# The Regional Crime Analysis Program (RECAP): A Framework for Mining Data to Catch Criminals

Donald E. Brown
Department of Systems Engineering
University of Virginia
Charlottesville, Virginia 22903

ABSTRACT

Most law enforcement agencies today are faced with enormous quantities of data that must be processed and turned into useful information. Two technologies provide the means to turn data into information: data fusion and data mining. Data fusion organizes, combines, and interprets information from multiple sources, and it overcomes confusion from conflicting reports and cluttered or noisy backgrounds. Data mining is concerned with the automatic discovery of patterns and relationships in large databases. This paper describes ReCAP, the Regional Crime Analysis Program that was built to provide crime analysts with both technologies.

## 1. INTRODUCTION

Law enforcement agencies are increasingly acquiring database management systems (DMBS) and geographic information systems (GIS) to support their crime analytic capabilities. In the most technologically advanced departments, crime analysts use these systems to manually search through data, link records, and plot the results on maps. This process can greatly improve crime analysis and aid in reducing and preventing crime, but it is very time intensive. Hence, current use of these technologies to discover spatial relationships and associations between crimes has been limited.

Data mining provides a way to leverage GIS and DBMS technology to support a broader range of crime analytic functions. Data mining is concerned with the automatic discovery of patterns and relationships in large databases. No field is in greater need of data mining technology than law enforcement. To see this we consider a very simple analysis of two metropolitan areas in Virginia.

The Charlottesville City and Albemarle County area is a relatively safe community within central Virginia. The number of reported criminal incidents in this area is typically on the order of 30,000. In contrast Richmond City alone has more than twice that number with approximately 100,000 per year.

Now, for most crimes we want to know if the report is related to any past incidents that might help us solve this current crime or related crimes. We are also interested in relationships with arrest records and jail information, but for this simple analysis we will consider only the incident database. Suppose that we look only at one year's worth of data. Further, for each report, we will want to look at various subsets of the reported fields, e.g. where the incident occurred and when. We want to perform the queries, plot the results, and look for spatial relationships. If each query takes3 minutes on average to retrieve and analyze and we do 5 queries per report, then the crime analyst must spend 450,000 minutes in Charlottesville and Albemarle and 1.5 million minutes in Richmond. Working 10 hour days on these queries and taking no time out for weekends and holidays, the single analyst in Charlottesville-Albemarle can look at less than half of these queries. Three analysts in Richmond working the same schedule can perform less than 15% of their queries. Obviously, this leaves no time to probe deeper than 15 minutes worth of analysis on each crime. Clearly, manual analysis of crime data is infeasible.

Our previous work [6] has shown that only data mining and data fusion algorithms can begin to keep pace with the information processing tasks faced by most crime analysts. We have developed algorithms that automatically process data from airborne and space-based sensors that produce higher data rates than those of law enforcement. However, the law enforcement task is complicated by the need to do semantic analysis on many of the data fields in the reports. Further, to more effectively relate crimes the analysts need to look at combinations of spatial, demographic, victim and other data.

This paper describes our software framework for building and applying data mining algorithms to crime analysis problems. In addition to more traditional approaches to data mining, our framework provides specific focus on spatial data mining. We justify this focus for several reasons. First spatial queries take the longest amount of time. Hence, if we can develop automated techniques to filter through possible spatial

0-7803-4778-1/98 $10.00 © 1998 IEEE

connections, the analyst can devote his or her time to the most important relationships. Essentially the computer will have filtered out the blind alleys and identified the most promising areas for further inquiry. Second, spatial analysis is harder to do than analyses based on attribute matching. The increased complexity derives from the need to explicitly and implicitly consider distance in spatial analysis. The distances may be to a school yard, a major highway, an exit ramp, and so forth. The distances can also be computed as line of sight, travel distance, and shortest travel distance, to name a few methods. Further, spatial analysis involves many different themes, such as income level in a neighborhood, zoning, and type of housing. Clearly the complexity of spatial analysis makes it an ideal area to leverage data mining technology to increase crime analytic productivity.

A third reason for focusing on spatial data mining is that it has the potential to yield important immediate benefits for crime analysis. Brantingham & Brantingham [5] argue that crimes have an inherently spatial component. Rosmo [28] goes further and indicates that geographic profiling can play a dominant role in solving major crimes.

Fourth, spatial analysis is key to law enforcement resource allocation. Effective planning requires an understanding of where and when resources are needed. Finally, if we can develop effective spatial data mining tools then we can also develop effective non-spatial tools. Many of the information processing principles we develop and employ will carry over to other aspects of the crime analysis mission.

In summary, the spatial data mining tools will provide law enforcement agencies with significant capabilities that they currently lack. Without these tools they will be forced to manage the flood of available data through only manual analyses of the most significant crimes.

## 2. A CLOSER LOOK AT THE PROBLEM

We look first at the spatial component of crime and then at existing tools for supporting spatial analysis. We conclude with a brief survey of analysis techniques from fields outside of law enforcement that have direct applicability to this proposal.

### 2.1. The Spatial Component of Crime Analysis

The spatial component of crime has been well documented and hence, the application of crime analytic techniques to understanding this component has also been significant. This subsection makes no claim to completeness in the discussion of this rich source of work. Instead our intent here is to provide a brief summary of this literature as a motivation both to our work and to the work of others that we present in Section 2.2.

One of the most complete discussions of spatial patterning in crime is contained in [5]. From the standpoint of this research we are particularly interested in what they call the microspatial component of crime or the choice of criminal locations by individual criminals. A number of researchers have documented and formulated descriptions for spatial decision making by criminals (see, for example, [4], [20], [24], [29]). Some have looked specifically at the question of distance from home to crime location (for example, [26], [27]). Taken together this impressive body of research shows that "target selection is a spatial information processing phenomenon." ([5] p. 344).

The evident spatial element of criminal behavior has meant that law enforcement personnel have for centuries included spatial analysis in their work. Weisburd and McEwen [31] provide a brief history of crime mapping in law enforcement going back to the early 1800s. The advent of GIS has made possible new types of spatial analysis of crime and created an explosion of interest in crime mapping (see, e.g. [2], [3], [12], [32]). This work suggests that we now have in place the mapping tools (GIS), the database tools, and some analytical methodology. However, most law enforcement agencies are using these tools as obvious computer-based versions of the old pin-mapping techniques. The combination of computer hardware and software provides us the opportunity to do considerably more with the data available. Not surprisingly many people have identified this need and built tools to address various aspects of it. We examine these tools in the next section.

### 2.2. Existing Systems for Spatial Analysis of Crime

A variety of tools have been developed to aid law enforcement in performing crime analysis and some of these specifically focus on spatial analysis. The emphasis of our work is on automating spatial analysis, but we want these tools to be used as widely as possible. Hence, our review here begins with more general crime analysis systems and then looks specifically at spatial crime analysis systems.

As we consider more general crime analysis systems, the question we need to answer is: Can we put spatial analysis data mining tools within the framework of existing or proposed crime analysis approaches? While a phase of our research will attempt to directly answer this question, a look at the literature in this area suggests a positive answer to this question. Many of the systems built over the last decade use either an expert systems or a model-based approach. Examples within the former category include the Armed Robbery Eidetic Suspect Typing (AREST) [1], VICAP [14], APES [7] and the

Baltimore County Burglary System [23]. None of these systems explicitly use spatial analysis. However, the general rule-based approach inherent in these systems can easily and directly incorporate the results of spatial data mining. For example, one of our data mining techniques (classification trees) produces rules. Thus, for agencies using systems like these, our spatial data mining tools can run in the background to produce rules for direct use in these systems.

Spatial data mining can also directly support model-based approaches. An early example of a model-based approach is in [13]. More recently Gottlieb, et al. [11] have codified this approach. They show how to employ simple statistical models to describe the behavior of criminals in an area. This idea combined with information processing techniques has been implemented in a wide variety of software products, such as Watson[13]mex, iGlass, ALTAanalytics, and SIUSS. All of these products find links between data elements but they do not specifically identify related suspects and criminal incidents nor do they do any spatial analysis. Once again, our spatial data mining tools can significantly enhance these existing products.

In addition to general crime analysis systems, a number of researchers, law enforcement agencies, and commercial companies have built spatial analysis support systems. Geake [9] describes the Crime Pattern Analysis System, which looks at spatial displays and summary statistics. The San Bernadino County Sheriff's Department [18] developed a program with similar capabilities. The San Diego police have also employed GIS to help understand crime patterns in their area [21]. However, none of these systems really provide spatial analysis tools for crime analysis. STAC, which was developed by Carolyn Block and described in [3], is a good example of a spatial analysis tool. Her approach provides a convenient computer aid for viewing gang related hot spots.

Many other spatial tools are listed in the books edited by Block, Dabdoud, and Fregly [3] and Weisburd and McEwen [32]. Also in February of this year a cluster conference on spatial crime mapping was held in Washington, D.C. and sponsored by the NIJ. This conference provided the opportunity to review the most important crime mapping work currently taking place in the U.S. under the sponsorship of the NIJ. We briefly summarize the key contributions from this conference as they relate to our proposed effort.

ICAM 2 [17] is a collection of data querying capabilities developed for the law enforcement officer to get simple information and reports very quickly and easily. Researchers at CUNY have investigated a number of techniques for innovative crime mapping [10]. Their strategy is to make these techniques user-friendly for COMPSTAT, precinct commanders, and Crime analysts. Rich [25] has built a system to make crime mapping easily available to community organizations. DeVoe and Nulph [8] have built SCAS which provides spatial display and reporting of crime incidents. LeBeau and Bennett [15] are working on spatial decision support systems to explore patterns arising from activity of known types, such as convenience store robberies. Finally Levine and Canter[16] are building CrimeStat, a set of spatial statistical tools that can be used with any GIS package on any platform.

The current systems have three significant shortcomings:

1. The existing systems and approaches do not explicitly support a user's need to spatially query a data set to relate and extract those data in which s/he is interested;
2. The analyses is not automated and cannot run autonomously, so the user must devote considerable time and effort to the process; and
3. The user must be proficient in GIS and mapping technology to use the system.

The crime analysis systems we describe in the next section provides a first approach to addressing all three problems. In particular, this approach provides the ability to relate data through a variety of spatial (and eventually non-spatial) attributes. These relationships can be explored manually now but we are extending the system to automatically learn by letting the machine mine the data. The ability to automate this process will greatly increase crime analytic effectiveness as the machine under guidance from the user explores and mines the data set. Finally, we let the computer handle the intricacies of the spatial modeling and present the user with the results. This frees the user from the need to be an expert in GIS and mapping.

## 3. A SOFTWARE SYSTEM FOR DATA FUSION AND DATA MINING OF CRIME DATA

The Regional Crime Analysis Program (ReCAP) system is a computer application designed to aid local police forces (e.g. University of Virginia (UVA), City of Charlottesville, and Albemarle County) in the analysis and prevention of crime. ReCAP works in cooperation with the Pistol 2000 records management system, which aggregates and houses all of the crime information from a region. Research and development has focused primarily on the individual components of the system - a database, geographic information system (GIS), and data mining tools. The fully functional system consists of the seamless integration of all components of the system. This seamless interface focuses the attention of those using the system on results of analysis or inquiries rather than on how to retrieve data. In particular, the components of the system intuitively perform the following functions:

The database of police incidents stores all necessary information.
The database query tools allow the extraction of required information.
The spatial tools provided by the GIS map out the crime incident data and provide spatial data mining and analysis.
The data mining and temporal tools provide insight into the nature of crime incident data through control charts, clustering analysis and forecasting.

The operational environment for ReCAP is a Local Area Network (LAN). The LAN consists of a server (with software), several secured network lines, and PCs located at each agency. The software comprising the new system is all Windows 95 and Windows NT-based. The regional server warehousing the regional data runs Novell NetWare operating system. The PCs at each department vary extensively in their capabilities but most of the PCs are Pentium-based.

The system has the following capabilities and functions:

- Hot Sheet: This report gives narrative summaries of the most important criminal activity in the region. Past Hot Sheets are accessible to the user.
- Summary Reports: These reports tally specific crime occurrences over a user-defined time and area.
- Map-Oriented Searches: These searchers provide a GIS display of the area along with plotted criminal activity. Crime data, taken from the secondary database, can be queried and displayed by type of crime, time/date of crime, location, suspect description, weapons involved, etc. The query engine user interface is developed in Visual Basic. These searches will provide the police departments with spatial analysis capabilities.
- Control-Charting: a control chart serves as a tool to alert the user when the process being monitored is out-of-control. Along with a time series plot, the control chart has three horizontal lines for reference. The centerline indicates the average or statistical mean level of nonconformity. The upper and lower control limits indicate the upper and lower range of variance in normal activity. If the time series plot falls above the upper limit or below the lower limit, the process is not in statistical control. An out-of-control result signifies that some sort of corrective action may need to be taken. The control charts can be run automatically at night, and the user can receive information about out-of-control areas every morning. The system can also run in autowatch, which means that it runs in the background looking for changing patterns of criminal behavior. If a change is detected, the crime analyst is then notified.
- Time-Charting: Patterns developing over time of day, day of the week, etc are plotted cyclically. Since many crimes have an indeterminate time of occurrence, the crime is plotted statistically using kernel density estimation. Each crime is distributed uniformly through the time range when it could have occurred. The officers on scene record these times into the incident report. This method is a great improvement over simply picking the midpoint of the time span. This tool allows an analyst to detect a significant rise in a certain crime level at for instance mid-afternoon when school lets out; or morning, when residents go to work leaving their homes unattended. As with control charting, the time-charting can also detect deviations from normal patterns of crime in an area.
- Hot Spot analysis using Kernel Density Estimation (KDE): Kernel density estimation is used to identify spatial patterns of crime within an area. Contour and color plots show the high density areas. A threshold density level can also be chosen, where higher densities are labeled at "Hot-Spots" and law enforcement can take action to reduce crime in that area. This analysis can be performed for any subset of crimes produced by a user's query.
- Cluster Analysis: Clustering algorithms search for statistically significant groupings of crimes in an area and alert the crime analyst user to potential problem areas. Currently partitional methods, k-means and nearest neighbor, are used to perform the clustering. ReCAP allows users to vary the resolution of the clusters found. As a user zooms into and area, the system will "dissect" existing clusters into smaller clusters. In this way, the user can establish statistical significance at whatever level of detail is desired.
- Detailed Inter-Modular Searching: These searches provide the capability to query the database for links between vehicles, suspects, warrants, etc. These searches are essential for making connections between related crimes.

Figure 1 summarizes the flow of information in the system, and the implementation tools involved in each component:

ReCAP Data Flow Diagram

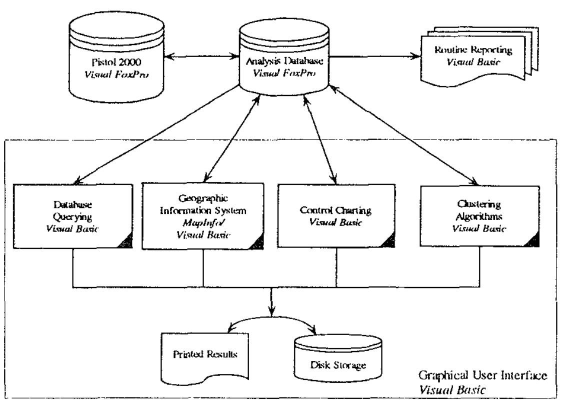
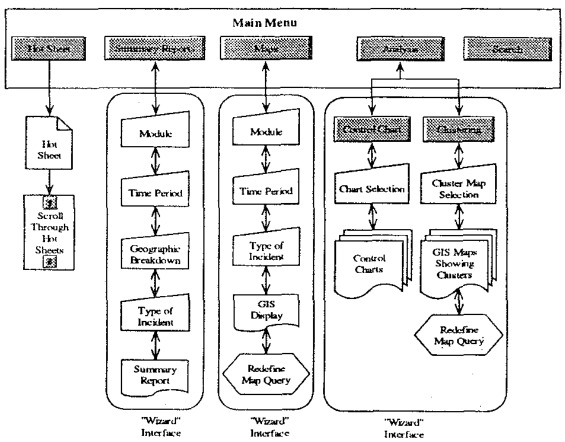



Figure 1. ReCAP Data Flow Diagram

Figure 2 shows how the user navigates through ReCAP to accomplish the various tasks.




## 4. CONCLUSIONS

The ability to quickly identify related records to generate investigative leads is obviously desirable in any police jurisdiction. Such a capacity not only produces information relevant to the identification and apprehension of criminal offenders, it also facilitates recognition of single incidents that are symptomatic of larger patterns of social disorder. Such recognition is critical to practical implementation of the Problem-Oriented Policing model and to the successful operation of Community-Oriented Policing practices such as Compstat. The need for this technology is particularly acute in urban areas where the human capacity to track and recognize long-term and even short-term relationships is overwhelmed by the sheer volume of records.

We have described ReCAP, a software system which provides crime analysts with the capability to mine data to catch criminals. The system provides spatial, temporal, and attribute matching methods for mining data.

## REFERENCES


[1] Badiru, A. B., J. M. Karasz, and R. T. Holloway. (1988) AREST: Armed Robbery Eidetic Suspect Typing Expert System. *Journal of Police Science and Administration.* 16:210-216.

[2] Block C.R. and Dabdoud, M. (1993) *Workshop on Crime Analysis Through Computer Mapping Proceedings*, Illinois Criminal Justice Information Authority, Chicago.

[3] Block C.R., Dabdoud,M., and Fregly, S. (1995), editors, *Crime Analysis Through Computer Mapping*, Police Executive Research Forum, Washington, D.C., 20036

[4] Brantingham, P. and P. Brantingham (1975), Spatial Patterns of Burglary, Howard *Journal of Penology and Crime Prevention*, 14, 11-24.

[5] Brantingham, P. and P. Brantingham (1984) *Patterns in Crime*, Macmillan Publishing Company, New York.

[6] Brown, D.E., C.L. Pittard, and A.R. Spillane (1992) "ASSET: a simulation test bed for evaluating data association algorithms," *Computers and Operations Research*, Vol. 19, No. 6, 479-493.

[7] Coady, W. F. (1987) "Investigating with APES." *Security Management.* 31:67-70.

[8] DeVoe, J. and Nulph, D. (1998) "Spatial Crime Analysis System – SCAS," NIJ Cluster Conference on the Development of Spatial Analysis Tools, February 26 – 27, Washington D.C.

[9] Geake, E. (1993), How PCs predict where crimes will strike, *New Scientist*, Oct. 23, 1993, p. 17.

[10] Goldsmith, V., McLafferty, S., Mollenkopf, J. and McGuire, P. "Identification, Development and Implementation of innovative Crime mapping Techniques and Spatial Analysis," NIJ Cluster Conference on the Development of Spatial Analysis Tools, February 26 – 27, Washington, D.C.

[11] Gottlieb, S., S. Arenberg, and R. Singh (1994) *Crime Analysis from First Report to Final Arrest*, Alpha Publishing.

[12] Harries, K.D. (1990) *Geographic Factors in Policing*, McGraw-Hill, New York, NY.

[13] Harris, D. H. (1978) "Development of a Computer-Based Program for Criminal Intelligence Analysis" *Human Factors.* 20 (February), 47-56.

[14] Icove, D. J. (1986) Automated Crime Profiling. *Law Enforcement Bulletin.* 55:27-30.

[15] LeBeau, J.L. and Bennett, D. (1998) "Spatial Decision Support Systems," NIJ Cluster Conference on the Development of Spatial Analysis Tools, February 26 – 27, Washington D.C.

[16] Levine, N. and Canter, P.(1998) "Development of Spatial Statistical Tools," NIJ Cluster Conference on the Development of Spatial Analysis Tools, February 26 – 27, Washington D.C.

[17] Lewin, J. (1998) "ICAM 2 Update," NIJ Cluster Conference on the Development of Spatial Analysis Tools, February 26 – 27, Washington D.C.

[18] Miller, T. (1994) Computers track the criminal trail, *American Demographics*, January, pp. 13-14.

[20] Molumby, T. (1976) "Patterns of Crime in a University Housing Project," *American Behavioral Scientist*, 20, 247-259.

[21] Newcomb, T. (1993) "High tech tools for the war on crime," *Governing*, August, pp. 20-21.

[22] Newman, O. (1972) *Defensible Space: Crime Prevention Through Urban Design*, Macmillan, New York.

[23] Ratledge, E. C. and J. E. Jacoby. *Handbook on Artificial Intelligence and Expert Systems in Law Enforcement.* New York: Greenwood Press, 1989.

[24] Repetto, T.A. (1974) *Residential Crime*, Ballinger, Cambridge, MA.

[25] Rich, T. (1998) "Introducing Community Organizations to Mapping and Crime Analysis," NIJ Cluster Conference on the Development of Spatial Analysis Tools, February 26 – 27, Washington, D.C.

[26] Rosmo, D.K. (1993) "Target Patterns of Serial Murders: A Methodological Model," *American Journal of Criminal Justice*, 17(2), 1-21.

[27] Rosmo, D.K. (1994) "Targeting Victims: Serial Killers and the Urban Environment," in *Serial and Mass Murder: Theory, Research, and Policy*, ed. by T.O'Reilly-Flemming and S. Egger, University of Toronto Press, Toronto.

[28] Rosmo, D.K. (1995) Overview: Multivariate Spatial Profiles as a Tool in Crime Investigation, in *Crime Analysis Through Computer Mapping*, ed. by C.R. Block, M. Dabdoub, and S. Fregly, Police Executive Research Forum, Washington, D.C. pp. 65-98.

[29] Scarr, H.A. (1973) *Patterns in Burglary*, 2nd Ed., U.S. Dept. of Justice, Washington, D.C.

[30] *Security Management* (1994) "Software Solution," 38 (March), 17.

[31] Weisburd, D. and McEwen, T. (1997a) "Introduction: Crime Mapping and Crime Prevention," in *Crime Mapping and Crime Prevention*, ed. by D. Weisburd and T. McEwen, Criminal Justice Press, Monsey, NY.

[32] Weisburd, D. and McEwen, T. (1997b), editors, *Crime Mapping and Crime Prevention*, Criminal Justice Press, Monsey, NY.