Automated Night/Day Standoff Detection, Tracking, and Identification of Personnel for Installation Protection
Brian E. Lemoff, Robert B. Martin, Mikhail Sluch, Kristopher M. Kafka, William McCormick and Robert Ice
WVHTC Foundation, 1000 Technology Drive, Suite 1000, Fairmont, WV, USA 26554

ABSTRACT

The capability to positively and covertly identify people at a safe distance, 24-hours per day, could provide a valuable advantage in protecting installations, both domestically and in an asymmetric warfare environment. This capability would enable installation security officers to identify known bad actors from a safe distance, even if they are approaching under cover of darkness. We will describe an active-SWIR imaging system being developed to automatically detect, track, and identify people at long range using computer face recognition. The system illuminates the target with an eye-safe and invisible SWIR laser beam, to provide consistent high-resolution imagery night and day. SWIR facial imagery produced by the system is matched against a watch-list of mug shots using computer face recognition algorithms. The current system relies on an operator to point the camera and to review and interpret the face recognition results. Automation software is being developed that will allow the system to be cued to a location by an external system, automatically detect a person, track the person as they move, zoom in on the face, select good facial images, and process the face recognition results, producing alarms and sharing data with other systems when people are detected and identified. Progress on the automation of this system will be presented along with experimental night-time face recognition results at distance.

Keywords: Face Recognition, SWIR, Night Vision, Surveillance, Biometrics, Active Imaging

1. INTRODUCTION

When the identification of individuals is discussed in the context of installation protection, it usually relates to confirming the identity of people wishing to gain access to the installation and is handled at entrances or checkpoints. Whether the checkpoint is manned or unmanned, and whether the identification uses cards or biometrics, it is assumed that the individual will be aware of and cooperate with the identification process, even if his intent is to falsify his identity, as lack of cooperation would not result in access. In contrast, the capability to covertly detect and identify individuals within a large radius of the installation, night or day, could provide another layer of security. Individuals known to pose a threat to the installation could be detected in the distance as they approach the checkpoint, allowing security time to prepare or to intercept. Alternatively, known bad actors could be detected and identified as they conduct clandestine surveillance of the installation. If installation security is on the lookout for one or more specific individuals thought to be in the vicinity, they could scan the area and possibly locate them. The technology could also be used to rule out possible threats; for example identifying a person approaching the facility after hours as a known good actor.

Under daylight or otherwise lit conditions, it is possible today for a guard using high-power optics to manually identify people at a distance if they are familiar to him or if he can refer to a short watch list of mug shots; however, automated identification at long range is not yet available. At night, or otherwise dark conditions, there is currently no technology that produces imagery that allows for long-range identification, either manual or automated. To address this capability gap, the West Virginia High Technology Consortium Foundation (WVHTCF), under a research contract from the Office of Naval Research (ONR), is developing the Tactical Imager for Night/Day Extended Range Surveillance (TINDERS), an active short-wave infrared (SWIR) imaging system that illuminates targets with an invisible and eye-safe SWIR laser beam. Goals for the project, from bright sunlight to total darkness, include: human detection and tracking at ranges up to 3 km; generating recognizable facial imagery at ranges up to 800 m; and identification through computer face recognition at ranges up to 400 m. When complete, the goal is for TINDERS to be a portable, easy to set-up system that can automatically detect, track, zoom in on, and image a moving person, and identify them through computer face recognition.
1.1 Background

Covert, long-range, night/day human identification requires the integration of several capabilities. First, a person must be detected and their location determined. Then, as people rarely stand still long enough to be identified, the person must be tracked as they move. Close up facial imagery must then be captured with sufficient resolution and quality to make a positive identification. This typically requires a minimum of 20 pixels between the eyes, or a resolution of roughly 3 mm per pixel, although resolution better than 1-mm per pixel is often stated as a requirement for high-performance computer face recognition. For the capability to work night and day, the imaging technology must be able to work under conditions ranging from bright sunlight to total darkness.

Whether the goal is computer face recognition or simply recognition by a human operator, visible-spectrum imagery will always produce the best result if conditions allow for a quality image to be obtained. Unfortunately, under nighttime or otherwise dark conditions, there is insufficient ambient illumination of the target to produce a visible image. A spotlight could be used, but this would not be covert, and the intensity required to produce a high-quality close-up facial image at long range would be damaging to the eye. There are a number of excellent technologies currently available for long range nighttime detection of personnel; however, none of them is suitable for long-range facial recognition.

Thermal or long-wave infrared (LWIR) imagery is an excellent tool that is widely used for nighttime detection of personnel; however, it does not produce recognizable facial imagery. LWIR imagery reveals the thermal profile of a person’s face, rather than skin surface texture and features, making it difficult to correlate with visible-spectrum facial imagery. In addition, as shown in Figure 1, the LWIR appearance of a person’s face will change depending upon the thermal conditions and the person’s metabolic state. Finally, thermal imagers are much better suited to wide-angle imagery, as a thermal imager capable of long-range close-up facial imagery (e.g. 2-mm per pixel at 300-m range) would require an extremely large lens aperture.

![Figure 1. Thermal infrared facial images of a single subject under different thermal conditions.](image)

This variability, along with the poor correlation between thermal facial images and visible-spectrum facial imagery, make it unlikely that thermal infrared imagery can be used to identify individuals based on a watch list of visible-spectrum facial images (mug shots).

Passive SWIR imagery is another excellent technology for day/night wide-area surveillance. Even with no moon and slightly overcast conditions, there is enough ambient night-glow to produce clear wide-angle imagery, such as that required for human or vehicle detection. Unfortunately, nighttime signal levels, even under a full moon are too low for close-up facial imagery at long range, such as that needed to recognize a person’s face at 100-m range. Because the amount of light hitting each pixel in the sensor is proportional to the total amount of light coming from the imaged area divided by the number of pixels in the image, the signal level in a passive image increases as the square of the field of view. For example, narrowing the field of view of an image by a factor of 10 reduces the signal level by a factor of 100.

Active near-Infrared (NIR) surveillance systems are available commercially from companies such as Vumii. These systems combine a long-range camera (conventional silicon CCD) with a NIR illuminator (typically around 800-nm wavelength) to produce high-quality, long-range imagery night and day. By illuminating the camera field of view with light that is invisible to the human eye, but close-enough to the visible spectrum to produce familiar-looking imagery, high-quality long-range imagery is possible. While such imagery should be sufficient for human identification, the illumination power required to produce quality facial images at ranges beyond 100-m creates a possible eye-safety
hazard to the target, whose face is being deliberately illuminated, and a severe eye-safety hazard (immediate and permanent damage to the retina) in close proximity to the illuminator. In addition, while virtually invisible to the human eye (at high intensity, 800-nm appears as a dull red glow), the NIR illumination is clearly visible with any night-vision goggle and most silicon-based cameras.

1.2 Active-SWIR imaging

Active-SWIR imagery at wavelengths > 1400 nm, particularly in the wavelength band most commonly used by the telecommunications industry for fiber-optic communication, overcomes the two primary limitations of Active-NIR imagery in that the eye-safe power levels are much higher, and it is completely invisible to night-vision goggles and humans. Eye safety can be evaluated using the ANSI Z136 and IEC 60825 standards. According to the standards, the maximum safe exposure level at SWIR wavelengths >1400 nm is 65 times higher than at 800 nm. Assuming an illuminator has an exit aperture 5-inches in diameter, an 800-nm laser beam would need to be below 0.2 W to be safe to the human eye, while a SWIR laser beam at standard telecom wavelengths could be as powerful as 13 W and still be safe to the naked eye. Such a powerful laser beam would even meet Class 1 eye safety requirements (safe even in the presence of magnifying optics up to 7X) at a target, where the illumination beam has diverged to at least 1 meter in diameter. All silicon-based cameras, such as CCD and CMOS imagers, are also completely insensitive to SWIR light at > 1400 nm.

Unlike thermal infrared, active-SWIR facial imagery produces repeatable imagery, showing only external surface features that scatter the illumination light. Figure 2 shows the same individual illuminated with visible white light and illuminated with an eye-safe SWIR laser. While the skin and hair pigmentation are quite different in the two images, the geometry of the facial features are the same. Thus, it should be possible to match a SWIR facial image against a database of visible-spectrum facial images using an appropriate computer face recognition algorithm. In addition, once a human operator becomes accustomed to the darker skin and lighter hair appearing in SWIR facial images, manual recognition of individuals based on SWIR facial images is possible.

Figure 2. (left) Facial image of an individual illuminated with visible-spectrum white light. (right) Facial image of the same individual illuminated with an eye-safe SWIR laser operating in a wavelength band commonly used for long-distance telecommunications. Note that hair appears white and skin appears dark in the SWIR image, but the same skin features, with the same shapes, are present in both images.

Early in the TINDERS project, visible and SWIR facial imagery similar to that shown in Figure 2 was collected from 56 subjects. An experiment was performed using a commercial face recognition software package, ABIS® System Face Examiner, from Identix (now MorphoTrust USA), in which a single SWIR facial image from each subject was matched against a database containing 1156 visible-spectrum facial images, including 1 visible image from each of the 56 subjects and 1100 visible images from the FERET facial database. The commercial software, which had been designed only to match visible images to other visible images, achieved a correct match for 40 out of 56 subjects, for a Rank 1 success rate of 71%. This indicates the overall feasibility of using active-SWIR imaging for long-range identification.

1.3 TINDERS Hardware

A conceptual illustration of the TINDERS hardware is shown in Figure 3, along with a photo of the current TINDERS research prototype. The TINDERS system consists of three physical units, an optical head, that sits on a pan-tilt (PT) stage, an electronics box that provides power, light (through and optical fiber), and communications to the optical head,
and a computer that runs the user interface, low-level camera control functions, system automation, and face recognition software.

![Diagram](a)

Figure 3. (left) Conceptual illustration of the TINDERS hardware. (right) Current TINDERS research prototype hardware.

The optical head includes both the SWIR illuminator optics and the imager. The imager and illuminator pan, tilt, and zoom together so that the illuminator beam is always just filling the imager field of view. This serves to maximize the image signal level and avoid wasted light. The imager and illuminator each have a 53X total zoom ratio (the imager has 10X optical, 5.3X digital, while the illuminator has 53X optical zoom). The illuminator light source, located in the electronics box, delivers a maximum power of 5W to the optical head through an optical fiber in the umbilical. This light source leverages commercial technology developed for the long-distance telecommunications industry.

![Image](b)

Figure 4. (left) TINDERS prototype at a December 2011 field experiment in which a long (> 50-ft) umbilical was used to connect the optical head to the electronics box located in a powered trailer some distance away. (center) TINDERS deployed atop a 35-ft mast during a December 2011 demonstration. For this implementation, the electronics box was located at the base of the mast, with the umbilical extending the length of the mast. (right) TINDERS can be controlled from any networked device with Windows Remote Desktop capability, such as the Windows tablet shown in the figure.

For installation protection, the TINDERS optical head can be deployed on a tripod or mast, as well as being permanently mounted to a building or other structure. The umbilical that connects the electronics box to the optical head includes power, data communications, and optical cables. When the electronics box is located near the optical head, as in Figure 3, a short umbilical (~ 15 ft) is typically used; however, longer umbilicals (> 50 ft) have been used when the TINDERS optics was deployed atop a mast or on a tripod located in a field far from any power source. In the current hardware configuration, the TINDERS computer is connected to the electronics box through an Ethernet cable. Cables as long as 300-ft have been used. An operator can control and monitor the system from the TINDERS computer console or via...
remote login from any networked device, such as a laptop or tablet that supports Windows Remote Desktop. Figure 4 shows examples of TINDERS deployment on a tripod and mast and remote operation from a tablet.

2. AUTOMATION

The practical use of TINDERS to provide long-range identification capability for installation protection will depend upon its ability to automate the process of detecting and following a human target while collecting optimized facial images, choosing the best images, and submitting them for facial recognition. Most of these automation functions have yet to be integrated into the TINDERS prototype; however, there is already good progress to report in this area.

2.1 Detection of personnel

Even at its widest zoom angle, TINDERS is still a narrow-angle sensor, so the first step in detecting people will be to point TINDERS in the general direction of a potential target. Once it is set up and calibrated, TINDERS has the capability to point and focus on any specified geographical coordinates within its range. A wide angle sensor, such as a ground moving target indicator (GMTI) radar system or wide-angle camera (visible, passive SWIR or thermal IR), can be used to provide initial detection of personnel within range of the installation, detected target coordinates can then be fed to TINDERS to provide initial cuing. Alternatively, an operator can scan TINDERS across areas of interest or let the system dwell at specific locations of concern, such as a roads or walkways leading up to the facility.

Common approaches to detecting personnel in surveillance video include change detection, motion detection, and cascade pattern recognition. Change detection works well for fixed surveillance cameras, where a static background image can be captured and compared to the live image. For a pan-tilt-zoom system such as TINDERS, this is not a viable approach. Motion detection is a good way to rapidly detect moving objects in video, but it cannot distinguish between a person and any other moving object. Cascade pattern recognition searches images for patterns that match a set of training images. This approach can be as specific as the training dataset, but the approach can also be time consuming, depending upon the complexity of the pattern and the range of search parameters. Motion detection and cascade approaches can be combined by using motion detection to narrow down the range of possible target locations in an image prior to starting a cascade search.

WVHTCF has developed a cascade to detect personnel in TINDERS imagery. Because feet and legs are often obscured by terrain and vegetation, the algorithm detects people from the waist up. This algorithm has proven successful at detecting people day and night at distances as high as 3 km. Figure 5 shows two examples of nighttime TINDERS person detection at distances of 91 m and 3 km.

Figure 5. Examples of successful TINDERS person detection, under dark nighttime conditions using an upper-body cascade pattern detector: (left) 100-m range; (right) 3-km range. Note that the identical algorithm was successful at both distances despite the obvious difference in image resolution.

2.2 Tracking and head/face detection

When used in an installation protection application, the concept is that TINDERS will initially detect personnel while in its widest-angle zoom setting. Once detected, an operator or an automated rule can select the target for tracking. At this point, the detection box from the upper-body detection will be sent to a tracking algorithm that could control the pan-tilt stage to keep the selected person centered in the imager field of view. Work to integrate a commercial tracking
algorithm into the TINDERS system is currently underway. If a person is beyond the 400-m upper limit for face recognition, tracking will continue at the widest zoom setting. Once a person comes within face recognition range (< 400 m), the concept is that the system will try to zoom in on the head while continuing to track their movement. At this point, the tracking will try to center the person’s head in the field of view. WVHTCF has developed another cascade algorithm that successfully detects faces in TINDERS imagery. Additional algorithms will also need to be developed to detect heads at different angles, for example side profiles and the back of the head, in order to continue to track a person’s head at the highest zoom setting. Figure 6 shows examples of successful face detection at both wide-angle and narrow-angle zooms. Notice that facial features do not need to be clearly visible in order for this algorithm to work.

Figure 6. Examples of successful TINDERS face detection, under dark nighttime conditions using a cascade pattern detector: (left) 100-m range, wide angle; (right) 350-m range narrow angle. Notice that the face is detected in the 350-m image even though eyes, nose, and mouth are not clearly visible. This is critical in order for the algorithm to be useful in tracking a moving face that may be blurred due to motion or poor focus. Similar algorithms will also be used to detect heads at other angles.

2.3 Face selection and queuing

In order for the identification process to be automated, TINDERS must automatically select video frames containing high-quality facial images suitable for use by face recognition software. For example, the 350-m image in Figure 6 would NOT be suitable for identification. Once the target distance and imager zoom level are within the limits of the face recognition capability, a face selection algorithm will be run that evaluates frames for facial image quality. WVHTCF has developed an algorithm that performs eye detection and nose detection, and uses the detected positions of the eyes and nose, together with the focus quality of the face to determine if the image is suitable for face recognition. Once integrated into the TINDERS system, the concept is that selected images will be ranked by quality and queued for submission to face recognition software. When the face recognition software is ready for a new submission, the best facial image in the queue will be submitted and processed for matching against the database of visible-spectrum facial images. As long as a single individual is being tracked, face shots will continue to be submitted to the face recognition software and the matching results accumulated, continually increasing the confidence level of any potential match. Figure 7 shows automated face, eyes, and nose detection for facial images at two distances.

Figure 7. TINDERS images selected for face recognition. Selection algorithm automatically detects the face, eye positions, and nose positions (all marked in the figure). Based on the relative positions of these features, the image is selected as suitable for face
recognition. A focus function will also be used to assess the relative quality of different images. Images above were acquired under dark nighttime conditions at distances of (left) 100 m and (right) 200 m.

3. FACE RECOGNITION

The TINDERS face recognition software leverages the commercial ABIS® System FaceExaminer software from MorphoTrust USA. As part of the TINDERS research program, researchers at MorphoTrust USA developed a pre-processing filter to apply to the SWIR facial images to improve the matching performance of the SWIR images to visible-spectrum images contained in the database. The automated face selection and queuing process described in the preceding section has yet to be implemented in the TINDERS prototype. The current TINDERS software allows the operator to submit multiple video frames to the face recognition software by clicking a button on the GUI. Face recognition results are then displayed in the TINDERS GUI. MorphoTrust USA is also continuing to work with WVHTCF on improving the performance of the SWIR-to-visible matching algorithms used in TINDERS.

3.1 Facial data collection

To develop and evaluate TINDERS face recognition capability, WVHTCF conducted two rounds of facial imagery data collection. In the first dataset, indoor facial imagery was collected in total darkness at distances of 50 m and 106 m for a total of 56 subjects. In the second dataset, only recently completed, outdoor imagery was collected under dark nighttime conditions at distances of 100 m, 200 m, and 350 m for 104 subjects. Sample images are shown in Figure 8.

Figure 8. (upper row) Example imagery from the first TINDERS dataset, collected indoors in total darkness at distance of (left) 50 m and (center) 106 m, along with (right) visible-spectrum facial image of the same person. (lower row) Example imagery from the second TINDERS dataset, collected outdoors under dark nighttime conditions at distance of (left) 100 m, (left center) 200 m, and (right center) 350 m along with (right) visible-spectrum facial image of the same person.

3.2 Identification Results

The first dataset was shared with two research groups at West Virginia University (WVU), who were working independently on SWIR-to-visible face recognition algorithms, as well as with the research team at Morpho Trust USA who developed the TINDERS face recognition software. Kalka, et. al. applied a pre-processing algorithm to the SWIR images before matching them to a visible-spectrum database using FaceIt G8 software from MorphoTrust USA. They achieved a Rank 1 success rate of 90% for the 50-m TINDERS images and 80% for the 106-m TINDERS images. Zuo, et. al. fused the results of the FaceIt G8 software with a face recognition algorithm developed by their group. With a 0.1% False Acceptance Rate, they achieved a Correct Acceptance Rate of 85% for the 50-m TINDERS images and 74% for the 106-m TINDERS images.

The same dataset was also used by researchers at MorphoTrust USA in their development of the face recognition software that is integrated into the TINDERS system. To evaluate their pre-processing filter, they processed 9 SWIR images for each subject at each distance, including 3 frontal neutral images, 2 frontal talking images, and 4 images with
a 10° pose angle. Each image was pre-processed and matched against a database containing visible-spectrum images of all 56 subjects. For each subject, the results of the 9 searches were fused by keeping the result with the highest matching score. With a 1% False Acceptance Rate, the pre-processed results achieved a Correct Acceptance Rate of roughly 70% at both 50 m and 106 m.

A proper statistical analysis of face recognition performance has not yet been completed for the second dataset; however, successful face recognition at both 200 m and 350 m range has been achieved with a performance level far better than random chance. Figure 9 shows screen shots of successful TINDERS face recognition at 200 m and 350 m, where the correct person is chosen out of a database containing visible-spectrum images of more than 1600 individuals. In these examples, TINDERS was playing back recorded video from the second dataset as if it were live. The operator clicked a button on the TINDERS GUI, which sent 11 video frames to the face recognition software for matching. In the case of the 200-m result, 8 of the 11 frames were detected as good faces and eye positions automatically marked. For the 350-m result, 9 good faces were detected and eyes automatically marked. The detected good facial images were then searched against the visible-spectrum facial database containing over 1600 individuals, and the results fused to produce an aggregate matching score for each of the 1600 candidates. The top 20 candidates are then displayed in rank order along the bottom of the screen. In both examples, the correct person was chosen as the top match. Aside from clicking the button on the TINDERS GUI to initiate the process, the operator did not need to interact with the system to produce the results. The entire process completed in < 20 seconds.

![Figure 9. Example screen shots showing TINDERS face recognition from video under dark nighttime conditions at distances of (left) 200 m and (right) 350 m. In both cases, the correct individual was chosen from a database containing visible-spectrum images of more than 1600 people.](image)

### 3.3 Approach to achieving high confidence identification

While the results shown in Figure 9 demonstrate the potential of the TINDERS system to positively identify individuals, day or night, at long ranges, practical implementation in an installation protection environment will require an automated approach to establishing identity at a high-confidence level. In more traditional implementations of computer face recognition, a single, high-resolution visible-spectrum facial image is matched with a very high confidence level against a large database of high-resolution visible-spectrum facial images. In the case of the TINDERS system, where the detected facial images are SWIR, rather than visible, and where long-distance resolution is often much lower than optimal, it is unlikely that a single detected SWIR facial image will ever produce a high-confidence match to a large visible-spectrum database.

For TINDERS, the key to achieving high-confidence identification will be the fusion of face recognition results from many SWIR video frames acquired as a single person is tracked. Each time a search is performed on a new facial image from the same tracked person, the differences in angle, expression, field of view, atmospheric conditions, and noise will result in different scores and rankings for each candidate in the database. If the person being tracked is actually in the database, then even if the results of any single search are inconclusive, accumulation of the results of many searches should allow conclusive identification. One fusion method often used, known as “maximum score fusion”, keeps only the highest score for each database candidate over all the searches. Other techniques use only the rankings of the database candidates, assigning points based on rank to a candidate each time they score, for example, among the top 5
candidates. Such a technique assumes that a genuine match will consistently rank highly, even if not always number 1 and that impostor candidates will tend to have a more random ranking. As more searches are performed, a genuine match should distance itself from the pack, allowing for a high confidence identification to be made. Investigating the optimal fusion method as well as methods for determining confidence levels will be an important research focus of the TINDERS project moving forward.

4. ACKNOWLEDGMENTS

This research was performed under contract N00014-09-C-0064 from the Office of Naval Research, with funding from the Deployable Force Protection Science and Technology Program, and with additional support from a subcontract from West Virginia University Research Corporation through Award No. 2010-DD-BX-0161 awarded by the National Institutes of Justice. The authors would like to acknowledge important technical contributions from Jason Stanley and Andrew Dolby, the MorphoTrust USA team, and the cooperation of the WVU Center for Identification Technology Research in the collection of the most recent facial dataset.

REFERENCES