186734

VISIONICS CORPORATION

FaceIt® FACE RECOGNITION TECHNOLOGY

Adaptive Surveillance: A Novel Approach to Facial Surveillance for CCTV Systems

Final Progress Report

January 28, 2001

NIJ Award No. 1999-LT-VX-K020

Table of Contents

Project Abstract

Accomplishments We have developed a surveillance system that uses real-time face recognition technology to increase the utility of currently existing CCTV-compatible surveillance software. The performance of the best existing surveillance system was dramatically improved by development of techniques for: (1) dynamic adjustment of video parameters in the region of the image containing a face and (2) tracking a face to acquire multiple images of it, across video frames.

The outcome of this work is a state-of-the-art, automated facial recognition surveillance system capable of providing immense value to the law enforcement community. The automation of the attentionally-taxing duty of surveillance lowers overhead, and thus, frees up resources for performance of other tasks. Use of this system will allow law enforcement to perform surveillance duties to a level of efficiency and precision beyond that which is possible at present. Ultimately, fewer crimes may be perpetrated and arrests made on the basis of surveillance may lead to a greater probability of conviction.

Research Design and Methodology Dynamic adjustment of video parameters involved three components. The first was automatic image quality evaluation. The second was compensation for various image characteristics which are suboptimal in the original image but could be improved using known image processing algorithms. The third component was real-time adjustment of imaging parameters via a feedback mechanism to the camera.

Tracking a face across multiple video frames allows for better performance of the system in terms of recognition. Current cameras have the ability to capture images at a very high rate, yet no currently available surveillance-CCTV system is taking advantage of this technology. Taking advantage of the statistical information that can be gleaned from multiple matching operations, and using a number of different views of the face that can be acquired over the course of time, has allowed us to improve accuracy of identification. We developed tracking algorithms, decision rules for defining an individual orbit and inference rules for summing information over time that allowed us to achieve a qualitative performance improvement.

Summary of Adaptive Surveillance Project

Accomplishments

1. Developed mechanisms for dynamic adjustment of video parameters in region of image containing a face – adaptive local control
2. Created algorithms that track a face to acquire multiple images of it, across video frames
3. Improved speed and accuracy of automatic face finding technology
4. Realized, through combination of above, an improvement in performance of facial one-to-many matching, i.e.identification

Outcome

State-of-the-art, automated face recognition surveillance system, capable of providing value to law enforcement community.

The automation of attentionally-taxing duty of surveillance may lower overhead and thus, free up resources for performance of other tasks.

This system will allow law enforcement to perform surveillance functions at level of efficiency and precision beyond that which is possible at present.

Highlighted Technical Advances

1. Dynamic Adjustment of Video Parameters

Simple, powerful idea: devote processing resources to only part of image of value for face detection and recognition, and thereby dramatically improve signal-to-noise in that area

Rationale: Using fast face detection, IQ evaluation around face, image enhancement and camera control feedback improved quality of image where it counts: on faces.

2. Tracking Across Multiple Frames

Idea: Exploit information in 4th dimension, **time**, a source that has not yet been tapped

Rationale: Development of improved tracking algorithms, decision rules for defining an individual orbit, and inference rules for summing information over time allowed us to achieve performance improvement.

Final Deliverables: Due and Complete on September 30, 2000

1. Working software module and documentation for a system that works as an actual surveillance tool for use by the law enforcement community (in conjunction with an existing CCTV set-up).

2. Spin-Off Application: FaceIt® DB

An offshoot of the surveillance system described above is a mechanism for post-event analysis of the video footage. This software module, to be used for off-line analysis, benefits from all of the previously outlined developments (Image Quality Evaluation, Dynamic Camera Adjustment, Tracking) and thus, allows relevant personnel to maximize information content gleaned from a given video segment.

Background

CCTV Surveillance: the Tool and its Limitations

Over the last several years CCTV technology has emerged as a useful tool for law enforcement. Many cities in the United States and many more town centers throughout the United Kingdom have installed CCTV cameras in public areas. Law enforcement agencies now rely on these surveillance systems to allow relatively small forces to police much larger areas. Investigators also rely on these systems in crime investigations and evidence gathering from taped footage. The end result has been a dramatic decrease in crime[1] and a restored sense of public safety in areas that have traditionally been crime ridden.

As the utility of CCTV grows, new challenges are emerging. Most significant among them is the tedious labor required to monitor the proliferating screens in a typical CCTV control room. In Newham, East London there are currently 240 cameras installed throughout the borough. It is impossible for a single human or even a team of half a dozen people to effectively monitor at all times the video output of all these cameras as it is fed into the borough's CCTV control room. However, without the ability to attend to the surveillance video and also to make accurate identifications from it, the utility of CCTV surveillance and its deterrence factor will diminish over time as criminals discover that most CCTV screens are going unmonitored by operators at the control centers.

[1] According to Bob Lack, the Head of Security for Newham borough in London, crime has dropped in their borough by 50-75% since they have installed CCTV surveillance systems. Testimonial was aired by Discovery Channel, DateLine NBC, etc. Similar reports have been issued on Baltimore CCTV systems.

Automated Human Detection and Facial Recognition

As was correctly anticipated by NIJ's solicitation, the answer to the CCTV challenge is the use of advanced software algorithms that can automate many of the monitoring tasks and will alert operators when "relevant" events occur. Among the class of algorithms that have been advanced commercially are real-time facial recognition systems.

Visionics has developed software called FaceIt® Surveillance that is designed to address surveillance needs in CCTV control rooms. The current surveillance system automatically detects faces from multiple video feeds and alerts the operator to the presence of human faces (via an audio signal). More importantl,y it automatically crops the faces from the video image and matches them against a "watch-list" database of police-designated individuals. The system alerts the operator if a match exceeds a certain operator-adjusted confidence threshold.

The search engine underlying FaceIt® Surveillance is the same database engine used by the mugshot systems in operation at many police departments in the United States and throughout the world. (See: List of Implementations). The basic matching technology was enhanced through the support of NIJ under a Gang Tracking Project. As a result of those enhancements, there is now powerful yet affordable facial matching technology that is fully commercial and is already available to local law enforcement groups as a commodity through several vendors (e.g. Imageware, PrinTrak, ANADAC and Public Sector Products, who rely on the FaceIt® engine to enable their mugshot and booking software systems.)

Several implementations and pilots utilizing FaceIt® Surveillance are in place. One example is the much-publicized and publically heralded system in Newham, London. This program began in October, 1998 and has yielded a tremendous amount of feedback that we use to continually improve performance. Another example installation is at a U.S. border crossing (agency and site

are unnamed as a condition of customer). FaceIt® Surveillance is also in operation in the CCTV control rooms of multiple casinos in the United States, where convicted card-counters comprise the security watch-list. Finally, numerous pilots are on-going via a partner, ANSER, as part of the SF-HIDTA initiative that NIJ is sponsoring. A major development program is now underway with DARPA[2].

The initial feedback from these fully implemented systems and pilots is encouraging and shows that our current automatic capture of faces in the scene and subsequent comparison of those against a watch-list are useful functions for which there is desperate need on the part of law enforcement agencies. However, due to the fact that many challenges are posed by operational scenarios in which surveillance is often performed, we felt that the existing system would benefit greatly from further development. It was our opinion that the technology had demonstrated significant promise and utility and that through a focused development initiative, the technology could be pushed towards accomplishing its full potential. At this point in time, having completed development of the Active Surveillance system, it is our contention that the technology is now much closer to that goal.

[2] The project with DARPA is aimed at developing a system that will identify non-cooperative and un-cooperative subjects at a distance, either alone against complex backgrounds or in crowds.

The Need to Advance the State of the Art

The ultimate problem faced by an automated surveillance system stems from the quality of the facial images captured by the software. In common surveillance situations, the imaging environment is uncontrolled; lighting and shadows are highly variable, glare can be problematic and viewing distance and angle available from a stationary video camera often do not allow for acquisition of images of sufficient quality. In addition, the pose of the individual relative to the camera is usually sub-optimal for facial imaging. The typically difficult imaging environment is not in itself an obstacle for face finding systems. FaceIt® includes advanced algorithms that can perform face finding under challenging viewing conditions. However, poor image quality does indeed affect matching accuracy and the effect is significant.

To improve the quality of video images available for use in computerized face recognition, we perceived a basic need for novel approaches to improving the signal-to-noise ratio of the image in the video stream. We produced an approach that was cost-effective because it was based on pure software and continues to rely on standard off-the-shelf cameras.

We developed intelligent software for improving signal-to-noise in CCTV video input through the use of two basic processing algorithms:

(1) Dynamic adjustment of video parameters in face regions:

Facial matching performance was improved by actively controlling image quality in the area of the video footage that is most valuable, i.e. the detected faces. For this component of the work we developed a software Controller module that automatically evaluates image quality when a face has been detected and performs real-time adjustment of imaging parameters via a feedback mechanism to the camera system.

(2) Tracking: multiple facial recognitions across multiple frames

The software utilizes "tracking" to allow the system to acquire multiple images of a given face as it moves from one frame to another. The system integrates matching information across these multiple instances in order to enhance the probability of a correct match.

In what follows, we give a detailed description of the work that was performed.

Conceptual Summary of Work

Dynamic camera adjustment:

The idea here is simple yet powerful—in facial recognition, the only part of the image of value is that part which contains the face. So by concentrating video resources on that area – as described next – we dramatically improved the signal-to-noise ratio in the image. For this system to work, we needed three basic algorithmic components:

(1) Automatic face detection in complex scenes

(2) Automatic image quality evaluation

(3) Controllable camera

The capability to perform automated facial detection in complex scenes has been well established and is highly developed in the FaceIt® developer kit. Among the tasks that we focused on was optimizing the speed of this capability for scenes with multiple faces. This component of the project also benefited from the ongoing effort for improvements in the face finding algorithms. (Visionics has recently discovered two new algorithms for detecting faces.)

The second capability is equally important—before we could improve image quality we needed to have a quantifiable assessment of it. We have previously developed a so-called AFQES module (Automated Facial Quality Evaluation System) which is software that takes as an input the sub-

region of the frame in which a face has been detected. The software automatically measures the quality of the image, providing as output a quantified assessment of a large number of factors. These include: spatial resolution, dynamic range within the face, lighting estimates, pose, blur, and other characteristics directly related to the probability that the face in the captured image will be successfully registered and accurately matched to an image in the watch list. (Successful registration, or "alignment", is a critical step in recognition and depends upon visibility of a sufficient number of landmarks on the face.) The AFQES module is proprietary to the Visionics technology, but has been subjected to extensive testing through the INS IDENT project, which proposes to incorporate the AFQES module into 931 sites in the US.

Once image quality has been assessed, the output of AFQES is used to compensate for defects in the image. There are two mechanisms by which we compensate and these comprise the automatic "Controller" developed for this project:

(1) intrinsic compensation in software

(2) adjusting camera parameters through feedback

Using the first mechanism, we utilized a host of image preprocessing techniques available to us that allow compensation for detected variability to create what is termed the "canonical image". These techniques include Kalman filtering, deblurring, pose compensation (in instances of deviation less than 30 degrees), histogram equalization and glare elimination.

The second mechanism utilizes information provided by the AFQES to dynamically modify the video resources. We developed an information inference model, which acts upon the output of the AFQES to continuously adjust a number of camera parameters including pan/tilt/zoom, gain control, back-light compensation and white balance. In more recent models of surveillance cameras, all of these parameters can be adjusted digitally and without any mechanical parts, thus allowing for rapid adjustment of camera resources.

Tracking of faces across images:

There was a dimension to performing facial surveillance from CCTV that had not yet fully been exploited prior to this project. In normal surveillance conditions, a suspect appearing in front of a camera will elicit more than a single picture of his or her face. Typical video capture technology allows for the possibility of 30 frames per second. This provides for the opportunity to capture the suspect's face 300 times in a typical ten second duration. No surveillance system existed that even attempted to take advantage of this fact. Our own current surveillance system simply treated each instance of the face as if it were fresh and new. We reasoned that if one were able to track a face as it moved from one frame to another and also to know that it is the same face even though other faces may have entered or exited the scene, then one could very significantly improve the chances of correct matching. This was hypothesized for two reasons:

(1) Higher confidence because of statistics: multiple recognitions can be performed on a given face and then the results can be analyzed using statistical detection theory. The distribution of the matching scores for the multiple recognitions can be used to lower the overall equal error rate and hence improve the probability of correct identification.

(2) As the person moves from one location to the next within the field of view of the camera, additional information in the face region will inevitably emerge from one frame to another. For example, as the system is locked on a face, the full face may become visible for a portion of the subject's path.

To accomplish our goals in this area, we developed the following:

(1) Tracking algorithms for multiple faces, even when the face is not frontal

(2) Rules for deciding how the orbits of individual people begin and end

(3) Recognition lock-on for the duration of the orbit

(4) Inference rules for summing information across frames and for using multiple measurements to improve recognition

Spin-Off Application

In addition, an off-shoot of the surveillance system described is a mechanism for post-event analysis of the video footage. This software module, to be used for off-line analysis, benefits from all of the previously outlined developments (Image Quality Evaluation, Dynamic Camera Adjustment, Tracking) and thus, allows relevant personnel to maximize the information content gleaned from a given segment of video.

List of Tasks

Task 1. Improved face finding speed when multiple faces are present in field of view
This entailed the exploration of some new head and face finding algorithms to decide on their suitability for this task. The initial head finding algorithms were completed re-written for optimization. The face finding technology was also improved by the addition of some new algorithms and clean-up of some of the existing code.

Task 2. Developed the Image Quality Feedback system
An automatic Controller was implemented which uses face detection as well as AFQES to decide what type of image adjustments are required. Algorithms were developed to compensate for any image factors not requiring camera adjustment.

Task 3. Designed control signals
Converted the output of the Controller to signals that are used to perform image adjustment including software transforms as well as camera feedback. An information inference model was developed to usee the output of the AFQES to continuously adjust camera parameters including pan/tilt/zoom, gain control, back-light compensation and white balance.

Task 4. Researched Hardware
Researched the performance of the surveillance system when connected to various hardware configurations.

Task 5. Tracking

5a) Developed techniques for combining information obtained by tracking a face from one video frame to the next

5b) Developed tracking algorithms for following multiple faces

5c) Formulated rules for deciding how the orbit of an individual person begins and ends

5d) Implemented a means for recognition lock-on for the duration of the orbit

5e) Derived inference rules for summing information across frames and for using multiple measurements to improve recognition

Task 6. Create software and supporting materials

Designed software package that incorporates all of the above elements and enables a user to install and begin operating the system. Modified previous graphical user interface to make it more intuitive. Wrote installation scripts and user manual.

Task 7. Developed off-line search system

Simultaneously developed off-line database search system for post-event analysis of the video footage (the spin-off application referred to above).

Research and Development

Research was performed in a number of key areas in order to optimize recognition performance for the task of automated surveillance. These areas are described below.

1. We worked extensively on improving face tracking. Improvements are observable in both speed and robustness. Detail is provided below.

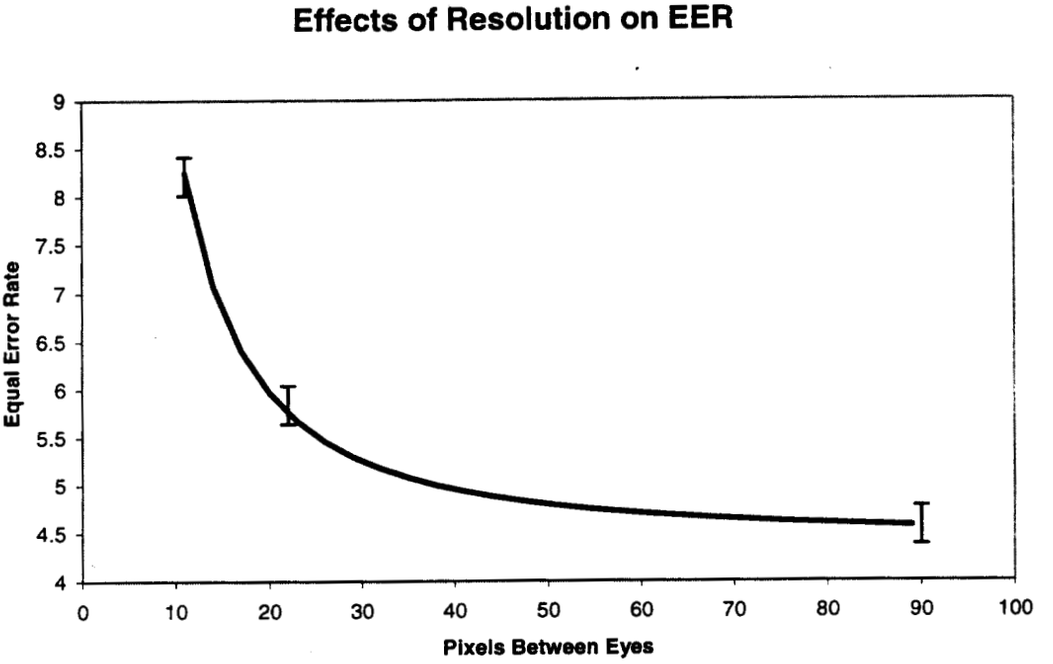
2. Parallelization concepts were used to speed up basic face finding on Intel-like platforms. We achieved a speedup of 150% in core face finding algorithms.

3. Image-processing ideas related to histogram equalization were implemented in order to improve the robustness of face recognition in arbitrary lighting conditions.

4. An improved background veto mechanism was developed in order to eliminate the need to search for faces in large static/background areas within the video field of view.

5. Recognition performance was improved via a number of approaches:
   A. Increase resolution via camera control

      The resolution of the facial image is strongly correlated with facial recognition performance. The graph below shows how the Equal Error Rate (EER) is related to the resolution of the images:



To increase resolution, camera control was developed to automatically pan, tilt, and zoom into a face. In order to zoom to a face, the 3D position of the face must be determined. We accomplish this by first finding the face at low resolution, then using the position and distance between the eyes; we can work backwards to estimate the global position of the face. One obvious problem with zooming in on a face is the restricted field of view. A second camera was added as a wide field of view camera to solve this problem. In the current implementation, the wide field of view camera finds faces and feeds that information to the pan/tilt/zoom camera. The wide field of view camera controls the zoomed camera.

This work required extensive camera testing and calibration since the wide field of view camera can be panned and zoomed arbitrarily with respect to the zoomed camera. If automatic zooming is not used, extra camera can be used to make a dual camera surveillance system with each pan/tilt/zoom camera working independently.

Since the wide field of view camera must find faces at very low resolution (less than 10 pixels between the eyes), an improved face finding algorithm was developed that was optimized for low-resolution video. The algorithm implemented is faster and more robust at low resolution than the previous face finder. The ability to find multiple faces in an image was also improved.

On the application side, the entire face recognition pipeline had to be re-designed to handle the added load of an extra camera. The redundancy filter was improved to better reduce the number of unnecessary searches.

B. Improve image quality via camera control

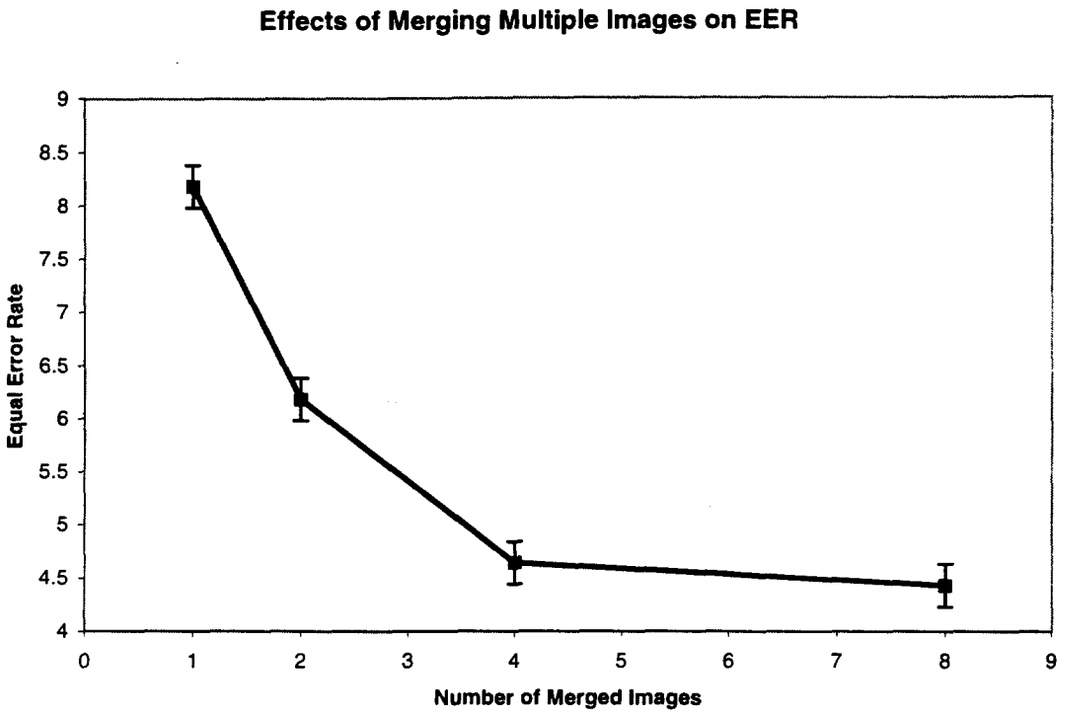
The rule of thumb for image quality is that the images should look good to the user. They should be sharply focused and there should be no over-saturation. We found that if the average 8-bit grayscale intensity of the inner facial region (around the eyes and nose) was near a value of 130, recognition performance was optimal. A feedback loop was developed that controls the video capture cards' (Winnov Videum PCI) brightness and contrast settings reflected the optimal settings for recognition.

C. Take advantage of consecutive video frames for face finding

Given that consecutive video frames are only a fraction of a second apart, it is intuitive that adjacent video frames in a stream would contain redundant information. Furthermore, the orbit of an object over time should be apparent. In order to improve face finding, we capitalized on these facts in order to track a face over time. This improves the speed of face finding and therefore increases the chance that a high quality facial image is captured. Integrating tracking with face finding was one of the major obstacles since the search for new faces must continue while simultaneously tracking a previously located face.

D. Take advantage of consecutive video frames to build a facial template for face recognition

Closely related to tracking issues is the fact that all of the information needed to recognize a person above a threshold is sometimes not found in a single image. This could be due to any number of extrinsic factors. This missing information can sometimes be constructed by using multiple images of the same person (determined by tracking or redundancy filtering). The graph below shows the improvement in recognition achieved by combining information obtained from multiple images of the same person.



E. Improve face recognition algorithms to compensate for face finding under environmentally-challenging conditions.

A new face recognition algorithm based on our existing recognition technology was developed in order to compensate for face finding when the latter is performing sub-optimally due to challenging environmental conditions. The algorithm was removed from the final product because it was found to perform poorly relative to the original algorithm. This was contrary to the results obtained with test data. The problem was identified to be faces containing a particular kind of pose, which are not in the data sets used for testing. The original algorithms worked better for faces pitched forward or backward, i.e. – nodding of the head). As a result, the new algorithm will undergo continued development and will be added to future products.

Summarized User Documentation

(Readers are requested to consult complete hard copy documentation for full, illustrated version.)

Database Format:

The databases are compatible with those created with the previous version of FaceIt® DB/Surveillance.

Recommended Hardware:

The system has only been tested with these minimum requirements:

Computer

- At least a dual CPU 700 MHz computer
- At least 256 MB of RAM
- At least 1 Gb of free HD space
- Windows 2000
- Display resolution of 1024x786 or greater
- At least 1 free serial port

Video Capture

- 2 Winnov Videum VO PCI capture cards (installed with drivers 2.9.1 or greater)
- 2 Sony EVI-D30 cameras
- 2 RCA video cables
- 1 camera control cable (9-pin to VISCA)
- 1 camera link cable (VISCA to VISCA)

System Setup:

1. Install the 2 Winnov Capture cards as detailed in the Winnov documentation.
2. Mount the 2 Sony cameras such that they are on the same horizontal plane, but there is about ½ inch between them. They should be mounted approximately eye height.
3. Connect an RCA video cable from the input of one of the capture cards to the output of one of the Sony cameras. Likewise, connect the input from the other capture card to the output of the other Sony camera.
4. Connect the 9-pin to VISCA cable from the computers serial port (COM 1) to one of the cameras VISCA input.
5. Connect the camera link cable from the VISCA output of the camera that is connected to the computer's serial port to the other camera's VISCA input.

6. Change the address of the last camera in the chain to camera 2 with the switch on the back of the camera (and make sure the other camera is set to camera 1).
7. Plug in the camera's and power them on. The hardware installation is now complete.
8. Install the FaceIt® ActiveSurveillance software.

After obtaining a license string, the software and hardware will be ready to use. You may need to calibrate the system for optimal performance.

Performance Specifications:

Using the recommended hardware the sustained rate of capture can be up to 1 face per second (without using redundancy filtering).

The burst rate of capture is approximately 10 faces per second.

Face capture distance can range between 2 and 200 feet.

Improvements on FaceIt® DB/Surveillance:

FaceIt® DB (offline recognition):

- Automated alignment of faces now uses algorithms designed to work with no prior knowledge of the scale of the face in the image.
- Faster video capture has been added. This allows one to grab images at a very high rate so as to ensure that all frames with a visible face will be digitized.
- Search options:
    - Vector only (scan) search is allowed with no second (intensive) pass. Fastest search. Not recommended for all operating scenarios, but works well for frontal images.
    - Option added to merge scores when using tracking or redundancy in surveillance

FaceIt® Surveillance:

- Added support for up to 2 cameras. This forced a complete re-write of the Face Capture pipeline including new algorithms and new surveillance threading architecture.
- Incorporated the use of time information into face capture
    - Face orbits are tracked from frame to frame to help improve face finding
    - Face recognition can take the orbit information from tracking and use it to construct a score using composite information from multiple video frames.

- Incorporated new head finding technology. Head finding is performed before we try to find the eyes. This is a critical step, because if no head is found, we will never find the eyes. In order to improve the face finding of small faces while at the same time improving the speed, a new class was implemented. This significantly helps when finding multiple faces in one frame. It also allows for integrating tracking technology into face finding technology.
- An improved redundancy filter was designed that better learns the most recently seen faces.
- Hardware control is used to improve the resolution of faces found. The face finding can be linked between 2 cameras. One camera is designated the controller and is used to scan a wide field of view, while the other zooms in on faces found by the controller in an attempt to get a higher resolution image of the face.
- Pan, tilt, and zoom control has been integrated into surveillance so that the operator can manually adjust the camera on the fly to adapt to changing situations by simply clicking on the video image. The ability to Home the camera was added as well.
- Added COMM port setup for camera control.
- Added camera geometry dialog to tweak camera configuration.
- Hardware control of the capture card allows the automatic adjustment of brightness and contrast for captured faces. The adjustment of the brightness and contrast improves the recognition performance.

Graphical User Interface

The following image is a screen capture of the application. In the middle is the live video display. The yellow circle overlaid shows a face that has been captured and is currently being tracked with the lock-on mechanism. On the left of the video display, faces captured over time are shown. (The redundancy filter is turned off, so multiple images of the same face will be displayed.) The colored horizontal bars under the face captures are indicators of the confidence that the captured face will lead to good recognition results.

On the right of the video display, face captures for which match searches have been conducted are displayed. Clicking on any one of those brings up a match results display, in which the most likely matches in the database are displayed in decreasing order of similarity (not shown here). Here, he colored horizontal bars under the face captures are indicators of the confidence that the system has a matching face in the database.

Concluding Statement

The combination of the software engineering and scientific research described above has resulted in a system that significantly extends the state-of-the art in face recognition-based surveillance. This system will be a valuable tool for law enforcement and intelligence personnel and for CCTV control room operators. We thank the National Institute of Justice for their generous support.