Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER
UNCLASSIFIED	507187

TITLE

COMPLEXITY ANALYSIS FOR NEAREST NEIGHBOR SEARCH. EXECUTIVE SUMMARY ONLY

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to:

National Defence
Research and Development Branch

Défense nationale
Bureau de recherche et développement

TECHNICAL MEMORANDUM 97/240
June 1997

EXECUTIVE SUMMARY ONLY

COMPLEXITY ANALYSIS FOR NEAREST NEIGHBOR SEARCH

P. Zakarauskas — J.M. Ozard

Defence Research Establishment Atlantic

Centre de Recherches pour la Défense Atlantique

Canada

National Defence
Research and Development Branch

Défense nationale
Bureau de recherche et développement

COMPLEXITY ANALYSIS FOR NEAREST NEIGHBOR SEARCH

P. Zakarauskas — J.M. Ozard

June 1997

Approved by: J.S. Kennedy:
Head / Spinnaker Section

TECHNICAL MEMORANDUM 97/240

Defence Research Establishment Atlantic

Centre de Recherches pour la Défense Atlantique

Canada

Executive Summary

Document: DREA Technical Memorandum 97/240

Title: Complexity Analysis for Nearest Neighbor Search

Authors: P. Zakarauskas and J. M. Ozard

Security Classification: Unclassified

Background

The Acoustic Analysis Group at Esquimalt Defence Research Detachment (EDRD) of Deferce Research Establishment Atlantic is investigating the use of Matched-Field Processing (MFP) for the detection and localization of acoustic sources on and in the ocean. MFP consists of matching the acoustic data received on an underwater acoustic array, to predictions of the acoustic field for all possible source ranges, depths and bearings, instead of predictions for bearings alone as is the case for conventional processing. Consequently MFP requires an environmental model. For the processing and environmental costs MFP enhances array gain and localization accuracy. Three requirements must be met to outperform conventional processing. First, the sound propagation must be accurately represented by the propagation models incorporated in the software being developed, second, the sound field sampling by a practical array must be adequate for enhancing performance, and third, the system must operate in real time. Real time execution can be achieved if the sound propagation modelling and the matching are sufficiently fast. The first two requirements have been shown to be met for several scenarios investigated.

The analysis described here is in support of the development of a prototype MFP system in a collaborative venture with the private sector. MacDonald Dettweiler is the prime contractor and the prime private contributor in the private sector.

This report describes an investigation into the speedup that can be achieved by partitioning predictions for the received data so that predictions most like the data can be selected for matching. Separate simulation studies at low signal-to-noise ratios have shown a speedup of a factor of ten. Some theoretical guidance was desired to assist this work.

Principal Results

Three methods for partitioning were investigated with similar speedups predicted under conditions of no noise. The robustness of the partitioning to nonuniformly distributed predictions for the received data was also investigated and found not to present a serious loss of speedup. Partitioning is appropriate for vertical arrays, however other methods for selecting predicted data before matching are indicated for horizontal arrays.

Defence Significance and Future Research

This memorandum provides theoretical support for the results obtained in separate studies where methods for speeding up matching for vertical arrays have been substantiated for several scenarios. In these separate studies it was found that for vertical arrays restricting matching to nearest neighbours of the data resulted in an order of magnitude increase in the search region or number of frequencies that could be covered in real time. However since MFP is likely to see increased application to horizontal arrays, alternate nearest neighbors methods for selecting the propagation predictions prior to matching require investigation.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)

DREA- Esquimalt Defence Research Detachment
P.O. Box 1012, Dartmouth, N.S. B2Y 3Z7

2. SECURITY CLASSIFICATION (Overall security of the document including special warning terms if applicable.)

UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title.)

Complexity Analysis for Nearest Neighbor Search

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

ZAKARAUSKAS, P. and OZARD, J.M.

5. DATE OF PUBLICATION (Month and year of publication of document.)

June 1997

6a. NO. OF PAGES (Total containing information. Include Annexes, Appendices, etc.)

45

6b. NO. OF REFS. (Total cited in document.)

18

6. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the reseach and development. include the address.)

Defence Research Establishment Atlantic
P.O. Box 1012, Dartmouth, N.S. B2Y 3Z7

9a. PROJECT OR GRANT NUMBER (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

1cc15

9b. CONTRACT NUMBER (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number mrst be unique to this document.)

DREA Technical Memorandum 97/240

10b. OTHER DOCUMENT NUMBERS (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; futher distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic annoucement of this document. This will normally correspond to the Document Availability (11). However, where futher distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

In this technical memorandum we present cost estimates for finding the k-nearest neighbors to a test pattern according to a Minkowski p-metric, as a function of the size of the buckets in partitioning searching algorithms. Prior to finding the nearest neighbor, the search space must be ordered using either product or hierarchical partitioning. It is known that fast nearest neighbors algorithms find the k stored patterns closest to a test pattern in a time which is asymptotically constant with the total number of patterns. The asymptotic total cost is derived explicitly for four types of searches based on three different partitioning schemes: the k-d tree partitioning, the ordered partitioning, and the product partitioning The asymptotic expected number of operations performed to find the nearest neighbor is presented as a function of the average number of patterns per bucket n and the dimensionality of the search space d, and is shown to contain a global minimum. Total cost is explicitly calculated for the Euclidian metric, $k = 1$, and realistic computation costs. The optimum value of n is weakly dependent on d, but differs between algorithms. A computer simulation with a uniform distribution of patterns supports the theoretical results.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title).

COMPLEXITY ANALYSIS
NEAREST NEIGHBOR
PARTITIONING
EUCLIDIAN METRIC
BAYES
TREE PARTITIONING
PRODUCT PARTITIONING
SEARCH ALGORITHMS

#507187