

**A Design Paradigm for
Long Range Iris Recognition Systems with
Sparsity Based Techniques for Iridal Texture
Enhancement**

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy
in the
Department of Electrical and Computer Engineering

Shreyas Venugopalan

B.Tech., Electronics & Telecommunications, University of Kerala
M.S., Electrical & Computer Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

27 May, 2015

THESIS COMMITTEE

Professor Marios Savvides, Advisor

Department of Electrical & Computer Engineering

Carnegie Mellon University

Professor Vijayakumar Bhagavatula

Department of Electrical & Computer Engineering

Carnegie Mellon University

Professor John Dolan

Robotics Institute

Carnegie Mellon University

Dr. Nalini Ratha

IBM Thomas J. Watson Research Institute

to Sarah Alexander, Venugopalan K.G. and Gita Venugopalan

Acknowledgments

I want to thank my parents and my wife for always being there and for being my pillars of support. They helped me keep my sanity and this thesis would not have been possible without their support and their wisdom!

My deepest gratitude to my advisor, Prof. Marios Savvides, who guided me and provided valuable advice throughout my PhD. I would like to thank him for providing me with the opportunity to work in his group and undertake this work. Without his support I could not have successfully brought this work to fruition.

I would also like to thank my thesis committee members, Prof. Vijayakumar Bhagavatula, Prof. John Dolan and Dr. Nalini Ratha for their valuable comments and feedback which has helped make this thesis a stronger representation of all the work that has gone into it.

Last, but not the least, I would like to thank DoD Biometrics Identity Management Agency as well as Cylab at Carnegie Mellon University for financially supporting this project during my PhD.

Abstract

The role of biometrics, as an integral component of security infrastructures, has been firmly established in today's world. Its ability to accurately identify people from a distance has served as an important life-saving tool in the arsenal of our men and women in uniform, out in war theaters; it has also served as an effective forensics tool for our police force and off-late, is being increasingly adopted to secure sensitive transactions such as those with financial institutions. Traditionally, fingerprint patterns and facial photographs have held sway over identity verification tools; the former, because of its uniqueness and the latter because of its ubiquity. However, over the past decade, iris patterns have also come to the fore because of the combined advantage of uniqueness across the populace as well as the ability to image these patterns from a distance. Today most commercial iris acquisition systems report recognition accuracies upwards of 90%, over entire populations. However, the caveat is that these systems are constrained to operate very close to the user, owing to the challenges involved in achieving sufficient resolution over the iride texture from large distances. This dissertation aims to push the boundaries of iris imaging devices and introduces theoretical and practical considerations in designing a novel long range iris recognition system. The aim is to design a system capable of acquiring enrollment quality iris images, similar in quality to those captured by short range systems such as IriShield, LG, PIER, HIIDE systems, at much larger stand-off distances. The challenges involved in acquiring images with sufficient spatial resolution, when subjects are too far out from the optimal acquisition range, are explored and potential solutions are described. First, we describe the design methodology behind creating a long range iris system, to aid the reader in understanding the key design considerations when using available commercial

off-the-shelf hardware. Competing objectives such as achieving high magnification while maintaining sufficient depth of field, using a sensor with large pixel pitch for low noise performance while resolving the required patterns, etc. are discussed. As a result of this work, the first long range iris acquisition system of its kind, capable of acquiring enrollment iris images from up to 9m and matching up to 12m, has been designed and built. We show the system, present lessons learned during the design process as well as empirical results for the imaging performance. The latter half of the dissertation develops an iris texture enhancement algorithm for restoring degraded iris images - specifically noisy, low-resolution and out-of-focus/blurred images, which are often encountered in the real world. The method is developed and evaluated using synthesized degraded images from the NIST Iris Challenge Evaluation (ICE) database. A significant improvement in verification performance is also seen, when using this enhancement method, on actual degraded images captured with the proposed long range iris system. The proposed enhancement method is modeled as a linear regression problem that exploits the sparsity of iris textures in an over-complete dictionary. Improvements in iris recognition performance using the enhanced textures are studied in the latter half of this dissertation. In summary, this work enhances and exceeds the capabilities of current iris imaging systems for the explicit purpose of accurately identifying threats using iride patterns, from distances larger than are currently possible. This will serve as a necessary tool in the arsenal of security establishments to identify threats, save lives and to protect our freedoms for generations to come.

Contents

1	Introduction	1
1.1	State-of-the-art in iris imaging	4
1.2	Overview of this dissertation	6
2	Design Considerations for a Long Range Iris Imaging System	11
2.1	Focal Length Estimation	12
2.2	Estimating the Spatial Frequency of interest	13
2.2.1	Limit to spatial resolution due to sensor pixels	15
2.2.2	Limit to spatial resolution due to lens diffraction limit	16
2.3	Depth of Field	18
2.4	Spatial Resolution Measurements using a USAF Resolution Target	20
2.5	Estimating Focus	23
2.6	Illumination Panels	27
2.7	Main System Software Components	31
2.8	Iris Recognition Engine	35
2.8.1	Iris Segmentation	36
2.8.2	Iris Occlusion Masking	37
2.8.3	Iris Feature Extraction and Matching	41
3	Enhancing spatial resolution in iris texture images	45

3.1	Sparse reconstruction in an over-complete dictionary	48
3.2	Proposed iris texture de-noising framework	51
3.3	Proposed spatial resolution enhancement framework	55
3.3.1	Generating a coupled texture dictionary model with exemplar samples	60
3.3.2	Super-resolving low resolution texture	61
4	Experiments and Results	75
4.1	Long range iris database used for our experiments	78
4.2	Long range subject verification performance	80
4.3	Long range subject identification performance	87
5	Concluding Remarks and Future Work	93
5.1	Improving the design of the proposed system	95
5.2	A note on computational imaging and role in biometrics	97

List of Figures

- 1.1 Long range iris acquisition system built as part of this dissertation 2

- 2.1 Comparison of spatial resolutions using different camera sensor-lens combinations, that are representative of the state-of-the-art in imaging equipments. The focal length used in each case was decided based on the magnification required, as discussed in section 2.1. Pixel size and other physical dimensions specific to each sensor can be found in the respective hardware manuals. As discussed in sections 2.2.1 and 2.2.2, the maximum resolvable spatial frequencies depend on the pixel pitch on the sensor, the diffraction limit imposed by the physical characteristics of the lens and the illumination wavelength. 14
- 2.2 Variation of depth of field for varying N 18
- 2.3 The USAF resolution target used in our experiment. The group indicated in (b) should ideally be resolved by any iris acquisition system adhering to quality standards prescribed in [1]. Fig. 2.4 shows a zoomed in crop of this group at various distances, using the proposed system. 21
- 2.4 Images of zoomed in crop of group 3 element 1 on the USAF resolution target and eye images captured at $9m$, $10m$ and $11m$, using a Canon EOS70D with a $f = 800mm$ lens. 22
- 2.5 Profile of the $f_{measure}$ for a user standing $u = 8m$ from the system 24

2.6	Eye images of two users captured at (a),(b) 9m, (c),(d) 10m and (e),(f) 11m	26
2.7	Visual comparison of iris images from the same user using (a) our system from 9m (b) LG system [2] from 30cm, (c) the PIER [3] from 15cm and (d) the iris on the move system [4] from 3m.	27
2.8	Uniformly illuminated pin-hole grid used in our experiment to measure the response of pixels behind the Bayer array of the camera sensor, to different wavelengths	28
2.9	Measured relative power output from a single LED in our LED array solution.	29
2.10	Response of various pixels on the camera sensor when imaging using (a)750nm (c)780nm (e)830nm and (g)850nm. For the latter two cases, every pixel shows equal response. Ideally, imaging should be done at these wavelengths. Iris of a user imaged using 20nm band-pass filters centered at the corresponding wavelengths are shown in (b), (d), (f) and (h).	30
2.11	Major software modules that make up the proposed long range iris recognition system.	32
2.12	Focus estimation, prior to high-resolution face and dual eye acquisition. Following a coarse focus estimation over the subject’s face, the region of interest (ROI) is set as the right eye. Fine focus, using a gradient descent-based strategy (section 2.5), is performed over the ROI. Finally a set of focus-bracketed high-resolution face images are captured.	32
2.13	Fiducial points on the face are estimated using the Modified Active Shape Model (MASM) approach [5]. The subset of points around each eye region are used to estimate the center of the corresponding eye. The eye regions with the highest focus measure (eqn. (2.8)), are used ultimately for iris feature extraction and matching.	33

2.14	The figure shows examples of facial landmarks detected using the algorithm proposed by Seshadri et al. The images used are from COFW database[6]. This algorithm is able to localize significant features on the face inspite of the presence of occlusions such as masks, head scarves etc. In this thesis, this approach robustly localizes the eye regions. Additionally, since the proposed system captures a high resolution face along with eyes using a single sensor, a facial recognition layer can easily be added to the system for multi-modal (face+iris) biometric authentication.	36
2.15	(a) shows the detected edge points when gradients are measured along lines radiating from pupil center, (b) shows the perpendicular bisectors m and n of lines AB and CD that join pairs of edge points. The point of intersection of such bisectors is taken as the limbic boundary center.	38
2.16	Various stages of the iris segmentation algorithm illustrated on an image acquired using our system. (a) shows the acquired eye image, (b) shows the image with specularly removed, (c) shows the median filtered image and (d) shows the detected iris boundaries.	39
3.1	Examples of degraded iris images captured using the long range iris system. Some of the factors that can cause degradation - longer subject stand-off distance, inadequate illumination, inadequate focus.	47
3.2	Examples of segmented normalized iris images reconstructed with varying levels of sparsity in an over-complete dictionary (see eqn. (3.1)). With increasing L , the reconstruction uses more atoms and hence we see a better visual reconstruction.	49

3.3	ROC curves comparing verification performance when using the original left eye images from the ICE database in the probe set (the blue curve) with the cases when reconstructed images are used in the probe set. The reconstruction is done using the over-complete dictionary in eqn. (3.1), with varying levels of sparsity L	50
3.4	Two examples of image denoising using the proposed method. The noise added is white gaussian noise with zero mean and variance 0.02.	53
3.5	Two examples of image denoising using the proposed method. The noise added is white gaussian noise with zero mean and variance 0.05.	54
3.6	ROC curves depicting iris verification performance using de-noised images from the ICE database. Noise is simulated in the probe image set for the experiment using additive white gaussian noise with varying levels of variance.	56
3.7	ROC curves depicting iris verification performance using de-noised images from the ICE database. Noise is simulated in the probe image set for the experiment using additive white gaussian noise with varying levels of variance.	57
3.8	Example of normalized iris textures from our database captured at 9m and 11m; (a) and (b) show images from two different participants in our study. Our aim is to increase the spatial resolution in \mathbf{I}_l to that in \mathbf{I}_h , in order to improve iris recognition scores. On visually inspecting the areas outlined in red, one can see the reduced resolution in the \mathbf{I}_l images	59
3.9	ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 20×20 Gaussian kernel with increasing values of variance (gaussian spread).	62

3.10	ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 20×20 Gaussian kernel with increasing values of variance (gaussian spread).	63
3.11	ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 30×30 Gaussian kernel with $\sigma_{blur} = 20$	64
3.12	Examples of left eye images from the NIST ICE database artificially blurred and enhanced using the coupled dictionary model described in this chapter. The two types of enhancements - with and without class-specific gradient constraints are shown.	67
3.13	Examples of left eye images from the NIST ICE database artificially blurred and enhanced using the coupled dictionary model described in this chapter. The two types of enhancements - with and without class-specific gradient constraints are shown.	68
3.14	Examples of left eye images from the NIST ICE database artificially blurred and enhanced using the coupled dictionary model described in this chapter. The two types of enhancements - with and without class-specific gradient constraints are shown. The level of blur in this case is higher, compared to Fig. 3.13. We can visually make out the recovered details in either case. . .	69
3.15	ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 20×20 Gaussian kernel with a variance of 8. The performance when using gradient constraints in eqn. (3.6) is compared with the performance without using the constraints.	71

3.16	ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 30×30 Gaussian kernel with a variance of 20. The performance when using gradient constraints in eqn. (3.6) is compared with the performance without using the constraints.	72
3.17	Distribution of genuine and impostor score distributions for the de-blurring experiment in Fig. 3.15. The thin line indicates the score distributions of the original left eye images. The thick dashed curves correspond to the blurred images and the thick solid curves to the de-blurred images using gradient constraint. The thin solid curves correspond to deblurring without the gradient constraint	73
3.18	Distribution of genuine and impostor score distributions for the de-blurring experiment in Fig. 3.16. The thin line indicates the score distributions of the original left eye images. The thick dashed curves correspond to the blurred images and the thick solid curves to the de-blurred images using gradient constraint. The thin solid curves correspond to deblurring without the gradient constraint	74
4.1	(a) Example of face images using our long range system at distances (from left to right) $9m$, $10m$ and $11m$. (b) and (c) show eye images extracted from (a).	77
4.2	This figure compares iris images from the same user using (a) our system from $9m$, with images captured using the (b) IrisID iCAM7000 system [7] from $30cm$, (c) the IriShield-USBMK2120[8] from $3cm$	79

4.3	ROC curves showing verification performance using left eyes, using our optimal Gabor filter. In all experiments, images from <i>9m</i> were used in the gallery. The enhancement used is the technique without the gradient constraint proposed in section 3.3.2.	81
4.4	ROC curves showing verification performance using left eyes, using the open source implementation in [9]. In all experiments, images from <i>9m</i> were used in the gallery. The enhancement used is the technique without the gradient constraint proposed in section 3.3.2.	82
4.5	ROC curves showing verification performance using right eyes, using our optimal Gabor filter. In all experiments, images from <i>9m</i> were used in the gallery. The enhancement used is the technique without the gradient constraint proposed in section 3.3.2.	83
4.6	ROC curves showing verification performance using right eyes, using the open source implementation in [9]. In all experiments, images from <i>9m</i> were used in the gallery. The enhancement used is the technique without the gradient constraint proposed in section 3.3.2.	84
4.7	ROC curves showing verification performance using the Iritech system and the IrisID system.	85

- 4.8 Examples of four normalized iris images captured at $11m$ and their enhanced counterparts. The enhancement is done using the gradient constraint method outlined by algorithm 3 in section 3.3.2. A low level of sparsity $L = 3$ is used in this algorithm to ‘de-noise’ the images. The top row shows how this method enhances gradients in the low resolution input images. In the bottom row, the enhanced images appear of lower resolution because of the de-noising effect when using a low value of L . Summarizing the observation across the database, irrespective of whether the input image appears blurred or noisy, the output-enhanced images always have a ‘soft focus’ (due to the de-noising effect of the algorithm) while still maintaining gradient information along particular directions (due to resolution enhancement portion of the algorithm). 88
- 4.9 Comparison of genuine and impostor score distributions for both the left-(a),(b) and right-(c),(d) eye images. (a) and (c) show the distributions for probe images from $10m$, while (b) and (d) correspond to $11m$. In all cases, the dotted lines indicate the original distribution and the solid lines indicate the distributions after image enhancement. Red indicates impostor distributions, while blue indicates genuine distributions. In all cases, we see a distinct improvement when enhanced images are used instead of the original. 90

List of Tables

1.1	Comparison of various features in state-of-the-art long range iris systems with the proposed system.	7
3.1	Identification performance with and without de-blurring using increasing hamming distance acceptance threshold values	73
4.1	Comparison of identification scores using original and resolution-enhanced images with an acceptance threshold of 0.4.	89
4.2	Cross-device identification performance using a hamming distance acceptance threshold of 0.3.	91

Chapter 1

Introduction

The use of machine vision hardware and associated algorithms in the security industry has progressed by leaps and bounds over the past decade. This technology has taken over many tasks which until now, were very labor-intensive and often required ‘humans in the loop’. It has been key in protecting our societies from criminal elements. There are security camera installations being set up all over cities, at high-security government facilities, at ports, in war zones etc. Algorithms are being designed using these cameras, to pick up anomalies and any potential security risk, well before an unfortunate incident can occur. Amongst these algorithms, biometrics has emerged as an *avant-garde* of choice. The ability to identify and track persons of interest based solely on the image of their face, their iris or their finger print has helped garner a lot of interest in this area.

We have seen several successful deployments of such biometric security systems for access control[10], for immigration/emigration at ports (for e.g. the US-VISIT program[11] or the UAE border control program[12][13]), for law and order[14] etc. In the current industrial and research environment, the most popular biometric modalities are human facial features, iris and fingerprint patterns. Fingerprint recognition predates the other modalities due to its early inception, as part of the Henry Classification system, in the late 19th century in British India. Face and iris patterns have of late come to the forefront,

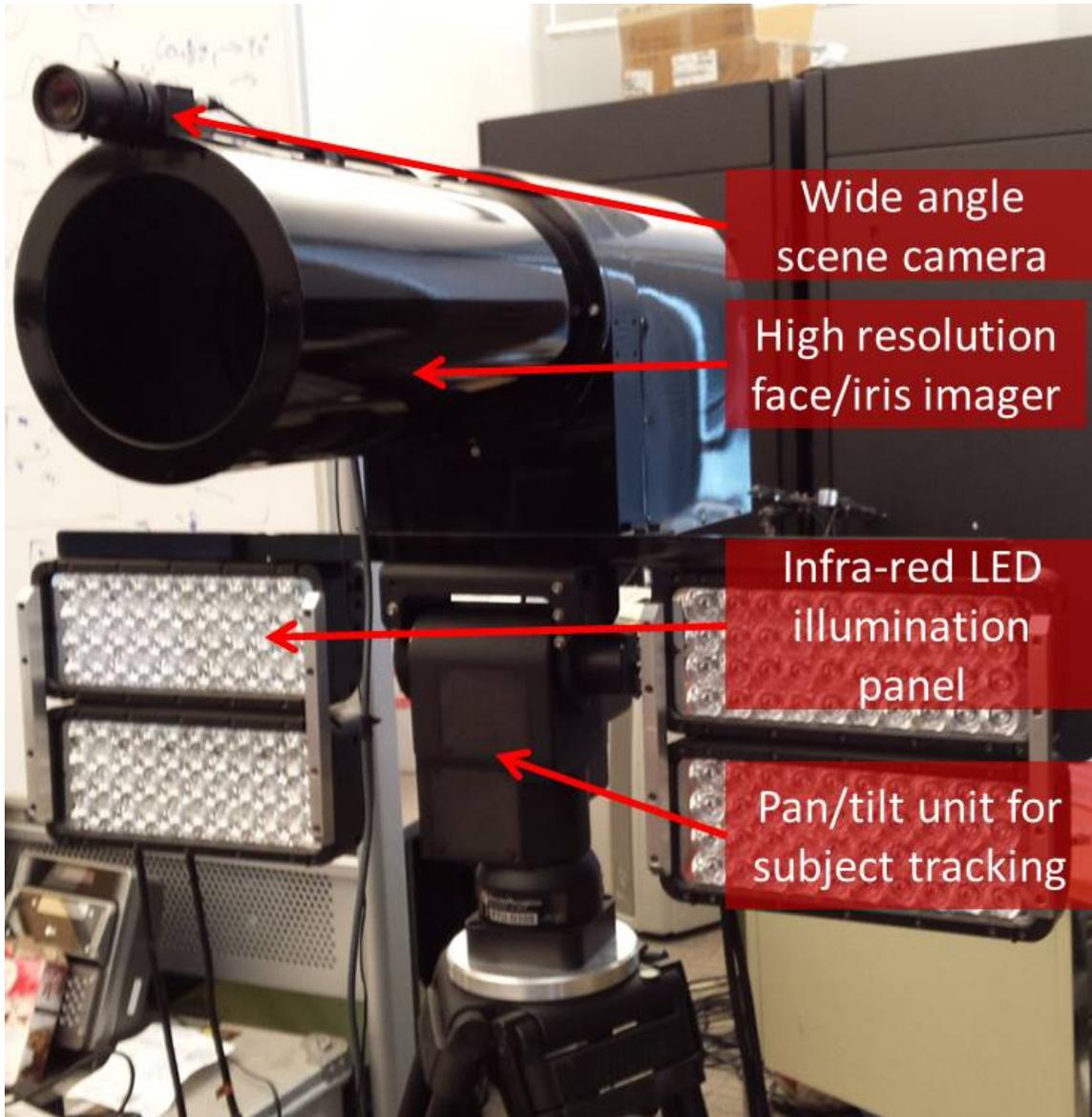


Figure 1.1: Long range iris acquisition system built as part of this dissertation

primarily due to the ability to image the associated biometric patterns from a distance. Of the two modalities, iris is preferred because they are thought to remain stable and constant over a person’s lifetime. Faces, on the other hand, change with age, with lifestyle, with facial hair and can be easily disguised or altered. The National Institute of Standards and Technology (NIST) has published the results of several studies that stand as a testament to the uniqueness of iride patterns across humans. NIST IREX III[15] and IREX IV[16] give further details on recognition accuracies using commercial iris recognition algorithms on existing iris recognition databases. A drawback of any iris-based biometric system, is the inability to image the iris beyond a certain system-to-user distance (also called ‘stand-off’ distance in this thesis), owing to the small size of the iris, when compared to a human face.

There has been very limited work to extend the range of iris based biometric systems, so that threats may be detected well before they reach our doorstep. This thesis aims to consolidate and address the major challenges involved, in designing a long range iris acquisition and recognition system. Based on the knowledge gained from this discussion, we build our own long range iris imaging system capable of acquiring iris patterns from up to $12m$ away, which is greater than other systems reported in literature. High quality iris images, for enrollment can also be acquired using the proposed system from up to $9m$ away. The key design challenges when using commercial off the shelf (COTS) equipment, to build this system, are described in this work. In addition, in the absence of such hardware, we also propose an algorithmic approach to enhance the spatial resolution of iris patterns in cases of low quality, degraded acquisitions. Iris recognition performance both with and without resolution enhancement is studied in the latter sections of this thesis. Before going into further details, we first present a brief review of the state-of-the-art in iris acquisition.

1.1 State-of-the-art in iris imaging

Long range imaging solutions for surveillance have been successfully used for vehicle tracking, vehicle forensics using number plates etc. [17]. However, the use of such systems for imaging biometric features such as face and iris has seen limited success. The primary reason for this, is that the object of interest in the latter cases is much smaller than a vehicle and hence imaging these objects, with the required resolution, poses a greater challenge. Take the iris for instance - the sphincter muscle that dilates/contracts the pupil based on the amount of light incident on the eye - it has a typical diameter of only around $12mm$.

As a result, most commercial and state-of-the-art systems operate at distances very close to the subject of interest. The IriShield system[8] is an example of a system on one end of the iris acquisition spectrum. This device needs to be held right next to the eye for proper image acquisition. Two other examples of such systems are the popular Hand-held Interagency Identity Detection Equipment (HIIDE)[18] and the Portable Iris Enrollment and Recognition Device (PIER)[19]. These systems are capable of acquiring high quality eye images, albeit only one at a time. In addition they require a high level of required cooperation from the side of the user and a high level of skill on the side of the operator. These systems are popular with the U.S. Army and other government agencies working out in the field. Two other examples of popular commercial iris acquisition devices, with greater stand-off distances ($0.1m-1m$) are the IrisAccess iCAM series [7], the LG IrisAccess series[2] and the Iritech IriMagic devices[20]. These devices have a small capture volume and require the user to be positioned within this capture volume, so that good quality eye images may be acquired. Many of them use voice prompts to direct the user into an optimal focal plane, prior to acquisition, without which iris acquisition usually fails. Often, these voice prompts may seem unintuitive and difficult to follow to an inexperienced user, which can lead to failure to acquire scenarios. These systems are popular in access control settings and can be found mounted at entrances of several office buildings. In addition

they are also deployed in airports and other points of ingress into countries.

Another category of systems involves the use of pan-tilt-zoom cameras. These cameras alleviate the constraint of a fixed capture volume, i.e., the user does not have to be at a fixed spot in front of the camera. Early attempts at using these cameras are reported by Oki IrisPass-M [21], Sensar R1 [22], Wheeler et al. [23]. These systems are based on the use of multiple cameras - a wide-angle scene camera to detect the face/eye in the scene and a second camera with a high-magnification lens, to resolve the fine patterns in the iris. Depth estimation is performed using a stereo-camera setup. This information helps in estimating the position of the user in 3D space and hence, may be used in the focus and pan/tilt modules. Venugopalan and Savvides [24] use a commercial off the shelf (COTS) pan-tilt-zoom camera to track the faces of subjects and to acquire irises when the subject is still. The subjects' irises may be acquired from stand-off distances of up to 1.5 meters. They use a single camera setup to acquire both the face and iris from subjects of different heights.

Two iris recognition systems that operate over relatively longer stand-off distances, are the Sarnoff Iris-On-the-Move (IOM) system[4] and the Retica Eagle Eye System[25]. The former reports a range of $3m$, while the latter reports a longer range of $5m$. Several other companies that advertise long range iris recognition, do so over shorter distances of up to $0.5m - 2m$; examples include - Morpho Iris at a Distance(IAD)[26], IrisID iCAM D1000[27], Terrain Biometrics Insight system[28] etc. In the case of the IOM system, iris patterns are captured while users walk through a portal, equipped with strobing near-IR illumination panels. The subject stand-off distance required during acquisition is 3 meters and the acquisition system has a fixed focal length, with a reported depth of field of $5cm$. IR beam sensors are used to trigger and to stop acquisition, as a user walks through the portal. Compared to traditional desktop/wall mount systems this device has the advantage of an increased stand-off distance and a reduced level of cooperation from the subject. However, iris acquisition fails if a users iris is not acquired within the fixed capture volume

of depth $5m$, around the $3m$ stand-off mark. The Retica Eagle Eye system [25] alleviates this problem to some extent. It uses a scene camera, a face camera, and two iris cameras. The capture volume of this setup is larger compared to the systems described so far, with a depth of $3m$ volume around a subject stand-off mark of $5m$. Terrain Biometrics'[28] described a system to perform iris recognition at a stand-off of $2m$. The system can be used to enroll as well as verify users whose heights vary between $0.9m - 1.9m$. They use adaptive optics with a multi-stage, real-time closed loop control system, in order to find the subject within a capture volume of depth $1m$ around the $2m$ stand-off mark.

This dissertation describes a system with a range and capture volume greater than any found currently in academia or industry. The purpose of the system is to push the boundaries of long range iris-based biometrics, while simultaneously reducing constraints on the user - by embedding subject tracking, dynamic focus estimation, quality estimation and image enhancement capabilities into the system. In short, instead of expecting the user to be fully cooperative, the system does the job of positioning, tracking and focusing prior to iris acquisition. Another question addressed here is - what if the image acquired is not ideal, due to low light, inadequate focus or due to subject motion? These issues are also explored in this thesis. A comparison of the imaging features of our system with those of commercial systems is given in Table 1.1. An image of the complete system is shown in Fig. 1.1.

1.2 Overview of this dissertation

The aim of this dissertation is twofold:

- to consolidate all the design challenges involved in building a long range imaging system, for the specific purpose of iris recognition and also to provide qualitative analyses of the same. We provide a design methodology for building a long range iris imaging system for unconstrained, unobtrusive iris acquisition.

Table 1.1: Comparison of various features in state-of-the-art long range iris systems with the proposed system.

System Features	Sarnoff IOM [4]	Retica Eagle Eyes [25]	Wheeler et al. [23]	Venugopalan et al. [24]	Proposed System
Height \times Width of Capture Volume (m)	0.2 \times 0.4	3 \times 2	not mentioned	not constrained	not constrained
Maximum user distance (m)	3	6	<1.5	1.6	12
Minimum user distance (m)	3	3	not mentioned	0.5	6
Single sensor for Iris/Face Acquisition?	Yes	No	No	Yes	Yes
Can capture enrollment iris images?	No	No	No	No	Yes
Dynamic focus adjustment	No	Yes	Yes	Yes	Yes
Iris acquisition from mobile subjects	Yes	No	No	No	Yes
Ability to track subjects	No	Limited	Yes	Yes	Yes

- In addition, in the absence of ideal hardware, due to physical limits of the available imaging equipment and due to non-ideal imaging conditions in the real world, there is always the possibility of acquiring iris images that are not of ideal quality. This dissertation explores methods to enhance low resolution, blurred and noisy images using signal processing techniques and shows how these enhancements can help in improving the recognition performance. These methods lay the foundation of enhancement techniques that can also be applied to images acquired using less than ideal imaging hardware.

The following summarizes the structure of this thesis:

- In chapter 2 the major design considerations are outlined. Quantitative analyses of various sensor-lens combinations that adhere to the design requirements are also presented. Factors such as focal length requirements, spatial resolution, depth of field and illumination wavelength are discussed amongst others.
- Chapter 2 also outlines the main components of the iris recognition engine used in the system. A method to estimate appropriate image filter parameters for feature extraction from acquired iris textures is also presented there.
- Chapter 3 describes methods to enhance low-quality iris images. Low quality acquisitions include cases of noisy eye images, acquired with less than ideal illumination, acquisition with improper focus settings and acquisition when the subject is outside the depth of field. This chapter investigates methods to de-noise images and to enhance the spatial resolution, while assuming a linear degradation model. Sample images from the NIST Iris Challenge Evaluation[29] database are used to study the suitability of the proposed methods to improve iris recognition scores.
- In chapter 4 iris verification and identification performance when using the proposed long range system, is presented. A database of eye images gathered using this system is used and we compare the performance with that using two popular short/medium

range systems ([30][8]). The high recognition accuracy observed in our experimental evaluation, shows the versatility of this design in capturing good-quality iris images over long distances and over an extended capture volume. In addition, in the same chapter, we investigate the potential for improving both verification and identification when employing the techniques outlined in chapter 3.

- Chapter 5 summarizes the work presented and proposes potential directions for future work which can enhance the capabilities of the system. In addition, thoughts on novel imaging systems that break free from the typical *camera obscura* model and how they can affect imaging for biometrics are also presented.

Chapter 2

Design Considerations for a Long Range Iris Imaging System

In this section we describe the basic design considerations for an imaging system, with special emphasis on a long range iris application scenario. There are many conflicting requirements when one tries to image small objects, such as an eyeball, from a large distance. System engineers often have to make a trade-off in one aspect or the other:

- increasing the obtainable magnification can result in a shallow depth of field at long distances
- increasing depth of field, by limiting aperture size, decreases the maximum spatial resolving power (due to the diffraction limit, as explained later in this chapter)
- increasing the resolving power using a smaller pixel pitch in the image sensor is an option; however, such sensors typically have a smaller area, thereby reducing the field of view (compare for instance, full-frame sensors [31] and crop-sensors [32])
- another point of contention is noise performance - sensors with larger pixel pitch (i.e. pixel side; the terms ‘side’ and ‘pitch’ are used interchangeably in this thesis) have better sensitivity to light and hence report better signal-to-noise ratios (SNR)

compared to those with smaller pixel pitches

Overcoming these conflicting challenges and minimizing any trial and error involved in choosing one's equipment, requires careful consideration of the algorithmic requirements, the imaging framework and the target object properties. In the following sub-sections, we present a quantitative analysis of our requirements and our choice of relevant parameters, so that we can acquire an image that satisfies the algorithmic requirements.

2.1 Focal Length Estimation

The focal length required for imaging, constrains the range of optics one can consider. Hence we consider this requirement first in our analysis. The focal length is decided based on the magnification required, which in turn, is based on the algorithm's requirements. Within the context of long range iris recognition systems, in order to be considered as a high-resolution iris image, most of the research in this field uses a typical pixel count of 200 pixels across the iris diameter, as per the ISO/IEC 19794-6 document [1]. In order to translate this value to a physical dimension, we require knowledge of the size of pixels in the sensor used. For the purpose of this example, consider we have access to a high-resolution full-frame digital camera such as the Canon EOS 5D Mark II. The sensor pixel side for this camera is $6.45\mu m$ [31]. Using this constraint for pixel side, we can estimate the image size required for the particular problem of high-resolution iris acquisition,

Typical diameter of the human iris, $h_o = 12mm$

Image size required, $h_i = 200pixels$

$$\begin{aligned} &= 200 \times (\text{side of a sensor pixel})mm \\ &= 200 \times (6.45 \times 10^{-3})mm \\ &= 1.29mm \end{aligned}$$

Once we have this value, the magnification (M) is easily estimated as the ratio of image size to object size.

$$M = \frac{h_i}{h_o} = \frac{1.29}{12} = 0.1075 \quad (2.1)$$

An approximate value of focal length f can be obtained from the following expression that relates f to magnification (this expression can easily be derived from the thin lens equation. Details can be found in a standard optics textbook such as[33]),

$$f = \frac{Mu}{1 + M} \quad (2.2)$$

where u is the object distance. Taking the example of a person standing $u = 12m$ away from the system, the required focal length is calculated to be $f = 1164mm$, using the above expression. Calculating the focal length in this manner for any image acquisition scenario, provides us an initial point, for considering viable optics and commercial off the shelf (COTS) systems that can be readily employed.

However, given f , there are quite a few options available to us, because there are several commercial lens manufacturers with lenses covering the entire focal length range. If we consider, for example, the Canon lens line-up, they market several lenses in the ‘super-telephoto’ range - 400mm, 600mm, 800mm along with focal length extenders. One can use several combinations of lenses along with focal length extenders, in order to obtain the required f . The choice we ultimately make, will depend on the spatial resolving power required. The latter in turn, depends on both the choice of sensor and the lens; this is discussed in the following sections.

2.2 Estimating the Spatial Frequency of interest

The ability of an optical system to resolve the required spatial frequency is related to the lens’ modulation transfer function (MTF), as well as the sampling rate imposed by the sensor used. As the imaging sensor is discrete, the resultant sampling of the lens’ image

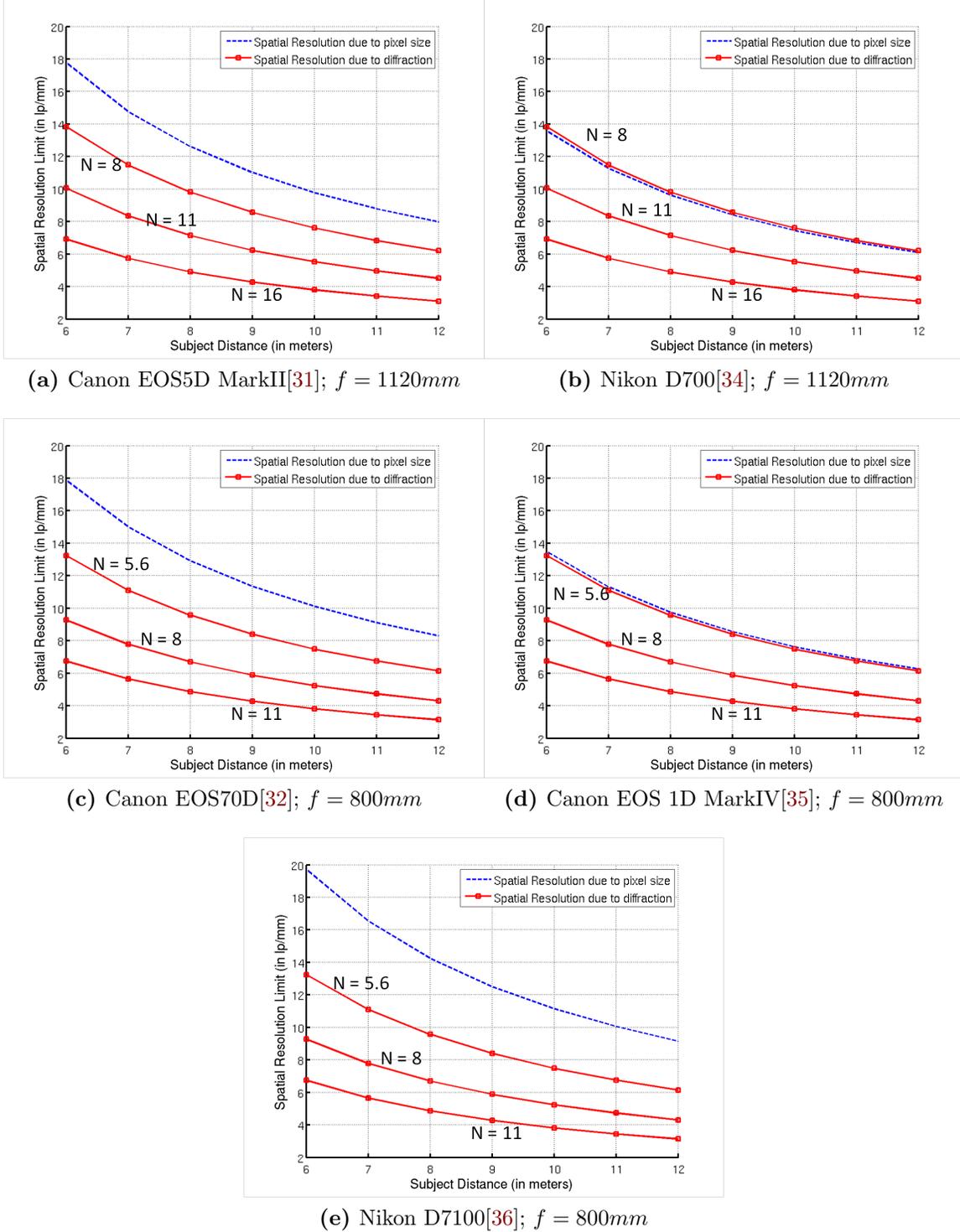


Figure 2.1: Comparison of spatial resolutions using different camera sensor-lens combinations, that are representative of the state-of-the-art in imaging equipments. The focal length used in each case was decided based on the magnification required, as discussed in section 2.1. Pixel size and other physical dimensions specific to each sensor can be found in the respective hardware manuals. As discussed in sections 2.2.1 and 2.2.2, the maximum resolvable spatial frequencies depend on the pixel pitch on the sensor, the diffraction limit imposed by the physical characteristics of the lens and the illumination wavelength.

circle, introduces limits to the size of the smallest element that may be resolved.

The first question then is, what is the spatial frequency, $f_{required}$, of interest in one's application? As specified earlier, for an iris image to be considered as high quality, the rule of thumb is to have 200 pixels across the diameter. Then, assuming a typical iris diameter of $12mm$, the smallest feature of interest is

$$u_{sample} = \frac{12}{200}mm = 60 \times 10^{-3}mm \quad (2.3)$$

This indicates that we need to resolve a spatial frequency of $\frac{1}{60 \times 10^{-3}}cycles/mm$. Then, as per Nyquist's sampling principle, we are assured to faithfully reproduce spatial frequencies of half this value, i.e. $f_{required} = \frac{1}{2 \times 60 \times 10^{-3}} = 8.33cycles/mm$. Given this requirement, what choice of imaging sensor can satisfy it? How does the choice of lens affect the minimum spatial frequency resolvable? The following two sub-sections deal with these questions.

2.2.1 Limit to spatial resolution due to sensor pixels

The pixel size (pixel pitch) of the sensor, a_{sensor} , determines the actual sampling rate at the object. Given the sampling element size a_{sensor} , the corresponding sampling size at the object plane can be determined using magnification M ,

$$a_{object} = \frac{a_{sensor}}{M} \quad (2.4)$$

Consider again, a full-frame sensor such as the one used in the Canon 5D Mark II [31] and an object distance $u = 12m$. In this case, $a_{sensor} = 6.45microns$ and $a_{object} = \frac{6.45}{M} = \frac{6.45}{0.1075} = 60microns$. Using this sensor, for iris imaging we are right at the limit for obtaining good-quality iris images at $u = 12m$ (as per the argument presented in the previous section). Fig. 2.1 shows the maximum resolvable spatial frequencies using different cameras at different user (object) distances (dashed blue lines). The plot was generated

based on the magnification required in each case, based on the pixel pitch of the individual sensors. If pixel pitch was the only constraint in resolving the desired spatial features, then we see that all these cameras would be ideal. However, the challenge of achieving fine spatial resolution is often complicated because of the effect of diffraction and the inability of a lens to resolve beyond a certain frequency. The effect is described in great detail in several optics textbooks and relevant publications such as [33] [37]. In this chapter, we will simply present results of interest, with respect to diffraction, for the iris acquisition application at hand. As shown in Fig. 2.1, the lens used differs in each case, based on the magnification required. The choice of lens, its maximum aperture size and the choice of wavelength used for illumination are factors which affect diffraction, as discussed next.

2.2.2 Limit to spatial resolution due to lens diffraction limit

When imaging an object, using light of wavelength λ and an aperture diameter of D , the angular diffraction limit in radians is given by

$$\theta_{diffraction} = 1.22 \frac{\lambda}{D} \quad (2.5)$$

The size of the smallest spot on the sensor due to the effect of diffraction is then approximately given by

$$a_{diffraction} = 1.22 \frac{f\lambda}{D} = 1.22\lambda N \quad (2.6)$$

where f is the focal length and N stands for the ratio of the focal length to aperture diameter, i.e., the f -number of the lens. In the case of our iris example, the typical imaging wavelength is in the near infra-red portion of the spectrum. If we continue with the example of using a Canon 5D MarkII sensor, from the previous section, we require a lens with $f = 1164mm$. This is achievable when using a Canon 800mm lens along with

a 1.4X focal length extender, which results in an effective focal length of $f = 1120mm$. When using the focal length extender, however, the effective aperture reduces in size and the f-number of the lens increases from $N = 5.6$ to $N = 8$. Using $N = 8$ and $\lambda = 850nm$ (which is the wavelength used in our system as discussed later in the dissertation), we compute $a_{diffraction} = 8.3microns$. Then the smallest spot size (sampling element size) on the object, due to diffraction $a_{object_diffraction} = a_{diffraction}/M = 77microns$. This is the smallest achievable sampling element size on the object and it translates to a maximum resolvable spatial resolution of $\frac{1}{2 \times 77} = 6.49cycles/mm$. We cannot hope to achieve better than this at a distance of $12m$, using this particular sensor-lens combination, as we have reached the *diffraction limit* of our imaging system.

Fig. 2.1 shows the maximum resolvable spatial frequencies due to diffraction (solid red lines), for different N , in the case of different sensor-lens combinations. In all cases, we use an $f = 800mm$ lens, the longest telephoto lens currently in the market. In the case of the full-frame sensors (Fig. 2.1(a) and (b)), we use an additional 1.4X extender, since the resultant focal length $f = 1.4 \times 800 = 1120mm$ reports better magnification when using these sensors, owing to their bigger pixel pitch. However, a drawback of using the extender is that, the effective aperture size decreases (indicated by higher values of N in these graphs; since N is the ratio of focal length to aperture size). This decrease in aperture size reduces the amount of light reaching the sensor, necessitating an increase in gain during image read-out - which can cause the image to get noisy in some cases. Using such analyses, we can decide on a sensor-lens combination that is able to both image spatial frequencies of interest and also performs at least as well as the diffraction limit in imaging. In the case of our system, we leave the aperture at the maximally open position (i.e. smallest f-number N), to achieve resolutions at the diffraction limit. We also wanted to avoid using an additional focal length extender (to limit the number of optical elements in the path of light, an increase in which, increases aberrations), so we chose to use a camera with a crop sensor (i.e. not a full-frame sensor) for our experiment. Specifically,

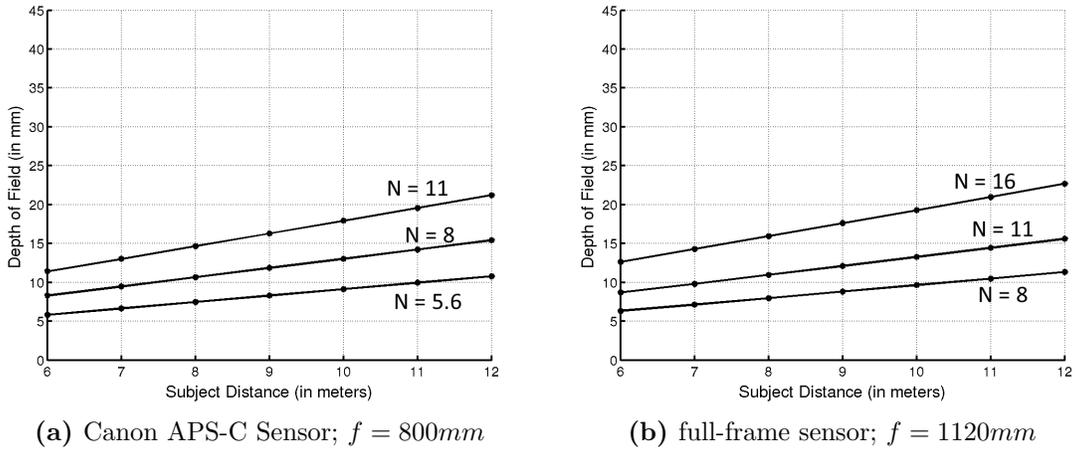


Figure 2.2: Variation of depth of field for varying N

we opted for Canon’s EOS70D used in tandem with a Canon 800mm lens. This system achieves the required magnification to obtain 200 pixels across the iris diameter. It is also able to resolve the required spatial frequency of 8.33cycles/mm as shown in Fig. 2.1. In addition, the Canon EOS70D has a well-documented C++ library for controlling the camera remotely(the associated SDK can be downloaded here[38]), which is necessary for our application.

2.3 Depth of Field

Depth of field of an optical system is the distance between the closest and the farthest objects in a scene, that are within acceptable focus. An iris acquisition system with a larger depth of field not only makes the process of acquisition easier, but also eases the constraints on the system user’s position. An approximate expression for depth of field of a system is given by

$$D_{dof} = \frac{2f^2u^2NC}{f^4 - N^2C^2u^2} \quad (2.7)$$

where,

N is the f-number of the lens - the ratio of the focal length to the aperture size.

u is the user distance.

C is the circle of confusion, i.e., the maximum acceptable size of the optical spot on the image plane when imaging a point source.

As per this expression, given a fixed object/user distance of u , the two other parameters we can tune to obtain a better depth of field are N and C . All articles written from the point of view of photography, prescribe a fixed C , based on the size of the sensor and the minimum spatial frequency the human eye needs to resolve on a standard $8.5'' \times 11''$ printout. For other imaging purposes, such as for iris recognition, we have to be more careful about the value of C used - in most cases it will not be the same as in the traditional use-case of professional photography.

As described earlier, for high quality iris acquisition, the spatial frequency of interest is 8.33cycles/mm and the corresponding smallest feature size is 60microns . This translates to a C of $60M$ for magnification M . The second parameter, N , is dependent on the application context. For a fixed focal length, N is inversely proportional to the size of the aperture. We can increase the depth of field by stopping down the aperture, in order to increase the depth of field. However, as pointed out in the previous section, this will increase the effect of diffraction and we run the risk of decreasing the spatial resolving power of our system. Fig. 2.2 quantifies the depth of field using the above expression and shows the variation for varying N for two different system configurations - when using a full-frame sensor along with an 1120mm lens (an 800mm lens with a 1.4X focal length extender) and when using a smaller Canon APS-C sensor with a lens of $f = 800\text{mm}$. These are the same sensor-lens configurations depicted in Fig. 2.1. We see that there is not too much of a difference in the depth of field achievable in either case. Our choice of using a crop sensor (found in the EOS70D) vs. using a full-frame sensor, for the iris acquisition application, was not dictated by the depth of field, since both cases are similar in this

regard. We try to overcome the effect of a shallow depth of field by using a two-stage focus strategy - one to fix coarse focus over the facial region and the second, a fine focus adjustment over the eye region. In addition, we capture a set of focus-bracketed images of the face and choose that image which reports a higher focus value. The metric used as well as the focusing mechanism are discussed later in section 2.5. As a concluding remark in this section, it is to be noted that, any effect of blur introduced by a shallow depth of field can be rectified using the resolution enhancement approach proposed in chapter 3. We show visual results of resolution enhancement as well as improvement in recognition performance, using the proposed methods, in chapters 3 and 4.

2.4 Spatial Resolution Measurements using a USAF Resolution Target

So far, we described the main design factors for a long range iris acquisition system. We determined the minimum feature size required to be resolved, to obtain a high quality iris image. This estimate helps one in deciding possible optical configurations to be used for various user distances. In this section, we experimentally show the spatial resolving power of our chosen system configuration, with the aid of a standard USAF resolution target. These measurements may be used to evaluate the imaging performance of any iris acquisition system, constrained to acquire iris patterns, with a pre-defined spatial resolution. A USAF resolution target (Fig. 2.3) is typically used to study the ability of an imaging system to resolve line pairs (variations, i.e., ‘cycles’ are represented as line pairs) that have a spatial frequency varying between $1lp/mm$ (line pairs per millimeter) and $228lp/mm$. $1lp/mm$ is represented by two $0.5mm$ dark lines separated by a $0.5mm$ bright line. We set up the resolution target against a white screen, illuminated using an infra-red LED array (the same used for our iris acquisition experiments[39]). This allows

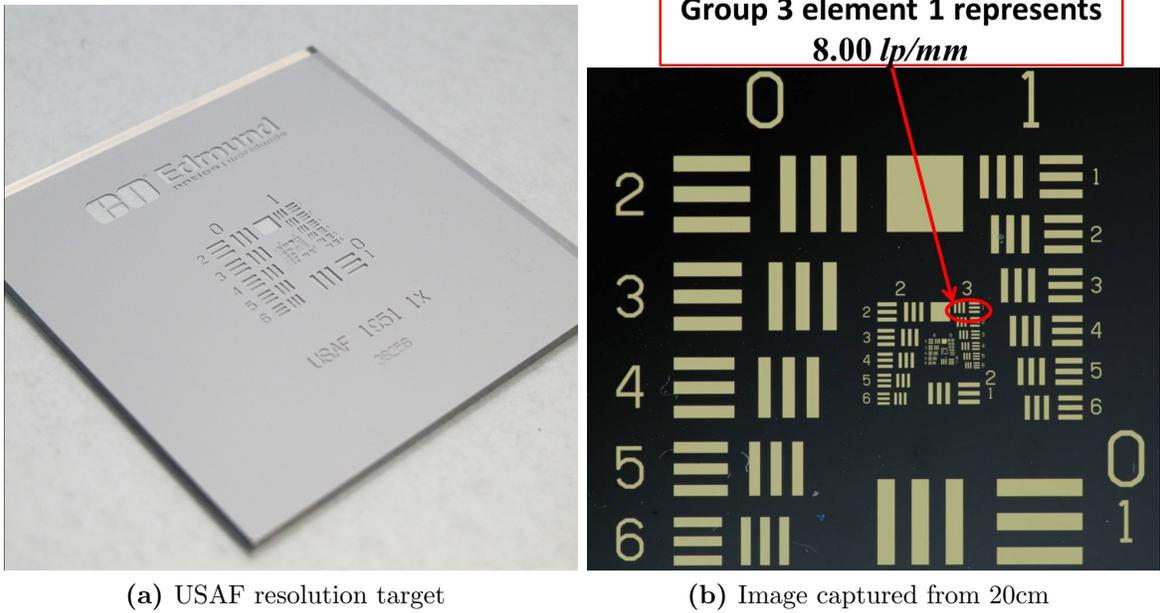
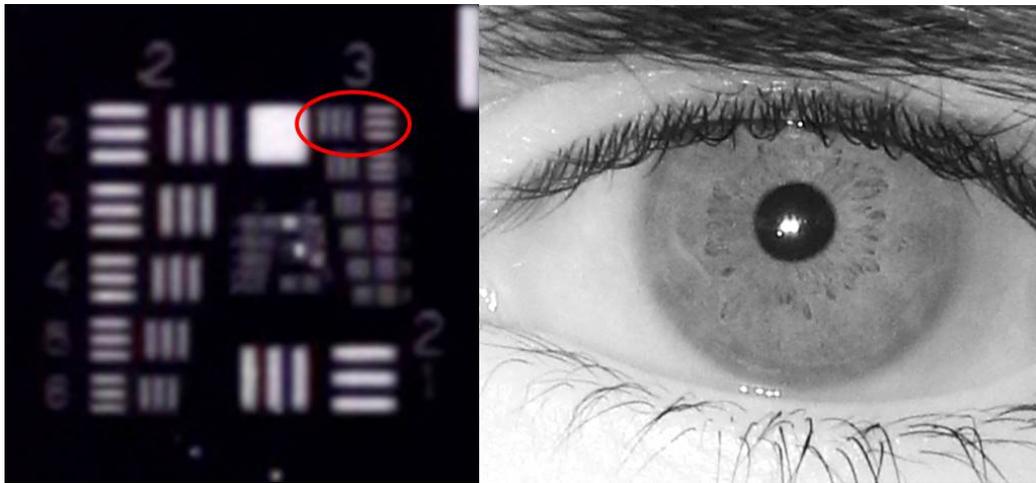


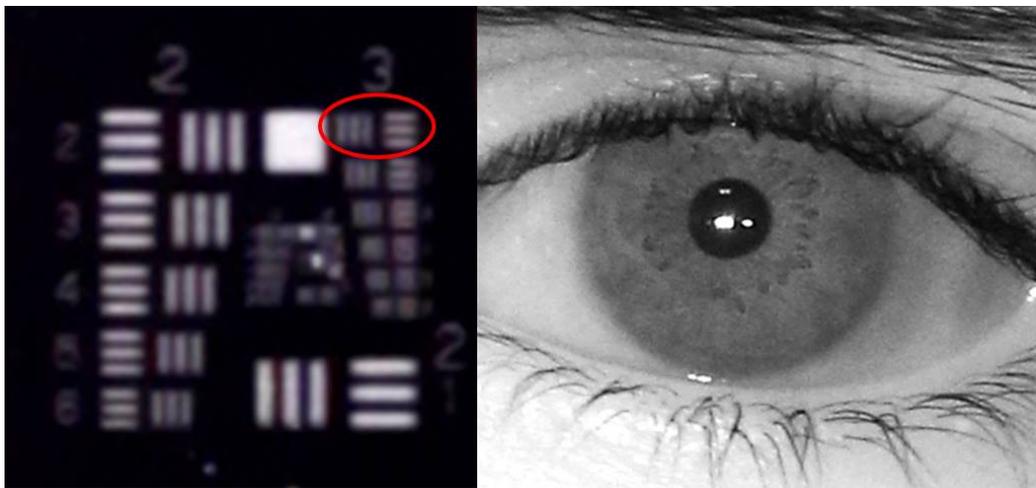
Figure 2.3: The USAF resolution target used in our experiment. The group indicated in (b) should ideally be resolved by any iris acquisition system adhering to quality standards prescribed in [1]. Fig. 2.4 shows a zoomed in crop of this group at various distances, using the proposed system.

us to achieve uniform illumination across the area of the resolution target. Fig. 2.3 shows an image of the target taken with a 180mm macro lens from a distance of 20cm. The bright lines in the target are the slits in the metal.

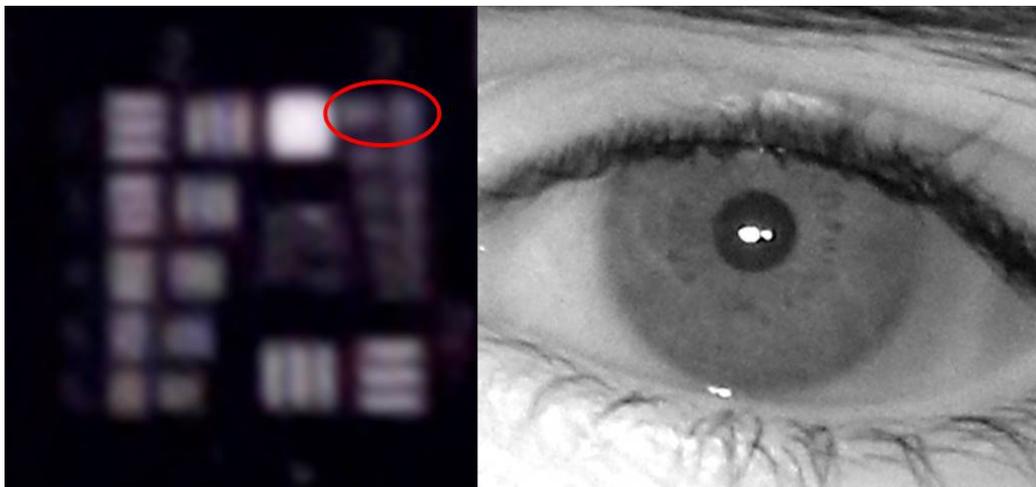
In section 2.2 we saw that an iris acquisition system should, ideally, be able to resolve line pairs with spatial frequency $8.33lp/mm$. This frequency is closest to $8.00lp/mm$ represented by the line pairs in group 3 element 1 of the resolution target (indicated in Fig. 2.3b). Fig. 2.4 shows images of the USAF target captured from distances of 9m, 10m and 11m. Eye images captured at corresponding distances are also shown. For the user distances of 9m – 10m, group 3 element 1 is resolved by our system, with reduced contrast. The corresponding eye images in Fig.



(a) images acquired from $9m$



(b) images acquired from $10m$



(c) images acquired from $11m$

Figure 2.4: Images of zoomed in crop of group 3 element 1 on the USAF resolution target and eye images captured at $9m$, $10m$ and $11m$, using a Canon EOS70D with a $f = 800mm$ lens.

2.5 Estimating Focus

As indicated earlier, one of the advantages of our system is the absence of a constraint on user position. We use a wide-angle scene camera to track the user and center him/her in the Canon EOS 70D frame. Following this, the key requirement is to achieve optimal focus over the user’s eye, in spite of the shallow depth of field. In this section we describe our auto-focus sub-system which interfaces with the lens focus motors. The system is driven by a spatial gradient-based focus measure algorithm; lens motor control is achieved by means of a lens interface module designed by Birger Engineering [40]. This module is used in tandem with a laser range finder[41], which initializes the gradient-based focus search strategy.

The focus measure over the frame $f_{measure}$ is given by the mean of the magnitude of the two-dimensional gradient at every pixel in the frame

$$f_{measure} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n |\nabla I_{i,j}| \quad (2.8)$$

where,

$$|\nabla I_{i,j}| = \sqrt{\left(\frac{\delta I_{i,j}}{\delta x}\right)^2 + \left(\frac{\delta I_{i,j}}{\delta y}\right)^2}$$

$I_{i,j}$ is the pixel intensity at (i, j) .

This $f_{measure}$ is used as the objective function in a gradient descent-based approach, to estimate the optimal focus position. An example of the focus measure variation, for a user distance of $8m$, is shown in Fig. 2.5. The x-axis represents the Birger mount position number x_{birger} (which maps directly to a lens focus setting) and the y-axis shows the corresponding $f_{measure}$ value. Since the user distance ranges from $6m - 12m$, the optimal

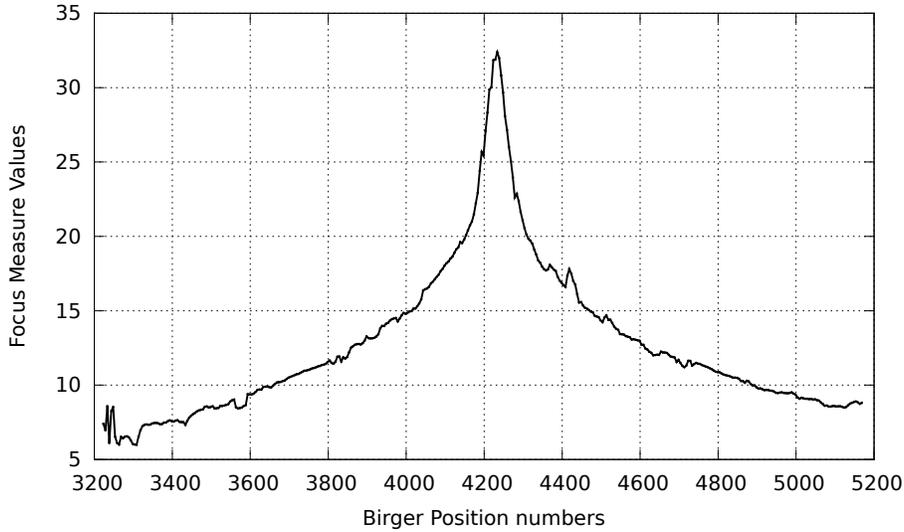


Figure 2.5: Profile of the $f_{measure}$ for a user standing $u = 8m$ from the system

point x_{birger} may be determined by solving the following optimization problem

$$\begin{aligned} & \underset{x_{birger}}{\operatorname{argmax}} \quad f_{measure}(x_{birger}) \\ & \text{s.t.} \quad 1000 \leq x_{birger} \leq 6000 \end{aligned} \quad (2.9)$$

The bounds on x_{birger} correspond to the minimum and maximum expected user position (for this example, $6m$ and $12m$ respectively). As seen from Fig. 2.5, this problem has a globally optimum (maximum) solution. This value is determined using a gradient descent (ascent in our case) search, initialized at one end of the user distance range. x_{birger} is initialized to 1000 and the position at the $(n + 1)$ th iteration of the gradient ascent search

$$x_{birger}^{n+1} = x_{birger}^n + \alpha \frac{\partial f_{measure}(x_{birger}^n)}{\partial x_{birger}} \quad (2.10)$$

α is the step size. When starting the search from the minimum user distance, i.e., $6m$, $\frac{\partial f_{measure}(x_{birger}^n)}{\partial x_{birger}} > 0$ always and hence we ‘climb the hill’. If at any iteration, we overshoot the global optimum point (i.e. the direction of the gradient changes), then we reduce the

Algorithm 1 Algorithm to Estimate Focus prior to Iris Acquisition

```
initialize  $x_{birger}^0 = 1000, \alpha = 500$   
while  $sgn\left(\frac{\partial f_{measure}(x_{birger}^n)}{\partial x_{birger}}\right) > 0$  do  
    Update  $x_{birger}^{n+1}$  using eqn (2.10)  
end while  
Set  $\alpha = -30$   
while  $sgn\left(\frac{\partial f_{measure}(x_{birger}^n)}{\partial x_{birger}}\right) < 0$  do  
    Update  $x_{birger}^{n+1}$  from eqn (2.10)  
end while  
Set  $\alpha = 8$   
while  $sgn\left(\frac{\partial f_{measure}(x_{birger}^n)}{\partial x_{birger}}\right) > 0$  do  
    Update  $x_{birger}^{n+1}$  from eqn (2.10)  
end while  
return  $x_{birger}$ 
```

value of α , change its sign and then proceed to the next iteration till we overshoot the optimal point again. This process is repeated till convergence is reached. The procedure followed in order to reach the optimal point is described by Algorithm 1. Typically, from our capture trials, we observe that the value of α will have to be changed three times before convergence is reached. The corresponding values of $\alpha \in [500, 30, 8]$. In order to speed up the focus search, instead of initializing the search at the minimum user distance, we initialize based on the output of a laser range finder. Specifically, we use the LDS30 laser range finder from Jenoptik [41], that has a range of $0.2m - 250m$, measured with a resolution of $1cm$. It returns reading at a rate of approximately 30 readings per second. This initialization is a coarse focus estimation to fix the focus close to the subject. Following this, we set the region of interest (ROI) of the high-resolution imager as the eye region and use algorithm 1 to set the optimal focus over the eye. This strategy allows us to quickly achieve optimal focus over the eye region and offsets the effect of the shallow depth of field.

Once the required focus value has been set, high-resolution images of the eye are captured. Further details regarding this system and the acquisition process have been pub-

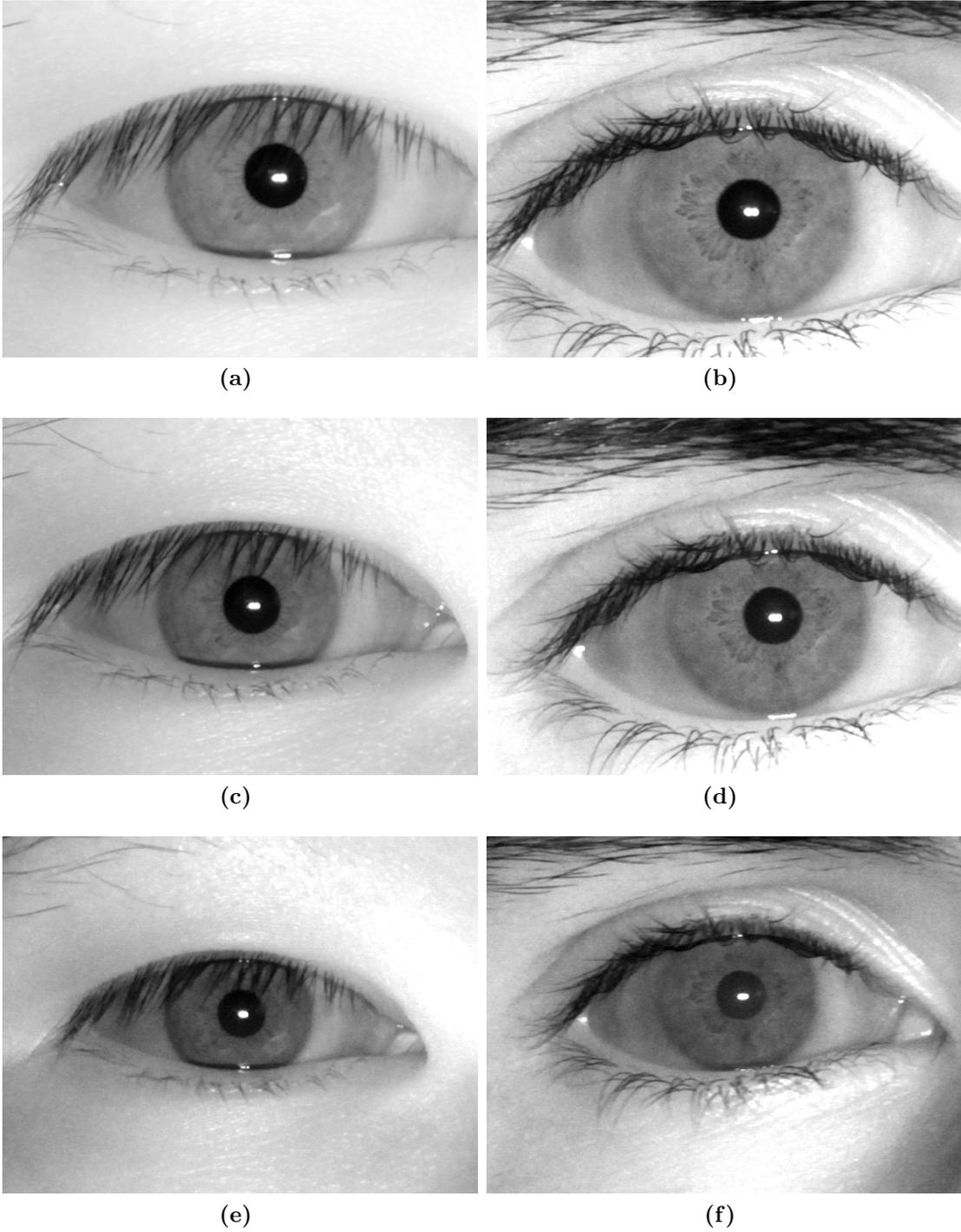


Figure 2.6: Eye images of two users captured at (a),(b) 9m, (c),(d) 10m and (e),(f) 11m

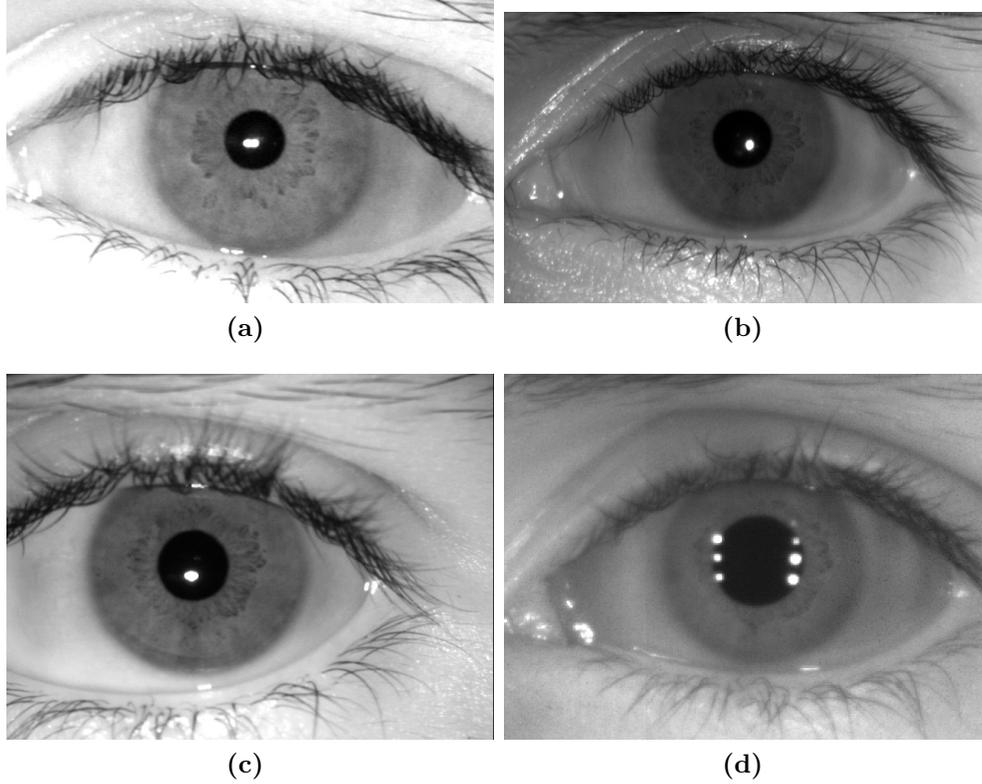


Figure 2.7: Visual comparison of iris images from the same user using (a) our system from 9m (b) LG system [2] from 30cm, (c) the PIER [3] from 15cm and (d) the iris on the move system [4] from 3m.

lished in [42]. A brief description of our process of iris segmentation, feature extraction and matching are described in section 2.8. Example eye images captured from various distances using this system are shown in Fig. 2.6. Fig. 4.2 shows an eye image acquired from 9m using the proposed system and visually compares it to images acquired from the same person, using other state-of-the-art COTS systems.

2.6 Illumination Panels

In this section we describe our choice of illumination for use in the system. In most cases, significant light in the visible spectrum incident on the iris, is absorbed while some portion of it is reflected off the cornea. However, most of the light in the infra-red wavelength



Figure 2.8: Uniformly illuminated pin-hole grid used in our experiment to measure the response of pixels behind the Bayer array of the camera sensor, to different wavelengths

of the spectrum is reflected back and can be imaged by the sensor. Ross et al.[43] have shown experimentally the validity of this argument and conclude that imaging the iris in the infra-red wavelength provides much of the discriminating information for an iris recognition system. We use a standard infra-red LED source such as the one described here [44] for our system. Four panels of these low power LEDs are optimally aligned such that the individual outputs superimpose over the required acquisition range. The system is also fitted with a filter wheel to switch between the visible spectrum and the infra-red spectrum as and when required. In order to determine the appropriate wavelength to be used for imaging an iris using our system, two main factors to be considered are as follows:

- Reponse of the red, green and blue pixels in the camera to different wavelengths. This is important since an unequal response from the Bayer array in the camera, will result in less information in one pixel when compared to its neighbor.
- Response of iris features to different wavelengths, which will affect the contrast obtained.

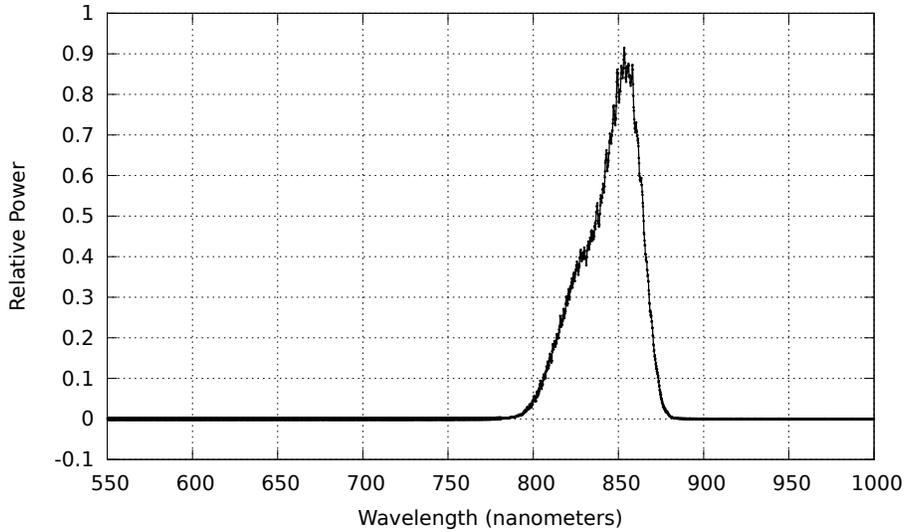


Figure 2.9: Measured relative power output from a single LED in our LED array solution.

As a first step, we performed an experiment to measure the response of the red, green and blue sensor pixels, when the scene is imaged at different wavelengths. We used a 75W halogen lamp (with 2900K color temperature) for this experiment. The target imaged was a uniformly illuminated pin-hole grid. The result is a diffuse surface which is color neutral over the imaging wavelengths. An image of the target is shown in Fig. 2.8. The target was then imaged using our system through a set of bandpass filters with wavelengths centered at $750nm$, $780nm$, $830nm$ and $850nm$. All these filters have a narrow band-pass of 20 microns so that we can determine whether any of the wavelengths elicit more information compared to the others from the iris region. Fig. 2.10 shows images of the pin-hole target captured. Using the same set of filters, we imaged the iris of a user, as shown in Fig. 2.10. We observed that the iris images appear similar when imaged at these wavelengths.

We opted to use a near infra-red illuminator with a spectrum centered at $850nm$, due to the equal response elicited from the Bayer array as well as due to the ready availability of LEDs at these wavelengths. Due to the limited quantum efficiency of the chosen sensor at these wavelengths, four $850nm$ LED panels are used. These panels are optimally aligned, such that the individual outputs superimpose, to project sufficient light $6m-12m$ from the

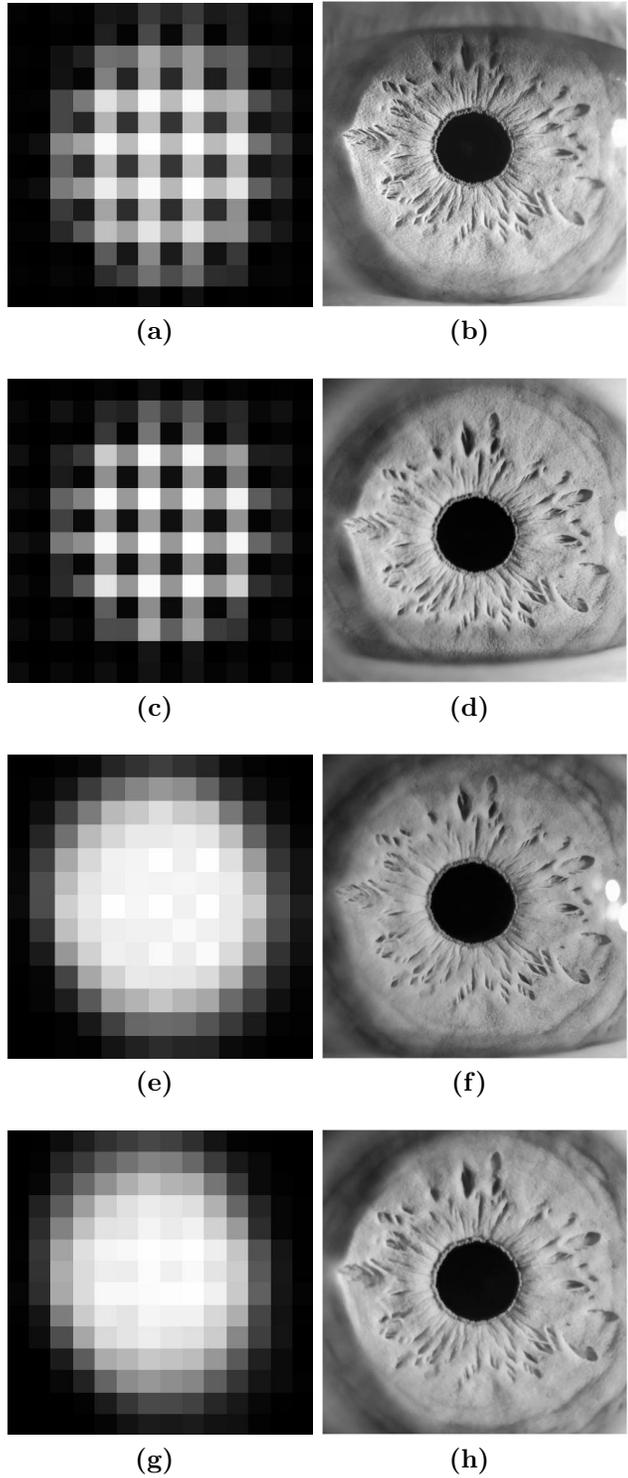


Figure 2.10: Response of various pixels on the camera sensor when imaging using (a) $750nm$ (c) $780nm$ (e) $830nm$ and (g) $850nm$. For the latter two cases, every pixel shows equal response. Ideally, imaging should be done at these wavelengths. Iris of a user imaged using $20nm$ band-pass filters centered at the corresponding wavelengths are shown in (b), (d), (f) and (h).

system.

2.7 Main System Software Components

As mentioned earlier, based on the design considerations earlier in this chapter, we built the long range iris system using a Canon EOS70D as the primary imager fitted with a Canon 800mm lens. A PointGrey Flea3[45] gigabit ethernet camera is used as a wide-angle ‘scene’ camera to track subjects in the frame. It has a resolution of 1.3MP and a frame rate of 31 *fps*. The high frame rate and the high speed gigabit ethernet link, allow us to detect and track persons of interest in real time. Once locked onto a subject, when the operator initiates biometric acquisition, this camera hands off control to the high-resolution primary imager for face and iris acquisition. The major software modules that make up the system are shown in Fig. 2.11. All of these modules were developed in C++, using publicly available libraries such as the Canon EDS SDK [38] and OpenCV[46]. Following is a brief description of each module:

1. **Face detection and tracking** - This module is necessary to track subjects of interest standing in front of the system. A Kalman filter-based face tracker is built on top of a face detector to smooth any variation in the detection output. The output of the tracker is used to control the pan/tilt mechanism on which the whole system is mounted.
2. **Focus estimation** - Focus estimation is based on the algorithm outlined in section 2.5. As mentioned in that section, we use a serial lens motor controller available from Birger Engineering [40] in order to control the lens motors independent of the camera used. Using a spatial gradient metric, we estimate whether the frame is in focus and direct the lens motors accordingly. The focus algorithm is initiated, when the subject is still and ready to be authenticated/identified. Initially a coarse focus is set on the face region. Following this, using the eye localization output from the

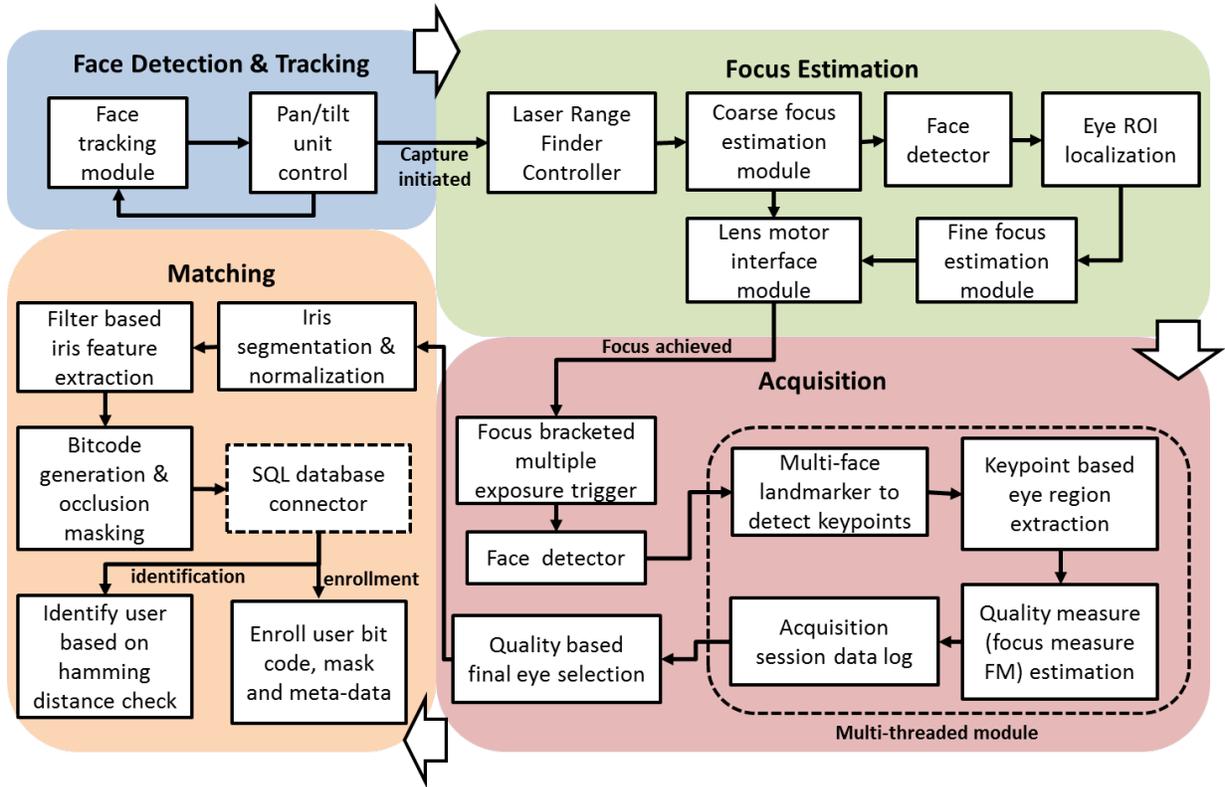


Figure 2.11: Major software modules that make up the proposed long range iris recognition system.

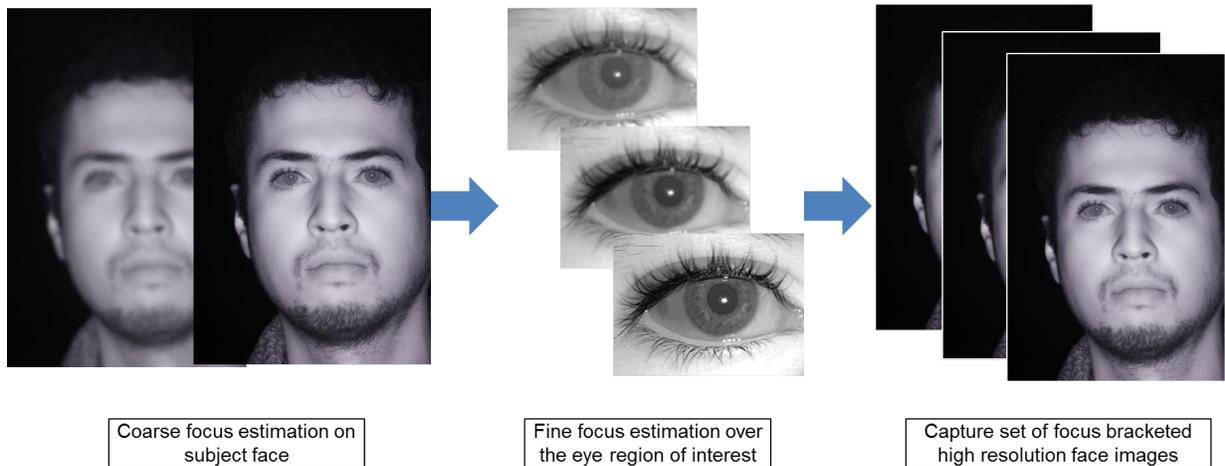
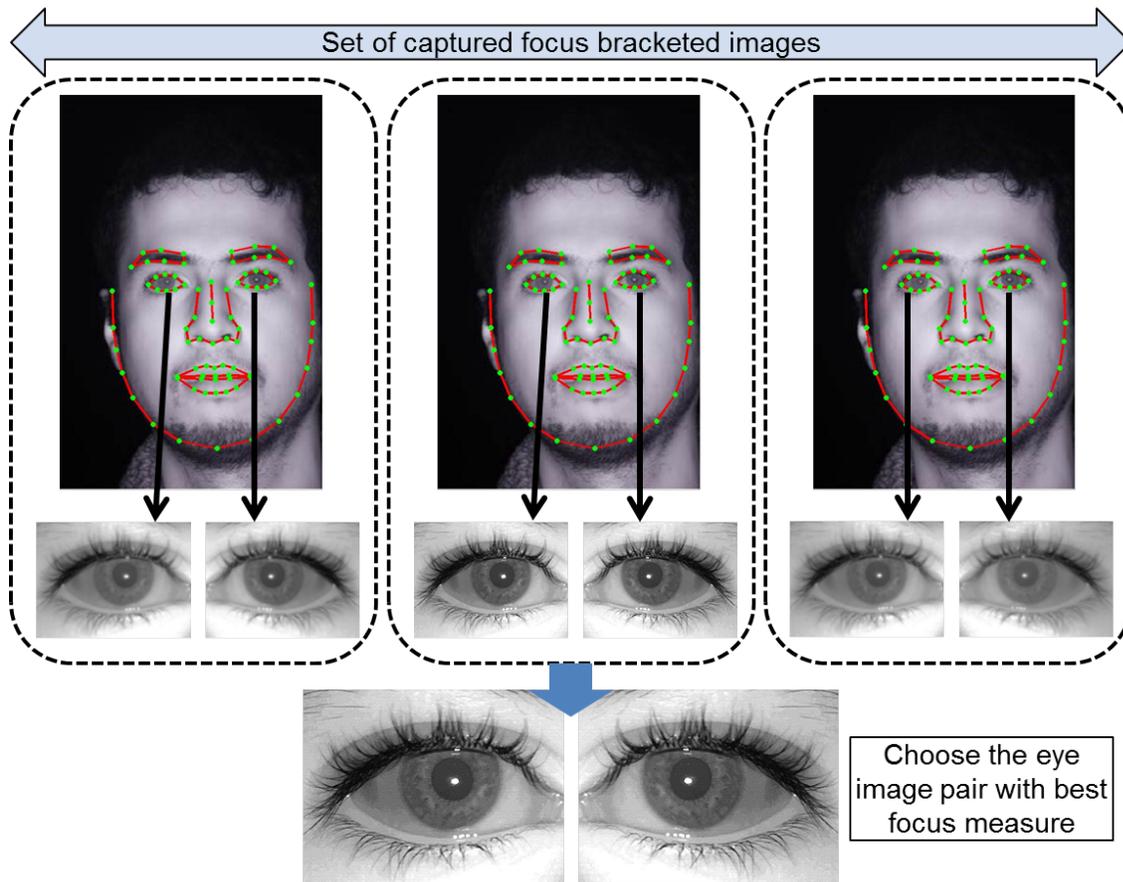


Figure 2.12: Focus estimation, prior to high-resolution face and dual eye acquisition. Following a coarse focus estimation over the subject's face, the region of interest (ROI) is set as the right eye. Fine focus, using a gradient descent-based strategy (section 2.5), is performed over the ROI. Finally a set of focus-bracketed high-resolution face images are captured.



22

Figure 2.13: Fiducial points on the face are estimated using the Modified Active Shape Model (MASM) approach [5]. The subset of points around each eye region are used to estimate the center of the corresponding eye. The eye regions with the highest focus measure (eqn. (2.8)), are used ultimately for iris feature extraction and matching.

face detection, the sensor's region of interest (ROI) is set as the left eye region, prior to continuing the focus estimation. By setting the ROI as the eye region, we make sure that focus is achieved over the eye, thereby minimizing the effect of a shallow depth of field. Fig. 2.12 illustrates this process.

3. **Iris Acquisition** - Once the focus is set, the camera captures a set of three face images in quick succession. Multiple images are captured to account for a subject blinking during acquisition or for any minor movements along the optical axis (which can result in the subject going in and out of focus). Following this, multiple threads are launched to process each face image without interrupting the main application thread. In each instance, the eye regions are extracted using a facial landmark detector, using modified active shape models (MASM) proposed by Seshadri *et al.* [5]; and assigned a focus value using eqn. (2.8) (see Fig. 2.13). All the data from each thread is logged in a sessions log, so that the operator can refer back to the images in the future if required. Finally, the best left and right eye images based on assigned focus measure values are handed over to the iris recognition module for further processing and subject identification.
4. **Iris Recognition** - Both the left and right eye images passed on by the iris acquisition module are then processed in the iris recognition engine. First the eye images are segmented, to locate the pupillary and limbic boundaries following which, the detected iris region is normalized and represented in polar co-ordinates. Feature extraction is done using an optimized Gabor filter, whose parameters are optimized to pick up characteristic texture features that can increase inter-class discrimination and decrease intra-class variation. Parameter optimization is done using a Markov chain Monte Carlo approach - simulated annealing. Further details regarding this parameter optimization approach are presented in section 2.8.3. The phase of the filter response is quantized to generate a bit code specific to that iris. Identifying

the subject in a database simply involves computing the hamming distance between this iris bit code and each bit code present in the database and determining the best match. Both the left and right eye bit codes are compared against the corresponding codes in the database. A subject is identified if his/her bit code reports minimum hamming distance, below a threshold value, when compared against the acquired irises' bit codes. A detailed description of this iris recognition engine and the components therein are presented next in section 2.8.

One of the advantages of our system is that it captures both the face image as well as the eye images using a single sensor. This opens up the possibility of adding a facial recognition engine on top of our iris recognition engine for multi-modal biometric authentication. As mentioned above, to localize features in the face, including the eyes, we use a facial landmark localization algorithm. This method is able to localize features even in images where significant portions of the face have been occluded. A few examples of faces captured in the visible spectrum, with occlusions, along with landmark localization are shown in Fig. 2.14. Even though not investigated as part of this thesis, the potential to use the proposed system for multi-modal biometric authentication opens up possibilities for application in even more unconstrained scenarios. One example could be to deploy the system to perform facial recognition, on subjects of interest, at very large distances ($200m$). When the subject is close to a sensitive region, iris based authentication may be performed, to accurately identify whether he/she is a target of interest. The iris recognition framework is explained next in this chapter.

2.8 Iris Recognition Engine

So far, we have considered the design challenges for an iris acquisition system and have outlined our design framework and related analyses. Following this, in this section we briefly outline the main components of our iris recognition engine - iris segmentation,

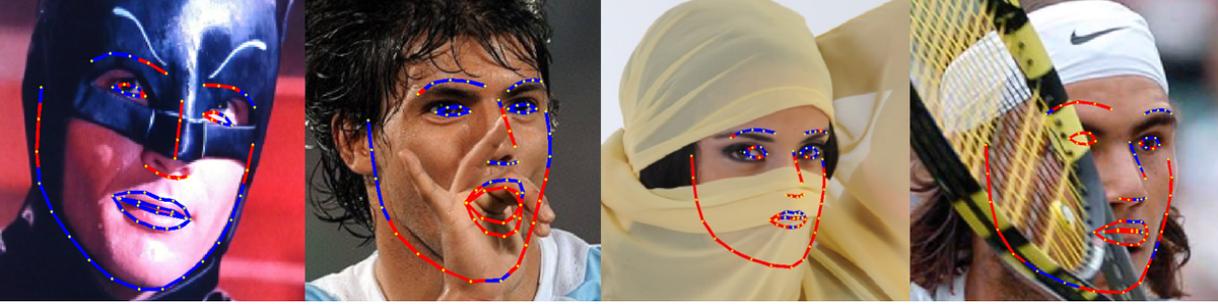


Figure 2.14: The figure shows examples of facial landmarks detected using the algorithm proposed by Seshadri et al. The images used are from COFW database[6]. This algorithm is able to localize significant features on the face inspite of the presence of occlusions such as masks, head scarves etc. In this thesis, this approach robustly localizes the eye regions. Additionally, since the proposed system captures a high resolution face along with eyes using a single sensor, a facial recognition layer can easily be added to the system for multi-modal (face+iris) biometric authentication.

feature extraction and template matching. The methods described in this section are used later in our iris verification and identification experiments to evaluate the performance of the proposed system.

2.8.1 Iris Segmentation

We follow a geometry-based method for determining pupillary and limbic boundaries in an eye image, similar to the method described in Li et al. [47]. Very briefly, this method involves the following stages:

2.8.1.1 Specular Reflection Removal

The first step is specular reflection removal in order to delineate the pupillary and limbic boundaries. These reflections are those caused by the reflection of the illumination source off the cornea. In our system, these occur within the pupillary region and hence can be removed by the following thresholding operation. If \mathbf{x}_{center} represents the center of the frame, we classify any point $I(\mathbf{x})$ as a specularity if $I(\mathbf{x}) > 0.9I_{max}$ & $\|\mathbf{x} - \mathbf{x}_{center}\| < r$. Here r is an empirically decided value (we use $r = 80$ pixels) and I_{max} is the maximum

intensity in the image. Following this, the regions are filled with the average pixel intensity from the neighborhood of the specularity. The whole image is smoothed using a median filter to remove any noise (see Fig. 2.16).

2.8.1.2 Locating the Pupillary Boundary

Since the pupil is the darkest region in the resulting image (Fig. 2.16(c)), it is localized using intensity thresholding. In order to eliminate false candidates such as eyebrows, eyelashes etc. we impose constraints on the height and width of the pupil region detected. These are empirically set to constant values. The center of the pupil is determined as the center of mass of the detected region. The pupil radius is the distance from this center to the furthest point in this region.

2.8.1.3 Locating the Limbic Boundary

The limbic boundary may be determined, using the determined pupil center as an initial point. This is done by drawing radiating lines from the pupil center and measuring gradients along these lines. The iris center is estimated by determining the point of intersection of perpendicular bisectors of lines joining all pairs of boundary points as depicted in Fig. 2.15. The radius is defined as the average distance from the center to all the boundary points. where, σ_x^2 and σ_y^2 represent the variances (the spread) of the Gabor function in the x and y directions respectively. λ represents the wavelength of the complex harmonic.

2.8.2 Iris Occlusion Masking

Once the segmented iris is converted to a normalized polar representation (see [48]) we mask out those regions occluded by eyebrows, eyelids, eyelashes etc. As in [49] Gaussian mixture models (GMM) are estimated for both occlusion pixels as well as iris pixels. A likelihood ratio can classify any given pixel as occlusion or iris based on these probability

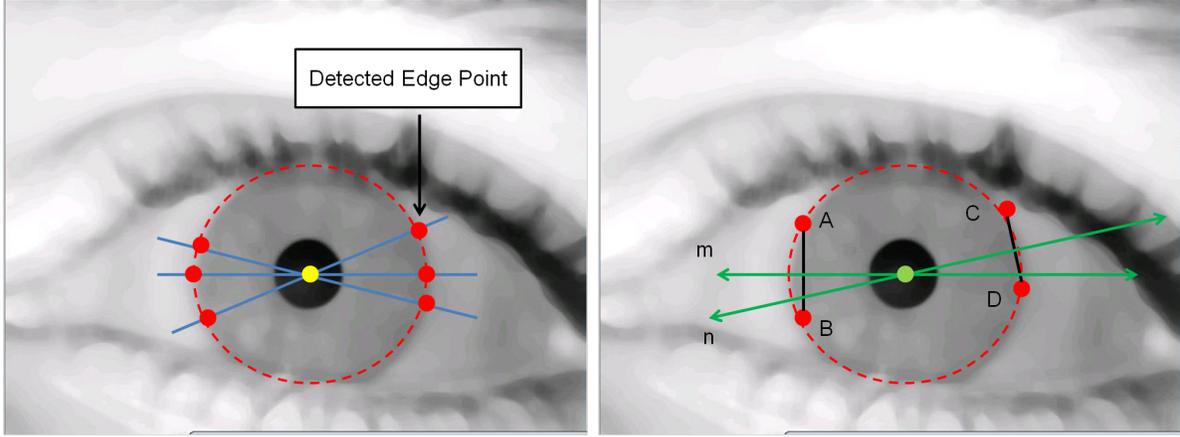


Figure 2.15: (a) shows the detected edge points when gradients are measured along lines radiating from pupil center, (b) shows the perpendicular bisectors m and n of lines AB and CD that join pairs of edge points. The point of intersection of such bisectors is taken as the limbic boundary center.

models. Given a pixel in a segmented, normalized iris image, the features used to train the GMM are - pixel intensity (I_{xy}), pixel location (x, y), mean (μ_{xy}) and standard deviation of intensity (σ_{xy}) in a 3×3 neighborhood. Let $\mathbf{x} = [I_{xy} \ x \ y \ \mu_{xy} \ \sigma_{xy}]^T$ be this feature vector at a pixel in the image. The probability mass function of this GMM is given by

$$P(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \mu_k, \Sigma_k) \quad (2.11)$$

assuming K Gaussians in the mixture are needed to fully represent the true feature distributions. μ_k, Σ_k are the parameters (mean and covariance) of the k th Gaussian in the mixture and π_k is the probability of occurrence of the k th Gaussian. The likelihood function and the corresponding log-likelihood function for N given texture pixels (say iris, for the purpose of this analysis) are:

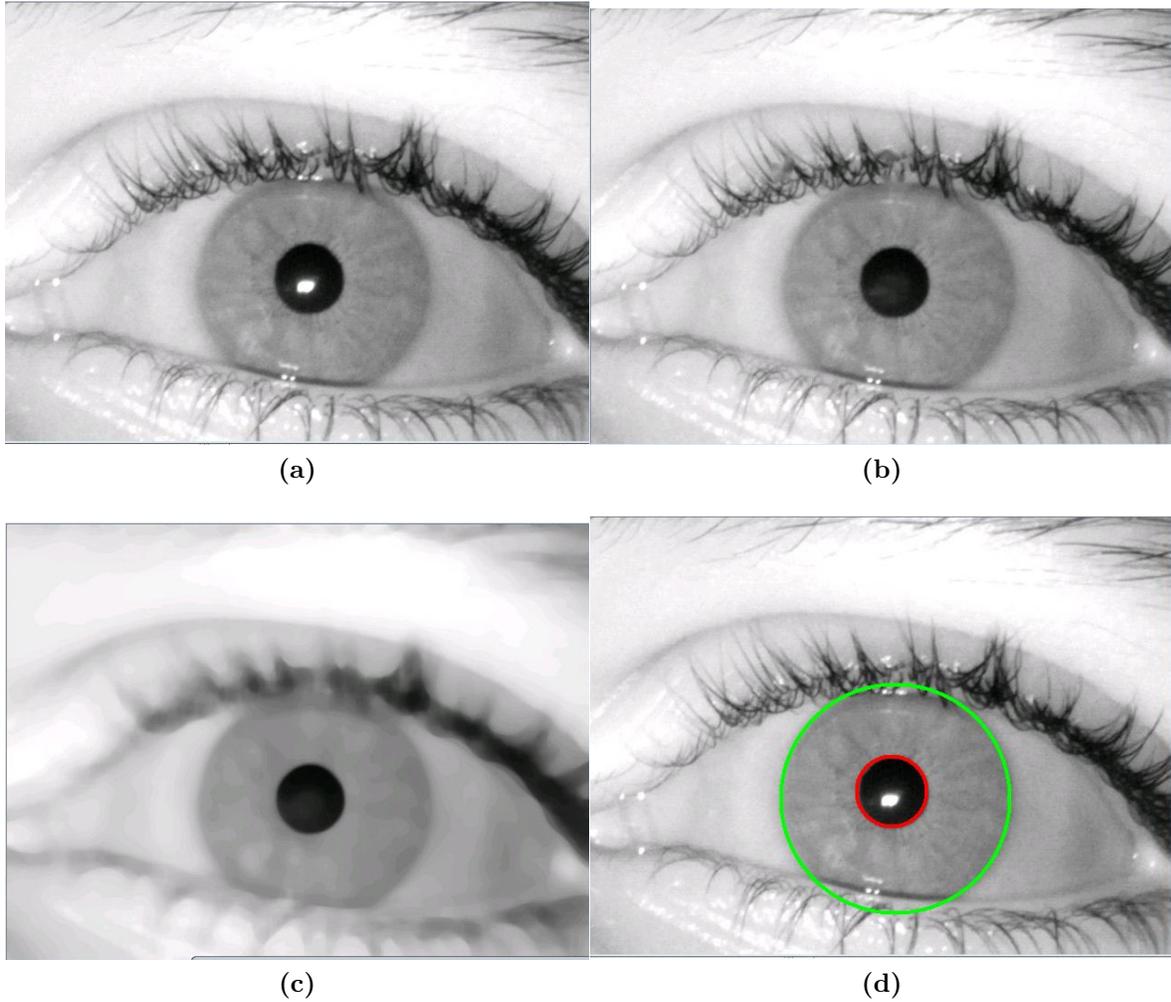


Figure 2.16: Various stages of the iris segmentation algorithm illustrated on an image acquired using our system. (a) shows the acquired eye image, (b) shows the image with specularities removed, (c) shows the median filtered image and (d) shows the detected iris boundaries.

$$\begin{aligned}
L(\mathbf{x}) &= \prod_{n=1}^N \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \mu_k, \Sigma_k) \\
l(\mathbf{x}) &= \sum_{n=1}^N \log \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \mu_k, \Sigma_k)
\end{aligned} \tag{2.12}$$

Let \mathbf{z}_k represent the k th Gaussian in the mixture with parameter set $\{\mu_k, \Sigma_k, \pi_k\}$. Using Bayes' rule, for a given pixel feature set \mathbf{x}_n , we have the posterior probability

$$P(\mathbf{z}_k | \mathbf{x}_n) = \frac{P(\mathbf{x}_n | \mathbf{z}_k) P(\mathbf{z}_k)}{P(\mathbf{x}_n)} \tag{2.13}$$

In order to get the maximum likelihood estimate for μ_k , we differentiate eqn. (2.12) w.r.t μ_k and set the derivative to zero. We get

$$\mu_k = \frac{\sum_{n=1}^N P(\mathbf{z}_k | \mathbf{x}_n) \mathbf{x}_n}{\sum_{n=1}^N P(\mathbf{z}_k | \mathbf{x}_n)} \tag{2.14}$$

Similarly, the expressions for Σ_k and π_k can be derived to be

$$\Sigma_k = \frac{\sum_{n=1}^N P(\mathbf{z}_k | \mathbf{x}_n) (\mathbf{x}_n - \mu_k) (\mathbf{x}_n - \mu_k)^T}{\sum_{n=1}^N P(\mathbf{z}_k | \mathbf{x}_n)} \tag{2.15}$$

$$\pi_k = \frac{\sum_{n=1}^N P(\mathbf{z}_k | \mathbf{x}_n)}{N} \tag{2.16}$$

An expectation maximization approach can be used to determine the parameter sets:

1. **E-step**-Using an initial set of values $\{\mu_1, \mu_2, \dots, \mu_k\}$, $\{\Sigma_1, \Sigma_2, \dots, \Sigma_k\}$ and $\{\pi_1, \pi_2, \dots, \pi_k\}$, calculate the probability given in eqn (2.13).
2. **M-step**-Using the calculated value from the E-step, new sets of values for the above parameters are calculated using eqns (2.14), (2.15) and (2.16).

These two steps are repeated till convergence. In this manner GMMs can be trained

to model both iris texture pixels and occlusion texture pixels. Given any pixel from the normalized iris image it can be classified as an iris pixel or an occlusion pixel using the following likelihood ratio test

$$\frac{L_{iris}(\mathbf{x})}{L_{occlusion}(\mathbf{x})} \begin{cases} > 1 & \text{iris pixel} \\ < 1 & \text{occlusion pixel} \end{cases} \quad (2.17)$$

where the suffix of the likelihood function $L((x))$ indicates the GMM model used.

2.8.3 Iris Feature Extraction and Matching

The feature extraction method used in our system, is a modified version of the method described by Daugman in [48]. This method projects the segmented and unwrapped iris image onto a set of complex valued 2D Gabor functions. The Gabor function $G(x, y)$ is a complex harmonic function windowed by a Gaussian function as shown below

$$G(x, y) = \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y} \exp \left[\frac{-1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + j \frac{2\pi}{\lambda} (x + y) \right] \quad (2.18)$$

The projection of the iris onto the Gabor basis is achieved by a simple 2D filtering operation of the iris texture using the Gabor functions. The parameters of the Gabor functions are optimized such that there is optimal inter-class discrimination during the recognition process. The filter responses are encoded and used for user identification/verification. The encoding of the extracted features involves a simple four level quantization operation applied to the phase of the filtered responses. In order to match two given iris codes thus generated, a hamming distance measure is used. Assume that we have to compute the hamming distance between two iris codes I_1 and I_2 . Let M_1 and M_2 be the associated iris masks generated as described in section 2.8.2. Then the hamming distance H between I_1 and I_2 is given by

$$H = \frac{\|(I_1 \otimes I_2) \cap M_1 \cap M_2\|}{\|M_1 \cap M_2\|} \quad (2.19)$$

Optimizing Gabor filter parameters

The Gabor filter parameters required to obtain very high user verification rates may be estimated by framing an optimization problem. For this purpose, we use the NIST Iris Challenge Evaluation Database [29]. We randomly choose a subset of this database to be the ‘gallery’ set and the remainder to be the ‘probe’ set. Given a Gabor filter G , we create ‘probe’ and ‘gallery’ iris bit codes. From the pair-wise hamming distance scores, we compute the genuine accept rates (GARs) as well as the false accept rates (FARs). The GAR at 0.1% FAR is chosen as the cost of using G . Our aim then is to choose that G which maximizes our cost. Let $\theta = \{\sigma_x, \sigma_y, \lambda\}$ be the parameter set that defines a G (see eqn.(2.18)). Then our optimization setup may be expressed as

$$\begin{aligned} \hat{\theta} &= \underset{\theta}{\operatorname{argmin}} f(\theta) \\ \text{subject to} \quad & \frac{I_{height}}{10} < \sigma_x, \sigma_y, \lambda < \frac{I_{height}}{4} \end{aligned} \quad (2.20)$$

where $f(\theta) = \frac{1}{N} \sum_{n=1}^N f_n(\theta)$. $f_n(\theta)$ represents one iteration of calculating the GAR value over the database for a randomly chosen probe and gallery set. We perform N such iterations and take the average as the final GAR value. I_{height} is the vertical dimension of the segmented, normalized iris texture. We obtain the solution to this problem by setting up a simulated annealing framework as discussed in [50]. This algorithm is a Markov Chain Monte Carlo (MCMC) based search of the parameter space. The following elements are used in this framework:

- The set $\Theta = \{\theta\}$ of all possible Gabor parameter sets θ

- A set $\Theta^* \subset \Theta$ of desired (global) maxima in $f(\theta)$ (since there may be more than one set θ with GAR values very close to each other and very close to the global maximum)
- Each Gabor parameter set $\theta_i \in \Theta$ is a state in a Markov chain. Let $\theta(i) = \Theta - \theta_i$ be the set of neighbors of θ_i . Let the probability of choosing a neighbor state θ_j be given by q_{ij} . In our implementation all states are equally probable. Hence all q_{ij} are equal.
- A non-increasing function in time $T(t)$ called the *temperature* at time t .
- An initial Gabor parameter set (initial state of the Markov chain) $\theta_{initial}$.
- The state $x(t)$ of the Markov chain at time t .

Starting from the initial state $x_0 = \theta_{initial}$, the evolution of the Markov chain with time t is based on the temperature $T(t)$. Using this approach, at equilibrium (i.e. when no state transitions occur and $T(t)$ is below a threshold) we will have obtained a θ that is very close to the global maxima. Convergence properties of this approach have been discussed in several works such as by Hajek [51] and Faigle *et al* [52]. Here we will only describe the procedure. Let state at time t , $x(t) = \theta_i$.

Choose a neighbor θ_j of current state θ_i at random, according to q_{ij} . Then $x(t+1)$ is chosen as follows

- If $f(\theta_j) > f(\theta_i)$ then $x(t+1) = \theta_j$
- If $f(\theta_j) < f(\theta_i)$ then

$$x(t+1) = \theta_j$$
with probability $P(x(t+1) = \theta_j | x(t) = \theta_i)$
else
$$x(t+1) = \theta_i$$

where,

$$P(x(t+1) = \theta_j | x(t) = \theta_i) = q_{ij} \exp \frac{-1}{T(t)} \max \{0, f(\theta_j) - f(\theta_i)\} \quad (2.21)$$

After every iteration, the temperature T is updated as $T(t+1) = \frac{T(t)}{1000}$. At the start of the ‘annealing’ process, T is high and the algorithm is ‘exploratory’ in nature. Hence, it does not get stuck in local maxima. As the algorithm proceeds, due to decreasing T , only those θ that report better GARs are accepted. When convergence is reached, $\theta \in \Theta^*$. In the remainder of this dissertation, the matching algorithm that uses the Gabor filter thus estimated, is referred to as the optimal Gabor filter algorithm. This method is used in chapters 3 and 4 to evaluate the efficacy of the long range system as a iris recognition tool as well as to evaluate the iris texture enhancement algorithms proposed. The next chapter describes a method that may be employed for de-noising as well as for enhancing the resolution of iris texture images, acquired using the system that was described in this chapter.

Chapter 3

Enhancing spatial resolution in iris texture images

As can be seen from sections 2.2 and 2.4, the spatial resolution of the acquired iris pattern, decreases as the person being imaged moves further away from the system. One of the additional tasks undertaken in this dissertation, is to address the issue of degraded iris images and explore a possible enhancement solution. Such images are often encountered in real world settings; a few examples acquired using the proposed system are shown in Fig. 3.1. The degradation may be due to low light acquisition - when there is lesser infra-red illumination available. Then owing to the lower quantum efficiency of the sensor at higher wavelengths, sensor noise can dominate the image. It can also be due to improper focus estimation prior to acquisition, due to atmospheric effects, shallow depth of field etc. This section explores the possibility of restoring such degraded iris textures, by solving a sparsity constrained linear regression problem.

Decrease in resolution will have a profound effect on the iris recognition performance, as we empirically show in section 4.2 and section 4.3. Factors such as noise due to low illumination acquisitions are also to blame for decreased performance at longer system-to-subject distances (or *stand-off distances*). To alleviate any drop in performance on

account of these factors, it is advantageous to have a stage in the iris recognition pipeline, that can enhance acquired images. This chapter visits the two important facets of image enhancement - de-noising as well as resolution enhancement (interchangeably called *super-resolution* in this text) - that are required in any long range iris acquisition system. The crux of the enhancement technique described in this section, is an inverse imaging problem solved using a dictionary based linear regression model. The inverse imaging model works using noisy, low-resolution texture that has been imaged by the long range system.

- **De-noising** - The proposed model can be used in both a de-noising context as well as in a resolution enhancement context. The de-noising algorithm is based on our observation that any given iris texture can be represented, with minimal reconstruction error, as a sparse linear combination of exemplar texture samples from an over-complete dictionary of textures \mathbf{D} . The over-complete dictionary is represented as a ‘fat matrix’ with more columns than rows i.e. more samples than dimensions. The columns of such a dictionary, are the exemplar textures and we refer to them as ‘atoms’. The number of rows depends on the dimensions of these atoms. The aim in the de-noising problem is to represent a given noisy texture patch, with a minimal number (sparse combination) of atoms. Since these atoms represent only iris textures and not noise, the intuition behind the method is that a sparse combination of atoms will represent only the essential texture variations and not noise. The method is discussed in section 3.2.
- **Super resolution** - The difference in the resolution enhancement method is that, it is based on generating a *coupled dictionary* model (two over-complete dictionaries instead of one), to map low resolution texture features to a higher resolution counterpart. In a nutshell, this means, one has to generate two representative iris texture dictionaries - a \mathbf{D}_l representing low-resolution patches and a \mathbf{D}_h representing the corresponding high resolution patches. The aim, in this approach, is simply to

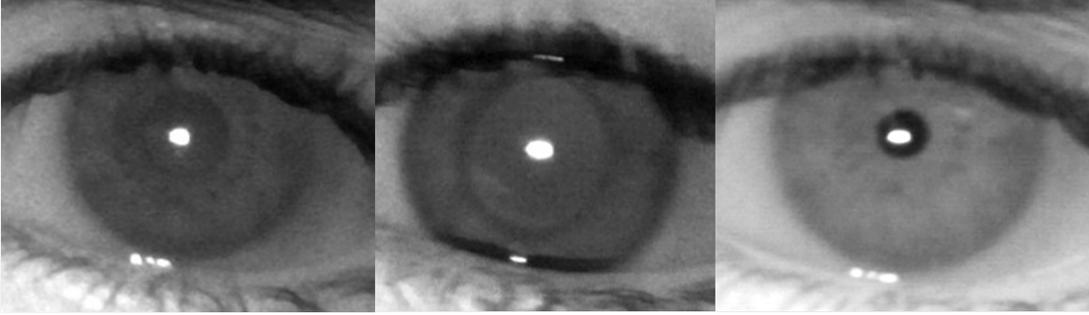


Figure 3.1: Examples of degraded iris images captured using the long range iris system. Some of the factors that can cause degradation - longer subject stand-off distance, inadequate illumination, inadequate focus.

determine a sparse linear combination of atoms in \mathbf{D}_l that best represents the observed low resolution texture patch and use the corresponding sparse coefficient set in \mathbf{D}_h , in order to estimate the corresponding high resolution patch. Preliminary work using this approach has been shown to work in the context of natural texture images by Yang *et al.* [53]. In addition to this approach, in this dissertation, a novel method of using class-specific gradient constraints, is also explored, when solving for the linear regression coefficients; we find that these constraints enhance particular class-specific texture features, that specifically help in improving iris recognition performance when using low resolution images. The resolution enhancement method is discussed in section 3.3.

To keep the approach general, the dictionaries used in this dissertation are built using publicly available iris databases and subsequently used in de-noising or super-resolving images captured using the proposed long range iris system; specifically, for our work we have used texture patches from right eye images from the Iris Challenge Evaluation (ICE) database that was released by the National Institute of Standards and Technology (NIST). We illustrate all the methods in this section using left eye images from the same database. Results of using these methods on images acquired with the proposed long range system are reported in the next chapter. Prior to describing the de-noising and resolution enhancement methods, the next section discusses over-complete dictionaries containing exemplar iris

texture ‘atoms’ and how any given iris texture patch may be represented as a sparse linear combination of these atoms.

3.1 Sparse reconstruction in an over-complete dictionary

It is a well known fact that, a high dimensional naturally occurring texture has a sparse representation in several orthogonal bases. These bases can be the DCT bases, wavelet bases or data dependent bases obtained using principal component analysis (PCA). This in fact forms the basis of several image compression schemes, one example of which is the well known JPEG2000 [54] format. Several image denoising methods such as [55][56][57] have also successfully made use of the implicit sparsity of images in a known basis. Under certain conditions, it is also known that the texture can be approximated with minimal error, by a sparse linear combination of atoms in over-complete dictionaries, with redundant data along the columns. This has been investigated in seminal works such as by Donoho *et al.* [58], Tropp [59], Fuchs [60]. For the purpose of this discussion, consider an iris texture \mathbf{x} , in vectorized form, such that $\mathbf{x} \in \mathbb{R}^M$; let the over-complete dictionary of choice be $\mathbf{D} \in \mathbb{R}^{MN}$, with $M \ll N$. If ϕ represents the vector of synthesis coefficients in \mathbf{D} for the image \mathbf{x} , the texture representation problem can be formulated as

$$\hat{\phi} = \arg \min_{\phi} \|\mathbf{x} - \mathbf{D}\phi\|_2^2 \text{ s.t. } \|\phi\|_0 < L \quad (3.1)$$

where L is the sparsity constraint on this under-determined coefficient estimation problem. Since the ℓ_0 -norm is discontinuous (hence non-differentiable), a gradient based minimization method is not possible in this case. Instead, we use a popular greedy approach - orthogonal matching pursuit (OMP) - to estimate the best set of coefficients $\hat{\phi}$ that satisfy the norm constraint. Before describing the enhancement algorithms in this section, we

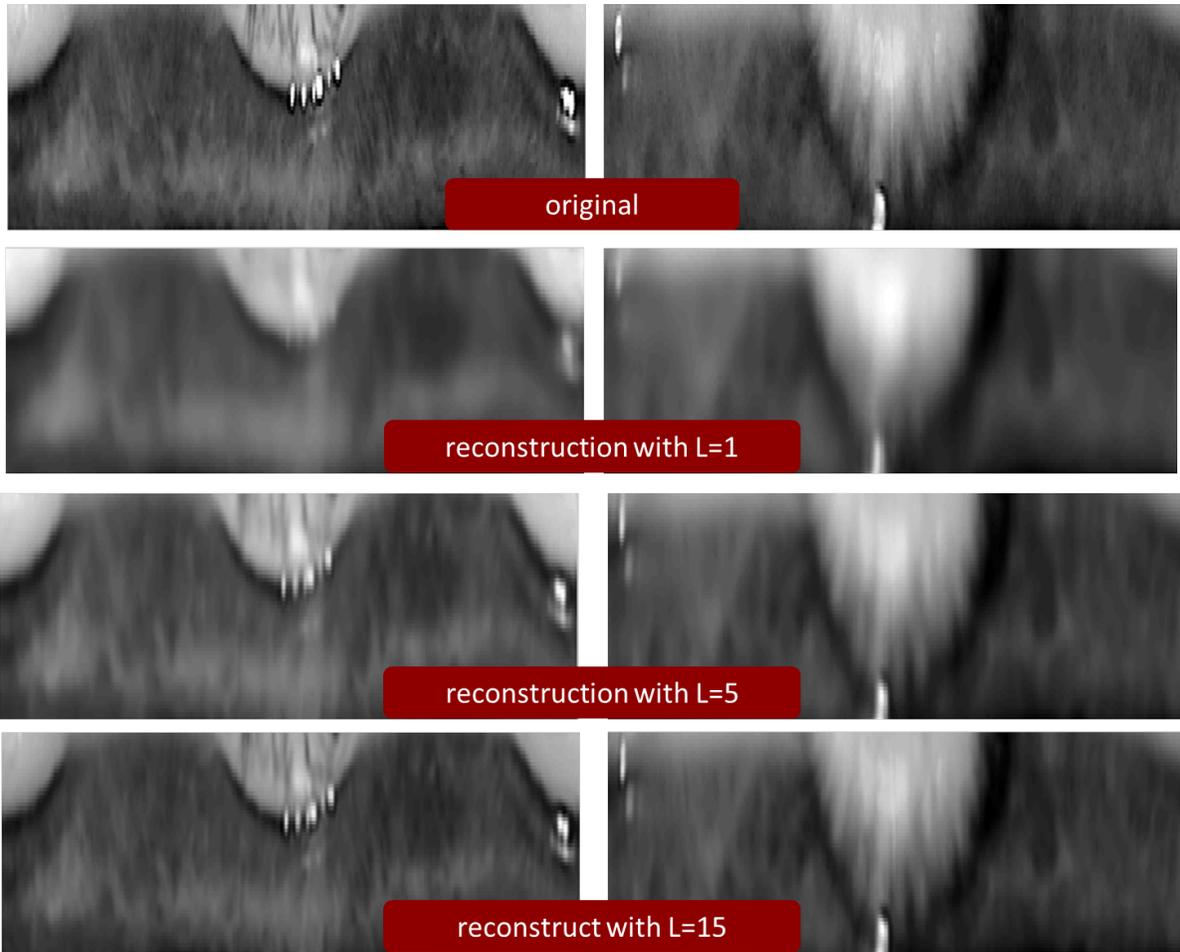


Figure 3.2: Examples of segmented normalized iris images reconstructed with varying levels of sparsity in an over-complete dictionary (see eqn. (3.1)). With increasing L , the reconstruction uses more atoms and hence we see a better visual reconstruction.

experimentally show that any iris texture can be represented with minimum reconstruction error, using a sparse linear combination of representative atoms in an over-complete dictionary. For the purpose of this experiment we build our dictionary \mathbf{D} of atoms using textures present in the segmented normalized right eye image set within the NIST Iris Challenge Evaluation database[29] (the segmentation and normalization method used in our work is outlined in appendix 2.8. Texture patches of size 20×20 are used in this work. The dictionary is built using the popular KSVD approach proposed by Aharon *et al.* in [61], with a sparsity level of $L = 15$ in eqn (3.1). KSVD, as described in [61] is

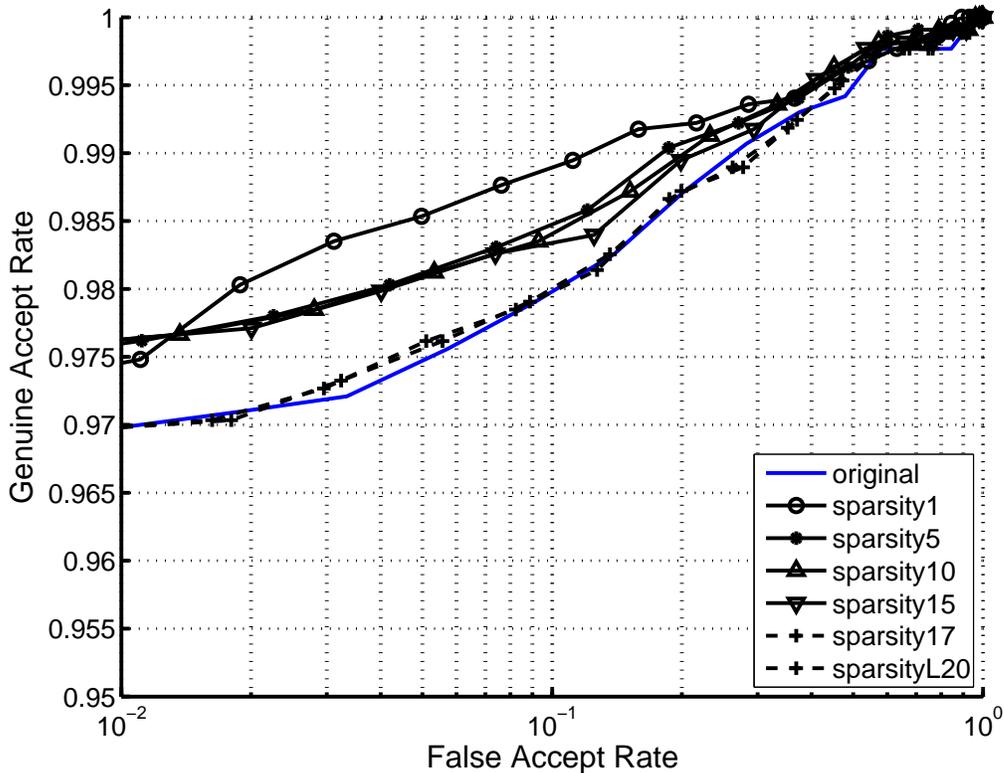


Figure 3.3: ROC curves comparing verification performance when using the original left eye images from the ICE database in the probe set (the blue curve) with the cases when reconstructed images are used in the probe set. The reconstruction is done using the over-complete dictionary in eqn. (3.1), with varying levels of sparsity L .

a generalized K-means algorithm. The K-means algorithms clusters data such that any data sample can be associated with one of K cluster centers. The KSVD algorithm on the other hand, allows any data sample to be represented as a sparse linear combination of these cluster centers. The dictionary of cluster centers, or ‘atoms’ as they are referred to, is generated starting with a representative set of exemplar data points. The algorithm iteratively determines the best K atoms for this purpose.

The left eye images are used to test the efficacy of sparse representation. We reconstruct every image in the left eye dataset using sparsity levels of $L = 1$, $L = 5$, $L = 10$ and $L = 15$. The reconstruction of each normalized left eye iris image is done patch by patch, using a set of overlapping 20×20 patches from the input image. Given a patch \mathbf{x} from an input

image, eqn. 3.1 is solved for a fixed value of L using OMP. The final reconstructed image is a composite of the individual patch reconstructions. The composite is built by alpha-blending the patch reconstructions. Fig. 3.2 shows examples of iris image reconstructions for various values of L . In addition, we use the dataset generated using various sparsity values in an iris verification experiment. The corresponding receiver operating characteristic (ROC) curves are shown in Fig. 3.3. These curves were generated using the original left eye images as gallery and the reconstructed images as probe. We see that even for very low levels of sparsity L , the reconstructed images hold sufficient class-specific discriminating information, such that they are able to be effectively used in iris recognition experiments. In fact, for low sparsity values, the verification performance is better than when using the original images in the probe set (represented by the blue curve). This is because, the sparse reconstruction in this dictionary has a de-noising effect on the data. The next section deals with this topic and shows, that this de-noising effect is advantageous especially in those cases when a large amount of noise is present in the data.

3.2 Proposed iris texture de-noising framework

One of the major challenges of long range iris acquisition systems, is the need for sufficient near infra-red illumination during imaging. Four near IR LED panels are used in the proposed system, to make sure that sufficient illumination is available. However, if a system designer is constrained by form factor size, or by power requirements, he/she may have to use a smaller source, which can result in noise in the acquired images. Under low light conditions, the primary sources of noise are electronic circuit noise and image sensor noise[62]. This noise can be approximated as additive white gaussian noise with zero mean and varying variance values - a higher variance depicting more noise. Examples of simulated noisy images from the ICE database are shown in Fig. 3.4 and Fig. 3.5. Algorithm 2 outlines how we can use the sparsity of iris textures to denoise a given noisy input image. Here

Algorithm 2 Algorithm to de-noise iris images

Input:Normalized iris image \mathbf{I} **Process:**Divide \mathbf{I} into K overlapping patches $\{\mathbf{x}_i | i = 1, 2, \dots, K\}$ **for** $i = 1, 2, \dots, K$ **do**

Solve

$$\hat{\phi}_i = \arg \min_{\phi} \|\mathbf{x}_i - \mathbf{D}\phi\|_2^2 \text{ s.t. } \|\phi\|_0 < L \quad // \text{ each } \mathbf{x}_i \text{ in this expression expressed in}$$

vector form

$$\text{Set } \hat{\mathbf{x}}_i = \mathbf{D}\hat{\phi}$$

end forInitialize $\hat{\mathbf{I}}$ as a matrix of zeroes.**for** $i = 1, 2, \dots, K$ **do**

$$\text{Set } \hat{\mathbf{I}} = \alpha \hat{\mathbf{x}}_i + (1 - \alpha) \hat{\mathbf{I}} \quad // \text{ before use in this expression each } \mathbf{x}_i \text{ reshaped into a}$$
$$2\text{-D patch}$$

end for**Output:**Return $\hat{\mathbf{I}}$ as the reconstructed normalized iris image; reconstructed with sparsity level L

again, we use the over-complete dictionary built using the right eye images from the ICE data set. Following this, we attempt to de-noise the left eye images from the data set. The intuition here is that, the dictionary is built using good quality (not noisy) iris texture patches from right eye images and hence it learns what natural iris texture looks like from this right eye ‘training’ set. It is not built to model noise. Hence, given a previously unseen iris texture patch, by enforcing sparsity during coefficient estimation, we force the reconstruction to resemble an iris texture and not the embedded noise. Examples of de-noised images using this algorithm with $L = 1$ and $L = 5$, for two different levels of noise are shown in Fig. 3.4 and Fig. 3.5. We see that, with increasing levels of sparsity, the reconstruction paradigm tries to model the high frequency noise variations as well. Hence we keep the sparsity levels low in our experiments with the de-noising algorithm. Using various levels of de-noised images, the iris verification experiment from the previous section is repeated. The corresponding ROC curves are shown in Fig. 3.6. We see from the figure that, up to a certain level of noise, the use of de-noised images using the proposed de-

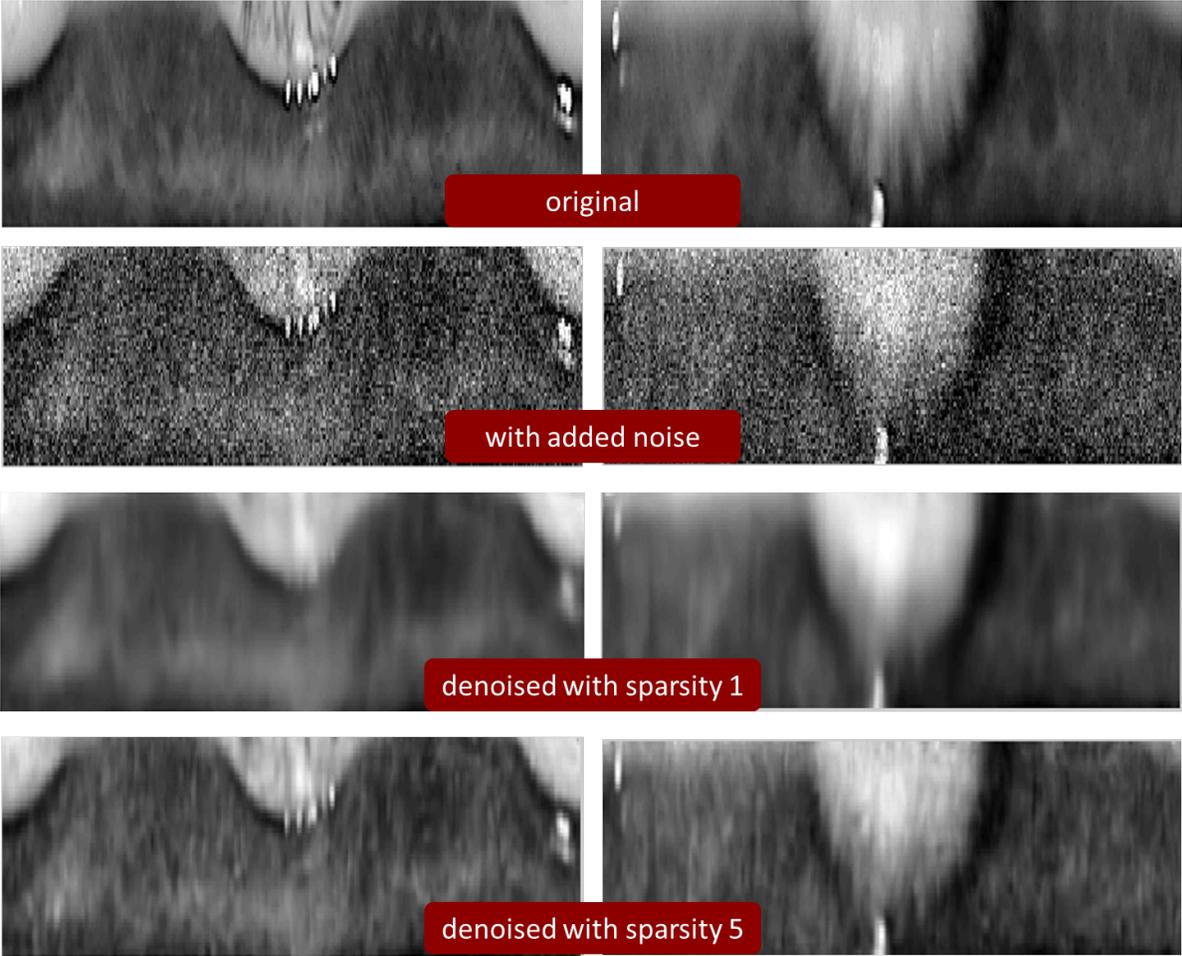


Figure 3.4: Two examples of image denoising using the proposed method. The noise added is white gaussian noise with zero mean and variance 0.02.

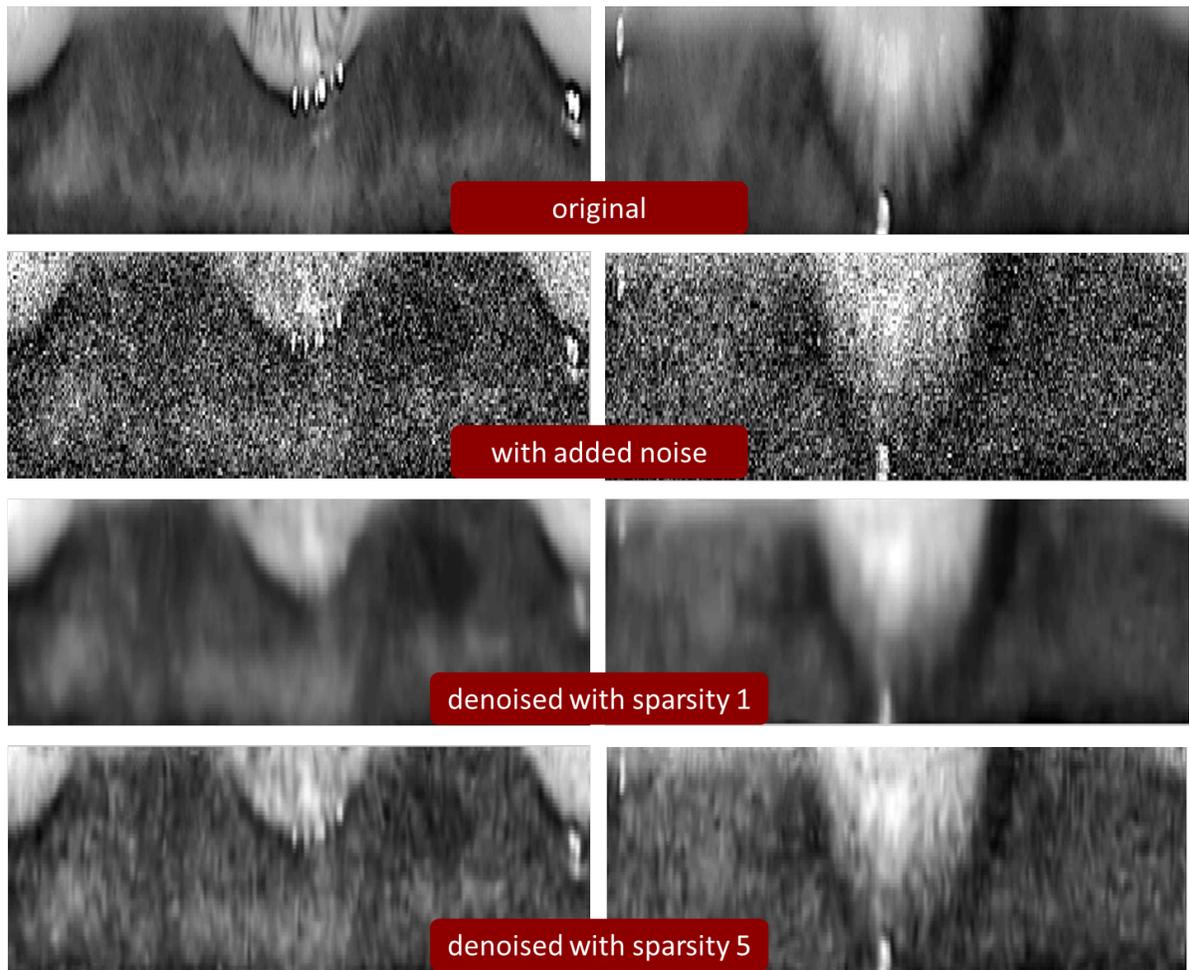


Figure 3.5: Two examples of image denoising using the proposed method. The noise added is white gaussian noise with zero mean and variance 0.05.

noising algorithm reports a better genuine accept rate (GAR) for all values of false accept rate (FAR). However, as the level of noise in the image increases, the sparse approximation framework will not be able to accurately model the underlying iris texture variations. Hence we see hardly any improvement in performance in Fig. 3.7b. In section 4.2, the same verification experiment is repeated using images captured using the proposed long range iris acquisition system.

3.3 Proposed spatial resolution enhancement framework

As a prologue to the discussion on resolution enhancement, we would like to remind the reader of the relationship between a high resolution image \mathbf{I}_h and its low resolution counterpart \mathbf{I}_l . The function that maps \mathbf{I}_h to \mathbf{I}_l can be modeled using the following linear relationship

$$\mathbf{I}_l = \mathbf{H}\mathbf{I}_h + \mathbf{n} \quad (3.2)$$

\mathbf{n} represents any noise present in the image. For the sake of this discussion, the primary source of noise is thermal noise and sensor noise, from the imaging sensor, when illumination intensity is low. The degradation matrix \mathbf{H} can be factorized as $\mathbf{H} = \mathbf{D}\mathbf{S}$ where \mathbf{D} models the loss in spatial resolution due to increased target distance (reduced magnification), and \mathbf{S} is a convolution matrix, representing the convolution of the original image with the point spread function (PSF) of the optical system. These terms correspond to the factors that affect spatial resolution, discussed in section 2.2. \mathbf{S} , for a diffraction-limited system, is the lens aperture function; which in its functional form, is expressed using a Bessel function of the first kind

$$S(\theta) = S_0 \left(\frac{2J_1(kr \sin \theta)}{kr \sin \theta} \right)^2 \quad (3.3)$$

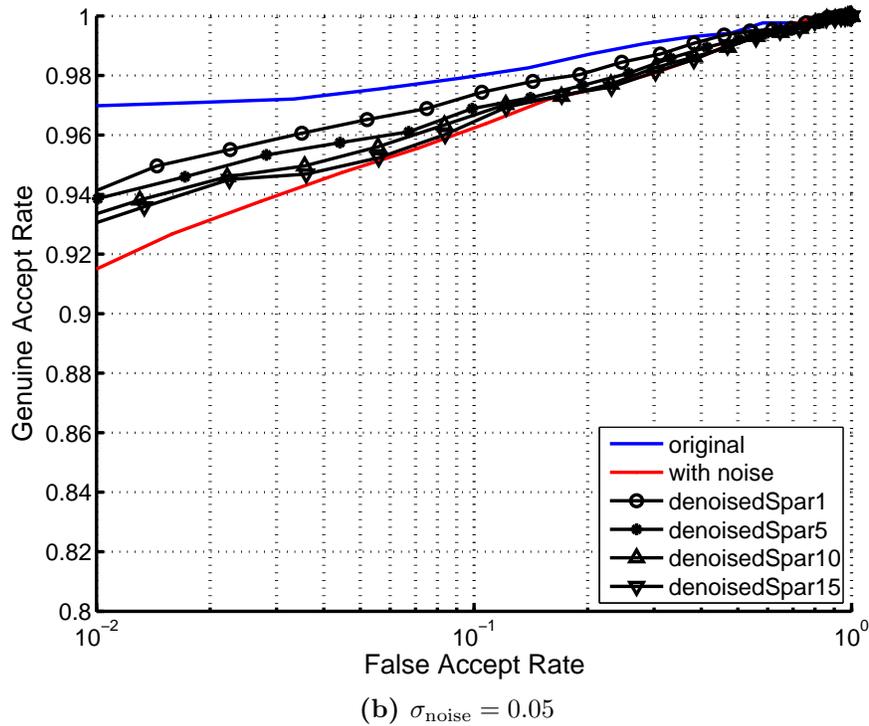
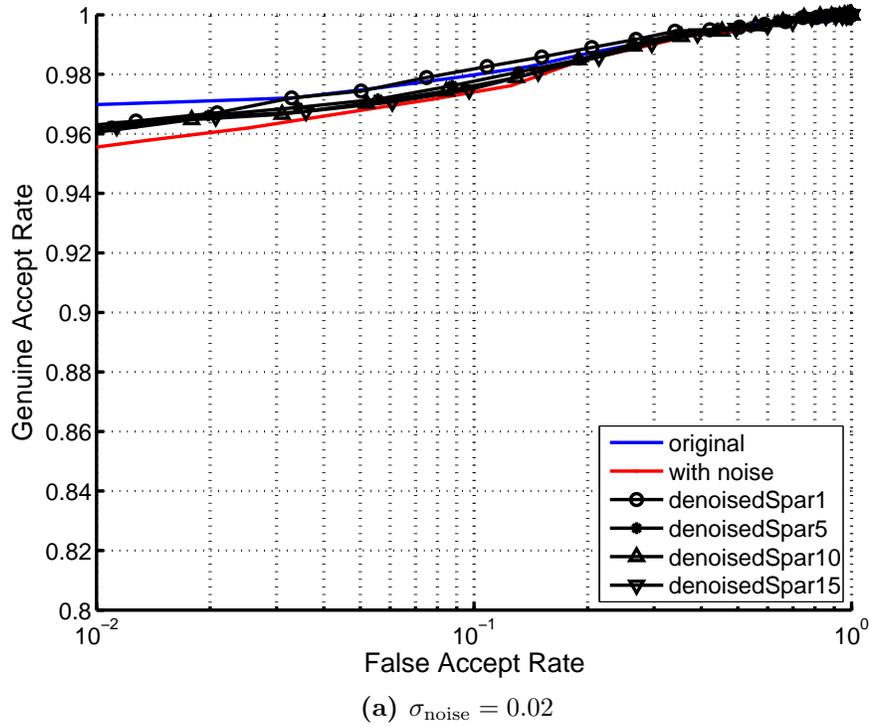


Figure 3.6: ROC curves depicting iris verification performance using de-noised images from the ICE database. Noise is simulated in the probe image set for the experiment using additive white gaussian noise with varying levels of variance.

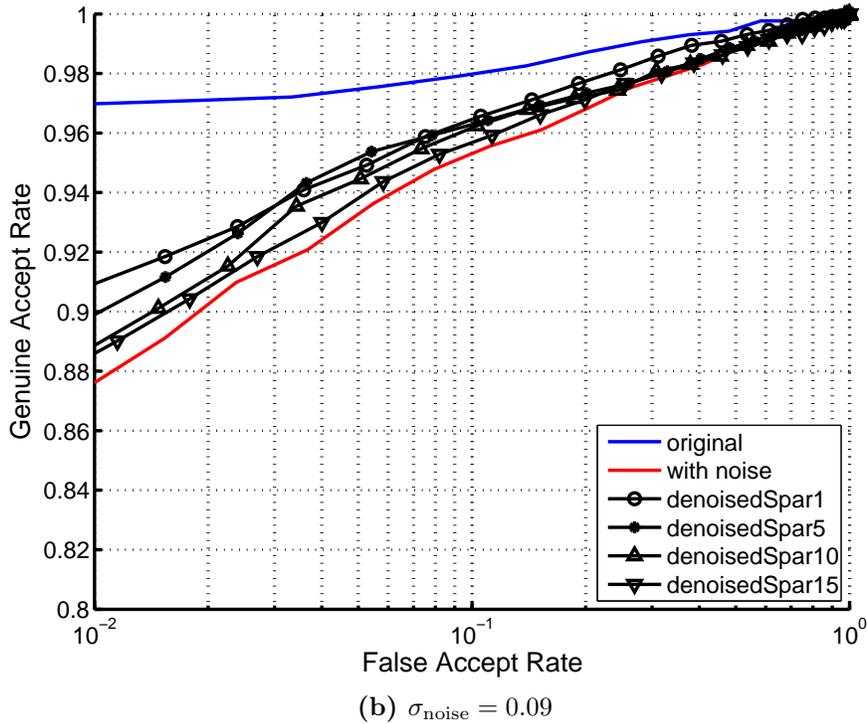
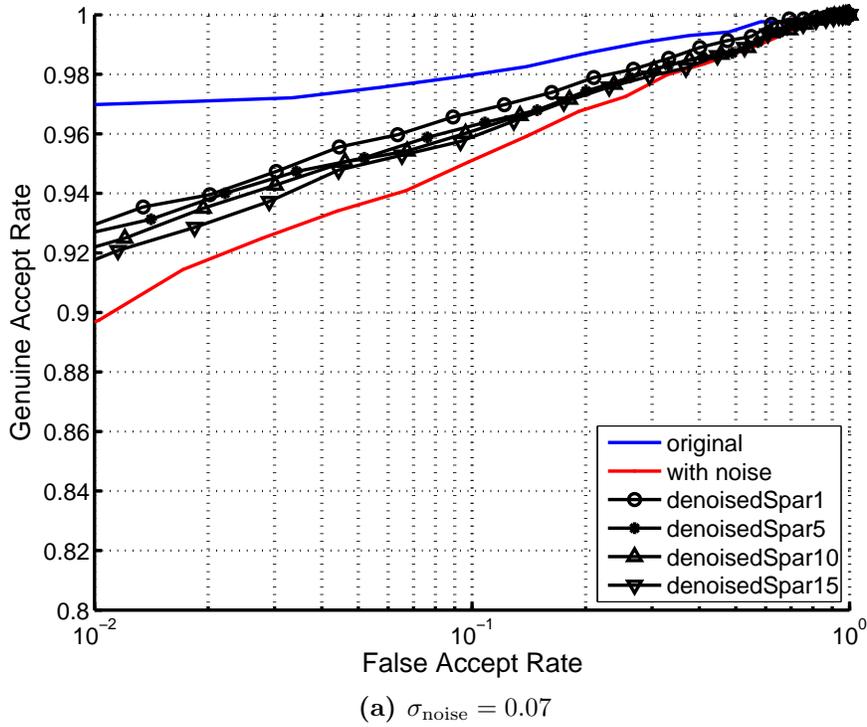


Figure 3.7: ROC curves depicting iris verification performance using de-noised images from the ICE database. Noise is simulated in the probe image set for the experiment using additive white gaussian noise with varying levels of variance.

This functional form is for a circular aperture, which is the case for the lens, with fully open aperture, used in this work. θ represents the angular separation of a point on the image plane from the optic axis; r is the radius of the aperture and $k = \frac{2\pi}{\lambda}$ is the wave number. As defined in section 2.2.2, $\lambda = 850nm$ is the wavelength at which the target is illuminated. Fig. 3.8 shows examples of the segmented and normalized iris textures \mathbf{I}_h and \mathbf{I}_l , from distances of $9m$ and $11m$ respectively. If there is an additional factor affecting resolution, say lens blur \mathbf{B} , then this can also be included in the degradation matrix as $\mathbf{H} = \mathbf{DSB}$. In this dissertation, as a first approximation, the effect of \mathbf{H} is modelled using a Gaussian kernel. The reason for this is that, in most cases, the ‘softening’ of images at larger distances is due to diffraction as described in section 2.2.2. This diffraction pattern (referred to as an *airy disc*), given by the lens aperture function (3.3), is very close to a squared sinc function. For a maximally open lens aperture, the side lobes of this function may be ignored and the main lobe can be approximated using a Gaussian kernel. To account for other effects such as de-focus, we simply increase the variance of the Gaussian kernel, to simulate a further decrease in resolution.

In this thesis, we propose to recover \mathbf{I}_h from \mathbf{I}_l by using a combination of exemplar texture sampling and sparse linear combinations of these exemplars. For the purpose of the discussion in this section, we ignore the presence of noise \mathbf{n} in eqn. (3.2), by assuming a de-noised image using the method outlined in section 3.2.

A high resolution counterpart $\hat{\mathbf{I}}_h$ for an observed \mathbf{I}_l is generated by

- breaking down \mathbf{I}_l into blocks (or patches; these two terms are used interchangeably in this thesis) $\mathbf{x}_l^i = \mathbf{P}_i \mathbf{v}_{\mathbf{I}_l} \in \mathbb{R}^W$; W being the dimension of the exemplars. Here $\mathbf{v}_{\mathbf{I}_l}$ is the 2-D image \mathbf{I}_l expressed in a vectorized form.
- estimating the high resolution counterpart block \mathbf{x}_h^i
- alpha-blending the resultant blocks to create the final composite $\hat{\mathbf{I}}_h$

Here, \mathbf{P}_i is a ‘patch selection’ matrix to select the i th image patch from image \mathbf{I} . The

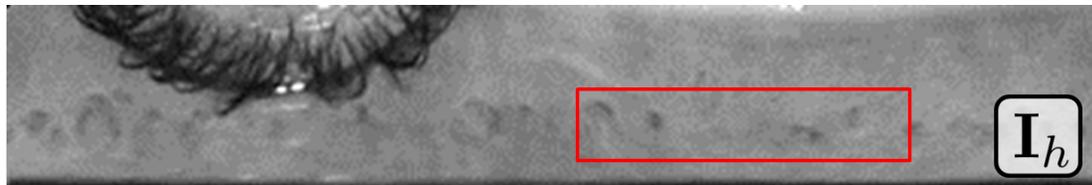


Image at 9 meters

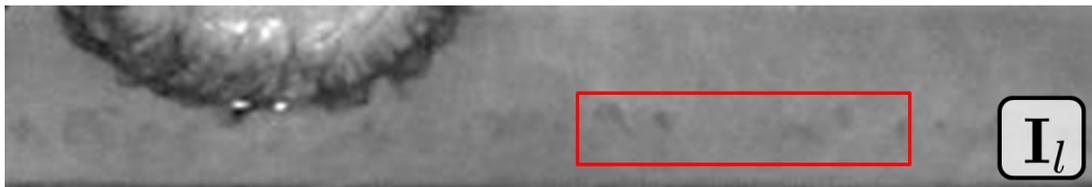


Image at 11 meters

(a)

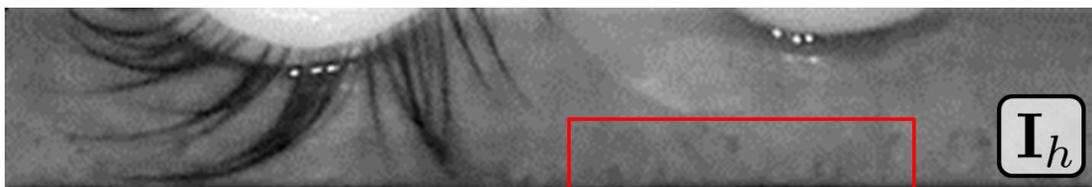


Image at 9 meters



Image at 11 meters

(b)

Figure 3.8: Example of normalized iris textures from our database captured at 9m and 11m; (a) and (b) show images from two different participants in our study. Our aim is to increase the spatial resolution in I_l to that in I_h , in order to improve iris recognition scores. On visually inspecting the areas outlined in red, one can see the reduced resolution in the I_l images

patch size used in this work is 20×20 .

3.3.1 Generating a coupled texture dictionary model with exemplar samples

The first step in our resolution enhancement algorithm is to populate a ‘texture dictionary’ with exemplar texture samples from available iris images. In this work, we do this by randomly sampling texture patches from iris images in the NIST ICE database [29]. We create two such dictionaries - \mathbf{D}_h , containing high resolution texture samples and \mathbf{D}_l , containing the corresponding low resolution counterparts. Both these dictionaries are over-complete i.e. the number of columns outnumbers the number of rows. Our super-resolution paradigm is based on the ability to reconstruct any given patch and its low resolution counterpart using the same set of sparse regression coefficients, in these over-complete dictionaries. We enforce a sparsity constraint to solve an under-determined problem, i.e. coefficient estimation in an over-complete texture dictionary. We observed that a sparse representation of any unseen iris texture patch in such an optimized dictionary, is able to capture all the minute variations, thus resulting in very low reconstruction errors.

The exemplar patches from ICE, expressed in vector form, are listed as the columns of a high resolution texture matrix \mathbf{X}_h . Following this, in order to emulate the loss of spatial resolution and thus to obtain a counterpart low resolution texture matrix, we express $\mathbf{X}_l = \mathbf{H}\mathbf{X}_h$. \mathbf{H} is approximated using a Gaussian kernel for the purpose of this study, in order to simulate loss of resolution with distance. In order to determine a \mathbf{D}_h and a \mathbf{D}_l that, using the same set of sparse coefficients, can minimize the reconstruction error of

patches in both \mathbf{X}_h and \mathbf{X}_l we pose the following optimization problem

$$\begin{aligned} \{\mathbf{D}_l, \mathbf{D}_h\} &= \arg \min_{\mathbf{D}_l, \mathbf{D}_h, \phi} \|\mathbf{X}_h - \mathbf{D}_h \phi\|_2^2 + \beta \|\mathbf{X}_l - \mathbf{D}_l \phi\|_2^2 \\ s.t. \quad &\|\phi\|_0 < L \\ &\beta > 1 \end{aligned} \quad (3.4)$$

On solving this optimization problem, for a fixed sparsity level L , the resulting dictionaries will be able to reconstruct, with minimum error, any given low resolution patch and its high resolution counterpart with the same set of sparse coefficients ϕ . Additionally, the weighting factor $\beta > 1$ is included to give more importance to the low resolution image. This is because, our aim is to eventually use our observation, i.e. a low resolution image, and super-resolve it to its high resolution counterpart. β ensures that more weight is given to the reconstruction error of our observation. In order to implement and solve eqn. (3.4), we use an excellent dictionary estimation toolbox available here[63].

3.3.2 Super-resolving low resolution texture

Once the dictionaries \mathbf{D}_h and \mathbf{D}_l have been determined, we can estimate the high resolution counterpart of a previously unseen low resolution patch \mathbf{x}_l , by solving a sparsity constrained ℓ_2 norm minimization problem similar to eqn. (3.4). Specifically, given a low resolution texture patch $\mathbf{P}_i \mathbf{I}_l = \mathbf{x}_{li}$, from any low resolution image \mathbf{I}_l , we first estimate its sparse representation in \mathbf{D}_l

$$\hat{\phi}_i = \arg \min_{\phi} \|\mathbf{x}_{li} - \mathbf{D}_l \phi\|_2^2 \quad s.t. \quad \|\phi\|_0 < L \quad (3.5)$$

Since eqn. (3.4) is designed such that $\hat{\phi}$ minimizes the reconstruction error of both a low resolution patch and its high resolution counterpart, the high resolution (or ‘super-resolved’) \mathbf{x}_{hi} is determined as $\mathbf{x}_{hi} = \mathbf{D}_h \hat{\phi}_i$. All such estimated \mathbf{x}_{hi} are blended together

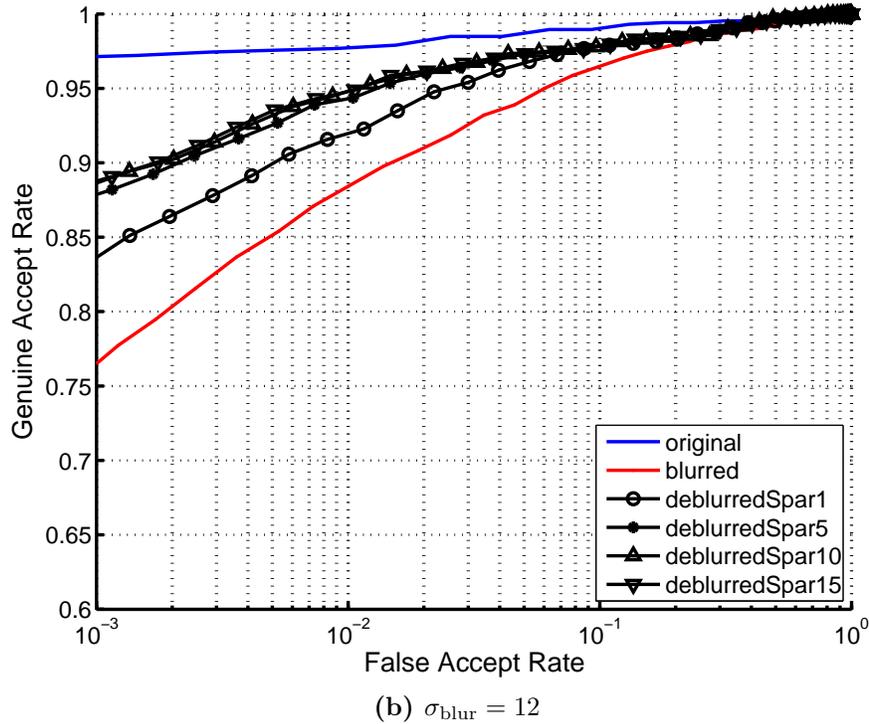
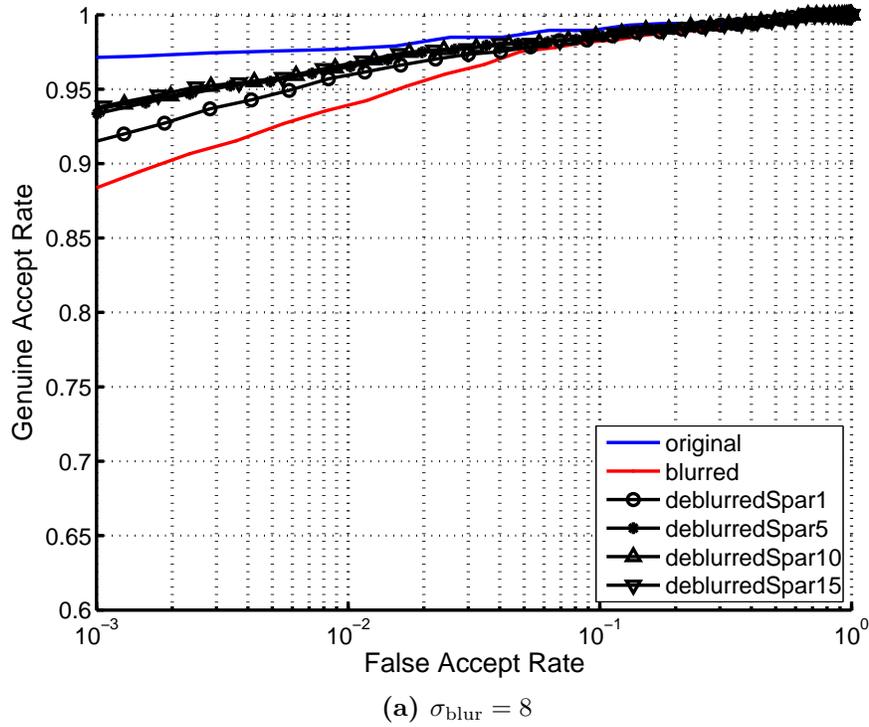
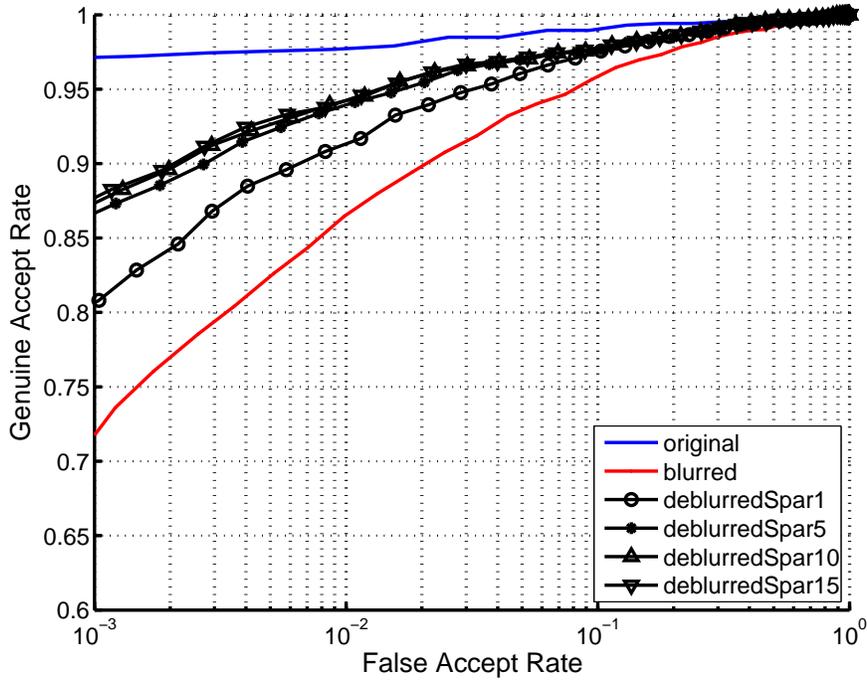
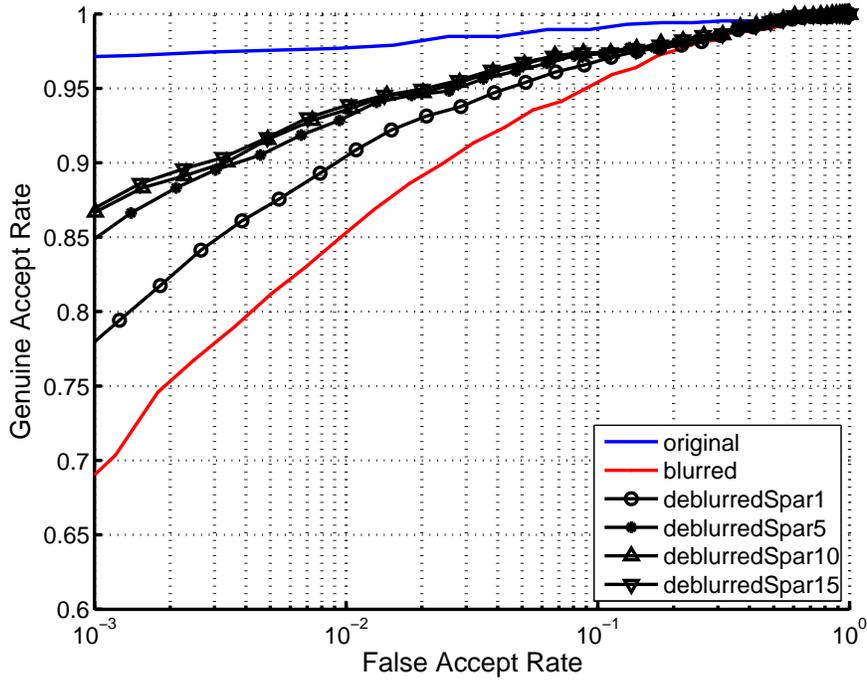


Figure 3.9: ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 20×20 Gaussian kernel with increasing values of variance (gaussian spread).



(a) $\sigma_{\text{blur}} = 16$



(b) $\sigma_{\text{blur}} = 20$

Figure 3.10: ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 20×20 Gaussian kernel with increasing values of variance (gaussian spread).

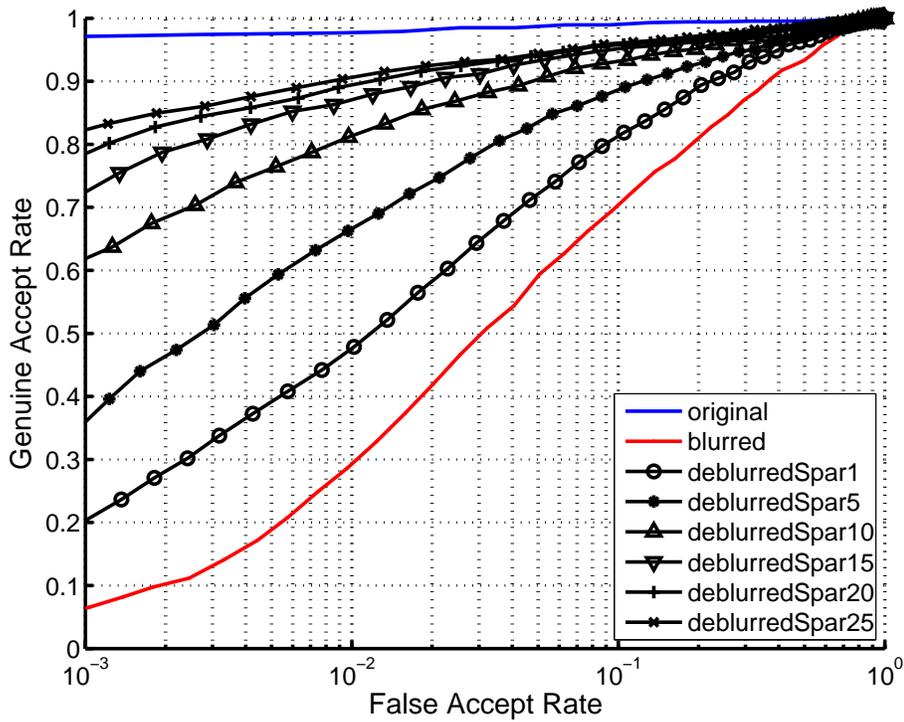


Figure 3.11: ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 30×30 Gaussian kernel with $\sigma_{blur} = 20$.

to obtain the final high resolution image \mathbf{I}_h . Fig. 3.12 and Fig. 3.13 show examples of left eye images from the ICE database that have been blurred using a 20×20 Gaussian kernel with variance of 8. The corresponding deblurred images are shown in the same figure. A comparison of the verification performance when using the blurred images vs. when using the deblurred images is shown in Fig. 3.9, Fig. 3.10 and Fig. 3.11 for varying values of blur (increasing the blur using increasing values of Gaussian kernel spread) and for varying values of sparsity parameter L . We see that increasing the sparsity during enhancement, results in better overall verification performances. This can be attributed to a better reconstruction of the high resolution texture. Next we describe how we can add a gradient constraint to the above problem to obtain a better estimate of the high resolution texture.

Using class-specific constraint

As previously mentioned, the proposed long range system is an ideal tool to detect potential targets and threats from a distance, by identifying them based on the iris pattern. To test the performance of the system in such a scenario, we set up an open set identification experiment, with a cut-off hamming distance threshold. However, the performance of the system drops as the system-to-target distance increases (the stand-off distance). To aid the identification process in our experiment, when using targets at larger stand-off distances, we add an additional constraint to the coefficient estimation problem (3.5) - *a class-specific gradient error term*. As an addition to the super-resolution approach proposed, to use the coefficient estimation problem in a class/subject identification framework, we can add an additional class-specific gradient constraint to eqn. (3.5). The advantage is that, now we force the image recovery problem (3.5) to estimate coefficients that promote gradients along certain desired directions, which are class-specific. As against an authentication/verification problem, in the case of an identification problem we compare an input probe image with every class' image in the gallery prior to deciding the correct class. By

Algorithm 3 Algorithm to enhance resolution and classify an input image \mathbf{I}_l , using class-specific gradient constraints

Input:

Probe image \mathbf{I}_l

Set of gallery images to compare against $\{\mathbf{I}_g | g = 1, 2, \dots, N\}$

Process:

Divide \mathbf{I}_l into K overlapping patches $\{\mathbf{x}_l^i | i = 1, 2, \dots, K\}$

for $g = 1, 2, \dots, N$ **do**

for $i = 1, 2, \dots, K$ **do**

 Solve

$$\hat{\phi}_{gi} = \arg \min_{\phi} \beta_1 \|\mathbf{x}_l^i - \mathbf{D}_l \phi\|_2^2 + \beta_2 \|\nabla \mathbf{x}_g^i - \nabla(\mathbf{D}_h \phi)\|_2^2 \text{ s.t. } \|\phi\|_0 < L$$

 Set $\mathbf{x}_h^i = \mathbf{D}_h \hat{\phi}_{gi}$

end for

 Initialize $\hat{\mathbf{I}}_{hg}$ as a matrix of zeroes.

for $i = 1, 2, \dots, K$ **do**

 Set $\hat{\mathbf{I}}_{hg} = \alpha \mathbf{x}_h^i + (1 - \alpha) \hat{\mathbf{I}}_{hg}$ //before using \mathbf{x}_h^i here, reshape it into a 2-D patch

end for

end for

Output:

Return $\hat{\mathbf{I}}_{hg}$ with minimal hamming distance to \mathbf{I}_g

first enhancing the probe image with class-specific constraints, we force the enhanced image to have gradient directions similar to the gallery image we are comparing with. The idea here is that, the enhanced image will be most similar only to the gallery image that belongs to the same class. The entire algorithm for the open set identification problem is given by algorithm 3. Assume that we are comparing an input image \mathbf{I}_l with a gallery image \mathbf{I}_g in our database (where $g = 1, 2, \dots, N$ i.e. there are N gallery images/people in our database). We solve eqn 3.5 for each patch i in the input image, with the new *class* ‘ g ’ specific gradient constraint as follows,

$$\begin{aligned} \hat{\phi}_i &= \arg \min_{\phi} \beta_1 \|\mathbf{P}_i \mathbf{I}_l - \mathbf{D}_l \phi\|_2^2 + \beta_2 \|\nabla \mathbf{P}_i \mathbf{I}_h^g - \nabla(\mathbf{D}_h \phi)\|_2^2 \\ &\text{s.t. } \|\phi\|_0 < L \end{aligned} \tag{3.6}$$

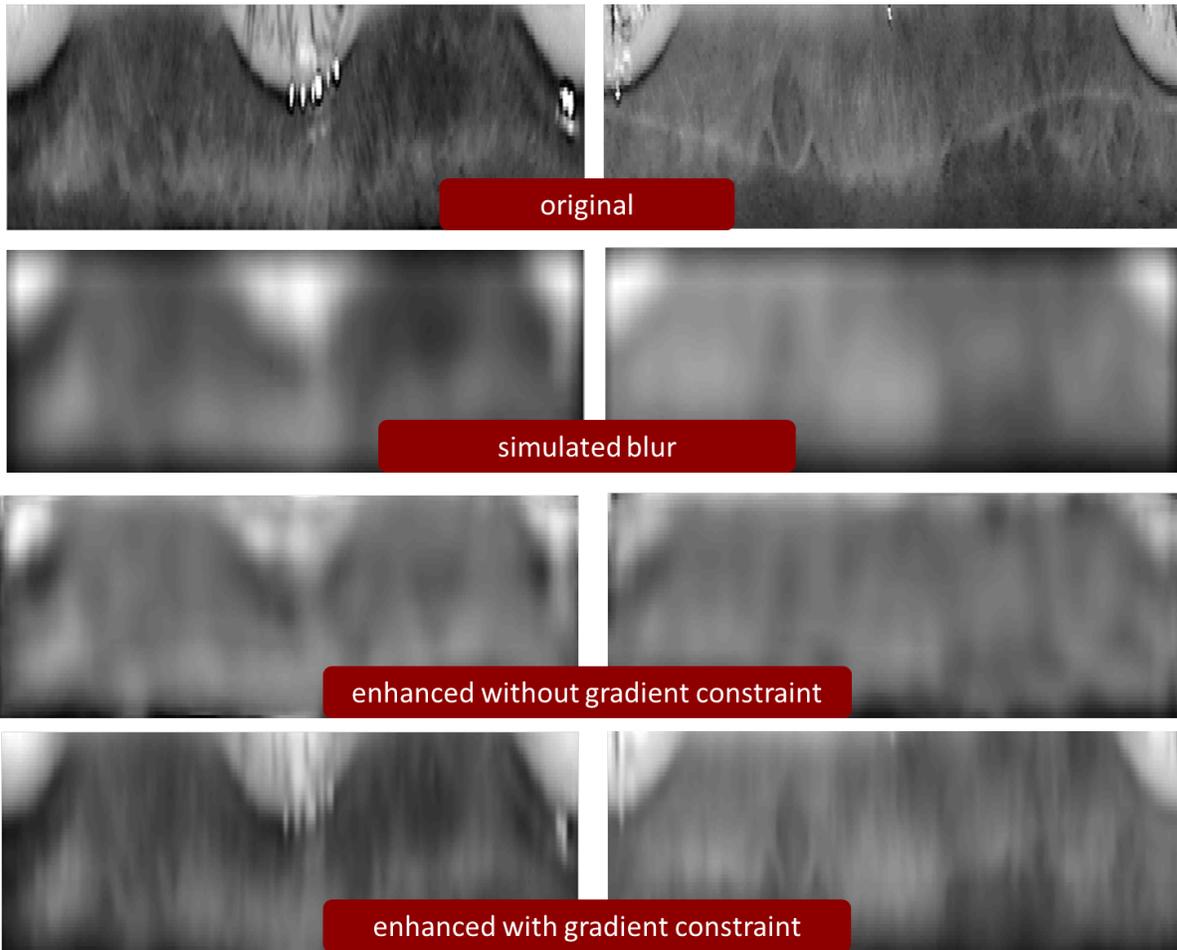


Figure 3.12: Examples of left eye images from the NIST ICE database artificially blurred and enhanced using the coupled dictionary model described in this chapter. The two types of enhancements - with and without class-specific gradient constraints are shown.

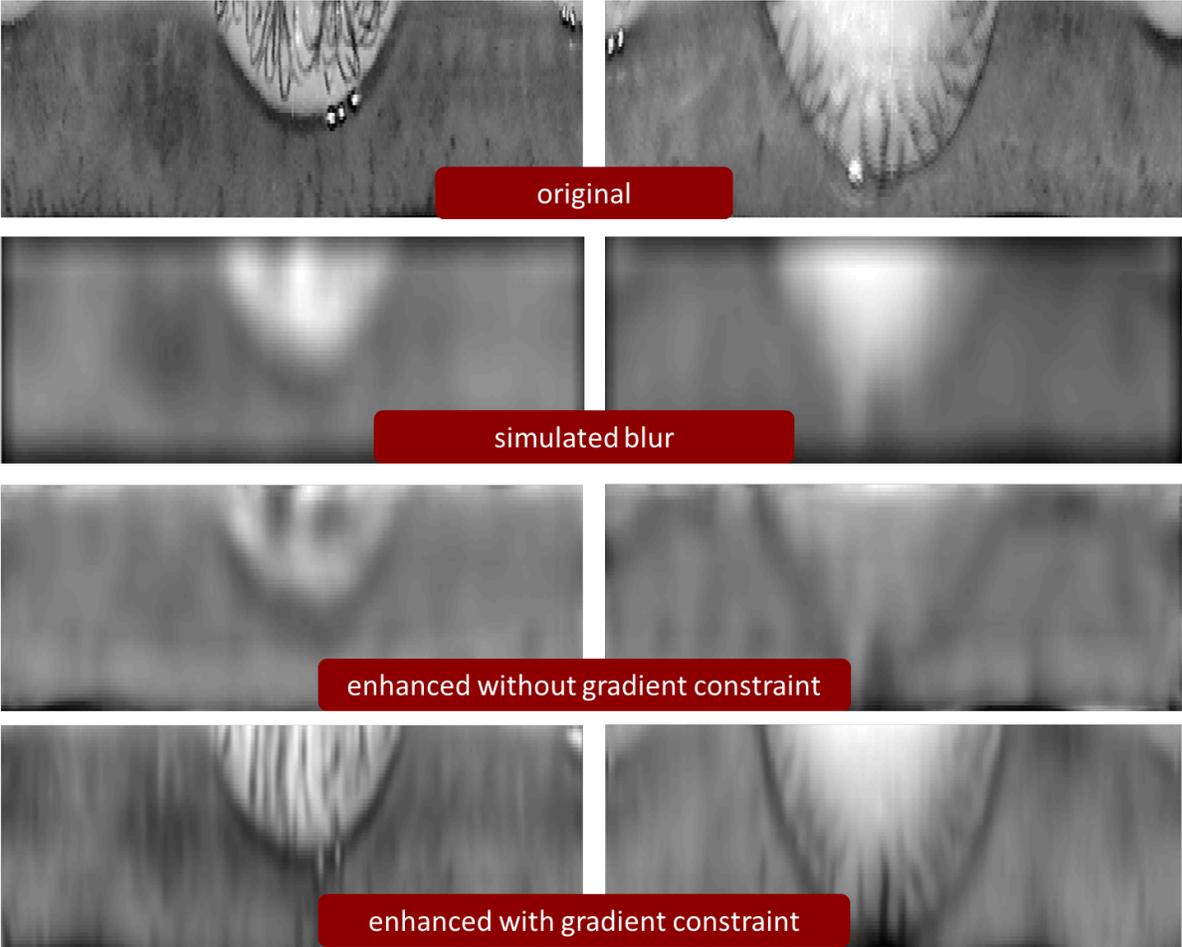


Figure 3.13: Examples of left eye images from the NIST ICE database artificially blurred and enhanced using the coupled dictionary model described in this chapter. The two types of enhancements - with and without class-specific gradient constraints are shown.

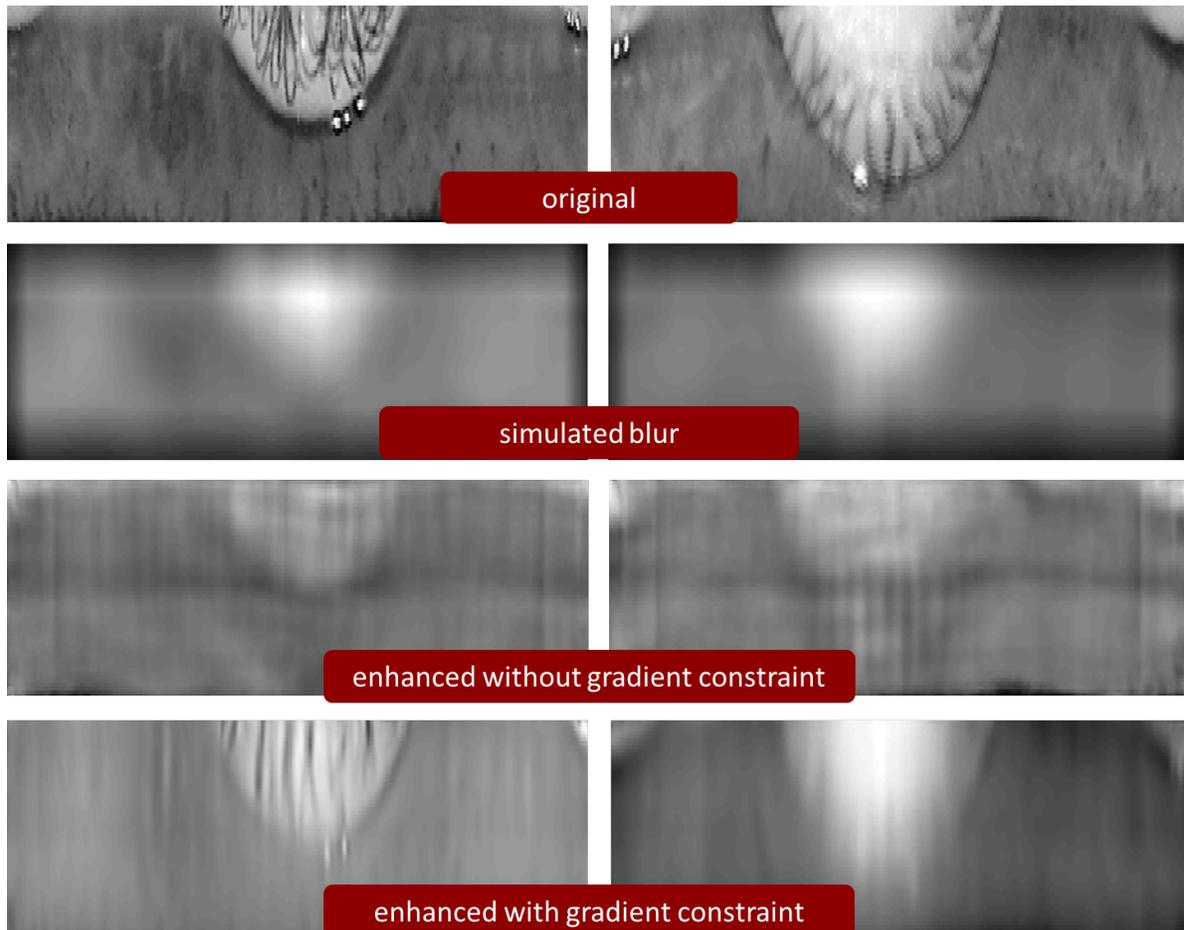


Figure 3.14: Examples of left eye images from the NIST ICE database artificially blurred and enhanced using the coupled dictionary model described in this chapter. The two types of enhancements - with and without class-specific gradient constraints are shown. The level of blur in this case is higher, compared to Fig. 3.13. We can visually make out the recovered details in either case.

\mathbf{P}_i as before, selects the i th patch from the image \mathbf{I} . On estimating all the high resolution patches $\{\mathbf{x}_h^i = \mathbf{D}_h \hat{\phi}_i | i = 1, 2, \dots, K\}$, the final *class-specific* high resolution iris texture $\hat{\mathbf{I}}_{hg}$ is estimated by alpha blending the patches (see algorithm 3). In this manner, we estimate $\{\hat{\mathbf{I}}_{hg} | g = 1, 2, \dots, N\}$, i.e. the high resolution estimates for all N gallery images. The advantage is that, now we have forced each estimate to have gradients similar to the gallery sample at hand. In this expression, empirically we set $\beta_1 = 1$ and $\beta_2 = 0.5$, to give more weight to our observation rather than the gradient from the gallery sample. We see that on matching these estimates to the gallery sample, a better identification score is obtained (identification results are reported in Table 3.1). It is to be noted here that, only in the case of matching gallery samples from the same class as the enhanced probe sample, will the match scores be maximally improved by enforcing the gradients. In all other cases, there is little improvement in match scores. Algorithm 3 summarizes the method of gradient constrained enhancement and subject identification given a probe image and a set of gallery images.

A few examples of artificially blurred iris textures (using a Gaussian blur kernel with zero mean and variance 8) from the ICE dataset and the enhanced- resolution counterparts with and without gradient constraints are shown in Fig. 3.12 and Fig. 3.13. On visually inspecting the enhanced images, we can clearly see an improvement in spatial resolution, compared to the original image. The distribution of genuine and imposter scores when running an identification experiment using these low resolution images and when using the enhanced counterparts are shown in Fig. 3.17. We see a distinct separation of the genuine and imposter distributions when using enhanced images, especially when using gradient constrained enhanced images. Fig. 3.15 shows the corresponding ROC curves when the match scores are used in a verification experiment. The results of the open set identification experiment, with a fixed acceptance threshold is given in Table 3.1. The identification rates when considering only the top match, the top 3, and top 5 matches are reported in this table, for different values of acceptance thresholds (i.e. hamming distance values).

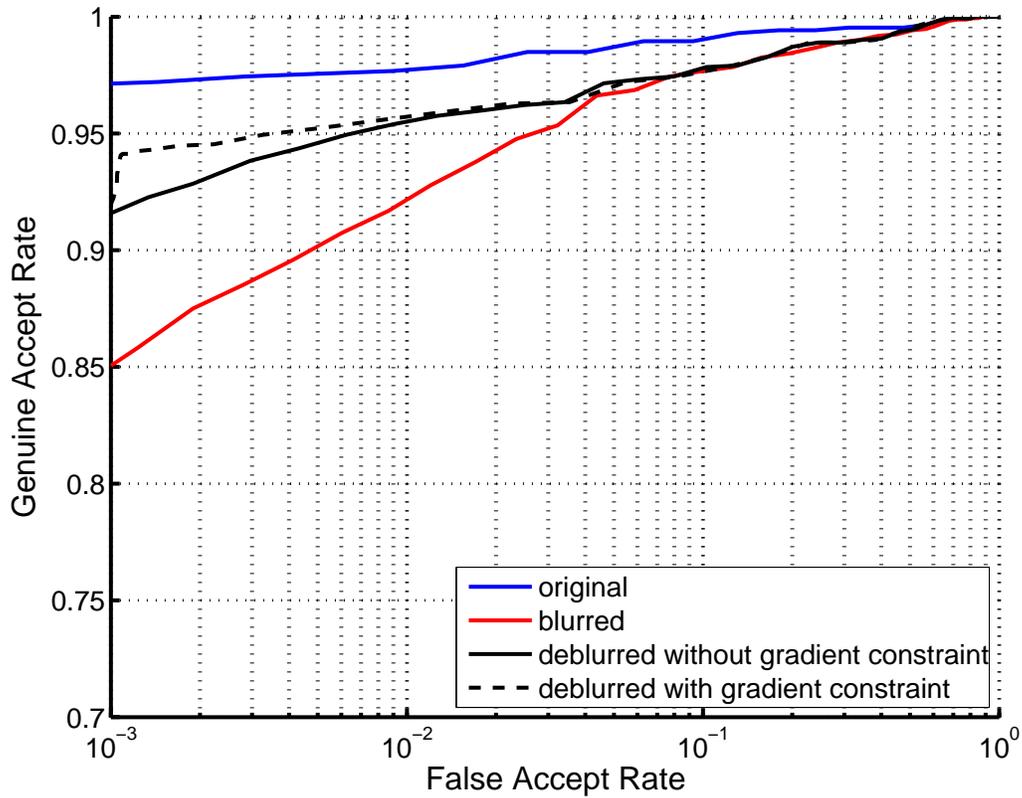


Figure 3.15: ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 20×20 Gaussian kernel with a variance of 8. The performance when using gradient constraints in eqn. (3.6) is compared with the performance without using the constraints.

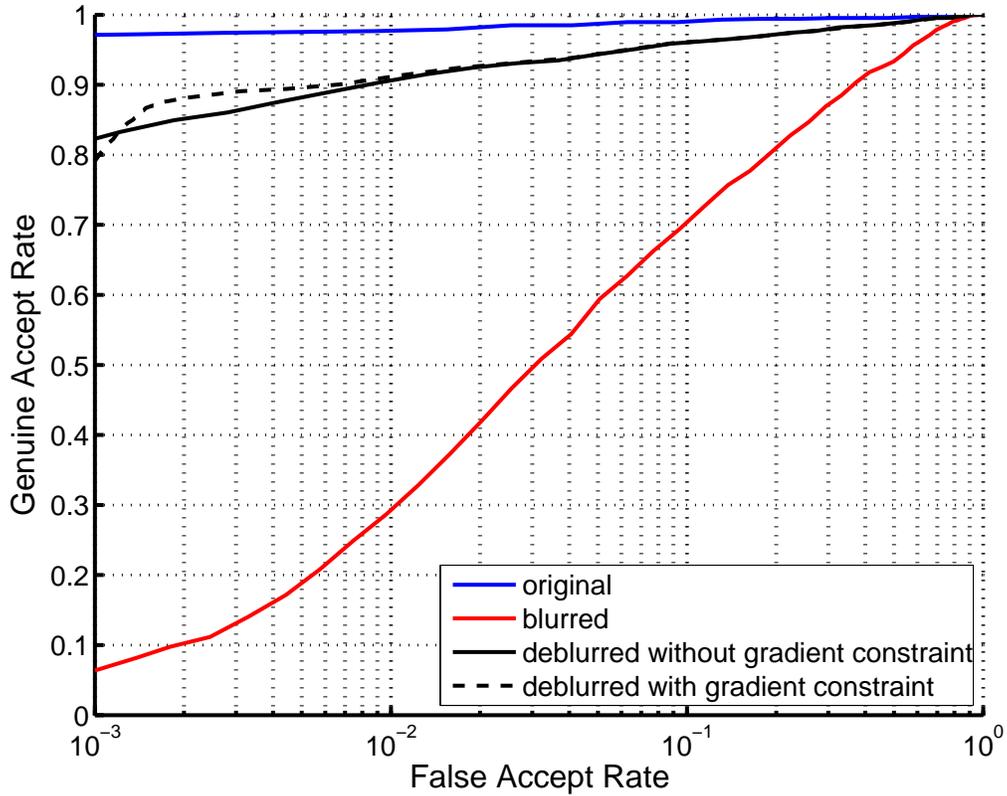


Figure 3.16: ROC curves comparing the verification performance using artificially blurred images from the ICE dataset and their enhanced counterparts. The blurring kernel is a 30×30 Gaussian kernel with a variance of 20. The performance when using gradient constraints in eqn. (3.6) is compared with the performance without using the constraints.

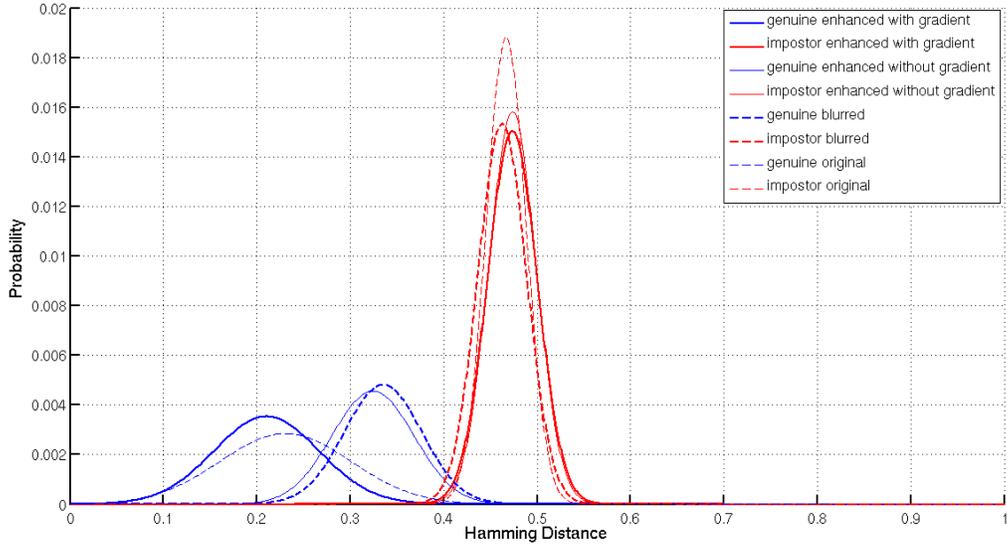


Figure 3.17: Distribution of genuine and impostor score distributions for the de-blurring experiment in Fig. 3.15. The thin line indicates the score distributions of the original left eye images. The thick dashed curves correspond to the blurred images and the thick solid curves to the de-blurred images using gradient constraint. The thin solid curves correspond to deblurring without the gradient constraint

In summary, the proposed method to enhance degraded iris images is based on solving a linear regression problem (3.1). By using low levels of sparsity, this equation can be used to reconstruct a de-noised version of any input noisy iris texture patch \mathbf{x} . In addition, a method to enhance the resolution of any iris texture degraded by a linear degradation operator \mathbf{H} was given in eqn. (3.5). By solving for the sparsest set of atoms in the low resolution dictionary \mathbf{D}_l , the corresponding high resolution texture can be estimated from the coupled high resolution dictionary \mathbf{D}_h . In our experiments we combine the de-noising

Table 3.1: Identification performance with and without de-blurring using increasing hamming distance acceptance threshold values

Threshold	Top match		Top 3 matches		Top 5 matches	
	blurred	enhanced	blurred	enhanced	blurred	enhanced
0.30	37.07%	84.91%	37.07%	85.13%	37.07%	85.13%
0.35	81.25%	89.44%	82.15%	90.95%	82.15%	90.95%
0.40	88.00%	88.00%	90.73%	91.16%	91.16%	92.10%

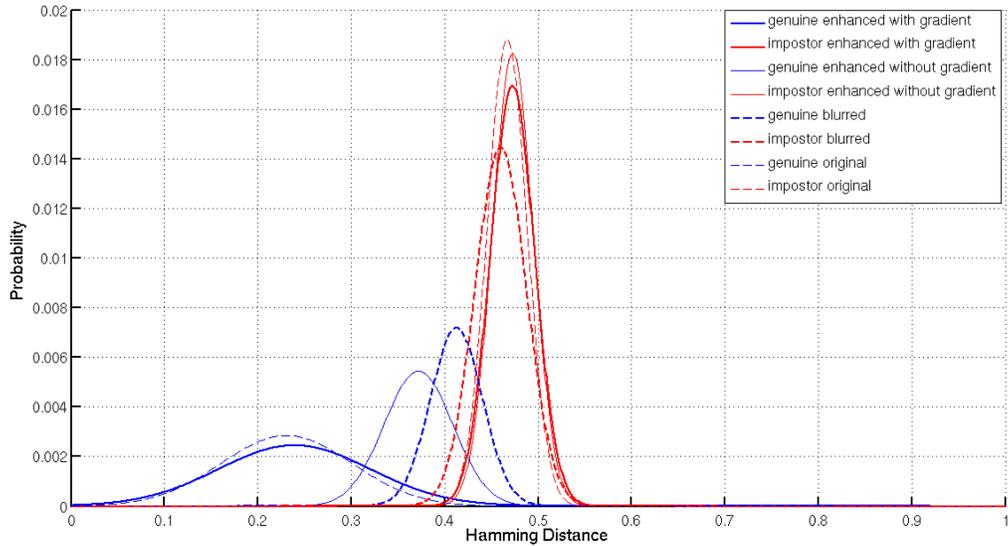


Figure 3.18: Distribution of genuine and impostor score distributions for the de-blurring experiment in Fig. 3.16. The thin line indicates the score distributions of the original left eye images. The thick dashed curves correspond to the blurred images and the thick solid curves to the de-blurred images using gradient constraint. The thin solid curves correspond to deblurring without the gradient constraint

and the resolution enhancement steps by using a low value of sparsity in eqn. (3.5). Finally, in eqn. (3.6) a method to encourage gradients in a particular direction, in the enhanced texture, is proposed. This expression is suited for a person identification setting, where we can present the enhancement problem in eqn. (3.5) with class-specific gradient constraints, while matching with the corresponding class' gallery sample. The algorithm to classify the sample is given by algorithm 3. In the next section, we use the enhancement paradigm described in this section and study its ability to enhance images acquired using the long range iris system described in the previous chapter.

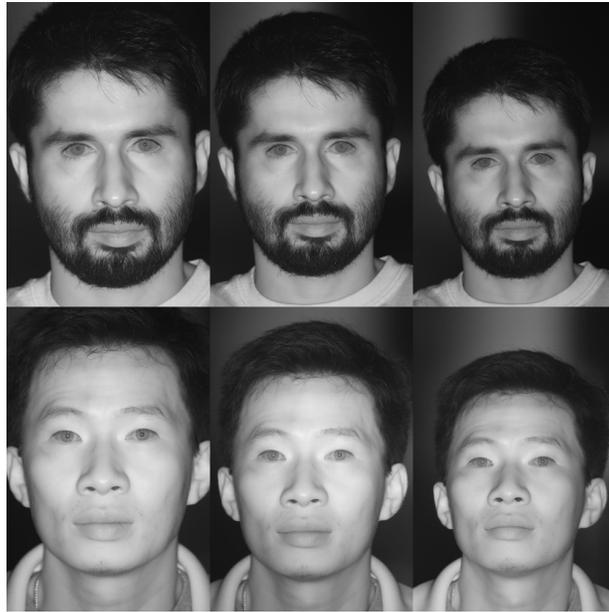
Chapter 4

Experiments and Results

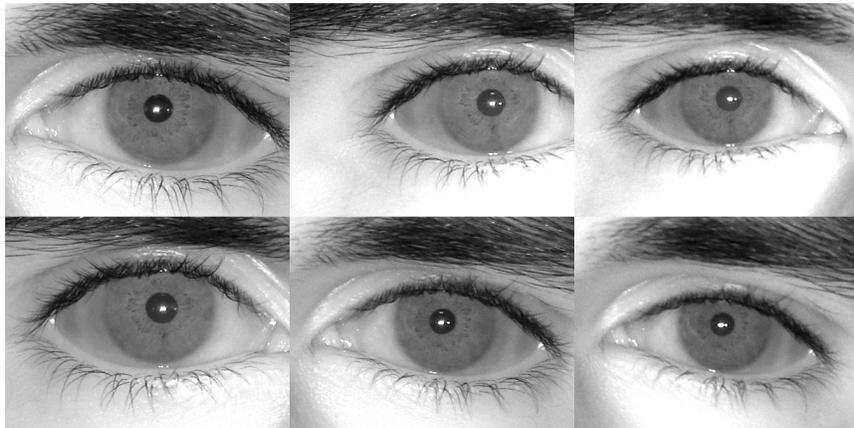
In this chapter results of iris recognition experiments, set up to test the efficacy of the proposed long range system, are presented. The experiments are run on a database of images acquired over the past few months - the database consists of face and eye images from 70 subjects. In addition to the data captured using the long range system, data were also captured with a short range and medium range system currently available in the market. For the purpose of this dissertation, a short range iris acquisition system is considered to be any system which has to be held right next to the eye for successful iris pattern acquisition. Examples of such systems include Iritech's IriShield-USBMK2120[8], the HIIDE system[18] and the PIER-T iris acquisition device[19]. These systems are capable of acquiring high quality eye images albeit only one at a time. In addition these devices require a high level of cooperation from the side of the user and a high level of skill on the side of the operator. Medium range devices refer to devices that operate at larger distances than these such as IrisID iCAM7000[30], Iritech IriMagic devices[20]. These devices are able to operate further away from the subject $20cm - 30cm$ from the subject. The short range system used in this work is Iritech's IriShield device and the medium range system is the iCAM7000. A comparison of iris verification using a database composed of images from all three systems can also be found in this chapter. A description of the database is presented

in section 4.1, following which the experiments are discussed. The primary motive of the experiments in this section is twofold

1. **to study the system's long range verification performance** - we do this by comparing our system's performance in a verification experiment to the performance of well-known commercial systems (which are known to excel in this scenario) in the same setting. The motive here is to measure the ability of the system to successfully verify whether a subject is indeed who he/she claims to be (such as in an access control setting). The results of this experiment are presented in section 4.2 and it is seen that, the proposed system performs as well as commercial short range/medium range systems. Using the proposed system, the verification performance is presented for various subject stand-off distances. The use of the enhancement methods introduced in chapter 3 is also explored, in order to improve the iris verification performance at longer stand-off distances. These results can also be found in section 4.2.
2. **to study the system's long range identification performance** - this is a study concerning the relevance of the system as a long range identification device. This ability is important in high security settings - airports, entrances to public rallies, check points in war zones - to name a few. If a potential threat is identified, based on his/her iris patterns, well before he/she is within range of sensitive areas, unfortunate incidents can be prevented. To test this ability, we run open set identification experiments, specifically in the context of subjects standing either $9m$, $10m$ and $11m$ away. We assume knowledge of a 'watch-list' of subjects, by enrolling subject iris images into a gallery, captured $9m$ from the system. The identification test is carried out using probe images captured at $9m$, 10 and $11m$. Here too, the use of the enhancement methods from chapter 3 is explored in order to improve identification performance. Further details regarding the database are presented in section 4.1. The long range identification experiments are presented in section 4.3.



(a)



(b)



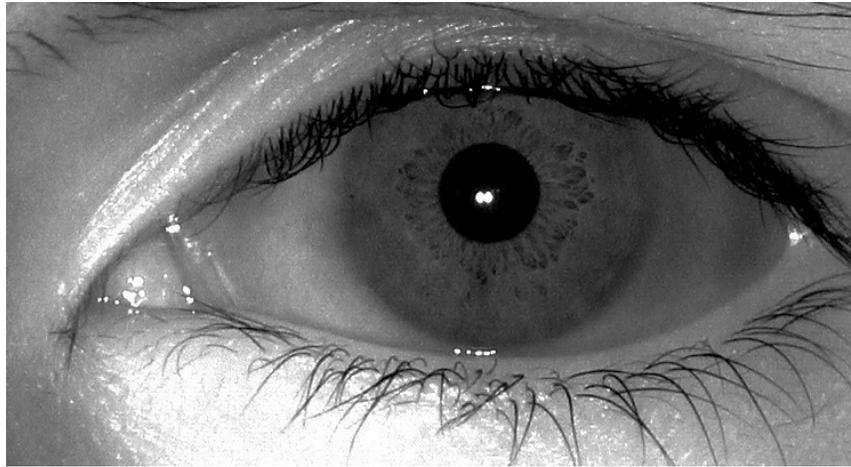
(c)

Figure 4.1: (a) Example of face images using our long range system at distances (from left to right) $9m$, $10m$ and $11m$. (b) and (c) show eye images extracted from (a).

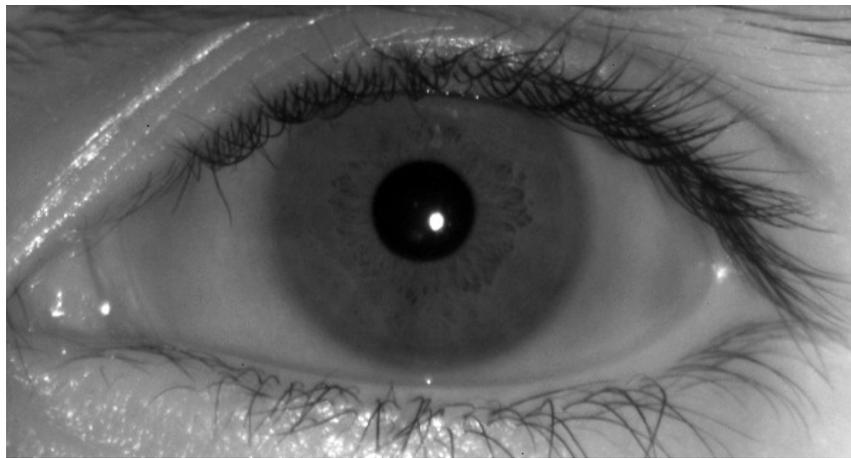
4.1 Long range iris database used for our experiments

In order to set up these experiments, we first gathered a database of 70 subjects. Database images were gathered using our system as well as two other commercial systems. In the context of this work, we wanted to use systems that could be used for both subject enrolment as well as verification. With that in mind, we chose to use two popular systems - the iCAM 7000 [30] and Iritech’s IriShield-USBMK2120[8], that are able to capture high quality images for enrollment as well as verification. **Long range data acquisition** - while acquiring images using our long range system, participants were asked to stand at distances of $9m$, $10m$ and $11m$ from the system. Five images of each participant’s face and eyes were captured at each distance. The number of participants in this study is 70, resulting in a total of 350 face images, left eye images and right eye images. Example images of faces and eyes acquired at these distances are shown in Fig. 4.1. With regards to the choice of illumination wavelength, as discussed in section 2.6, a lot of the light in the visible spectrum incident on the iris, is absorbed while some portion of it is reflected off the cornea. Most of the light in the infra-red wavelength of the spectrum is reflected back and can be imaged by the sensor. An infra-red LED panel[39] is used in our system. Four panels of these low power LED panels are optimally aligned such that the individual outputs superimpose over the required acquisition range. The system is also fitted with a infra-red band-pass filter to be able to image within a $100nm$ band centered about the $850nm$ wavelength.

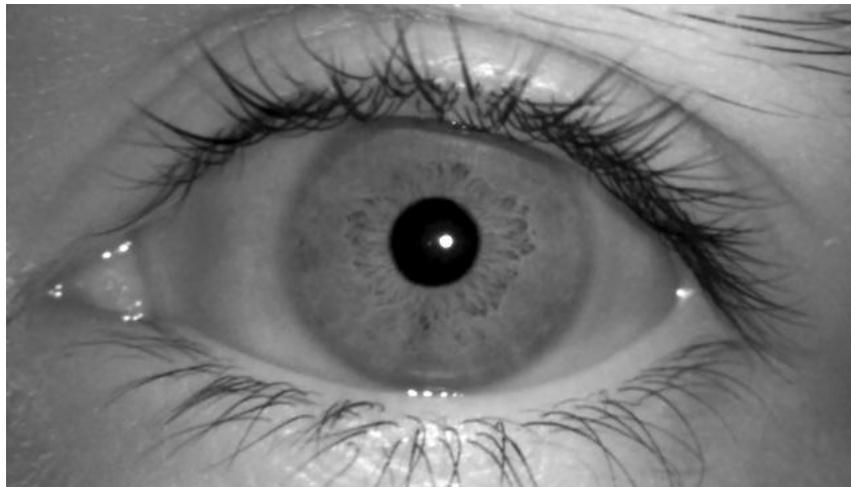
Medium & short range data acquisition - this database consists of left and right eye images acquired using the iCAM7000 and the IriShield-USBMK2120. It includes the same 70 participants as in the long range database and we capture three images per person. In total, from each device, we have collected 210 left eye and 210 right eye images. The iCAM7000 captures images from a distance of approximately $30cm$ from the subject, while the IriShield-USBMK2120 requires that the device be held right next to the subject’s eye.



(a)



(b)



(c)

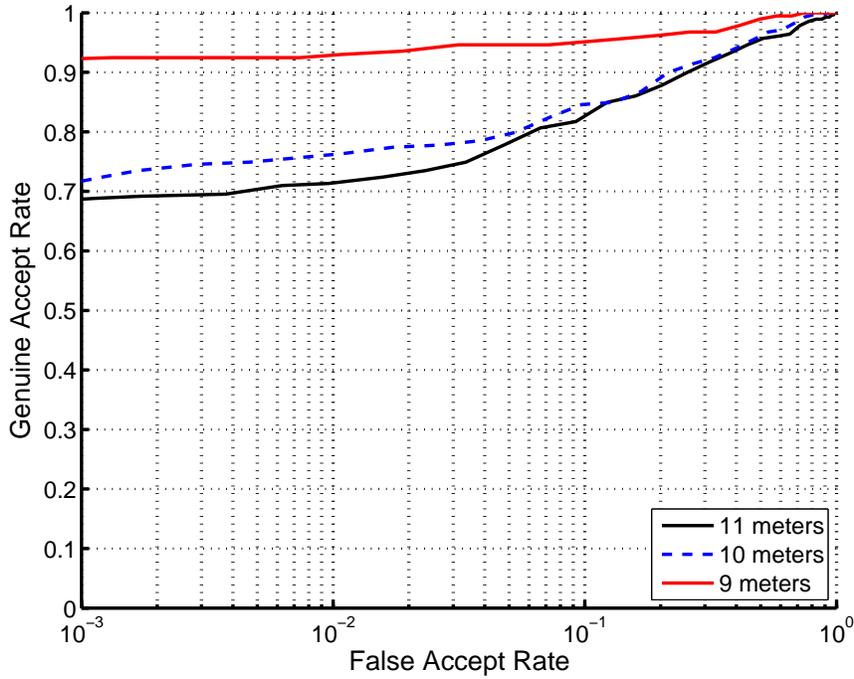
Figure 4.2: This figure compares iris images from the same user using (a) our system from 9m, with images captured using the (b) IrisID iCAM7000 system [7] from 30cm, (c) the IriShield-USBMK2120[8] from 3cm.

Fewer images are captured using the short/medium range systems, when compared to the long range system, since the former systems operate over a shorter range and involve a lot of cooperation from the user. The iCAM7000 for example requires the subject to follow audio prompts to position himself/herself at the ideal spot before imaging is initiated. Because of this, the images acquired vary very little and all images acquired are of high quality - as against the long range system, where the subject can blink or can move out of the depth of focus. It should be pointed out, however, that since these constraints are not enforced in the long range system, user interaction with the system is far easier and the system does the job of selecting the best iris images from the set of five collected during acquisition. Fig. 4.2 shows example eye images of the same participant acquired using all three systems.

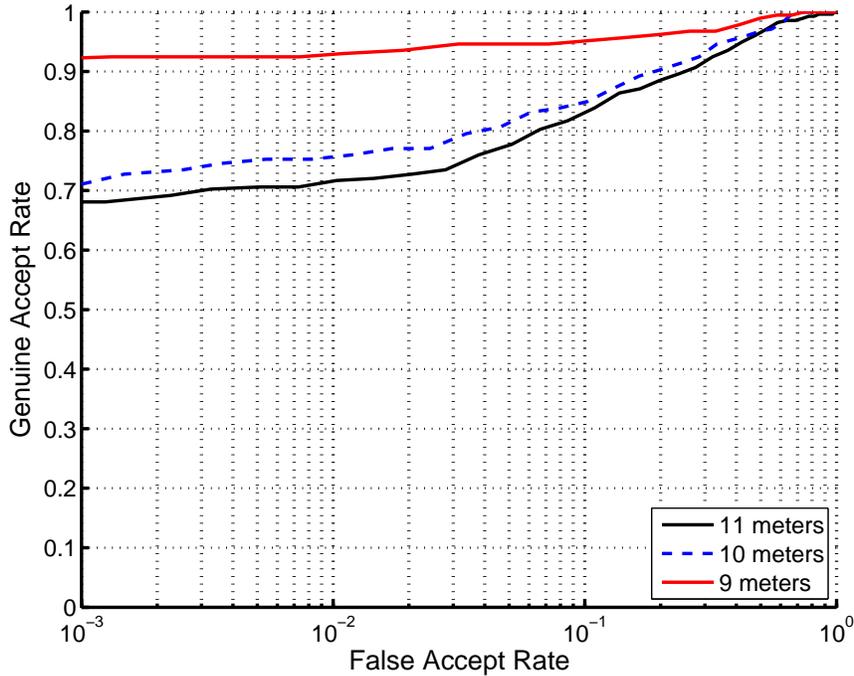
4.2 Long range subject verification performance

This section studies the efficacy of the proposed system as an iris pattern verification tool. We assume that the subject presents credentials claiming to be a person ‘A’ and study whether his/her acquired iris pattern can correctly confirm if he/she is indeed person ‘A’. The iris feature extraction and matching techniques were outlined in section 2.8.3.

In order to test the proposed system, we *enroll* subjects in a gallery, using only an eye image captured from the 9m mark. Three probe sets (the credentials presented by the user) are generated - one using three other images captured from 9m, another using three images from 10m and the last set using three images from 11m. Fig. 4.3 and Fig. 4.4 show the receiver operating characteristic (ROC) curves for verification experiments when using only the left eye; Fig. 4.5 and Fig. 4.6 show the curves when using only the right eye. The performance when using both the optimal Gabor filter described in section 2.8.3 as well as when using a popular open source algorithm [9] are shown. An ROC curve plots the genuine acceptance rates (GAR) versus the false acceptance rates (FAR) for varying

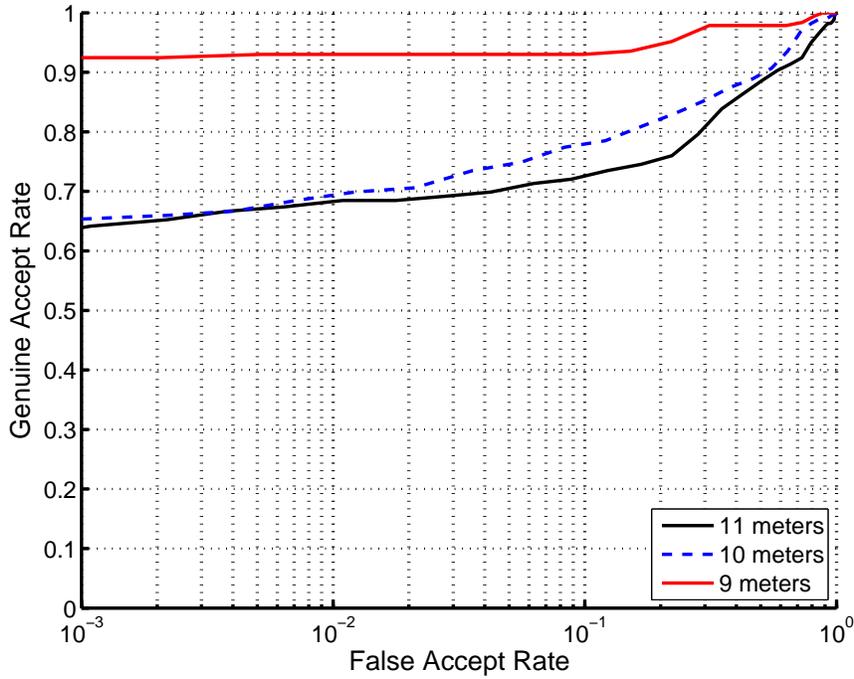


(a) using original probe images at 10m and 11m

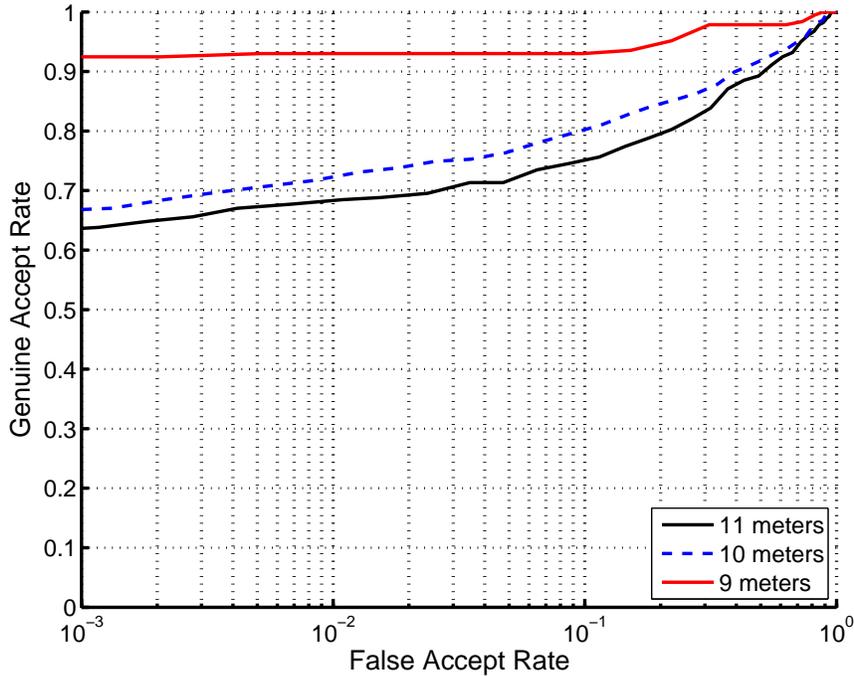


(b) using enhanced probe images at 10m and 11m

Figure 4.3: ROC curves showing verification performance using left eyes, using our optimal Gabor filter. In all experiments, images from 9m were used in the gallery. The enhancement used is the technique without the gradient constraint proposed in section 3.3.2.

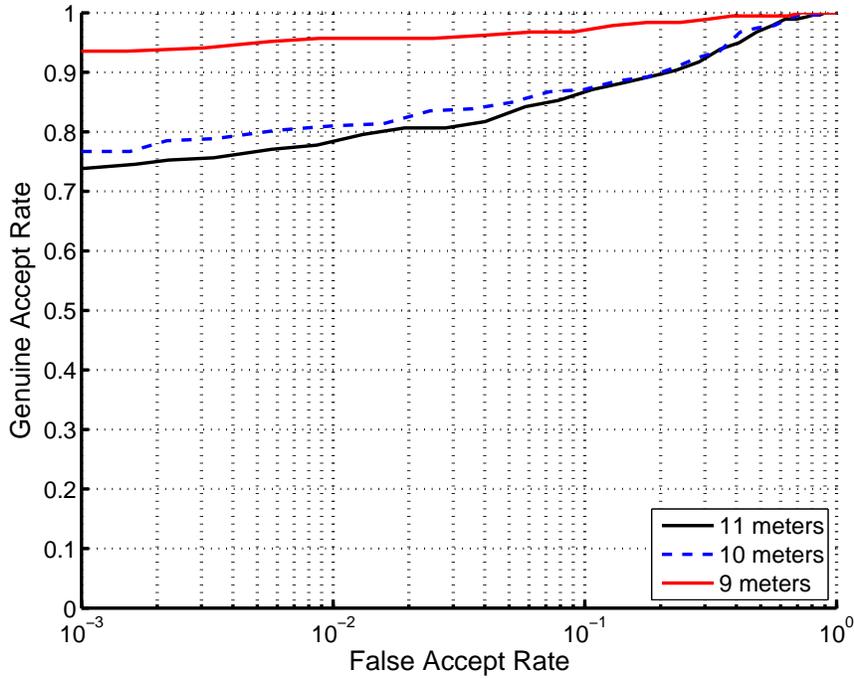


(a) using original probe images at 10m and 11m

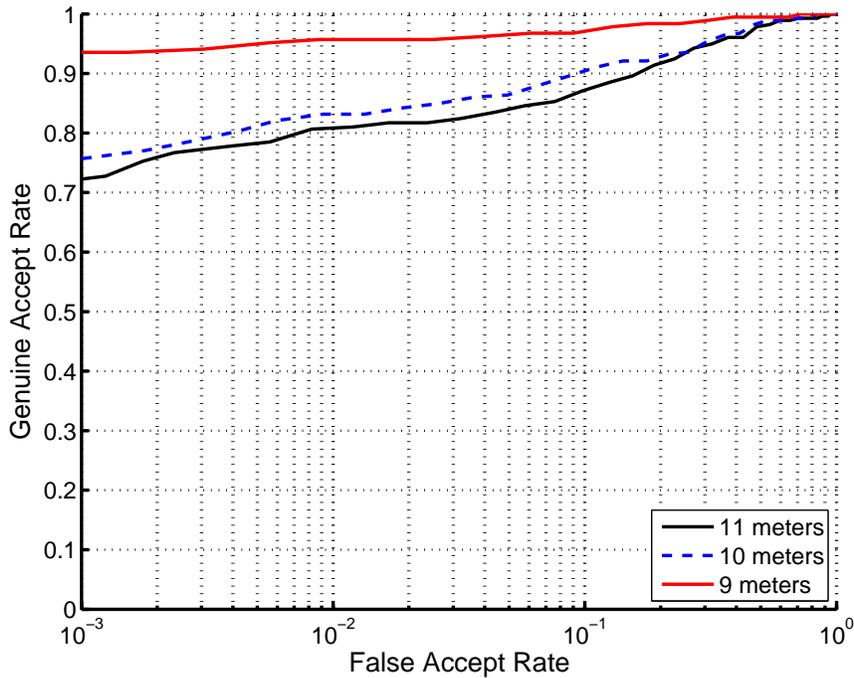


(b) using enhanced probe images at 10m and 11m

Figure 4.4: ROC curves showing verification performance using left eyes, using the open source implementation in [9]. In all experiments, images from 9m were used in the gallery. The enhancement used is the technique without the gradient constraint proposed in section 3.3.2.

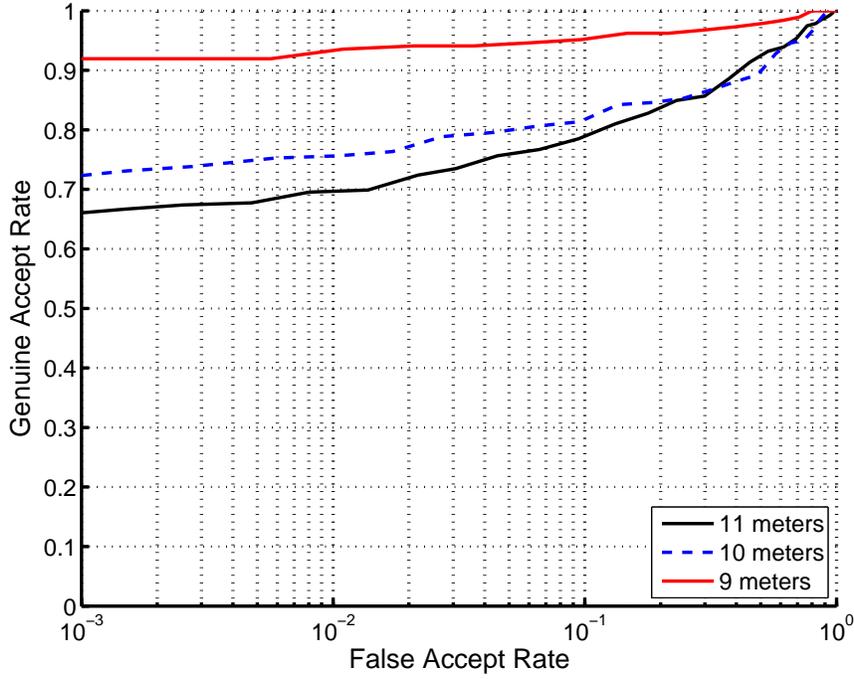


(a) using original probe images at 10m and 11m

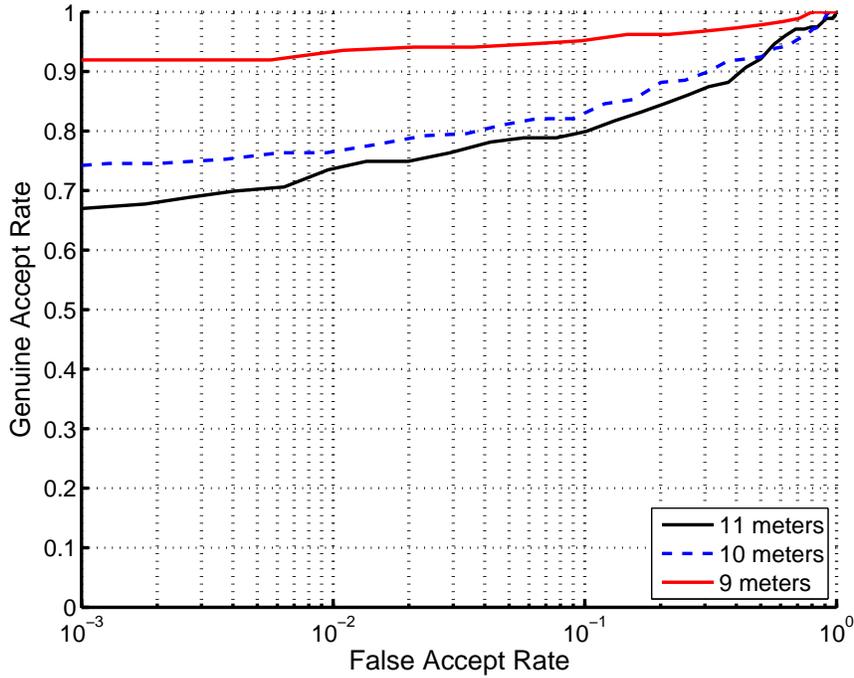


(b) using enhanced probe images at 10m and 11m

Figure 4.5: ROC curves showing verification performance using right eyes, using our optimal Gabor filter. In all experiments, images from 9m were used in the gallery. The enhancement used is the technique without the gradient constraint proposed in section 3.3.2.



(a) using original probe images at 10m and 11m



(b) using enhanced probe images at 10m and 11m

Figure 4.6: ROC curves showing verification performance using right eyes, using the open source implementation in [9]. In all experiments, images from 9m were used in the gallery. The enhancement used is the technique without the gradient constraint proposed in section 3.3.2.

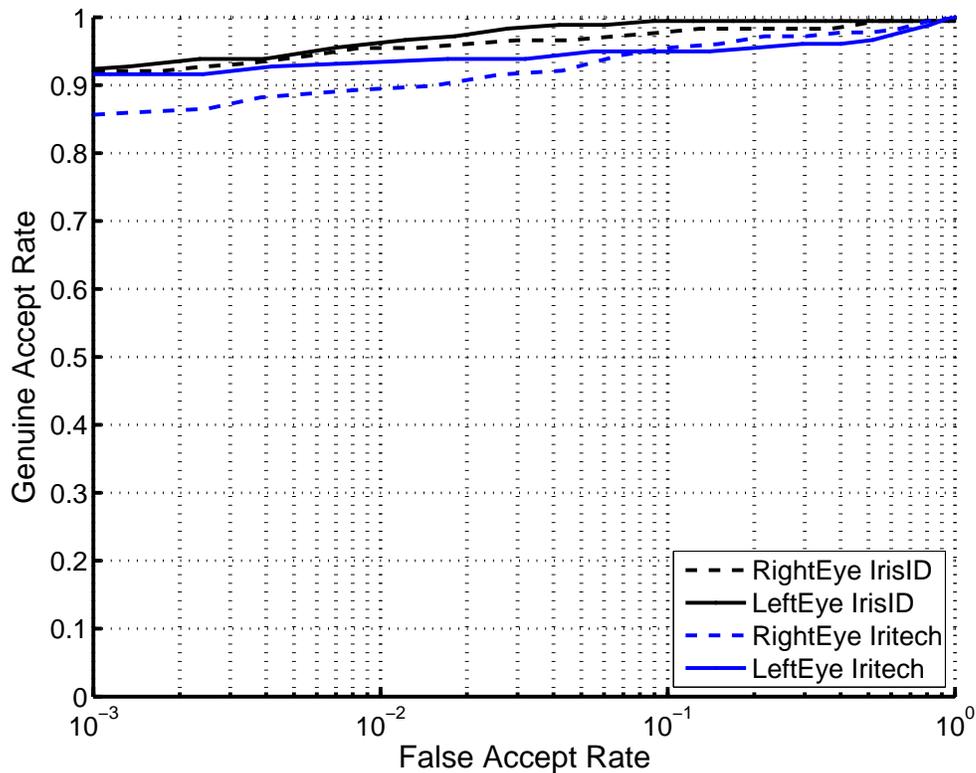


Figure 4.7: ROC curves showing verification performance using the Iritech system and the IrisID system.

values of hamming distance thresholds. By adjusting the distance threshold, based on the application at hand, the system operator can decide how many false accepts he/she is willing to tolerate for a particular number of genuine accepts. Ideally, we would require high GARs for very low FARs. Fig. 4.5a and 4.6a show the corresponding results when using only the right eye images. In either case, we see that the system is able to maintain very high values of genuine accepts for very low values of false accepts.

To serve as a benchmark, we used the data captured using both commercial systems, to perform the same verification experiments. These systems are popular in access control and verification scenarios in industry. One image per person is used in the gallery and the remaining are used in the probe set in either case. Fig. 4.7 shows the verification performance when using left eye images and right eye images acquired using both these

systems. It is noteworthy that there is very little difference in performance, when using either of these close range systems, compared to using the proposed long range acquisition system. Specifically, we see GARs greater than 90% when using the Iritech and IrisID systems at low values of FAR (between 1% and 10%). The same level of performance is seen when using images from our system which have been acquired at the same distance as the gallery samples (9m in the case of this experiment). The probe images gathered from the 9m mark consistently report GARs above 90%, for very low values of FAR (0.1%). The caveat of the system though, as pointed out earlier, is a slight degradation in performance at 10m and 11m - this is to be expected due to the decrease in spatial resolution and increase in noise as discussed in section 2.2.

Fig. 4.3b, Fig. 4.4b, Fig. 4.5b and Fig. 4.6b show the verification performances when the images at 10m and 11m were enhanced using the methods proposed in the previous chapter. For the purpose of this experiment, we have used eqn. (3.5) for the purpose of resolution enhancement and de-noising. A low level of sparsity ($L = 3$) is used for the purpose of de-noising each patch in a given normalized iris image. The coefficients thus estimated are used in the high resolution over-complete dictionary \mathbf{D}_h to estimate the corresponding high resolution patch. The coupled dictionaries are estimated in the same manner as that described in section 3.3.1. The degradation matrix \mathbf{H} in this case is approximated using a Gaussian blur kernel to mimic the observed degradation when imaging a USAF resolution target, specifically when comparing Fig. 2.4a to Fig. 2.4b and Fig. 2.4c. A general upward trend is seen in the case of both the 10m and 11m curves when using the enhancement method, especially in the case of Fig. 4.4b and 4.6b. The reason for not seeing a greater improvement is because, linear degradation is an approximation of what happens to the eye as we move further back. There is a non-linear deformation component, due to pupil dilation as the subject moves further back from the system. The pupil dilation is due to less light reaching the eye, especially if the ambient illumination is limited (which was the case in our data acquisition setting). In addition, during the normalization of the iris

texture, there is some amount of interpolation during the unwrapping of the iris. This too is not taken into account by this approximation.

4.3 Long range subject identification performance

In this second set of experiments, as mentioned previously, we test the performance of our system when used for long range subject identification. The ability to tell who a person is, with a high level of confidence, is important for threat detection and subsequent access control. Traditionally, an iris pattern is ideal for this application because of the unique iris signature associated with each person. This test shows the potential for long range iris based identification as against its traditional role as an authentication/verification tool (which was presented in the previous sub-section).

To test this application, we use the enrollment images from $9m$ in the identification experiment, to try and identify probe images from $9m$, $10m$ and $11m$. The experiment was set up as an open set identification experiment, such that only those comparisons with hamming distance values less than a fixed threshold (in this case 0.3) are considered. If any probe image reports an $H > 0.3$, we consider that person as not belonging to the database. Table 4.1 reports the identification accuracy when considering the top match and the top three matches for all probe images with $H < 0.3$. We first consider the results reported under the ‘original’ columns in that table, which uses the images prior to any enhancement. As is clear from the results, the identification accuracy is high when the user is close to the system ($9m$) and subsequently decreases with distance. The ‘enhanced’ column in this table reports the results of using enhanced images instead of the original ones. The method used for identification in this case was outlined by algorithm 3 in section 3.3.2. Past the ‘top 3’ case, the accuracy does not improve, since no other match falls below the threshold value. Hence, this is not reported in the table.

We see a distinct improvement across the board, when using the enhanced images.

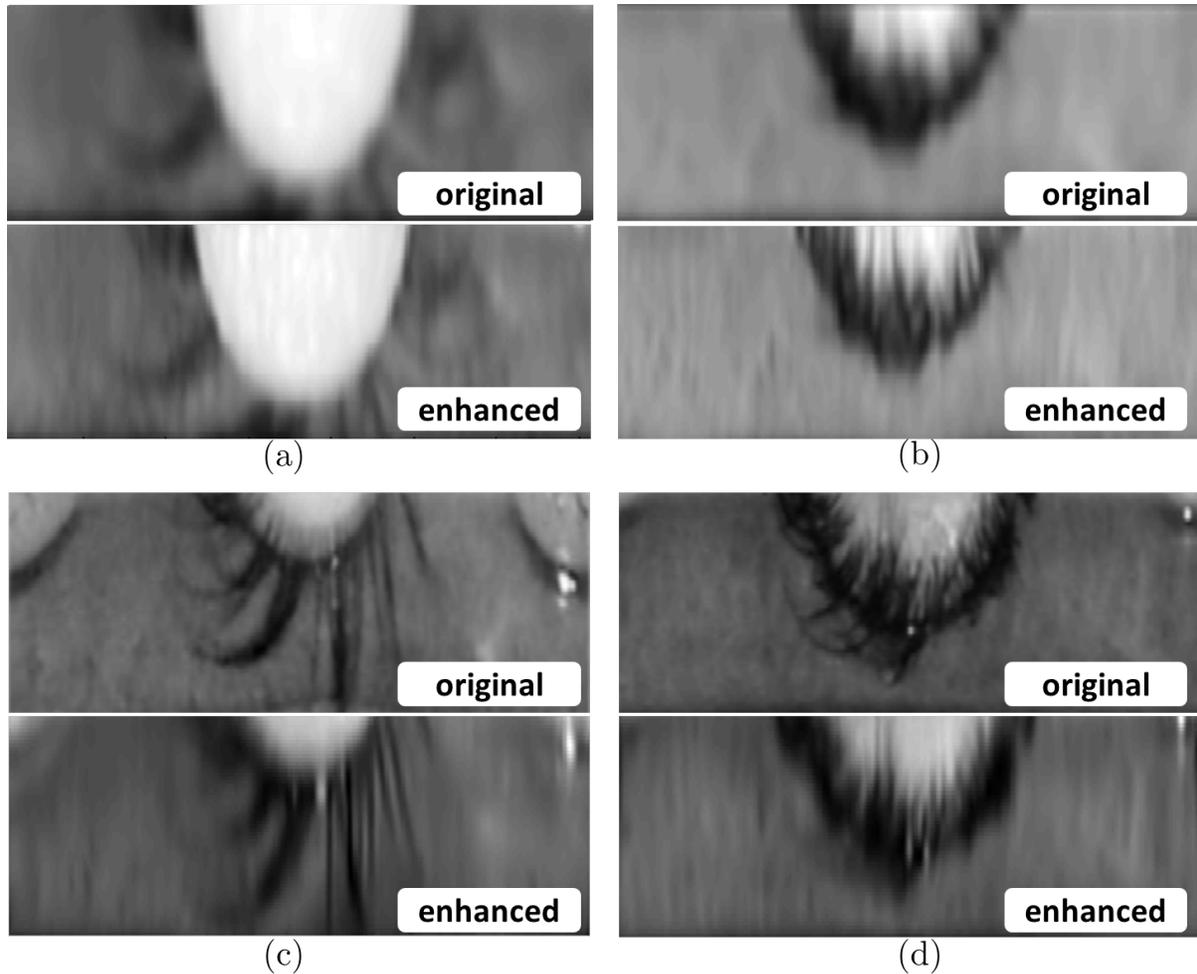


Figure 4.8: Examples of four normalized iris images captured at $11m$ and their enhanced counterparts. The enhancement is done using the gradient constraint method outlined by algorithm 3 in section 3.3.2. A low level of sparsity $L = 3$ is used in this algorithm to ‘de-noise’ the images. The top row shows how this method enhances gradients in the low resolution input images. In the bottom row, the enhanced images appear of lower resolution because of the de-noising effect when using a low value of L . Summarizing the observation across the database, irrespective of whether the input image appears blurred or noisy, the output-enhanced images always have a ‘soft focus’ (due to the de-noising effect of the algorithm) while still maintaining gradient information along particular directions (due to resolution enhancement portion of the algorithm).

Table 4.1: Comparison of identification scores using original and resolution-enhanced images with an acceptance threshold of 0.4.

Distance	Top match		Top 3 matches		Top 5 matches	
	original	enhanced	original	enhanced	original	enhanced
10m	73.3%	70.3%	75.2%	78.2%	75.2%	80.6%
11m	64.8%	70.9%	73.3%	75.8%	75.2%	81.2%

A few examples of enhanced images are shown in Fig. 4.8. We see that in case of low resolution texture images (the top row in the figure), some amount of enhancement results in textures with more gradient information. In the case of images that still have gradient information, there seems to be more of a de-noising effect and the enhancement results in a comparatively ‘softer’ image. A low level of sparsity $L = 3$ is used in algorithm 3; this results in simultaneous de-noising and resolution enhancement for all input texture images. This is why irrespective of whether the input image is blurred or whether it is noisy, the output enhanced images always have a ‘soft focus’ (due to the de-noising effect of the algorithm) while still maintaining gradient information along particular directions (due to the resolution enhancement by the algorithm). Fig. 4.9 compares the genuine and impostor score distributions when using the original probe images from 10m and 11m and compares them to the distributions when using enhanced images. Here too, for enhanced images, we can see a visible shift in the genuine distribution to the left (the desired direction for genuine scores), while the impostor distribution reports approximately the same distribution as the original case. Table 4.1 reports identification results when using both the original set of probe images and the enhanced set.

Identification using VeriEye SDK - In addition, we conducted an identification experiment using the VeriEye SDK from Neurotechnology[64]. This algorithm was one of the best performing algorithms as reported by NIST in the IREX II report[15]. In this experiment again, one image per person captured from 9m was used in the gallery and three images per person captured from 9m, 10m and 11m were used in the probe set. The

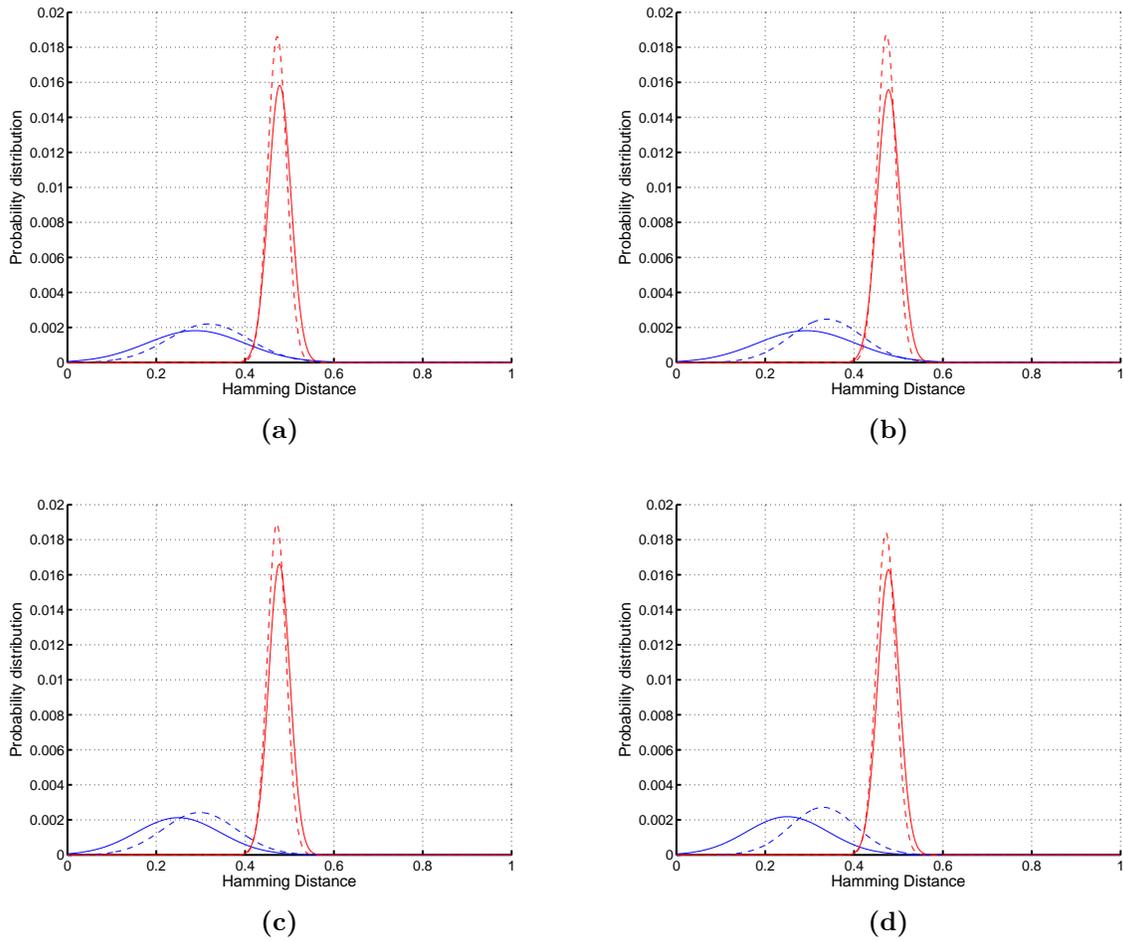


Figure 4.9: Comparison of genuine and impostor score distributions for both the left-(a),(b) and right-(c),(d) eye images. (a) and (c) show the distributions for probe images from $10m$, while (b) and (d) correspond to $11m$. In all cases, the dotted lines indicate the original distribution and the solid lines indicate the distributions after image enhancement. Red indicates impostor distributions, while blue indicates genuine distributions. In all cases, we see a distinct improvement when enhanced images are used instead of the original.

Table 4.2: Cross-device identification performance using a hamming distance acceptance threshold of 0.3.

Distance	Top match		Top 3 matches		Top 5 matches	
	original	enhanced	original	enhanced	original	enhanced
9m	36.0%	41.8%	36.0%	60.1%	36.0%	78.8%
10m	30.0%	30.0%	30.0%	57.5%	30.0%	70.6%
11m	28.1%	28.1%	28.1%	58.8%	28.1%	72.6%

identification rates reported by the algorithm, using the top match were 74.5%, 61.3% and 45.5% for the three probe sets respectively. **Subject identification across devices** - Another important application context for any biometric system, is the ability to enroll people using one device and identify them at a different location using another device. This is especially significant for surveillance scenarios, where security personnel maintain a database of high risk individuals - with the aim of identifying them later, at a different location, using a different acquisition system. This problem is particularly difficult for iris based biometric systems, due to the different properties of the optics and sensors used across acquisition systems and differences in imaging environments. However, as a first order approximation one can assume that these systems differ only in spatial resolving power. Within the context of this approximation, we ran an additional experiment, to explore identification across devices, using enrollment images from the short range system in our study - IriShield-USBMK2120. One image per person, from our database of 70 subjects is used as a gallery sample. Three separate probe image sets, as before, gathered from 9m, 10m and 11m are used in this identification experiment. Each probe set has 3 images per person, as in the previous experiment. Table 4.2 shows the identification results. Here too, we show the results before enhancement ('original' column) and after enhancement ('enhanced' column). The observed results are encouraging and show the potential of using the proposed enhancement method in such scenarios. However, a more accurate degradation model, that takes into account the different camera response functions has to be developed, to improve the identification scores further.

Chapter 5

Concluding Remarks and Future Work

The world of today is one of great political, economic and technological advances. Unfortunately, as has always been the case in history, the betterment of life always comes with a caveat of increasing insecurity. Today's democratic societies face an ever increasing need to develop better tools to fight crime and terrorism, to better protect their way of life and their resources. It is impossible to always have a 'human in the loop' system for security, just because of the sheer size of our population and the sheer volume of resources that need protection. Machine vision as a result has been firmly rooted as the tool of choice in any security infrastructure and has become indispensable in the fight to keep societies safe. It has provided us the ability to have a virtual eye of sorts everywhere, even in locations where a human presence may be difficult or even harmful.

Over the past two decades, machine vision based security tools have seen further progress with the influx of biometrics. The aim of using these tools is to identify people by measuring physical traits, specific to the person in question; and thus detect any potential threats well before they have a chance to cause harm. However, the technology and the state-of-the-art are severely limited with respect to the range they can operate

over; the amount of cooperation required from the side of the person being imaged and even in terms of the image quality required by the back end algorithms that make up the system's intelligence.

The aim of this dissertation is to address a sub-set of these problems with specific emphasis on the popular iris biometric. The iris has come to the fore as the biometric of choice, over the past decade, because of its uniqueness across the population just like fingerprints. Unlike fingerprints however, there is lesser likelihood of a change to a person's iris pattern due to injury, illness or age. It is also possible to image the iris from a greater stand-off distance, unlike fingerprints, which require proximity between the sensor and the subject. Owing to its small size ($12mm$) however, it is challenging to image iris patterns from long distances and hence all the sensors currently in the market operate close to the subject of interest. The main contribution of this dissertation was to break away from this conventional stand-off distance and to discuss and address the challenges in designing a long range iris acquisition and recognition system. Chapter 2 outlined these challenges and solutions to achieve high quality iris acquisition from longer distances than what is possible currently in the commercial and academic space. Quantitative analyses were also presented with respect to different lens/sensor combinations that may be used in such a system. The final solution presented is capable of acquiring high resolution face and dual eye images from users standing at distances of $6m$ to $12m$ away. Examples of images acquired and the spatial resolution achievable were presented in the same chapter.

Chapter 4 reported iris recognition performance in both a verification experiment as well as an identification experiment, using a database of images gathered with the long range iris acquisition system. The performance is similar to that reported using commercial short range and medium range systems, as reported in the same chapter. One observation to note was the decrease in performance with increasing subject distance (compared to the distance at which the gallery images were acquired). Solutions for de-noising and super-resolving iris images captured using this system at greater stand-off distances were

presented in chapter 3. The solution was based on representing an iris texture using a linear combination of a sparse set of texture atoms from an over-complete dictionary. A computationally efficient greedy algorithm, namely orthogonal matching pursuit, was used to estimate this sparse set. In that section, a method to estimate a coupled dictionary model - a low resolution dictionary \mathbf{D}_l coupled to a high resolution dictionary \mathbf{D}_h - was presented. We saw how this model minimizes the reconstruction error for a low resolution patch and its high resolution counterpart simultaneously. As a result, by estimating a sparse coefficient set to reconstruct a given, previously unseen, low resolution iride texture patch in \mathbf{D}_l , the high resolution counterpart can be estimated by simply using the same coefficient set in \mathbf{D}_h . In addition, we showed how, using class specific gradient constraints, one can force the reconstruction to have gradients in a particular direction. Both these reconstruction paradigms were evaluated in the same chapter. It is seen from the experiments in chapter 4 that this method improves both the verification and identification results when using less than ideal iris images acquired with the system. This enables one to not only push the range of the acquisition system, but also to acquire images in less than ideal imaging conditions, when for example - one is forced to use non-ideal hardware (low quality optics/sensors), there is noise in the image due to low illumination, there is blur/de-focus due to inadequate focus estimation *etc.* - all of which are problems encountered in the real world.

5.1 Improving the design of the proposed system

A promising future direction in terms of system functionality is developing a module to identify subjects on the move. Preliminary work has been reported in [42] where we describe how focus estimation may be performed for users moving at a walking pace. With current and future developments in this direction, we will be able to acquire iris images from users moving at a faster pace, thus reducing constraints further on the user's end. This will enhance the throughput of the current system, enabling it to capture enrollment

iris images from users on the move, with very little co-operation from them. A second method to achieve this, is to use a range finder to track the position of the subject and to keep the focal plane for acquisition fixed at a choke point through which subjects will have to pass. This choke point may be the point of immigration in airports, can be controlled entrances to public rallies etc. If the focal plane is fixed here, then acquisition needs to be performed only when the subject is in the vicinity of this point and we are always assured of getting in focus images. This is another method to increase throughput using long range systems, while minimizing cooperation required from subjects. Another useful feature is the ability to enroll/identify multiple users at the same time. This tool will be useful when trying to pick out individuals from a crowd and accurately identifying them, serving as an effective localization tool for city-wide surveillance. One way of achieving this is to use wide-angle lenses during acquisitions. Other distortions, not considered in this dissertation, such as barrel distortion have to be considered in such systems. An alternate method of capturing multiple subjects in one exposure, is to use catadioptric systems such as [65]. These cameras will be able to capture panoramas around the position of the camera, following which subjects of interest may be tracked and biometric patterns acquired. This will require, in addition to eliminating barrel distortion, a study into distortions resulting from imperfections in the mirror.

In addition to all these improvements, the current work can also be extended using new imaging techniques that have been developed over the past decade. These techniques use a combination of both novel optics as well as back-end computation to break free of the traditional ‘camera obscura’ model of imaging. A few thoughts on these methods as well as how the ideas proposed in this dissertation apply to them, are presented next.

5.2 A note on computational imaging and role in biometrics

At the outset, I would like to point out that the analyses presented earlier on in this dissertation is general and can be used in designing any biometric system and not just one for the iris in particular. A major point to be noted here is that, all of these are based on the traditional *camera obscura* model of imaging, which involves an object, sensor and a lens in between, to focus light rays in a required fashion. The principles of imaging are in essence the same as that of a standard pin hole camera - a very restrictive sampling of the complete set of rays, of the light field, emanating from any given scene/object.

There has been a surge of work that has attempted to break free of this traditional camera model by taking advantage of advances in computing and image processing. Computational cameras, as they are referred to in the literature, use a combination of novel optics and computations to produce a final image. The aim of most of these cameras is to find better ways to record all of the light rays from a scene instead of just a sub-set which is captured using traditional cameras. In the case of all these cameras, the captured image is optically coded and may not be meaningful in its raw form. The inverse computation module at the back end has a model of the optical setup and the related point spread function (PSF), which it uses to decode/deconvolve the captured image to estimate a new type of image that holds more benefits than an image from an conventional imaging system. A few examples of work on such systems are:

- Seminal work in building a light field camera (plenoptic camera) was done early in the late 20th century by Adelson and Wang[66]. Their design was based on placing an aperture behind the lens to sample only a sub-set of the light rays in each exposure. By varying the position of this aperture, they were able to show how to image the same scene from slightly varying view points. This preliminary design was used to

build a single-lens stereo system.

- Several other researchers in the field have developed variations of the design proposed by Adelson and Wang, some prominent ones being works by Okano *et al.* [67] and Naemura *et al.* [68], who use graded-index (GRIN) microlens arrays in front of the sensor to sample subsets of rays in a single exposure. Another approach to capture an entire light field in a single exposure was proposed by Wilburn *et al.* [69], using an array of cameras.
- One of the most popular systems for capturing light fields that takes inspiration from the design proposed in [66] is that proposed by Ng *et al.* in [70]. This design has been commercialized and is being currently marketed by Lytro [71]. This design is based on using a microlens array at the focal plane of a primary lens. The sensor captures the intensity from multiple light rays passing through each point on the primary lens during one exposure. By integrating over a sub-set of light rays Ng *et al.* show how we can recover images for different focal length values. Both [70] and [72] provide a detailed description of the camera design and associated theory.
- Another general class of imaging systems uses a form of encoding light rays at the pupil plane. An optical element such as phase plates or coded apertures are placed at, or close to, the pupil plane of a traditional lens. An important work in this category was the wavefront coding technique put forth by Dowski and Johnson [73]. This work uses an aspheric lens instead of the primary lens at the pupil plane, which produces a depth-independent blurred image on the sensor. The PSF of this aspheric lens is known beforehand. They use this PSF to deconvolve the actual image from the raw image on the sensor. The application of this technique extends the depth of field compared to traditional imaging systems.
- A system similar to the above is coded aperture photography proposed by Levin *et al.* in [74], which breaks the aperture up into smaller segments and thus is able to

simulate the effect of a low-pass filter with a larger support. In effect, the PSF of this system has non-zeros where the conventional system has zeros. Because of this, they are able to recover greater detail using inverse imaging techniques, when compared to using a standard fully open aperture.

All the advantages accrued using these novel imaging techniques can directly be carried over into a long range biometric system. Most of the analyses presented in chapter 2 still holds in the case of these cameras. Take the case of light field cameras such as that described by Ng *et al.* [70]:

- The aim of this camera is to mitigate the effect of depth of field by providing the ability to re-focus after the image has been acquired. This is done by sampling only a sub-set of rays and intergrating them over a limited support on the actual sensor, to simulate varying positions of a sensor when using a traditional camera.
- The choice of focal length of the primary lens is still dictated by the choice of magnification as discussed in section 2.1
- The maximum spatial resolution resolvable also remains the same, since it depends on sensor pixel size and the aperture size of the primary lens as described in section 2.2.
- The new optical element introduced in the path of light in the case of this camera is the micro-lens array placed at the focal plane of the primary lens. This optical element will have two effects. First, it makes the discussion of depth of field in section 2.3 obsolete, because of the ability to re-focus after image acquisition. The size of the sensor pixels and the density will determine the directional resolution of the camera i.e. how finely can the light field be sampled. This will ultimately decide how fine of a focal stack can be re-created on the object side. Hence, smaller pixels will ultimately aid our spatial resolution as well as in achieving perfect focus on the focal plane on which the iris pattern resides.
- A second implication of using this optical element is a change in the field of view of

the device. In effect, the sensor serves as a plane to record light ray information from several smaller imaginary sensors placed at varying locations behind the primary lens. By choosing the location of the imaginary smaller sensor, we integrate over the corresponding set of light rays on our main sensor, to simulate the image that would have been formed on the smaller sensor. Because we integrate over only a sub-set of these light rays, the imaginary sensor images only a portion of the image circle at the output of the primary lens. The effect is the same as using a crop sensor such as Canon EOS70D[32] vs. using a full frame sensor such as Canon EOS5DMark2[31]. In order to account for the loss in field of view, many light field cameras have started opting to use medium format sensors, which are bigger in dimension such as [75]

The analyses are simpler in the case of computational cameras that use pupil plane encoding such as [73] and [74]. The effect in these cameras is a modification of the PSF as recorded by the sensor. This involves an additional term in the image formation expression eqn. 3.2. The inbuilt inverse imaging in these cameras, removes the effect of this linear operator and for all practical purposes we can consider these as imaging systems with an unconventional depth of field. However, in practice there will always be a difference in the recovered image due to imperfections in the materials used to generate the modified PSF. Empirical analyses will be required to determine the extent to which this will affect the biometric patterns imaged.

This dissertation has summarized the essential concepts and the pitfalls that have to be considered while designing a long range biometric imaging system, with specific emphasis on iris imaging. These concepts are generic and will aid in building new age biometric systems using existing imaging hardware as well as using emerging technologies. Future work in this area, as discussed in this chapter, will help to extend the reach of security systems in a manner that will help us identify and remove potential threats before they can cause harm to our societies. Further developments in this technology will go a long

way in aiding and protecting our men and women in uniform who put their lives on the line and guard our way of life day in and day out.

Bibliography

- [1] “Information technology – Biometric data interchange formats – Part 6: Iris image data,” ISO 19794-6, 2011.
- [2] “Lg IrisAccess available at,” <http://www.lgiris.com/ps/products/irisaccess4000.htm>. 1.1
- [3] “PIER-T iris Acquisition System available at,” <http://www.l1id.com/pages/49-pier-t>.
- [4] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W. Zhao, “Iris on the Move: Acquisition of Images for Iris Recognition in less constrained environments,” *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1936–1947, nov. 2006. 1.1
- [5] K. Seshadri and M. Savvides, “Robust modified active shape model for automatic facial landmark annotation of frontal faces,” in *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, Sept 2009, pp. 1–8.
- [6] X. P. Burgos-Artizzu, P. Perona, and P. Dollár, “Robust face landmark estimation under occlusion,” in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ser. ICCV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 1513–1520. [Online]. Available: <http://dx.doi.org/10.1109/ICCV.2013.191>
- [7] “IrisAccess iCAM available at,” www.irisid.com. 1.1

- [8] “Irishield-usb-mk2120u available at,” <http://www.iritech.com/products>. 1.1, 1.2
- [9] L. Masek and P. Kovesi, “MATLAB Source Code for a Biometric Identification System Based on Iris Patterns,” The School of Computer Science and Software Engineering, The University of Western Australia., 2003.
- [10] “Biometrics Play Important Role in Access Control Market,” <http://findbiometrics.com/biometrics-play-important-role-in-access-control-market/>. 1
- [11] “New Biometric Technology Improves Security and Facilitates U.S. Entry Process for International Travelers,” https://www.dhs.gov/xlibrary/assets/usvisit/usvisit_edu_10-fingerprint_consumer_friendly_content_1400_words.pdf. 1
- [12] A. N. Al-Raisi and A. M. Al-Khoury, “Iris recognition and the challenge of homeland and border control security in UAE,” *Telematics and Informatics*, vol. 25, no. 2, pp. 117 – 132, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0736585306000360> 1
- [13] “Multi-biometric e-Border in the UAE,” <http://www.planetbiometrics.com/article-details/i/2645/desc/morpho-signs-biometric-eborder-contract-with-the-uae/>. 1
- [14] “The Future of Corrections: How Can Mobile Biometric Technology Revolutionize the Arrest and Booking Process?” http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display_arch&article_id=3585&issue_id=122014. 1
- [15] P. Grother, M. Ngan, G. W. Quinn, W. Salamon, J. R. Matey, G. Fiumara, and C. Watson, “IREX III - Performance of Iris Identification Algorithms,” National Institute of Standards and Technology, Tech. Rep., April 2012.
- [16] G. W. Quinn, P. Grother, and M. Ngan, “IREX IV: Part I - Evaluation of Iris Identification Algorithms,” National Institute of Standards and Technology, Tech. Rep., August 2013. 1

- [17] “Applications of Automatic Number Plate Recognition,” http://www.anpr.net/anpr_09/anpr_applicationareas.html. 1.1
- [18] “HIIDE Series 4: Hand-held Interagency Identity Detection Equipment,” <http://www.envimet.com/en/security/biometrics-secure-credentialing/biometrics/mobile-id-hiide-4.html>. 1.1
- [19] “PIER-T Portable Iris Enrollment and Recognition Device,” <http://www.envimet.com/en/security/biometrics-secure-credentialing/biometrics/mobile-id-pier-t.html>. 1.1
- [20] “IriMagic Series,” <http://www.irittech.com/products/hardware/irimagic%E2%84%A2-series>. 1.1
- [21] “OKI Irispass available at,” <http://www.oki.com/en/iris/>. 1.1
- [22] G. Guo, M. Jones, and B. P., “A System for Automatic Iris Capturing,” Mitsubishi Electric Research Laboratory <http://www.merl.com/publications/TR2005-044/>, 2005. 1.1
- [23] F. Wheeler, A. Perera, G. Abramovich, B. Yu, and P. Tu, “Stand-off iris recognition system,” in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, 29 2008-oct. 1 2008, pp. 1–7. 1.1, 1.1
- [24] S. Venugopalan and M. Savvides, “Unconstrained iris acquisition and recognition using cots ptz camera,” *EURASIP J. Adv. Signal Process*, vol. 2010, pp. 38:1–38:20, February 2010. [Online]. Available: <http://dx.doi.org/10.1155/2010/938737> 1.1, 1.1
- [25] F. Bashir, P. Casaverde, D. Usher, and M. Friedman, “Eagle-eyes: A system for iris recognition at a distance,” in *Technologies for Homeland Security, 2008 IEEE Conference on*, may 2008, pp. 426–431. 1.1
- [26] “Iris at a Distance,” <http://www.morpho.com/identification/news-and-publications-686/iris-at-a-distance-the-power-behind-the-iris>. 1.1

- [27] “iCAM D1000,” <http://www.irisid.com/productssolutions/hardwareproducts/icamd1000/>. 1.1
- [28] “Terrain Biometrics Insight DUO,” <http://www.terrainbiometrics.com/high-throughput/products/insight-duo/>. 1.1
- [29] P. Phillips, K. Bowyer, P. Flynn, X. Liu, and W. Scruggs, “The Iris Challenge Evaluation 2005,” in *IEEE Int’l Conf. on Biometrics: Theory, Applications and Systems*, Oct. 2008, pp. 1–8. 1.2
- [30] “iCAM7000 available at,” <http://www.irisid.com/irisaccess7000>. 1.2
- [31] “Canon EOS 5D Mark II available at,” http://www.usa.canon.com/cusa/consumer/products/cameras/slr_cameras/eos_5d_mark_ii.
- [32] “Canon EOS 70D available at,” http://www.usa.canon.com/cusa/consumer/products/cameras/slr_cameras/eos_70d.
- [33] S. Ray, *Applied Photographic Optics: Lenses and Optical Systems for Photography, Film, Video, Electronic and Digital Imaging*. Focal, 2002. [Online]. Available: <http://books.google.com/books?id=cuzYl4hx-B8C>
- [34] “Nikon D700 available at,” <http://www.nikonusa.com/en/Nikon-Products/Product-Archive/dslr-cameras/D700.html>.
- [35] “Canon EOS 1D Mark IV available at,” http://usa.canon.com/cusa/consumer/products/cameras/slr_cameras/eos_1d_mark_iv.
- [36] “Nikon D7100 available at,” <http://www.nikonusa.com/en/Nikon-Products/Product/dslr-cameras/1513/D7100.html>.
- [37] J. W. Goodman, *Introduction to Fourier Optics*. Roberts and Company Publishers, 2005.
- [38] “Canon Digital Camera Software Development Kit available at,” http://www.usa.canon.com/cusa/consumer/standard_display/sdk_homepage.

- [39] “Infra-Red LED Panel 850nm available at,” <http://www.larsonelectronics.com/p-47335-infrared-led-light-emitter-40-leds-120-watts-850-or-940-nm-1300l-x-180w-spot-beam-9-42v.aspx>.
- [40] “Birger engineering,” <http://www.birger.com/>.
- [41] “Laser Distance Sensor LDS 30 available at,” http://www.jenoptik.com/internet_en_laser-distance-sensor-lds30.
- [42] S. Venugopalan, U. Prasad, K. Harun, K. Neblett, D. Toomey, J. Heyman, and M. Savvides, “Long Range Iris Acquisition System for Stationary and Mobile Subjects,” in *Biometrics (IJCB), 2011 International Joint Conference on*, Oct. 2011, pp. 1–8. 2.1
- [43] C. Boyce, A. Ross, M. Monaco, L. Hornak, and X. Li, “Multispectral iris analysis: A preliminary study51,” in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, june 2006, p. 51.
- [44] “Infra-Red LED 850nm available at,” <http://www.sparkfun.com/products/9469>.
- [45] “PointGrey Flea3 GiGE camera,” <http://www.ptgrey.com/flea3-usb3-vision-cameras>.
- [46] “OpenCV - Open Source Computer Vision,” <http://opencv.org/>.
- [47] Y. hui Li, “Robust Long Range Iris Recognition from Video using Super Resolution,” Ph.D. dissertation, Language Technology Institute, Carnegie Mellon University, 2010.
- [48] J. Daugman, “Probing the Uniqueness and Randomness of Iriscodes: Results from 200 billion Iris Pair Comparisons,” *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1927–1935, nov. 2006.
- [49] Y. hui Li and M. Savvides, “A Pixel-Wise, Learning based approach for Occlusion Estimation of Iris Images in Polar Domain,” in *ICASSP, 2009*, pp. 1357–1360.
- [50] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by Simulated Annealing,” *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

- [51] B. Hajek, “Cooling Schedules for Optimal Annealing,” *Mathematics of Operations Research*, vol. 13, no. 2, pp. pp. 311–329, 1988.
- [52] U. Faigle and W. Kern, “Note on the Convergence of Simulated Annealing Algorithms,” *SIAM journal on control and optimization*, vol. 29, no. 1, pp. 153–159, 1991.
- [53] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, Nov 2010.
- [54] “Overview of JPEG2000,” <http://www.jpeg.org/jpeg2000/index.html>.
- [55] J.-L. Starck, E. Candes, and D. Donoho, “The curvelet transform for image denoising,” *Image Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 670–684, Jun 2002.
- [56] D. L. Donoho and I. M. Johnstone, “Ideal denoising in an orthonormal basis chosen from a library of bases,” *Comptes Rendus Acad. Sci., Ser. I*, vol. 319, pp. 1317–1322, 1994.
- [57] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, Dec 2006.
- [58] D. Donoho, M. Elad, and V. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6–18, Jan 2006.
- [59] J. Tropp, “Greed is good: algorithmic results for sparse approximation,” *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, Oct 2004.
- [60] J.-J. Fuchs, “On sparse representations in arbitrary redundant bases,” *Information Theory, IEEE Transactions on*, vol. 50, no. 6, pp. 1341–1344, June 2004.
- [61] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation,” *IEEE Transactions on Signal Pro-*

- cessing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [62] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson Education, 2002.
- [63] “Implementation of the K-SVD and Approximate K-SVD dictionary training algorithms,” <http://www.cs.technion.ac.il/~ronrubin/software.html>.
- [64] “VeriEye SDK,” <http://www.neurotechnology.com/verieye.html>.
- [65] <https://shop.360fly.com/products/360fly> last accessed on 5/19/2015.
- [66] E. Adelson and J. Wang, “Single lens stereo with a plenoptic camera,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 99–106, Feb 1992. 2.2
- [67] F. Okano, J. Arai, H. Hoshino, and I. Yuyama, “Three-dimensional video system based on integral photography,” *Optical Engineering*, vol. 38, no. 6, pp. 1072–1077, 1999. [Online]. Available: <http://dx.doi.org/10.1117/1.602152> 2.2
- [68] T. Naemura, T. Yoshida, and H. Harashima, “3-d computer graphics based on integral photography,” *Opt. Express*, vol. 8, no. 4, pp. 255–262, Feb 2001. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-8-4-255> 2.2
- [69] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, “High performance imaging using large camera arrays,” *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005. [Online]. Available: <http://doi.acm.org/10.1145/1073204.1073259> 2.2
- [70] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, “Light Field Photography with a Hand-Held Plenoptic Camera,” Tech. Rep., Apr. 2005. [Online]. Available: <http://graphics.stanford.edu/papers/lfcamera/> 2.2
- [71] “Lytro Cameras,” <https://www.lytro.com/>. 2.2
- [72] R. Ng, “Fourier slice photography,” *ACM Trans. Graph.*, vol. 24, no. 3, pp. 735–744,

- Jul. 2005. [Online]. Available: <http://doi.acm.org/10.1145/1073204.1073256> 2.2
- [73] E. R. Dowski, Jr. and G. E. Johnson, “Wavefront coding: a modern method of achieving high-performance and/or low-cost imaging systems,” vol. 3779, pp. 137–145. [Online]. Available: <http://dx.doi.org/10.1117/12.368203> 2.2
- [74] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, “Image and depth from a conventional camera with a coded aperture,” *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007. [Online]. Available: <http://doi.acm.org/10.1145/1276377.1276464> 2.2
- [75] “Pentax 645Z Medium Format Camera,” <http://www.imaging-resource.com/PRODS/pentax-645z/pentax-645zA.HTM>.