Face Recognition at a Distance System for Surveillance Applications

Frederick W. Wheeler, Richard L. Weiss and Peter H. Tu

***Abstract*— Face recognition at a distance is concerned with the automatic recognition of non-cooperative subjects over a wide area. This remote biometric collection and identification problem can be addressed with an active vision system where people are detected and tracked with wide-field-of-view cameras and near-field-of-view pan-tilt-zoom cameras are automatically controlled to collect high-resolution facial images. We have developed a prototype active-vision face recognition at a distance system that we call the Biometric Surveillance System. In this paper we review related prior work, describe the design and operation of this system, and provide experimental performance results. The system features predictive subject targeting and an adaptive target selection mechanism based on the current actions and history of each tracked subject to help ensure that facial images are captured for all subjects in view. Experimental tests designed to simulate operation in large transportation hubs show that the system can track subjects and capture facial images at distances of 25–50 m and can recognize them using a commercial face recognition system at a distance of 15–20 m.**

I. INTRODUCTION

For a large variety of commercial, security and defense applications there is a need to recognize people in large open areas with dimensions of 10–20 m or even much larger, without subject cooperation. Face recognition is the most viable biometric modality for such applications. The face is generally visible, and ordinary cameras and optical equipment can image the face from a distance without the knowledge of the subject. This field is called Face Recognition at a Distance (FRAD). However, in addition to distance, key additional challenges are the large coverage area and biometric recognition with imperfect facial images. Face recognition at a distance will enable watch-list recognition for security at terminals and critical infrastructure, white-list recognition to detect intruders, and re-recognition for pervasive person tracking over a large area covered by an intelligent camera network.

For large coverage areas, FRAD can be accomplished with a multi-camera active vision system. One or more Wide Field Of View (WFOV) cameras image a large area to allow the detection and localization of persons. One or more Narrow Field Of View (NFOV) cameras are then actively controlled to capture facial images with high resolution using pan, tilt and zoom (PTZ) commands.

In this paper we will describe an active-vision FRAD system we call the Biometric Surveillance System. We begin with reliable detection and tracking of persons from fixed camera surveillance video (WFOV). Subjects are tracked in the ground-plane of a real world metric coordinate system with a Kalman filter, enabling the prediction of the location where their facial image can be captured by an automatically controlled PTZ camera (NFOV). The person tracking system handles many simultaneous subjects in view. A priority scoring system is used to select a person to target for imaging, causing the system to rapidly move from person to person, ensuring that facial images are collected from all subjects in the coverage area. Collected facial images are processed by a commercial face recognizer. In performance evaluations, persons are detected and tracked at distances in the 25–50 m range and recognized at distances of 15–20 m.

An important and novel feature of this system is the way biometric recognition is integrated with a persistent and reliable ground-plane person tracker. The tracker can maintain a lock on an individual for the entire time they appear in the WFOV camera. Captured facial images and recognition results are all associated with a tracked subject

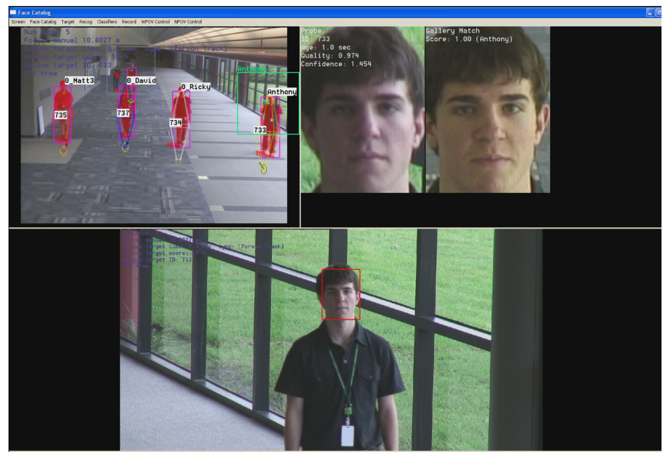

Fig. 1. A highly annotated user interface shows tracked subjects in WFOV video (upper left), the NFOV video with face detection results (bottom), and all captured facial image with face recognition results as they are processed (upper right). Each tracked subject in the WFOV video is labeled with a name determined via face recognition on facial images acquired during tracking.

This report was prepared by GE Global Research as an account of work sponsored by Lockheed Martin Corporation. Information contained in this report constitutes technical information which is the property of Lockheed Martin Corporation. Neither GE nor Lockheed Martin Corporation, nor any person acting on behalf of either;
a. Makes any warranty or representation, expressed or implied, with respect to the use of any information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or
b. Assume any liabilities with respect to the use of, or for damages resulting from the use of, any information, apparatus, method, or process disclosed in this report.

Frederick W. Wheeler (wheeler@research.ge.com), and Peter H. Tu (tu@research.ge.com) are with GE Global Research, Niskayuna, NY, USA. Richard L. Weiss (richard.weiss@lmco.com) is with Lockheed Martin IS&GS Civil, Orlando, FL, USA.

Fig. 2. The Biometric Surveillance System prototype on a large portable cart with two raised camera nodes (left), and a close-up view of one node (right).

ID in internal data structures. This enables the accumulation of identity information over a long time frame and allows the target selection system to choose tracked subjects who are not yet identified. The system has many configurable operating modes, including an auto-enrollment feature, and network-based sharing of auto-enrollment data to enable re-identification of subjects as they move from one camera zone to another. Fig. 1 shows a screen shot of the prototype application with several persons tracked and identified, Fig. 2 shows the system hardware in use, and Fig. 3 shows some examples of captured facial images from an outdoor test session.

II. PRIOR WORK

There has been a considerable amount of earlier effort on the active vision approach to face recognition at a distance. In this section we review prior work in this field, focusing on efforts related to automatic targeting and camera control for the purpose of facial image capture and recognition.

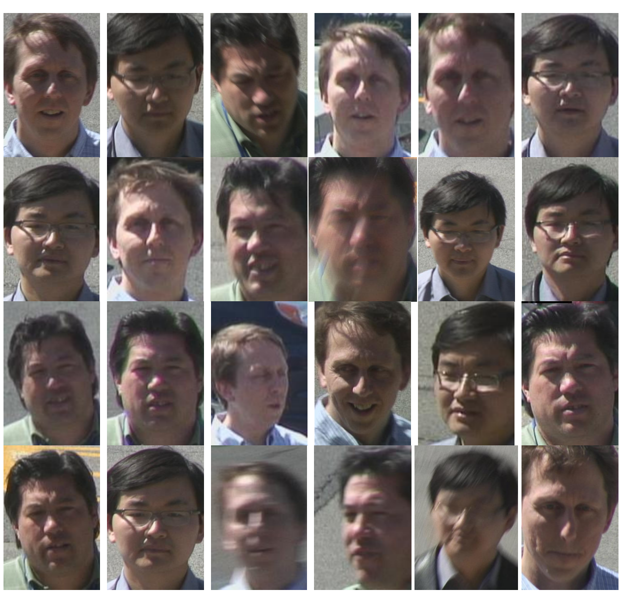

Fig. 3. A small subset of the 100's of facial images captured by the Biometric Recognition System over a few minutes at a 10–20 m range outdoors. This sampling shows the pose variation, illumination variation, and blurring that can occur when imaging non-cooperative subjects.

Stillman et al. [1], in some very early work, developed an active vision system for person recognition using a pair of WFOV cameras and a pair of NFOV cameras. This system was developed for human computer interaction applications and worked under some limited conditions over a range of a few meters, detecting people based on skin color, using triangulation to determine 3D locations, and automatically controlling the NFOV cameras to capture facial images. Greiffenhagen et al. [2] developed a real-time two camera face capture system where an overhead omnidirectional camera is used as the WFOV camera to determine subject location and a PTZ camera is then directed at the subject's face.

The Distant Human Identification (DHID) system was developed by Zhou et al. [3] and collects video of humans at a distance, for identification via both face recognition and gait. This system uses a single WFOV camera with a 60° field of view. Persons are tracked out to a distance of about 50 m. Short zoomed-in video sequences are captured for gait recognition and images are collected for face recognition. A two-camera face capture at a distance system has also been developed by Marchesotti et al. [4], with persons detected in the WFOV video using a blob detector and a NFOV camera that is panned and tilted to acquire short video clips of subject faces.

A *face cataloger* system has been developed and described by Hampapur et al. [5], [6]. For person detection, this system uses two geometrically calibrated WFOV cameras with overlapping views of a 6 m by 6 m capture area. A 2D multi-blob tracker is applied to each WFOV camera view and a 3D multi-blob tracker uses these outputs to determine 3D head locations in a real-world coordinate system.

Prince [7], [8], Elder [9] et al. have developed an approach to face capture at a distance with a goal of being robust to pose and partial occlusion of subjects. They make use of a stationary WFOV camera with a 135° field of view and a NFOV camera with a 13° field of view. For robustness to occlusion, faces are detected in the WFOV camera view instead of whole bodies. Faces are detected using a combination of motion detection, background modeling and skin detection. The NFOV PTZ camera is then directed to the detected faces for higher resolution facial image capture.

Bellotto et al. [10] describe an architecture for face capture with an active multi-camera surveillance system. Person trackers are associated with each WFOV camera, and high-level reasoning algorithms share data via an SQL database. Person detections from WFOV trackers are used to automatically assign and target NFOV cameras to particular subjects. Using the NFOV cameras, faces are tracked and the NFOV camera follows the face with a velocity control system.

Also at GE Global Research, Krahnstoever et al. [11] have developed a multi-camera tracking framework and prototype face capture at a distance system for surveillance applications. Four WFOV cameras with overlapping viewpoints are used for person detection and tracking over a $300\,m^2$ coverage region. Tracking is performed in a real-world coordinate frame, and controls the targeting of four NFOV PTZ cameras that surround the monitored region. In follow-on work, Yu et al. [12], have used this system for long-term monitoring of groups of people. Face recognition established an identity for tracked individuals and the system evaluates the degree of close interaction between individuals to learn the social network.

While several prior efforts have had the same goal of capturing facial images over a wide area for recognition at a distance there are a few novel aspects of our system that we will focus on in this paper. The person detection and tracking component of the system uses a Kalman filter for tracking in the real-world coordinate system ground plane. This enables tracking across momentary occlusions, and also allows the prediction of subject location. Instead of targeting the NFOV camera to where the subject is now, the system predicts the subject location at a target time about 0.5–1.0 sec. in the future and directs the NFOV PTZ camera to that location and holds until the target time arrives. Thus, facial images are captured even for rapidly moving and running subjects.

Because the ground-plane person tracker is reliable, information collected about each subject can be accumulated. The system stores all attempts to capture a facial image, captured facial images, facial image quality measurements and recognition results in a data structure associated with a tracked subject ID. This enables the system to perform auto-enrollment, where subjects who are not recognized may be automatically enrolled, with a simple numeric identifier, so that they may be re-recognized later by the system.

III. SYSTEM

A. Hardware

A single unit of the system is called a *node* and consists of a co-located pair of WFOV and NFOV cameras with a high-end but standard workstation containing Matrox® frame grabbers to accept the video streams. Each camera is a Sony EVI-HD1, which is configurable to several video modes, has an integrated lens, and has pan, tilt, zoom and focus settings controlled by the computer using a serial line VISCA™ interface. The WFOV camera is operated at a resolution of 640 by 480 at 30 Hz (NTSC), and is held at a fixed pan, tilt and zoom. The NFOV camera is operated at a resolution of 1280 by 720 at 30 Hz, and its pan, tilt and zoom setting (up to 10x optical) are actively controlled. Fig. 4 shows a system diagram of the main processing components of the multi-threaded software application for the system. In this section we will describe these components.

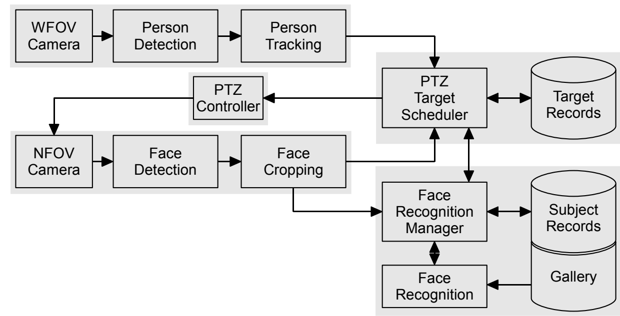

Fig. 4. System diagram showing main computational components of the Biometric Surveillance System. Shaded regions identify groups of components that are executed in a separate processing thread.

B. Person Detection and Tracking

People moving in the field of view of the stationary WFOV camera are detected and tracked. The camera is stationary, so a background subtraction approach is used for moving object detection. We use an adaptive parametric model for the distribution of each color component of each pixel. Grayscale video may be used as well, but the additional degrees of freedom of color video increases the person detection rate. Any pixel that does not conform well enough to the model is declared to be a foreground pixel. The internal and external parameters, which include the focal length, principal point, location, and orientation of the WFOV camera, are determined through a calibration procedure. This yields a mapping from real-world metric coordinates to the WFOV camera video frames. With the assumption that people are walking upright and using the camera calibration information, feasible regions that may contain a whole person in the video are readily determined. Clusters of foreground pixels that match these feasible regions are assumed to be detected persons. The person detection process operates on the WFOV video at about 10 Hz. An extended Kalman filter [13], [14] is applied to these detected persons in the ground-plane. This makes the system robust to momentary occlusions and provides a velocity estimate for each tracked subject, allowing for the prediction of future subject locations.

C. PTZ Controller

The NFOV camera view is calibrated with respect to the WFOV camera view. With the NFOV camera in its home position, with its pan and tilt angles at 0° and its zoom factor set to 1, point correspondences are used to estimate a homography relating the location of a point in the WFOV camera view to the same point in the NFOV camera view.

The NFOV camera is additionally calibrated to determine how the pan, tilt and zoom settings affect its field of view. A subtle but critical part of this calibration is the camera's *zoom point*. The zoom point is the pixel location that always points to the same real-world point as the zoom factor is changed. This is nominally the center of the image, but not precisely, and in our experience the exact point can vary from device to device of the same model. Even a small offset of the zoom point can affect targeting accuracy when a high zoom is used for distant subjects. With this calibration information, once a target location and region size is determined in the WFOV video frame, it is straightforward to determine the NFOV pan, tilt and zoom settings that will result in that region filling the full NFOV video frame.

D. Target Scheduler

Multiple subjects may be detected and tracked in the low-resolution WFOV video. The system uses a priority mechanism to select a target for high-resolution face capture using the automatically controlled NFOV PTZ camera. The targeting priority is based on both the history of the subject and the current state. For each tracked subject a *target record* of prior actions and results is maintained. This record includes the number of past targeting attempts, the number of facial images captured by the face detection module, and the number of successful recognitions by the face recognition manager module. From the person tracker Kalman filter several current state parameters are determined. These parameters are: the distance from the camera node to the subject, the direction cosine, and the speed of the subject. The direction cosine is the cosine of the angle between travel direction and camera direction and indicates the extent to which the subject is facing the camera node.

Using this information, each tracked subject is scored, and the subject with the highest score is selected as the next target for facial image capture. To produce the score, for each parameter a multiplicative factor is applied and the result is clipped to a certain range and added to the score. Table I shows the complete set of parameters and factors. For example, the subject's direction cosine is multiplied by the factor 10, clipped to the range [-8,8] and added to the score. Thus, subjects facing the camera have this increase in priority. Similarly, a subject's speed in m/sec. is multiplied by 10.0, and clipped to the range [0,20]. Subjects moving quickly are more likely to leave the coverage area and may be of greater interest, so they are given an increased priority. The number of capture attempts, successful facial image captures and successful recognitions for a subject all have a negative factor. Each of these events reduces a subject's priority to prevent the system from repeatedly targeting the same subject. A subject for which many facial images are already captured will have a reduced priority. Further, the clipping range for each of these factors prevents any single factor from dominating the score. The overall effect of this process is to favor the targeting of subjects moving more quickly toward the camera node who have not yet been satisfactorily imaged. In practice, this target selection mechanism causes the system to move from subject to subject, with a tendency to target subjects from which we are most likely to get new and useful facial images.

Parameter	Factor	Clipping Range
Direction cosine	10	[-8,8]
Speed (m/sec.)	10	[0,20]
Prior capture attempts	-2	[-5,0]
Prior face captures	-1	[-5,0]
Times recognized	-5	[-15,0]

TABLE I

PARAMETERS USED TO SCORE AND PRIORITIZE TARGETS, WITH MULTIPLICATIVE FACTORS AND CLIPPING RANGE. THIS PARAMETER SET WAS PRODUCED EXPERIMENTALLY AND IS NOT OPTIMIZED.

Once a subject is selected, the Kalman filter from the person tracker is used to predict the location of the subject's face at a specific fixed target time about 0.5–1.0 sec. in the future. The NFOV camera will zoom in and point to this location and hold its position until the target time has passed. This strategy allows for time to complete the pan, tilt and zoom change and for any camera vibration to settle. When the subject moves through the view of the NFOV camera as predicted, the subject's face and upper body will pass through the center portion of the NFOV video. Near the the target time, facial images will be detected and collected by the separate face detection module operating on the NFOV video. Once the target time has passed, a new target is immediately selected. The system thus directs the NFOV camera to a new person about every 1–2 sec.

Besides selecting a target and determining where to point the NFOV camera, the system must also select a NFOV camera zoom factor. There is a trade-off between the zoom factor and the probability of successfully capturing a facial image. Increased zoom will result in higher resolution facial images, but with any error in subject tracking or targeting, there is a higher chance that facial capture will fail because the subject does not pass through the resulting smaller field of view of the NFOV camera. The system uses an adaptive approach for zoom factor selection. If a subject has no successful face captures, the initial face resolution goal will be a modest 30 pixels eye-to-eye. Then, each time a facial image is successfully captured at a particular resolution, the resolution goal is increased by 20%. The system targets and images each subject repeatedly, so the facial image resolution tends to rapidly increase. The facial image resolution goal and the subject distance determine the zoom factor of the NFOV camera.

The NFOV camera has both automatic and manual focus modes. The subject distance is used to set the focal distance of the NFOV camera. Automatic focus also works reasonably well for the NFOV camera, but can occasionally focus on a distinctive object that is much further or closer to the camera than the targeted subject, causing the subject to be out of focus. Using knowledge of subject location and distance eliminates this problem. With our particular camera equipment and typical subject distances, the depth of focus

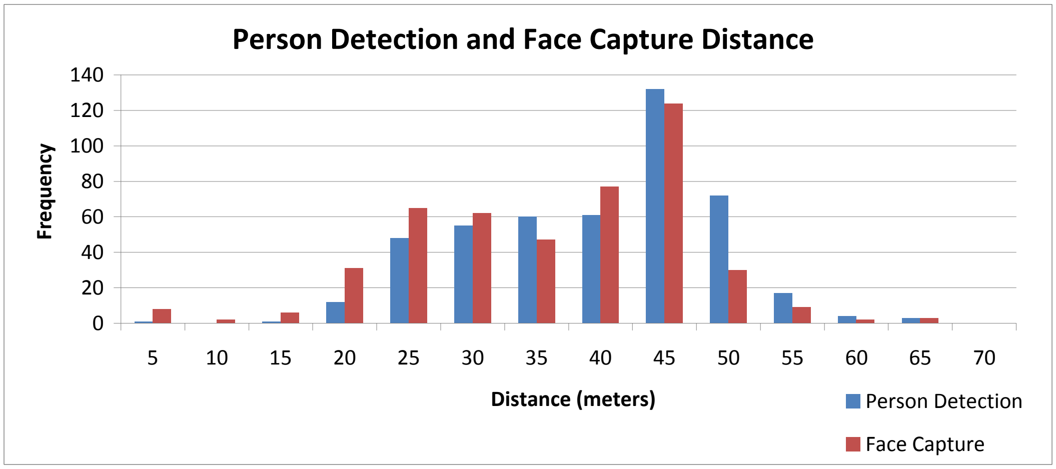

Fig. 5. Histograms of distance to first person detection and first face capture for test subjects.

is fairly large. So subject distance does not need to be known very accurately, and the focal distance of the camera is only adjusted if the desired focal distance changes by more than an adjustable threshold of a few meters.

E. Face Detection and Cropping

As the target scheduler directs the NFOV camera, the NFOV video stream is continuously monitored for facial images. In each NFOV frame, the Pittsburgh Pattern Recognition FT-SDK is utilized to detect faces. This process operates at about 10 Hz, with frames skipped to maintain real-time operation. If there is more than one detection in a single frame, we use only the most central face in the image, since it is more likely to be the face of the targeted subject. A detected face is cropped from the full frame image and passed asynchronously to the face recognition manager. The target scheduler is also informed of the face capture, so the target record can be updated.

F. Face Recognition

When the face recognition manager receives a new cropped facial image, a facial image capture record is created and the image is stored. The face recognition manager queries the target scheduler to determine which tracker subject ID a facial image came from, based on the capture time of the image. The target scheduler keeps a record of which subject was targeted at any given time in the past. With this information a *subject record*, keyed by the tracker ID number, is retrieved or created if it does not already exist, and the facial image capture record is associated with the subject record.

Facial images can be captured by the system at up to about 10 Hz, but face recognition generally takes 0.5–2 sec. per image, depending on the algorithm. Recognition cannot keep up with capture, so the face recognition algorithm is operated asynchronously. The system can be interfaced to Cognitec FaceVACS®, Identix FaceIt® SDK, Pittsburgh Pattern Recognition FTR-SDK, or an internal research face recognition system. In a processing loop, the face recognizer is repeatedly applied to the most recently captured facial image not yet processed, and results are stored in the facial image capture record. Unprocessed facial images are abandoned after about 30 sec. Face recognition can use a stored gallery of images, manual enrollments, automatic enrollments or any combination. Face recognition attempts and results are stored in the appropriate facial image capture record in a subject record and displayed in the application. Face recognition results are passed to the target scheduler so the appropriate target record can be updated. Face recognition results are also used by the GUI to annotate the WFOV video with the names of recognized tracked subjects.

The optional auto-enrollment feature of this system makes use of the subject records accumulated for each tracked subject. This is a configurable rule-based process with the goal of automatically enrolling a tracked subject when we have confidence that the subject is not already enrolled and we have a facial image suitable for enrollment. An auto-enroll candidate must have at least one facial image that has a quality score from the face recognition system exceeding a threshold, and recognition using the current gallery must have been attempted with at least 4 different facial images without success. This criteria helps ensure that the subject is indeed unknown. The most recent facial image capture for an auto-enroll candidate must have occurred at least 4 seconds ago. This criteria prevents auto-enrollment when a subject is still in view and additional facial images might still be collected. If a subject is an auto-enroll candidate, the facial image with the highest quality score is selected and enrolled in the face recognition gallery, after an optional user confirmation.

Nodes may be networked with a star topology and can communicate gallery images and data to share manual enrollments and automatic enrollments. This sharing of automatically enrolled subjects enables the tracking of subjects from one node's coverage area to another.

G. Summary of Operation

Subjects entering the area monitored by the WFOV camera are detected via background subtraction, tracked using a Kalman filter, and assigned person track numbers. A target score is calculated for each tracked subject that is currently in the WFOV camera's view using the method described in Section III-D. The subject in view with the highest total score calculated from the target's direction, speed of travel, and modified by previous face capture attempts, face capture successes, and recognition successes is selected as the next target for the NFOV camera. Scoring is implemented such that those subjects moving toward the camera at a high speed and for whom facial capture has not been attempted or succeeded often will have higher priority over slower moving targets for whom facial capture or recognition has already occurred several times.

Once target scores are calculated, the highest scoring subject is chosen, and the NFOV camera is moved to point at the predicted location of the subject's face 0.5–1.0 seconds in the future and held until that time. This gives the NFOV camera time to move to the location and prevents vibration or motion blur from affecting the images. At the target time, as the target moves through the field of view of the NFOV camera, the face detector will rapidly find and extract several facial images which are stored and queued for further processing. The number of capture attempts and successful facial image captures are stored with the subject record for the track. The number of successful facial recognition attempts is not immediately known due to the required processing time, but will be added to the subject record later. Once the target time has passed, target scores are recalculated for all tracked subjects still in the WFOV camera's field of view. The previous target now has a lower score due to the capture attempt just made, so the NFOV camera will likely be assigned to a different target, but this is not always the case.

While the system is targeting subjects and capturing new facial images, facial recognition processing is occurring asynchronously on the facial images previously captured and stored. The system currently supports multiple different commercial facial recognition software packages, any of which could be configured to process the collected images. Successful facial recognition attempts are passed to the GUI to be displayed to the operator and are also passed to the target scheduler so that the subject record can update the number of successful recognitions. If the auto-enroll feature is enabled the target may also be enrolled in the gallery for future matching.

IV. PERFORMANCE

In order to test the range capabilities of the Biometric Surveillance System an experimental set up was designed to mimic conditions that may be found in transportation hubs such as airports or subway stations. The system was stationed at fixed locations both indoors and outdoors. Indoor testing was primarily conducted in different sections of a long and wide (approximately 9 m) corridor illuminated primarily with natural light through windows and skylights with very little artificial lighting. This corridor is comparable in scale to an airport terminal and can be seen in Fig.1. Lighting conditions at this location varied due to the time of day and weather conditions, ranging from bright when the weather was sunny to somewhat dim when the weather was cloudy or night was approaching. Outdoor testing was conducted with the cameras facing both toward and away from the sun. Experimental trials were in sessions spaced approximately one week apart over the course of about six weeks giving some degree of variation in overall subject appearance with respect to person detection and tracking.

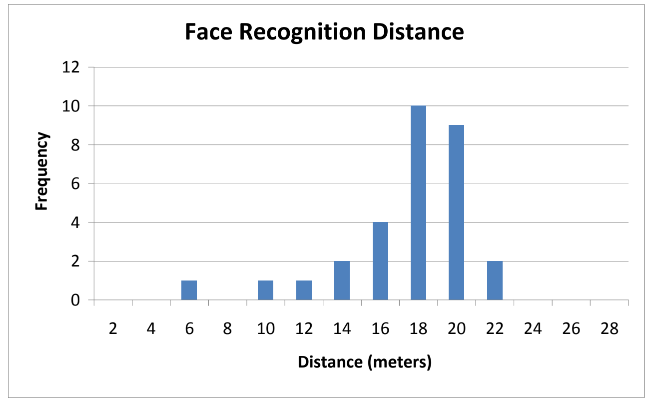

Fig. 6. Histogram of distance to first successful face recognition for test subjects.

A real-time integrated hardware/software system such as this can only be evaluated in a live scenario. In these experiments, test subjects walked in a natural manner through the corridor or outside area, beginning out of range of the cameras and moving in their general direction. An operator then signaled the subject to record the distance when the system first detected them, captured an image of their face, and successfully matched their face to an image of the subject in the system gallery, if that occurred, using periodically spaced distance markers.

For the first experiment to evaluate person detection and facial image capture performance, Table II provides data collection details, while Fig. 5 shows histograms of the distance to the subject at the the initial person detection (and tracker initialization) and initial face capture. Though

Parameter	Value
Subjects	2
Total Trials	466
Trials (Subject A)	233
Trials (Subject B)	233
Trials (Indoor)	386
Trials (Outdoor)	80
Experiment Days	6
Non-Subjects in Range	0–8

TABLE II

PARAMETERS OF THE EXPERIMENTAL TESTS.

the behavior of the test subjects was controlled there were varying numbers of other individuals walking through the area in all directions, as well as standing still in view of the cameras. These other individuals were not part of the experiment or even aware it was taking place. The number of other subjects in view had little effect on capture range of the system due to the scoring penalties for multiple captures of the same subject and the ability of the system to rapidly cycle through potential subjects. Observationally we noted equivalent person detection and face capture ranges for non-controlled passersby, though in this experiment data was only recorded for the test subjects. During the 466 trials completed for this experiment there were only 1 failed person detection and 8 failed face captures, usually due to the subject or their face being obscured in some manner.

For a second experiment to evaluate face recognition performance, Fig. 6 shows a histogram of the distance to the subject at the initial facial recognition using the Cognitec FaceVACS® facial recognition algorithm with a recognition threshold of 0.5. This experiment included 30 trials, all of which resulted in successful recognition of the subject. The face recognition gallery was created with 241 facial photos taken from a gallery of publicly available still face images as well as 27 locally enrolled images, including 3 of each subject in different conditions and the rest from other volunteers.

Table III shows the mean and standard deviation values for person detection, face detection, and facial recognition for the experiments outlined above. As can be seen, the average range for person detection and face capture is about twice that of facial recognition. This is expected due to the need for high quality facial images for facial recognition which is a limitation of current facial recognition capabilities. At higher ranges the captured facial images are more subject to motion blur from the movement of the subject and are also smaller in total size, making it more difficult for face matching software to perform adequately.

V. Conclusion

Face recognition at a distance with an automatically controlled camera system such as this Biometric Surveillance System enables biometric recognition at a significant range without the subject's knowledge or cooperation and offers a distinct advantage in many applications. While other biometric modalities such as finger and iris recognition typically require the cooperation of a subject and can be captured usually at distances of less than a meter, this Biometric Surveillance System allows tracking of subjects at distances of 25–50 m and recognition at distances up to 20 m. With the inclusion of a networked system of nodes once a subject is identified or enrolled they can be successfully labeled and tracked by any node with a coverage area that the subject enters.

The development of the Biometric Surveillance System has focused on facial image capture through subject tracking, target selection and automatic PTZ camera control, with the goal of providing high quality facial images to third-party face recognition software. The system can acquire 10 or more facial images per second in bursts. Faster facial recognition will improve overall performance by providing more immediate results and making use of more of the incoming data. The same benefit could come from rapid quality analysis as a preprocessing step. If facial image quality can be evaluated very quickly, the best facial images can be selected for face recognition processing.

Face recognition at a distance, over a wide and open coverage area has a large number of security and other applications. The experimental results demonstrate the viability of this prototype system for real-world applications. It would also be fairly straightforward to extend the range of a single system node by a factor of two or more through the use of improved cameras and optics. Of course, the effective coverage region can be extended indefinitely through the use of multiple networked nodes.

Capture Type	Mean (m)	Standard Deviation (m)
Person Detection	37.3	9.4
Face Capture	33.6	10.7
Facial Recognition	16.5	3.5

TABLE III

Mean and standard deviation of distances to initial person detection, face capture and face recognition.

References

[1] S. Stillman, R. Tanawongsuwan, and I. Essa, "A system for tracking and recognizing multiple people with multiple cameras," in *Proc. of 2nd Intl. Conf. on Audio-Vision-based Person Authentication*, pp. 96–101, 1998.

[2] M. Greiffenhagen, V. Ramesh, D. Comaniciu, and H. Niemann, "Statistical modeling and performance characterization of a real-time dual camera surveillance system," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, (Hilton Head, SC), June 2000.

[3] X. Zhou, R. Collins, T. Kanade, and P. Metes, "A master-slave system to acquire biometric imagery of humans at distance," in *ACM International Workshop on Video Surveillance* (ACM, ed.), December 2003.

[4] L. Marchesotti, L. Marcenaro, and C. Regazzoni, "Dual camera system for face detection in unconstrained environments," in *Proc. of the International Conference on Image Processing (ICIP)*, (Barcelona, Catalonia, Spain), 2003.

[5] A. Hampapur, S. Pankanti, A. Senior, Y.-L. Tian, L. Brown, and R. Bolle, "Face cataloger: multi-scale imaging for relating identity to location," in *Proc. of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 13–20, July 2003.

[6] A. Senior, A. Hampapur, and M. Lu, "Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration," in *IEEE Workshop on Application of Computer Vision (WACV)*, vol. 1, pp. 433–438, Jan. 2005.

[7] S. Prince, J. Elder, J. Warrell, and F. Felisberti, "Tied factor analysis for face recognition across large pose differences," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 970–984, June 2008.

[8] S. Prince, J. Elder, Y. Hou, M. Sizinstev, and E. Olevsky, "Towards face recognition at a distance," in *Proc. of the IET Conf. on Crime and Security*, pp. 570–575, June 2006.

[9] J. H. Elder, S. Prince, Y. Hou, M. Sizintsev, and Y. Oleviskiy, "Pre-attentive and attentive detection of humans in wide-field scenes," *Int. J. Computer Vision*, vol. 72, pp. 47–66, 2007.

[10] N. Bellotto, E. Sommerlade, B. Benfold, C. Bibby, I. Reid, D. Roth, C. Fernández, L. V. Gool, and J. Gonzàlez, "A distributed camera system for multi-resolution surveillance," in *Proc. of the ACM/IEEE Intl. Conf. on Distributed Smart Cameras (ICDSC)*, 2009.

[11] N. Krahnstoever, T. Yu, S.-N. Lim, K. Patwardhan, and P. Tu, "Collaborative real-time control of active cameras in large scale surveillance systems," in *Proc. Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications (M2SFA2)*, October 2008.

[12] T. Yu, S.-N. Lim, K. Patwardhan, and N. Krahnstoever, "Monitoring, recognizing and discovering social networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, (Miami, FL), June 2009.

[13] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House Publishers, 1999.

[14] N. Krahnstoever, P. Tu, T. Sebastian, A. Perera, and R. Collins, "Multi-view detection and tracking of travelers and luggage in mass transit environments," in *In Proc. Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 2006.