

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

**Document Title: Artificial Neural Network System for
Classification of Offenders in Murder and Rape
Cases**

Author(s): L. J. Kangas

Document No.: 190984

Date Received: October 25, 2001

Award Number: 97-IJ-CX-K007

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

**Opinions or points of view expressed are those
of the author(s) and do not necessarily reflect
the official position or policies of the U.S.
Department of Justice.**

190984 c.2

Artificial Neural Network System for Classification of Offenders in Murder and Rape Cases

Summary

Prepared by Lars Kangas for the
National Institute of Justice,
Office of Science and Technology
Under Award Number 97-IJ-CX-K007

When a serial offender strikes, it usually means that the investigation is unprecedented for that police agency. The volume of incoming leads and pieces of information in the case(s) can be overwhelming as evidenced by the thousands of leads gathered in the Ted Bundy Murders, Atlanta Child Murders, and the Green River Murders. Serial cases can be long-term investigations in which the suspect remains unknown and continues to perpetrate crimes. With state and local murder investigative systems beginning to crop up, it will become important to manage that information in a timely and efficient way by developing computer programs to assist in that task. One vital function will be to compare violent crime cases from different jurisdictions so investigators can approach the investigation knowing that similar cases exist.

The "Artificial Neural Network System for Classification of Offenders in Murder and Rape Cases" project developed two software prototypes that demonstrate developed algorithms for analyzing and comparing large databases of crime data. The CATCH (Computer Aided Tracking and Characterization of Homicides) and CATCHRAPE software applications analyses homicide data and sexual assault data, respectively. Both applications are similar, although they analyze different databases. "CATCH" will from hereon refer to both applications.

CATCH was developed to provide crime analysts enhanced means for interpreting large databases of crime data. These databases store a large number of crimes with each case described in a large number of details. The huge volumes of information collected for crimes during investigations make it difficult to compare a crime against other crimes by the investigators alone. CATCH using advanced algorithms running on computers facilitates the investigators need to compare crime data and to find crimes that have similar patterns.

The development of CATCH was made possible with the HITS (Homicide Investigation Tracking System) database system. Police involved in the infamous Green River and Ted Bundy murder investigations in the State of Washington developed HITS more than ten years ago to enable computer-based analysis of murders. The database now contains several thousand violent crimes primarily from the Pacific Northwest. The HITS database system is similar in the crime details it stores to the VICAP system at FBI and the violent crime database system operated by the RCMP in Canada.

The algorithms in CATCH learn from crime data. HITS provided the required historical crime data for CATCH to learn offender behavior so that new unsolved crimes can be characterized for analysis and comparison.

PROPERTY OF

National Criminal Justice Reference Service (NCJRS)
Box 6000
Rockville, MD 20849-6000

The CATCH software assists with crime investigations by assessing likely characteristics of unknown offenders, by relating a specific crime case to other cases, and by providing a tool for clustering similar cases that may be attributed to the same offenders. An investigator working with CATCH is working against the collection of crimes in the HITS database. The software speeds up the database access to these crimes, improves the views of the crimes, and organizes the crimes according to similarity.

CATCH is a collection of tools that assist the crime analyst in the investigation process by providing advanced data mining and visualization capabilities. These tools include clustering maps, query tools, geographic maps, timelines, etc. Each tool is designed to give the crime analyst a different view of the case data.

The clustering tools are based on artificial neural networks (ANNs). The ANNs in CATCH learn from a subset of existing crime cases (approx. 5000 homicides, 3000 sexual assaults) in the HITS database using parameters describing modus operandi (MO) and signature characteristics of the offenders. The ANN clusters cases onto two-dimensional visual representations, according to the similarity of offender patterns. The proximity of cases within the feature map allows the analyst to identify serial homicides/sexual assaults.

A case being investigated is compared to the solved adjacent cases in the visual representation of the database. A set of one or more cases is analyzed in detail to determine whether a linkage exists between the cases. For example, is it one offender that is responsible for a series of crimes? The investigator goes between the different tools in CATCH and analyses when and where the crimes were committed. The analyst can focus on specific details that he or she believes are important for determining whether there is a connection between the crimes. The analysis work continues using a set of cases where individual cases can be dropped or new cases can be added from the database. All accesses to the database are fast and simple using point and click mouse methodology.

Access to a database system, like the HITS database, requires strict adherence to a query language to retrieve information. This language is slow and difficult to use even for an experienced database professional. SQL is the most common language for accessing a database. CATCH removes the need for an investigator to know this language by dynamically creating the queries in the SQL language on demand, as the investigator accesses the database. The investigator can concentrate on the information in the databases, rather than the access method. The investigator only sees the information in CATCH, not the query language, as in the previous version of software accessing the HITS database.

CATCH characterizes an unknown offender by comparing his or her crime to known offenders of other crimes where the behaviors of the offenders have been similar. The characterization can give some clues to personal information about the unknown offender to expedite the investigation process. This information includes, level of education, type of work, physical appearances such as neatness, misuse of alcohol and drugs, etc. The underlying assumption is that two individuals that otherwise behave similarly, may also commit crimes with similar

behavior. The validity of this assumption has been proven to some degree by FBI's behavioral science unit and others.

The characteristics of offenders developed in CATCH are hints at best and differs from most of the other information, which are factual and retrieved from the database. With the exception for the characterization tool, CATCH was not designed to modify or generate data describing the crimes in the database. It only organizes the existing facts to provide more meaning to the investigators.

CATCH was developed using visionary information technologies to advance crime analysis. These technologies include methods for analyzing data, visualizing information, and accessing data. Research staff drew from past experience in these technologies and pushed the envelope just a little further by combining these technologies in one application. The application has received considerable attention in the media as being new and innovative.

The investigations of crimes involve analysis of offender behavior. A set of automated technologies can make the work of investigators more efficient, they cannot replace the investigators. Although technologies are becoming more advanced, a software tool should still only be in support of the human in-depth understanding of behavior. The quintessential objective in the research was to facilitate the work of investigators.

The development of CATCH would not have been possible without the collaboration Battelle enjoyed with Washington State Attorney General's Office, Criminal Division. The experience of the crime analysts and the extensive experience they have in operating a database of violent crimes were invaluable to the project. These crime analysts were the drivers that recognized the need to research the possibility of using advanced technologies to make their work more efficient.

CATCH demonstrates the value of incorporating new technologies and methods that can aid the work in crime investigations. The scope of the project was limited to specifically show how the artificial neural network technology can support a multivariate or information fusion approach to finding significant patterns in crime data. Other research and development efforts concentrate on equally important aspects of crime analysis, such as the use of graphical information systems. Although CATCH was developed as a self-sufficient application and includes some crime mapping and other tools of a complete application for crime analysts, it did not explore or exhaust the possibilities of those tools.

Two versions of CATCH were developed, CATCH and CATCHRAPE for homicide data and sexual assault data, respectively. These applications were deployed at Washington State Attorney General's Office, Criminal Division for evaluation. Although the first versions of CATCH were configured specifically for the HITS database of violent crimes, they can be applied against other crime databases through relatively minor changes in the software.

Status

PNNL finished Version 1.0 of CATCH (Computer Aided Tracking and Characterization of Homicides). A second application CATCHRAPE to analyze sexual assaults was also completed with a subset of the functionality of CATCH for murders.

A parallel project at PNNL, sponsored by Washington State, redesigned the HITS database after the completion of CATCH, Version 1.0. A development of CATCH, Version 1.1 was started to accommodate the new HITS database. This version is operational at PNNL using an estimated 95% of the crime details from the new version of the HITS database. (The decision to redesign the HITS database came very late in the CATCH development, when most of the project time and funds had been spent. Version 1.1 has not been completed for installation at Washington State AGO to run against the current HITS database).

Washington State AGO started an evaluation of CATCH. The evaluation was not extensive due to loss or availability of key staff at AGO, which included Dr. Robert Keppel's retirement and Ken Hanflen's departure.

Preliminary evaluations by Dr. Keppel, et. al., suggest that the clustering algorithms and visualization tools in CATCH have the potential to add considerable value to crime analysts.

Presentation of CATCH to NIJ

PNNL is scheduled to present CATCH at an "In Process Review" arranged by Mr. William Deck, NLECTC Southeast to Mr. Trent DePersia, NIJ, et.al., May 22, 2000 in Denver, Colorado.

Description of CATCH

CATCH was developed to assist crime investigations by assessing likely characteristics of unknown offenders, by relating a specific crime case to other cases, and by providing a tool for clustering similar cases that may be attributed to the same offenders.

CATCH is a collection of tools that assist the crime analyst in the investigation process by providing advanced data mining and visualization capabilities. These tools include clustering maps, query tools, geographic maps, timelines, etc. Each tool is designed to give the crime analyst a different view of the case data.

The clustering tools in CATCH are based on artificial neural networks (ANNs). The ANNs learn to cluster similar murder and sexual assaults cases residing in a database. The clustering algorithm is applied to parameters describing modus operandi (MO), signature characteristics of the offenders, and other parameters describing the victim and offender. The proximity of cases within a two-dimensional representation of the clusters allows the analyst to identify similar or serial murders and sexual assaults.

Future Work

We finished a first draft proposing a new version of CATCH, which incorporates additional tools that have been identified from the current research and development. The first set of tools in Version 1 of CATCH was concentrated on researching the value of using artificial neural networks to cluster similar cases. The new tools will provide the crime analysts a more complete suite of tools.

Presentations of CATCH

- L.J. Kangas, K.M. Terrones, R.D. Keppel, R.D. La Moria. 1999. "Computer Aided Tracking and Characterization of Homicides & Sexual Assaults (CATCH)." SPIE AeroSense '99, Orlando, FL, April 5-9.
- L.J. Kangas. 1999. "CATCH." Justice Research and Statistics Association, Washington, DC, September 30.
- L.J. Kangas, K.M. Terrones. 1999. "Data Mining and Visualization for Crime Analysis: CATCH." Presented at 8th Advanced Information Processing & Analysis Steering Group, Washington DC, March 23-24.
- S.V. Shoemaker, L.J. Kangas, 1998. "Washington Courts Database Conversions." Information Sharing Technology Workshop, Georgia Institute of Technology, May 6.

Media Exposure of CATCH

- "Computer Takes a Byte out of Crime." Popular Mechanics, March 2000.
- "CATCH Program." TECH beat, Fall 1999.
- "New Way to CATCH Bad Guys." <http://www.govtech.net/onlineservices/news/newspage.shtm?story=June181999NewWaytoCATCHBadGuys>, June 18, 1999.
- "Catch as CATCH Can." MIT's Magazine of Innovation Technology Review, March/April 1999
- "Buying Computers and Software" (CATCH software described) Law Enforcement Technology, April 1999
- "CATCHing Criminals." http://newsweek.com/nw-srv/issue/15_99a/tnw/today/cs/cs01th_1.htm, April 9, 1999
- "Thinking Like a Serial Killer." http://www.wired.com/news/print_version/technology/story/19940.html?wnpg=all, June 1, 1999
- "David Lawrence talking with Lars Kangas, Sr. Research Engineer at Pacific Northwest National Labs" <http://www.audiocast.net/geekradio>, June 7, 1999.

Artificial Neural Network System for Classification of Offenders in Murder and Rape Cases

L. J. Kangas

October 2000

This project was supported under award number 97-IJ-CX-K007 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Justice.



LEGAL NOTICE

This report was prepared by Battelle Memorial Institute (Battelle) as an account of sponsored research activities. Neither Client nor Battelle nor any person acting on behalf of either:

MAKES ANY WARRANTY OR REPRESENTATION, EXPRESS OR IMPLIED, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, process, or composition disclosed in this report may not infringe privately owned rights; or

Assumes any liabilities with respect to the use of, or for damages resulting from the use of, any information, apparatus, process, or composition disclosed in this report.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by Battelle. The views and opinions of authors expressed herein do not necessarily state or reflect those of Battelle.

Artificial Neural Network System for Classification of Offenders in Murder and Rape Cases

L. J. Kangas

Contracting Officer: J. E. Miles
MSIN: K3-40
P.O. Box 999
Richland, WA 99352
(509) 365-4467

Program Manager: S. W. Martin

October 2000

Prepared for
National Institute of Justice,
Office of Science and Technology

Battelle Memorial Institute
Pacific Northwest Division
Richland, Washington 99352

About the author

Mr. Kangas has BS and MS degrees in computer science and three degrees from the Marine Academy. He has developed several information analysis and diagnostic systems based on artificial intelligence technologies, such as artificial neural networks, fuzzy logic, genetic algorithms, and model based reasoning. These systems include medical diagnostics and monitoring, turbine engine diagnostics, speech recognition, speaker identification, graph data extraction, spectral analysis in environmental sensing, ground penetrating radar signal analysis, forensic handwriting analysis, and crime characterization. He has authored or co-authored more than twenty publications and holds patents on two software applications using artificial neural networks. Mr. Kangas has been at PNNL since 1991. He also lectures artificial neural network and other computer science courses as an adjunct faculty at Washington State University Richland, Washington.

Executive summary

When a serial offender strikes, it usually means that the investigation is unprecedented for that police agency. The volume of incoming leads and pieces of information in the case(s) can be overwhelming as evidenced by the thousands of leads gathered in the Ted Bundy Murders, Atlanta Child Murders, and the Green River Murders. Serial cases can be long-term investigations in which the suspect remains unknown and continues to perpetrate crimes. With state and local murder investigative systems beginning to crop up, it will become important to manage that information in a timely and efficient way by developing computer programs to assist in that task. One vital function will be to compare violent crime cases from different jurisdictions so investigators can approach the investigation knowing that similar cases exist.

The "Artificial Neural Network System for Classification of Offenders in Murder and Rape Cases" project developed two software prototypes that demonstrate algorithms for analyzing and comparing large databases of crime data. The CATCH (Computer Aided Tracking and Characterization of Homicides) and CATCHRAPE software applications analyses homicide data and sexual assault data, respectively. Both applications are similar, although they analyze different databases. In this report, "CATCH" will refer to both applications.

CATCH was developed to provide crime analysts enhanced means for interpreting large crime databases. These databases store information on a large number of crimes, with each case described in a large number of details. The huge volume of information collected during criminal

investigations make it difficult for investigators to compare one crime against others. CATCH, using advanced algorithms running on computers, facilitates the investigators need to compare crime data and to find crimes that have similar patterns.

The development of CATCH was made possible with the HITS (Homicide Investigation Tracking System) database system. Police involved in the infamous Green River and Ted Bundy murder investigations in the State of Washington developed HITS more than ten years ago to enable computer-based analysis of murders. The database now contains several thousand violent crimes primarily from the Pacific Northwest. The HITS database system is similar in the crime details it stores to the VICAP system at FBI and the violent crime database system operated by the RCMP in Canada.

The algorithms in CATCH learn from crime data. HITS provided the required historical crime data for CATCH to learn offender behavior so that new unsolved crimes can be characterized for analysis and comparison.

The CATCH software assists with crime investigations by assessing likely characteristics of unknown offenders, by relating a specific crime case to other cases, and by providing a tool for clustering similar cases that may be attributed to the same offenders. An investigator working with CATCH is working against the collection of crimes in the HITS database. The software speeds up the database access to these crimes, improves the views of the crimes, and organizes the crimes according to similarity.

CATCH is a collection of tools that assist the crime analyst in the investigation process by providing advanced data mining and visualization capabilities. These tools include clustering maps, query tools, geographic maps, chronology maps, etc. Each tool is designed to give the crime analyst a different view of the case data.

The clustering tools are based on artificial neural networks (ANNs). The ANNs in CATCH learns from a subset of existing crime cases (approx. 5000 homicides, 3000 sexual assaults) in the HITS database using parameters describing modus operandi (MO) and signature characteristics of the offenders. The ANN clusters cases onto two-dimensional visual representations, according to the similarity of offender patterns. The proximity of cases within the feature map allows the analyst to identify serial homicides/sexual assaults.

A case being investigated is compared to the solved adjacent cases in the visual representation of the database. A set of one or more cases is analyzed in detail to determine whether a linkage exists between the cases. For example, is it one offender that is responsible for a series of crimes? The investigator goes between the different tools in CATCH and analyses when and where the crimes were committed. The analyst can focus on specific details that he or she believes are important for determining whether there is a connection between the crimes. The analysis work continues using a set of cases where individual cases can be dropped or new cases can be added from the database. All accesses to the database are fast and simple using point and click mouse methodology.

Access to a database system, like the HITS database, requires strict adherence to a query language to retrieve information. This language is slow and difficult to use even for an experienced database professional. SQL is the most common language for accessing a database. CATCH removes the need for an investigator to know this language by dynamically creating the queries in the SQL language on demand, as the investigator accesses the database. The investigator can concentrate on the information in the databases, rather than the access method. The investigator only sees the information in CATCH, not the query language, as in the previous version of software accessing the HITS database.

CATCH characterizes an unknown offender by comparing his or her crime to known offenders of other crimes where the behaviors of the offenders have been similar. The characterization can give some clues to personal information about the unknown offender to expedite the investigation process. This information include, level of education, type of work, physical appearances such as neatness, misuse of alcohol and drugs, etc. The underlying assumption is that two individuals that otherwise behave similarly, may also commit crimes with similar behavior. The validity of this assumption has been proven to some degree by FBI's behavioral science unit and others.

The characteristics of offenders developed in CATCH are hints at best and differs from most of the other information, which are factual and retrieved from the database. With the exception of the characterization tool, CATCH was not designed to modify or generate data describing the crimes in the database. It only organizes the existing facts to provide more meaning to the investigators.

CATCH was developed using visionary information technologies to advance crime analysis. These technologies include methods for analyzing data, visualizing information, and accessing data. Research staff drew from past experience in these technologies and pushed the envelope just a little further by combining these technologies in one application. The application has received considerable attention in the media as being new and innovative.

The investigations of crimes involve analysis of offender behavior. A set of automated technologies can make the work of investigators more efficient, they cannot replace the investigators. Although technologies are becoming more advanced, a software tool should still only be in support of the human in-depth understanding of behavior. The quintessential objective in the research was to facilitate the work of investigators.

The development of CATCH would not have been possible without the collaboration Battelle enjoyed with Washington State Attorney General's Office, Criminal Division. The experience of the crime analysts and the extensive experience they have in operating a database of violent crimes were invaluable to the project. These crime analysts were the drivers that recognized the need to research the possibility of using advanced technologies to make their work more efficient.

CATCH demonstrates the value of incorporating new technologies and methods that can aid the work in crime investigations. The scope of the project was limited to specifically show how the artificial neural network technology can support a multivariate or information fusion approach to finding significant patterns in crime data. Other research and development efforts concentrate on

equally important aspects of crime analysis, such as the use of graphical information systems. Although CATCH was developed as a self-sufficient application and includes some crime mapping and other tools of a complete application for crime analysts, it did not explore or exhaust the possibilities of those tools.

Two versions of CATCH were developed, CATCH and CATCHRAPE for homicide data and sexual assault data, respectively. These applications were deployed at Washington State Attorney General's Office, Criminal Division for evaluation. Although the first versions of CATCH were configured specifically for the HITS database of violent crimes, they can be applied against other crime databases through relatively minor changes in the software.

Project description

The project developed new technology for investigators working with crime analysis. The nature of this work has changed rapidly in recent years. Resources have become increasingly scarce, requiring investigators to work more efficiently. Computers are being incorporated in the investigative work with the expectations that they would automate some work tasks. The reality has been that most computers in the law enforcement arena are only used for storing crime data. Thus, the computer automation in law enforcement has in the last several years concentrated on how to collect and store data. The need for efficiently processing the data has not been addressed to the same extent.

The move of investigative data from file folders to the storage on computers has resulted in the investigators now having more data available than before. This change requires that the investigators be given tools to efficiently data to make use of it.

This project developed information processing technology for large crime databases to solve some of the investigative needs. Databases in law enforcement are growing in size with some of them already having thousands crimes. The growth changes both how the data is accessed and the methodology used in processing the data in investigations. The project demonstrated the use of leading edge technology for efficiently accessing and processing these large databases.

Technology was developed for both accessing data rapidly in large databases through mining and visualization paradigms. Once the significant data have been retrieved from the databases, it also has to be processed to extract pertinent information. Methodology was demonstrated that lets the investigators organize and view the data in tools that enhance the information content. The views of the data are sorted and enhanced with color according to relevance of crime details when comparing several crimes.

The methodology for processing large databases developed in the project is based on recent research that has shown the need for radical changes. Past methodologies fail when data sets required in processing grow large. The new methods unfortunately cause a hurdle of acceptance to overcome. Like with any other major change in technology, this means that the practitioners may not immediately embrace these radical changes, but given time and experience, it is expected that the new methods will be accepted in the information processing communities.

Another significant benefit in moving the investigative work from hardcopy files systems to computers is that information can be shared easier and thus be made available to more investigators, without geographic boundaries. With the increased mobility of the populations,

the serial offenders also commit their crimes over larger geographic regions. Several serial murderers have been known to commit their crimes across state lines and in different corners of the country. Automated recognition of crime patterns gives investigators better means for finding these offenders. There is also value in comparing the pattern of one offender to those of other offenders in other parts of the country. Human behavior tend to repeat and much can be learned from computer applications that focuses on the patterns of criminal behavior.

Scope and methodology

The objective in the design of CATCH was to use artificial neural networks to deduct useful information and present the information in a meaningful way to the investigators. The sheer volumes of data residing in large crime databases make it near impossible for individuals to process the data for significant patterns such as those exhibited by serial offenders. Computers on the other hand don't suffer from the same information overload. They also have a speed advantage. Computers have the disadvantage that they aren't as good as people in pattern recognition. It was thus an imperative objective in the design of CATCH to reduce the number of crimes that the investigators had to consider at any one time. The set of crimes with a potential significance to a case being investigated also had to be presented to the investigators with some added value compared to views of raw data only.

There are three main concepts in CATCH that provides value to the investigators. These are the clustering algorithm, database mining, and database visualization. The clustering algorithm organizes the database according to similar patterns in the crimes. The database mining allows

the investigator to retrieve crimes that may have significance without having to read through all crimes. The database visualization makes it possible to understand the content of the database at a high level.

Clustering algorithm

CATCH uses artificial neural networks (ANNs) for analysis. The benefit of ANNs is often described by means of their information (sensor) fusion capabilities. Information fusion is the process of extracting information from a several data sources in parallel. More information can frequently be gained by this approach compared to processing each data source individually. Another benefit with some ANNs is their ability to extract non-linear information from data.

The clustering algorithm in CATCH is based on self-organizing maps (SOMs). These networks are also called self-organizing feature maps or Kohonen networks after the inventor Professor Teuvo Kohonen (Kohonen 1997; Kaski 1997).

The SOMs belong to the unsupervised neural network class, meaning that the network is not provided any labels that describe the data vectors during a learning phase. Instead, the SOM organizes data vectors into clusters of similar data in regions on a two dimensional map. Two dimensions provide a convenient visual representation, though it is not a requirement.

The HITS Unit staff at the Attorney General of Washington, Criminal Division, use standard forms to record the large number of details describing each crime, which are then entered into the HITS database. CATCH processes these crime details and generates data vectors for numerical analysis. Each data vector includes more than 200 details of each crime.

The SOM in CATCH has 4096 cells organized as a 64 by 64 grid (see Exhibit 1 below). The learning phase assigns each crime to exactly one of these cells. The specific cell to which each crime is assigned is based on a clustering algorithm applied during a learning phase. Similar crimes are placed in closer proximity to each other. Identical or nearly identical crimes may be placed in the same cell. Some cells may not be assigned any crimes during the learning phase but these cells may be assigned new crimes as they are entered into the database between

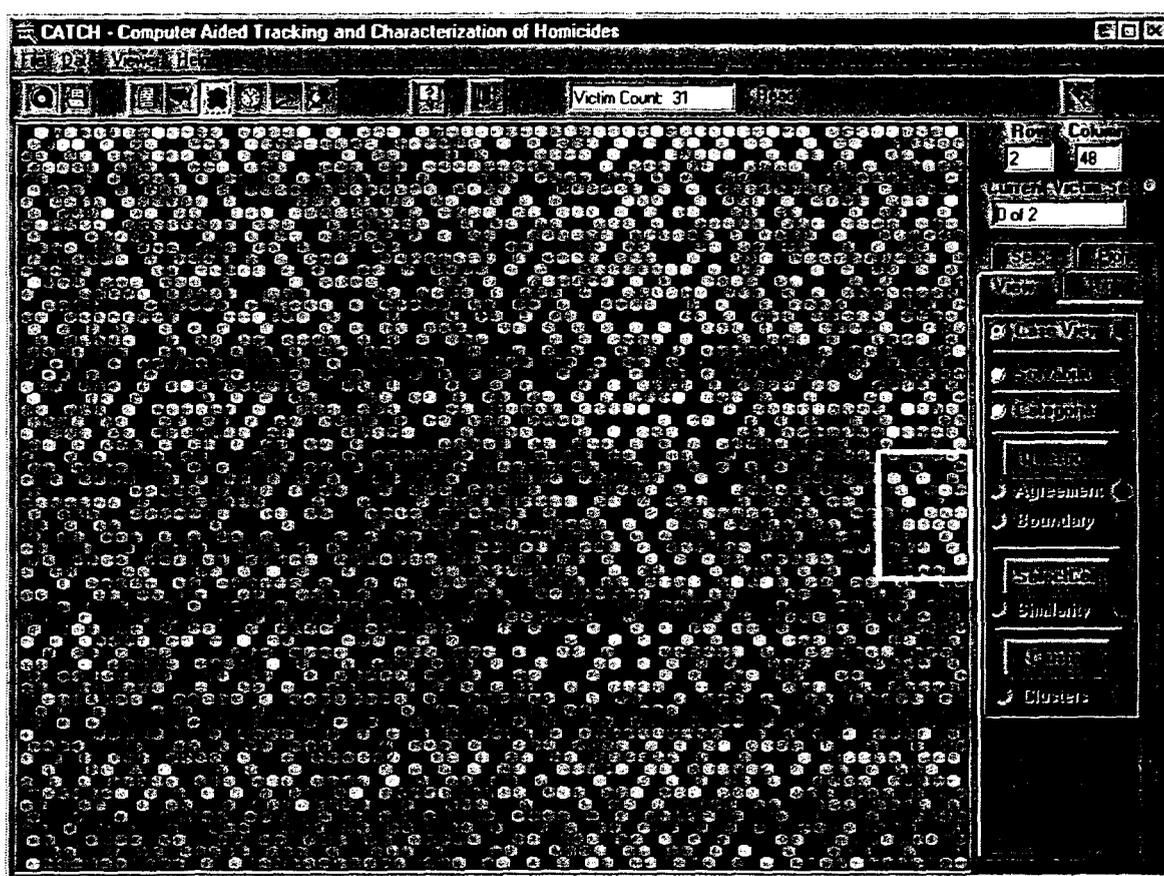


Exhibit 1. The self-organizing map in the exhibit represents about 5,000 murders in the HITS database. Each of the cells in the 64 by 64 map typically contains eight or less crimes. The black cells contain no crimes. The lighter the cell color the more crimes are in the cell. (The cells are colored in different shades of green in the application.) The overlaid light rectangle contains light colored cells (in yellow) that are selected into a current set of crimes being analyzed. (The example set of clustered crimes in the exhibit is the Green River Murders believed to be by one serial offender).

retraining of the SOM. The SOM should periodically be retrained when a sufficient number of new cases are added to the database to take advantage of all the crime data available.

Database mining

The tools in CATCH are of two types. First, there are database mining tools to give the crime analyst a better understanding of the content of the database. Second, there are tools that let the analyst retrieve and compare specific crimes.

The self-organizing map is like a window into the database. Each crime in the database has a location on the SOM and the clusters on the SOM link together similar crimes in the database. Thus, the database can be mined for related crimes through the SOM. These mining tools which use the SOM include a search tool that lets the analyst select a combination of crime details and see where on the SOM there are crimes for which these details hold true (see Exhibits 2 and 3 below). Another tool allows the analyst to select one crime case and see where in the SOM there are other similar crime cases based on any combination of the details describing the crimes. CATCH allows the analyst to add to, remove from, or crop the current set of crime cases by selecting areas of cells in the SOM, while mining the database.

The “starmap” of crimes in CATCH is shown in Exhibit 4. This representation of all crimes in the database is a three-dimensional cube, where the data vectors describing the crimes have been reduced down to three eigenvalues. The cube is viewed by selecting any two of the dimensions. Although a significant amount of information is lost when high dimensional data is reduced to a

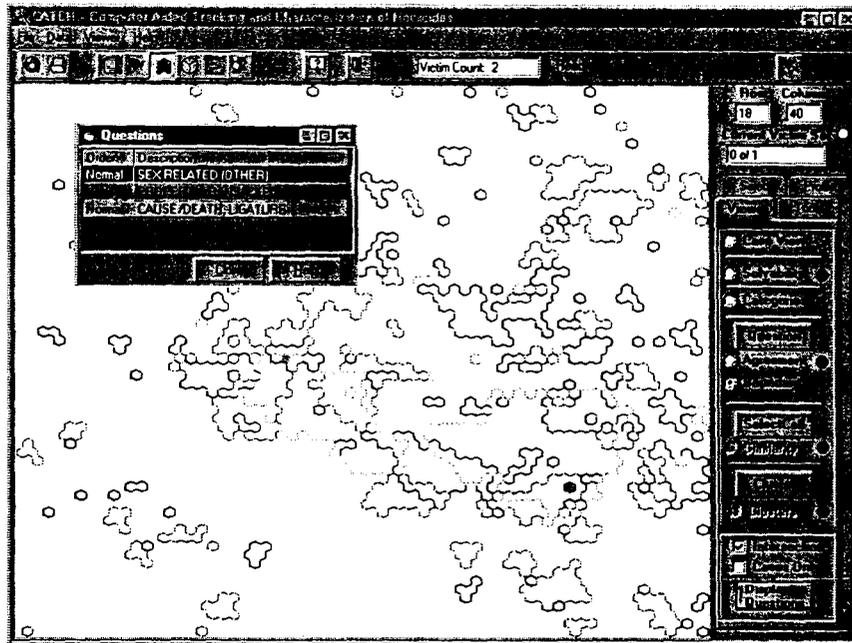


Exhibit 2. The SOM is overlaid by boundaries around areas of common crime details. The small window shows which details are selected and the color coding of the boundaries. The user can select crimes that are in the unions of the bounded areas, shown as light colored cells in the example.

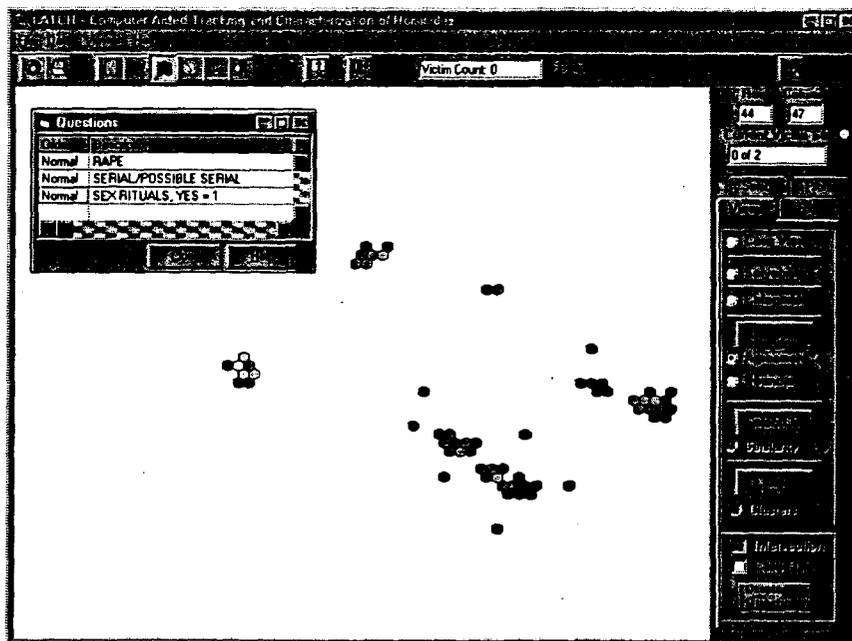


Exhibit 3. The depicted tool emphasizes cells containing crimes for which all selected details correctly describe the crimes. The cells in the SOM are colored lighter according to the correlation of the selected crime details, i.e., lighter cells have higher correlation with the selected crime details.

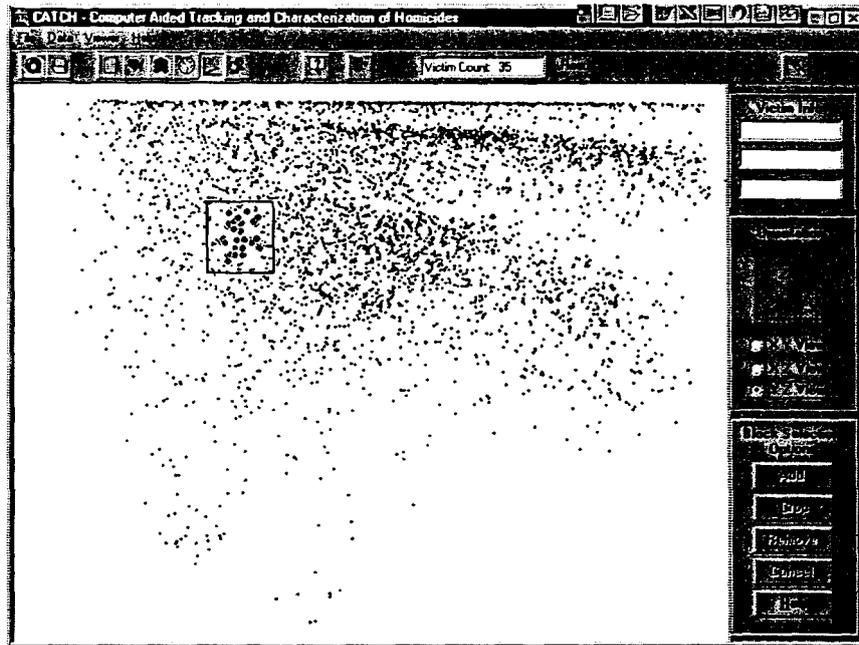


Exhibit 4. The exhibit shows all the crime data vectors as points in a three-dimensional eigenspace. The cube of crimes is viewed in any two of the three dimensions. This cube of crimes gives an alternate view of the clusters and structure of the crimes in the database. Similar crimes form denser areas of “stars” in the cube. The highlighted crimes, within the overlaid rectangle, are selected into the current working set.

few dimensions, the visualization of the data still conveys significant structure of the data in the database. The user can select volumes in the cube to retrieve, remove, and crop crimes to and from the current set of crimes being analyzed.

The geographic map in CATCH is shown in Exhibit 5 with crimes placed as pins at the locations where they were committed. This map also allows the user to select an area and retrieve all the crimes in that area or the user can crop or remove crimes from the current working set of crimes.

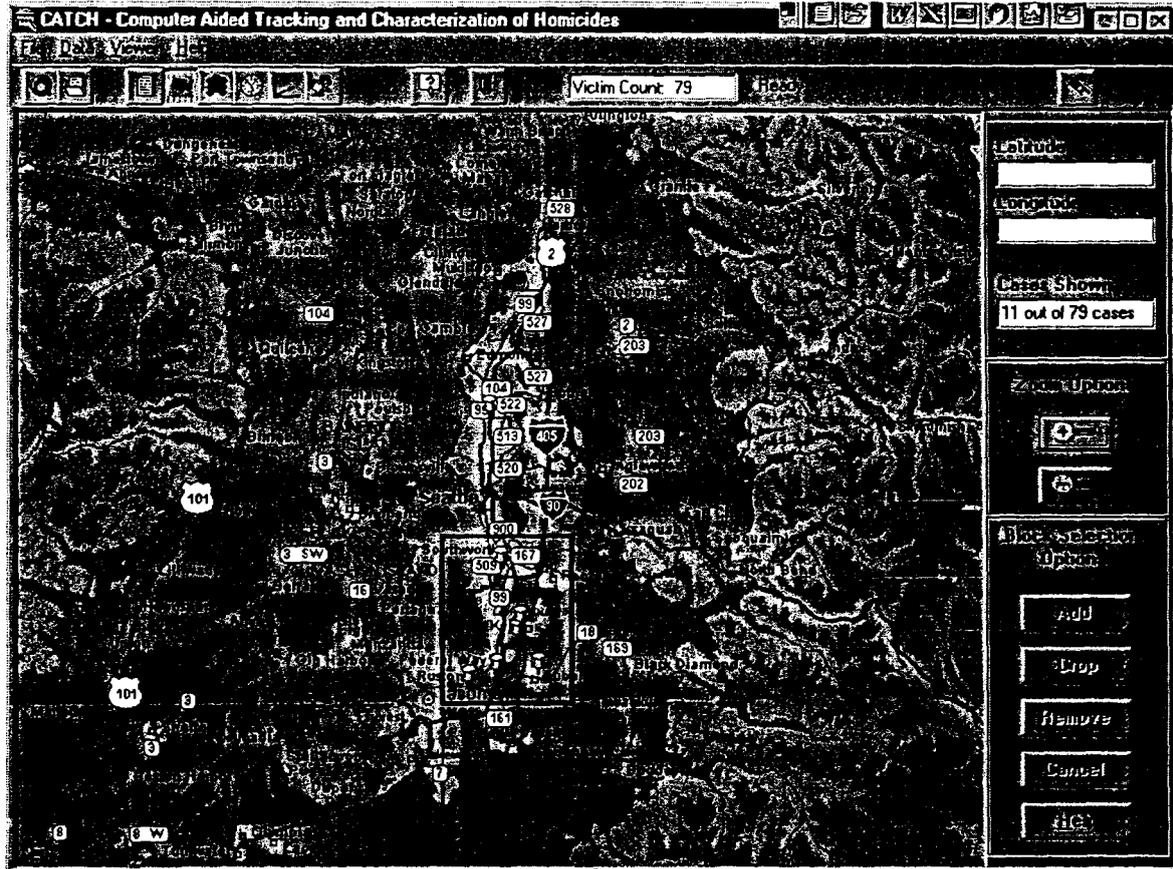


Exhibit 5. The geographic map tool places the current set of crimes on the map as pins (see examples in the rectangle). The user can select pins to view additional information about specific crimes.

The tools described above and some additional tools, e.g., a time line tool, allow the crime analyst to retrieve crime data from the database without having to use queries. CATCH automatically generates SQL queries to retrieve requested information from the user's interaction with graphical representations of the data. Thus, although CATCH allows the use of queries in a specific query tool, it has been designed so that a user is removed from having to work with queries when mining the database.

Database visualization

While the data mining tools are used for rapidly focusing on a set of crimes that may be related, the data visualization tools become the priority for more in-depth analysis of crime data. Some graphical data visualization capabilities of CATCH have been partially described above with the data mining tools. This section describes a few of the tools that allow the users to view, analyze, and compare details describing different crimes. (Because CATCH processes sensitive information it is necessary to restrict the images of these tools in the exhibits).

Most of the data visualization tools in CATCH show the crime details in grids that are enhanced by color and order of significance. The color enhancement in the grids is used to give the user improved perception of the data without having to focus on numerical values. Typically, grid values representing crime details are lighter in color if the crime detail has a higher numerical value or if the crime detail holds true for a specific crime. The grids can also be sorted to bring more significant details to the top of the grids. The significance of each detail is dynamically computed in the sorting algorithms.

Analytical software process

CATCH provides the crime investigator with a set of tools, some of which are described above. These tools enable the investigators to efficiently view crime details and compare crimes against each other. The objective in working with CATCH is to find a set of related crimes. This set of crimes may be related because they appear to be committed by the same offender.

The investigator in the analytical process uses point-and-click methods that generate SQL queries to retrieve an initial set of one or more crimes from the database. The investigator moves

between the tools with the set of crimes, which are analyzed from different perspectives. One tool may suggest that a specific crime isn't related to the others. Another tool may suggest that there could be other crimes in the database that should be considered in the set. Each of the tools allows the investigator to remove crimes or add new crimes from the database to the current working set. The investigator stops when a set of crimes is found that may be related. This set can then be the focus of additional work using evidence that is not stored in the database.

Analysis and discussion

CATCH was developed to identify serial offenders by recognizing that serial offenders tend to repeat certain aspects of their crimes. Because the neural network algorithm clusters similar data vectors, we expect the crimes by the same offenders to be clustered close together. The graphs in Exhibits 6 and 7 below show the summary of distances found between any pair of crimes committed by the same known offenders, for murders and sexual assaults, respectively. Distances are measured as the number of cells between two crimes on the self-organizing map. A distance of zero indicates that both crimes are in the same cell; a distance of one indicates that the two crimes are in adjacent cells, etc.

The results shown in Exhibit 6 are based on 189 serial murders committed by 81 known offenders. The graph shows that 50 percent of serial murders by the same offenders are within 15 cells from each other. The results shown in Exhibit 7 are based on 412 serial sexual assaults committed by 154 known offenders. The graph shows that 50 percent of serial sexual assaults by the same offenders are within 8 cells from each other.

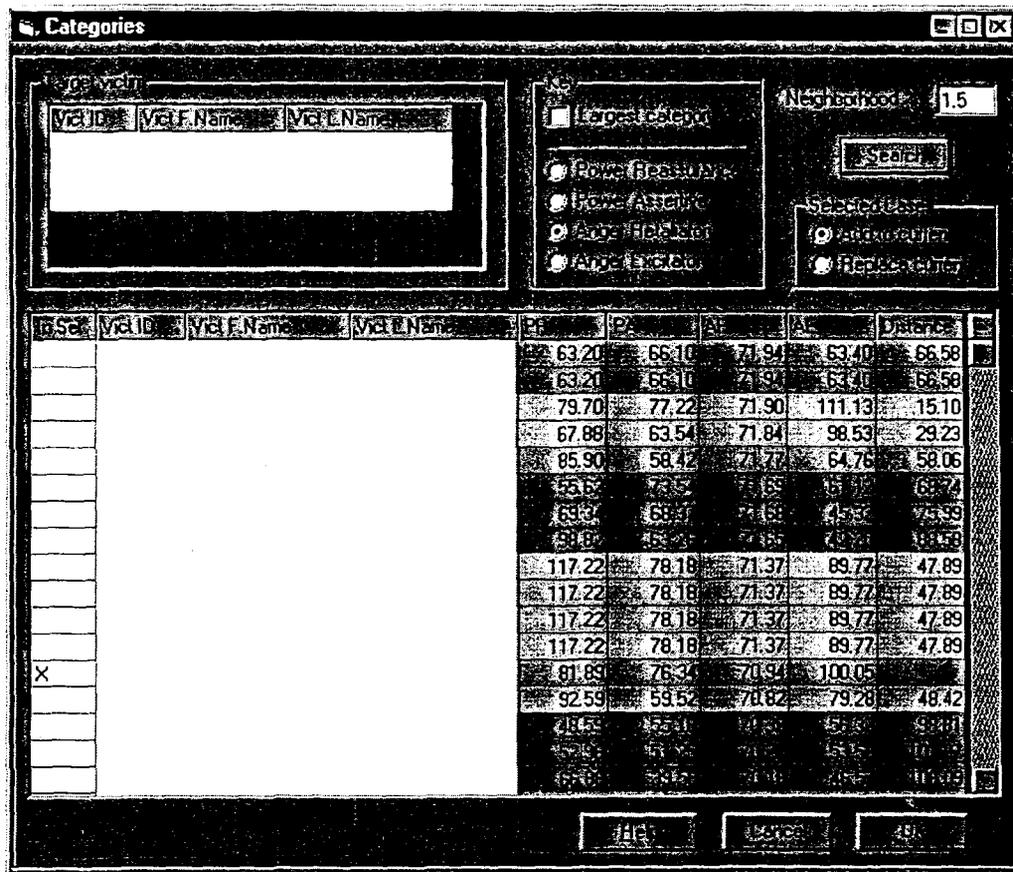


Exhibit 6. The solid line in the graph shows the probability of finding two murders by one serial murderer n number of cells apart. Fifty percent of the related serial murders are found within 15 cells of each other. The dashed line in comparison shows the distance between the same murders, as they would appear if randomly placed into cells in the self-organizing map. The confidence is greater than 99% against these two probability distributions having the same mean (two-tailed t-test).

Although the graphs above show significant clustering capabilities for serial crimes by the same offenders, there are still possibilities for optimizing the algorithm. A weighting scheme was applied to the crimes details that determine how much each detail contributed to clustering together similar crimes. The weighting scheme applied to the results presented here was based on experience investigators intuition. This approach was selected to establish a benchmark.

Category Questions

Victim Name: [] Offender Name: []

Q#	Descriptor	P1	P2	A1	A2	Tot	P1	P2	A1	A2	Tot
149a	UNUSUAL RITUAL, YES = 1	94.60		40.54	67.57	202.71	94.60		40.54	67.57	202.71
161	BLINDFOLD, YES = 1	73.17		73.17		146.34	73.17		73.17		146.34
164	VICT. RE-DRESS, YES = 1	57.13		57.13		114.26					
142	DIS PHONE/ALARM/UTILS	35.48	35.48	35.48		106.44	35.48	35.48	35.48		106.44
140a	STAGING/POSING, YES = 1					73.17					73.17
173b	BLUNT FORCE, EXTREME =										
139a	BODY DISP., OPENLY										
162	FACE COVERED, YES = 1										
186c	UN. ASSAULT, OFFENDER										
176	BLUNT WOUNDS, 20 = 1										
174	NO. OF STABS, 25 = 1										
143b	PROPERTY, DISTURBED										
172c	CAUSE/DEATH, BLUNT FOF										
136	RECOV. AT VICT. RESIDEN										
163	VICT. CLOTH, DRESSED = 1										
190	SEX ASSAULT, YES = 1										
128	DIST. ICS TO RECOVERY										
191a	TYPE SEX, VAGINAL										

Exhibit 7. The solid line in the graph shows the probability of finding two sexual assaults by one serial rapist n number of cells apart. Fifty percent of the related serial sexual assaults are found within 8 cells of each other. The dashed line in comparison shows the distance between the same sexual assaults, as they would appear if randomly placed into cells in the self-organizing map. The confidence is greater than 99% against these two probability distributions having the same mean (two-tailed t-test).

Another approach to the weighting scheme worth considering is to compute the weighting based on the known serial crimes in the databases. This would ensure that the solved serial crimes cluster yet closer together and the assumption can be made that also unsolved serial crimes would cluster tighter together. The assumption will hold if the behaviors of offenders in unsolved crimes are similar to the behaviors of offenders in solved crimes.

Conclusion and implication of findings

Version 1 of CATCH was developed with funding from National Institute of Justice to provide crime analysts enhanced means for interpreting large databases of crime data. These databases store a large number of crimes and each crime is described in a large number of details.

Washington State made CATCH possible by providing the project with access to the HITS database system. The HITS database of violent crimes was initially funded by National Institute of Justice and developed by the state circa fifteen years ago. It has since been maintained by the state.

Battelle developed CATCH in collaboration with the Washington State Attorney General's Office (AGO), Criminal Division, using the state as a test site. Investigators at AGO made some initial evaluations of the first installed versions of both CATCH and CATCHRAPE. The evaluation was terminated due to both retirement and attrition of staff among the principal collaborators at the state.

CATCH is currently not operational against the HITS database due to a change in the database system at the state. A complete rewrite of the HITS database system was completed after the project was completed. An estimated 95% of those changes at the state were accounted for in a CATCH version 1.1. A relatively small effort would be required to align CATCH with the latest HITS database system operating at the state and make it operational again.

The preliminary evaluation of CATCH suggests that it provides considerable value to the crime analysts. Without addressing the value of the analytical tools in CATCH, it is obvious that the data access and data viewing capabilities have improved considerably compared to the previous application used against the HITS database. It was also recognized in the development of CATCH that there are some additional tools that would add considerable value to the crime analysts. First implementations of some of these tools were made in version 1, although these were not within the initial scope.

The first version of CATCH was concentrated on researching the value of using artificial neural networks to cluster similar cases. Both the clustering algorithms and visualization tools in CATCH have the potential to add considerable value to crime analysts. A next version of CATCH would incorporate additional tools that have been identified from the current research and development. The production version of CATCH would include these new tools that would provide the crime analysts with a more complete suite of tools.

Endnotes

MEDIA EXPOSURE

- “Artificial Intelligence Helps Solve Serial Crimes”
Bill Siuru, Law and Order, June 2000.
- “C.A.T.C.H. Computers and Artificial Intelligence Track Down Serial Killers”
ComputerEdge, April 14, 2000.
- “CATCH”
Breakthroughs. Science, Technology, Innovation
Pacific Northwest National Laboratory, Spring 2000.
- “CATCH Program”
TECH beat, Fall 1999.
- “New Way to CATCH Bad Guys”
<http://www.govtech.net/onlineservices/news/newspage.shtm?story=June181999NewWaytoCATCHBadGuys>, June 18, 1999
- “Catch as CATCH can” MIT's Magazine of Innovation Technology Review, March/April 1999

- "Buying Computers and Software" (CATCH software described)
Law Enforcement Technology, April 1999
- "CATCHing Criminals"
http://newsweek.com/nw-srv/issue/15_99a/tnw/today/cs/cs01th_1.htm, April 9, 1999
- "Thinking Like a Serial Killer"
http://www.wired.com/news/print_version/technology/story/19940.html?wnpg=all,
June 1, 1999
- "David Lawrence talking with Lars Kangas, Sr. Research Engineer at Pacific
Northwest National Labs"
<http://www.audiocast.net/geekradio>, June 7, 1999

PRESENTATIONS

L.J. Kangas, K.M. Terrones, R.D. Keppel, R.D. La Moria. 1999. "Computer Aided Tracking and Characterization of Homicides & Sexual Assaults (CATCH)." Presented at SPIE AeroSense '99, Orlando, FL, April 5-9.

L.J. Kangas, K.M. Terrones. 1999. "Data Mining and Visualization for Crime Analysis: CATCH." Presented at 8th Advanced Information Processing & Analysis Steering Group, Washington DC, March 23-24.

REFERENCES

1. S. Kaski, *Data Exploration Using Self-Organizing Maps*, Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 82, Espoo, 1997.
2. T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Berlin Heidelberg, Second Edition, 1997.

BIBLIOGRAPHIES

- G. Copson, "Coals to Newcastle? Part 1: A Study of Offender Profiling: Police Research Group Special Interest Series," Paper 7, Home Office Police Department, London, 1995.
- G. Copson, R. Badcock, J. Boon, and P. Britton, "Articulating a Systematic Approach to Clinical Crime Profiling," *Criminal Behaviour and Mental Health*, 1997.
- J. E. Douglas, A. W. Burgess, A. C. Burgess, and R. K. Ressler, *Crime Classification Manual*, Lexington Books, NY, 1992.
- V. J. Geberth and R. N. Turco, "Antisocial Personality Disorder, Sexual Sadism, Malignant Narcissism, and Serial Murder," *Journal of Forensic Sciences*, Vol. 42, No. 1, pp. 49-60, 1997.
- V. J. Geberth, *Practical Homicide Investigation: Tactics, Procedures, and Forensic Techniques*, CRC Publishing, Boca Raton, Florida, Third Edition, 1996.
- R. D. Keppel and J. P. Weis, "Time and Distance as Solvability Factors in Murder Cases," *Journal of Forensic Sciences*, Vol. 39, No.2, pp. 386-401, 1994.
- R. D. Keppel, *Signature Killers*, Pocket Books, NY, 1997.
- R. D. Keppel, "Signature Murders: A Report of Several Related Cases," *Journal of Forensic Sciences*, Vol.40, No.4, pp. 658-662, 1995.
- R. D. Keppel, *The Riverman: Ted Bundy and I Hunt the Green River Killer*, Pocket Books, NY, 1995.

Appendix

CATCH Software Manual

CATCH: Computer-Aided Tracking and Characterization of Homicides

Software manual

Version 1.0

CATCH is developed by Battelle Memorial Institute for National Institute of Justice in collaboration with Attorney General of Washington, Criminal Division. It is a prototype as a proof of principle concept of how the application of Artificial Neural Network (ANN) technology can aid in analysis of serial sex murder cases. This is not a commercial-ready software product.

DISCLAIMER

This material was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the United States Department of Energy, nor Battelle Memorial Institute, nor any of their employees, MAKES ANY WARRANTY, EXPRESS OR IMPLIED, OR ASSUMES ANY LEGAL LIABILITY OR RESPONSIBILITY FOR THE ACCURACY, COMPLETENESS, OR USEFULNESS OF ANY INFORMATION, APPARATUS, PRODUCT, SOFTWARE, OR PROCESS DISCLOSED, OR REPRESENTS THAT ITS USE WOULD NOT INFRINGE PRIVATELY OWNED RIGHTS.

Table of Contents

1.0	Introduction	1
2.0	Getting Started.....	1
2.1	Checking for new cases	1
3.0	Main Screen.....	2
3.1	Proceeding From The Main Screen	2
3.1.1	Selecting Cases.....	2
3.1.2	Viewing Case Information	3
4.0	Query.....	4
4.1	SQL Tab.....	4
4.2	Defined Tab	5
4.3	Point & Click Tab	5
5.0	Examine Case.....	6
6.0	List Viewer.....	6
7.0	Map Viewer.....	7
7.1	Navigating on the Geographic Map.....	7
7.2	Adding/Removing Cases	8
8.0	Self-Organizing Map Viewer	9
8.1	Navigating On The SOM	10
8.1.1	Moving the Mouse.....	10
8.1.2	Selecting Cells.....	10
8.1.3	Selecting Blocks of Cells	10
8.1.4	Right Clicking on a Cell.....	10
8.2	Control Panel Tabs.....	10
8.3	Select Tab.....	11
8.4	Profile Tab	12
8.4.1	Closest Category.....	12
8.4.2	Category Weighting	14
8.5	Views Tab	15
8.5.1	Case View.....	15
8.5.2	Solvability	15
8.5.3	Categories.....	16
8.5.4	Boundary and Agreement.....	16
8.5.5	Choosing Questions.....	16
8.5.5.1	Order and Tolerance	17
8.5.6	Boundary Display.....	18
8.5.7	Agreement	19
8.5.8	Options Available in Boundary Display.....	20
8.5.9	Cluster	21
8.6	Tools Tab	23
8.6.1	Search	24
8.6.2	Case	25
8.6.3	Case Data.....	26
8.6.4	Cell Values	26
9.0	Time Line Viewer	27

10.0	Principal Components Viewer.....	29
11.0	Characterization.....	30
11.1	How to Use the Characterization Screen	32
11.2	Report Format	33
12.0	Importing and Exporting	33
13.0	Admin.....	34
14.0	Exiting	34
	Appendix A: Theory	35
A.1	Glossary	35
A.2	Determining where a case will be in the SOM map	38

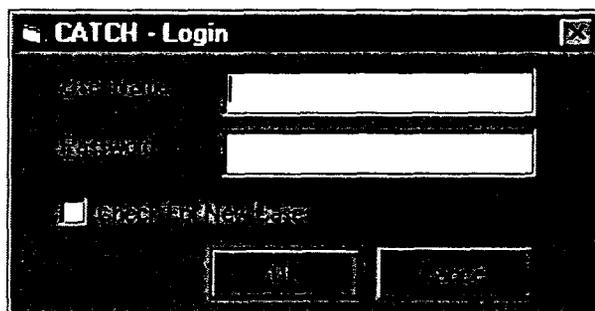
1.0 Introduction

The functionality that is available in this version of CATCH is:

- Searching the HITS database for cases by name, victim ID, or certain characteristics.
- Viewing the report information for selected cases.
- Viewing cases by location.
- Viewing cases on a self-organizing map in which you can:
 - View case information for selected cells.
 - View weight information for selected cells.
- Search for a victim or suspect.
- Choose a set of questions and display the corresponding areas in the map, by two methods (Agreement and Boundary).
- Viewing cases on a time line.
- Viewing cases on a principal components graph.
- Creating reports to view or print.
- Importing and exporting cases between CATCH and CATCHRAPE.

2.0 Getting Started

- Start the program by double-clicking the CATCH icon. The login window will then be displayed (see below).
- Enter your UserID and your Password.
- Click on the checkbox if you want CATCH to search for new cases. See below for a complete explanation.
- Click OK.
- The main screen will then be displayed.



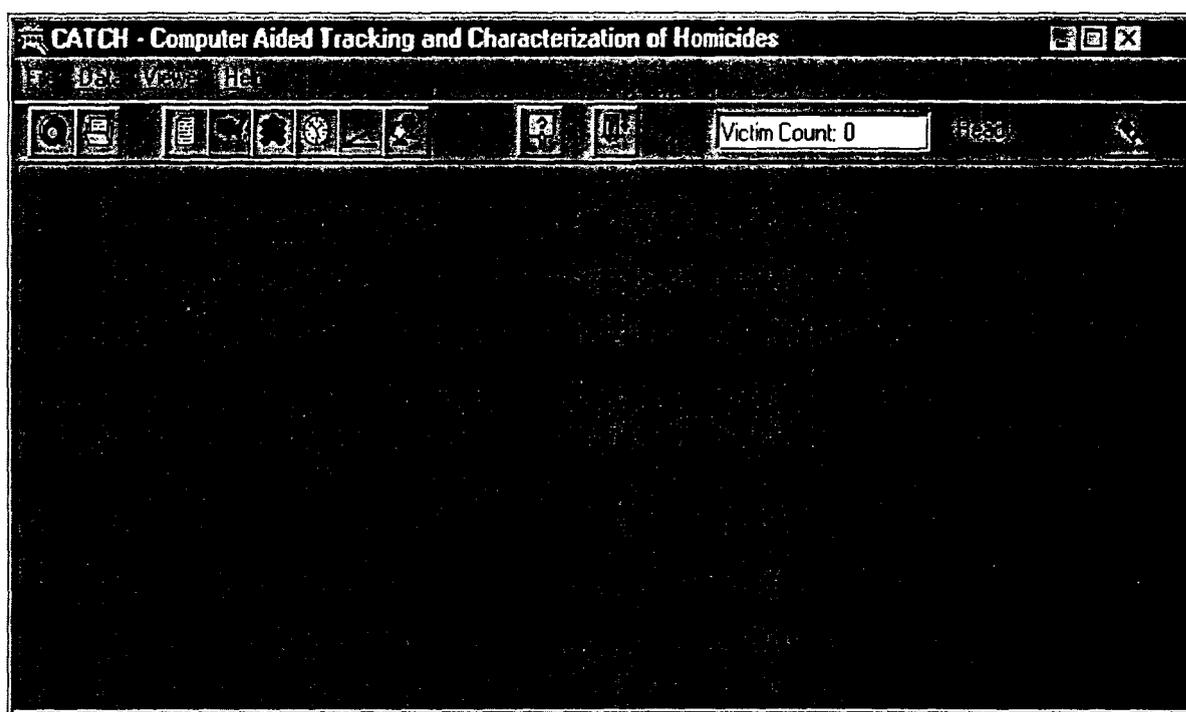
2.1 Checking for new cases

If the Check For New Cases checkbox is checked when OK is pressed, the CATCH program will search the HITS database for all recently entered murder cases. The program will then include those cases into the neural network. This procedure will take time to search the entire database and perform calculations, therefore checking this box is not advised at every login.

3.0 Main Screen

The main screen has four menus at the top: **File**, **Data**, **Viewer**, and **Help**. Underneath the menus is the toolbar. The buttons in the toolbar are: **Query**, **Examine Case**, **List Viewer**, **Map Viewer**, **Self-Organizing Map Viewer**, **Time Line Viewer**, **Principal Components Viewer**, **Characterization**, **Help**, and **Exit**. Holding the mouse over any button will cause a box to appear with the name of that button.

To the right of these buttons is the **Victim Count**, which contains the number of cases in the current victim set. On the far right is the **Reset** button, which when clicked will return you to the main screen and set the victim count to zero.



3.1 Proceeding From The Main Screen

3.1.1 Selecting Cases

From the main screen, you can go in many directions. First you will want to query the database and retrieve cases that you will work with. Retrieving cases can be done at any time during you session and can be done in a variety of ways. For instance, you can:

- **Query** the database for cases using SQL queries or pre-defined queries
- Select a case by victim name using **Examine Case**.

- Select cases by a geographic location in the **Geographic Map**.
- Select cases on the **Time Line**.
- Select cases, by a variety of methods in the **Self-Organizing Map (SOM) View**.
- Select cases in the **Principal Components View**.

No matter how and when you select cases, those cases will become part of the *current victim set*. The current victim set will be mentioned throughout this document and can be thought of as the set of cases that you are working with at one time. Cases can be added and/or removed from the current victim set by many different methods. These methods will be discussed later.

3.1.2 Viewing Case Information

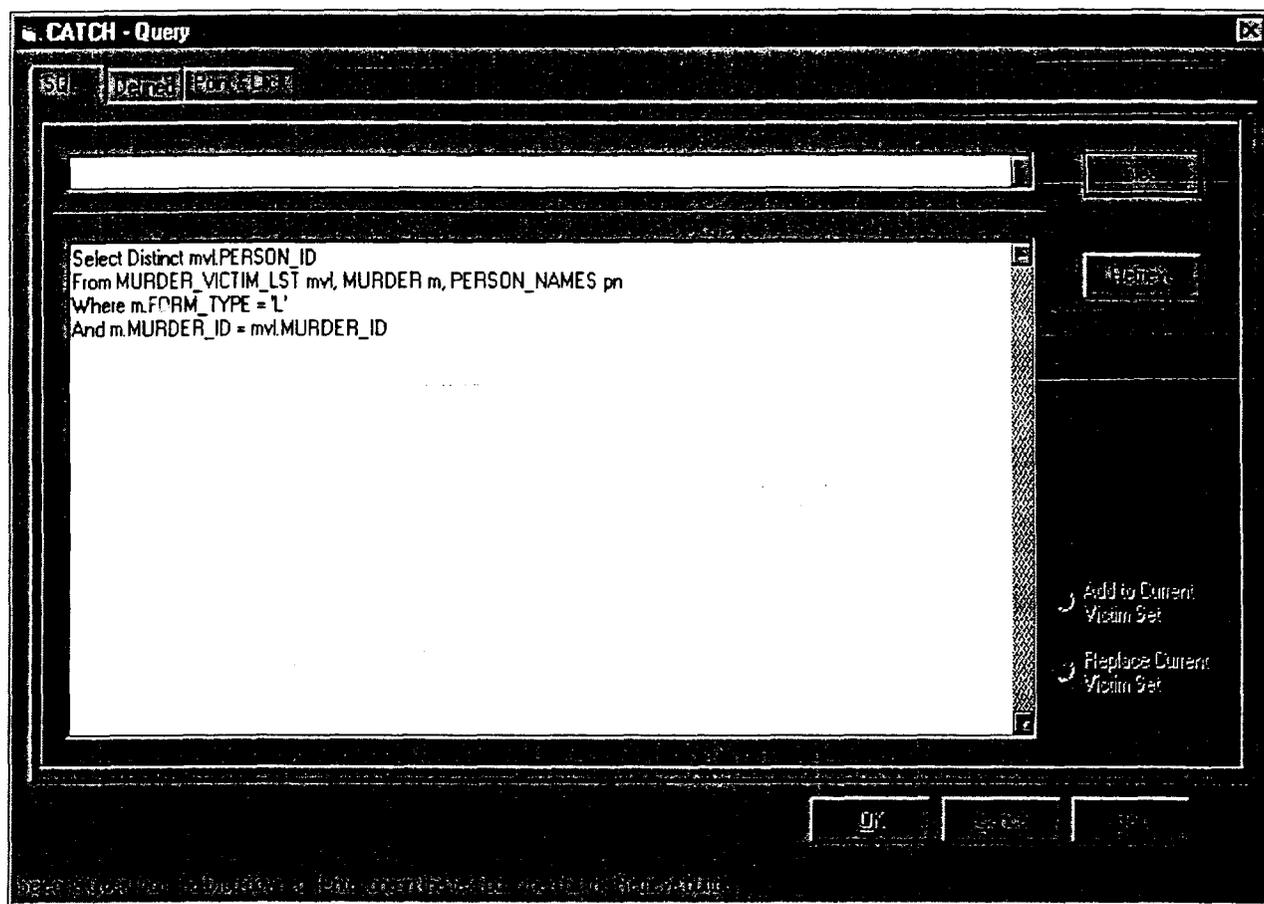
Information associated with the cases in the current victim set can be viewed in several ways. These views will be discussed in detail later, but the following is a list of possibilities:

- The **List Viewer** displays case information in a grid form.
- The **Case View** displays case information in report form.
- The **Geographic Map Viewer** plots cases according to incident sites.
- The **SOM Viewer** displays case and neural network information in a variety of ways.
- The **Time Line** plots cases according to incident dates.
- The **Characterization** displays characterizing case information in report form.

4.0 Query

Enter the query screen either by choosing **Query** from the **Data** menu or by pressing the **Query** button in the toolbar. A screen will appear that looks similar to the exhibit below. QUERY will allow you to query the database for cases using SQL queries or pre-defined queries. You can add one or multiple victims to the current victim set, or replace the current victim set.

There are three tabs in this screen: **SQL**, **Defined**, and **Point & Click**. You can move between tabs by clicking once on the tabs. Under each tab, there is a window with a small arrow to the right. This is referred to as a combo box. By clicking on the arrow, a list will appear and you can then choose an item in this list. Any item that is chosen will appear in the large window below the combo box. This large window is called the SQL window. The SQL code is written in this area and can be edited and/or cleared.



4.1 SQL Tab

The SQL Tab allows you to write your own queries using the SQL language. This requires a familiarity with SQL and the design of your database. The procedure is as follows:

1. From the combo box select one of the four options. The first lets you search for a victim with a last name similar to one you type in. The second lets you search for a victim with a certain last name. The third lets you search for a victim by the first few letters of their last name. The fourth lets you search by the victim's ID.
2. When the line you selected appears in the SQL window, replace the last name, the letters, or the victim ID with the one you are searching for.
3. Click **Retrieve** to search. The number of cases retrieved will appear in the lower left corner of the screen. The cases are *not* in the current victim set yet.
4. To perform a new search, click **Clear** and choose another option from the drop-down menu.
5. Choose either **Add to Current Victim Set** or **Replace Current Victim Set** and press **OK** to exit. Now the cases will be in the current victim set and they will be displayed in the **List Viewer** (see below for more details).

4.2 Defined Tab

This tab allows you to choose from a list of pre-defined queries. Follow the steps below.

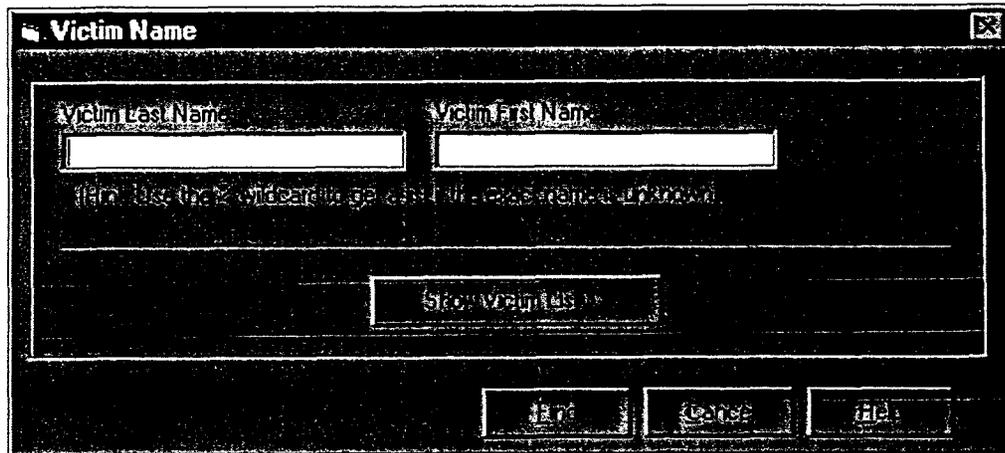
1. Choose one of the four options from the drop-down menu at the top.
2. If you choose to search by hair color, replace 'RED' in the SQL window with the hair color you are searching for.
3. If you choose the last option, you can replace 'NO' with 'YES' in the bottom box to search for victims who were not blindfolded.
4. Click **Retrieve** to search.
5. To perform a new search, choose another option from the drop-down menu.
6. Choose either **Add to Current Victim Set** or **Replace Current Victim Set** and press **OK** to exit. Now the cases will be in the current victim set and they will be displayed in the **List Viewer** (see below for more details).

4.3 Point & Click Tab

This functionality is not available at this time.

5.0 Examine Case

Enter this screen by selecting **Examine Case** from the **Data** menu or by clicking the **Examine Case** button on the toolbar. **Examine Case** will allow you to add cases to the current victim set by searching for a victim's name. The procedure follows:



1. Enter either the first or last name of the victim (or both). Use the % symbol to search by the first few letters of the name. Or press **Show Victim List** to view a list of all of the murder victims in the database.
2. Press **Find** to begin searching.
3. A new window will appear showing a list of all names matching the specified search criteria (see below).
4. In this window, select a name from the list by clicking it with the mouse.
5. Select **Add to Current Victim List** or **Replace Current Victim List**.
6. Press **OK**. The cases are now in the current victim set and will be displayed in the **List Viewer** (see below for more details).

6.0 List Viewer

Enter this screen by choosing **List** from the **Viewer** menu or by clicking the **List Viewer** button on the toolbar. It is also entered automatically after altering the victim set in the **Query** screen or the **Examine Case** screen.

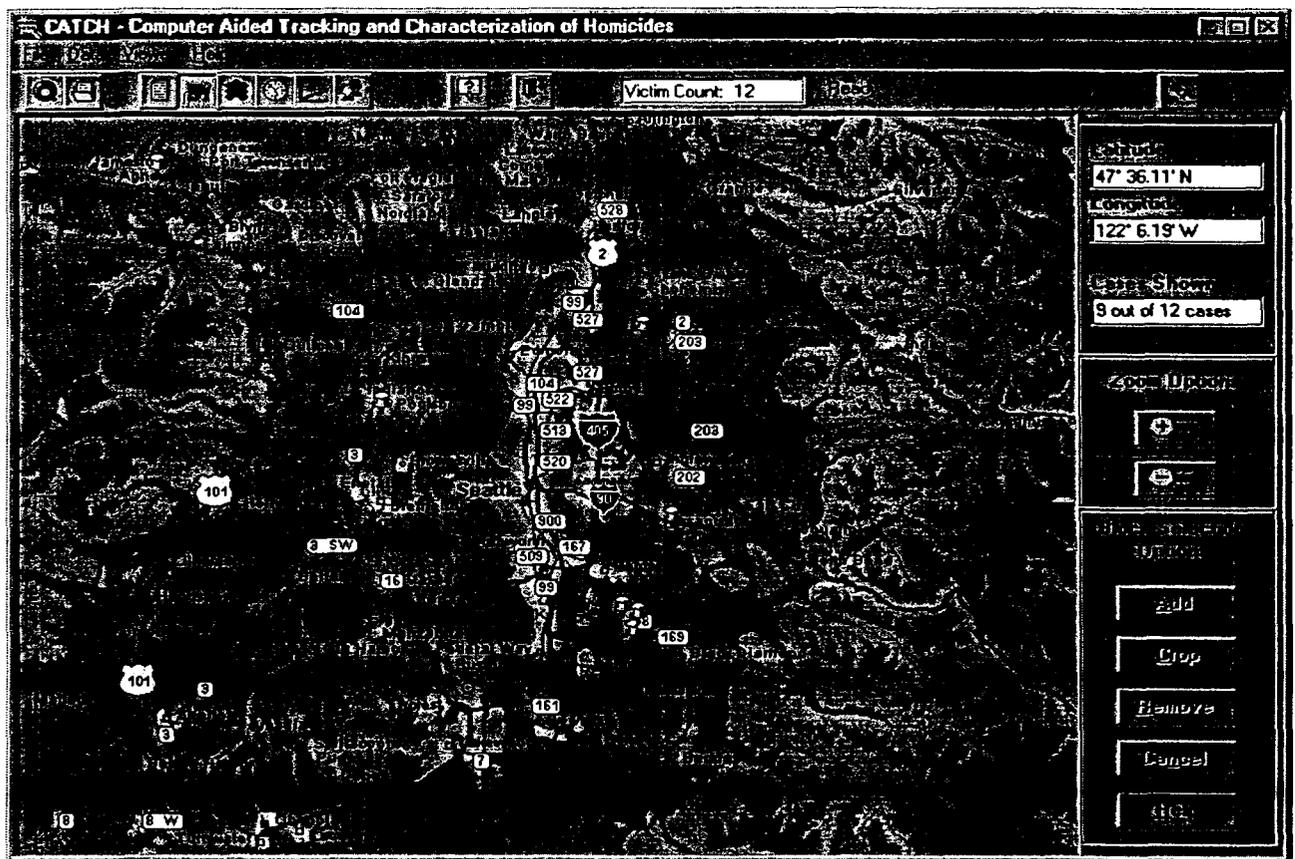
The information for the cases currently in the current victim set will be displayed. The information is organized into different categories which are shown on the tabs at the top of the screen. Click on the different tabs to view the information.

Columns can be resized by placing the mouse between the column headings until you see an arrow pointing left and right. While holding the mouse button down, drag the arrow left or right to resize the column.

A case can be removed from the current victim set in the list viewer. First, click on the row that contains the case. Then, click the **Delete** button at the bottom of the screen. This will remove the case from the grid and from the current victim set.

7.0 Map Viewer

Enter this view by choosing **Map** from the **Viewer** menu or by clicking the **Map Viewer** button on the toolbar. This view displays the cases in the current victim set, represented by a thumbtack, according to geographic location. See below.



7.1 Navigating on the Geographic Map

Resting the mouse over one of the thumbtacks causes a box to appear with the victim ID, name, latitude and longitude of the case.

Double-clicking on any case will bring up a report summarizing the information for that case. The information is organized into different categories. Click on the tabs at the top to view different information. See **How to Use the Characterization Screen** (Section 11.1) below for more information about viewing and printing the report.

At the top of the right hand side of the screen, there are three text boxes. The top two boxes display the **Latitude** and **Longitude** of the current position of the mouse. The **Cases Shown** box displays how many of the cases in the current victim list are shown. For example, if the box reads "9 out of 12", this indicates that there are 12 cases in the current victim set, but only 9 are displayed on the map. Some cases may not be shown because their latitude and longitude locations are not available in the database.

The next grouping of controls are **Zoom Options**. The (+) and (-) buttons allow you to zoom in and zoom out on the map. The **Zoom Options** currently allows you to toggle among a set of maps. The maps that are available are of the Pacific Northwest, the State of Washington, the Seattle/Tacoma area, and the Kent/Renton area.

7.2 Adding/Removing Cases

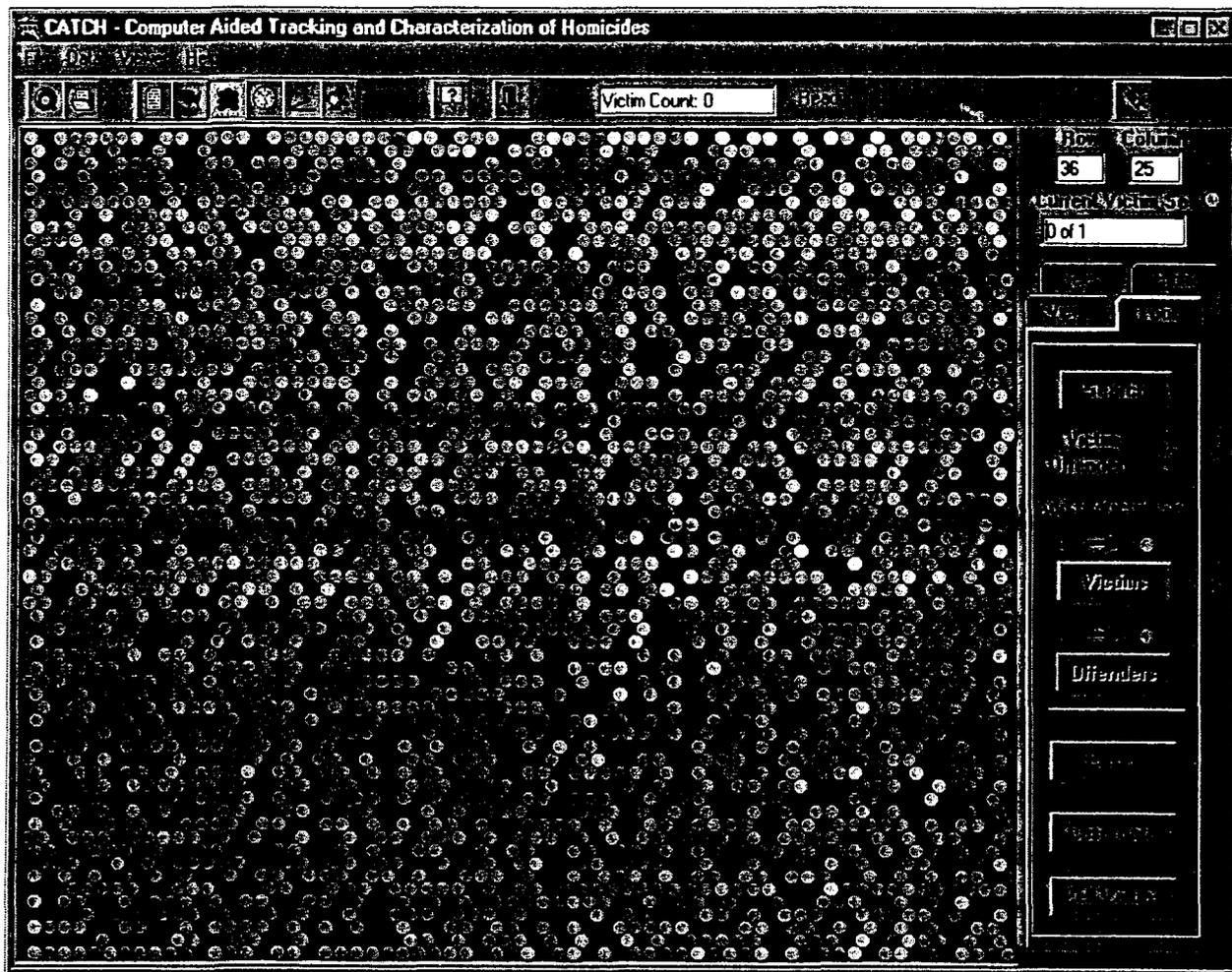
Cases can be added to or removed from the current victim set. A region of the map can be selected by pressing the **Ctrl** key while holding down the left mouse button and dragging the mouse to form a rectangle. Then cases that are contained within that region can be added or removed. The controls that allow this are:

- **Add** adds all cases within the selected region.
- **Crop** removes all cases outside the selected region.
- **Remove** removes all cases inside the selected region.
- **Cancel** unselects the region.

8.0 Self-Organizing Map Viewer

Enter this view by choosing **Self-Organizing Map** from the **Viewer** menu or by clicking the **Self-Organizing Map Viewer** button on the toolbar. The screen will display a self-organizing map that displays *cells* of different shades of green. The neural network has analyzed each case in the database and placed them into a cells based on how the investigator answered the questions on the HITS long form. For more on how the cases were placed, refer to the **Theory** section located in Appendix A. The color of each cell represents the density of cases in the cell (the more cases in the cell, the lighter the shade of green). Black denotes that there are no cases in the cell. Remember that even if a cell contains no cases, it does have weights (a reference vector) associated with it. Refer to the **Theory** section.

The control panel is located on the right hand side of the screen. It contains **Row**, **Column**, and **Current Victim Set** text boxes. It also contains a series of tabs labeled: **Views**, **Tools**, **Select**, **Profile**. By clicking on each tab, a panel appears displaying the controls that are available. These controls will be explained in detail below.



8.1 Navigating On The SOM

8.1.1 Moving the Mouse

When you move the mouse around on the map, the row-column location of the cell that the mouse is on is displayed **Row** and **Column** boxes in the control panel. The **Current Victim Set** box tells you how many cases are in that cell and it tells you how many on those cases are in the current victim set. For instance, if the box reads "1 of 2", this indicates that there are 2 cases in the cell and 1 of those cases is in the current victim set.

8.1.2 Selecting Cells

You may click with the mouse on any cell that contains cases. Clicking once on a cell will turn the cell yellow, which denotes a "selected" cell. The cases in this cell are now in the current victim set. Clicking a second time on a cell will turn it gray. A gray cell is no longer selected; the case is no longer in the current victim set, but it indicates that you have previously viewed that cell. Finally, clicking once more on a gray cell will turn it back to its original color.

8.1.3 Selecting Blocks of Cells

A block of cells can be selected at one time by first positioning the mouse arrow to a corner location of where you want to define the block. Then, while holding the **Ctrl** key down on the keyboard, push the left mouse button, move the mouse to the opposite corner of the block (a rectangle will appear around the included cells), and release the left mouse button. After the mouse button is released, the **Select Tab** on the control panel will appear revealing the controls available for selection. Refer to Section 8.2 for an explanation of the **Select Tab**.

8.1.4 Right Clicking on a Cell

Right-clicking on any cell will bring up a window with a list of cases in that cell. Any cases that are currently in the victim set will be highlighted in yellow. Clicking on any line of the list will highlight that line. Clicking on any highlighted line will unselect it. When the desired cases have been selected (or unselected), click **OK**. This action either adds the cases to or removes the cases from the current victim set, depending on whether the cases were selected or unselected. If **Cancel** is pressed, no changes will be made to the victim set.

8.2 Control Panel Tabs

As was stated above, the control panel is located in the right hand side of the **Self-Organizing Map Viewer**. The control panel is composed of four tabs, each containing controls that allow the user to locate the functionality available to aid analysis on the SOM. The following discussion explains the functionality on those tabs.

8.3 Select Tab

The Select tab contains controls for adding and removing cases from the victim set.



The following controls are available on the **Select Tab**.

Block selection controls:

- Intersection Only checkbox
- Add
- Crop
- Remove
- Cancel

Current Victim Set Controls :

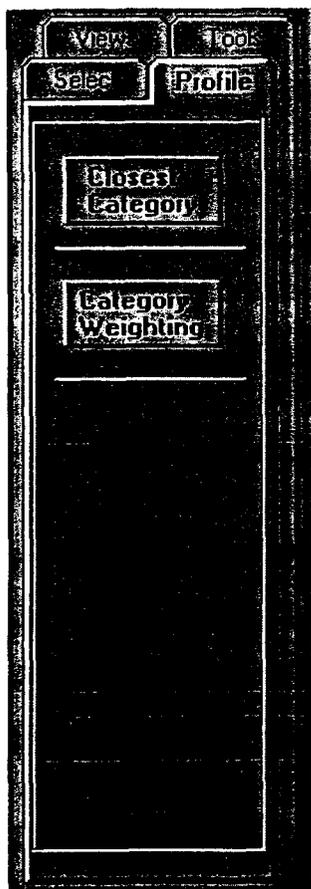
- Remove Current Victim Set
- Remove Old Set

The **Intersection Only** check box applies only in block selection while in the **Agreement** and **Boundary** Views. These views will be discussed later.

The **Add** button adds all cases in cells that are in the specified region. The **Crop** button removes all cases outside the specified region. The **Remove** button removes all cases inside the specified region. Finally, the **Cancel** button unselects the region.

The **Remove Current Set** button removes all cases in the current victim set and turns their cells gray. The **Remove Old Set** button turns gray cells back to their original color.

8.4 Profile Tab

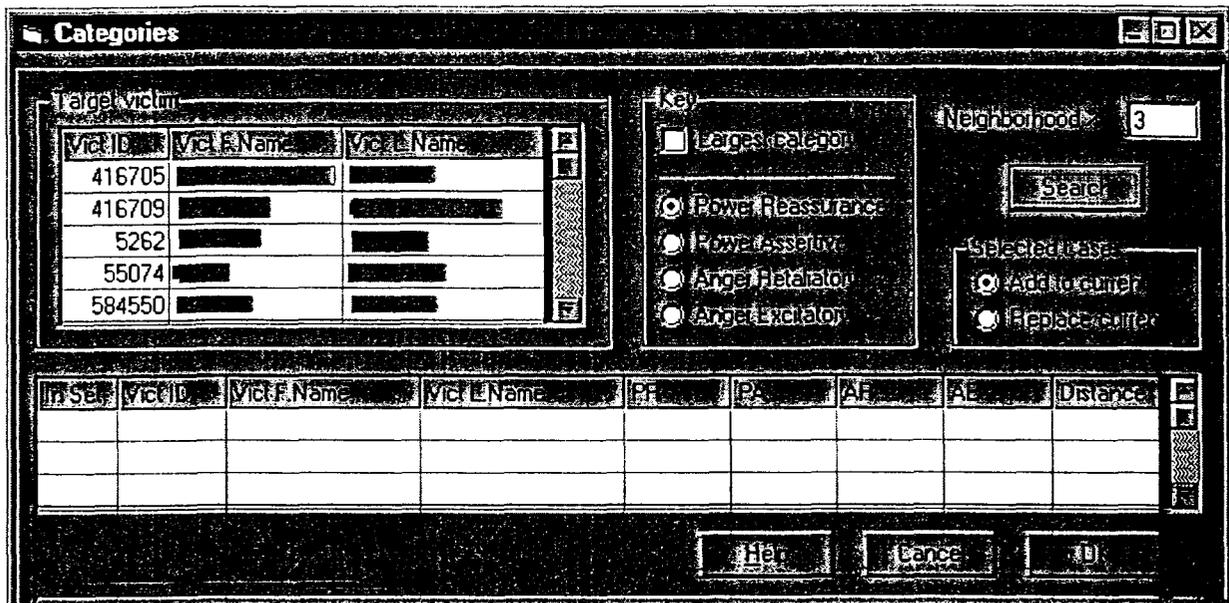


The following command buttons are available on the **Profile Tab**. Their functionality is discussed below.

- Closest Category
- Category Weighting

8.4.1 Closest Category

Closest Category involves the serial sex murder categories: power reassurance, power assertive, anger retaliatory, and anger excitatory. Once cases that are categorized as serial sex murders, are in the current victim set and the **Closest Category** button is pushed, a window will appear similar to the exhibit below. For one victim, **Closest Category** will search the database for all other serial sex murders that have a specified category value within +/- % of the case chosen. This range is defined as the neighborhood. The user must specify the victim, the category that the search will be based on, and the neighborhood.



The procedure is as follows:

1. Choose one victim, by clicking on the victim name in the Target Victim box.
2. Choose the category. Either check the Largest Category checkbox or click on one of the four categories list below. By checking the Largest Category, the program will look at all four category values for the chosen victim and use the largest of those categories.
3. Adjust the search neighborhood. The default is set at 3%
4. Click Search.

The grid will then be filled with all other cases that have a category value within the specified neighborhood. For example. If you choose a victim, the Power Reassurance key, and a 3% neighborhood, the program will find the PR value for that case. Then it will search for all other cases that have a PR value of (+/-) 3% of the target case's PR value. All of the cases that satisfy this condition are displayed in the grid. If the case happens to already be in the current victim set, an "X" will be in the IN SET column. The heading for the column of the key that was searched on will be highlighted blue. The target case will be colored in yellow and the rest of the rows will be colored according to the magnitude of the distance column. The distance column is calculated by the difference between each case and the target case for all categories.

You can add any of these cases to the current victim set by following the below steps:

1. Click in the In Set column in the corresponding row. An "X" will appear in the column.
2. Choose either Add to Current or Replace Current
3. Click OK.

8.4.2 Category Weighting

Category Weights also involves the serial sex murder categories: power reassurance, power assertive, anger retaliatory, and anger excitatory. Once cases, that are categorized as serial sex murders, are in the current victim set and the **Category Weighting** button is pushed, a window will appear similar to the exhibit below. For one victim or two victims, **Category Weighting** will display the category values for each question.

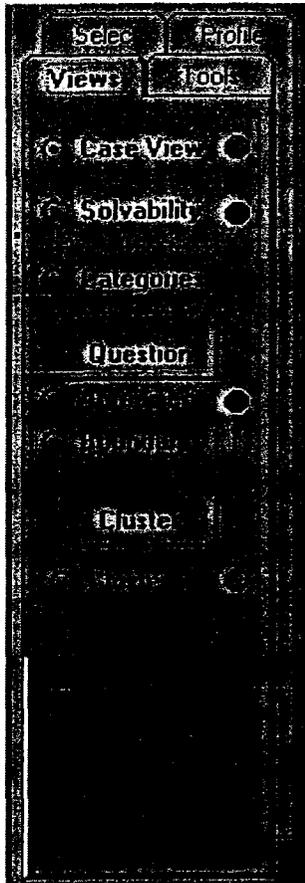
The procedure is as follows:

1. Choose One Case or Two Cases from the Cases Viewed box.
2. Choose the correct number of cases from the Target Victim(s) grid.
3. Click retrieve.

The case(s) will appear in the grid at the bottom of the screen showing the category values for every question. For each question, the sum of the category values will be displayed in the Total column.

8.5 Views Tab

The Views tab contains controls that change how the cells are represented on the screen.



The views available are:

- Case View
- Solvability
- Categories
- Agreement
- Boundary
- Clusters

8.5.1 Case View

Case View is the default view that is displayed when the **Self-Organizing Map Viewer** is chosen. The color of each cell represents the density of cases in the cell (the more cases in a cell, the lighter the color). Black denotes that there are no cases in the cell.

8.5.2 Solvability

Solvability View colors the cells based on how many cases in the cell are solved. The cells with the highest ratio of solved cases to total cases are white, those with the smallest ratio are the darkest red. When the mouse is held over a cell and the map is in solvability view, the percent of solved cases within that cell is displayed at the bottom right of the screen.

8.5.3 Categories

Categories View colors the cells based on the four categories of serial sex murders: power reassurance, power assertive, anger retaliatory, and anger excitatory. If a cell contains cases that where the offenders fall into one of these categories, the cell is colored based on which category is most represented in that cell. The cells where none of the offenders fall into one of the categories are colored gray. The colors for the different categories are shown at the bottom right corner of the screen when you are in **Category View**.

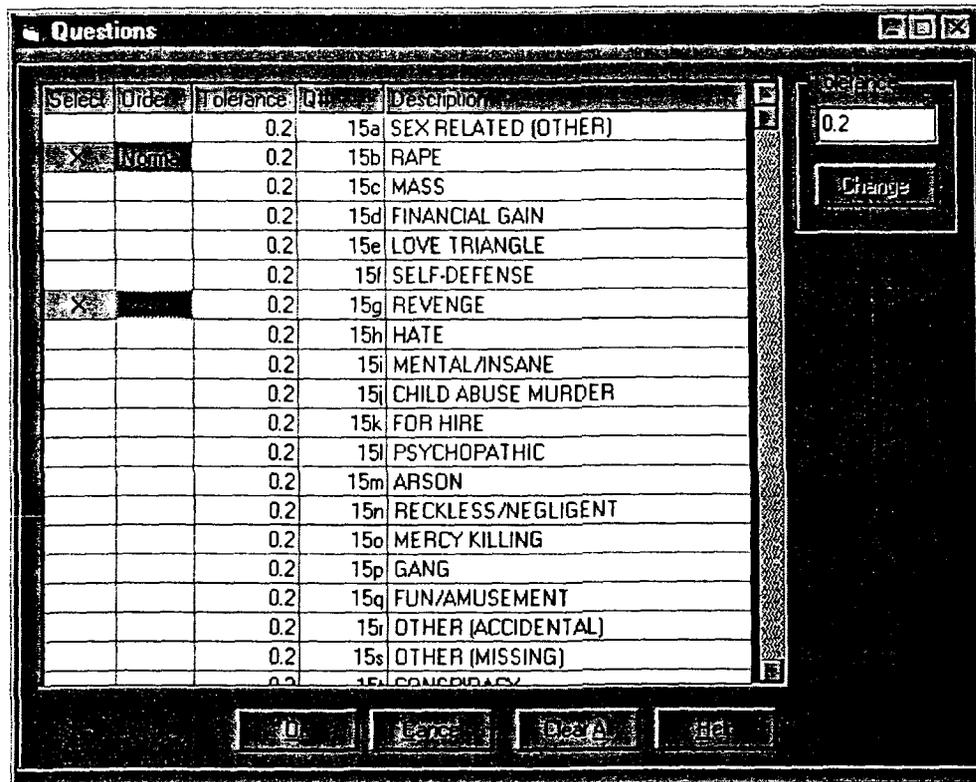
When the mouse is held over a cell, the number of offenders in that cell that fall into each category is displayed beneath the abbreviation and color in the bottom right corner.

8.5.4 Boundary and Agreement

You can choose one or more questions, long form questions, that are used in the neural network and display the regions on the map that are associated with these questions. For example, you may be interested in finding where on the map serial murders have clustered or where both serial and rape murders have clustered. In the areas of the map where these cases have clustered, the cells weights (reference vector) are high for those particular questions. There are two ways of displaying these areas, **Boundary and Agreement**. **Boundary** outlines the area where the cell weight for each question is high. **Agreement** calculates a correlation between the weights of the chosen questions and then colors the cells based on the correlated weight.

8.5.5 Choosing Questions

The **Question** button lets you choose the questions used in the **Boundary and Agreement Views**. Clicking on the **Question** button makes a window pop up that displays a grid containing the questions used in the neural network (see below).



The first column is labeled Select, which contains a “X” and the cell is highlighted yellow if the question is selected, and blank if it is not selected. The second column is the order of the selection. Normal order is highlighted blue and invert order is highlighted red. The third column contains the tolerance for the order. The tolerance is used in conjunction with the order to determine the regions included in the display. The last two columns contain the question number and question description. The order and tolerance will be explained later.

First, let us describe how to choose a question. To choose a question, click in the select column in the same row as the question. Then the order needs to be chosen by clicking in that cell and the tolerance can change by highlighting the tolerance cell and typing in a new value. The tolerance has to be between 0 and 1. The tolerances for all of the questions can be changed at once by typing the tolerance in the box in the upper right corner of the questions window and pushing the **Change** button. The **Clear All** button on the questions window will deselect all of the selected questions and reset the tolerances to their original values. Finally, when all choices have been made, the **OK** button will cause the map to redraw.

The questions that were chosen can now be viewed by clicking the **Display Questions** button. The colors in this grid correspond to the colors used in the boundary display.

8.5.5.1 Order and Tolerance

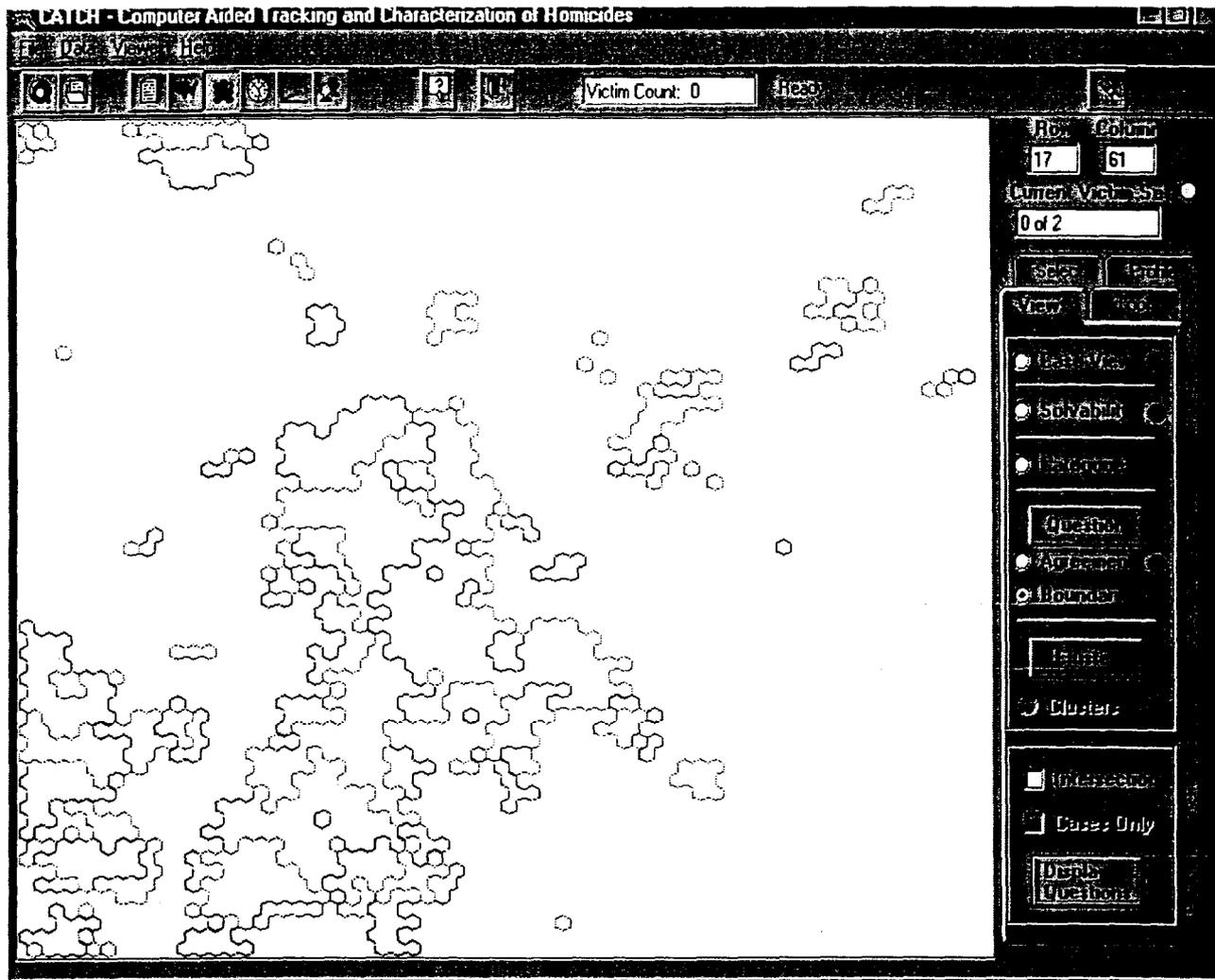
Let’s explain the order and tolerance in terms of an example using real case information. For instance, you may want to view where certain ages of victims appear in the map. So you chose question #36, “V. age = 5-60” (victim age), in normal order and the tolerance set at 0.2 (20%). If

the ages of people ranged from 0 to 100 years old, the normal order would produce a display including all cells with people ranging in age from 20-100 years old (excluding the lower 20%). These cells would either be colored by the age of the person; the older the person, the lighter cell color (agreement display) or enclosed by a boundary (boundary display). The reverse order would produce a display including all cells with people ranging in age from 0-80 years old (excluding the upper 20%). The boundary display would enclose these cells by a boundary. In the agreement display, the cells would also be colored by the age of the person, but instead the younger the person is denoted by the lighter cell color. The above example used real case information to illustrate order and tolerance, but the CATCH uses the neural network weights to display the map.

Expressing the same example in terms of weights, the agreement display would produce a map that would still be colored according to age, but the higher the weight for the age question of a cell; the lighter the cell color. If the same question was chosen by invert; the higher the invert of the weight for the age question of the cell ($1 - \text{weight}$), lighter the cell color. When more than one question is selected, a correlated weight is calculated using the weights from all of the selected questions. This correlated weight is then used in the display. In the boundary display, if the order is normal and the weight of the question is above the tolerance it is included within the boundary. In the invert order, if the invert of the weight for a cell is above tolerance, then the cell is included within the boundary.

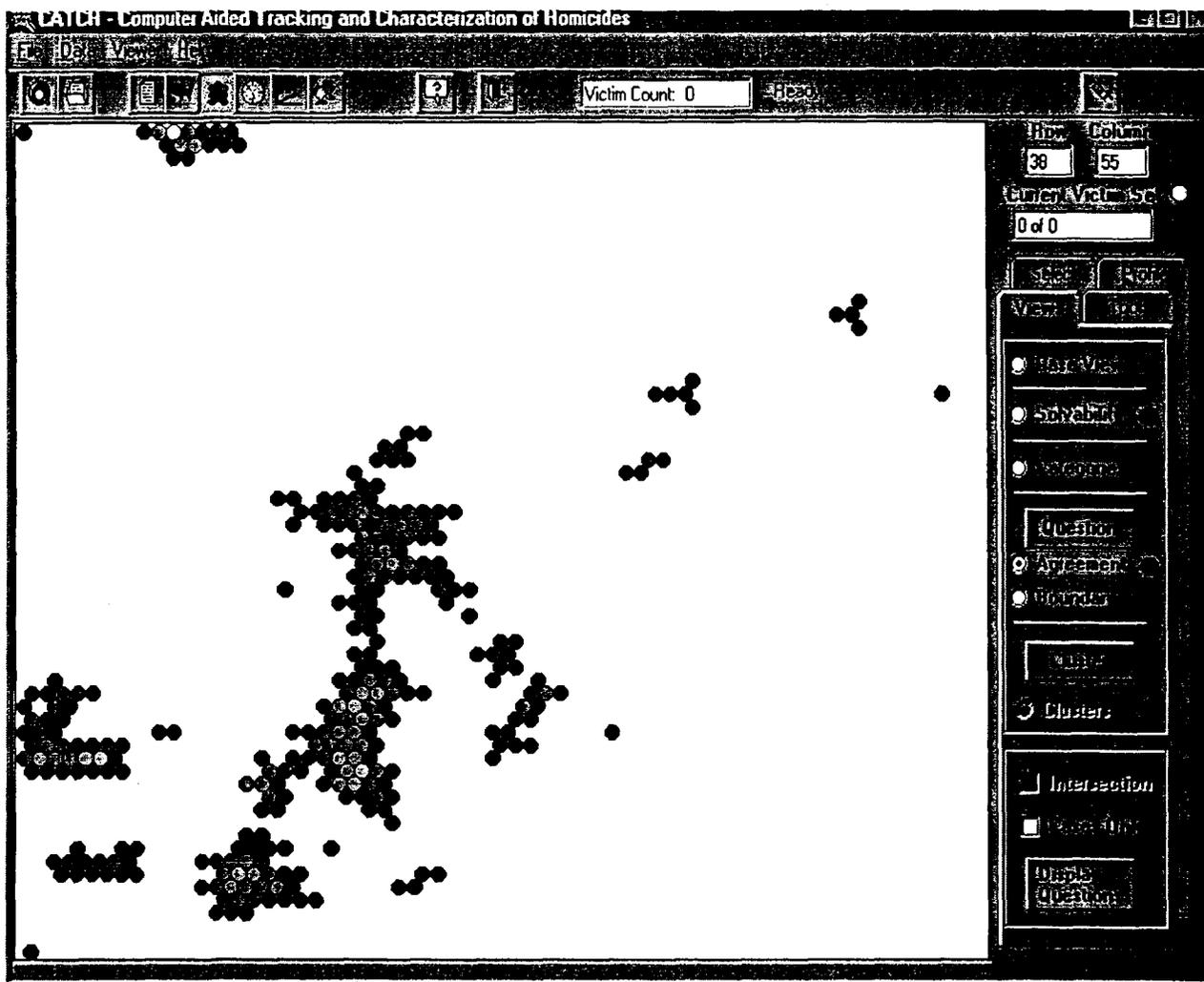
8.5.6 Boundary Display

The display that appears after the closing of the questions window is the boundary display. The boundary display outlines the area where the cell weight (reference vector, see the **Theory** section) for each question is high. Each cell contained inside of these areas does not necessarily contain cases that apply to the specific question selected. But they are located in an area where the cells weights for the question are high. For instance, if the user chose question #15b, rape murders, and displayed the boundaries. Not all of the cells contained inside of the boundaries have specific rape murder cases in them, but they are located in an area of high rape murder probability, where the weight for question #15b is high. When two or more questions are selected, the boundaries are displayed in different colors. These colors correspond to the colors displayed in the grid that appears when the **Display Questions** button is pushed. See the example below.



8.5.7 Agreement

After one or more questions are selected and the boundary display is showing, you can switch to the agreement display. To change to the agreement display, click on the radio button located next to Agreement. The agreement display calculates a correlation between the cell weights of the selected questions and then colors the cells based on the correlated weight. If the correlated weight in a given cell is above the tolerance and is very high, the cell is colored in a light shade of blue. Likewise if the correlated weight in a cell is above the tolerance is very low, it will be a darker shade of blue. If the cell is black, then the correlated weight was not above the tolerance and is not included in the agreement display.



8.5.8 Options Available in Boundary Display

Intersection

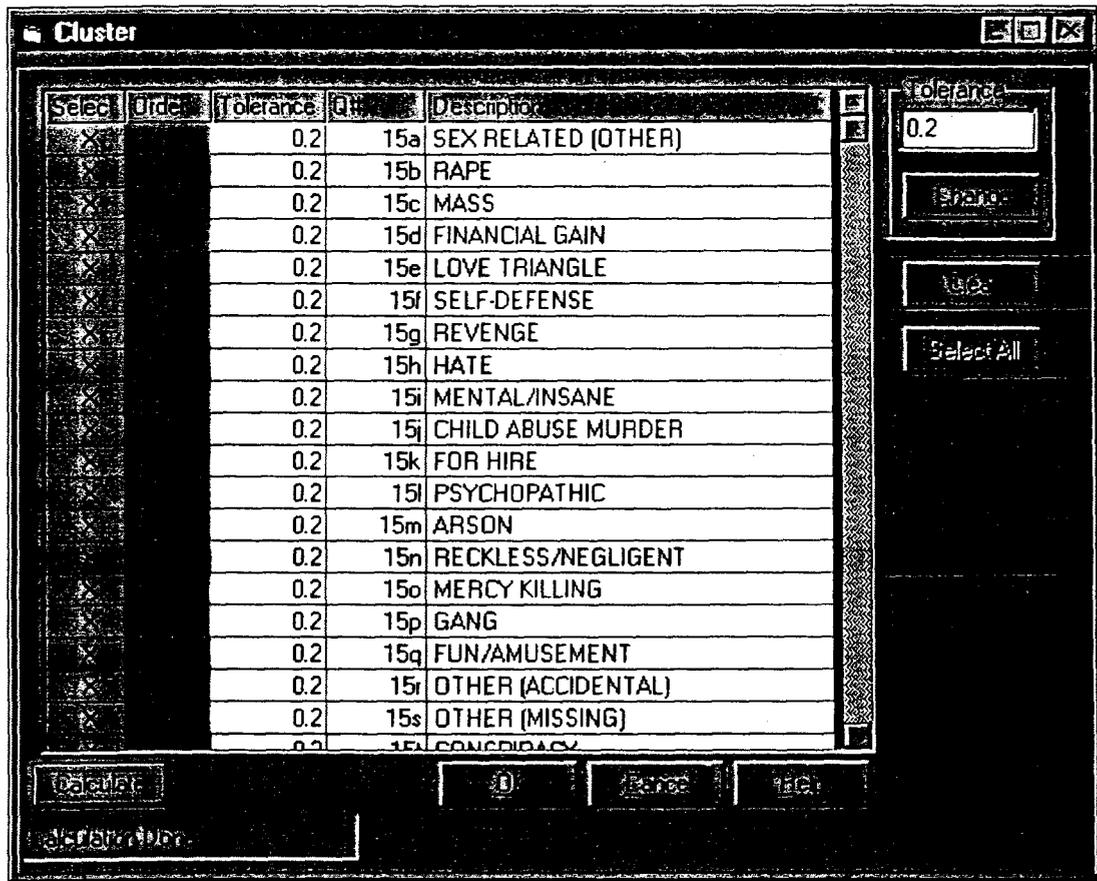
When two or more questions are selected and the boundaries for the questions are displayed, there is an option to activate the cells in the intersection of the boundaries. The checkbox is at the bottom right of the screen. Clicking on the checkbox will redraw the display with the cells in the intersection drawn in white. To deselect the cells in the intersection, click the checkbox again.

Cases Only

When intersection is checked, the checkbox labeled **Cases only** can also be checked. This selects only the cells in the intersection that actually contain cases. This checkbox can also be toggled on and off.

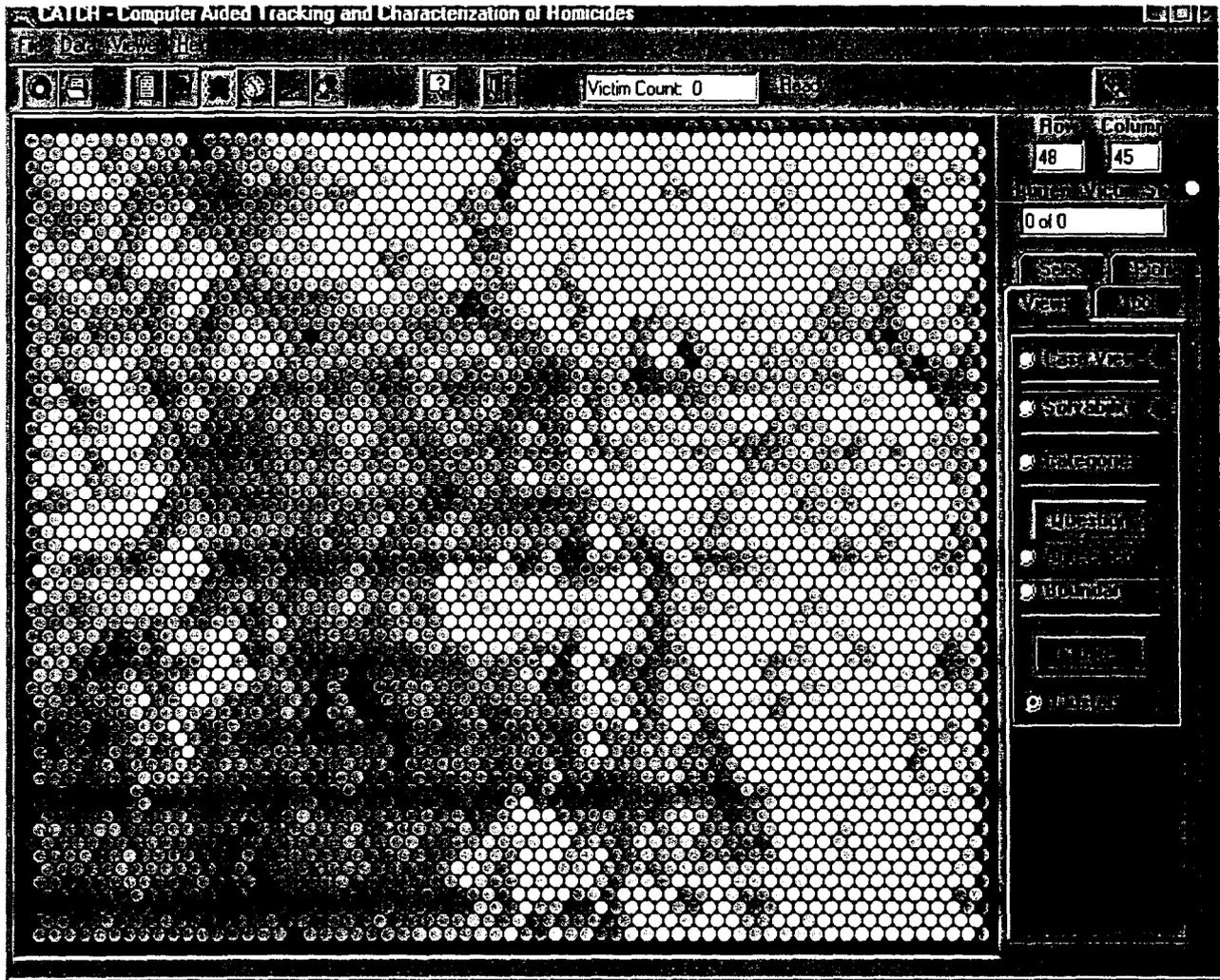
8.5.9 Cluster

To display the **Cluster View**, push the **Cluster** button. The **Cluster** button allows you to visually represent the similarity of each cell to the surrounding cells based on the questions you choose. When you click the **Cluster** button, a screen (similar to the exhibit below) appears for you to choose which question(s) are to be used in the cluster comparison.



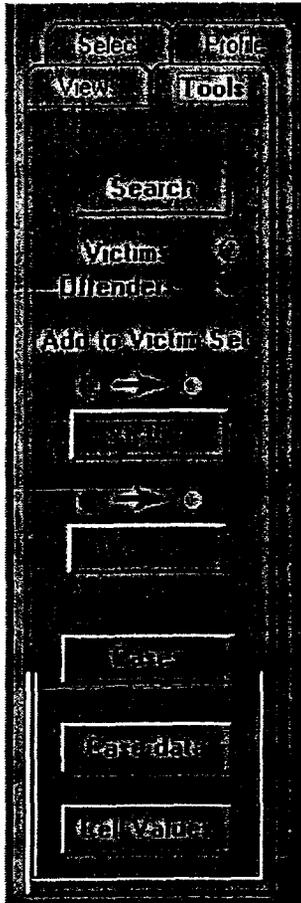
The columns in the grid that are used in Cluster are Select and Description. The Tolerance and Order columns are not used. To select a question, click in the Select column next to the question. The select box will turn yellow and contain an 'X' when it is selected. Select individual questions by clicking on the question row. All of the questions are chosen by default. To unselect all questions click **Clear**. To select all the questions at one time, click the **Select All** button. You can choose as many questions as you want. When you are finished selecting questions, click the **Calculate** button. When the computer is finished calculating click **OK** to exit the screen. To exit the screen at any time without changing the self-organizing map, click the **Cancel** button. The self-organizing map will be redrawn to represent the similarities between the cells based on the questions you have chosen. The similarities are determined by calculating the Euclidean distance between the reference vector of each cell with each of its

neighboring cells. The specific weights in the reference vectors that are used in the Euclidean distance calculation are the weights associated to the selected questions. Those cells that are most like their surrounding cells will be colored white, and the cells that are most different from their surrounding cells will be colored dark red. (See below). This feature allows you to characterize portions of the self-organizing map as being associated with certain questions.



8.6 Tools Tab

On the Tools tab you can search the SOM for a specific victim or offender, or view the case data or cell data for the cases in the victim list.

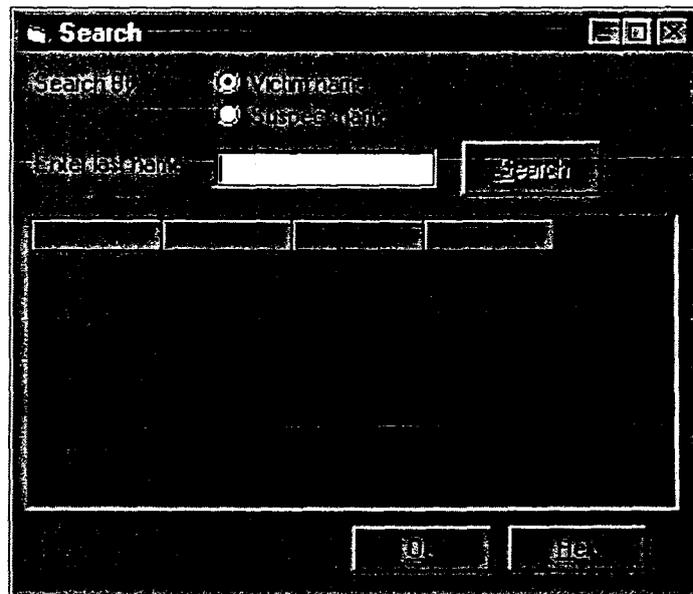


The tools available are:

- **Search**
- **Cases**
- **Case Data**
- **Cell Values**

8.6.1 Search

Search allows you to find victims or suspects on the SOM. First, you must choose either to search by victim name or suspect name (click the appropriate radio button). See the exhibit below. Then, type the last name of the victim/suspect in the **Enter last name** field. You need not type the full last name. After entering the name, you may click the **Search** button, or press <Enter> on your keyboard. A list of names will appear in the grid for all victims/suspects whose names begin with the text entered. You can scroll down the list to find the name of the desired person; then click on the name. This action will select the cells which contain this person; i.e., these cells will be re-colored on the map. Suspect cells and victim cells will be colored orange and purple, respectively.



You may perform as many searches as desired before closing this window, including changing from victim searches to suspect searches and vice versa. If two or more names are chosen of the same type, i.e., two victim names or two suspect names, then the color of the cells from the previous searches will change. In the current selection, the appropriate cell will always be colored in the bright standard color (orange or purple) and the cells from a previous selection will be changed to a dark shade of this color. Therefore, a distinction is always made between the most recent selection and previous selections.

After the search window is closed, the SOM map on the main screen will be redrawn with the appropriate colored cells. Right-clicking on one of the orange or purple cells will display the cases in that cell, with the selected cases highlighted in orange or purple.

To add the searched cases into the current victim set, the cells must be selected. Each individual case can be added by right clicking on the cell and choosing the case from the grid. In addition, all of the cells from the searches can be selected at once using the two buttons located just below

the **Search** button. Clicking on the **Offenders** button will turn the suspect cells from orange to yellow. Now information is available for these cells. The same action can be done for the victim cells with the **Victims** button.

8.6.2 Case

When there are cases in the current victim set and the **Case** button is clicked, a window will pop up that displays the location of each selected cell, the murder ID, the victim ID, the victim name, the offender ID and the offender name. The cases in the victim list are highlighted in yellow. This list of information can be sorted according to any of the columns by double-clicking on that column. This screen also shows the information for the import list. Importing cases will be discussed later in this manual.

Case information

Row Col	Murder ID	Victim ID	Victim Name	Offender ID	Offender Name
53.35	5873	307891		339948	
55.35	1284	3945		3946	
55.35	2015	5200		46756	
56.34	240	1926		505949	
56.34	4970	330040		38104	
53.33	5995	348571		348572	
52.32	2455	139577		139578	
52.32	6365	323755		356472	

Import Export

Row Col	Murder ID	Victim ID	Victim Name	Offender ID	Offender Name
---------	-----------	-----------	-------------	-------------	---------------

8.6.3 Case Data

When there are cases in the current victim set and the **Case Data** button is clicked, a window will appear that displays each case's individual weights for each of the questions (case vector, see **Theory** section). See the exhibit below. Double-clicking in the grid will sort the list based on which questions have the highest average value for all cases displayed; that is, the question with the highest weights will appear at the top, then the question with the next highest weights, etc. A case can be removed from the victim list by clicking at the top of the column on the victim name or ID so that the column is selected, then clicking the **Remove** button. The column will then become gray but will not be removed from view. However, this case is removed from the current victim set.

Case data				
	Victim Person ID	328266	328323	328324
	Victim First Name			
	Victim Last Name			
155	SEX RELATED OFFENSE			
156	RAP			
156	MAS			
158	FINANCIAL GAIN			1.00
158	JOY TRIANGLE			
159	REVENGE			1.00
157	RAV			
157	REVENGE			

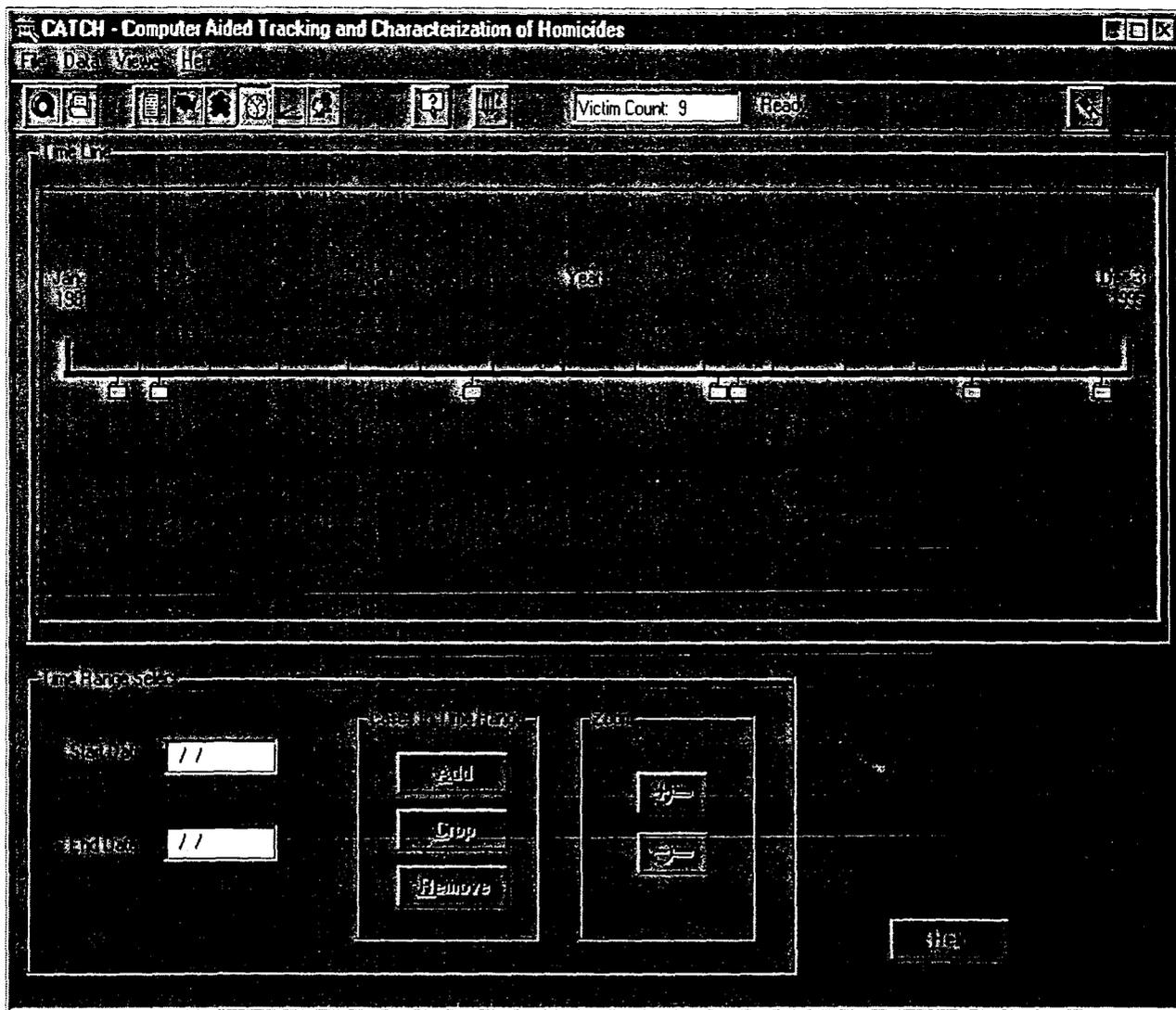
8.6.4 Cell Values

When there are cases in the current victim set and the **Cell values** button is clicked, a window will pop up that displays a question description, each cell's weights for all questions (labeled at the top with the cell's row and column location), and the average weight for each question. This is the reference vector for each cell (refer to the **Theory** section). The magnitude of the weights is displayed in each cell along with a color indicator. The color is a visual representation of the value of the weights, the lighter the color the higher the magnitude of the weight. If the cell contains the text "n/a", that indicates that the question was non-applicable in that cell. The questions can be sorted for any column by double-clicking in that column. This arranges the questions so that the highest values in the chosen column are at the top.

Cell values				
ID	Description	14, 28	14, 29	Average
15a	SEX RELATED (OTHER)	1000	1000	1000
15b	PAP	1000	1000	1000
15c	MAGE	1000	1000	1000
15d	FINANCIAL GAIN	1000	1000	1000
15e	LOVE TRIANGLE	1000	1000	1000
15f	SELF-DEFENSE	1000	1000	1000
15g	REVENGE	1000	1000	1000
15h	HATE	1000	1000	1000
15i	MENTAL ILLNESS	1000	1000	1000
15j	CHILD ABUSE MURDER	1000	1000	1000
15k	FOR HIRE	1000	1000	1000
15l	PSYCHOPATH	1000	1000	1000
15m	ARSON	1000	1000	1000

9.0 Time Line Viewer

Enter this screen by choosing **Time Line Viewer** from the **Viewer** menu or by clicking the **Time Line Viewer** button on the toolbar. This view displays the cases in the current victim set, represented by a file folder, according to when the crime was committed. Upon entering the view, if there are no cases in the victim list, the program will ask you for starting and ending dates for the time line.



Holding the mouse over one of the folders causes a box to appear with the victim ID, name, and date of the case.

Double-clicking on any case will bring up a report summarizing the information for that case. The information is organized into different categories. Click on the tabs at the top to view different information. See **How to Use the Characterization Screen** (Section 11.1) below for more information about viewing and printing the report.

The dates used to plot the cases are the date of death, the date the body was found, the date the victim was last seen, and the date of the report. The program checks the date of death first. If that information is available, it plots the case according that date. If the information is missing, the program then checks the date the body was found. If that date is missing, the program proceeds through the other dates in the above list until a date is found. Then the case is plotted according to that date.

A time period can be selected in one of two ways: by holding the **Ctrl** key down while pressing the left mouse button and dragging the mouse, or by typing dates into the **Start Date** and **End Date** boxes below the time line. After a range has been selected, four actions can be performed.

1. The **Add** button adds all cases in the database within the time period selected.
2. The **Crop** button removes all cases that are outside the selected time period.
3. The **Remove** button removes all cases within the time range selected.
4. The **Zoom In (+)** button enlarges the area of the time line selected. Once this has been done, the **Zoom Out (-)** button allows you to return to the original time line.

10.0 Principal Components Viewer

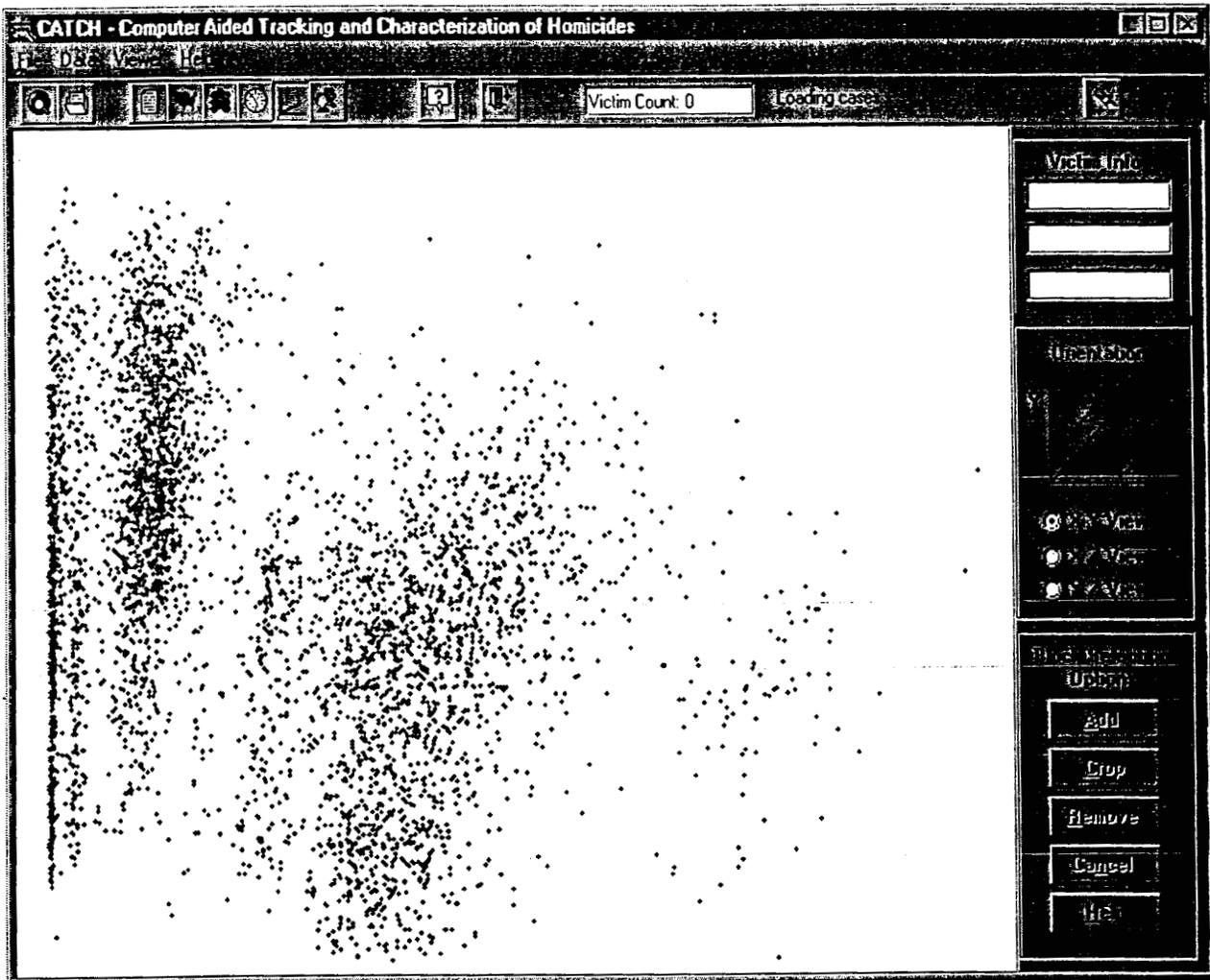
Enter this screen by choosing **Principal Components** from the **Viewer** menu, or by clicking the **Principal Components Viewer** button on the toolbar.

Principal Components View plots the eigenvalues of each case vector. The program reduces each 225 element case vectors to a three element eigenvector (X,Y,Z). The eigenvalues are the three elements in this vector. The purpose of this procedure is to reduce the case vector into a smaller, manageable number of dimensions (in this case three dimensions) that we can plot while still retaining as much of the information as possible.

Each case is represented by a blue dot. If the case is in the current victim set, it is represented by a yellow dot. Clicking the right mouse button on or near a case in the current victim set causes the victim ID, and first and last name to appear in the boxes under **Victim Info** in the upper right portion of the screen.

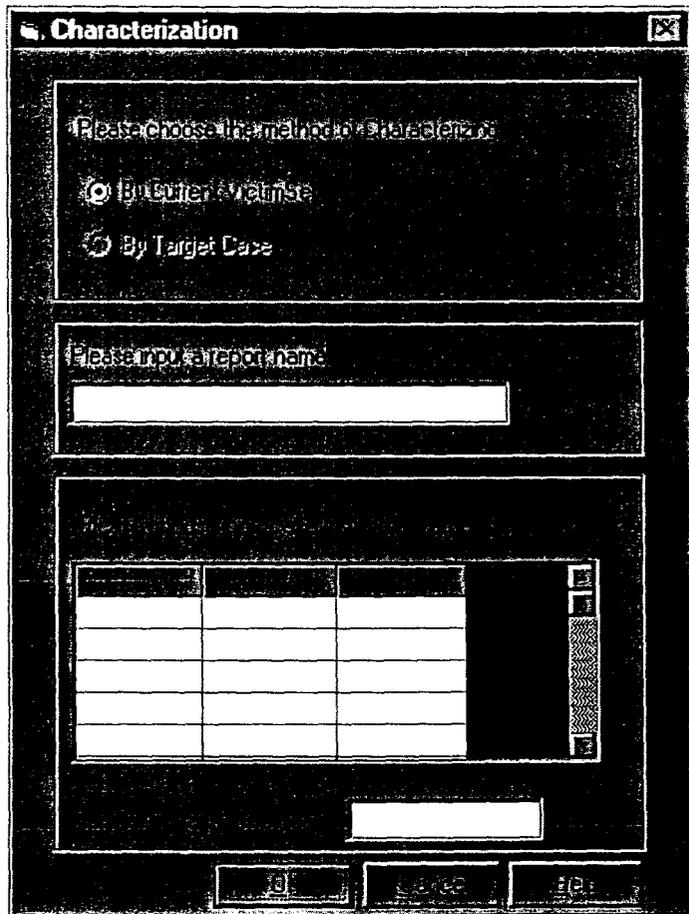
You can view the eigenvalues in different two dimensional orientations: (X,Y), (X,Z), (Y,Z). Simply click on the appropriate button on the right side of the screen.

You can add, remove, or crop victims to/from the current victim set by holding the **Ctrl** key down, holding the left mouse button down, and moving the mouse. A rectangle will appear on the screen. Then you can use the **Add**, **Crop**, **Remove**, or **Cancel** buttons to perform the appropriate actions.

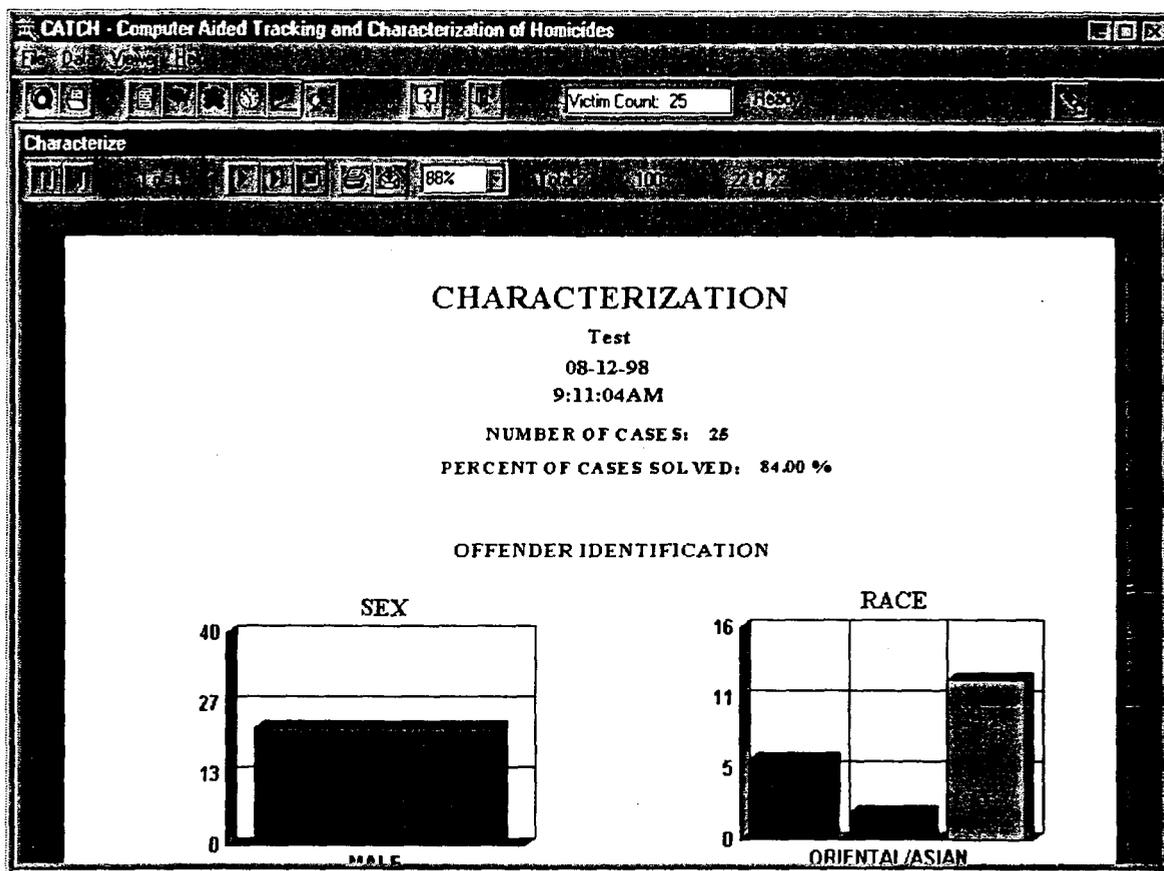


11.0 Characterization

Characterization displays characterizing case information in report form. To enter this view, choose **Characterization** from the **Viewer** menu, or by clicking the **Characterization** button on the toolbar. This will bring up a screen that asks you to choose between current victim set or target case as a method of characterizing. At this time choosing a target case is not available so **Current Victim Set** will be automatically selected. (The box at the bottom of the screen has to do with a target case, so you will not be able to input anything there.) You may also choose a name for your report by typing in the text box. At this point you may click **Cancel** to return to the **Main** screen, or **OK** to see a report.



When you click OK, the computer will take a few minutes to input the data for the cases from the database. When the report appears, it will depict the characteristics of the cases in the victim list in graphical and tabular format. See below.



11.1 How to Use the Characterization Screen

To view the pages, click on the page with the mouse and then push the up and down arrows or the page up and page down keys on the keyboard to move up or down on one page. Use the **Go to Next**, **Go to Previous**, **Go to Last**, and **Go to First** buttons on the toolbar to move to a different page in the report. In order to make the page larger or smaller, choose a different size on the **Display Size** drop-down menu.

To print the report, click the **Print** button on the toolbar. A window will appear that allows you to print the entire report, or certain pages, choose the number of copies, and choose whether or not to collate the copies. The report will automatically be printed to the default printer, which is shown at the top of the window. Click the **OK** button when you have made all your choices and the report will be printed. Choose the **Cancel** button to exit the **Print** screen without printing.

The **Export** button allows you to convert the information in the report to another format and save it to your hard drive or send it to your e-mail. When you click the **Export** button a window will appear that allows you to choose what format you want to save the information in (including a Word document, an Excel spreadsheet, or a database file) and whether to save it on the hard drive or send it to e-mail. After you have made your choices click **OK**. If you have chosen to save the file to a disk the computer will ask for a name for the file. Type in a name and choose what directory to save it in then click **Save**. If you are choosing to send the file via e-mail, you

will be asked for the address of the recipient and a message. If you are exporting the file to an Exchange Folder, you will be asked for the name of the profile. After you have made all your choices, a screen will be displayed showing the progress while the file is being exported. At this point you can cancel by clicking the **Cancel Exporting** button.

11.2 Report Format

At the top of the report you will see the report name or "Characterization" if you did not input a name, the date the report was prepared, the time, the number of cases in the victim list, and the percent of cases solved. The date and time will appear at the top of each page of the report.

The first section of the report is the **Offender Identification** section. This consists of graphs representing the offenders' sex and race, and also shows the average age of the offenders.

The next section is the **Offender Physical Characteristics** section. This section shows the average height and weight of the offenders, has graphs showing build, hair color, hair length, hair shade, facial hair, eye color, tattoos, and grooming of the offenders, and tables that deal with offenders' glasses and where their tattoos were created.

Next is the **Offender Lifestyle** section. In this section there are graphs showing marital status, who the offenders were living with, and how many were employed. There are also tables showing offender lifestyle, criminal history, occupation, group member, and their serial sex category.

The last section is the **Vehicle** section. This section contains a graph showing how many cases involved vehicles, and tables showing vehicle model type, body type and condition.

12.0 Importing and Exporting

Cases can be imported and exported between CATCH and CATCHRAPE. In order to import or export cases, you must be in the **Self-Organizing Map** screen. To export cases, choose **Export** from the **Data** menu. The computer will export the cases that are currently in the victim list. They are now available to be imported into CATCHRAPE. (Please note: This is different from exporting from the **Characterization** screen. When you export from the **Characterization** screen, you are creating a file that you can view outside of CATCH and CATCHRAPE. When you export from the **Characterization** screen, that information is not made available to import into CATCHRAPE.)

If you have exported cases from CATCHRAPE, then you can import them into CATCH while in the **Self-Organizing Map** screen. Choose **Import** from the **Data** menu. The computer will automatically find the cases exported from CATCHRAPE and import them into CATCH. The computer will determine where the rape cases would fall on the self-organizing map if they were murders, and these cells are colored red. This allows you to see if the rape cases exported from CATCHRAPE are similar to the murder cases in the victim list based on their proximity on the self-organizing map.

To view the case information and position, choose the **Cases** button on the **Tools** tab. The cell position, rape ID, victim ID, victim name, offender ID, and offender name for each of the imported cases will be displayed in the lower box.

13.0 Admin

The **Admin** option under the **File** menu allows you to perform neural network administration. The neural network administrator will be the only user of this functionality.

14.0 Exiting

To quit the CATCH program, choose **Exit** from the **File** menu, or click the **Exit** button on the toolbar, or close the window by clicking the 'X' button in the upper right corner of the screen. Please note that all cases that have been selected will not be saved when the program is exited, so when CATCH is opened again there will be no cases in the victim list.

Appendix A: Theory

A.1 Glossary

Question Weights - The question weights are weights applied to each question that denote the *importance* of that question. The current weighting scheme weighs all of the questions equal (1.0) except for certain victim questions; they receive no weight (0.0) as requested by the client.

Reference Vectors - The reference vector is composed of *weights* associated with a *cell*.

As a result of the training of the neural network, each cell has a reference vector of weights associated with it. For our purpose of understanding, the vector is simply an organized list of numbers and can be represented in the following form:

<Reference vector> = < 0.02, 0.3, 0.5, 0.02, 0.00, 0.01, 0.4, 0.06,.....>

The reference vector for each cell is unique. In the vector, there is a weight for every input variable used in the neural network. In our case, these variables are questions.

In CATCH, the reference vector for any selected cell can be displayed by pushing the **Cell Values** button. The exhibit below shows the reference vectors for cells (12,21), (14,39), and (21,26). The three cells were selected by clicking on those cells on the SOM map (they were highlighted yellow) and then the **Cell Values** button was pushed. Each column of numbers is the reference vector for that cell. Every reference vector has 225 elements, the number of input variables (the number of questions and question alternatives).

The magnitude of the weights is an indication of what type of cases may be found in that cell. For instance, if the weight in a cell reference vector for REVENGE is very high and the weight for GANGS is very low, it is very likely to find cases in that cell where the crimes involved revenge but not gangs. Finally, the last column shows the average of the cell weights of each row. This aids in finding commonalties between cells.

ID	Descriptor	12, 21	14, 39	21, 26	Average
15a	SEX RELATED (OTHER)	0.0	0.0	0.0	0.0
15b	HATE	1.0	1.0	1.0	1.0
15c	MAS.	0.0	0.0	0.0	0.0
15d	FINANCIAL GAIN	0.0	0.0	0.0	0.0
15e	LOVE TRIANGLE	0.0	0.0	0.0	0.0
15f	SELF-DEFENSE	0.0	0.0	0.0	0.0
15g	REVENGE	0.0	0.0	0.0	0.0
15h	HATE	1.0	1.0	1.0	1.0
15i	MENTAL/INSANE	0.0	0.0	0.0	0.0
15j	CHILD ABUSE/MURDER	0.0	0.0	0.0	0.0
15k	FOR HIRE	1.0	1.0	1.0	1.0
15l	SYNOPSIS (ATH)	0.0	0.0	0.0	0.0
15m	ARSON	0.0	0.0	0.0	0.0

Case Vector - The case vector is composed of normalized values associated to real data for a specific case.

For each case, CATCH looks on how each question was answered. The answer for each question is normalized (assigned a value between 0 and 1) and these normalized values make up the case vector for that case. Every case has one case vector and its length is 225 elements long, just like the reference vector. The case vector for all selected cases can be displayed by pushing the Case Data button.

The Case Data display is shown below for five cases. Each column of numbers is the case vector for that case. In this example, for the case involving victim 2881, notice that the elements for HATE and FOR HIRE are high (1.0). This indicates that those questions were answered "yes" and element SEX RELATED (OTHER) was answered "no" (represented by a 0.0).

Case data		5230	48920	3945	2881	10322
	Victim Person ID					
	Victim First Name					
Question	Victim Last Name					
15a	SEX RELATED (OTHER)					
15b	RAP					
15c	MASS					
15d	FINANCIAL GAIN					
15e	LOVE TRIANGLE					
15f	SELF DEFENSE					
15g	REVENGE					
15h	HATE				1.00	
15i	MENTAL INSTAB.					
15j	CHILD ABUS. MOBILE					
15k	ROBBERY				1.00	
15l	PSYCHOPATHIC					
15m	ARSON					
15n	RECKLESS/NEGLIGEN.					
15o	MERCY KILLING					
15p	BANK					
15q	FOR AMUSEMENT					
15r	OTHER ACCIDENTAL					
15s	OTHER MISSING					
15t	CONSPIRACY					
15u	DOMESTIC VIOLENCE	1.00	1.00			1.00
15v	SERIAL POSSIBLE SERIAL					
15w	REX OF ANGE			1.00		1.00
15x	ROBBERY					

A.2 Determining where a case will be in the SOM map

To determine which cell will contain a specific case, the Euclidean distances between the case vector for that case and the reference vector for each cell is computed. The Euclidean distance is a way of finding the distance between two vectors.

For instance:

<Reference vector> = < 0.02, 0.3, 0.5, 0.02, 0.00, 0.01, 0.4, 0.06,.....>

<Case vector > = < 0.5, 0.25, 0.8, 0.00, 0.05, 0.4, 0.7, 0.00,.....>

$$\text{Euclidean distance} = \sqrt{(0.02 - 0.5)^2 + (0.3 - 0.25)^2 + (0.5 - 0.8)^2 + (0.02 - 0.00)^2 + (0.00 - 0.05)^2 + \dots}$$

The Euclidean distances between the specific case vector and *all* reference vectors (every cell) are calculated. **The cell reference vector that resulted in the smallest Euclidean distance is the "winner" and that case will be located within that cell.**

PROPERTY OF
National Criminal Justice Reference Service (NCJRS)
Box 6000
Rockville, MD 20849-6000