09 JAN 1987

RR 6582

NUSC Reprint Report 6582
15 December 1986

Data Fusion in a Multisensor–Multicontact Environment

K. F. Gong
M. L. Graham
Combat Control Systems Department

A. G. Lindgren
University of Rhode Island

Naval Underwater Systems Center

Newport, Rhode Island / New London, Connecticut

Approved for public release; distribution unlimited.

Reprint of a paper presented at the 20th IEEE Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 1986.

ℓℓ6582
9 #2

Data Fusion in a Multisensor-Multicontact Environment

by

Allen G. Lindgren
Dept. of Electrical Engineering
University of RI
Kingston, RI 02881
(401) 792-2505

Kai F. Gong
U.S. Naval Underwater Systems Center
Newport, RI 02841
(401) 841-3846

Marcus L. Graham
U.S. Naval Underwater Systems Center
Newport, RI 02841
(401) 841-2432

ABSTRACT: The problem of estimating the trajectories of an unknown number of contacts in a multisensor-multicontact environment is considered. The data consists of independent and nonassociated estimates or features extracted from locally associated measurements. To avoid the computational difficulties of maximum likelihood estimation, a clustering or decision-directed approach to unsupervised learning is employed. Data samples are clustered and trajectory estimates formed via a hierarchical tree with a criterion for identifying an estimate of the number of clusters or contacts. Complications arise when system observability poses a problem. An algorithm is presented for scenarios where two independent data segments are required for observability. The estimation technique manages the assignment of the data segments by delaying decisions and converting decision making to the resolution of conflicts. An example of such a problem, the simultaneous estimation of trajectories from locally-associated/globally-nonassociated angle-of-arrival data, is detailed and experimental results presented.

I. Problem Statement

This paper is concerned with simultaneously estimating the parameters of an unknown number of dynamic systems. The data is comprised of independent and nonassociated samples. The problem is illustrated in Fig. 1, where it is desired to estimate the state or initial conditions of known dynamic systems. This problem, further detailed in Fig. 2a, is typical of trajectory estimation in a multisensor-multicontact environment. The measurement environment involves unlabelled samples and, generally, no prior knowledge of the parameter set or the number of contacts is available. An alternate problem, typical of texture analysis and image segmentation, is illustrated in Fig. 2b. This problem, where the desired x_i are the ARMA or Markov parameters, is detailed in [1]. In the trajectory estimation problem, the measurements received at various sensors may fade from one contact to another providing unlabelled data streams. The data is assumed to be locally associated (supervised) permitting retention of the information, most critical to the task at hand, in a reduced data set; e.g., sufficient statistics or features. This data set $A = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n)$ is realized through local processing of the raw data, y (k). The probability of a given data stream originating with a particular system is not known a priori. The problem is then, given a data set A with known probability structures, find $x = (x_1, x_2, \ldots, x_c)^T$ where c is the estimated number of contacts.

II. Mathematical Formulation

For data streams with a large number of samples, the local estimate is approximately Gaussian with mean α_i, and Fisher information matrix F_i [1-3]. If each local estimate is labelled (supervised), then the measurements associated with a given system can be clustered and the likelihood function is given by

$$p(A|x) = \prod_{j=1}^{c} \left[\prod_{\alpha_i \in A_j} p(\alpha_i|x_j) \right] \tag{1}$$

where A_j represents the cluster associated with the j^{th} system. For simplicity in presentation it is assumed that α_i is a direct measure of the x_j; extension to the more general problem follows readily. Thus, the maximum likelihood solution for this problem is given by

$$\sum_{\alpha_i \in A_j} F_i (\alpha_i - x_j) = 0 \tag{2}$$

for j = 1,2,...,c. When the global association is not known (unsupervised samples), attention focuses on the mixture density

$$p(\alpha_i|x) = \sum_{j=1}^{c} p(\omega_j)\, p(\alpha_i|x_j,\omega_j), \tag{3}$$

and

$$p(A|x) = \prod_{\alpha_i \in A} p(\alpha_i|x), \tag{4}$$

where ω_j is the event that α_i is associated with the jth system and the $p(\omega_j)$ are mixing parameters (probability of the event ω_j). When c and $p(\omega_j)$ are known, it is readily shown [4,5] that the maximum likelihood estimate (MLE) is given by the solution to the c-coupled equations

$$\sum_{\alpha_i \epsilon A} w_j(\alpha_i, x)\ F_i(\alpha_i - x_j) = 0, \qquad (5)$$

where $w_j(\alpha_i x) = p(\omega_j|\alpha_i, x)$ has the effect of windowing away data not in the vicinity of x_j. It is instructive to view the form of $w_1(\alpha, x)$ for two identically Gaussian distributed clusters with means $x_1 = -d/2$ and $x_2 = d/2$, and variance σ^2. For this case $w_1(\alpha) = [1 + \exp(\frac{\alpha}{\sigma} \cdot \frac{d}{\sigma})]^{-1}$, and the transition from the asymptotic value of 1 for negative α to the asymptotic value of zero for positive α is seen to be abrupt for large relative separations of the cluster centers ($d = x_2 - x_1 >> \sigma$). Thus, under these conditions the problem reduces to one of partitioning A into c disjoint subsets $A_1\ A_2\ A_3 \ldots A_c$, each representing a cluster with "most similar" samples. The more difficult problem of close or overlapping clusters requires maximization of the mixture density. In generalizing the problem to include unknown mixture parameters the constraints $p(\omega_j) \geq 0$, $\Sigma p(\omega_j) = 1$ are imposed. Maximization of (4) is a nonlinear iterative process that is computationally intensive and concern exists regarding the convergence to a global solution. With c also unknown, the alternative approach of employing a decision-directed or clustering procedure is attractive [1,4]. Criteria for measuring the quality of the data partitioning is critical to the overall performance [4]. To evaluate a partitioning $R = [A_1, A_2, \ldots A_c]$, the criterion

$$p(A|R) = \pi_{\alpha_i \epsilon A_j} \left[\sum_{j=1}^{c} \frac{n_j}{n}\ p(\alpha_i|x_j, A_j) \right] \qquad (6)$$

was suggested [1], where n is the total number of measurements, n_j is the number of measurements assigned in A_j, and x_j is obtained via (2).

III. A Hierarchical Approach to Maximum Likelihood Clustering

To begin the estimation or data fusion process, we employ the insight gained in the previous section on the natural way $w_j(\alpha, x)$ tends to partition the measurement space. Hence, we seek a partition that minimizes the criterion

$$\sigma_w^2 = \sum_{j=1}^{c} \sum_{\alpha_i \epsilon A_j} ||\alpha_i - x_j||^2_{F_i}. \qquad (7)$$

Note the relationship of (7), using unlabelled data, to the log likelihood associated with the distribution in (1). Note, also, that the problem is complicated for unknown c by the fact that increasing the number of clusters monotonically decreases σ_w^2. Criteria for identifying an estimate for c is discussed later

A hierarchical approach to minimizing σ_w^2 is to combine nearest neighbors, by an appropriate metric, into a new point (cluster) and continue the process until all points have been combined. Thus, a hierarchical tree is formed starting with n (singleton) clusters, that reduces the number of clusters by one at each step. At each step the criterion (7) is minimized and, therefore, for any selection of c, an appropriate partition is achieved.

It is readily shown that the metric that achieves the above result is

$$d^2_{min} = \min_{k,l} ||x_k - x_l||^2_{\mathcal{F}_{kl}} \qquad (8)$$

where

$$\mathcal{F}_{kl} = \left[F_k^{-1} + F_l^{-1} \right]^{-1} \qquad (9)$$

Here, the points x_k and x_l with Fisher matrices F_k and F_l are combined and replaced by the new point

$$x_{kl} = F_{kl}^{-1}\left[F_k x_k + F_l x_l\right] = x_k + F_{kl}^{-1} F_l x_l \qquad (10)$$

and new Fisher matrix

$$F_{kl} = F_k + F_l. \qquad (11)$$

With the use of rough screening tests, these computations can be kept reasonable. In this manner, global estimates, x_j, for each of the formed clusters are automatically computed in the agglomerative process, see Fig. 3.

IV. Refined Clustering and Estimation Algorithms

Selecting a criterion function is critical to both cluster formation and for deciding on the number of clusters present. Measures with a minimum variance flavor are popular and all generally produce the same good results when the clusters are compact and well separated. When the clusters are close together, or overlap, the performance degrades rapidly and the results become increasingly sensitive to the criterion or metric employed.

To place the problem in perspective, a qualitative analysis based on identically distributed Gaussian clusters shows that both the hierarchical tree approach to clustering (partitioning) described in Section III, and simplified criterion for determining the number of clusters, c, perform well for separations of the cluster centers on the order of $d \leq \sqrt{12}\,\sigma$

A possible criterion for determining c is to minimize

$$\sigma_c^2 = g(c)\sigma_w^2 = g(c) \sum_{j=1}^{c} \sum_{\alpha_i \epsilon A_j} ||\alpha_i - x_j||_{F_i}^2, \quad (12)$$

where g(c) is a monotonically increasing function of c. On the other hand, when the centers are very close together, $d \leq \sqrt{12/n}\,\sigma$, smaller estimation errors result by assuming one cluster even when two or more clusters are present. At this point, separation of the clusters is no longer reasonable. Consequently, it is noted that much research on this subject is directed at achieving efficient techniques that work in the approximate range $d/\sigma = \sqrt{12/n}$ to $\sqrt{12}$, see Fig. 4. It is reported that use of mixture density ML techniques [6] extends the limits of reasonable performance by approximately 3 db, i.e., down to separations on the order of $d = \sqrt{6}\,\sigma$. Currently, separation below this level is questionable and at best sensitive to a priori models and dependent on data rich environments. Therefore, we restrict the scope of this paper to the consideration of cluster separations that permit the agglomerative hierarchical methods of Section III to fuse data from angle-of-arrival sensors. It is noted that divisive techniques (cluster splitting) are available to refine the results. Here, dimensionality reduction to one dimension is achieved by projection pursuit or by projections onto the dominant eigenvector of the resulting scatter matrix of the formed cluster. These projections can be used to efficiently compute new trial partitions of the agglomeratively formed clusters to maximize the mixture likelihood (6). Use of the estimate, x_j, to initialize maximum likelihood estimation algorithm has also been suggested. However, little change in the estimates is expected for widely separated clusters and degradation of the information content of the data (due to the nonassociated nature of the data [5]) occurs when the clusters overlap.

This paper extends the results of the single-contact trajectory estimation problem with segmented data [3] to the multisensor-multicontact problem. In the application to trajectory estimation from angle-of-arrival data the measurements, α_i, are a nonlinear function of the state of the body under observation, i.e., $\alpha_i = \alpha(x)$. Similar to the single-contact problem [3], for the multisensor-multicontact problem, we seek the x_j such that

$$\sigma_w^2 = \sum_{j=1}^{c} \sum_{\alpha_i \epsilon A_j} ||\alpha_i - \alpha(x_j)||_{F_i}^2 \quad (13)$$

is minimized.

V FUSING ANGLE-OF-ARRIVAL DATA

The decision-directed approach is now applied to the specific problem of estimating trajectories from angle-of-arrival data in a multisensor-multicontact environment. For associated data, the estimation problem is detailed in [3,7,8]. Data streams are locally processed to extract (sufficient) statistics or features. This reduced data set is then used to represent the raw data at higher levels in hierarchical approach for associating the elements within this reduced data set. For simplicity, contact motion is assumed to be constant velocity while observer motion is confined to constant velocity "legs" separated by velocity shifts (maneuvers). Because of the nature of the bearings-only observation process, two independent data segments are required for the contact trajectory to be observable [3,7,8]. The association criteria are based in the contact state-space and the following procedure is employed:

•Sufficient statistics, α_i, are extracted by locally processing segments of bearing data.

•Primary cluster estimates, x_j, are formed from all combinations of independent data segments taken two at a time.

•These primary estimates serve as the observations in the contact state-space used in the clustering process described in section III. Because each primary observation requires two data segments, individual data segments may appear in more than one cluster. This is minimized through selective cluster pruning.

•At each level of the hierarchical procedure the cluster set is sorted and clusters which appear as subsets of others are removed. Primary clusters whose segment sets are divided between two non-primary clusters (consisting of more than two segments) are also removed as pairings of assumed non-associated segments. Pruning in this manner greatly reduces the required number of levels in the hierarchical procedure.

•At each level of the tree, once two points have been combined, all segments associated with the combined points are reprocessed to refine the estimate by minimizing

$$\sigma_w^2 = \sum_{\alpha_i \epsilon A_j} ||\,\alpha_i - \alpha(x)\,||_{F_i}^2 \quad (14)$$

and the new information matrix is determined. This new (weighted) cluster center and Fisher matrix are used in the next step of the process.

•This process continues until the estimated number of clusters is declared. At this point some data segments may appear in more than one cluster presenting a conflict in association. Due to the sequential nature of the tree it is also possible for a segment to be wrongly assigned to one or more clusters

•Conflicts are resolved by iteratively

reassigning segments to clusters based on minimizing the contribution to the residual error (13).

•The resulting (weighted) cluster centers are the estimated states or trajectories of the contacts. Confirmation of a contact is not possible when a cluster contains only two data segments.

IV. EXPERIMENTAL RESULTS AND COMMENTS

Experiments were conducted on simulated data. The test scenario is given in Fig. 5 where there are a total of 5 nearly stationary contacts being tracked. There are 5 observer "legs" along the path O-A-O-B-O-C; data are received on all contacts on the first leg, the first four contacts on the second leg, the first three on the third, etc. This provides a total of 15 segments of bearing data for fusion. A nominal zero-mean white gaussian noise is added to the bearing measurements. All estimates are formed at initial time, and range is normalized to the true range. Three separate test runs are presented: case I with all five contacts present, case II with contacts 1,2,3, and case III with contact 1 alone.

Fig. 6 is a scatter plot of the primary clusters for case I where estimates which violate an upper-bound speed constraint have been pruned. This represents the first level in the hierarchical tree. As weighted nearest neighbor clusters are grouped and the cluster sets that appear as subsets of others are pruned, the total number of clusters is reduced. Fig. 7 is a scatter plot of the set of cluster centers at the 8th level of the tree. Fig. 8 is for the 11th level, where the number of segments associated with each cluster center is also indicated. The total residual power versus the number of clusters is shown in Fig. 9. Behavior of the residual power in the vicinity of the correct value of c may be exploited along with other clues or critera to estimate c. In this experiment the criterion used for estimating the number of clusters was simply to minimize (12) with $g(c) = c$. Plots of this function versus the number of clusters are also shown in Fig. 9. While this criterion is appealing for its intrinsic simplicity, improved performance requires a more refined technique such as use of the mixture density (6). Using this criterion, good results were obtained in cases I and II. However, two closely-spaced clusters were estimated for case III, where only a single contact was present, due to the increase in residual power when a single cluster is formed. This effect was dependent on the particular noise sequence. Fig. 10 illustrates the clustering for all three cases after the iterative optimization of the segment assignments. No change is made in the estimated number of clusters, and segments are simply reassigned as necessary to remove conflicts. It is important to note that the fourth and fifth contacts in the case I scenario represent unconfirmable (with only two data segments associated to it) and unobservable (only a single segment associated with it) contacts, respectively. In this case, clusterings for contacts 1,2 and 3 represent confirmed contacts, where the number of data segments assigned to each cluster is as indicated on the figure.

While the approach taken is direct, a certain elegance may be found in that decisions on the association of data segments with contacts are not made until it is necessary to do so. The observability problems of using bearings-only data require the formation of a primary observation set prior to implementation of the clustering procedure. However, through selective pruning at each level of the hierarchical tree, the observation set is quickly reduced by removing subset and "illegal" states. Thus, while the primary observation set may be quite large, the hierarchical tree quickly converges. The subsequent evaluation, on a segment by segment basis, of the residual power contribution of each segment to each cluster permits resolution of any conflicts as well as optimizing the segment assignment. Performance of this optimization phase is dependent on the proper estimation of the number of contacts present, and further work is required in this direction. When the number of clusters is correctly estimated, the clustering generated by the tree is generally quite good and needs little refinement. However, the construct of the tree does allow conflicting segment assignments and resolution of these is still necessary.

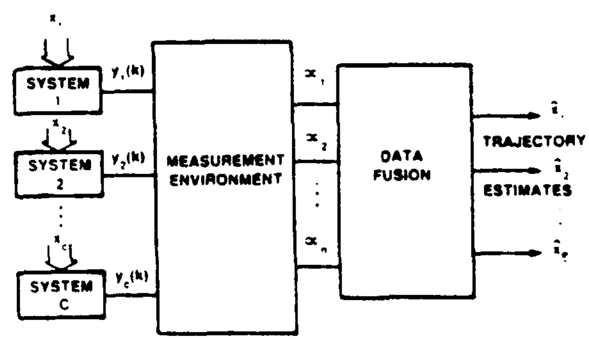

Figure 1
Data Fusion System

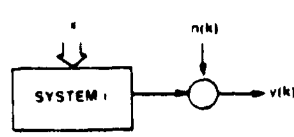

Figure 2a
Trajectory Estimation Model

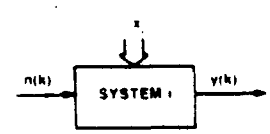

Figure 2b
Texture Recognition Model

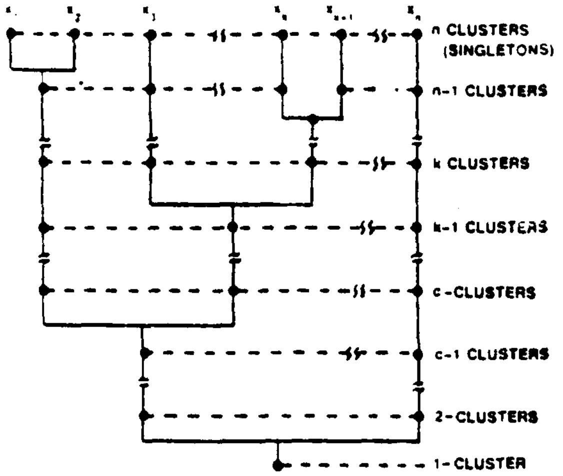

Fig. 3. Hierarchical Tree

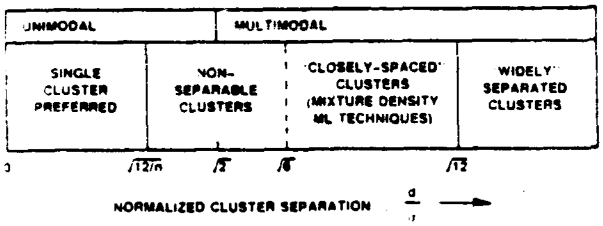

Fig. 4. Qualitatively determined transition between various cluster classifications: analysis of approximate boundaries based on the behavior of two identically distributed Gaussian Clusters

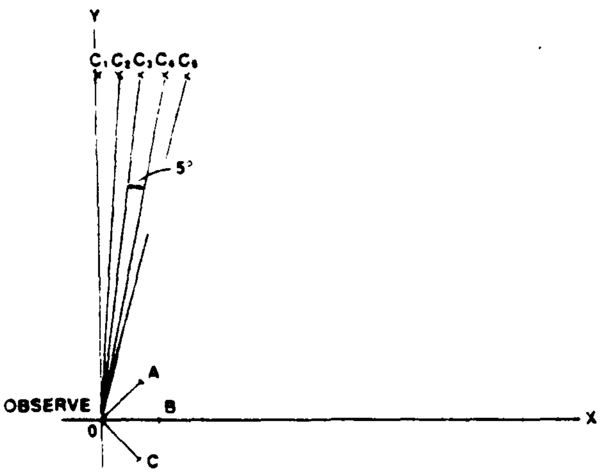

Fig.5. Test Scenario: Contacts spaced 5° apart, observer traverses path O-A-O-B-O-C

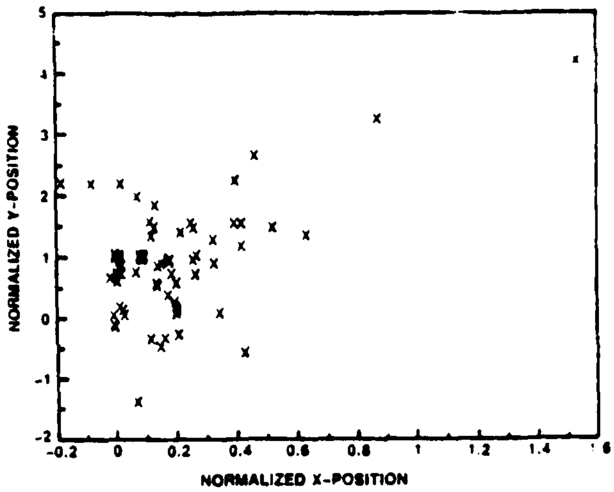

Fig. 6. Plot of cluster centers at 1st level (case 1)

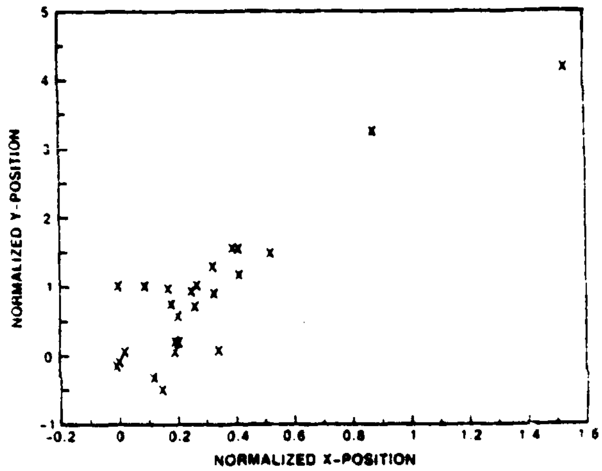

Fig. 7. Plot of cluster centers at 8th level (case 1)

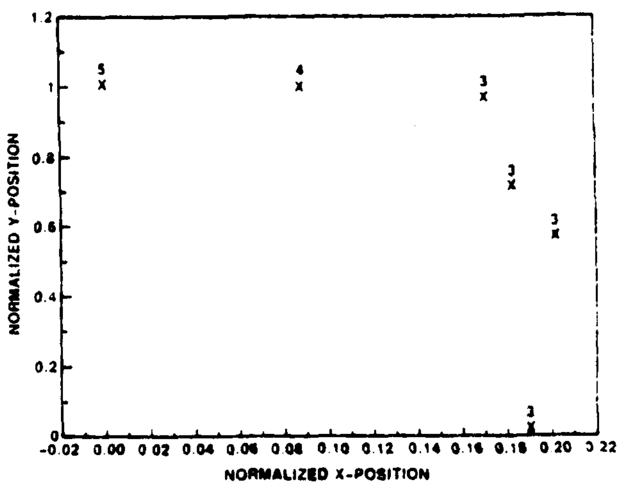

Fig. 8. Plot of cluster centers at 11th level. Numerical values indicate number of data segments assigned to each cluster (case 1).

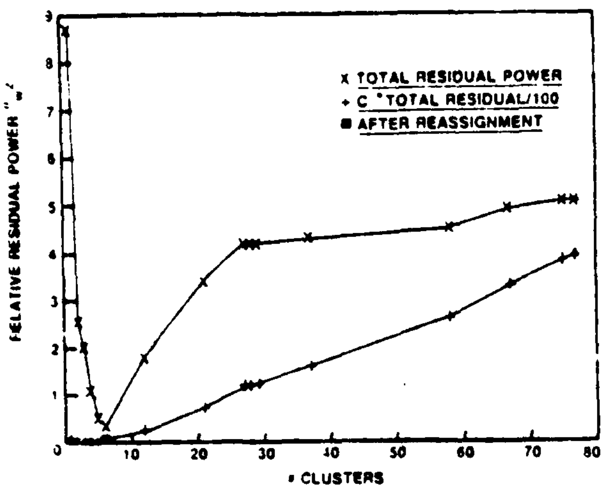

Fig. 9a. Residual-power characteristics (case 1)

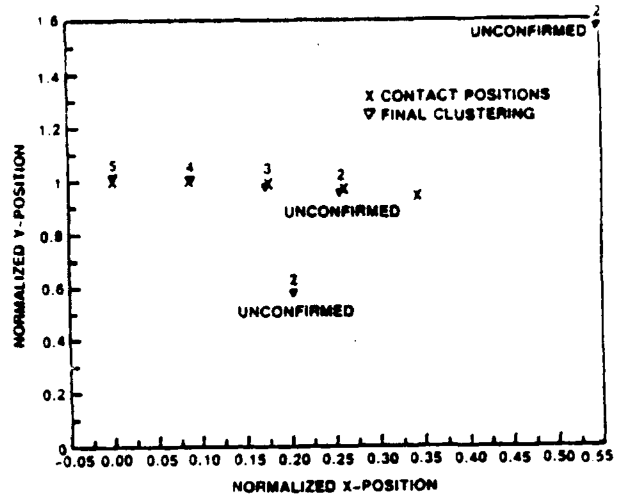

Fig. 10a. Final cluster centers (case 1). Numerical values indicate number of data segments assigned to each cluster.

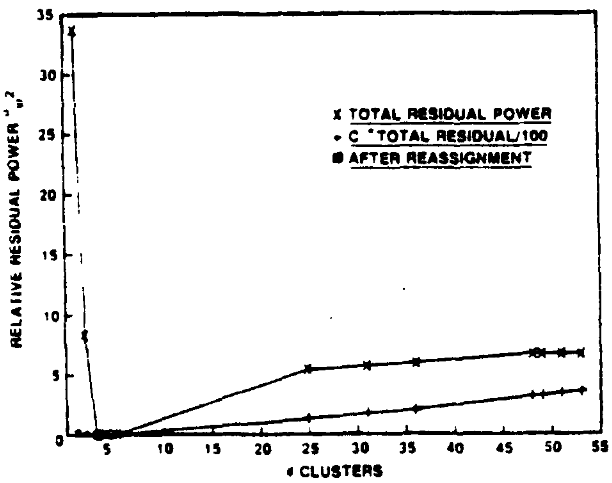

Fig. 9b. Residual-power characteristics (case 2)

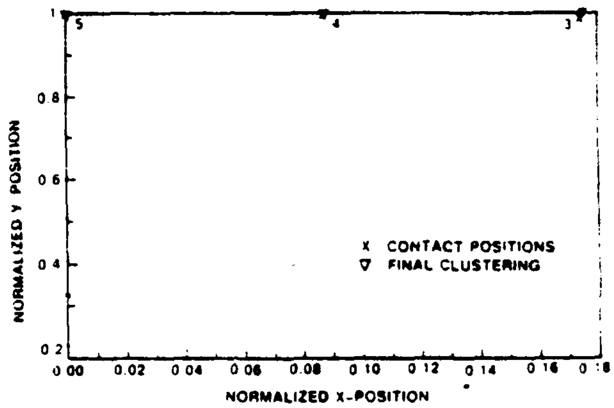

Fig. 10b. Final cluster centers (case 2). Numerical values indicate number of data segments assigned to each cluster.

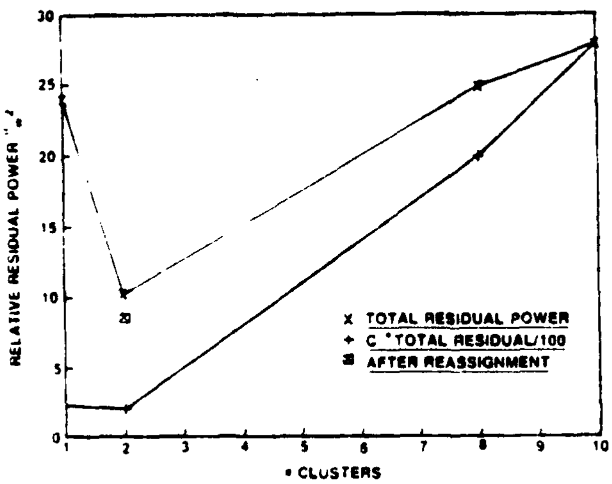

Fig. 9c. Residual-power characteristics (case 3)

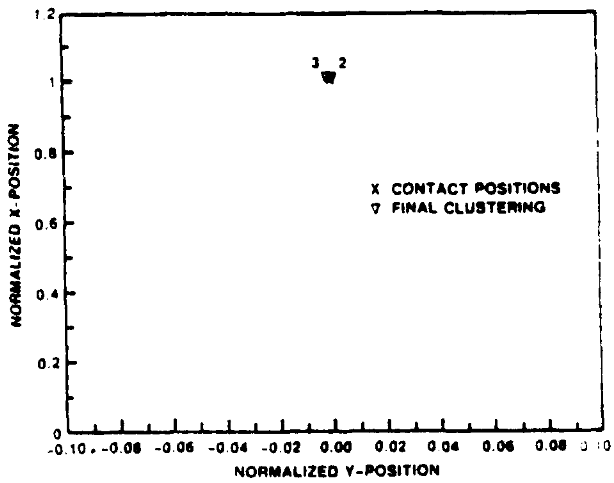

Fig. 10c. Final cluster centers (case 3). Numerical values indicate number of data segments assigned to each cluster

REFERENCES

1. F. S. Cohen, Z. Fan, "System Identification from Unlabelled Samples" Proceedings of the 20th Asilomar Conference, Pacific Grove, CA, Nov. 1986.
2. D. V. Lindley, "The Use of Prior Probability Distribution in Statistical Inference and Decisions," Proceedings of the Fourth Berkely Symposium of Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press, Vol. 1 1961.
3. A. G. Lindgren, K.F. Gong, and M. L. Graham, "Nonlinear Parameter Estimation with Segmented Data: Trajectory Estimation with Biased Measurements." Proceedings 19th Asilomar Conference on Signals, Systems and Computers, Nov. 1985.
4. R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis. John Wiley and Sons Inc., New York 1973.
5. A. G. Lindgren, S. C. Nardone and J. Irza, "Trajectory Estimation with Uncertain and Nonassociated Data." Transactions on Aerospace Electronics, Vol. AES, No. 1, Jan. 1986.
6. F. C. Cohen and Z. Fan, "System Identification from Unsupervised Samples," to be presented at the Conference on Intelligent Systems, Netherland, Dec. 1986.
7. S. C. Nardone, A. G. Lindgren and K.F. Gong, "Fundamental Properties and Performance of Conventional Bearings-Only Target Motion Analysis." IEEE Transaction on Automatic Control, Aug. 1984.
8. D. J. Murphy, "Noisy Bearings-Only Target Motion Analysis," Ph.D. thesis, Northeastern University, Boston, MA 1970.

Acknowledgment

This study was conducted under the Underwater Combat Control Technology Program sponsored by the Office of Naval Technology, program manager D. Houser (OCNR-232), NUSC program manager J.S. Davis (Code 221).