}$ . A dominant point  $B_i^m$  in  $P^m$  is considered to be matched against a point  $B_j^s$  in  $P^s$  if the following criteria are satisfied :

$$T_r < \frac{R^m \left( B_{i-1}^m, B_i^m, B_{i+1}^m \right)}{R^s \left( B_{j-1}^s, B_j^s, B_{j+1}^s \right)} \ \Box \ \frac{1}{T_r}$$
(a)

$$\theta^m \left( B_{i-1}^m, B_i^m, B_{i+1}^m \right) - \theta^s \left( B_{j-1}^s, B_j^s, B_{j+1}^s \right) \Big| \Box T_\theta \quad (b)$$



Fig. 3. The dominant points of . (a) Reference gun. (b) Reference hammer.

Dominant points on the scene and reference object boundaries which satisfy criteria a and b are linked under a matched pair. If three consecutive matched point exist, the amoung scaling  $T_{xy}$  and angular rotation  $A_{\theta}$ necessary for aligning the three dominant points of the reference contour with their matched counterpart on the scene boundary are calculated. In the example given in Fig 3, 4 a base of matched point pairs and a set of isolated points are detected

$$Q_{ms} = \{ (1,6), (2,7), (3,8), (4,9), (5,10), (6,16), (7,17), (8,5) \}$$
(4)



Fig. 4. The dominant points of occluded objects.

## IV. Results

The first pattern set of test is done for a rate of recovery of 5%, we represented in the Fig 5 some examples of images of the recognition basis.



Fig. 5. Dominant points for patterns recognition with  $\zeta = 5\%$ .

We are going to give results of the architecture recognition that have the best behavior:

rate-rec	$\zeta{=}5\%$			
Node	Proto1	Proto2	Proto4	Proto5
S1	0.0010	0.9987	0.6814	0.9954
S2	0.7510	0.2512	0.2521	0.0045
S3	0.2121	0.0021	0.0145	0.5198

Table 1: Architecture NN(14, 10, 3)

where S1, S2, S3 states of neurons of the exit layer to the key of pitchfork, to the hammer and the gun respectively.

We have increase the rate of recovery again from 20 to 80% (figures 6 and 7) of a shape with regard to an another one in the scene and we got recognition results as in table 2, 3 and localisation in Fig 6 and 7.



Fig. 6. Dominant points for patterns recognition with  $\zeta = 20\%$ .

rate-rec	$\zeta{=}40\%$			
Neurone	Proto1	Proto2	Proto3	Proto4
S1	0.0412	0.0498	0.6742	0.8951
S2	0.1845	0.6852	0.7589	0.0174
S3	0.2899	0.1478	0.0005	0.0893

Table 2: Architecture RN(14, 30, 3)



Fig. 7. Dominant points for patterns recognition with  $\zeta = 5\%$ .

rate-rec	$\zeta{=}60\%$			
Neurone	Proto1	Proto3	Proto4	Proto6
S1	0.3251	0.4214	0.6014	0.9785
S2	0.1452	0.2851	0.1054	0.0125
S3	0.3140	0.0214	0.0154	0.0312

Table 3: Architecture NN(18,70,3)

The following table show results of the recognition gotten according to the rate of recovery :

	$ au_{loc}$		
rate-rec	Fourier	AR	AR+PCA
5%	83.3%	92.94%	95%
20%	72.3%	79.11%	84%
40%	66.6%	76.6%	79.22%
60%	50%	62%	72%

Table 4: Evaluating the performance of the ANNs withAR, AR+PCA compression and comparisons using theFourier descriptors

Of after results presented we can conclude that the rate of localization decreases progressively whenever the rate of recovery increases.

#### V. Conclusion

The originality of this work doesn't reside in the application of neural networks to the recognition of isolated shapes but there utilization for the occluded shape recognition. Otherwise, a comparative survey, can show that the time of development of a recognition system was distinctly in favor of neural networks. in this paper, several artificial neural network based approches for the recognition of 2D dimensional objects represented by translation, scal and rotation invariant coefficients representations were introduced. The compression by AR prediction+PCA gave better results then the compression by Fourier descriptors. Besides the utilization of the PCA improved results of the recognition distinctly and decreased the time of training. The polygon approximation of the boundaries of the objects play an important role in both-hypothesis localisation and verification. The scheme has been adopted to analyze scenes of multiple handtools overlapping each other. In the majority of cases, the identity of each handtool in an image is correctly determined with a similarity score reflecting its possibility of existence. The encouraging results suggested for the recognition and localization of complicated scenes of overlapped object images.

## References

- [1] Kazamarand Evariste. Performance numériques des algorithmes rapides pour les systèmes linéaires structures Toeplitz. Theses, Application en traitement du signal, septembre 1993.
- [2] K. Berradja. Analyse par composantes principales et réseaux de neurones pour la reconnaissance de formes incompletes. Mémoire de magister, Département d'Electronique USTO (1997).
- [3] J. Karhunen & J. Joutsensalo. Generalisation of principale component. IEEE Signal processing. vol 21, 552-561 (1994).
- [4] Chao. J & S. Tsujii. Estimation and synthesis of singular vector AR processes over components spaces. IEEE Signal processing (1979).
- [5] P. W. M. Tsang, P. C. Yuen and F. K. Lam. Classification of partially occluded objects using 3-point matching and distance transformation. Pattern Recognition, Vol 27 No 1, pp 27-40, 1994.
- [6] H. Henocq & G. Burel. Reconnaissance d'objets 3D par analyse de Fourier de la silhouette. Revu Technique Thomson-CSF, Vol 25 No1. Mars 1993.
- [7] H. Temimi & N. Berrached. Reconnaissance de formes planes superposées par réseaux de neurones. ICEEE 1994.

- [8] Paul J. Werbos. Backpropagation Trough Time: What It Does and How Do It. Proc of IEEE, vol. 78, No. 10, pp 1550-1560. October 1990.
- [9] Bernard Widrow and Michael A. Lehr. 30Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation. Proc of IEEE, vol.78, No.9, pp 1415-1442, September 1990.
- [10] S. Narayan, G. A. Tagliarinri & E. W. Page. Enhancing MLP networks using a distributed data representation. IEEE Trans. System Man and Cybernetics, Vol 26, 143-147 (1996).
- [11] J. Herault. Réseaux neuromimétiques. Editeur Hermes 1995.
- [12] J. S. Lee, C. H. Chen, Y. N. Sun & G. S. Tseng. Occluded objects recognition using multiscale features and hopfield neural network. *Pattern Recognition 30*, 113-122 (1997).
- [13] Peter. W. M. Tsang, P. C. Yuen & F. K. Lam. Recognition of occuded objects. Pattern Recognition 25, 1107-1117 (1992).
- [14] Stan Z. Li. Invariant representation, matching and pose determination of 3D space curves under similarity transformations. *Pattern Recognition 30, 447-458 (1997).*
- [15] M. H. Han & D. Jang, The use of maximum curvature points for the recognition of partially occluded objects, *Pattern Recognition 23, 21-33 (1990)*.
- [16] O. Yanez-Suarez and M. R. Azimir-Sadjadi. Unsupervised clustering in Hough space for identification of partially occluded objests. IEEE Trans PAMI 9, 946-950 (1999).
- [17] M. W. Koch and R. L. Kachyap. Using polygons to recognize and locate partially occluded objects. IEEE trans. Pattern analysis. Mach. Intell. PAMI-9, 483-494 (1987).