COMPUTER - ASSISTED MAPPING FOR CENSUS COLLECTION - APPLICATION EXPERIENCES

Joel Z. Yan, Geocartographics Subdivision,
Data Processing Division, Statistics Canada
D. Ross Bradley, Geography Division, Statistics Canada

ABSTRACT

The Census of Canada, often referred to as the largest peacetime operation carried out by the Government of Canada, takes place once every five years. Its objective is to count all persons living in Canada at their usual place of residence on Census Day. To carry out a task like this requires some 35 to 40,000 census representatives (enumerators) to drop off questionnaires to about 10,000,000 households during a very short but specific time period. To ensure an orderly coverage, maps must be provided to census representatives so that precise geographic areas are covered once and only once throughout the entire land mass of Canada. For the 1981 Census, some 300 collection maps were successfully produced using a prototype computer mapping system. Given promising results, census management approved the development of a production system for computer-assisted mapping, with the intention of producing about 8000 enumeration area maps for the 1986 Census.

This paper describes the computer-assisted mapping system being developed, and application experiences from the 1981 census and a volume pilot study conducted in January 1983.

RÉSUMÉ

Le recensement quinquennal du Canada est souvent présenté comme la plus grande opération du gouvernement canadien en temp de paix. Son objectif est de recenser toutes les personnes à leur résidence habituelle le jour du recensement. La réalisation de cette tâche requiert de 35 à 40,000 énumérateurs pour remettre les questionnaires à environ 10,000,000 ménages durant une très courte mais spécifique période de temps. Pour assurer une couverture méthodique, des cartes doivent être fournies aux recenseurs qui leurs permettent de couvrir une fois et une seule fois l'ensemble du Canada. Pour le recensement de 1981, quelque 300 de ces cartes qui servent à la collecte de données ont été produites avec succès en utilisant un prototype de système de cartographie automatisée. À la suite de ces résultats prometteurs la direction du recensement a approuvé la mise en place d'un système pour la production des 8,000 cartes des aires d'énumération du recensement de 1986.

Cet exposé décrit le système de cartographie assisté par ordinateur, présentement en développment, son expérimentation lors du recensement de 1981 et du projet pilote de production de masse réalisé en Janvier 1983.

KEYWORDS: Computer Cartography, Automatic Text Placement, Human Factors, Urban Mapping

1. INTRODUCTION

To carry out a household survey based on area sampling methodology, good maps are necessary to ensure that enumerators or interviewers know exactly the geographic area that must be covered in the collection of data. This statement is true whether it be a sample survey such as the Labour Force Survey (56,000 households) conducted on a monthly basis or whether it be the Census of Canada (10,000,000 households) with its objective of counting every person usually resident in Canada at his or her usual place of residence on Census Day.

The need for easy to read, large scale maps by thousands of enumerators, not especially trained in map reading, is more than evident. To provide these maps is a large costly Statistics Canada operation, carried out by a team of clerical staff and draftspersons, consuming approximately 80 person years and $1,500,000 for the 1981 Census.

For several years Statistics Canada has been investigating the use of automated mapping techniques to reduce the cost of this operation and to improve the consistency of map bases used for the collection, processing, and retrieval of census data.

2. THE FEASIBILITY STUDY

2.1 Objective

In the fall of 1979, a study was authorized by senior management at Statistics Canada to explore the feasibility of producing, on a highly selective basis, census collection maps by semi-automated means, which could be tested for operational adequacy during the collection of the 1981 Census, and for which their use would have a high potential for subsequent censuses. The study was initiated in the interest of saving human and financial resources through the use of technology, and the integration of activities.

The results of this study have been reported in detail by the authors elsewhere (Bradley 1981; Yan, 1982)

The first activity was to determine the scope of the study and to decide on the selection of sites where the results would be tested for operational adequacy.

2.2 Scope

Three separate areas were chosen. They were (i) a small portion of the City of Kingston, Ontario, (ii) a large section of the city of Gatineau in Quebec, and, (iii) the entire Federal Electoral District of Windsor West. There areas were chosen because of their proximity to Ottawa, the familiarity of the areas by the staff who were to work on this task, and the good quality of the Area Master Files from which this information would be retrieved. But what is an Area Master File and from where does it come?

2.3 The Area Master File

Area Master Files contain a logical representation of all city streets and other geographic features such as railroad tracks, rivers, and municipal boundaries in machine readable form. The Area Master File (A.M.F.) is a product of the Geocoding system which was introduced more than 10 years ago to enable retrieval of census data for user specified areas.

Large urban areas (population 50,000 and over) are divided into block faces. A block face is one side of a street between two successive intersections. These block faces are small enough that when aggregated they become a good approximation for a user identified area. Each block face is assigned a central coordinate, to which files of households, or persons can be coded i.e., geocoded. When a user needs information from a geocoded file, he outlines his area of interest on a map. This area is digitized and becomes a special "query area". All coordinates falling within this area are aggregated and statistical data from the Census are tabulated for those block faces. This process is described in the booklet "Facts by Small Areas" (Statistics Canada, 1972).

Area Master Files now exist for virtually all urbanized areas of the 36 tracted centres in Canada of 50,000 population and over. This constitutes coverage of over 60% of the Canadian population.

2.4 Computer-Assisted Mapping System

It was determined that the quality of maps plotted from the current A.M.F. software was inadequate for census enumeration because the street network was shown as single lines. A system development activity was undertaken to overcome the shortcoming of the plotted maps. Given the short time frame available, there was not time to develop a production system for the 1981 census. Blocks were formed and shrunk on the main frame computer to create double line street patterns. Then AMF's were clipped by polygon and street names were placed using a minicomputer. Thereafter, street names were shifted, and additional text not resident on the input file was added interactively using the

AUTOMAP cartographic system. Three or four cycles of text editing and plotting were required to produce a final quality map. About 2.5 hours of interactive editing were required per Census Tract(CT).

2.5 Results

An evaluation and cost benefit study was conducted of the 32 census tracts maps and 200 enumerator maps prepared by computer-assisted means (see Yan, 1982).

The maps were found to be more than operationally adequate in quality. Even with the high interactive edit times, the costs of producing the maps from the AMF for the prototype system were similar to costs of manual drafting for preparation of the base maps from scratch. Furthermore, the average person time per map was reduced from 17 to 9 person-hours. However, the cost for updating the base map and producing the final CT diagram appeared less with the traditional method.

The actual costs and person hours (PH) measured are given below:

	TRADITIONAL METHOD		PROTOTYPE COMPUTER ASSISTED METHOD	
Preparing the CT Base	$218	17PH	$181	9PH
Updating the Base	$ 65		-	
Producing the diagram	$ 18	0.7PH	$100	4.5PH

A further cost saving of 30 to 40 per cent was predicted with development of a production system for computer-assisted mapping.

Secondly, a savings in the geocoding census capture operations was expected given that the CT map used for collection would be consistent with the A.M.F., which is used for processing and retrieval of the census data . Moreover, a consistent geographic base would be used for urban areas.

Thirdly, computer-assisted mapping could eliminate the double updating operation of street patterns: both on the CT maps and on Area Master Files.

Given these anticipated benefits, the following decisions were made in April of 1982:

(1) Develop and implement a production system for computer-assisted collection mapping to reduce costs and manpower requirements of the current prototype system.

(2) Produce a large number of base maps for the 1986 Census by the computer-assisted method, including all maps requiring redrafting and all maps for new tracted centres.

(3) Maintain the traditional method of producing collection maps for updating and as backup and work-sharing with the computer-assisted system.

(4) Move towards utilizing the AMF as the unique base for urban areas and extend it to cover the entire urbanized core of cities of population 50,000 and over.

(5) Extend the cartographic content of the AMF base to include additional features for mapping.

(6) Investigate potential links with other areas in Statistics Canada, and other government departments and agencies for mapping needs via automated production methods.

(7) Review these decisions in early 1983 based on results of application experiences.

3. DEVELOPMENT OF A PRODUCTION SYSTEM

3.1 Objective

Development of a production system was undertaken to reduce the cost and increase the through put and reliability of the prototype system. It was subsequently decided to utilize the system for the production of about 1200 CT maps encompassing some 8000 enumeration area maps for the 1986 Census. These maps must be provided for verification in the field by early in 1985.

3.2 Improvement to the Prototype System

In May 1982 development of a production system began in earnest. Priority was concentrated on reducing the cost while increasing the throughput, reliability, and quality.

Major enhancements which were included during the period May to December 1982 are described below.

3.2.1 Block Formation System

The system was completely written to use in-core processing as compared to disk I/O. The result was a 100-fold increase in speed and corresponding reduction in price.

This also meant that debugging of the program could proceed much more quickly because tests could be turned around on-line rather than overnight.

3.2.2 Polygon Clipping System

The clipping program was rewritten and moved to the main frame computer. It became the first step in this process, utilized to subdivide a large A.M.F. into more manageable units before block formation.

3.2.3 Cartographic Prompter

An online prompter was developed to review the output from clipping, and for each polygon to query for cartographic control information including map scale, and map title information.

3.2.4 Model Loading and Plotting System

This system was retained on the mini-computer in order to facilitate on-line viewing and edit of the maps. However, a much streamlined batch procedure was developed which greatly reduced the requirements for operator intervention. The result was an 80% reduction in the cost of processing and connect time from approximately $20 to $4 per map. At the same time, a much improved text placement system was developed. Consultation with cartographers and draftspersons played a useful part in this enhancement. Some of the major elements included:

(a) plotting the name centered along features every eight inches based on the final plotscale.

(b) breaking the name into composite words and attempting to fit each word into a street segment rather than the complete name.

(c) ensuring that names of dead end streets do not cross over the junction streets.

(d) keeping track of the number of names which do not fit and coding them for special processing.

(e) providing the option of plotting addresses at block face centroids or street intersections.

The objective was to come as close as possible to producing acceptable quality CT maps without a need for cartographic editing which had consumed about 2 hours per map with the prototype system.

3.2.5 Map Touch-up System

We recognized it would be unlikely, if not impossible, to produce CT maps which would require no change or "touch-up" given the size of some CT's, the density of street patterns, and the requirement that the maps fit on a single plotter sheet of about thirty by forty inches in size. There were two types of touch-up expected: displacement of text to prevent overlap, and addition of cartographic information not currently contained on the digital input files. These changes could usually be made manually on the final plotted maps, but given the need to plot such maps at regular intervals, (at least twice per census), an automated touch-up which could be made once digitally, and utilized subsequently seemed preferable. But, under what circumstances were touch-ups required? It was decided to conduct a volume test to examine this issue.

4. THE VOLUME PILOT STUDY

4.1. Objectives

To conduct a volume pilot study in order to measure the progress made in the system development activity and to determine further developments and improvements required to ensure an effective and efficient production system. Some of the other specific objectives were:

(1) To examine a wide variety of computer-assisted collection mapping situations with respect to feasibility and economy of production;

(2) To ascertain the amount of manual intervention (i.e., touch up) required for computer-assisted collection mapping;

(3) To test the current prototype production system with current inputs and determine what pitfalls and procedural problems exist;

(4) To refine cost, throughput, and time estimates;

(5) To determine the adequacy of the AMF and CT boundary file for computer-assisted collection mapping .

4.2. Scope

The initial intention was to conduct the test for a period of 4 weeks elapsed time, 1 shift, and produce as many maps as possible in that time. Given the availability of the plotter and staff, the plan was reduced to 2 weeks.

The test involved utilizing the system as it existed with no manual intervention or touch-up. A detailed map checking procedure was developed involving careful review of the maps produced by three groups: draftsmen familar with the specifications required, the system team, and the AMF maintenance personnel.

4.3 RESULTS

4.3.1 Throughput

The test was terminated after one week of the planned two weeks because there was sufficient volume and the expected printer had not arrived.

- 207 CT maps were produced in the 6 shifts from Jan. 21 to 28th on the HP - 1000 minicomputer and from Jan. 18th to Jan. 25th on the Amdahl mainframe computer.

- An additional 76 maps were produced from Jan. 31 to February 2nd.

Throughput was approximately:

45 CT's per shift through the prompting and block formation phase.

32 CT's per shift through the model loading and text placement phase.

4.3.2 Costs

Average production costs in dollars per map are given below:

	Average Cost
Preparation Phase	$ .50
Clipping	0.15
Prompting	0.77
Double Line Formation	0.12
Splitting by CT	1.28
Model Loading and Processing	4.03
Plotting	5.18
General Overhead	1.66
Total Cost	**$12.84**

4.3.3 Quality of Outputs

The quality of the maps has improved significantly from earlier tests, but still requires some further improvements. In general, the urban core CT maps at scales of 1:2400, and 1:4800 were quite good, the rural CT maps at scales of 1:7200 and up were too crowded.

4.3.4 Problems with the Current System

Overall, processing with the current system went quite smoothly. Most operational problems that occured were resolved. Production controls went well.

4.3.5 Lessons Learned

From a detailed check of the volume test maps, a list of problems was drawn up. There were nineteen problem areas identified related to the input boundary and street network files, and fifteen problem areas related to the double line mapping system itself. Analysis and resolution of the problems is now underway.

Most of the problems with the inputs arise from the fact that the AMF was designed and created primarily as a geographic base for block face centroids, rather than as a cartographic base. Some issues such as the addition of parks, railway spur lines, and divided highways can be handled easily within the current framework. Other cartographic shortcomings identified can be handled by increasing the number of feature types to include ramps, cliffs, fences, hydro lines, overpasses etc. It is clear that in some cases a location for plotting the names will have to be stored as well as the coordinates which define the feature itself. Some issues such as the addition of the French accents on names will require a more detailed cost-benefit analysis.

Most of the system-related problems were related to the automatic placement of text by the system. Many streets were too short for the complete name to fit at the plot scale requested. In other cases, the names of features close together overlapped. In Quebec city, for example, 42% of the names did not fit when plotted at 1:4800, whereas at only 16% did not fit when plotted at 1:2400. In the future, the plot scale will be reviewed based on the number of names which "appear to fit". Some of the decisions taken which should improve the quality of text placement in the future include:

(a) leaving dead end streets open so the street name will not overlap the line at the end of the street.

(b) shifting text slightly above most single line features (e.g., railways, creeks) but slightly below all boundary type features (e.g., city limit names) to minimize the chance of text overlap.

(c) checking for adjacent segments of similar slope for plotting longer names.

Resolution of these problems is only beginning. Significant progress is expected over the next year, before the final maps have to be produced for the 1986 Census.

Some of the lessons learned which could be generalized to other computer-assisted mapping applications are enumerated below.

Lesson 1: Utilize the automated system for what it does best. Don't insist that the system produce the final map completely.

Lesson 2: Insist that the computer-produced map can be touched up or updated manually.

Lesson 3: Set up a project team which includes experienced personnel from both drafting/cartography and systems development and ensure a high level of commuication. Let the cartographers provide the map specifications.

Lesson 4: Keep the system friendly, responsive and easy to use, so that it will more likely be viewed as a tool and not as a threat by drafting personnel.

5. FUTURE APPLICATIONS

5.1. Collection Mapping

Even assuming that with touch-up and other processing, the cost increases to $60 per map from the current $13, the computer-assisted mapping system should still provide a savings of approximately $150 per map, components to drafting maps from scratch by the traditional method. This savings for 1200 maps of approximately $180,000 should cover the development cost for a production system. Assuming that a cost effective touch-up system can be developed, and that the input files can be extended in cartographic content, the future of computer-assisted collection mapping at the Census of Canada should be bright. If production of the 1200 maps goes well for the 1986 Census, the number of census collection maps produced by the computer-assisted method is likely to increase for the 1991 Census.

5.2. Other Mapping Applications

Given that the system has been developed taking into account a generalized set of requirements, there are applications to other areas besides Census. The system has been designed to produce a set of maps to user specificationsfrom any AMF, and any set of polygon boundaries. A number of different applications have recently been demonstrated including:

- CT and Forward Sortation Area (F.S.A.)-double line maps with postal code shown at block face centroid for Canada Post corporation;
- CT maps showing number of voters per block face to meet the election needs of the Chief Electoral Officer of Canada;
- CT maps showing data variables such as household and population count per block face for field collection purposes.

In the future, one could imagine that a user could request an urban map by a computer-assisted method tailored to his specifications. He or she could select from a menu that included:

- the polygons or study areas of interest
- the urban area or AMF to be processed
- the census data variable to be displayed by block or block face
- the map scale, map title and layout
- single or double line street patterns
- addresses to be plotted or not
- other cartographic features required

6. THE FUTURE

The future for computer assisted mapping appears to be bright. Technology would appear to be in place to allow for the production of large volumes of maps for the specific purpose of collecting information in an orderly manner. Costs for putting a system in place may appear to be high, but the benefits which can accrue will probably render these costs to be quite affordable. The important factor to be considered is that many institutions should be in position to take advantage of this technology. Their needs while different have many striking similarities. The Labour Force Survey and the Census of Canada collect information from households for the most sought after statistical series in Canada. Enumerations for election purposes perform a similar function to those from Statistics Canada,

Canada,and one which allows an election to take place in an organized fashion. Canada Post, in carrying out its mandate to deliver the mail in an effective manner requires a map for control and administrative purposes. The same can be said for municipalities as they provide the many essential services, especially in large urgan areas. They must though be a change in attitude amongst cartographers. They must realize the potential of computer assisted cartography, and take the initiative to ensure that they specify the system to be produced, rather than abicating this aspect to the computer specialists.

The task of coordinating these needs is formidable, but one which has to be undertaken. There must be cooperation among the many users so that the expensive systems to be developped can be afforded by some sort of system of cost sharing. No longer can any duplication of effort be afforded.

REFERENCES

1. GRDSR: Facts by Small Areas, Statistics Canada, June 1972.

2. Bradley, D.R., "Computer Assisted Collection Maps for the 1981 Census of Canada", proceedings of the National Advisory Committee on Control Surveys and Mapping Seminar on the "State and Utility of Computer-Assisted Map Production", November, 1981.

3. Yan, J.Z. "Evaluation Report of Computer-Assisted Collection Maps for Census", 44 pp., Statistics Canada, March, 1982.