Long Range Facial Image Acquisition and Quality

Terrance E. Boult[1,2] and Walter Scheirer[1,2]

Abstract This chapter introduces issues in long range facial image acquisition and measures for image quality and their usage. Section 1, on image acquisition for face recognition discusses issues in lighting, sensor, lens, blur issues, which impact short-range biometrics, but are more pronounced in long-range biometrics. Section 2 introduces the design of controlled experiments for long range face, and why they are needed. Section 3 introduces some of the weather and atmospheric effects that occur for long-range imaging, with numerous of examples. Section 4 addresses measurements of "system quality", including image-quality measures and their use in prediction of face recognition algorithm. That section introduces the concept of failure prediction and techniques for analyzing different "quality" measures. The section ends with a discussion of post-recognition "failure prediction" and its potential role as a feedback mechanism in acquisition. Each section includes a collection of open-ended questions to challenge the reader to think about the concepts more deeply. For some of the questions we answer them after they are introduced; others are left as an exercise for the reader.

1 Image Acquisition

Before any recognition can even be attempted, they system must acquire an image of the subject with sufficient quality and resolution to detect and recognize the face. The issues examined in this section are the sensor-issues in lighting, image/sensor resolution issues, the field-of view, the depth of field, and effects of motion blur.

Vision and Security Technology Lab, University of Colorado at Colorado Springs, Colorado 80918, firstinitial.lastname@vast.uccs.edu · Securics Inc., Suite 200 1867 Austin Bluffs Parkway, Colorado Springs CO 80918 firstinitial.lastname@securics.com

Pre-print of chapter to appear in the book *Handbook of Remote Biometrics: for Surveillance and Security*. The original publication is available at http://www.springerlink.com.

1.1 In the beginning: Let There Be Light

To recognize a face one needs an image with visible features, which requires that we collect an image with sufficient light levels and quality. Understanding the impact of illumination variation, or normalizing to reduce it, is by far the most well studied of the issues associated with lighting and face-recognition [17, 1, 8, 10, 29, 5]. While this type of work is very important, it is more focused on algorithms and not acquisition, and hence not covered in this chapter. This section will focus on illumination aspects associated with acquisition, in particular, collecting and measuring light.

When working at close range in daylight conditions, the issue of sufficient lighting is not a critical concern. However, as one starts looking at long-range face-based recognition, especially for 24 hour "surveillance", assuring sufficient light level is critical. Addressing this raises 2 unique issues: how to measure those light levels, and what sensors to use to collect in lower light and/or long range settings.

Long-range face needs very long focal lengths, often in the range 800-3200mm. Combining distance with the inherent limits on optics results in high F-Numbers levels. For example the Questar Ranger 3.5 which is a portable telescope used in long range surveillance, provides 1275-3500mm focal lengths, but it comes at a cost of light, with the 89mm(3.5inch) providing F13.2 at 1175mm and F35 at 3500mm.[1] Recalling that each F-stop is a 50% loss of light, this telescope will measure intensity that is orders of magnitude smaller than that measured with a more traditional F4 lens used for close/moderate range face recognition. This need for light is even more exasperated by the need for faster shutter speeds to avoid motion blur issues that will be described later in this chapter. Understanding the available lighting for long-range settings thus is far more important than for standard face recognition. A question then is how to report light levels for long range experiments, especially for low-light conditions. This is important not just for scientific experimentation, but for practical concerns if one wants to determine if conditions are sufficient for a particular system to operate.

The most common measure for low-light imaging is in terms of *lux*. Lux is a measure of illuminance (the accumulated light energy reaching a surface), and measures how much light is in the scene. Given the lux reaching a surface, and the bi-directional reflectance function of the material/subject, one can estimate the luminous flux (the light leaving the surface in a particular direction). Luminous flux is measured in *lumens*. One can also compute the Luminous emittance which is the luminous flux per unit area emitted by a source. Luminous emittance, like luminance, is measured in lux. Given the Luminous flux one can use field-of-view of the lens and its F-stop to estimate the amount of light reaching the sensor from the targets. When the models are done right, this can be effective in predicting the light reaching the camera and hence the response of the sensor.

Unfortunately, to use this approach for long-range low-light imaging there are a number of difficulties. First, the reflectance function of the face varies considerably across the population. More significantly, the reflection is directional and is

[1] http://www.company7.com/questar/surveillance/querange.html

impacted significantly by self-shadowing, so measuring the scene irradiance with a traditional lux meter is not very effective without accounting for reflectance, shading and shadowing which requires a detailed calculation after measurement, making it difficult to use without advanced computer models. Finally, an issue especially important for low light settings is that, even higher end hand-held light sensors are only effective down to .01lux. These sensors use a light-to-voltage conversion that makes them good for bright scenes. But even though their accuracy is officially rated at plus or minus .01lux, in practice, it is quite tenuous below .1lux. In many of our field experiments, the available lux sensors report underflow or zero (it is too dark for them to operate). There are higher end NVIS lux meters, such as the ANV-410 and TSP-410, but these are significantly more expensive and still have the issues of not providing sufficiently directional measurements to measure light that will reach the sensor.

There is an alternative, which is to directly measure the light leaving the face in the direction of the sensor: *luminance*. The candela per square meter ($\frac{cd}{m^2}$) is the SI unit of luminance; nit is a non-SI name also used for this unit. A candela is a lumen per steraidan (solid-angle), so a $\frac{cd}{m^2}$ (nit) is equivalent to a $\frac{lumen}{m^2}$ sr), where as a lux is a $\frac{lumen}{m^2}$ there is no simple conversion between lux and nits without using knowledge of the view subtended by the source (face), which varies with distance. Luminance is valuable because it quantities describe the "brightness" of the source and does not vary with distance, whereas illuminance in lux (the "light" falling on a surface) must be manipulated to estimate how much light there is to measure. Putting it another way, illuminance is a good measure to use when asking how well people or cameras can function anywhere within a dimly lit environment, but luminance is the better measure to use for how well they can view a particular target (see [14]). The question then is how to effectively measure luminance for long-range face, especially if experimenting in low-light conditions.

To address these problems, and provide for a simple in-field measurement, we have adapted a different type of measurement sensor. Using a sensor originally designed for "sky quality" measurements or "sky darkness" measurements, we have a device that can operate at much lower light levels and can measure a narrow enough FOV to capture just the data of the face. The sensor being used is the SQM-L[2], based on the TAOS TSL237S sensor, which is a light to frequency converter. The SQM-L has an added lens so that the Full Width Half Maximum (FWHM) of the sensor is $\sim 20^\circ$. The sensitivity to a point source $\sim 19^\circ$ off-axis is a factor of 10 lower than on-axis and fall off faster beyond that. We will be experimenting with adding a component for further restriction of the field of view. The SQM-L sensor reports Magnitudes/arcsecond, which is an astronomical unit of measurement, but which is easily converted into $\frac{cd}{m^2}$. If we let s be the SQM-L value reported, then $\text{Luminance}\frac{cd}{m^2} = 108000 \times 10^{-0.4*s}$.

We use the SQM-L for long-range face experiments by having the subject look toward the camera, so it has appropriate lighting falling on the face, and them aiming the sensor on the center of their face while holding the sensor about 18 inches

[2] http://unihedron.com/projects/darksky/

away. At this range a face subtends approximately 18°, i.e. the sensor is measuring the light leaving the face and little else (but some care has to be used if there are distant lights behind the subject, and not shadow the face from any light sources with the hand/sensor). We call this measurement the "face luminance" and consider it the most useful overall lighting measurement for estimating performance of a long-range face-system in low-light conditions. This is really a measurement of luminance but it can be converted to luminous flux using the area of a face and the solid angle subtended by a face from the target range, which is simple scaling.

Fig. 1 Example of low-light long-range EMCCD imagery. The measured scene illuminance for the left image was .01lux, and illuminance was not measurable for the other two images. The measured Face Luminance left to right, was 0.089 nits, .0768 nits and 0.015 nits respectively.

For example in figure 1, we have long range images in low-light conditions. The images were obtained at approximately 100m with an F5.6 Sigma 300mm-800mm. Capture occurs under star-light conditions 60, 90 and 120 minutes after sunset, with a street light 100m off on the subject's left. We prefer the face luminance measurement approach because works in the low-light setting where we want to operate and it already accounts for the complex lighting/face shape interactions and is easily converted to a direct measure of the luminous flux heading in the direction of the camera. It is also very easy to "measure" in the field: hold the sensor, face the direction of the camera and push the button on the sensor and hold (maybe up to 60 seconds if it's really dark), then read the measurements from the unit's LEDs. These measurements are more repeatable and reliable than using simple lux-based estimations of overall illumination and then trying to convert it to lux at the sensor.

The second major issue impacting light levels for acquisition is the inherent imaging system sensitivity. This is significantly impacted by the sensor. Again, since long-range face is generally for surveillance, there is a general need to consider low-light conditions.

One approach often suggested for dealing with low-light settings is the use of Infra-Red sensors for face recognition. For long wave IR (8-14 microns), the human

body is an light source and such images could be collected in total darkness. While there has been some significant progress in the area of LWIR face (see [25, 23, 4, 11]), we believe LWIR is too limited for long range face for several reasons. First, the need for long focal length lenses and high-resolution sensors for long-range face - the combination of which are simply not available for LWIR. The resolution issues will be discussed in the next section. The second limitation of LWIR is that since long-range face is usually for non-cooperative subjects; LWIR requires specialized enrollment whereas visible recognition can use standard intelligence photos.

For comparison, consider Figure 2, which shows example images (close range) of 3 different types of sensors: a standard visible image, an intensified image and a thermal image. This dataset can be obtained, for US researchers, from the author. An interesting open research question is the development of a LWIR recognition system that can operate with visible image galleries. That is, with some intitial work in the area converting thermal into visible images addressed in [6].

Fig. 2 Left to right shows the same subject in a normally lit visible light camera, a low-light intensified imagery and in LWIR thermal imagery. The Intensified imagery was obtained using an American Eagle 603U which is a GenIII+ intensifier (specs are the same as PVS-14 commonly used by the US Military). The intensified image was captured by a IQ-EYE smart camera with 1280x1024 resolution. The thermal (LWIR) sensor is an NYTEK WEB-50 Micro-Bolometer, 8-14micron sensor, with images captured from the analog 640x480 video output. The visible images were captured from an IQ-EYE 1megapixel sensor. Images have faces with 80 pixels between eyes, which is the lower end of is expected for good recognition.

The alternative for low-light operation is to use some type of intensified imagery. There are a few alternatives within this group ranging from the very common tube-based intensifier optically coupled to a CCD sensor to an intensified CCD, to the current generation of Electron Multiplying CCD. In our early work in low-light we used tube-based intensifiers coupled with a CDD. One disadvantage of this is the blurring induced by the micro-charge plates of the intensifier and the visible "channel" artifacts, which have also been noted by other researchers (see Figure 1 in [22]). Our more recent work in long-range face/surveillance [24] has moved to using EMCCD technology, based on a Salvador Imaging camera using the TC 285 Chip. This provides 1004x1002 pixel images with 8μ pixels and an overall quantum efficiency of 65%. This sensor can operate from full sunlight down to star-light conditions. While the full details of our long-range low-light experiments are

beyond the scope of this chapter, Figure 1 provides some examples of cropped face data collected with an 800mm Sigma F5.6 lens at more than 100m under very low light conditions. Except in the first of these photos, the naked eye camera operators could not even see the subject was there let alone recognize them. These images show there is potential for long-range low-light face recognition using EMCCD technology.

1.2 Resolution: what does it mean and how much do we need?

In acquisition, presuming we have sufficient number of photons, the next most important issue is the resolution of the target. While it is quite common to hear people talk about resolution in terms of number of pixels, it is more accurate to talk about the effective resolution. One can formally define this using the Modulation Transfer Function (MTF) of the imaging system, which can account for both blur and contrast loss. Under some simplifying assumptions one can decompose the MTF into the product of the optical (lens) MTF, the sensor geometry MTF, and the diffusion MTF [9].

The ability of a lens to resolve detail is usually determined by the quality of the lens, though some very high end lenses and telescopes are diffraction limited. The effective aperture of the lens diffracts the light rays so a single point in space forms a diffraction pattern in the image, which is known as the Airy disk. If the system is not diffraction limited, then other lens artifacts produce patch such that different rays leaving a single scene point to not arrive at a single point in the image, giving rise to what is called the "circle of confusion", even though it can be a far more complex shape. Ideally, the circle of confusion will be smaller than a sensor pixel.

Most MTF tables provided by lens manufacturers (see [3]), will show the MTF as a function of image position or distance from the center of the image. MTF values above .6 are considered satisfactory, while some lenses such as the Canon EF 400mm f2.8 IS USM, which we use for some of our long range experiments, have a circle of confusion of .035mm and MTF values above .9 over the whole field of view. Even when extending with the Canon 2xII extender (making it it an 800mm F5.6 lens), the MTF is above .7 everywhere. In general, zoom lenses will have lower MTF because of the more complex lens designs limit the optimization. (Note: you can buy adaptes for C-mount to Canon lenses, with complete rs232 based control of lens parameters such as focal distance, aperature, and stabalization parameters. These adapters are open air but because the increase the separation to adapt the 35mm format to C-mount they may degrate the MTF).

It is important that when working with long-range biometric the lenses be match, or over qualified for the sensor choice. Modern high-quality lenses are multiple element multi-coated designs optimized by the manufacturers for particular sensor choices and with particular wavelengths in mind. If you can see vignetting, spatially varying blur (when focused on a flat target) or color "fringe" artifacts, find a better

lens. It is also important to note that few lenses are optimized outside of the visible range, so be particularly careful in choices if working in the NIR range.

In the remainder presume that the optics are properly adapted to the sensor such that the overall MTF is not significantly limited by the optics, atmospheric or by motion blur because if those are the limiting factors, it makes little sense to discuss sensor "resolution". In practice, the the important consideration is that the blur of the system is less than a pixel, otherwise the image can be effectively down-sampled by a factor of the blur, and not loose significant information. If you are doing long-range biometrics, the minimum is to measure your effective blur, or you can waste a lot of time working on issues which are limited by blur. In short a large sensor/image size with a blurred image is not providing the resolution you might think.

Assuming good optics, resolution for long range face becomes a question of ensuring enough pixels on face to support recognition, and sufficiently above the minimum needed for recognition to deal with the loss of resolution do to atmospheric turbulence. Formal models for atmospheric loss have been derived in the literature. See [27]. Diving into those models is beyond the chapter, but using such models can estimate an atmospheric blur level for long-range face, and expand the resolution requirements by an equivalent amount. We have routinely expanded the 60 pixel IPD, used for close range face recognition, to 80 pixel IPD for our 100m experiments.

Given the desired goal of 80 pixels between the eyes, and an average physical size of 4 of between 60mm and 72mm, combined with pixel size (for example, 8 micron for visible spectrum sensors, 15 micron for LWIR) and size of the sensor (in number of pixels), one can then estimate the focal length needed to produce an image with the necessary spatial resolution on the subject. Deriving the formula is left as an exercise for the reader. In doing so, don't forget to account for any "adaptors", e.g. converting a 35mm camera lens into a C-mount is a change in format and back-focal distance that impacts the effective focal length.

In reviewing Table 1, you should note that lenses for visible sensors up to 1000mm are readily available and up to 3500mm available as special order via fields "telescopes". Most intensified CCDs are only 640x480 and LWIR sensors are only 320x240 (though there are exceptions for both, there are no 1280x1024 LWIR sensors). Long wave IR lenses up to 300mm are available and up to 1000mm is a special order (and massive).

	1280x1024	640x480	320x240
Range (m)			
50	333	625	1250
100	667	1250	2500
150	1000	1875	3750
200	1333	2500	5000
250	1667	3125	6250
300	2000	3750	7500

Table 1 Focal lengths needed to achieve 80 pixel average inter-pupil distance for different sensor sizes.

1.3 The working volume: Depth of Field and Field of view

The working volume, the region where the subject is in focus and within the filed of view, is clearly important for the acquisition system design. Depth of field defines the ranges around the focus distance where in subjects will be sharp focus. DOF increases with decreasing lens aperture and decreases with focal length, so for long-range face it is a much more significant issue. The depth of field for a lens is not symmetric, with different formula for the distance in front of the focus plane and behind.

Formally we one can derive these as:

$$\text{Front depth of field} = \frac{d \cdot F \cdot a^2}{f^2 + d \cdot F \cdot a} \tag{1}$$

$$\text{Rear depth of field} = \frac{d \cdot F \cdot a^2}{f^2 - d \cdot F \cdot a} \tag{2}$$

where f is the focal length , F is the F number, d is the diameter of the circle of confusion, and a is the subject distance from the first principal of the lens to subject. Note that if $f^2 < d \cdot F \cdot a$, the rear depth of field is considered infinite.

The important things to note here is that the depth of field decreases with the square of the focal length and for long focal length lenses can be quite short. Focus is further exasperated by the fact that for long-range face the optical axis usually does not intersect a ground plane where the target will be, because they will be walking well above the ground, and thus there is no way to easily pre-focus the image. Fast auto-focus or having subjects "walk through" the DOF region are the most common choices.

While DOF is directly impacted by distance, the FOV of a lens is not. Ignoring blurring, the field of view necessary to maintain sufficient resolution for long-range face is actually the same as that needed for near-field "non-cooperative" subjects. The increased resolution requirements to account for atmospheric blur does change it, but the change is effectively the same as requiring a larger inter-pupil distance in pixels.

The more significant difference is that there are many near-field face applications presuming cooperative subjects at effective choke-points to limit subject positioning. With non-cooperative subjects, a larger field of view is needed to allow for subject movement. This is especially acute in maritime biometrics where the subjects, and the sensor, may be moving with the waves.

The FOV is defined by the combination of the sensor resolution and the focal length. Presuming a focal length just sufficient for the minimum resolution, provides the maximum FOV. Again, one can easily derive, via basic geometry, the FOV, the associated pixel resolution on the sensor, and the effective physical size at the Focal point of the working volume. Example figures are shown in table 2, with the derivation of the formula left to the reader. In deriving the table, we presumed an I'D of 80 pixels and an overall head size of 160 pixels, and assured the head is within the frame. Figure 3 puts that data into perspective, and also shows how the FOV affects

"time of target" if one is using a stationary camera aimed at a choke point. It's not just that the larger sensor gives a larger FOV, the larger FOV translates into more frames on target as they cross the larger FOV.

Sensor resolution	2048x1520	1280x1024	640x480	320x240
Usable size in pixels	1888x1320	1120x824	480x280	160x40
Usable physical FOV (in ft)	5.8' x 4.3'	3.6' x 2.6'	1.5' x 0.9	' 0.5' x .1'
Allowed height variation	25in	15.6 in	5.4 in	0.6 in

Table 2 Usable Resolution, and size of FOV; the maximum size can reasonably be used for face recognition. Conservative estimates are half the sizes/times shown.

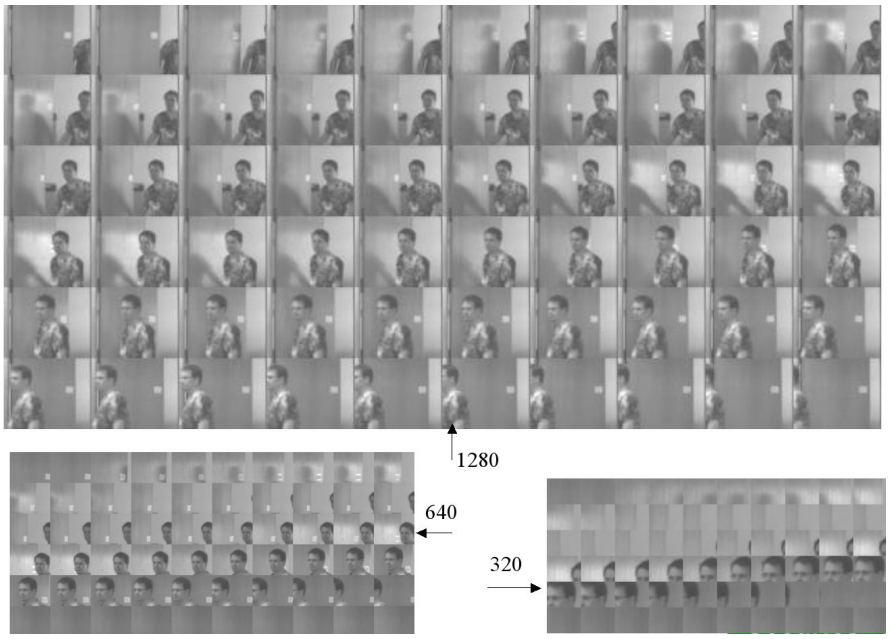

Fig. 3 Example showing image sequences of a subject exiting a doorway and how the sensor resolution and FOV affect the effective number of frames where there is sufficient face data for potential recognition. Note how 640x480 is just large enough for a head, and would not capture good data for someone significantly taller or shorter. The 320x240 sensor is relatively useless.

1.4 Motion Artifacts

The last significant "sensor" issue to be discussed on acquisition is motion artifacts including motion blur. In any face-based system with non-controlled subjects the issue of subject motion must be addressed. This section addresses some of those motion artifact issues.

At first one might again presume that these issues are the same for long-range face as they are for any non-cooperative subject. That is, in part, correct. However the long focal lengths necessary for long-range face can mean even a slight vibration in the sensor mounting can produce far more signficant results. The vibrations near field imaging on a basic camera mount on the wall might produce unnoticable interlace artifacts, but the same vibration magnified by a 800mm lens might tear the image appart and seem as here simple: *Do not even think of using an interlace camera for long-range face recognition.*

Beyond interlace artifacts, there are two major artifacts which impact long range face. While they also impact near-field face recognition, the fact that non-cooperative distance subjects can be moving faster, or that the long-range sensor could be on a moving platform induces greater potential for these issues to be objectionable.

The first of these issues is the well known motion blur. It can occur because of platform motion, including vibrations, as well as because of subject motion. We have found that for long-range face with walking subjects, most of the gait cycle will have noticeable vertical motion blur with a significant reduction at the top of the stride. Figure 4, shows an example with both a clear face image and various images showing motion blur. These images were with an 800mm f5.6 lens with a shutter speed of 1/30 of a second. These types of issues are further exasperated if attempts are made to use slower shutter speeds, or if the camera or subject are on a moving platform such as a ship.

Fig. 4 Example of motion blur. Subject is moving at a walking pace toward the EMCCD camera. Images are taken at approximately 100M from the camera at dusk. The top of the walking stride produces minimal motion blur. (Scene has approximately .04 lux, yielding face lumens of 0.115 nits.)

A less well known issue, which may not be obvious at first, arises with modern CMOS sensors that use a rolling shutter. Before we describe the issue, go study the images in figure 5 and see if you can discern what it is. Note the images have a very fast integration time speed, (1/10000 of a second), so if you thought it was motion blur, think again. And it's not a depth of field issue either.

Fig. 5 Examples of rolling shutter artifact

There are two primary reasons to use a rolling shutter. First it saves one transistor per cell compared to a true "snapshot" shutter, and secondly it allows the integration time to be almost equal to the frame rate without significant buffering or fast read out circuits. The concept is quite simple, think of it as having two pointers to sensor pixels, both "rolling down" the sensor. One pointer is for readout of data, and the other is the the reset or erase operation. The time difference between a row's erase and next read defines the effective integration time (shutter speed). Each pixel sees (and accumulates) the light for the same exposure time (from the moment the erase pointer passes it till it is read out), but that happens at different times.

All this sounds good - a wider range of integration times at a lower cost. So what is the problem? Looking back figure 5 again, we'll give you a hint. The wall to the subject's left is a normal doorway, it's a vertical edge and "straight". Your cell phone camera is almost certainly a rolling shutter CMOS sensor - you can try some experiments on your own see how significant the skew, warp or wobble can be.

The issue for rolling shutters is that even with a short integration time, the shutter is capturing data at different times for the top and bottom of the image. In the example the camera was subjected to horizontal motion fast enough that the top of the image saw the wall in a different position than the middle or the bottom. Now ask yourself what that would be doing to your face recognition algorithm, and you'll start to appreciate the issue and probably think twice about rolling shutter sensors, even if they are the cost effective solution for getting multi-mega pixel arrays.

This first section has reviewed the early, and more static aspects of image acquisition for long-range face. The next section examines approaches for controlled experiments for long-range evaluation, which is a necessary precursor before we can get into the impacts of weather and atmosphere.

2 Photo-heads: Controlled Experiments in Long-Range Face

Even after the images are acquired the atmosphere and weather impacts can be critical for long-range face acquisition. Studying them is a challenge as it is hard to collect enough data under varying conditions. To address this we designed a specialized experimental setup we call Photo-heads. The setup of the initial photo-head experiment is shown in Figure 6, and example images in Figure 7. This "photo-head" data is unique in that it is a well-known set of 2D images (FERET) that were displayed on a special LCD and then re-imaged from approximately 94ft and 182ft. (We are currently implementing another photo-head setup at much greater distances, with 3D animated imagery). At these distances we needed a very long FOV lens, for which we used Phoenix 500mm zoom lenses (for 35mm cameras), with C-Mount adapters and Panasonic PAL cameras. The marine LCD was 800x600 resolution with 300NITS and a special anti-reflective coating. For display the FERET face images were scaled up for display. As one can see from the examples in Figure 7, which are all from the same subject, the FERET data has a range of inter-pupil distances, poses and contrasts. This re-imaging model allows the system to control pose/lighting and subjects so as to provide the repeatability needed to isolate the effect of long-distance imaging and weather. As one can see, the collection produced images sufficient for identification but with the types of issues, e.g. loss of contrast and variations in size, that one would expect in a realistic long-distance collection. All experiments herein used FaceIt (V4), the commercial face-recognition system from Identix. This algorithm was one of the top performers in the National Face Recognition vendor tests [16]. These tests were completed un the 2001-2004 time frame.

This photo-head dataset is well suited formally study the issues to be encountered in using biometrics for long-range "uncooperative" subjects in surveillance video. One of the most controlled variations is what we call "self-matching" the probe and the gallery are based on the same image, except that the probe has been subject to the long-range (re)imaging process, atmospheric distrubances and the weather. The self-matching experiments are tightly controlled - they have exactly the same pose and subject lighting conditions. For initial testing we used a camera at approximately 15 ft and the rank-one self-matching performance was over 99%, showing the re-imaging process and LCD are not a significant issue. We then moved to the real photo-head collections. We generally ran each data set, which includes 1024 images, with 4 images of each subject, every 15 minuets, with collections over 4 months. The resulting 1.5TB of photo-head data was included in the DARPA HBASE, and subsets of the data are available from the authors. With 4 images per subject we can

Fig. 6 The Photo-head experimental setup. Two cameras are positioned at two different distances from a mounted weather-proof LCD display on a rooftop. Data capture occurred from dawn till dusk. Experiments were conducted over two years, capturing weather for all seasonal conditions.

use the BRR technique [13] to estimate standard errors and statistical confidence. All our graphs include such error bars, though for clarity it is often show only for the first plot point as it usually does not vary much as we change "rank" in the CMS curves.

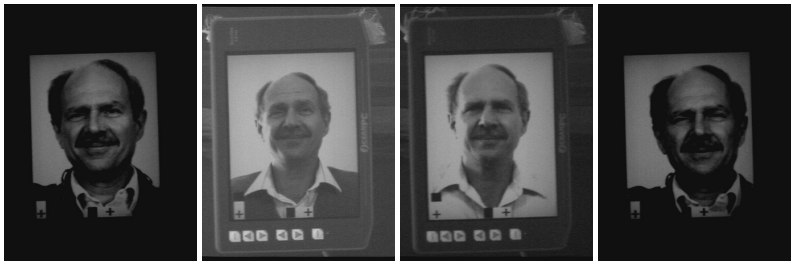

Fig. 7 Example Photo-heads: four different views of the same gallery subject taken at 100ft. Moving left to right: image taken at dawn, mid-morning, early afternoon, and evening.

3 In the Middle: Atmospheric and Weather

The obvious utility of a photo-head setup is the ability to capture outdoor conditions at all time of the year. Clearly, harsh weather conditions will have a significant impact on recognition performance, but, even seemingly good conditions can have unexpected impacts on recognition performance, depending on the interaction of atmospherics. Figure 8 shows the visual impact of weather in three different conditions captured during the photo-head collection. Note the images are rotated for display, the white on the left edge of the middle image is the snow building up on the top of the display.

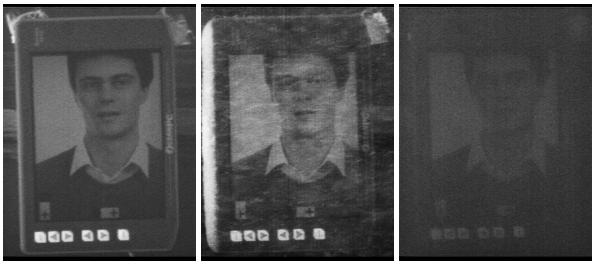

Fig. 8 From left to right: clear conditions, snow conditions, and rain conditions.

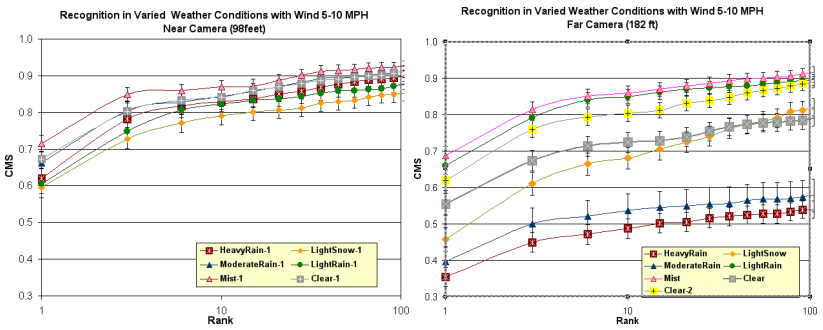

Fig. 9 CMC curves under various weather settings with self-matching.

The two graphs in Figure 9 show the impact of different weather conditions on face recognition. These are semi-log cumulative match curves with error bars from BRR. The curve shows the recognition rate on the vertical axis and the log "rank" used to decide correct recognition on the horizontal axis. Rank-N recognition means the person was within the top N scores of the systems.

Two things should be apparent from these graphs. First, looking at rank-1 recognition (or even rank-3), off the shelf systems are not sufficient for these ranges, even

under the best of weather conditions and ideal pose/expression. Recall, these are self-matching experiments, only the imaging systems, atmospheric and weather are stoping it from being identical images for probe and gallery).

Second, and not suprising, the far camera, at approximately 182ft, was much more significantly impacted by the variations in weather. (The best weather rank-1 recognition at 182ft was $< 70\%$). While it is not show here, increasing wind even more significantly impacted the system, in part because at these ranges even a small deflection of the camera causes significant blur and may take the face out of the sensors field of view. (With these long FOV lenses, we needed 30" housings which that increase wind loading.) These graphs are computed over more than 20,000 images and with the "controls" of the photo-head collections we know the images are identical, thus the variations are not artifacts of individual errors, pose or expression changes. The techniques of [15] improved performance slightly, not statistically significantly, in large part because they dont address blur or geometric distortion, only contract and dynamic range.

There are also some intially suprising results within these curves. If you look at the far camera results, you will see that light-rain and mist are statitically better than "clear" days. Can you generate a plausable hypothesize why "clear" days were not better? We controled for reported wind speed, so it is not that.

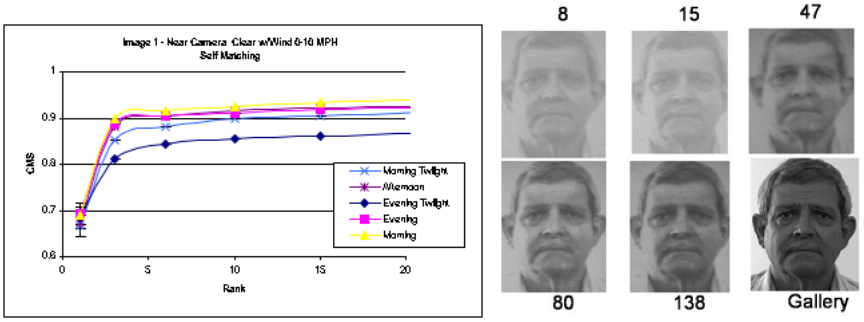

Fig. 10 On the left: variations over the time of day. On the right: recognition rank for various "quality" images

In addition to variations due to obvious weather effects, our experiments also showed that there were variations due to time of day. Atmospherics, such as thermal waves, can have a significant impact on recognition performance. Figure 11 shows thermal activity in difference images computed from a base frame and subsequent frames from a sequence timed over an entire day. The four images (two from the far camera and two from the near) shown are from two successive captures only a few minutes apart. Note the significant variation in the far camera between the two capture instances. Beyond atmospherics, rapid natural lighting changes, such as when the sun is shining down on the scene, and then is quickly hidden by a passing cloud, can also impact the collection. The significant variation visible in the

first image is likely due to this effect. But what about the others. You can see from the "structure" of the differences that its not just a shifting of the image significant differences are up and to the left of edges in the upper left , but down and tot he right on the lower/right part of the image. The "difference" patterns are more like localized zooms, probably caused by atmospheric lensing from thermals.

The impact of atmospherics and natural lighting changes on the far camera's recogntion rate is shown in Figure 10. These differences are statistically significant. Note that to reduce the impact of pose and lighting variations, these images are using the exact same image on the display as in the recognition database the only variations between the probe and the gallery are those caused by the imaging system. Recall that indoors at 15ft, the performance on this type of data is nearly perfect. Even with this very strong constraint we see that at 182ft on a clear and low-wind day, for Rank N recognition the performance of one of the best commercial algorithms of its day is below 65% with $N < 4$ and still below 80% recognition rate even when N is 10. Again, these are averaged over hundreds of trials with 1024 images per trial, so this is not a sampling artifact.

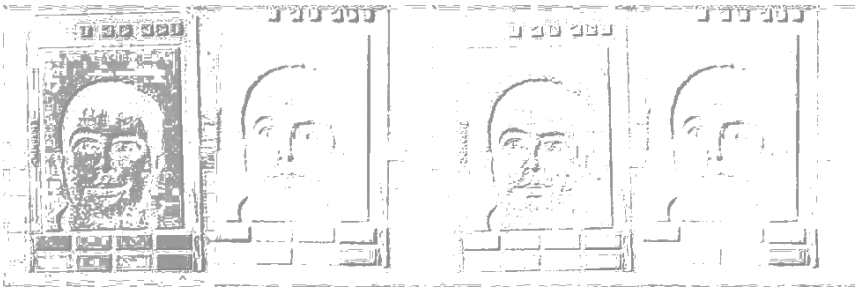

Fig. 11 Difference images highlighting thermal activity and natural lighting changes for a sequence of frames captured several minutes apart. The first and third images are produced by the far camera at 200ft, and the second and fourth images are produced by the near camera at 100ft.

A first guess might be that the weather impacts the raw "image quality", which is determining the performance. We examined various measures of facial image quality and (to our surprise), many of the errors had nothing to do with human perceived or measured image quality. While better quality images generally did do better, it was not as strongly related as one might hope. The right half of Figure 10 shows some examples of the recognition rank (i.e. where the imaged ended up when probes are sorted by match score) for a collection of images from a "same image" experiment. Rank and image quality in this set were inversely correlated. A detailed discussion of problematic issues with quality is presented in the next section.

Our research set off to find the causes of this unexpectedly poor performance. After considerable investigation we hypothesized the poor performance was due in large part to error in localization of the eyes. In [18] we presented an analysis of this theory. To definitively show the cause we added registration markers within

ourt photo-head data to allow us to transform the original eye coordinates to provide eye-locations in the captured images. The graphs above show the recognition performance (with error bars) for the off-the-shelf FaceIt algorithms and when use forced FaceIt to use the correct eye positions. The results on both cameras where statistically significant, and when the eyes are corrected the performance both far and near cameras are similar. These results are, of course, highly optimistic because the data for correction is artificial calibration points and secondly this is self-matching, with the same image as probes and gallery so the near perfect recognition is to be expected. It is important to note that the "eye-locations" being discussed are not just an question of where in the image the eyes appear but how that position related to where it should be in the image. In the "good quality" images of Figure 10, the corrected eye position is not in the middle of the eye! Atmospheric turbulence and lensing effect can distort the face image to the point that to work properly the system needs to use a different eye position for its coordinate system and normalization procedures. Many of the computed eye locations were visibly off the eye, and the average difference between the computed and FaceIt eyes was 6 pixels.

4 In the end: Measuring Quality

As mentioned in the previous section, an obvious guess is that weather and atmospherics reduce the raw "image quality", which is why they reduce performance. In order to study this potential impact, we formally defined quality and tried to study its relation to performance. We experimented with multiple measures of "quality", including blur and contrast in various ways. We eventually defined a blind signal to nosie ratio estimator for facial image quality, based on concepts from [28]. The concept is that statistical properties of edge images change with quality and have been showen to correlated with underlying signal to noise ratios. In our experiments, our derived measure is, under general conditions, better correlated with recognition rates than the other quality measures examined.

To derive this measure, suppose the probability density function of an edge intensity image, $\parallel \nabla I \parallel$ is given by $f \parallel \nabla I \parallel (\cdot)$ whcich is assume to have mean μ. The histogram of edge intensity image I can be modeled as a mixture of Rayleigh probability density functions, and that can be used to show that an estimate the signal-to-noise ratio (SNR), is given by

$$QS = \int_{2\mu}^{\infty} f \parallel \nabla I \parallel (r) dr$$

It has been proven that the value of QS for a given noisy image is always smaller than the value of QS for that image with less noise. Zhang and Blum also show it it can estiamte blur and is overall correlated with Signal to Noise ratio.

Choosing a fixed sized window around the eyes (examples are shown in Figure 12), we can define the Face SNR image quality as:

Fig. 12 Window around eyes for various images qualities.

$$Q' = \frac{\sum edge\ above\ 2\mu\text{'s } pixels}{\sum edge\ pixels} \simeq \int_{2\mu}^{\infty} f \parallel \nabla I \parallel (r)dr \tag{3}$$

This Face SNR IQ estimage, the ration of number of pixels above twice the mean strength to the total number of edge pixels, is easily calculated and can be shown to be a good approximation to QS.

The results from this estimator are well correlated with recognition rate. We took the images classified them into 5 bins using Q', and then examined the recognition rate for each subset. Figure 13 shows the correlation between quality and recognition rate, with overall correlations of 0.922 and 0.930 for two different galleries of photo-head data. Beyond the Face-SNR IQ estimate, we also performed experiments with multiple levels of blur, contrast, and multi-metric fusion - none were better than the blind SNR estimate. While at first this might seem signifcant, you should recall we were looking to understand/mitigate the impacts of atmospherics and wanted to use the quality predict, on a per image basis, if an image was going to be successful for quality. Unfortunately, a stong correlation was not sufficient for a good predictor.

We concluded that "quality" is indeed found in the recognition performance, not on what we "like" to imagine in some preconceived concept of quality or even our our blind SNR estimates. Interestingly, recent NIST studies [7] [2] on quality assessment come to this same conclusion. For the iris work in [7], three different quality assessment algorithms lacked correlation in resulting recognition performance, indicating a lack of consensus on what image quality actually is. In the face recognition work [2], out of focus imagery was shown to produce better match scores.

We had already shown that on an individual image level both precieved and measured quality could be inversly related to rank, but also showed that quality was positively correlated with overall recognition scores. We are not alone in this observation. We note that more recently [2] showed that, on a per instance basis, what is visually of poor quality produced good recognition results. good, it was not sufficient for per-image predictor. Reflecting upon this issue of quality a bit deeper, we began to wonder how to predict if an image would be successful and also how compare different measures of "quality" for face recognition.

The concept is using some measure of the system to predict if a particular input image will be (or is) successfully classified by the system. That is, we could threshold on quality and say any quality less than 2 will fail. With such a model we can compare the usefulness of different image quality measures.

The question then becomes to measure use the effectiveness of each predictor. Since this is measuring system performance, this then suggests that for a comparison of measures what is needed is some form of a Receiver Operator Characteristic

Fig. 13 In this plot, larger is better for quality. Correlations for blind SNR-based face image quality to recognition rate are 0.922 and 0.930. Experiments were also performed with multiple levels of blur, contrast, and multi-metric fusion. None were better than the blind SNR estimate.

(ROC) analysis on the prediction/classification performance. In [12] and [21] we define 4 cases that can be used as the basis of such a analysis. Let us define:

1. "False Accept", when the prediction is that the recognition system will succeed but the ground truth shows it will not. Type I error of the failure prediction and Type I or Type II error of the recognition system.
2. "False Reject", when the prediction is that the recognition system will fail but the ground truth shows that it will be successful. Type II error of failure prediction.
3. "True Accept", wherein the underlying recognition system and the prediction indicates that the match will be successful.
4. "True Reject", when the prediction system predicts correctly that the system will fail. Type I or Type II error of the recognition system.

The two cases of most interest are Case 2 (system predicts they will not be recognized, but they are) and Case 1 (system predicts that they will be recognized but they are not). From these two cases we can define the Failure Prediction False Accept Rate (FPFAR), and Failure Prediction Miss Detection Rate (FPMDR) (= 1-FPFRR (Failure Prediction False Reject Rate)) as:

$$FPFAR = \frac{|Case2|}{|Case2| + |Case3|} \tag{4}$$

$$FPMDR = \frac{|Case1|}{|Case1| + |Case4|} \tag{5}$$

With these definitions, the performance of the different reliability measures, and their induced classifier, can then be represented in a Failure Prediction Receiver Operating Characteristic (FPROC) curve, of which an example is shown in figure 14. Implicitly, various thresholds are points along the curve and as the quality/performance threshold is varied, predictions of failure change the FPFAR and

FPMDR just as changing the threshold in a biometric verification system varies the False Accept Rate and the Miss Detect Rate (or False Reject Rate). High quality data, which usually matches better, will generally be toward the upper right, with low failure prediction false alarms (and lower failures overall), but when good quality data does fail it is harder to predict it so more are missed. Lowest quality data is usually toward the bottom right, with few missed failure predictions, but more false predictions, as poor quality more often results in marginal but correct matches.

The advantage of using the FPROC curve as opposed to simple CMC or ROC curves with the data segmented by quality (or any other predictor variable) is two fold: First it allows for a more direct comparison of different measures on the same population, or a the same quality measure on different sensors/groups. Second, segmentation of data to generated CMC/ROC curves inflates the measure since it means the quality i data is not interacting with quality j data. Furthermore, it is not practical to compare measures or sensors when each one generates multiple ROC curves, especially if trying to compare multiple different "quality" measures. The FPROC evaluation approach allows us to vary the quality threshold over the gallery and see how it impacts prediction, while still maintaining a mixed gallery of qualities. The FPROC curve requires an "evaluation" gallery, and depends on the underlying recognition system's tuning, sensors, and decision making process.

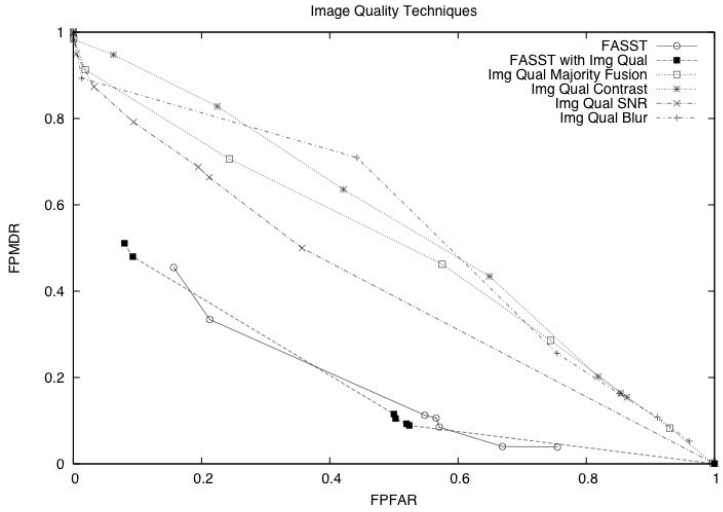

Fig. 14 FPROC for 4 different image quality techniques on 12,000 images, compared with the post-recognition Failure Analysis from Similarity Surface Theory (FASST) technique, with an without out image quality as a feature dimension.

The impact of switching approaches from a standard multiple CMC/ROC evaluation of image quality to the FPROC representation is noted in figure 14, where three different image quality techniques and a simple image qualityfusion scheme

are plotted. The underlying data is 12,000 images obtained in varied weather conditions outdoors. As can be seen, while our Face SNR estimage out performs the other quality measures in prediction, none of the image quality techniques are very powerful at predicting failure. Thus, while image quality is a well correlated with recognition overall, it can fare poorly on a per image basis where significant pose, lighting, contrast, and compression are allowed. In essence, any unconstrained setting where data collection is taking place.

Early on in our "quality" analysis, we introduced a compelling alternative approach [12], which was to learn to predict when a system fails and when it succeeds, and classify individual recognition instances using the learning as a basis. Based on the decisions made by a machine learning classification system, a Failure Prediction Receiver Operator Characteristic Curve can be plotted, allowing the system operator to vary a quality threshold in a meaningful way. Failure prediction analysis of this sort has been shown to be quite effective for single modalities [12], fusion across sensors for a single modality [26], and across different machine learning techniques [19] [21]. The FPROC quality prediction resuls of [12] are compared with basic image quality predictions in figure 14 and are clearly signficantly better.

Since the early observation on image quality, we have continued to build the alternative approach in the form of post-recognition analysis of the recognition score distributions. We call this analysis *Failure Analysis from Similarity Surface Theory*. Let S be an n-dimensional similarity surface composed of k-dimensional feature data computed from similarity scores. The surface S can be parameterized by n different characteristics and the features may be from matching data, non-matching data or a mixed set of both.

Similarity Surface Theorem 4.1 *For a recognition system, there exists a similarity surface S, such that surface analysis around a hypothesized "match" can be used to predict failure of that hypothesis with high accuracy.*

While the (empirical) similarity surface theorem 4.1 suggests that shape analysis should predict failure, the details of the shapes and their potential for prediction are unknown functions of the data space. Because of the nature of biometric spaces, the similarity surface often contains features at multiple scales caused by matching with sub-clusters of related data (for example, multiple samples from the same individual over time, from family members, or from people in similar demographic populations). What might be "peaked" in a low-noise system, where the inter-subject variations are small compared to intra-subject variations, might be flat in a system with significant inter-subject variations and a large population. These variations are functions of the underlying population, the biometric algorithms, and the collection system. Thus, with theorem 4.1 as a basis, the system "learns" the appropriate similarity shape information for a particular system installation. We have applied the FASST technique to a variety of different data sets, with implementations utilizing different learning techniques and underlying features generated from the recognition scores [21, 19, 26, 12]. Even if we cannot get good predictors from just image face quality data, the "quality" of face data for recognition *can* be learned from the distribution of scores after matching. Further, we have demonstrated a multi-modal fusion

approach [20] for this sort of failure prediction, which is able to enhance recognition performance beyond the best performing multi-modal fusion algorithms.

5 Conclusions

In this chapter, we looked at the issues in image acquisition that must be considered for effective long-range, facial recognition. As we have seen, both obvious and very non-obvious issues arise in all aspects of the image acquisition process. We discussed working volume and resolution issues that designer must consider. Lighting is always a challenge for outdoor acquisition, and problems multiply in low-light conditions. We have had good success measuring "face luminance" as opposed to scene lux or illuminance. Further, a mega-pixel EMCCD sensor with high resolution has provided us with images of sufficient quality and spatial resolution for standard face recognition, which overcomes many of the problems faced when LWIR systems. In general, with today's technology, cheaper components (low-resolution, interlaced sensors, rolling shutters, cheap lenses) will often hurt performance, in spite of their bargain price tag.

Designing a long-range facial recognition system requires extensive testing for validation. Our photo-head setup provided much insight into the effects of weather and atmospherics on long-range data acquisition. Not only did we learn of the impact on raw recognition scores, but also the limitations of image quality, which has been a traditional indicator of performance. Our observations have led us to define a new paradigm for image assessment, based on post-recognition score analysis. We believe this post-recognition analysis is a critical component to enhance performance, along with proper equipment selection and system design.

6 Face Image Acquisition Exercises

1. Using the standard formulas for illumination and luminance (or irradiance and radiance, your choice), sketch out the steps needed to determin the amount of light reacing the sensor for a face that is 100m away from the sensor. Using this determine if a 2Mega Pixel Camera, with 11micron pixels, fitted with a cannon EOS 400mm f2.8 lens with a 2x adapter could operate at difference scene light levels.
2. Derive a formula for the necessary focal length for long range face recogntion, at distance d, using a sensor with 1280x1024 pixels each of p microns accross. State any assumptions you need to make along the way.
3. Derive a formula for the operational volume, both width of the FOV and the Depth of field, for long range face recogntion, at distance d, using a sensor with 1280x1024 pixels each of p microns accross with a cannon EOS 400mm lens with a 2x adapter set at maximum aperature.

4. For a camera at 200m from a subject of interest, who is exiting a door, what is the necessary sensor size to have at least 3 seconds of video with sufficient resolution for face recognition (and temporal fusion). It is useful to ask for 3 seconds to ensure some frames at the top of the gate where there is minimal motion blur. State any assumptions you need to make along the way.
5. For a subject walking at normal speed, determine the shutter speed to ensure their walking does not produce more than a .5 pixel motion blur.
6. Consider the design of the Photohead experriment. List 4 limitations of the experimental design and suggest alternative designs that overcome these limitations (while on a university/student budget :-)).
7. In the definition of Face SNR IQ, we constrained it to a narrow region around the eyes and nose. Discuss the advantages and disadvantages of this windowing.

Acknowledgments

This work was supported in part by the DARPA HID program, ONR contract #N00014-00-1-0388, by NSF PFI Award # 0650251, DHS SBIR NBCHC080054, ONR STTR N00014-07-M-0421 and ONR MURI N00014-08-1-0638.

References

1. Adini, Y., Moses, Y. and Ullman, S.: Face Recognition: The Problem of Compensating for Changes in Illumination Direction. IEEE Trans. on Pattern Analysis and Machine Intelligence **19** (1997), no. 7, 721-732.
2. Beveridge, R.: Face Recognition Vendor Test 2006 Experiment 4 Covariate Study. Presentation at the NIST MBGC Kick-off Workshop (2008).
3. Canon: Optical Termoninology. The EF Lens Work III, Canon Inc., Lens Products Group, 2006, 192-216.
4. Chen, X., Flynn, P.J. and Bowyer, K.W.: IR and Visible Light Face Recognition. Computer Image and Vision Understanding **99** (2005), no. 3, 332-358.
5. Chen, T., Yin, W., Zhou, X., Comaniciu, D. and Huang T.: Total Variation Models for Variable Lighting Face Recognition. IEEE Trans. Pattern Analysis Machine Intelligence **28** (2006), no. 9, 1519-1524.
6. Dou, M.S., Zhang, C., Hao, P.W. and Li, J.: Converting Thermal Infrared Face Images into Normal Gray-Level Images. The 2007 Asian Conference on Computer Vision, 2007, II: 722-732.
7. Flynn, P.: ICE Mining: Quality and Demographic Investigations of ICE 2006 Performance Results. Presentation at the NIST MBGC Kick-off Workshop (2008).
8. Georghiades, A.S., Kriegman, D.J. and Belhumeur, P.N.: Illumination Cones for Recognition under Variable Lighting: Faces. Proc. of 1998 IEEE Conf. on Computer Vision and Pattern Recognition, 1998, 52-58.
9. Hoist, G.C.: CCD Array, Cameras, and Displays. SPIE Optical Engineering, 1996.
10. Jacobs, D.W., Belhumeur, P.N. and Basri, R.: Comparing Images Under Variable Illumination. Proc. of 1998 IEEE Conf. on Computer Vision and Pattern Recognition, 1998, 610-617.

11. Kong, S.G., Heo, J., Abidi, B.R., Paik, J.K. and Abidi, M.A.: Recent Advances in Visual and Infrared Face Recognition: A Review. Computer Vision and Image Understanding **97** (2005), no. 1, 103-135.
12. Li, W., Gao, X. and Boult, T.: Predicting Biometric System Failure. Proc. of the IEEE Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS 2005), 2005.
13. Micheals, R. and Boult, T.: Efficient evaluation of classification and recognition systems. Proc. of 2001 IEEE Conf. on Computer Vision and Pattern Recognition, 2001, I: 50:57.
14. Marasco, P. and Task, H.: The Impact of Target Luminance and Radiance on Night Vision Device Visual Performance Testing. Helmet- and Head-Mounted Displays VIII: Technologies and Applications. Edited by Rash, Clarence E.; Reese, Colin E. Proceedings of the SPIE, **5079** (2003), 174-183 .
15. Narasimhan, S. and Nayar, S.: Contrast Restoration of Weather Degraded Images. IEEE Trans. on Pattern Analysis and Machine Intelligence **25** (2003), no. 6, 713-724.
16. Phillips, P.J., Grother, P., Micheals, R., Blackburn, D., Tabassi, E. and Bone, M.: Face Recognition Vendor Test 2002 (FRVT 2002). National Institute of Standards and Technology, NISTIR 6965, 2003.
17. Phillips, P.J. and Vardi, Y.: Efficient Illumination Normalization of Facial Images. Pattern Recognition Letters **17** (1996), no. 8, 921-927.
18. Riopka, T. and Boult, T.: The Eyes Have It. ACM Biometrics Methods and Applications Workshop, 2003, 33-40.
19. Riopka, T. and Boult, T.: Classification Enhancement via Biometric Pattern Perturbation. IAPR Conference on Audio- and Video-based Biometric Person Authentication (Springer Lecture Notes in Computer Science) **3546** (2005), 850-859.
20. Scheirer, W. and Boult, T.: A Fusion Based Approach to Enhancing Multi-Modal Biometric Recognition System Failure and Overall Performance. In Proc. of the Second IEEE Conference on Biometrics: Theory, Applications, and Systems, 2008.
21. Scheirer, W., Bendale, A. and Boult, T.: Predicting Biometric Facial Recognition Failure With Similarity Surfaces and Support Vector Machines. In Proc. of the IEEE Computer Society Workshop on Biometrics, 2008.
22. Socolinsky, D., Wolff, L. and Lundberg, A.: Image Intensification for Low-Light Face Recognition. In Proc. of the IEEE Computer Society Workshop on Biometrics, 2006.
23. Socolinsky, D., Wolff, L., Neuheisel, J. and Eveland, C.: Illumination Invariant Face Recognition Using Thermal Infrared Imagery. Proc. of 2001 IEEE Conf. on Computer Vision and Pattern Recognition, 2001, I: 527:534.
24. Vogelsong, T., Boult, T., Gardner, D., Woodworth, R., Johnson, R. C. and Heflin, B.: 24/7 Security System: 60-FPS Color EMCCD Camera With Integral Human Recognition. Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VI. Edited by Carapezza, Edward M. Proceedings of the SPIE **6538** (2007), 65381S.
25. Wilder, J. and Phillips, P.J. and Jiang, C. and Wiener, S.: Comparison of Visible and Infrared Imagery for Face Recognition. Proc. of the IEEE Conf. on Automated Face and Gesture Recognition, 1996, 182-187.
26. Xie, B., Boult, T., Ramesh, V. and Zhu, Y.: Multi-Camera Face Recognition by Reliability-Based Selection. Proc. of the IEEE Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS 2006), 2006.
27. Yitzhaky, Y., Dror, I. and Kopeika, N.: Restoration of Atmospherically Blurred Images According to Weather Predicted Atmospheric Modulation Transfer Function (MTF). Optical Engineering **36** (1997), no. 11.
28. Zhang, Z. and Blum, R.: On Estimating the Quality of Noisy Images. Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1998, 2897-2900.
29. Zhao, W. and Chellappa, R.: Illumination-Insensitive Face Recognition using Symmetric Shape-from-Shading. Proc. of 2000 IEEE Conf. on Computer Vision and Pattern Recognition, 2000, I: 286-293.