A New Edge-Grouping Algorithm for Multiple Complex Objects Localization

Yuichi Motai

Intelligent Media Laboratory
Department of Electrical and Computer Engineering
College of Engineering and Mathematics
University of Vermont
33 Colchester Avenue, Burlington, VT 05405, U.S.A.
ymotai@emba.uvm.edu
http://www.uvm.edu/~ymotai

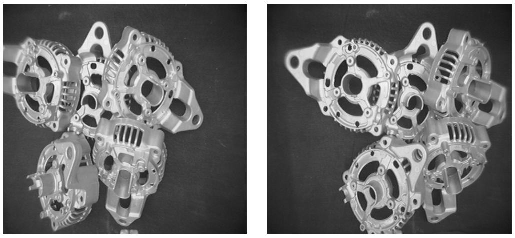
Abstract. We present a new algorithm that provides an efficient localization method of elliptic industrial objects. Our proposed feature extraction inherits edge grouping approaches. But instead of utilizing edge linkage to restore incomplete contours, we introduce criteria of feature's parameters and optimize the criteria using an extended Kalman filter. Through a new parameter estimation under a proper ellipse representation, our system successfully generates ellipse hypotheses by grouping the fragmental edges in the scene. An important advantage of using our Kalman filter approach is that a desired feature can be robustly extracted regardless of ill-condition of partial edges and outlier noises. The experiment results demonstrate a robust localization performance.

1 Introduction

Robust localization of industrial objects for assembly tasks is an issue in the robot vision community. The difficulty of feature extraction for complex objects is never ameliorated under real assembly conditions. In order to achieve reliable automatic tasks, it is essential to develop robust feature extraction and its correspondence. Typical stereo views of a workspace are shown in Fig. 1, where automobile parts, called Alternator Cover, are our target objects to be manipulated by robot hand. In this paper, we propose a new algorithm that is capable of localizing these complex industrial parts by a single camera mounted on a robot hand that uses stereo views.

Numerical model-based vision systems have been developed to estimate object pose for robotic manipulation. In such algorithms [3],[13],[17], the correspondence search between model features and scene features needs to be solved first, and the precise estimation of 3D object pose needs to be then accomplished second. In our previous approach [9], for example, the features generated by region growing are used for matching first between the object scene and the correspondent model. However lack of perfect feature extraction occasionally fails the localization, due to lighting conditions, shading, occlusion, and sensory noise.

R. Orchard et al. (Eds.): IEA/AIE 2004, LNAI 3029, pp. 1194-1203, 2004.



(a) (b)

Fig. 1. Target objects of stereo views (a) left and (b) right.

A number of studies on this problem have shown that edge grouping methods [6] [7] are promising to extract salient features. Such algorithms aim to utilize global salient structures of object contours, inspired perceptual organization. Based on a set of edges, optimal curve fitting using Kalman estimation is an important extension [4], [15].

Our feature extraction method inherits these edge grouping approaches. In order to cope with partial occluded poses of the cluttered objects as shown Fig. 1, grouping is essential to estimate salient/high-level features from local/low-level edges. Regardless of the complexity of the object, edge grouping approaches are possible [7], and several partial edges are only cues to extract the salient features in severe conditions [4]. However when a target object is too complicated, it is difficult to extract features by other extraction methods such as region growing [9]. In the automobile assembly tasks, our target shape, ellipse is a large class of curved objects in industry. Currently our focus will be limited to 3D planar elliptical features, as exemplified by the curved silhouettes on the industrial object shown in Fig. 1. The first contribution in this paper is to extract salient ellipse feature to represent this complex object class through a new edge-grouping.

To detect elliptic geometrical features, the least mean square method is frequently used to fit the ellipse boundary [1], [3], although this method is very weak for outlier data. Main alternative methods are to utilize Hough transformation [5], [14] and moments method [16], although these methods handle to extract ellipses under some limited scattered images. Our approach handles noises and outliers by the extended Kalman filter to optimize specific ellipse parameters directly. Thus our second contribution is to derive a proper ellipse representation for a Kalman estimation under the new grouping algorithm to generate ellipse hypotheses.

In the following sections, we will first present the overall strategy for our object localization system. Then our main focus in this paper, feature extraction method by grouping edge will be presented in details. Subsequently feature matching will be described. Finally, experimental results with the system's evaluation will be shown.

2 Overall Strategy of System

In our feature-based bin-picking system, first of all the object model registration is required by a human-in-the-loop with a graphical editor through image-to-image and pose-to-pose correspondences [10]. After the 3D vision model is acquired in this manner, the same objects are randomly stacked. The goal of on-line localization system described here is robust 3D pose calculation of each object. Once the feature matching between model and scene is achieved, then the 3D translation and rotation from the model coordinate to the scene coordinate is computed using quaternion approach [2]. The robot gripper can pick up the object or perform a peg-in-hole alignment through its 3D localization.

In order to complete reliable matching results, the program first extracts smooth curves for fitting large ellipse, as seed features, based on the object model. For the object shown in Fig. 1, at least half numbers of holes among extracted regions should be matched in left and right images to look for an optimal solution based on the geometric constraint equation. The system generates the hypotheses of the large ellipse candidates, and then selects other holes which will verify the estimated pose. The verification can be done by looking for supporting features of the model, for example of this object, small holes.

Among many modules in our system, one of the main focuses in this paper is feature extraction module by a robust edge grouping technique. The feature extraction essentially determines robustness and accuracy of the object localization [9]. In our formalism, a *primitive* is defined as a salient entity of features by which matching between models and scenes can be directly established. For a representation of an industrial object called Alternator Cover, ellipses are such primitives. In the scene, however, such salient primitives may not be extracted perfectly. For example, the ellipses may be broken into smooth convex curves in the scene as shown in Fig. 2. Therefore we need to prepare immediate minimum entities extracted from images, and call such low-level/local image features for *fragments*.

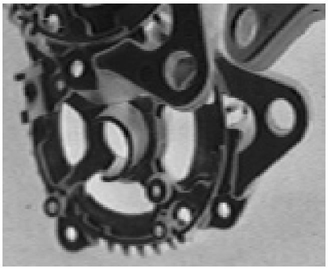


Fig. 2. False case of ellipses with broken smooth convex curves (a) a captured image (b) incomplete primitives.

Our edge grouping strategy is first to gather the fragments based on the attributes, such as size, convexity, and gray level similarity. The system then checks the elliptic curves by the number of fragments participating in forming the ellipses. For each group of fragments, the system estimates the parameters of hypothesized ellipses using iteration of Kalman filtering. In order to formalize the criteria of parameters, in the next section we will describe the representation of ellipse.

3 Representation of Ellipses

One of our contributions in this paper is a parametric representation and estimation of ellipses in the images which will be suitable for edge-based feature extraction. We reconsider this shortcoming of previous approaches [4], [15], and propose a proper parametric form for Kalman filter estimation. Although previous researchers represent ellipses by parametric equations of either

$$au^2 + buv + cv^2 + du + ev + f = 0 \tag{1}$$

or

$$\frac{(u\cos\theta - v\sin\theta - u_0)^2}{a^2} + \frac{(u\sin\theta + v\cos\theta - v_0)^2}{b^2} = 1\,. \tag{2}$$

where (u,v) be any arbitrary point on an ellipse in the image, a and b represent the lengths of the longer and shorter axes of the ellipse, θ the orientation of the ellipse, and (u_0, v_0) the center of the ellipse.

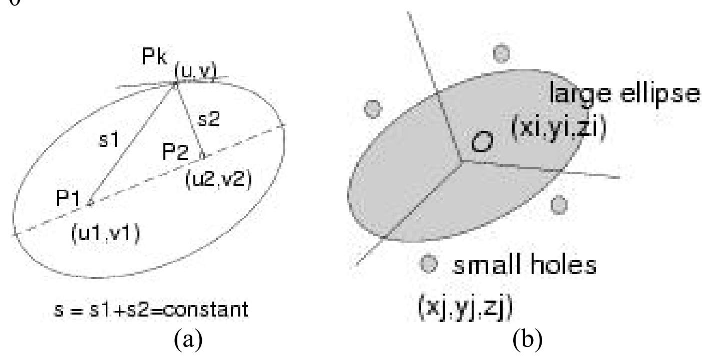


Fig. 3. (a) Ellipse model (b) Seed feature (3D ellipse) and supporting features (small holes).

The first representation (1) includes arbitrary quadratic forms other than ellipses, and therefore it is difficult to deal with geometric constraints over an ellipse by a set of parameters (a,b,c,d,e,f). Because the parameters are no explicitly bounded, - wide range of parameters - there is no guarantee to generate elliptic curve.

The second representation Eq. (2) is more intuitive, but we have observed the instability of estimating the orientation parameter θ when a and b are getting close, since the orientation becomes arbitrary as a regular circle. The parameters (a, b, u_0, v_0, θ) cause the serious problem of an instable extraction.

We have exploited a different approach to represent ellipses which will be useful to estimate all proper parameters of elliptic features extracted from the image. As shown in Fig. 3, we represent the ellipse by two focal points P_1 and P_2 and the sum s of the distances (s_1, s_2) from the two focal points to any boundary points P_k. Let (u_1, v_1) and (u_2, v_2) be image coordinates of two focal points P_1 and P_2, and (u, v) be the image coordinate of arbitrary boundary point P_k. Then

$$f \equiv \sqrt{(u-u_1)^2 + (v-v_1)^2} + \sqrt{(u-u_2)^2 + (v-v_2)^2} - s = 0 \tag{3}$$

Our contribution using Kalman filter approach includes this proper ellipse representation to derive criterion function. The parameterization in Eq. (3) is very important when we apply Kalman filtering. In this specific parametric representation, if two focal points get close, the two focal points become simply coincident; therefore no instability for the parameter estimation can be observed. So our problem is how to estimate five ellipse parameters $\boldsymbol{p} = (u_1, v_1, u_2, v_2, s)$ from primitives extracted from the images.

4 Feature Extraction: Edge Grouping

Our feature extraction starts from 2D image processing, in which the local edges are extracted as fragments, described in subsection 4.1. Subsequently the group of fragments is generated as hypothesis of ellipse, described in 4.2.

4.1 Extraction of Fragments

After Canny edge detector is applied to an image, the edges are finely sampled to local segments--called *fragments*. These sampled edges are defined by tracking edges. Fragments along curves in the 2D image scene are automatically generated by the system in the following manners:

1. Thin the edge pixels so that the edge tracking can be performed.
2. Extract endpoints and junction points in the thinned edge map.
3. Track the edge map and extract high curvature points along edge curves.
4. Divide the long smooth curves into at least two components to avoid the accidental coincidence of merged curves.
5. Register curve segments as fragments.

Note that the selection of high curvature points is done by smoothing the curve along its original form. If the deviation of the smoothed curve from the original curve is higher than some threshold and is maximal, then the system registers these points as high curvature points. As a base of low-level feature, the system decomposes the curve into smaller pieces if the curve is long enough and occupies a large angle for the ellipse formation. The decomposing is very useful for avoiding accidental coincidence, by chance, two different curves are merged due to the viewpoint ill-conditions.

4.2 Extraction of Ellipse Hypotheses

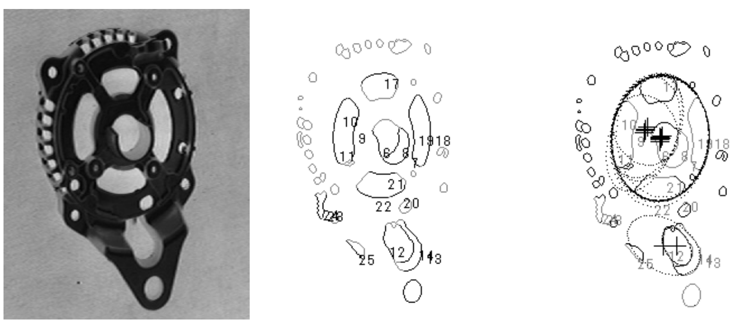
Fig. 4 illustrates the procedure to extract ellipse candidates in the scene image based on grouping fragments. The system checks the elliptic curves by the number of fragments participating in forming the ellipses. As we have discussed in the previous section, a single curve extracted from the image may not necessarily correspond to a perfect ellipse. An ellipse may be broken into several fragments. Therefore, we deal with the grouping of the fragments which potentially constitute an ellipse. The grouping of fragments is decided on the following constraints:

1. *size:* The size of the ellipse in the image is limited. For each group of fragments, the combined curves must be smaller in size than some threshold based on the object model.

2. *convexity:* Any pair of fragments must not violate the convexity when these fragments are combined.

3. *gray level similarity:* Any pair of fragments must possess gray level similarity. Either internal or external region has the similar gray level. Note that this is a typical case for an industrial object when the object is composed of parts of homogeneous color. If the object region is homogeneous along the elliptic curve, then two fragments i and j must satisfy the gray level similarity constraint:

$$\frac{|\mu_i^{Internal} - \mu_j^{Internal}|}{\sigma_i^{Internal} + \sigma_j^{Internal}} < \varepsilon \quad \text{or} \quad \frac{|\mu_i^{External} - \mu_j^{External}|}{\sigma_i^{External} + \sigma_j^{External}} < \varepsilon \tag{4}$$



(a) (b) (c)

Fig. 4. Edge grouping sequential process (a) Object scene, (b) Curve fragments extraction along edges, (c) Group of fragments extracted by aggregation through elliptical parameter estimation.

In our current implementation, the system generates an ellipse hypothesis based on how many fragments are chosen for grouping, from a single fragment to four fragments. The system first estimates initial parameter of each ellipse and then updates for verifying that ellipse. For each group of fragments for an ellipse candidate, the system verifies whether or not these fragments certainly constitute an ellipse in terms of the parameters. More specific procedure is as follows:

- **Generation of Initial Parameter Estimation**

Given a set of points along group of the fragment curves, $\boldsymbol{p} = (u_1, v_1, u_2, v_2, s)$ is to be estimated. First of all, the system generates an initial estimate of $\boldsymbol{p}$, and then applies the Kalman filter to update the parameter $\boldsymbol{p}$ [8]. In order to compute the initial estimate of $\boldsymbol{p}$, we first compute the centroids and the moment of inertia for the image points participating in the fragment set. The initial estimates of (u_1, v_1) and (u_2, v_2) are computed as the above centroids. The initial estimate of s is computed by the sum of the lengths of two axes spanned by the moment of inertia. We also associate the covariance matrices for these parameters. The covariance matrices are assigned on the basis of the experiments.

- **Verification of Ellipse Formation**

After the system obtains the initial estimate of the ellipse parameter $\boldsymbol{p}$, the system selects representative points from the fragments. This is done by equally selecting points along the boundary curves. In our current implementation, the system selects at least 16 points for each fragment. By applying the Kalman filter to the constraint equation of Eq. (3) for every selected boundary point (u, v), the system updates the ellipse parameter $\boldsymbol{p}$.

5 Feature Matching

We utilize an object model, consisting of ellipse (seed feature) and small holes (supporting feature) illustrated in Fig. 3 (b). In the first step, the 2D results of edge grouping extraction (described in Section 4.2) are reconstructed in 3D. Hypothesis generation of objects based on seed features (Large Ellipses) are used for matching as following procedures:

i. Given groups of fragments, generate hypothesis that optimally fits to the each hole by computing elliptic parameters.

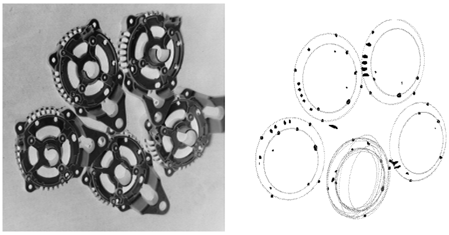
ii. For each hypothesized ellipse in the left, look for an ellipse in the right image which will correspond to the left one by considering the epipolar constraint.

iii. By epipolar constraint, estimate the seed feature position in the 3D space. Check whether or not this feature will support the estimated pose of the object.

iv. Apply attribute constraints of the correspondent model, such as area, circularity, shape complexity, perimeter, and average gray level with deviation, many hypothesized regions are pruned out.

As you expect, there exist several mismatches of generated hypothesis, which should be removed by verification using supporting features. In the second step, hypothesis verification of objects is used for supporting features (Small Holes).

Two fine-line ellipses in Fig. 5 (b) represent where the small holes should be given the estimated poses. These supporting features are extracted by a different extraction method, such as region-based method called split and merge segmentation process [9], because this method is just useful to extract small holes as you see Fig. 5(b). The second steps are described more details as follows:



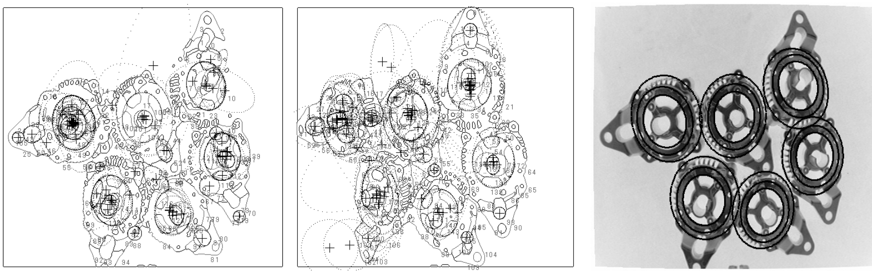
(a) (b)

Fig. 5. (a) 2D captured view (b) Small holes extracted along hypothesized regions.

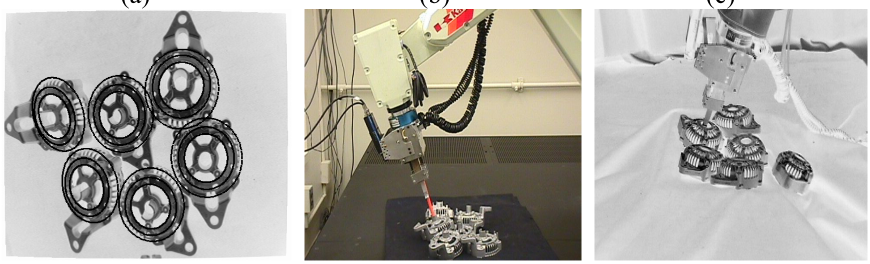
i. Based on the 3D object pose hypothesis, generate constraint regions of 3D supporting features where the supporting features should be in the left and right images, by projecting the 3D supporting features onto the 2D images with a given pose associated with the hypothesis.
ii. Select 2D features associated with 3D supporting features in the left and right images within the constraint regions.
iii. Estimate the 3D positions of 3D supporting features based on the stereo correspondences of 2D features obtained in previous step **ii**.
iv. Verify the hypothesis by considering the geometric constraints of 3D supporting features: These 3D supporting features should be compatible with the hypothesized 3D object pose given in step **i**, as well as the 3D supporting features satisfy the geometric constraints, e.g., distance between 3D supporting features, orientation between 3D supporting features. Also the number of supporting features should exceed a user specified threshold.
v. Find an optimal solution, if multiple solutions exist within a certain portion of the workspace. (This may happen due to the edge grouping). For each solution, we approximate the object space occupancy. If multiple solutions share the space occupancy, then select the optimal solution from such shared solution sets. The optimality is based on the geometric constraints of the 3D seed feature and 3D supporting features, which is associated with the fitting error.

6 Experimental Results

We mounted a monocular camera on the robotic manipulation gripper to generate a 3D object model by capturing multiple images of the object from different viewpoints. Our wrist-mounted robotic vision systems consisted of a Sony *DC-47* monocular *1/3 inch* CCD camera with Pulnix Lens of focal-length *16 mm*, a Kawasaki *JS-10* and a PC. We used only two views to localize the object for automatic localization. Curve shaped *4 - 6* pieces objects, automobile industrial Alternator Cover, were used for automatic localization experiments, *35* times, total *144* pieces. The typical 3D translation errors were less than *3 mm* and rotation errors were less than *5 degree*. The outline diameter was *65 mm* for Alternator Cover of symmetric circle outline. These localization results were verified through the robotic manipulation. For example, the error of *4 mm* and *8 degree* was within the tolerance range for our robot manipulation of that Alternator Cover. Since we already calibrated all the robotic coordinate transformations [11], the robot hand can manipulate the localized object shown in Fig. 6 (e) (f). Fig. 6 (a) (b) shows typical results of the edge grouping of left and right 2D image, and (c) (d) shows the 3D model projected onto the 2D object scene, -- bold ellipses represent the estimated poses from all candidates, which seems to graphically be lined up well.



(a) (b) (c)



(d) (e) (f)

Fig. 6. (a) A typical experimental result of Alternator Cover localization (a) left view of edge grouping, (b) right view of edge grouping, (c) left view of 2D superimposition of model onto scene, (d) right view of 2D superimposition of model onto scene, (e) robot manipulation demonstration from far view, (f) robot manipulation demonstration from close view.

We also evaluated the success rates of the system into the following categories:

- successful localization case: localization completed with a success robotic manipulation (within tolerance range) [82.6%]
- inaccurate localization case: localization completed with false manipulation (outside tolerance range) [1.4%]
- incomplete localization case: no localization output [16.0%]

7 Conclusions

We developed a new vision-based localization system for 3D elliptic industrial objects. Our salient feature extraction from complex objects extended traditional edge grouping approaches. More specifically, our contributions of this system were (1) to establish a competent edge-grouping method to generate ellipse hypotheses in a complex object and

(2) to derive an efficient ellipse representation for Kalman estimation. Using each group of edge fragments, the system estimated the parameters of ellipse hypotheses using an extended Kalman filtering. The advantage of Kalman estimation was that a desired feature could be robustly extracted regardless of ill-condition of partial occlusions and outlier noises. For optimizing the criterion, we introduced a proper parametric representation of an ellipse feature to achieve a stable result. The evaluation experiments verified that our feature extraction and its matching method were robust for an object manipulation.

References

1. Chatterjee, C. and Chong: E.K.P., Efficient algorithms for finding the centers of conics and quadrics in noisy data, Pattern Recognition, vol. 30, (1997) 673-684
2. Faugeras, O.D.: Three-Dimensional Computer Vision, MIT Press, (1993)
3. Fitzgibbon, A, Pilu, M. and Fisher, R. B.: Direct Least Square Fitting of Ellipse, IEEE Trans. of Pattern Analysis and Machine Intelligence, Vol. 21, (1992) 476-480
4. Guichard, F. and Tarel, J: Curve finder combining perceptual grouping and a Kalman like fitting, Proc. of IEEE International Conf. on Computer Vision, Vol. 2, (1997) 1003-1009
5. Hough, P.V.C.: Method and means for recognizing complex patterns, US Patent, No. 3 069 654, (1962)
6. Huttenlocher, D. P. and Wayner, P.C.: Finding convex edge groupings in an image, International Journal of Computer Vision, Vol. 8, (1992) 7-27
7. Jacobs, D.W.: Robust and efficient detection of salient convex groups, IEEE Trans. on Pattern Analysis and Machine Intelligence, 4:30 PM 10/31/2003Vol. 18, (1996) 23-37
8. Kosaka, A and Kak, A.C.: Fast Vision-guided Mobile Robot Navigation using Model-based Reasoning and Prediction of Uncertainties, Computer Vision, Graphics, and Image Processing - Image Understanding, Vol. 56, (1991) 271-329
9. Kosaka, A. and Kak, A.C.: Stereo Vision for Industrial Applications, Handbook of Industrial Robotics, edit. S. Y. Nof, John Wiley & Sons, Inc., (1999) 269-294
10. Motai, Y and Kak, A.C.: An Interactive Framework for Acquiring Vision Models of 3D Objects, IEEE Trans. of Systems, Man, and Cybernetics - Part B: Cybernetics, in press, (2003)
11. Motai, Y and Kosaka, A: SmartView: Hand-Eye Robotic Calibration for Active Viewpoint Generation and Object Grasping, Proc. of IEEE Int. Conf. of Robotics and Automation, (2001) 2183-2190
12. Motai, Y and Kosaka, Y.: Concatenate Feature Extraction for Robust 3D Elliptic Object Localization, Accepted in 19th ACM Symposium on Applied Computing, AI track, (2004)
13. Niwakawa, M., Onda, T. and Fujiwara, N., Vision-based handling system using model-based vision and stereo ranging, 1998 Second International Conference on Knowledge-Based Intelligent Electronic Systems, Vol. 2, (1998) 199-204
14. Olson, C.F.: Improving the generalized Hough transform through imperfect grouping, Image and Vision Computing, Vol. 16, (1998) 627-634
15. Porrill, J.: Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter, Image and Vision Computing, Vol. 8, (1990) 37-41
16. Voss, K. and Suesse, H.: Invariant Fitting of Planar Objects by Primitives, IEEE Transaction of Pattern Analysis and Machine Intelligence, Vol. 19, (1997) 80-84
17. Yoshimi, B.H. and Allen, P.K.: Active uncalibrated visual servoing, Proceeding of IEEE International Conference on Robotics and Automation, Vol. 1, (1994) 156-161