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

Document Title: Algorithmic Image Matching (AIM): Project Analysis for Santa Ana Police Department

Author(s):   Santa Ana Police Department

Document No.:   181955
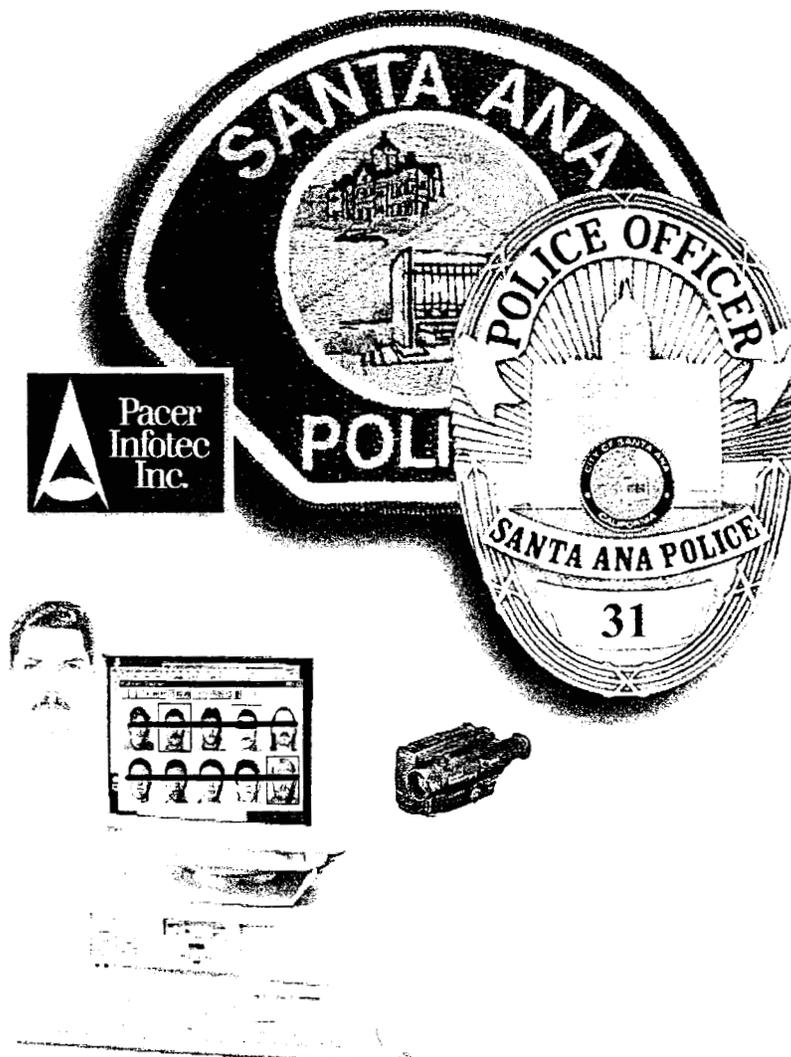
Date Received: April 19, 2000

Award Number: 97-IJ-CX-K011

Algorithmic Image Matching
(AIM)

Technology Research And Development Partnership
Projects For Community Policing

Submitted to: National Institute of Justice

Submitted by: Santa Ana Police Department

In Partnership with: Pacer Infotec, Inc.

Algorithmic Image Matching (AIM)

Project Analysis For
Santa Ana Police Department

Id #: 97-IJ-CX-K011
(Replaces Id #: 97-IJ-CX-0019)

February 10, 1999

1.0 Executive Summary

Many organizations throughout the United States have adopted an electronic mug imaging system. These systems have proven to be a reliable source for storing and retrieving biographical and image information. Most have restrictions on the physical descriptions and types of searches available.

Pacer Infotec Inc. partnered with the Santa Ana Police Department to develop, integrate, implement, test, and evaluate a mug identification imaging system that exceeds the current generation of mug imaging systems. During this partnership, Pacer Infotec Inc. seamlessly integrated an algorithmic face-recognition engine into its existing *MugMatch* identification-imaging product. In addition to searching for suspects based on physical characteristics, Santa Ana Police Department can use photographs, composite drawings, and video surveillance clips as the template image and sort the resulting images into similar and dissimilar images. This technology assists investigators in identifying suspects by focusing on the most probable set of images.

In order to properly stress test the system and assess its overall usability, Santa Ana Police Department's existing data was imported into the *MugMatch* environment. After analyzing Santa Ana Police Department's current mug imaging system's data structure, Pacer Infotec Inc. designed and tested an import utility that successfully transferred all sixty thousand mug data records and images into *MugMatch*. Since the *Phantomas* face recognition engine relies on algorithmic mathematical computations, the graphs denoting the three main facial characteristics of an image: left eye, right eye, and mouth, needed to be automatically generated. A function was created to generate all 60,000 approximate face-finding *Phantomas* graphs and successfully executed.

The next phase entailed training investigators on how to effectively use *MugMatch*. A half-day course was provided in-house and course materials and user guides were distributed to students. A Santa Ana Police Department investigator was appointed as the *MugMatch* system administrator and took on the role to lead the system evaluation effort. During this phase, database cleanup efforts were made to fix existing mug images of individuals whose head was not photographed in an upright position and verify the automated placement of *Phantomas* graph cross points. As the Santa Ana Police Department's *MugMatch* system administrator became an expert user of the system, several enhancements were suggested to Pacer Infotec Inc. project staff members. Many of the valuable suggestions were designed, programmed and implemented into the latest release of *MugMatch*.

Several test cases were issued to test the efficiency and accuracy of the system. The volume of data, approximately 60,000 mug records, (over 3 gigabytes of information) showed no signs of system performance impairment. Both fast client hardware and open network throughput effect the retrieval response times. As for usability, investigators were able to access the *MugMatch* system and perform inquiries and *Phantomas* picture-to-picture comparisons after being instructed on how to create and edit queries and manipulate query results.

During the evaluation process, *MugMatch* proved to be a useful tool in assisting investigators in identifying a suspect. It enables investigators to use a digital image, digital composite sketch, or a frame captured from a surveillance video and automatically sort the images into similar and dissimilar faces. It lets investigators focus on the most probable set of images and quickly create high-quality *MugMatch* lineups. In closing, the majority of the test cases issued revealed that the system's ranking capability was able to identify the true image match of a template's digital composite sketch within the top 5 percent on average. This project has proven *MugMatch* to be an advantageous tool to any investigation.

2.0 Objective of *MugMatch* Software

The objective of *MugMatch* is to assist law enforcement officials with suspect-image booking, witness Identification of suspects, and investigative work. It efficiently enables forensic investigation agencies to maintain a central repository of color images, physical descriptions, personal Identification, and arrest information of individuals tracked by the agency. With this information, agencies can search and identify unknown suspects by using *MugMatch's* robust textual query feature, coupled with the high-tech *Phantomas* face-recognition engine.

In a suspect booking, descriptions using standard terminology are entered by trained law enforcement officials. *MugMatch* expedites biographical data entry by allowing customized pick lists and auto-fill data fields. Electronic image capture devices create high-quality photo images. *MugMatch* can capture and store an unlimited number of images and types of images for a single suspect record.

MugMatch provides a user-friendly case management environment. It gives law enforcement officials the ability to search rapidly through mug images by issuing robust textual database inquiries for suspects. The search capability allows investigators to include various types of physical descriptions, including tattoos and aliases. Furthermore, a range of values may also be associated with the search criteria, such as "Age Today between 25 and 30" or "Hair Color is either Brown or Sandy". The case management environment enables agencies to generate high quality suspect lineups in quick response to queries. Case notes may be recorded and retrieved throughout an investigation. For easy access, additional evidence such as photographs, composite sketches, or video surveillance clips can be stored to a case. Information regarding witnesses and interview sessions may be entered and also stored to the related case. In addition to its textual database query capabilities, *MugMatch* offers a remarkable face-recognition engine. This powerful picture-to-picture search engine, *Phantomas*, is embedded in the *MugMatch* environment.

Phantomas is an automatic facial comparison and identification-imaging engine. A suspect identification task often began with a composite sketch, which led to a witness looking through possibly thousands of images. This task has proven to be a tiresome and time-consuming

2

responsibility. *Phantomas* takes an input image, either a facial image, a digital composite sketch, or a frame captured from a surveillance video, and automatically sorts the images in the *MugMatch* database into similar and dissimilar faces. Therefore, the witness can focus his/her attention on the important details.

Phantomas utilizes neural face-recognition algorithms that permit rapid response times. *Phantomas* is capable of searching through 10,000 images in approximately five minutes. A large database is likely to contain duplicate images. *Phantomas* greatly simplifies database administration since it detects duplicates through its face comparison capability.

Phantomas operates by detecting the face within an image. It then places a "graph", a right-angled grid, across the face. Each graph will contain nodes that are the cross points representing the characteristic features: left eye, right eye, and center of mouth. The *MugMatch* user, a person who accesses the system, may override the system generated cross points and manually place the cross points on a facial image to ensure accuracy. These features are calculated using mathematical operations. The results are stored in the *MugMatch* database and will be used in obtaining the qualitative characterization of that face. They are of mathematical-abstract nature; therefore, it is not possible to reconstruct a face out of the graph values or to recognize particular biological features (i.e. a big nose or a small mouth). Only the whole characteristic of the face is evaluated and not apparent changes such as whether the person has a beard or no beard, glasses or no glasses, or even a new hairstyle.

2.0.1 *MugMatch* System Architecture

The *MugMatch* software communicates with the *MugMatch* database. The *MugMatch* data resides on a designated database server. The server is in constant operation and communicates to *MugMatch* client workstations via local or wide-area network links. The *MugMatch* system contains several basic components:

- *Capture station* enters suspect information and captures mug images.
- *Retrieval station* searches the *MugMatch* database by specifying various descriptive characteristics of a suspect and examining the results.
- *Update station* enables information to be modified.
- *Administration station* performs administrative duties for the caretakers of the *MugMatch* system.

3

Printer Database Server Retrieval Station

Local or Wide

Area Network

Capture Station Administration Station Update Station
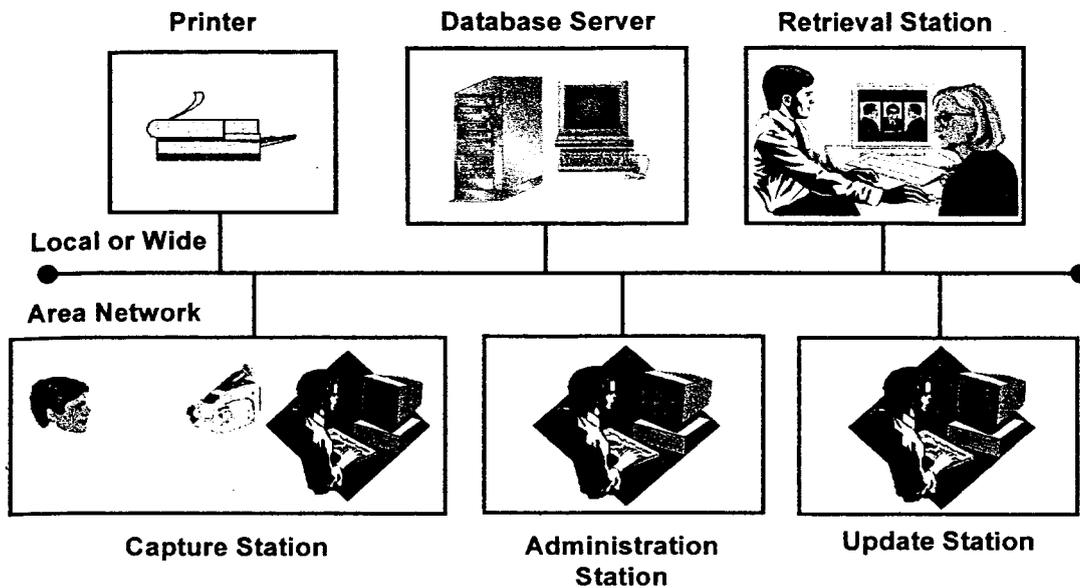
In smaller implementations, a site may choose to just use a single workstation for retrieval, updates, and administration. The capture station differs from other components since its configuration includes a digital camera.

3.0 Project Environment

The following section states the purpose and describes the environment of the research and development project.

The project consisted of developing, testing, implementing and evaluating a state-of-the-art mug imaging system for the Investigations unit of Santa Ana Police Department. The project entailed developing and testing a robust textual search capability coupled with a sophisticated face recognition engine. The *Phantomas* face-recognition engine utilizes neural face recognition algorithms enabling images to be sorted into similar and dissimilar faces.

3.0.1 Statement of Problem

In the last decade the city of Santa Ana has incurred significant demographic changes. Santa Ana's dramatic demographic change is also contributed to the influx of immigrants. Santa Ana's daytime population has also increased since it is home to the majority of Federal, State, and private agencies providing social and governmental services to Southern California communities. With over 300,000 residents and changing demographics, Santa Ana's ability to provide effective police and social services to the community is continually challenged.

A police investigator may investigate several related crimes. Investigations often involve searching for suspects in a mug shot database. Due to the high volume of images in Santa Ana's original mug database, this could make the identification task tedious for witnesses, victims and investigators. After viewing a few hundred similar images, a person's attention span will most likely dwindle.

4

Many businesses have attempted to prevent crime by installing video surveillance cameras. Sometimes an investigator is able to use the suspect surveillance video clip. For cases where no photograph evidence exists, an investigator may rely on a composite sketch based on the physical description given by a victim and witnesses. The video surveillance clip or composite sketch is displayed among other law enforcement officials at roll call briefings, perhaps circulated among other law enforcement agencies, or distributed to the media. Unfortunately, this has proven to be time-consuming and allows the suspect to act out additional crimes and continue to terrorize communities.

The Santa Ana Police Department has expressed interest in a computer technology that assists investigators in the process of identifying suspects in a reduced time frame than those of prior identifications. The *MugMatch* software offers investigators the ability to search for images and rank resulting images according to similarity to the selected template image. The textual database search and face-recognition engine lets investigators, victims, and witnesses focus on the most probable set of suspect candidates.

3.0.2 Santa Ana Police Department – Investigation's System Environment

The following section describes the environment before and after installation of the *MugMatch* software at Santa Ana Police Department.

3.0.2.1 Current Generation of Santa Ana Police Department's Mug Imaging System

The current generation of Santa Ana Police Department's mug imaging system stores and retrieves images and biographical data. The database server's hardware is antiquated and the operating system is an outdated version of UNIX. The mug shots are captured during the booking process and taken in an environment with good lighting. Until recently, many mug images captured were of the individual with their head tilted slightly to either the left or right as opposed to a centered and upright position. The booking officer then enters the descriptive and personal information.

The underlying database is proprietary, hence, not a commercial available product. Due to the database's complex data structure, this makes it impossible for persons other than the vendor to generate ad hoc reporting using commercial available tools. The database contains three categories of data: Adult, Juvenile, and Registrant. Consequently, an investigator must first select which database category to access before continuing.

A total of three client workstations are available for retrieving information or creating a person record. Since the client application runs on a UNIX platform, it is not currently possible to obtain the information at each investigator's desktop computer. This has caused some lack of efficiency since an investigator must walk over to a mug imaging workstation and occasionally wait until another investigator is finished processing a query.

The current imaging system supports a finite set of physical description data elements: eye color, facial hair, glasses, hair color, hair length, gender, race, body height and weight. Person records are generated at the time of booking. The system limits the booking officer to capturing one

5

image per booking record. If additional images are needed to depict a tattoo or special marking, additional records are created. Since multiple images may not be captured and associated to a single person record, duplicate images are currently dispersed throughout the database.

An investigator generates a group of suspects by issuing information describing the pursued individual. Based on the information gathered by victims, witnesses, or surveillance cameras, the investigator builds a query. The investigator enters the information into the provided dialog. Queries consist of text information only, which search the database for records matching the supplied criteria. The current system is restricted to just one instance of a particular type of information. For example, the query may search for "white males with brown hair" as oppose "white males with either brown or black hair". From the resulting set, the system allows investigators to build a high-quality lineup and direct it to the printer.

3.0.2.2 *MugMatch* Environment at Santa Ana Police Department

The following identifies the deliverables made to Santa Ana Police Department at the time of installation:

Qty	Item	Vendor/Model or Remark
2	Pentium Personal Computers	Dell Pentium 233 – server has 128 MB RAM and client has 32 MB RAM. Both include monitors, network cards, controller cards, video capture boards.
1	Database software	Sybase SQL AnyWhere with 4 user license.
1	Laser printer	LexMark Optra C
1	Scanner	Hewlett Packard Scanjet
1	Video camera	Sony CCD-TR86
1	VCR	Sony SLV-685HF
1	Frame grabber	Integral Flash Point 128 Light
1	TV/Monitor	20" Mitsubishi CS-20103
2	MugMatch software	Includes online help text.
N/A	MugMatch training & support	Class material and a user guide was supplied to each student who attended the course.

For evaluation purposes, only two workstations were needed: a database server and client workstation. The database server is solely used as the central *MugMatch* data repository. The client workstation is used to access and edit information. The database server and client workstation both run Microsoft NT workstation. Both are connected to Santa Ana Police Department's local area network via TCP/IP. Throughout the duration of the software evaluation period, several trips were made to Santa Ana Police Department in effort to upgrade the *MugMatch* software and provide assistance to investigators.

3.0.2.2.1 *MugMatch* Evaluation – Phase I

Phase I of the project entailed analyzing the current mug imaging system's data structure and importing the data into a single *MugMatch* database. Since the current mug imaging system is based on a proprietary file structure, a custom import utility was developed so the three categories

6

of data could be successfully transferred into a Sybase SQL AnyWhere *MugMatch* database structure. After developing and testing the import utility, all of Santa Ana Police Department's sixty thousand mug data records and images were successfully exported into the *MugMatch* database.

Upon exporting the data into *MugMatch*, the *Phantomas* graphs needed to be defined. This entailed modifying the system administrator's application, Maint, by designing and programming a function that automatically calculates the *Phantomas* graphs. This function generated an approximate face-finding *Phantomas* graph for all sixty thousand images in the *MugMatch* database.

3.0.2.2.2 *MugMatch* Evaluation – Phase II

Phase II of the project consisted of providing training to investigators on how to use *MugMatch*. In addition to the user guide, a half-day course was structured to train investigators. Pacer Infotec provided in-house class instruction and provided course material. The class was taught in a computer classroom at Santa Ana Police Department so each student could practice the exercises on-line.

3.0.2.2.3 *MugMatch* Evaluation – Phase III

During this stage of the project a *MugMatch* system administrator was identified and trained on how to become the caretaker of the system. Investigators exercised *MugMatch* to evaluate its overall usability. To ensure the system's accuracy, it was suggested that one or more individuals be appointed to review the *Phantomas* system generated face-finding graphs. Many of the mug shots captured from Santa Ana Police Department's current mug imaging system displayed individual's with their head slightly tilted either to the left or right. To obtain optimal results in a picture-to-picture search, the *Phantomas* graphs needed to be rotated to an upright position and cross points redefined. *MugMatch* offers a utility enabling image rotation and a separate function for editing the *Phantomas* graph.

In effort to complete this task, Pacer Infotec adopted the task. Approximately, twenty-five percent of the graphs were reviewed and modified if necessary.

3.0.2.2.4 *MugMatch* Evaluation – Phase IV

During this phase, the appointed system administrator, investigator Mark Strohman, became acquainted with *MugMatch* and assembled a requirement and enhancement list of features he thought would be useful to investigations. Pacer Infotec analyzed the list and were educated on the type of features that would make investigators more efficient in doing their detective work. Pacer Infotec designed, programmed and implemented several new features suggested by Santa Ana Police Department's *MugMatch* system administrator. The following summarizes the features implemented in the latest *MugMatch* release:

- It is now possible to create a query that matches the description of a selected image in a query results viewer. From the speed menu, choose "Create AutoQuery" and a new query will be created with a description that matches the selected image. The query property dialog is opened to permit editing of the new query. When the query is saved, it is associated with the

7

case from which the query results originated. The default name of the query is "AutoQuery - NAME" where NAME is the name of the person from which the query is derived. There are a number of items that can be included in the system-generated query. The user is given the opportunity to select these items when the AutoQuery is first created.

- The person locator dialog was modified to provide more flexibility. The new person locator capability allows the user to search for several characteristics simultaneously, including last name, first name, date of birth, gender, SSN, driver's license, race, and age. The results are displayed in a list which is appended to the locate dialog during the search. For each entry in the result list, details can be obtained. Optionally, a locator query can be executed which shows details or images according to the Options setting.

- It is now possible to rearrange the images in a lineup builder using drag and drop techniques. To move an image from one frame to another, click on the image with the left mouse button, hold the button down, drag the mouse to the frame where the image is desired and release the mouse button. If you drag one image on top of another, the original image is cleared and the image being dragged replaces it.

- A new feature has been added to case viewer. It is now possible to capture and save images so that they are associated with a case. An additional branch appears in the case viewer to display these images.

- In the query results viewer, it is now possible to easily skip to the next/previous page containing a suspect. There are two new buttons in the query result viewer toolbar that are used for this purpose.

- A progress bar has been added to the *Phantomas* Sort and *Phantomas* Compare functions. It is now possible to cancel these operations prior to completion, in which case the associated query results are unchanged.

4.0 Case Findings

To measure how well *MugMatch* performed; several test cases were issued and executed at Santa Ana Police Department. For some scenarios, composite sketches of known suspects were obtained from Santa Ana Police Department's FaceKit workstation. These images were saved in a TIFF format and imported into a *MugMatch* case for efficient retrieval. If the suspect did not already exist in Santa Ana Police Department's mug database, then a record was created and the actual mug shot or related photograph was scanned into the *MugMatch* database. The following cases reflect a sampling of tests performed at Santa Ana Police Department.
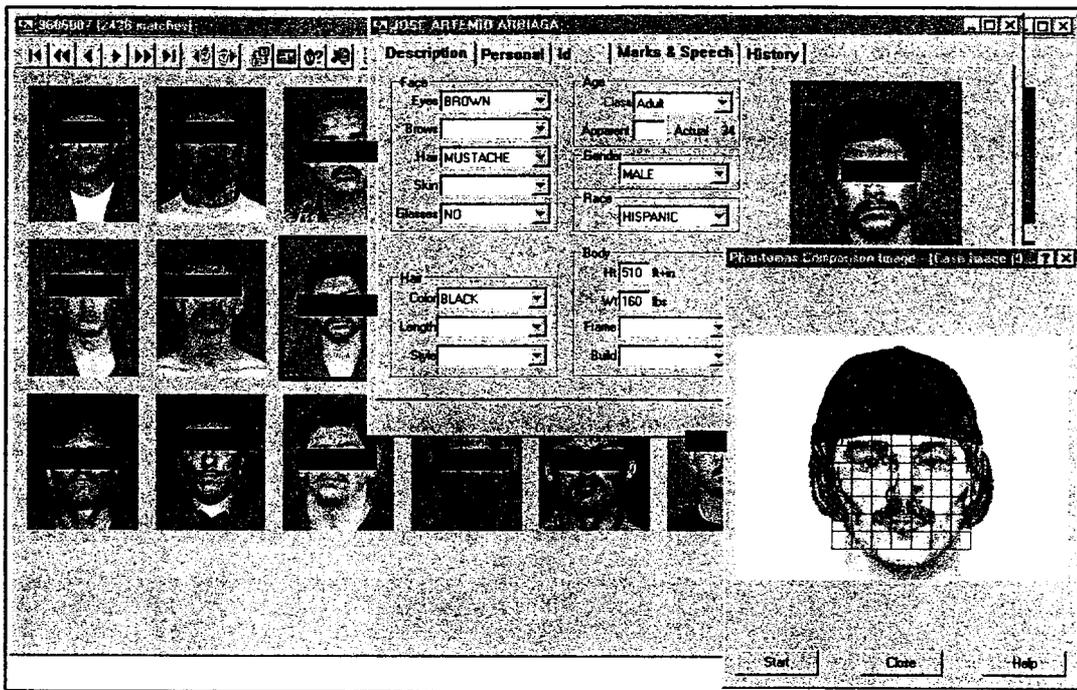
Test Case 1

Query parameters:

Attribute	Criteria
Race	= Hispanic
Gender	= Male
Age Today	Between 28 and 40 years old
Height	Between 5'08" and 6'00"
Weight	Between 150 and 175 pounds
Hair Color	= Black

Number of matches. 2,426.

Phantomas template source. FaceKit composite sketch image as a TIFF image.

Result The *Phantomas* facial comparison ranked the suspect in position 59 out of 2,426.
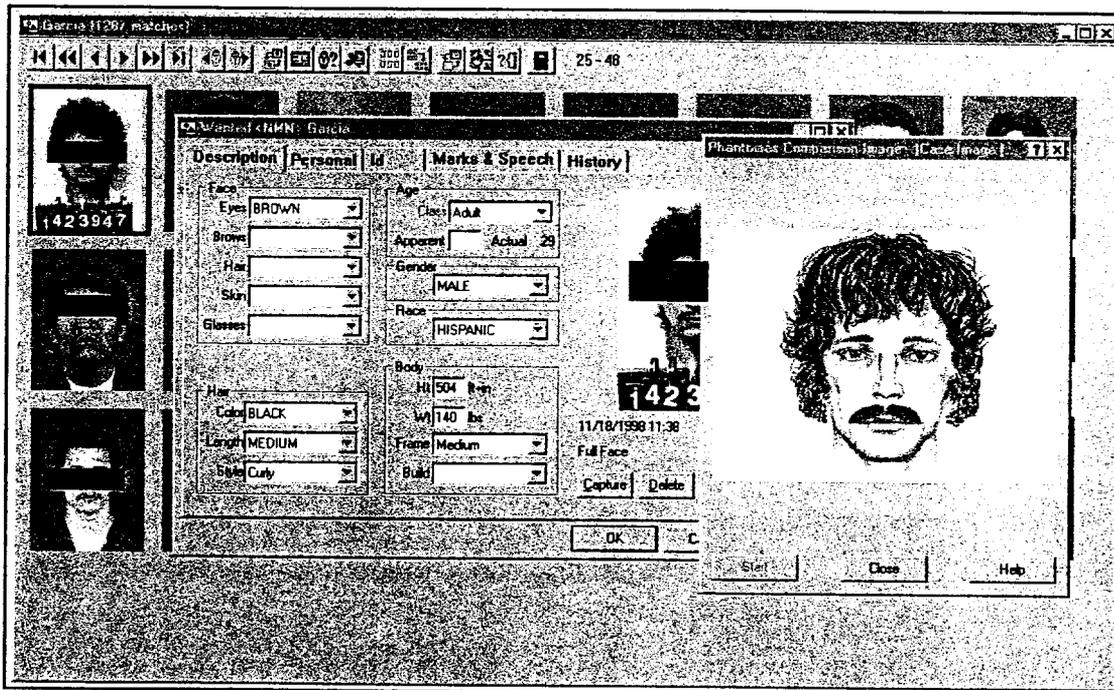
Test Case 1 environment

10

Test Case 2

Query parameters

Attribute	Criteria
Race	= Hispanic
Gender	= Male
Age Today	Between 28 and 31 years old
Height	Between 5'02" and 5'05"
Weight	Between 130 and 150 pounds

Number of matches: 1,287.

Phantomas template source: FaceKit composite sketch image as a TIFF image.

Result: The *Phantomas* facial comparison engine ranked the suspect as 25th out of 1,287 returned matches.
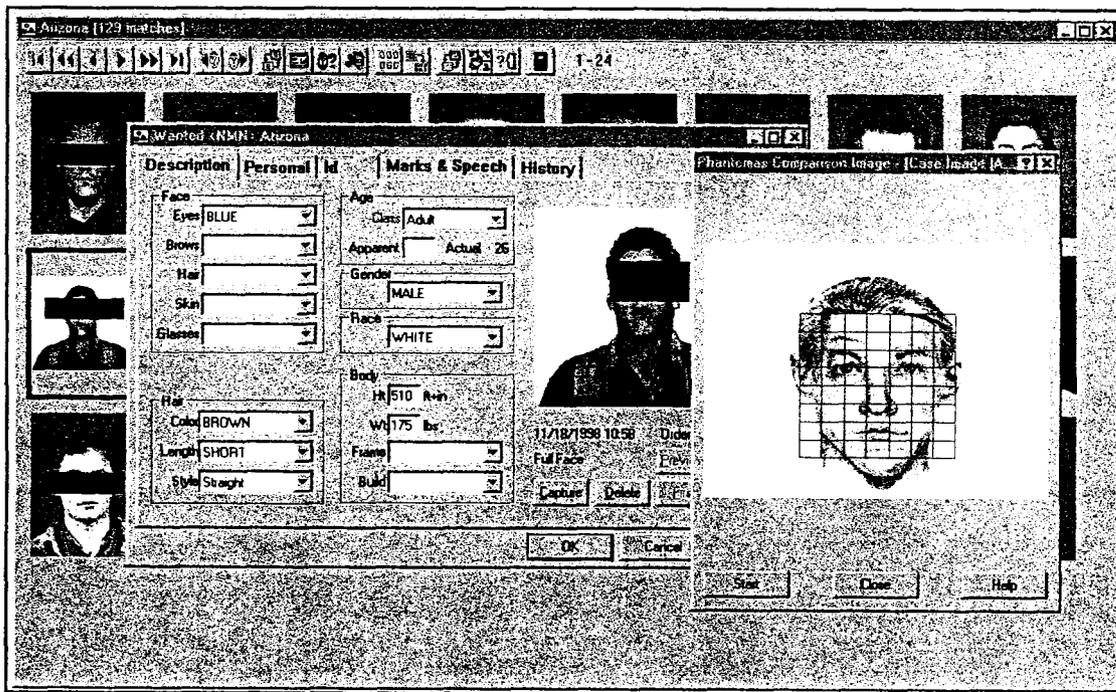

Test Case 2 environment

Test Case 3

Query parameters:

Attribute	Criteria
Race	= White
Gender	= Male
Age Today	Between 21 and 30 years old
Height	Less than 6'00"
Weight	Less than 200 pounds
Eye Color	= Blue

Number of matches: 129.

Phantomas template source: FaceKit composite sketch as a TIFF image.

Result: The *Phantomas* facial comparison engine ranked the suspect as 9th out of 129 returned images.
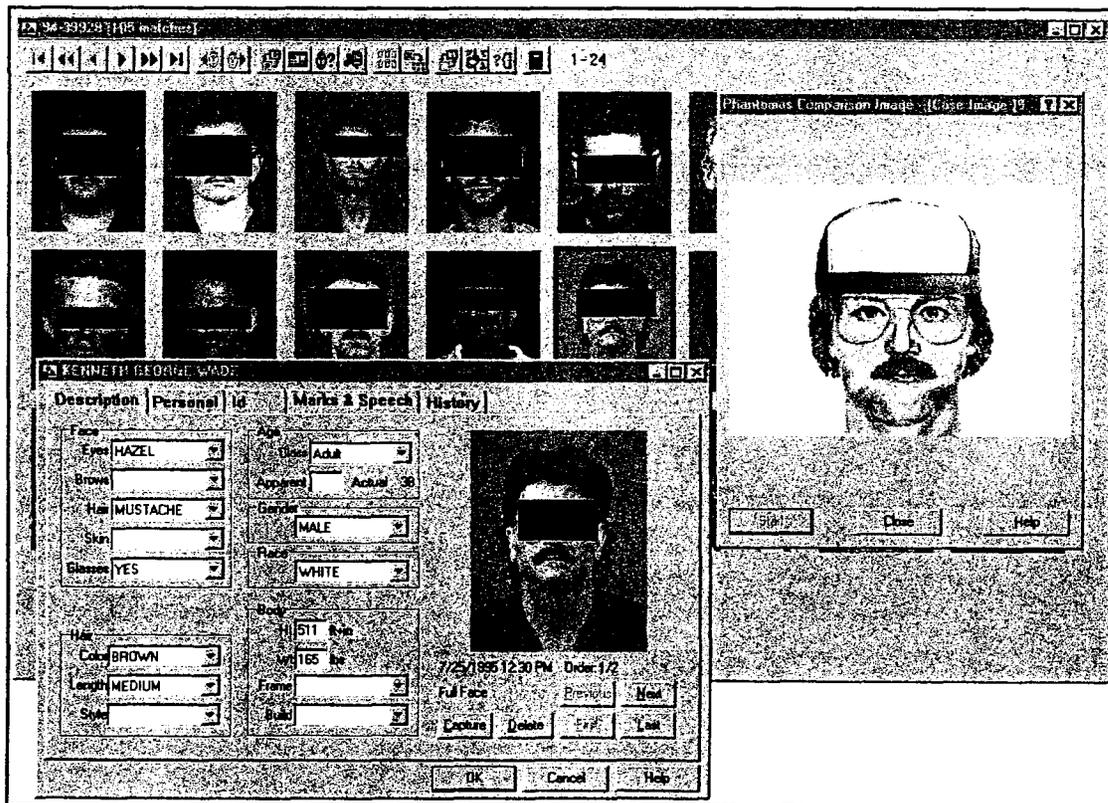

Test Case 3 environment

12

Test Case 4

Query parameters:

Attribute	Criteria
Race	= White
Gender	= Male
Age Class	= Adult
Height Color	= Brown or Sandy
Facial Hair	= Mustache
Eyewear	= Yes

Number of matches: 105.

Phantomas template source: FaceKit composite sketch image as a TIFF image.

Result: The *Phantomas* facial comparison ranked the suspect in the 13th position out of 105 returned images.

Test Case 4 environment

13

5.0 Conclusion

Pacer Infotec Inc. is proud to have partnered with Santa Ana Police Department to create the powerful mug identification imaging system, *MugMatch*. Pacer Infotec Inc and Santa Ana Police Department personnel worked closely together to achieve the goal of developing, integrating, implementing, and evaluating the enhanced *MugMatch* system.

During the evaluation period, it has been demonstrated that *MugMatch* is a useful tool in assisting investigators in identifying a suspect. Several time saving features such as the *Phantomas* picture-to-picture search, robust query editor, and new AutoQuery function lets an investigator concentrate on the most important details of a case. Instead of manually searching through approximately 60,000 mug images, *MugMatch's* various tools let an investigator review a manageable set of images.

Of the Santa Ana Police Department mug database, a typical query will usually return several thousand images. *MugMatch's Phantomas* face-recognition engine lets the comparison search automatically rank the resulting images according to how similar they are to the input template image. This significantly reduces the amount of time spent analyzing each image, hence decreasing the chance of becoming fatigued. The test cases performed at Santa Ana Police Department revealed that the system's face recognition capability can locate the true image match using a digital composite sketch within the top 5 percent of returned images. This project has proven *MugMatch* to be an advantageous tool to any investigation.