High Magnification and Long Distance Face Recognition: Database Acquisition, Evaluation, and Enhancement

Conference Paper · September 2006

DOI: 10.1109/BCC.2006.4341635 · Source: IEEE Xplore

CITATIONS
10

READS
183

5 authors, including:

Yi Yao
SRI International
49 PUBLICATIONS **988** CITATIONS
SEE PROFILE

Besma Abidi
Phelps2020, Inc.
169 PUBLICATIONS **2,887** CITATIONS
SEE PROFILE

Natalia Schmid
West Virginia University
104 PUBLICATIONS **1,634** CITATIONS
SEE PROFILE

HIGH MAGNIFICATION AND LONG DISTANCE FACE RECOGNITION: DATABASE ACQUISITION, EVALUATION, AND ENHANCEMENT

Yi Yao[1], *Besma Abidi*[1], *Nathan D. Kalka*[2], *Natalia Schmid*[2], *and Mongi Abidi*[1]
[1]The University of Tennessee, Knoxville, Tennessee, 37996
[2]West Virginia University, Morgantown, West Virginia, 26506

ABSTRACT

In this paper, we describe a face video database obtained from **L**ong **D**istances and with **H**igh **M**agnifications, **IRIS-LDHM**. Both indoor and outdoor sequences are collected under uncontrolled surveillance conditions. The significance of this database lies in the fact that it is the first database to provide face images from long distances (indoor: 10m~20m and outdoor: 50m~300m). The corresponding system magnification is elevated from less than 3× to 20× for indoor and up to 375× for outdoor. The database has applications in experimentations with human identification and authentication in long range surveillance and wide area monitoring. The database will be made public to the research community for perusal towards long range face related research. Deteriorations unique to high magnification and long range face images are investigated in terms of face recognition rates. Magnification blur is proved to be an additional major degradation source, which can be alleviated via blur assessment and deblurring algorithms. Experimental results validate a relative improvement of up to 25% in recognition rates after assessment and enhancement of degradations.

1. INTRODUCTION

Substantial developments have been made in face recognition research over the last two decades and multiple face databases have been collected. They include the FERET [1], AR [2], BANCA [3], CMU FIA [4], and CMU PIE [5], to name a few. Face images or videos, typically visible RGB or monochromatic, are recorded under different resolutions, illuminations, poses, expressions, and occlusions. These databases, mostly collected from close distances and with low and constant camera zoom, are well suited for applications with controlled distances, such as identity verification at access points. The rapidly increasing need for long range surveillance and wide area monitoring calls for a breakthrough in face recognition from close-up distances to long distances and accordingly from low and constant camera zoom to high and adjustable zoom. The database described herewith serves this purpose and provides the research community with a standard testing foundation for long range face related research.

Before continuing our discussions, we give a definition of different ranges of system magnifications and observation distances for near-ground surveillance (both indoor and outdoor), as seen in Table 1.

Table 1 Definition of magnification/distance ranges.

Range	**Low/short**	**Medium**	**High/long**	**Extreme**
Magnification	1×~3×	3×~10×	**10×~30×**	**>30×**
Distance	<3m	3m~10m	**10m~100m**	**>100m**
Coverage	Most existing databases		**Our database**	

Our database has the following distinguishing characteristics. (1) According to the above definitions, most of the existing face databases fall into the category of low magnification with a few of them achieving medium magnification. The database collected by the University of Texas at Dallas involves medium distance sequences [6]. Their parallel walking videos start from a distance of 13.6m and their perpendicular walking videos are collected from a distance of 10.4m. However, their camera zoom remains low and constant. In comparison, our database aims at high to extreme magnifications and long to extreme distances. For indoor sequences high magnifications (10× to 20×) are used while for outdoor sequences extreme magnifications are obtained with a maximum of 375×. As a result, degradations induced by high magnification and long distance, such as magnification blur, are systematically presented in the data. (2) Our database closely resembles the real near-ground surveillance conditions (illumination changes caused by non-uniform roof light, air turbulences, and subject motion) and, more importantly, includes the effect of camera zooming, which is commonly ignored by other databases to date. Furthermore, sequences with various combinations of still/moving subjects and constant/adjustable camera zoom are collected for the study of individual and combined effects of target and camera motions.

In this effort, we also study the degradations in face recognition rates introduced by increased system magnification and observation distance. Apart from illuminations, poses, and expressions, magnification blur is identified as another major deteriorating source. To describe the corresponding degradations, a face image quality measure is developed. Several enhancement

algorithms are then implemented and their performances compared.

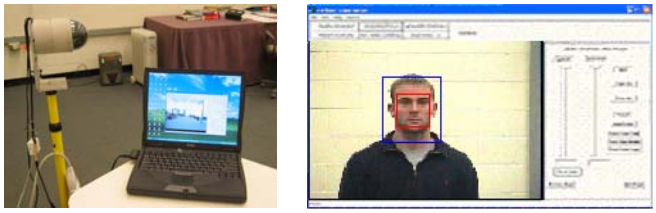
The remainder of this paper is organized as follows. Section 2 elaborates on our high magnification face video database. Face quality measures and enhancement algorithms are discussed along with experimental results in sections 3 and 4. Section 5 concludes this paper. The availability of this database is provided in section 6.

2. DATABASE DESCRIPTION

Our database collection, including indoor and outdoor sessions, began in February 2006 and is scheduled to be finished in July 2006. The final deliverable contains frontal view face images collected with various system magnifications (10×~375×), observation distances (10m~300m), indoor (office roof light and side light) and outdoor (sunny and partly cloudy) illuminations, still/moving subjects, and constant/varying camera zooms. Small expression and pose variations are also included in the video sequences of our database, closely resembling the variations encountered in uncontrolled surveillance applications.

2.1. Indoor sequences

For the indoor sequence collection, the observation distance was varied from 10m to 16m. Given this distance range and an image resolution of 640×480, a 22× system magnification is sufficient to yield a face image with 60 pixels between the subject's eyes[1]. Therefore, a commercially available PTZ camera (Panasonic WV-CS854) is used, as shown in Figure 1.



(a) (b)

Figure 1. The indoor sequence collection system: (a) Panasonic PTZ camera and (b) GUI. The blue and red rectangles depict the position and image area of the subject's face and T-zone (including eyes, nose, and mouth), respectively.

Our indoor database includes both still images (7 images per subject) and video sequences (6 sequences per subject). Still images are collected at uniformly distributed distances in the range of 10m to 16m with an interval of 1m approximately. The corresponding system magnification varies from 10× to 20× with an increment of 2×, achieving an approximately constant face image size to eliminate influences of resolution. Still images with low magnification (1×) are also taken from a close distance (1m), forming a reference image set. The achievable face recognition rates using this image set provide a reasonable performance reference for evaluating degradations caused by high magnifications.

The observation distance and system magnification are two major factors, to which this effort is devoted. Meanwhile, the effect of composite target and camera motions are included to achieve a close resemblance to practical surveillance scenarios. Therefore, the indoor video sequences are recorded under the following three conditions: (1) constant distance & varying system magnification, (2) varying distance (the subject walks at a normal speed towards the observation camera) & constant system magnification, and (3) varying distance & varying system magnification. Conditions 1 and 2 concentrate on the individual effect of camera zooming and subject motion while the combined effect can be observed in condition 3. In addition, system magnification is varied so that a constant face image size is obtained in condition 3. These video sequences can be used for studies of resolution, target motion, and camera zooming.

The above still images and video sequences are collected under fluorescent roof lights with full intensity (approximately 500Lux) and include a certain degree of illumination changes caused by the non-uniform distribution of the roof lights. Our indoor database also considers large amount of illumination changes under high magnification. A halogen side light (approximately 2500Lux) is added and a sequence is recorded as the intensity of the roof lights is decreased from 100% to 0, which creates a visual effect of a rotating light source.

The gallery images are collected by a Canon A80 camera under controlled indoor environment from a distance of 0.5m. The image resolution is 2272×1704 pixels and the camera's focal length is 114mm (magnification: 2.28×). Table 2 summarizes the sequence specifications. Figures 2 and 3 illustrate sample images of one data record in the database.

Table 2 Indoor sequence specifications.

Still images							
Mag. (×)	1	10	12	14	16	18	20
Distance (m)	1	9.5	10.4	11.9	13.4	14.6	15.9
Video sequences							
Conditions				**Mag. (×)**		**Distance (m)**	
Constant distance & varying system Mag.				10→20		13.4 and 15.9	
Varying distance & constant system Mag.				10 and 15		9.5→15.9	
Varying distance & varying system Mag.				10→20		9.5→15.9	
Illumination				20		15.9	

The indoor session has 55 participants (78% male and 22% female). The ethnic diversity is defined as a collection of 73% Caucasian, 13% Asian, 9% Asian Indian, and 5% African Descent. The image resolution is 640×480 pixels. For video sequences, our database provides uncompressed

[1] A minimum distance of 60 pixels between eyes is recommended by FaceIt® for successful face recognition.

frames in the format of BMP files at a rate of 30 frames per second as well as the associated AVI files compressed using Microsoft MPEG 2.0 codec. Each video sequence lasts 9 seconds. The total physical size for storage is 84 GB, with 1.53 GB per subject.

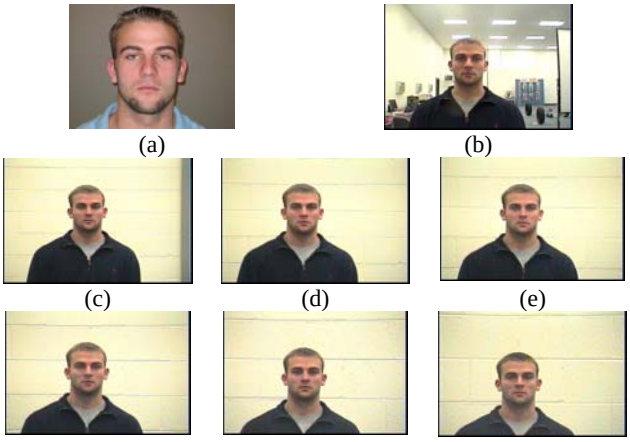


Figure 2. A set of still images in one data record: (a) gallery, (b) 1× reference, (c) 10× , 9.5m, (d) 12×, 10.4m, (e) 14×, 11.9m, (f) 16×, 13.4m , (g) 18×, 14.6m, (h) 20×, 15.9m. Probe images have approximately the same face image size.

Figure 3. A set of sample frames from collected sequences in one data record.

2.2. Outdoor sequences

For the outdoor sequence collection, a composite imaging system is built where a Meade ETX-90 telescope (focal length: 1250mm) is coupled with a Canon A80 camera (focal length: 38~114mm) via various eyepieces following an afocal connection (Figure 4). In order to achieve the required resolution (60 pixels between the subject's eyes) at long to extreme distances (50m~300m), three eyepieces are used: Meade 4.7mm, Meade 26mm, and Celestron 40mm. Accordingly, the achievable system magnification range is 24×~600×. Our outdoor database is still in the collection phase and is scheduled to be finished in July 2006. Table 3 lists the image sequences being collected.

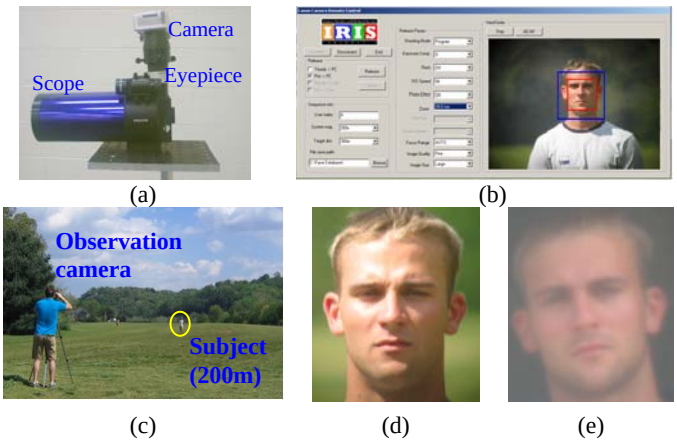


Figure 4. The outdoor sequence collection system: (a) composite imaging system, (b) GUI, and (c) data collection scene. Still images: (d) 62×, 50m and (e) 375×, 300m, severely blurred by high magnification and air turbulences.

Table 3 Outdoor sequence specifications.

Mag. (×)	62	125	187	250	312	375
Distance (m)	50	100	150	200	250	300

3. FACE IMAGE QUALITY ASSESSMENT

In face detection and tracking, cost functions are used to describe the probability of an area being a face image. The term face quality assessment was first explicitly used by Identix [7], where a face image is evaluated according to the confidence of detectable eyes, frontal face geometry, resolution, illumination, occlusion, contrast, focus, *etc*. Kalka *et al.* applied the quality assessment metrics, such as lighting (illumination), occlusion, pixel count between eyes (resolution), and image blurriness, originally proposed for iris to face images [8]. Xiong *et al.* developed a metric based on bilateral symmetry, color, resolution, and expected aspect ratio (frontal face geometry) to determine whether the current detected face image in a surveillance video is suitable to be added to an on-the-fly database [9].

Long distance and high magnification introduce severe and non-uniform blur, which is unique to our database in comparison with most existing databases collected from close distances and with low magnifications. Thus, in addition to the aforementioned metrics, our first priority is to examine the effect of high magnification blur on face image quality and face recognition rates. We begin our study with the assessment of high magnification blur and describe the corresponding enhancement techniques, including image deblurring and illumination compensation,

in sections 3 and 4, respectively. Still images are used in the following experiments to exclude blurs from other sources such as subject motion, camera zooming, and improper focus.

3.1. Image sharpness vs. system magnification

Sharpness measures are traditionally used to evaluate out-of-focus blur. However, since image noise level increases with system magnification, conventional sharpness measures, sensitive to image noise, are not applicable. To avoid artificially elevated sharpness values due to image noise, adaptive sharpness measures are proposed and have been applied to high magnification systems to quantize magnification blur [10].

Adaptive sharpness measures assign different weights to pixel gradients according to their local activities. For pixels in smooth areas, small weights are used. For pixels adjacent to strong edges, large weights are allocated. Adaptive sharpness measures can be divided into two groups, separable and non-separable. Separable adaptive sharpness measures only focus on horizontal and vertical edges while non-separable measures include the contributions from diagonal edges. For separable approach, two weight signals are constructed, a horizontal $L_x(x,y)=[I(x+1,y)-I(x-1,y)]^p$, and a vertical $L_y(x,y)=[I(x,y+1)-I(x,y-1)]^p$, where $I(x,y)$ denotes the image intensity at pixel (x,y) and p is a power index determining the degree of noise suppression. The Tenengrad sharpness measure [11], for instance, then becomes $S=\sum_M\sum_N(L_xI_x^2+L_yI_y^2)$, where *M*/*N* denotes the number of image rows/columns and I_x/I_y represents the horizontal/vertical gradient image obtained via the Sobel filter. For non-separable methods, the weights are given by $L(x,y)=[I(x-1,y)+I(x+1,y)-I(x,y-1)-I(x,y+1)]^p$ and the corresponding adaptive Tenengrad is formulated as $S=\sum_M\sum_N L(I_x^2+I_y^2)$.

Figure 5 shows the computed sharpness values (non-separable adaptive Tenengrad measure with $p=2$). The mean sharpness values are obtained by averaging the values within each image set across different subjects. Each image set consists of face images collected at the same observation distance and with the same system magnifications. They present a clearer view of the overall behavior with respect to system magnification. As expected, image sharpness decreases gradually as system magnification increases.

3.2. Face recognition rate vs. system magnification

In the following experiments, the same set of gallery images (Figure 2(a)) is compared against different sets of probe images, each set consisting of face images collected at the same observation distance and with the same system magnification. We employ FaceIt® [12] as our evaluation tool and focus on the first rank performance and the overall cumulative match characteristic (CMC) measure for a numerical comparison of various CMC curves. The CMC measure is a quantified measurement of a CMC curve defined as $Q_{CMC}=\sum_{i=1}^{K}C_i/i$, where K is the number of ranks considered and C_i denotes the percentage of probes correctly recognized at rank i. In the following experiments, $K=10$ is assumed.

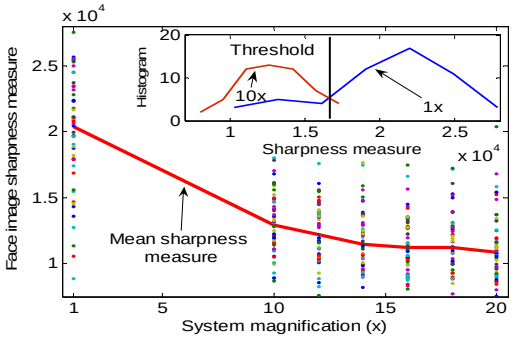


Figure 5. Sharpness measures for face images collected with different system magnifications. Dots represent the sharpness measures computed from face images of different subjects. The mean sharpness measures decreases as system magnification increases.

The relationship between face recognition performance, characterized by CMC, and system magnification is illustrated in Figure 6 and Table 4. It is obvious that deterioration from limited available fine details causes the face recognition rates to drop gradually as the system magnification increases. From magnification 10× to 20×, the CMC measure declines from 74.3% to 58.8%. There exists significant performance gap between the low (1×) and high (20×) magnification probes, which reveals that magnification blur is an additional major degrading factor in face recognition. This performance gap is to be compensated for by image post-processing. The decrease in face recognition rates caused by magnification blur is consistent with the behavior of image sharpness measures. Therefore, we could use sharpness measures as an indicator not only for the degree of magnification blur but also for the recognition rate.

4. FACE IMAGE ENHANCEMENT

4.1. Face image deblurring

Apart from numerous image deblurring algorithms, such as adaptive unsharp masking (UM) [13] and regularized image deconvolution [14], algorithms are proposed especially for face deblurring by making use of known facial structures. Fan *et al.* incorporated the prior statistical models of the shape and appearance of a face into the regularized image restoration formulation [15]. A hybrid recognition and

restoration architecture was described by Stainvas *et al.* [16], where a neural network is trained by both clear and blurred face images. Liao *et al.* applied Tikhonov regularization to Eigen-face subspaces to overcome the algorithm's sensitivity to image noise [17]. In our implementation, a wavelet based method is selected primarily because of its ability of multi-scale processing. From our experiments, the Haar level 1 wavelet transform appears to be the most promising candidate. The wavelet based enhancement algorithm proceeds as follows. The vertical, horizontal, and diagonal detail coefficients are thresholded to remove noise while the approximate coefficients undergo an unsharp masking to enhance image details. Afterwards, adaptive gray level stretching is applied to improve the contrast of the facial features.

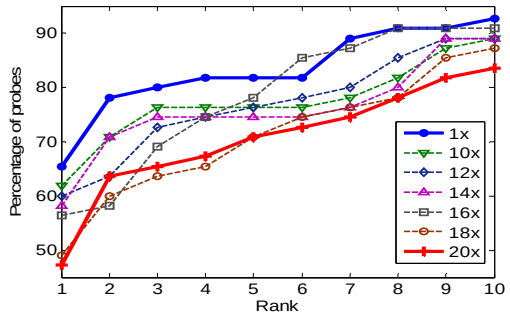


Figure 6. CMC comparison across probes with different system magnifications. Face recognition rates drop gradually as the system magnification increases.

Table 4 CMC measure and rank 1 performance comparison across probes with different system magnifications.

System Mag. (×)	**1**	10	12	14	16	18	**20**
CMC measure (%)	**74.3**	69.7	67.3	67.5	64.9	59.4	**58.8**
CMC at rank 1 (%)	**65.5**	61.8	60.0	58.2	56.4	49.1	**47.3**

In the following experiments, different probe sets are obtained by enhancing the same image set via various enhancement methods, including UM, regularized deconvolution, Liao *et al*'s Eigen-face [17], and wavelet based methods. In addition, two probe sets, the original face images and the 1× reference face images, are also tested and their performance serves as comparison references. The same experiments are repeated for image sets at different magnifications and similar observations are obtained. In the interest of space, only the comparison based on the 20× image set is shown (Figure 7 and Table 5). The wavelet based method is able to achieve the most improvement with a relative increase of 25% in CMC measure, yielding a performance comparable to the 1× reference. With proper post-processing, the degradation in face recognition rates caused by magnification blur can be successfully compensated.

In this work, we intend to use sharpness measures to predict face recognition rates at different system magnifications and determine whether a post-processing is necessary. Therefore, a threshold is derived. If the sharpness measure of the input image is below the given threshold, a post-processing is carried out. Based on the assumption of Gaussian distribution and Maximum Likelihood (ML) estimation, the threshold is defined as the mean sharpness measures of 1× and 10×: 16600, as shown in Figure 5. This threshold is obtained empirically and is application dependent. With sharpness measure selection (SMS), two samples from the original images meet the minimum requirement and hence no post-processing is carried out. The resulting performance is identical to the case where all images are processed, which verifies the suitability of the derived threshold.

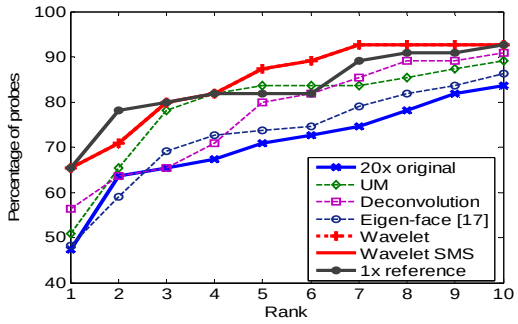


Figure 7. CMC comparison across probes processed by different enhancement algorithms. The performances of the wavelet methods with and without SMS are identical and outperform other tested methods.

Table 5 CMC measure and rank 1 performance comparison across probes processed by different enhancement algorithms.

Image	CMC measure (%)	CMC at rank 1 (%)
20× original	58.8	47.3
Eigen-face [18]	59.7	48.2
UM	64.3	50.9
Deconvolution	65.3	56.4
Wavelet	**73.6**	**65.5**
Wavelet SMS	**73.6**	**65.5**
1× reference	74.3	65.5

4.2. Illumination compensation

To evaluate the effect of varying illuminations, we employ algorithms from a class of well-known illumination normalization techniques, namely Jobson *et al*'s single scale retinex (SSR) [18] and Wang *et al*'s Self Quotient Image (SQI) [19]. In the interest of space, only the experimental results with 90% roof light are illustrated (Figure 8 and Table 6). Despite high magnifications and long distances, the SQI method is proved to be efficient and robust.

5. CONCLUSIONS

A unique face database including still images and video sequences collected from long distances and with high system magnifications was collected. This database features various types of degradations encountered in practical long range surveillance applications, with emphasis on magnification blur, which was addressed and identified as a

major degradation source in face recognition for the first time. Apart from existing face quality measures, a special metric evaluating face quality degradation caused by high magnification was applied and its efficiency in distinguishing low and high magnification images and predicting face recognition rates was illustrated. Image enhancement algorithms were implemented to show that degradations in face recognition rates introduced by magnification blur can be efficiently compensated for by applying the proper deblurring algorithms, such as wavelet based processing and that illumination change can be successfully balanced by using the SQI method. An improvement of 25% and 68% in CMC measure was achieved via magnification deblur and illumination compensation, respectively.

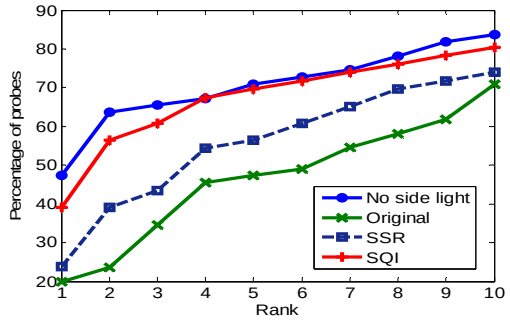


Figure 8. CMC comparison across probes processed by different illumination normalization algorithms. Illumination condition: 90% roof light + side light.

Table 6 CMC measure and rank 1 performance comparison across probes processed by different illumination compensation algorithms.

Image (20×)	Original	SSR	**SQI**	No side light
CMC measure (%)	31.4	38.5	**52.7**	58.8
CMC at rank 1 (%)	20.0	23.9	**39.1**	47.3

6. AVAILABILITY

Requests for the database will be entertained via FTP server or on hard disk, in a similar manner to our thermal-visual [20] database.

ACKNOWLEDGMENTS

This work was supported by the DOE University Research Program in Robotics under grant #DOE-DEFG02-86NE37968 and NSF-CITeR grant #01-598B-UT.

REFERENCES

[1] P. J. Philips, H. Moon, P. J. Rauss, and S. Rizvi, “The FERET evaluation methodology for face recognition algorithms”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[2] A.M. Martinez and R. Benavente, “The AR Face Database”, CVC Technical Report #24, June 1998.

[3] E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, B. Ruiz, and J.-P. Thiran, “The BANCA database and evaluation protocol,” *Audio- and Video-Based Biometric Person Authentication*, Guilfore, UK, pp. 625-638. Jun. 2003.

[4] R. Goh, L. Liu, X. Liu, and T. Chen, “The CMU face in action database”, *Int. Workshop on Analysis and Modeling of Face and Gestrues, in conjunction with Int. Conf. on Computer Vision*, Beijing, China, pp. 254-262, Oct. 2005.

[5] T. Sim, S. Baker, and M. Bsat, “The CMU pose, illumination, and expression database”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615-1618, Dec. 2003.

[6] A. J. O’Toole, J. Harms, S. L. Snow, D. R. Hurst, M. R. Pappas, J. H. Ayyad, H. Abdi, “A video database of moving faces and people,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 812-816, May 2005.

[7] P. Griffin, “Understanding the face image format standards”, *ANSI/NIST Workshop*, Gaithersburg, MD, *2005.*

[8] N. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, “Image quality assessment for iris biometric”, *SPIE Symposium on Defense and Security, Conf. on Human Identification Technology III*, Apr. 2006.

[9] Q. Xiong and C. Jaynes, “Mugshot database acquisition in video surveillance networks using incremental auto-clustering quality measures”, *IEEE Conf. on Advanced Video and Signal Based Surveillance*, Miami, FL, pp. 191-198, Jul. 2003.

[10] Y. Yao, B. Abidi, and M. Abidi, “Digital imaging with extreme zoom: system design and image restoration”, *IEEE Conf. on Computer Vision Systems*, New York, Jan. 2006.

[11] E. P. Krotkov, *Active computer vision by cooperative focus and stereo*, New York: Springer-Verlag. 1989

[12] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face Recognition Vendor Test 2002, Evaluation Report".

[13] G. Ramponi, “A cubic unsharp masking technique for contrast enhancement”, *IEEE Trans. on Signal Processing*, vol. 67, no. 2, pp. 211-222, Jun. 1998.

[14] T. G. Chan and C. K. Wong, “Total variation blind deconvolution”, *IEEE Trans. on Image Processing*, vol. 7, no. 3, pp. 370-375, Mar. 1998.

[15] X. Fan, Q. Zhang, D. Liang, and L. Zhao, “Face image restoration based on statistical prior and image blur measure”, *Int. Conf. on Multimedia and Expo*, vol. 3, Baltimore, MD, pp. 297-300, Jul. 2003.

[16] I. Stainvas and N. Intrator, “Blurred face recognition via a hybrid network architecture”, *Int. Conf. on Pattern Recognition*, Barcelona, Spain, pp. 805-808, Sept. 2000.

[17] Y. Liao and X. Lin, “Blind image restoration with eigen-face subspace”, *IEEE Trans. on Image Processing*, vol. 14, no. 11, pp. 1766-1772, Nov. 2005.

[18] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Processing*, vol. 6, no. 7, pp. 965--976, Jul. 1997.

[19] H. Wang, S. Z. Li, Y. Wang, "Face Recognition under Varying Lighting Conditions Using Self Quotient Image," *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Seoul, Korea, pp. 819-824, May 2004.

[20] http://www.imaging.utk.edu/files/IRISDB_0703_03.zip.