

Harnessing Auxiliary Information: New Methods to Improve Person Identification

By

Samarth Bharadwaj

Advisors

Mayank Vatsa, PhD

Richa Singh, PhD

Afzel Noore, PhD (External)

Thesis submitted to the
Indraprastha Institute of Information Technology-Delhi (IIIT-Delhi)
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science & Engineering
July, 2015

Keywords: Biometric Quality, Online Context Switching, Quality based Enhancement, Social Context aided Face Recognition, Distance Metric Learning.

Copyright© S. Bharadwaj, 2015

Abstract

Large scale biometric identification systems still lack the versatility to handle challenging situations such as adverse imaging conditions, missing or corrupt data, and non-conventional operating scenarios. It is well understood that in different operating conditions, evidence of identity obtained from different sources is disparate. In such cases, additional ‘situational’ cues can be utilized to improve the performance and robustness. The primary emphasis of this thesis is the formulation of new methods to utilize situational cues such as quality of input biometric samples, social cues of co-occurrence, and other background information towards more inclusive biometric systems.

Biometric sample quality assessment during capture and its integration into the recognition system improves performance and reduces the failure-to-enroll rates. The first contribution of this thesis is an in-depth survey along with statistical evaluation of different concepts and interpretations of biometric quality in multiple biometric modalities. The thesis also investigates the effectiveness of holistic representations of faces for classifying them into different quality categories that are derived from matching performance. The experiments on the CASPEAL and SCFace databases containing covariates such as illumination, expression, pose, low-resolution, and occlusion, suggest that the representations can efficiently classify input face images into relevant quality categories and be utilized in face recognition systems. An assessment based quality enhancement framework is also presented that showcases the effectiveness of quality assessment metrics for parameter selection in a denoising method to enhance performance and reduce computational time.

Multi-modal biometric recognition systems combine evidence from multiple sources of information for improving the recognition performance. Existing multi-modal biometric recognition techniques are, however, unable to provide required levels of accuracy in uncontrolled noisy capture environments. Such algorithms do not adequately scale to variations in data distribution that occur due to changing deployment conditions. The second contribution of this thesis is an adaptive context switching algorithm coupled with online learning to address both these challenges of multimodal biometrics. The proposed framework uses the quality of input images to dynamically select the best biometric matcher or fusion algorithm to verify the identity of an individual. The proposed algorithm continuously updates the selection process using online learning to address the scalability and accommodate the variations in data distribution. The results on the WVU multimodal database and a large real world multimodal database obtained from a law enforcement agency show the efficacy of the proposed framework.

Humans are efficient at recognizing familiar faces even in challenging conditions by deducing social context between individuals in group photos. The identity of the person in a photo, in such cases, is inferred based on other individuals present in the same photo; using the known or deduced social context between them. The third contribution of the thesis is a novel algorithm to utilize co-occurrence of individuals as the social context to improve face recognition. Association rule mining is utilized to infer multi-level social context among subjects from a large repository of social transactions. The results are demonstrated on the G-album and on the real-world SN-collection pertaining to 4675 identities that is prepared for the purpose of this research from a social networking website. An anonymized version of the dataset with match scores from a commercial system is also made available. The results of the proposed approach show that association rules extracted from social context can be used to augment face recognition and improve the identification performance.

The availability of a large number of unlabelled images from various sources facilitates semi-supervised approaches to improve the performance and robustness of recognition systems. As the fourth contribution, this thesis introduces a novel learning based approach to face recognition towards an affordable and friendly biometric for *newborns*. Biometric recognition of newborns is an opportunity for the realization of several useful applications such as improved security against swapping and abduction, accurate census and effective drug delivery. The proposed approach couples learning based encoding method via deep neural networks with a one shot similarity distance metric formulated with an online SVM to match effective features with low semantic gap. To evaluate the approach, the largest publicly available database of 96 newborns is collected from various hospitals to study face recognition and is also made available to other researchers. Several existing face recognition approaches and commercial systems are also evaluated on a common benchmark protocol. The proposed approach provides state-of-the-art identification and verification performance on the newborns database.

I dedicate this thesis to my late maternal grandmother, Dr. T. Mangamamba, and to my paternal grandmother, T. Sitamahalakshmi

Acknowledgements

“There are wavelengths that people cannot see, there are sounds that people cannot hear, and maybe computers have thoughts that people cannot think”

— *Richard Hamming*

The journey of this unlikely PhD candidate at IIIT-Delhi has been as much an internal pursuit of self as the pursuit for knowledge that it is meant to be. During the course of this journey, I was blessed with several interesting people who, as Douglas Adams put it and I audaciously rephrase, met me by design as though human connections are exempt from chance and the laws of Physics that govern all other things in this universe. I would like to take this opportunity to express my deepest sense of gratitude to each of them.

Ever since I first met Dr. Mayank and Dr. Richa, I felt a sense of purpose and determination about them that still motivates me. Being privy to some of the circumstances under which they have achieved their success and accolades, their spirit and rigor in converting opportunities to results is truly inspiring. Above all other qualities is their mentoring that has constantly encouraged me to pursue my intuitions and to stiffen my sinews. I continue to benefit from their sound advice and also their hearty friendship. I would like to express my immense appreciation to Prof. Afzel Noore for the opportunity to visit West Virginia University and to work with him. Witnessing his ability to apply critical thinking in solutions to problems, his wisdom and points-of-view on research and academia, are treasures that I brought back and will forever hold dear.

I would like to acknowledge Dr. A. Martinez, Dr. M. Grgic, Dr. R. Gross, Prof. S.G. Shan, Dr. A. Gallagher, and Dr. S. K. Singh for granting access to various datasets. I also thank the Department of Information Technology,

Government of India for their support. I would also like to thank the faculty and staff at IIIT-Delhi for all the help and feedback. I also appreciate my team at IBM Research and look forward to exciting collaborations.

My uncle and aunt, T. Sudhakar and T. Shailaja, took me in as nothing less than their third and most demanding son, I feel blessed to have family in Delhi. Your thoughts on an extraordinary range of subjects and your work is a continuous inspiration to me, dear *Babai*, and I am also extremely sorry for denting your car! My family Dr. T. Rambabu, Dr. G. Swarnabala, and dear sweet little PhD candidate Anahita Bharadwaj, your love and support is unbound, and cannot be captured in words. My love to my dearest and to the gang, particularly Himanshu, Kuldeep, Tejas, and Anush for all the good times and for also listening and lending your more evolved brains to bounce-off and filter my thoughts. How I would have stuck at it through the years without all of your support is beyond me. I hope this thesis convinces me of my own abilities and leads me towards doing good work. I sincerely apologize to all those who have suffered for my faults and shortcomings.

Contents

Dissemination of Research Results	1
List of Figures	3
List of Tables	11
1 Introduction	13
1.1 Overview of Biometrics	13
1.2 Factors that Improve Biometric Performance	16
1.2.1 Quality of a Biometric Sample	17
1.2.2 Adaptive Multibiometrics	17
1.2.3 Harnessing Auxiliary Information	18
1.3 Research Contributions	18
2 Biometric Quality Assessment	21
2.1 Introduction	22
2.1.1 Quality Assessment During Enrolment	24
2.1.2 Quality Assessment During Recognition	25
2.2 Biometric Quality: Factors, Degradations, and Features	28
2.2.1 Factors that Influence Biometric Quality	29
2.2.2 Degradations in Biometric Images	30
2.2.2.1 Image-based Degradations	30
2.2.2.2 Biometric-modality-specific Degradations	32
2.2.3 Image-based Features	32
2.2.4 Naturality, Fidelity, and Utility in Biometric Quality	34
2.3 Literature Review: Quality Assessment in Fingerprint, Iris, and Face	35
2.3.1 Fingerprint Quality Assessment	36
2.3.2 Iris Quality Assessment	39
2.3.3 Face Quality Assessment	41

2.3.3.1	Still-face Images-based Techniques	42
2.3.3.2	Video-based Techniques	44
2.3.4	Evaluating Quality Assessment Approaches	46
2.4	Analysis of Quality Metrics	47
2.4.1	Database and Evaluation Protocol	47
2.4.2	Experimental Analysis	50
2.5	Can Holistic Representations be used for Biometric Quality Assessment? . .	57
2.5.1	Quality Assessment of Face Biometric	58
2.5.1.1	Face Quality as Match Score Predictor	59
2.5.1.2	Holistic Image Representations	60
2.5.1.3	Quality Labels based on Face Matcher	62
2.5.2	Experiment and Analysis	63
2.5.2.1	Experiments with Noisy Images	66
2.5.2.2	Experiments with Pose	67
2.6	Quality Assessment based Denoising to Improve Recognition Performance .	69
2.6.1	Image Denoising with BayesShrink	72
2.6.1.1	Candidate Parameter Set	73
2.6.2	Proposed Quality Assessment based Denoising Framework	73
2.6.2.1	Training	74
2.6.2.2	Testing	74
2.6.3	Experimental Results	75
2.6.3.1	Recognition Experiment	75
2.7	Discussion	77
3	QFuse: Online Learning Framework for an Adaptive Biometric System	81
3.1	Introduction	81
3.2	Literature Review	84
3.3	QFuse: Quality Based Context Switching with Online Learning	86
3.3.1	Quality Assessment	88
3.3.2	Context Switching Algorithm	90
3.3.2.1	Training the SVMs	90
3.3.2.2	Online Learning	91
3.3.2.3	Context Switching during Verification	94
3.4	Experimental Results	95
3.4.1	Unimodal Matchers	95

3.4.2	Database and Experimental Protocol	95
3.4.3	Results and Analysis	96
3.5	Summary	103
4	Aiding Face Recognition with Social Context Association Rule based Re-Ranking	105
4.1	Introduction	105
4.1.1	Related Work	106
4.1.2	Research Contributions	108
4.2	Social Context Aided Face Recognition	109
4.2.1	Building Context from Tagged Faces	109
4.2.2	Social Context Association Rules from Tagged Faces	111
4.2.3	Context Propagation and Re-Ranking	112
4.3	Database and Protocol	114
4.3.1	The G-album	114
4.3.2	SN-collection	114
4.4	Experimental Results and Analysis	116
4.5	Summary	119
5	Learning Based Encoding and Distance Metric Approach to Newborn Face Recognition	121
5.1	Introduction	121
5.1.1	Related Work	123
5.1.2	Research Contributions	124
5.2	Newborn Faces: Characteristics and Database	126
5.2.1	Characteristics of Newborn Face	126
5.2.2	Newborn Face Database and Challenges	127
5.3	Learning based Encoding and Distance Metric Approach	129
5.3.1	Preliminaries	129
5.3.1.1	Stacked Denoised Autoencoder	130
5.3.1.2	Learning based Distance Metrics: One Shot Similarity using LDA	131
5.3.2	Proposed Approach	134
5.3.2.1	Domain Specific Representation via SDAE	134
5.3.2.2	Problem Specific Distance Metric Learning via One Shot Similarity with 1-class-online-SVM	135

5.4	Performance Evaluation and Analysis	137
5.4.1	Experimental Protocol	138
5.4.2	Results of Experiment 1: Benchmarking Existing Feature Extractors	139
5.4.3	Experiment 2: Results of the Proposed Algorithm	141
5.4.4	Experiment 3: Effectiveness of Problem Specific Information	144
5.5	Summary	144
6	Conclusions and Future Work	147
A	Image Quality Assessment	151
B	Biometric Standards	157
	Bibliography	159

Dissemination of Research Results

Journals

- J.1 **S. Bharadwaj**, H.S. Bhatt, R. Singh, and M. Vatsa, *Face Recognition for Newborns based on Learning based Representations and Distance Metrics*, under review in TIFS.
- J.2 **S. Bharadwaj**, H.S. Bhatt, R. Singh, and M. Vatsa, *Online Context Switching Algorithm for Adaptive Biometric System*, Elsevier Pattern Recognition (PR), 2015.
- J.3 **S. Bharadwaj**, M. Vatsa, and R. Singh, *Biometric Quality: A Review of Fingerprint, Iris, and Face*, EURASIP Journal on Image and Video Processing, 2014(1), 34.
- J.4 H.S. Bhatt, **S. Bharadwaj**, R. Singh, and M. Vatsa, *Recognizing Surgically Altered Face Images using Multi-objective Evolutionary Algorithm*, IEEE Transactions on Information Forensics and Security (TIFS), Vol. 8, No. 1, pp. 89-100, 2013.
- J.5 H.S. Bhatt, **S. Bharadwaj**, R. Singh, and M. Vatsa, *Memetic Approach for Matching Sketches with Digital Face Images*, IEEE Transactions on Information Forensics and Security (TIFS), Vol. 5, No. 5, pp. 1522-1535, 2012.
- J.6 R. Singh, M. Vatsa, H.S. Bhatt, **S. Bharadwaj**, A. Noore and S.S. Nooreydzan, *Plastic Surgery: A New Dimension to Face Recognition*, IEEE Transactions on Information Forensics and Security (TIFS), Vol. 5, No. 3, pp. 441-448, 2010.

Peer-reviewed Conferences

- C.1 **S. Bharadwaj**, M. Vatsa, and R. Singh, *Aiding Face Recognition via Social Context Association*, In Proceedings of IEEE/IAPR Joint Conference on Biometrics (IJCB), 2014.
- C.2 G. Goswami, **S. Bharadwaj**, M. Vatsa, and R. Singh, *On RGB-D Face Recognition using Kinect*, In Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2013. (**Best Poster Award**)
- C.3 **S. Bharadwaj**, M. Vatsa, and R. Singh, *Can Holistic Representations be used for Face Biometric Quality Assessment?*, In Proceedings of IEEE International Conference on Image Processing (ICIP), 2013. (*Oral paper*).
- C.4 **S. Bharadwaj**, T. Dhamecha, M. Vatsa, and R. Singh, *Computationally Efficient Face Spoofing Detection with Motion Magnification*, In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2013.
- C.5 H.S. Bhatt, **S. Bharadwaj**, R. Singh, M. Vatsa, A. Noore and A. Ross, *Quality Driven Biometric Classifier Selection Framework for Improved Performance*, In Proceedings of IEEE International Joint Conference on Biometrics (IJCB), 2011.
- C.6 H.S. Bhatt, **S. Bharadwaj**, R. Singh, M. Vatsa, A. Noore and A. Ross, *On Co-training Online Biometric Classifiers*, In Proceedings of IEEE International Joint Conference on Biometrics (IJCB), 2011 (**Best Poster Award**).
- C.7 **S. Bharadwaj**, H.S. Bhatt, M. Vatsa, and R. Singh, *Quality Assessment based Denoising to Improve Face Recognition Performance*, In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2011.
- C.8 H.S. Bhatt, **S. Bharadwaj**, R. Singh and M. Vatsa, *Evolutionary Granular Computing Approach for Recognizing Face Images Altered due to Plastic Surgery*, Face & Gesture, 2011.

- C.9 H.S. Bhatt, **S. Bharadwaj**, R. Singh and M. Vatsa, *On Matching Sketches with Digital Face Images*, In Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2010.
- C.10 **S. Bharadwaj**, H.S. Bhatt, R. Singh, M. Vatsa, and S.K. Singh, *Face Recognition for Newborns: A Preliminary Study*, In Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2010 (**Best Poster Award**).
- C.11 M. Vatsa, R. Singh, **S. Bharadwaj**, H.S. Bhatt, and A. Noore, *Matching Digital and Scanned Face Images with Age Variation*, In Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2010.
- C.12 **S. Bharadwaj**, H.S. Bhatt, M. Vatsa, and R. Singh, *Periocular Biometrics: When Iris Recognition Fails*, In Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2010.
- C.13 M. Vatsa, R. Singh, A. Tiwari, **S. Bharadwaj**, and H.S. Bhatt, *Analyzing Fingerprint of Indian Population using Image Quality: A UIDAI Case Study*, International Workshop on Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB), 2010.
- C.14 H.S. Bhatt, **S. Bharadwaj**, R. Singh, and M. Vatsa, *Face Recognition and Plastic Surgery: Social, Ethical and Engineering Challenges*, In Proceedings of International Conference on Ethics and Policy of Biometrics and International Data Sharing (ICEB), 2010.

Other Publications

- O.1 H.S. Bhatt, **S. Bharadwaj**, R. Singh, and M. Vatsa, *Plastic Surgery and Face Recognition*, Encyclopedia of Biometrics, 2nd Edition, 2013.
- O.2 **S. Bharadwaj**, T.I. Dhamecha, M. Vatsa, and R. Singh, *Face Anti-spoofing via Motion Magnification and Multifeature Videolet Aggregation*, IIITD-TR-2014-002, 2014.
- O.3 H.S. Bhatt, **S. Bharadwaj**, R. Singh, and M. Vatsa, *Recognizing Surgically Altered Face Images using Multi-objective Evolutionary Algorithm*, IIITD-TR-2011-006, 2013.

List of Figures

1.1	Several modalities have emerged in biometrics, however this thesis chiefly discusses three prominent modalities: fingerprint, iris and face.	14
1.2	Pipeline of a typical biometric system consists of a capture sequence (probe), detection & pre-processing, feature extraction, matching and decision modules.	14
1.3	Various sources available for combining evidence; multi-sample, multi-modal, multi-instance, multi-algorithm, and multi-sensor.	15
2.1	Variation in quality. A biometric system may encounter samples of a wide range of quality. Effective quality assessment metrics that are indicative of these variations are therefore essential to an automated biometric system. .	22
2.2	Image quality vs biometric quality. While the images in (a) are of poor <i>image</i> quality, the images in (b) may have lower <i>biometric quality</i>	23
2.3	Pipeline of a typical biometric system. This consists of a capture sequence (probe), detection and preprocessing, feature extraction, matching and decision modules.	25
2.4	Utilizing biometric quality assessment for context switching. Framework for (a) a quality-driven biometric image enhancement, based on [23], and (b) quality-based multiclassifier selection, proposed in [29].	26
2.5	Sample images of varying quality. (a) Fingerprint, (b) iris (from WVU multimodal database), and (c) face (from SCface and CAS-PEAL face databases) illustrating the wide range of quality that a biometric system can encounter with different image and biometric specific degradations. . .	31
2.6	Four image features are primarily used for estimating quality of biometric images. Orientation, intensity statistics, power spectrum, and wavelet transform.	33

2.7	Three aspects of quality assessment: naturality, fidelity, and utility, in a typical biometric pipeline.	35
2.8	Poor quality fingerprint samples often lead to spurious minutia.	36
2.9	A fingerprint image (a) and corresponding Fourier transform (magnitude component after shifting) (b). The ridge information manifests as a bright band. Chen <i>et al.</i> [40] use the difference of two Butterworth filters to obtain a soft bandpass filter that captures the strength (and thereby quality) of the ridges.	37
2.10	Samples of poor iris segmentation on images obtained from CASIA-V4 iris database.	41
2.11	Face images illustrating different levels of biometric quality.	42
2.12	Genuine and imposter score distribution for (a) face, (b) fingerprint, and (c) iris matchers on the WVU multimodal dataset. (d) Receiver Operating Characteristic (ROC) curve illustrates the verification performance of the respective matchers indicating the overall quality of the database.	48
2.13	Relation between match scores obtained from NBIS fingerprint matcher and various quality metrics. Relation between match scores obtained from NBIS fingerprint matcher (z -axis) and various quality metrics (a) SE, (b) ES, (c) A, (d) B, (e) Z, (f) NFIQ, (g) RE, (h) BR for genuine (green) and imposter (red) match pairs. The x -axis pertains to gallery quality, while y -axis pertains to the probe quality. The scattering indicates that ES, A, B, Z, RE, and BR quality metrics can characterize genuine scores.	52
2.14	Relation between match scores obtained from a commercial face matcher and various quality metrics. Relation between match scores obtained from a commercial face matcher (z -axis) and various quality metrics [(a) SE, (b) ES, (c) A, (d) B, (e) Z, (f) P, (g) F, (h) BR] for genuine (green) and imposter (red) match pairs. The x -axis pertains to gallery quality, while y -axis pertains to probe quality. The scatterplot indicates that A, B, Z, F, and BR quality metrics can characterize genuine scores.	53

2.15	Relation between match scores obtained from a commercial iris matcher and various quality metrics. Relation between match scores obtained from a commercial iris matcher (z -axis) and various quality metrics [(a) SE, (b) ES, (c) A, (d) B, (e) Z, (f) DF, (g) MB, (h) O, (i) I, (j) SR, (k) PC, (l) Q, (m) BR] for genuine (green) and imposter (red) match pairs. The x -axis pertains to gallery quality while y -axis pertains to probe quality. The scatterplot indicates that ES, A, B, Z, SR, PC, and BR quality metrics can characterize genuine scores of match pairs. However, DF, O, I, and Q are unable to characterize genuine match scores.	54
2.16	The cumulative density functions (CDF) between genuine score and quality metrics for (a) face, (b) fingerprint, and (c) iris modalities. The plots compare the distribution of each quality metric with the corresponding genuine score distribution.	56
2.17	Results of the regression test. MSE of the regression test with genuine scores and quality metrics accumulated over 10 times cross-validation. Even with a small number of training samples, a linear model can predict match scores of genuine pairs showing that quality scores can be indicative of matching performance.	57
2.18	Face images of varying quality encountered by a face recognition systems. .	58
2.19	Face images with degradations exhibit more <i>roughness</i> , evident from the surface plots (z -axis is pixel intensity (I)). Roughness can be captured with holistic features and may be indicative of biometric quality.	60
2.20	The training process of the proposed approach.	62
2.21	(a) The empirical cumulative density function (ECDF) of the z -normalized match scores (-20 to 160), (b) the number of samples per quality bin obtained for training.	63
2.22	Verification performance of testing images when segregated into quality bins (left) and when lower quality bins are discarded (right) using a) Gist and b) HOG.	64
2.23	Sample images of four quality bins obtained from the proposed approach (common to both Gist and HOG).	65

2.24	a) Confusion matrix of the performance of GIST descriptor b) Accuracy of SVM in classifying each of the eight noise/blur artifacts and 1 uncorrupt class labels. Labels are defined as 1- Gaussian blur, 2- Motion blur, 3- Uncorrupt (original), 4- White noise, 5- Localvar noise, 6- Poisson noise, 7-Salt & Pepper noise, 8- Speckle noise, and 9- Sharp.	67
2.25	a) Frontal illumination images of a subject in all nine viewpoints from the MultiPIE dataset, b) a bar graph of the HOG and SVM pose estimator, and c) similar graph with PHOG descriptor.	68
2.26	a) All nine poses from the SC-Face dataset for a single subject, b) a bar graph of the HOG and SVM pose estimator on the SC-Face dataset, and c) A similar bar graph with PHOG and SVM	69
2.27	Irregularities due to different types of noise degrade the quality of face images significantly. Sample images from the AR face dataset with synthetic noise.	70
2.28	CMC curve of local binary pattern (LBP) + χ^2 matcher: Identification performance decreases when noise is added to the probe images	71
2.29	The training scheme of the proposed assessment based denoising framework. The process presented here is for a single noised training probe set; this process is repeated for all the five data driven noises.	73
2.30	The testing scheme of the proposed assessment based denoising framework .	75
2.31	Sample noisy (top) and denoised (bottom) images obtained using the proposed framework.	76
2.32	Scatter plot of the quality score of training data. The illustrated class labels correspond to the best parameters selected for denoising. This indicates that images with a certain set of quality scores require a specific parameter for best denoising.	76
2.33	CMC of LBP based face recognition using proposed parameter selection framework and each parameters without selection. The selection framework slightly improves performance and reduces computation time.	76
3.1	Illustrating examples of multimodal images from the same subject with varying quality (a) fingerprint, (b) iris (the last two rows demonstrate the unwrapped iris images and occlusion mask that indicate iris feature occlusion), and (c) face images.	82

3.2	General concept of the proposed QFuse algorithm that selects a single information source or fusion of multiple sources based on its reliability and the quality of information.	83
3.3	Sample images and their corresponding quality assessment scores for (a) face, (b) fingerprint, and (c) iris images. These quality scores are obtained after segmentation and utilized by support vector machines for context switching.	87
3.4	Illustrating the proposed quality based serial context switching algorithm. .	89
3.5	Illustrating the process of assigning labels: genuine and impostor match score distributions are used to assign labels to the input gallery-probe quality vector $Q = [Q_g, Q_p]$ during SVM training.	92
3.6	ROC curves of the proposed quality based context switching algorithm QFuse and comparison with different unimodal matchers, sum-rule fusion and Vatsa <i>et al.</i> [187] on the WVU database [45].	97
3.7	ROC curves of the proposed quality based context switching algorithm and comparison with different unimodal matchers and sum-rule fusion on the LEA database.	97
3.8	Improvement with online learning over batch/offline learning on the a) WVU multimodal database [45] and b) LEA database.	100
3.9	Illustrating sample cases where unimodal a) fingerprint, b) iris, and c) face matchers is selected.	101
3.10	Illustrating a sample case where multimodal fusion is selected.	102
4.1	In this illustration, context derived from two images help confirm identity of a face in third. <i>A professor meets a couple of PhD students, A and T, after an invited talk at a university. Later at a conference, he is unable to recognize A but infers his identity when he sees A and T seated together.</i> . .	106
4.2	Three problem domains: (a) traditional face recognition, (b) photo album organization (based on events, locations and people) with imperfect (manual) annotations and (c) reliable face tagging approaches of large number of uploaded photos, which is the focus of this work.	107

4.3	In case of challenging imaging condition, prior knowledge of association of subjects may induce additional information to improve face recognition. The proposed approach generates context association rules, $\{Michelle, Sasha\} \rightarrow Barack$ and $\{Michelle, Barack, Sasha\} \rightarrow Malia$, based on prior co-occurrence to aid face recognition.	109
4.4	An illustration of the proposed multi-level association rule mining to derive social context to improve face identification.	110
4.5	A visualization (Fruchterman Reingold layout) of 100 association rules [75] with the highest <i>lift</i> obtained from the G-album dataset [61]. Lift is defined as $\frac{\text{Supp}(X \cup Y)}{\text{Supp}(X)\text{Supp}(Y)}$, greater lift is indicative of more informative association rules. The vertices are shaded proportional to the in-degree of each rule, i.e., the number of rules the consequent participates in.	113
4.6	Mean confidence vs. cardinality of X of a social association rule ($X \rightarrow Y$) shows more confidence as $n(X)$ increases.	116
4.7	Identification accuracy obtained on a) G-Album database and b) SN-collection database.	117
4.8	Illustrating three cases from the G-album dataset where initially a face is recognized incorrectly at rank-1. With the application of the proposed context association rule based re-ranking, using the correctly recognized faces and social context, rank-1 performance is achieved for these samples. .	118
5.1	Sample images pertaining to three newborns from the Newborn Face Database illustrating large facial variations during capture. We term newborns as <i>unintentionally un-cooperative</i> users of face recognition.	123
5.2	Identification form used to capture the footprints of the newborn and fingerprint of the mother. Image is taken from http://www.amazon.com/Briggs-Newborn-Attachment-Signatures-Verification/dp/B002C18XRW	124
5.3	This research presents a two-stage learning framework using domain-specific and problem-specific unlabeled instances to perform face recognition in newborns.	125
5.4	The craniofacial characteristics of an infant face	126
5.5	Some challenging images from the IIITD Newborns face database indicate the non-cooperative nature of newborns that hinder traditional face recognition approaches.	128

5.6	Images from the newborns face database pertaining to two subjects. The figure illustrates the high intra-class variation in the images captured in two sessions posing challenges in detection and recognition.	128
5.7	Sample images of (a) twins and (b) newborns with <i>lanugo</i> , from the New-born face database.	129
5.8	One Shot Similarly: A and B are the two instances to be matched. The distance between them is computed using the classifier output obtained using each instance and the same negative background set I	132
5.9	The training phase, is performed by effectively utilizing abundant domain-specific instances and limited problem-specific instances to learn a representation and distance metric.	132
5.10	The testing phase of the proposed approach where each input pair to be matched is converted into previously learned representation with the separate encoder learnt for each patch. Next the representations are matched with the online SVM based one-shot similarity metric approach.	133
5.11	The performance of key-point detection based techniques such SURF is misleading. The detected keypoints are unreliable both in genuine and imposter pairs. This may be attributed to unstable edge information presented due to face wrinkles.	141
5.12	Experiment 2: ROC and CMC curves to evaluate the effectiveness of individual components of the proposed algorithms. The experiments are performed with four gallery images per subject. The graphs are best viewed in color.	142
5.13	Experiment 3: ROC and CMC curves of the proposed algorithm ($SDAE + OSS(SVM_{online})$) to evaluate the effectiveness of using problem specific data. The results are computed with 1000 adult face images as background samples. The graphs are best viewed in color.	143
6.1	This thesis establishes the advantages of leveraging auxiliary sources of information to enhance the performance and reliability of a biometric system. This figure illustrates the connections made in this work between the biometric recognition pipeline with four auxiliary sources, namely, biometric quality, reliability and context, social networks, and unlabeled background images obtained from the web. The jagged lines show future scope of utilizing an auxiliary source.	148

A.1	A typical Full Referenced (FR) bottom-up image quality assessment system based on error visibility.	153
A.2	Portrait image altered by various distortions. a) original image b) mean luminance shift c) contrast stretch d) impulse noise e) Gaussian noise f) blur g) compression h) spatial shift i) scale j) rotation. While most images have the same mean square error (MSE), there is s drastic difference in the visual quality. A motivating example towards SSIM quality index obtained from [198].	153

List of Tables

2.1	Different interpretations of quality in biometrics from literature	24
2.2	Various behavioral, environmental, and operational factors that effect quality of biometric sample	30
2.3	A representative list of fingerprint quality assessment algorithms	36
2.4	A representative list of iris quality assessment algorithms	40
2.5	A representative list of face quality assessment algorithms	42
2.6	Various representative quality metrics considered in this study	50
2.7	Spearman correlation between face quality scores	50
2.8	Spearman correlation between fingerprint quality scores	51
2.9	Spearman correlation between iris quality scores	51
2.10	Summary of a representative list of existing approach in face quality.	59
2.11	Summary of databases.	63
2.12	Performance of COTS on each quality bin.	65
2.13	Different noise and blur artifacts used in the experiments with their associated parameters.	66
2.14	Different noise artifacts used in the experiments with their associated parameters.	75
2.15	Average computation time for denoising an image with each image enhancement parameter and Rank-1 efficiency with testing-gallery-set. All values computed in Matlab with a Dual-core CPU and 2 GB RAM	77
3.1	Verification accuracy of individual matchers, fusion algorithms, and QFuse at 0.01% false accept rate (FAR).	100
3.2	Comparing the verification accuracy and computational time of QFuse when the training is performed in online and offline manner.	100
3.3	Illustrating percentage of instances processed by individual components involved in the proposed context switching algorithm.	102

4.1	Experimental protocol of both the databases.	115
4.2	Identification accuracy (%) on two databases.	117
5.1	Newborn recognition using different biometrics. *For newborns and toddler recognition.	122
5.2	Rank-1 identification accuracies (%) for experiment 1. The results are re- ported in terms of average accuracy with standard deviation over times cross validation. *Commercial systems rejected a portion of the samples based on quality.	139
5.3	Verification accuracy (at 0.1%FAR) and standard deviation on the newborn face database for experiment 1. *Commercial systems rejected a portion of the samples based on quality.	140

Chapter 1

Introduction

Unique identification of individuals is necessary in several application domains such as forensics, law enforcement, governance, access control, commerce and entertainment. Government agencies frequently verify the identity before providing information or resources, usually during entitlement programs, immigration, voter registration, and welfare schemes. It is also critical to establish correct identity to ensure information security in personal devices, logins, network security, secure documents or databases, and medical records. Similarly, proof of identity is required in many commercial establishments such as banks, hospitals, and shops. The entertainment industry also verifies user identity to provide user-based content and virtual or augmented reality.

Biometrics, as part of *Identification Science*, is the measurement of certain key features of physical or behavioral traits to uniquely identify a person. Figure 1.1 shows sample images from some popular biometric modalities. By intelligently integrating biometrics with correct policies and procedures, it can be deployed for large scale country-wide applications to potentially eliminate the need for paper work or ID cards as well as to cater the needs of identity management in civil and law enforcement applications. It can be expected that coming years will witness several functioning, robust, and reliable large scale biometrics systems in deployment along the lines of UAE iris based security system and India's Aadhaar project.

1.1 Overview of Biometrics

Figure 1.2 illustrates the recognition pipeline of a generic biometric system. The process is inspired by the human visual system which hierarchically extracts and processes data starting from coarse to dense features. The first step in a typical biometric system involves capture of a biometric image (source) followed by segmentation of the region of

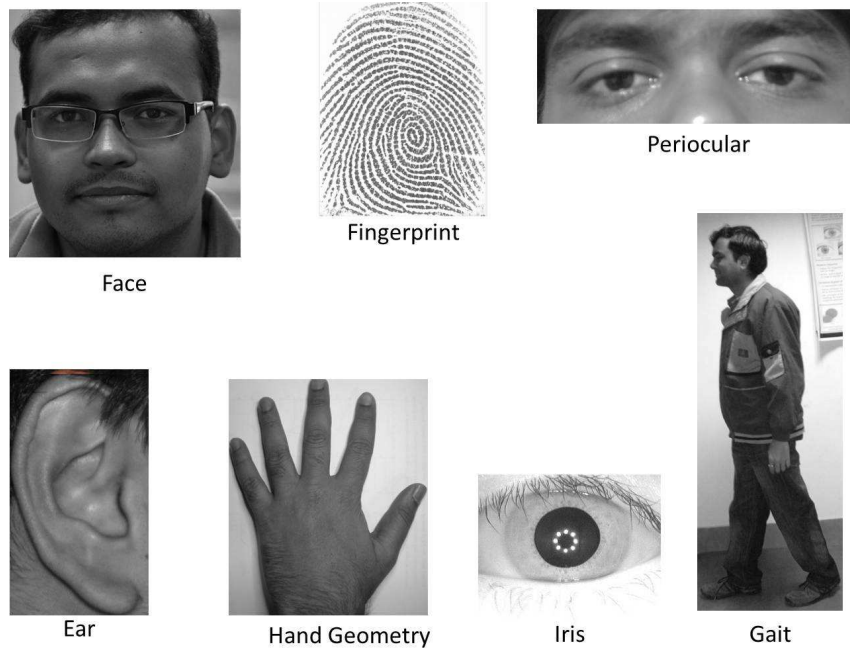


Figure 1.1: Several modalities have emerged in biometrics, however this thesis chiefly discusses three prominent modalities: fingerprint, iris and face.

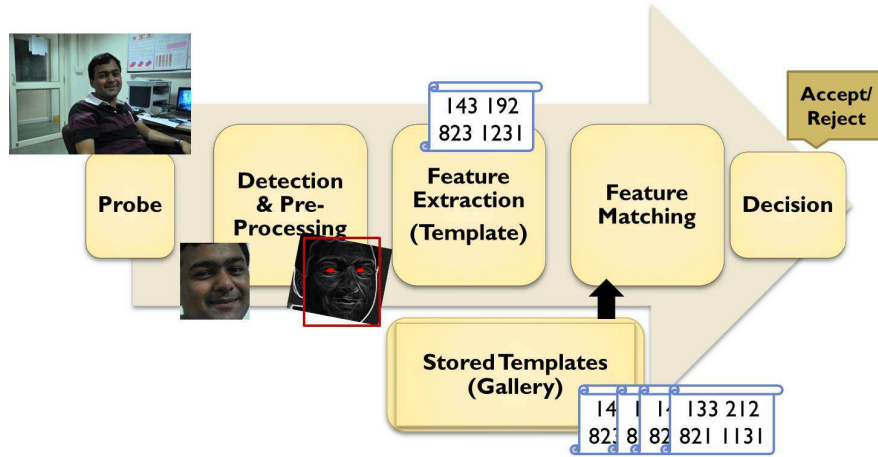


Figure 1.2: Pipeline of a typical biometric system consists of a capture sequence (probe), detection & pre-processing, feature extraction, matching and decision modules.

interest from the input image. The segmented image is preprocessed to verify and enhance the biometric trait present in the image. Next, discriminating features are extracted and matched with a stored template with identity and the matching processes produces a decision. Jain *et al.* [88] discuss seven fundamental properties of a biometric modality, namely, **Universality**: A characteristic of all/most individuals, **Distinctiveness**: Possess a sufficiently unique characteristic, **Permanence**: Sufficiently invariant over a period

of time, **Collectivity**: Measurable and easily collectable characteristic of an individual, **Performance**: Accuracy and speed of recognition, resources used, and environmental factors that affect performance, **Acceptability**: Social and personal acceptance of the system, **Circumvention**: No/limited spoofing or fraudulent methods of fooling the system. Different biometric modalities exhibit advantages and disadvantages when compared based on these characteristics. Hence, there is no *ideal* biometric modality but several *admissible* biometric traits [87]. The relevance of a biometric modality is always derived from the intent, when certain characteristics may become more important than others, leading towards combining multiple biometric systems.

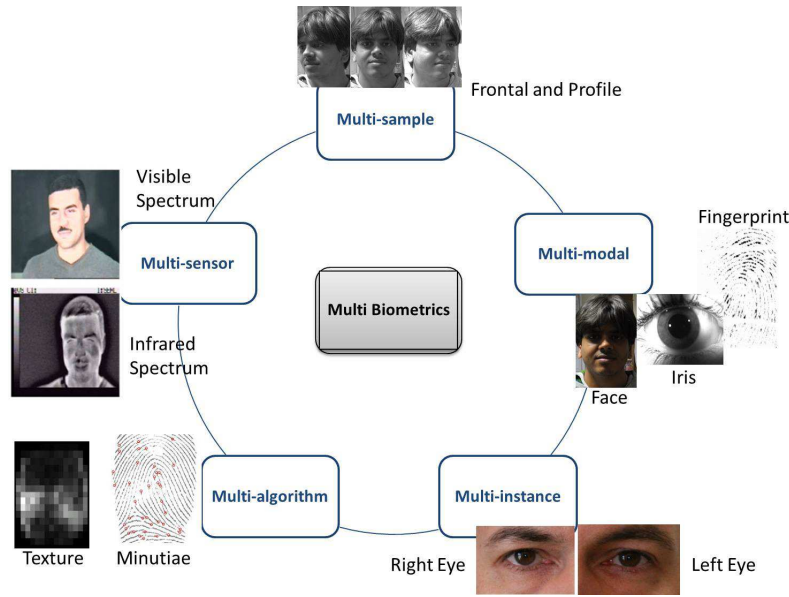


Figure 1.3: Various sources available for combining evidence; multi-sample, multi-modal, multi-instance, multi-algorithm, and multi-sensor.

An important direction of research and one that has obtained considerable focus in the literature is the study of utilizing (*fusing*) evidence of identity from multiple biometrics, referred to as *multibiometrics*. Such systems offer additional benefits over uni-modal counterpart such as resiliency to sensor malfunction or spoofing, universality, greater resilience to noise, fault tolerance, and improved accuracy [159].

As illustrated in Figure 1.3, multibiometric approaches can be categorized depending on the data or process used. A **Multi-sensor** approach is used to capture a biometric trait using different types of sensors. For instance, Chen *et al.* [39] fuse face images obtained from thermal infrared and visible light camera by integrating evidence (at score and rank level) to improve matching performance. Marcialis *et al.* [116] use information from an

optical and a capacitive fingerprint sensor in a fusion framework to improve recognition. A **Multi-sample** approach combines multiple samples of the same modality such as multiple images (frames) of a person’s face obtained in a video sequence from a surveillance system. **Multi-instance** approach combines different instances of the same biometric, for example, fingerprints from different fingers or iris scans from both eyes of the same person. Different modalities captured from the same individual can provide several layers of evidence of identity in a **Multi-modal** approach. Moreover, the use of physically uncorrelated traits such as face and gait, is expected to improve recognition performance. A **Multi-algorithmic** approach can be used to combine recognition algorithms that inspect different aspects of the same modality. Additionally, **Meta-data** approaches leverage additional auxiliary sources of information and combine them with biometric systems to improve performance.

1.2 Factors that Improve Biometric Performance

Fusion of biometric modalities has become an integral part of research in biometrics as the countermeasure to inherent limitations of individual biometric traits. Despite the advantages of multibiometrics, poor quality input, large intra-class variations, limited or unrepresentative training data, can cancel out any theoretical advantages of fusion. Additional issues must also be considered in the development of multibiometric systems such as the cost incurred due to the addition of a new entity in the framework vs. the improvement in accuracy and/or usability. The availability and the reliability of the information used for fusion is also important. Further, using multibiometrics may also increase time of enrollment (for example, multiple contact capture based modalities), computation time (processing and matching several biometric samples) and cost of deployment of the system (cost of installation and maintenance of multiple sensors). However, by considering the *quality* of a biometric sample and utilizing various orthogonal evidences of performance in a contextual setting, it is possible to alleviate the aforementioned concerns of traditional multibiometric systems. Further, the availability of a large number of domain specific examples available on the web, allows to further improve performance by leveraging these unsupervised examples. Next, we briefly describe three potential directions that can help in alleviating the limitations of biometric systems.

1.2.1 Quality of a Biometric Sample

The quality of a biometric sample is interpreted differently throughout literature [40, 60, 91, 102]. Most commonly, it is assumed that image quality indicates the *usefulness* of the biometric image in recognition. It is well established that the environmental distortions such as noise, blur, adverse illumination, and compression affect the performance of state-of-the-art recognition algorithms. However, existing image quality metrics generally encode only *visual perception* of a biometric sample. Ideally, it is desirable to design a biometric quality metric that, given an image, can measure the proficiency of an input sample in recognizing an individual (irrespective of recognition algorithms). Quality assessment is also an important component of fusion and adaptive biometric frameworks. Current research uses certain image processing algorithms that are able to assess image degradations due to noise, compression or illumination. However, a quality metric that is biometric specific and entails a greater insight of the usefulness of the biometric sample in consideration, can improve the performance of these systems by providing more discernible quality cohorts.

1.2.2 Adaptive Multibiometrics

Multibiometric systems that are intended for deployment in challenging real-world settings such as airports, national borders, and railway stations must maintain robust performance and low computation time in these non-ideal conditions. A primary concern is of degraded and missing data. The quality of probe image may degrade due to large illumination variations, improper interaction with the sensors (pose variations) or different kinds of noise or blur due to limitations of capture sensors. Further, multibiometric fusion schemes may not handle situations when the quality of probe image is not optimal and when all modalities can not be captured, thereby, performance degrades. Marcialis *et al.* [115] proposed a fusion technique that eliminates the need for all biometric modalities to be captured at once. Serial fusion of face and fingerprint achieves significant reduction in the verification time while maintaining high accuracy. This adaptive nature is triggered by the confidence of prediction from match score distribution of genuine and impostor scores. However, the *trigger* must be some additional meta-information such as the quality of gallery and probe samples. In recent literature of multibiometrics, quality based fusion techniques have gained enormous attention. Vatsa *et al.* [187] proposed a parallel context switching framework that uses image quality and case-based context switching for selecting an appropriate uni-modal classifier or fusion algorithm. Bhatt *et al.* [29] proposed a serial

framework of quality based classifier selection. The ordering of the classifiers in this serial framework is such that the strongest classifier is allowed the first attempt for identification.

1.2.3 Harnessing Auxiliary Information

Humans are naturally exposed to abundant samples of faces and utilize the experience to learn and develop a superior understanding of the structure and context of a faces. The ability to learn from large samples of domain specific information, termed as *background* information, lends to robust face recognition. Further, humans are also able to leverage the social context of co-occurrences of faces, events, and locations to further enhance person recognition capabilities. Moreover, the advent of sources of image data such as search engines and social networking platforms have led to availability of a large number of unlabelled images. This facilitates semi-supervised approaches to improve the performance and robustness of recognition systems. The images allow to improve the understanding of structure of face and tailor the understanding to problem specific needs.

1.3 Research Contributions

This thesis presents new methods to expand the applicability of biometric systems, particularly face recognition, in unconstrained scenarios. Several aspects of biometrics such as holistic biometric quality assessment of face images, online learning based context switching, derived social context, deep learning and distance metric learning are used to improve the purview of adaptive biometrics. The approaches discussed here are evaluated on standard and reproducible benchmarks on publicly available databases and improve upon state-of-the-art performance. The major contributions of this thesis are as follows:

1. **Biometric Quality Assessment:** Different directions of quality assessment in biometrics are collated towards a unified framework with respect to three primary modalities, *viz.*, iris, fingerprint, and face. Various factors and degradations that influence quality in biometrics are presented along with a general quality framework. An experimental analysis of different quality metrics and corresponding relevance to match scores provide a better understanding of the behavior of biometric quality metrics with respect to matching performance. In this experiment it is observed that in place of using an arbitrary set of quality metrics, a careful selection with respect to match scores can provide additional benefits to biometric systems.

Face Quality Assessment: This research also explores a new direction in *face quality assessment using holistic representations* that are designed to encode commonalities

in a large collection of scene images for image classification. As opposed to biometric features, that encode *unique* attributes of an image, scene recognition techniques effectively encode abstract and categorical features of an images. These features are used to learn the *usability* of a face image and segregate face images into abstract categories that are indicative of quality or definite categories such as pose.

Assessment based Quality Enhancement: A probe image may also contain noise due to environmental conditions, incorrect use of sensors or transmission error. The performance of recognition severely depletes when the probe image is contaminated with noise. Denoising techniques can improve recognition performance, provided the correct parameters are used. Context switching can be utilized for effective parameter selection framework where the optimal parameter set is selected for denoising using quality assessment algorithms with low complexity.

2. **Context Switching Framework for Adaptive Biometric System:** Human beings effortlessly process information from multiple sources and utilize the information for decision making. It is well understood that in different operating scenarios, information from some sources may be more useful than others. Hence, a mechanism is required to efficiently combine or circumvent diverse information based on situational cues obtained from the information sources under different conditions. This research proposes an online context switching algorithm that incorporates quality of images in the dynamic selection of unimodal classifiers and their fusion.
3. **Aiding Face Recognition with Social Context Association Rule based Re-Ranking:** This research also aims to broaden the scope of face recognition in consumer photos using social context. Rather than binary cues that have been explored in literature, such as $\{friend, no-friend\}$, the proposed approach infers association between groups of individuals from multi-level social cues such as co-occurrence of people in consumer photos, to improve face identification. These context cues are used to re-rank face recognition results to improve the overall performance. To evaluate the performance of the proposed algorithm, a large dataset is mined from a leading social networking site consisting of 160,264 images from 4675 connected users.
4. **Learning Based Encoding and Distance Metric Approach to Newborn Face Recognition:** Biometric recognition of newborn babies is an opportunity for the realization of several useful applications such as improved security against

swapping and abduction, accurate census and effective drug delivery. This research explores the possibility of using face recognition towards an affordable and friendly biometric modality for newborns. We present a learning framework to handle the large variation in newborn faces, that first learns a *domain-specific* representation of the human face with a deep neural network learning architecture. Next, the learned representation of two input faces are matched with an online SVM formulation of one shot similarity, to match extracted features with low semantic gap. We introduce the largest publicly available database of 96 newborns collected from various hospitals to study existing and proposed face recognition techniques.

Chapter 2

Biometric Quality Assessment

Biometric systems encounter variability in data that influence capture, treatment, and usage of a biometric sample. It is imperative to first analyze the data and incorporate this understanding within the recognition system, making assessment of biometric quality an important aspect of biometrics. This chapter first presents a survey of different concepts and interpretations of biometric quality so that a clear picture of the current state and future directions. Several factors that cause different types of degradations of biometric samples, including image features that attribute to the effects of these degradations, are discussed. A survey of the features, strengths, and limitations of existing quality assessment techniques in fingerprint, iris, and face biometric are also presented. Finally, a representative set of quality metrics from these three modalities are evaluated on a multi-modal database with respect to match scores obtained from the state-of-the-art recognition systems. Next, this research presents a face quality metric to quantitatively measure the usability of an image as a biometric sample. The experiments on the CAS-PEAL and SC-Face databases containing covariates such as illumination, expression, pose, low-resolution and occlusion by accessories, suggest that the proposed algorithm can efficiently classify input face image into relevant quality categories and be utilized in face recognition systems. Finally, a parameter selection framework is presented to obtain an optimal parameter set for denoising using quality assessment algorithms with low complexity. Quality score based parameter selection is evaluated on the AR face dataset and the experiments suggest that the proposed framework improves the performance both in terms of accuracy and computation time.

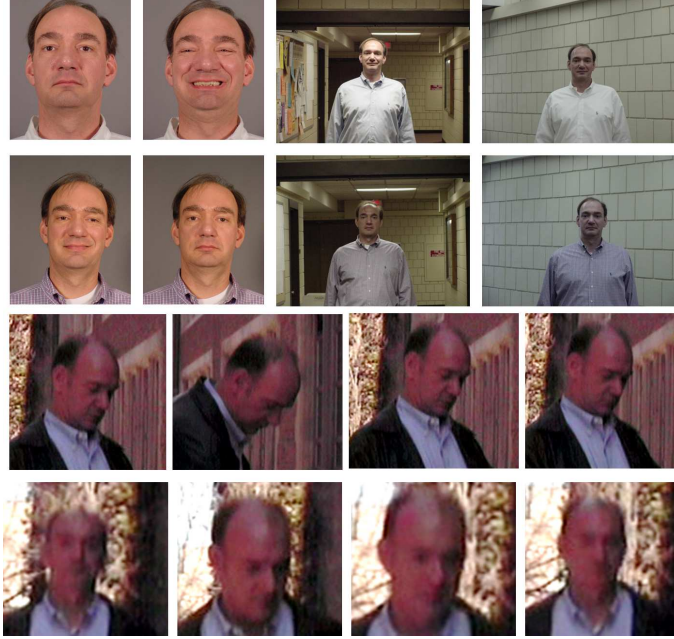


Figure 2.1: Variation in quality. A biometric system may encounter samples of a wide range of quality. Effective quality assessment metrics that are indicative of these variations are therefore essential to an automated biometric system.

2.1 Introduction

Biometrics, as an integral component in identification science, is being utilized in large-scale biometrics deployments such as the US VISIT, UK IRIS project, UAE iris-based airport security system, and India's Aadhaar project. These far-reaching and inclusive delivery systems not only provide a platform to assist and enhance civilization but also offer new research directions. An important research challenge among them is the measurement of *quality* of a biometric sample. Biometric systems, like other applications of pattern recognition and machine learning, are affected by the quality of input data. Therefore, it is important to quantitatively evaluate the quality of a sample that is indicative of its ability to function as a biometric. In our opinion, quality of a biometric is beyond measuring the quality of the image itself. While a sample's quality is susceptible to irregularities during capture or storage, it may also have low quality by its very nature. For instance, as shown in Figure 2.1, an input biometric sample may possess a wide range of *quality*.

Quality assessment (QA) of an *image* measures its degradation during acquisition, compression, transmission, processing, and reproduction. Several QA algorithms exist in image processing literature, which pursue different philosophies, performance, and applications. A majority of these methods are motivated towards accurate *perceptual* image

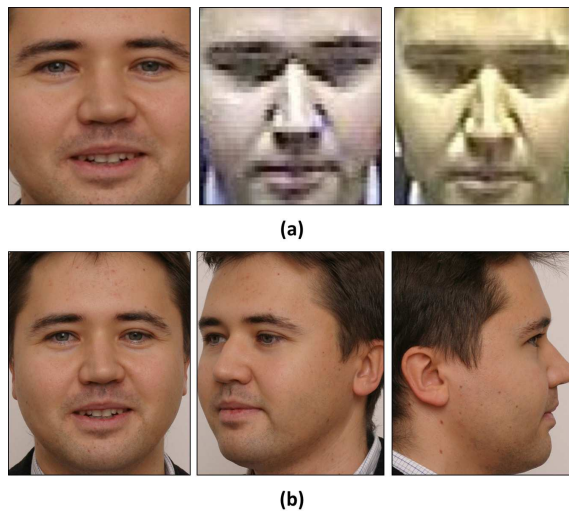


Figure 2.2: Image quality vs biometric quality. While the images in (a) are of poor *image* quality, the images in (b) may have lower *biometric* quality.

quality i.e., quality as perceived by the sophisticated human visual system (HVS). These approaches require an in depth understanding of the anatomy and psychophysical functioning of the human cognitive system. Several perceptual quality metrics are surveyed by Wang and Bovik [197] and Lin and Kuo [110]. On the other hand, the quality of a biometric sample is interpreted differently throughout literature [21, 40, 91, 99, 102, 142, 152, 213]. A summary of these interpretations is provided in Table 2.1. In general, biometric quality is defined as an indicator of the *usefulness* of the biometric sample for recognition, as illustrated in Figure 2.2. It is well established that environmental distortions such as noise, blur, and adverse illumination, affect the performance of state-of-the-art recognition algorithms. However, existing image quality metrics that measure such degradations encode only a part of the information that can measure the overall quality of a biometric sample. Hence, a clear distinction must be made between Perceptual Image Quality Assessment (PIQA) and Biometric Quality Assessment (BQA). PIQA research attempts to understand why human subjects prefer some images to others [37, 66]. The task is complex and involves multiple disciplines, including an understanding of the HVS. On the other hand, BQA provides an initial estimate of the ability of a sample to function as a biometric. We therefore define biometric quality as

The quality of a biometric sample is a measure of its efficacy in aiding recognition of an individual, ideally, irrespective of the recognition system in use.

Table 2.1: Different interpretations of quality in biometrics from literature

Reference	Modality	Interpretation of quality in biometrics
Chen <i>et al.</i> [40]	Fingerprint	A global measure of the strength of ridges
Grother and Tabassi [142]	Fingerprint	Suitability for automatic matching
Youmaran and Adler [213]	Face	The decrease in uncertainty of identity due to a given sample
Kryszczuk <i>et al.</i> [99]	Face	Conditionally relevant class predictors
Beveridge <i>et al.</i> [21]	Face	A measurable and actionable predictor of performance
ISO/IEC standards [78]	Face	Biometric data that adheres to best capture practices
Kalka <i>et al.</i> [91]	Iris	The measurement of various degradations known to affect iris recognition
Kumar and Zhang [102]	Knuckles	Confidence of generating reliable matching scores from the user templates
Poh and Kittler [152]	General framework	Degree of <i>extractability</i> of recognition features
BioAPI [178]	General framework	Biometric data that provides good performance for the intended purpose

In literature, quality assessment metrics are widely used in the formulation of biometric techniques. As illustrated in Figure 2.3, quality metrics can be used at various stages of the recognition pipeline to improve performance and usability of biometrics in challenging conditions. The application of quality metrics can be during both enrolment and recognition phases. Since enrolment phase is the best opportunity to re-capture a sample to maintain the overall quality of the gallery set, the quality of input sample is an important consideration. On the other hand, the quality of a probe sample during recognition phase is utilized in different methodologies to improve the recognition performance. Some important applications and evaluation metrics of quality assessment techniques in biometric systems are described here.

2.1.1 Quality Assessment During Enrolment

Quality feedback during enrolment is critical in collecting high-quality gallery data. It is common, especially in large-scale biometric systems, to have a supervised enrolment process as in the case of the India’s Aadhaar project. An active quality feedback enables the collection officer to evaluate and maintain quality standards during the enrolment process [203]. It can also be a performance measure for the collection apparatus and

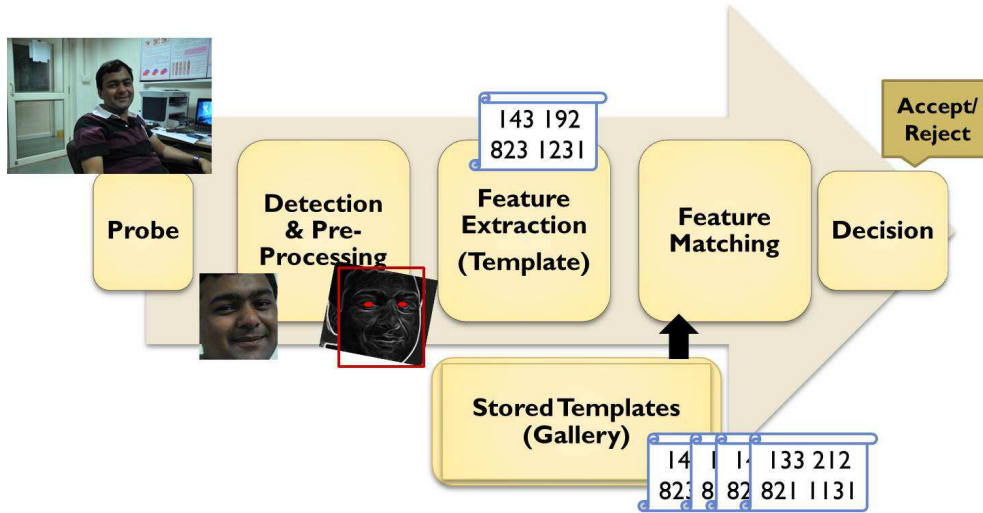


Figure 2.3: Pipeline of a typical biometric system. This consists of a capture sequence (probe), detection and preprocessing, feature extraction, matching and decision modules.

procedure employed for data capture [190]. Aggregated quality may also be used to create timeline along with historical or geographical meta-data for other analysis.

2.1.2 Quality Assessment During Recognition

Quality assessment and feedback during verification can help mitigate false alarms. A verification system can choose not to perform matching if the quality score is below a threshold, depending on the computation time of matching and the overhead of re-acquisition of data. Most modern fingerprint and iris sensors are now bundled with active quality-control mechanisms. Identification is inherently a computationally expensive process, hence, it is a good idea to use quality assessment (computationally less expensive) to improve system usability. For example, quality can be used in negative identification, where it is in the interest of the subject to provide a poor quality sample. The subject may then be persuaded to provide better quality samples without having to wait for misleading and incorrect identification result from the system. Further, in the recognition pipeline, quality is used at different stages/levels of a biometric system:

- *Preprocessing:* A probe sample may contain degradations due to environmental conditions, incorrect use of sensors, or transmission error. The performance of recognition systems severely depletes in such cases. Image restoration techniques can improve image quality, provided that the correct parameters are used [23]. Quality-assessment-based selection of parameters for image enhancement shows marked im-

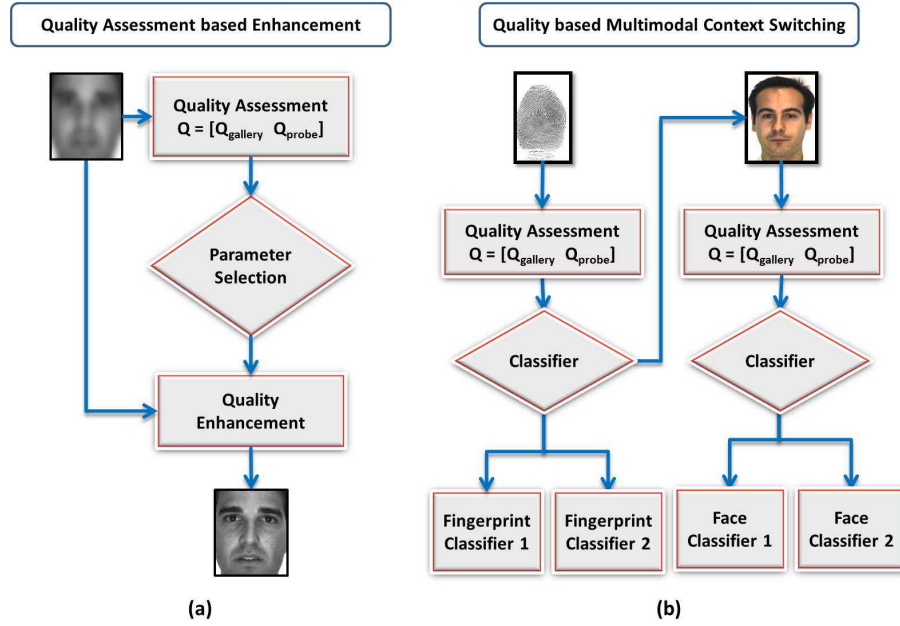


Figure 2.4: Utilizing biometric quality assessment for context switching. Framework for (a) a quality-driven biometric image enhancement, based on [23], and (b) quality-based multiclassifier selection, proposed in [29].

provement in the recognition performance of the resultant biometric sample, when compared to using generic parameters. Also, biometric images obtained from different uncorrelated or orthogonal bands of the spectrum can provide different amounts of information, as demonstrated by Vatsa *et al.* [188] with the face and iris [189]. An illustration of a quality-assessment-based image enhancement framework is presented in Figure 2.4a.

- *Recognition*: Poh *et al.* [151], Kryszczuk *et al.* [98, 99], and Poh and Kittler [152] have shown that while quality assessment scores are used for perceptual understanding of the sample or performance prediction, they also possess some discriminatory ability. Their experiments show that incorporating quality assessment values as additional features can improve the recognition performance. Similarly, quality-augmented product of likelihood ratio fusion scheme has shown to improve the performance [124]. Grother and Tabassi [142] have studied the relationship between quality and recognition accuracy in fingerprints and suggested that quality scores can help in predicting the similarity scores.
- *Context switching*: Context-switching frameworks dynamically select classifiers and/or distance metrics based on the quality of the sample. A serial framework for quality-

based context switching is illustrated in Figure 2.4b. Recent literature [29, 167, 187] demonstrates the advantages of context switching of a biometric recognition pipeline based on the feedback from quality assessment algorithms. Vatsa *et al.* [187] propose a parallel context switching framework that uses energy in sub-bands, activity level, and pose angle for selecting the appropriate uni-modal classifier or fusion algorithm. Sellahewa and Jassim [167] present a simple thresholding-based adaptive fusion approach on illumination estimation from first-order statistics. Bhatt *et al.* [29] propose a serial framework of quality-based classifier selection using both image quality and biometric-specific quality metrics. Alonso-Fernandez *et al.* [9] present a quality-based context switching framework to improve sensor inter-operability in fingerprint biometric. Poh and Kittler [152] propose a unified framework for fusion of biometric classifiers at match score level by incorporating quality measures. This framework is based on a Bayesian perspective and can be used both as a generative and discriminative classifier.

- *Decision*: Quality assessment scores can also aid decision-level fusion. By providing quality priors to maximize selective or cumulative combination of decision, the notion of strong or weak classifiers can become subject specific. Hence, the primary concern of using decision-level fusion schemes, discussed in [50], can also be eliminated. For rank-level fusion, Abaza and Ross [1] propose a weighted variant of Borda count rank aggregation technique using quality assessment scores. An empirical evaluation [101] shows the applicability of nonlinear rank-level fusion as well, particularly in palmprint biometrics.
- *Sample Update or Replacement*: Another interesting application of quality scores is in the replacement or addition of a confirmed probe sample to the gallery based on its quality. While this procedure has the risk of gallery contamination, it can elevate important concerns of temporal variations of biometric data, such as facial aging.
- *Decision update*: Researchers are exploring the use of *online or incremental learning* approaches to improve the decision boundary of the classifiers even in deployment phase [29, 30, 171]. A major concern in such systems is to select suitable samples to learn incrementally. For instance, modifying decision boundary based on all the incoming samples may be computationally expensive. Further, online learning on outlier samples can adversely affect the system performance. One area of focus is

towards using quality of the sample to determine whether the sample is suitable for classifier update.

The applications show that active involvement of quality assessment beyond the capture stage of the biometric pipeline encourages the formulation of complex and accurate biometric quality assessment. Hence, BQA is an important aspect of biometrics research that can lead towards robust and user-friendly biometric recognition systems. The aim of this survey is to collate different directions of quality assessment in biometrics towards a unified framework with respect to three primary modalities, *viz.*, iris, fingerprint, and face. Section 2.2 discusses various factors and degradations that influence quality in biometrics. Image features used in quality assessment to evaluate the effect of those degradations are also presented along with a general quality framework. Section 2.3 presents a review of recent literature in biometric quality assessment pertaining to fingerprint, iris, and face modalities. Evaluation protocols inspired by different applications that are indicative of the metric’s performance are also presented. Section 3.4.3 presents an experimental analysis of different quality metrics and corresponding relevance to match scores providing a better understanding of the behavior of biometric quality metrics with respect to matching performance. In this experiment it is observed that in place of using an arbitrary set of quality metrics, a careful selection with respect of match scores can provide additional benefits to biometric systems. Finally, we also discuss the salient finding from our experimental evaluations and literature as well as future scope and directions. Additionally, a brief overview of perceptual image quality assessment is presented in Appendix 1 and quality metric standards prevalent in biometrics literature are discussed in Appendix 2.

2.2 Biometric Quality: Factors, Degradations, and Features

An observer’s perspective in assessing quality is an important aspect of QA [94]. For instance, the perception of an image can change with respect to the subject, the photographer, or by the interpretation of some third party. Similarly, the quality of a biometric sample can depend on the acquisition system and the technology used for matching. For meaningful prediction of quality, the ideal pursuit is towards a quality metric that is consistent across any type of degradation and matching techniques. However, pragmatic solutions utilize some understanding of the degradation and matching techniques in their formulation.

This section describes the cause and effects of factors that influence quality of biometric samples. Further, the image features that are typically used in automatic image analysis

of biometric samples are studied. Finally, a general framework for quality assessment in biometrics is presented.

2.2.1 Factors that Influence Biometric Quality

It is important to appreciate the effects of various factors that affect quality to develop better assessment algorithms. While some factors are unavoidable, others may be inherent limitations of the biometric itself. These factors are either user traits or interactions between user and sensors:

- *User traits* Some important factors that influence the quality of a biometric sample during capture process can be classified as *behavioral* and *physiological* traits of the human users [10]. Behavioral traits may include motivation levels, cooperation, and fears. Physiological traits include facial hair or sensitivity to light. While some behaviors of users can be restricted, it is at the cost of usability and increased inconvenience. Further, unavoidable factors such as age, social customs, gender, and injuries can impair the quality of the captured sample. For instance, fingerprints obtained from older age groups is of lower inherent biometric quality (due to worn ridges) when using different commercial fingerprint systems [120].
- *User-sensor interaction and operational constraints* The second important factor that influences the quality of contact capture (closed/near field of view) based biometrics, such as fingerprints, palmprints, iris, and retinal, is the interaction between users and sensors. The *usability* of the sensor is crucial to quality. Sensors with active user feedback that are portable and easy to use ensure good user-sensor interaction, resulting in better quality captures. However, environmental factors such as temperature, humidity, and background influence this interaction, adversely affecting the quality of a biometrics. Other factors that affect the quality of a biometric sample are operational constraints particularly in the use and maintenance of (touch-based) sensors and training of handlers. For instance, Aadhaar project uses different types of sensors and operational procedures in accordance with the climatic conditions of different regions of India. In such cases, controlling conditions, policies, and guidelines during operation play a significant role.

Table 2.2 presents some possible causes of each of the aforementioned factors. These factors have varying degrees of adversarial effect on the performance of a captured biometric sample. Uncooperative users, such as in criminal cases, pose an additional challenge to

Table 2.2: Various behavioral, environmental, and operational factors that effect quality of biometric sample

Factors	Possible causes
User traits	Tiredness, distractions, motivation, cooperation, fear, makeup, appearance, facial hair, clothes, or hats
User-sensor interactions	Indoor/outdoor, background, temperature, humidity, illumination, and ambient noise
Operational	Familiarity, quality feedback, sensor cleaning, supervising operator, and time between acquisition

effective data collection processes. It is worthwhile to understand the different degradation processes that result from these factors.

2.2.2 Degradations in Biometric Images

In order to better understand quality assessment in biometrics, it might be useful to closely inspect the different artifacts that commonly manifest in biometric images. As illustrated in Figure 2.5, these degradations are either virtues of an image or of the biometric modality itself.

2.2.2.1 Image-based Degradations

Image degradations are manifested by the property of capture devices and conditions, irrespective of the biometric being captured:

- **Blurring:** Image blurring is a common phenomenon that occurs due to incorrect focus (object is outside the depth of field), motion, or certain environmental factors. Blurring effects edge information, which is vital to biometric recognition, particularly the minute edges of iris patterns.
- **Illumination:** Uniform lighting is essential for the capture of a good quality biometric. Conversely, adversely directed lighting drastically affects the performance of iris and face.
- **Noise/Compression:** An image may contain noise due to environmental factors, incorrect use of sensors, and transmission error. Noise contamination drastically affects the performance of recognition systems. Depending on the compression levels, various image encoding techniques produce artifacts such as blockiness and ringing effect.

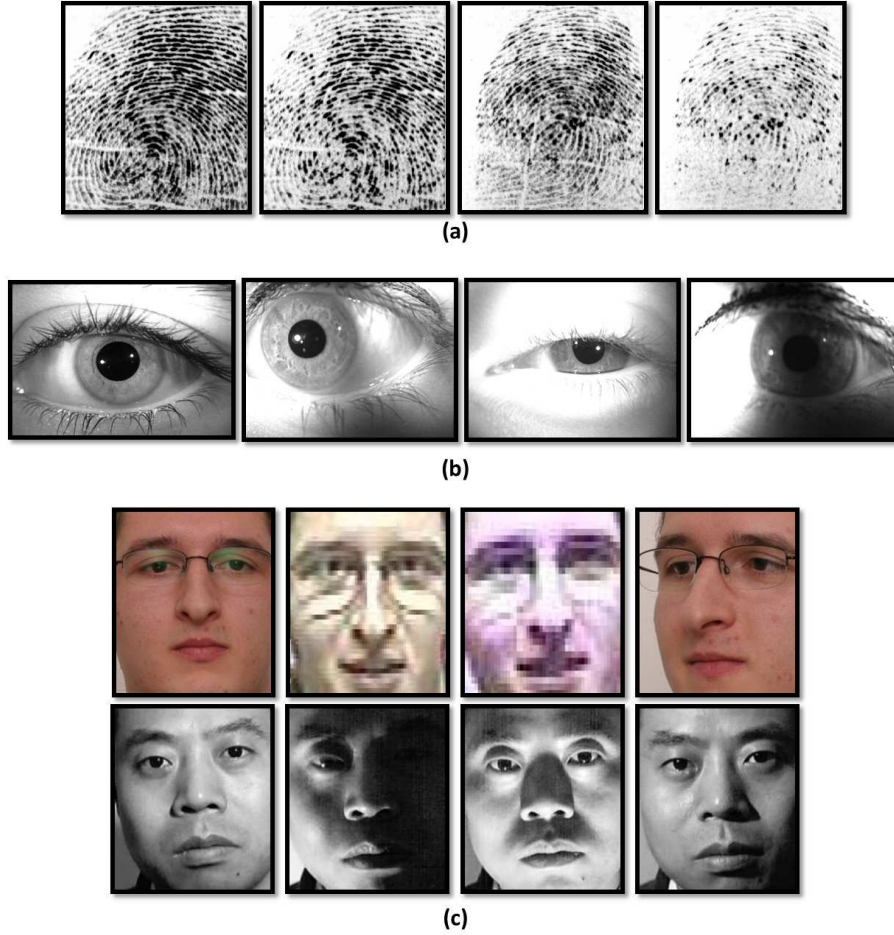


Figure 2.5: Sample images of varying quality. (a) Fingerprint, (b) iris (from WVU multi-modal database), and (c) face (from SCface and CAS-PEAL face databases) illustrating the wide range of quality that a biometric system can encounter with different image and biometric specific degradations.

- Optical distortions: Nonconformity to rectilinear projection causes distortion in the captured image. Such distortions may occur due to various environmental factors or due to the functioning of sensors. Further, difference in the sensor models also results in different distortion profile, degrading recognition performance [160].

The aforementioned degradations usually occur due to the limitation of sensor technology or environmental conditions. As the constraints on user during capture are relaxed, the impact of these factors on the performance of systems increases drastically. Therefore, estimation and analysis of these factors are critical for building robust and nonintrusive biometric systems.

2.2.2.2 Biometric-modality-specific Degradations

Biometric degradations occur as a consequence of the nature of the biometric modality being captured. For example, face and iris biometrics have multiple degrees of motion and hence pose angle at which a captured image can affect quality. Murphy-Chutorian and Trivedi [121] survey several head-pose estimation techniques. Fingerprints exhibit pose variations in terms of fingerprint orientation that may result in a partial prints. Biometric data from unconstrained environment is plagued with occlusion or missing information. Common causes in case of face include accessories and facial hair. Erroneous data can also arise from medical conditions, scars, or skin deformations (due to temperature or dryness).

Certain degradations may be difficult to measure, for example, the aesthetic changes of the face brought about by hair style or makeup. Beveridge *et al.* [21] introduce the notion of *measurable covariates*, a subset of different degradations that are easy to estimate from an image. Note that measurable covariates can be properties of the image (edge density measures) or of the subject (inter-eye distance). Further, properties such as region of interest, focus of camera, and also expression, glasses, and clothing that can be controlled to some extent (at the cost of usability), are termed as *actionable*. Nonactionable covariates include age, gender, and race. Accurate assessment of measurable and actionable covariates of biometrics must be the focus of quality assessment techniques. Current research primarily focuses on using image processing techniques to assess image features that indicate quality. These different *image features* are examined next.

2.2.3 Image-based Features

The aforementioned degradations manifested in biometric samples can be assessed using image features that are computationally inexpensive to compute. Automatic QA is primarily addressed by analyzing spatial and temporal features that are indicative of the image content. Features that are used extensively in current literature can be broadly divided into four categories (as shown in Figure 2.6):

- *Orientation* features are obtained from edges in the image. In case of the iris and face, edge information is widely used as features for recognition. Blurring, illumination, and noise degrade edge information thereby affect performance. Hence, orientation information can provide a good indication of the quality of a biometric sample.
- *Power spectrum* is a temporal measure of the power of the image signal. This measure is an indication of the amount of information present in an image region.

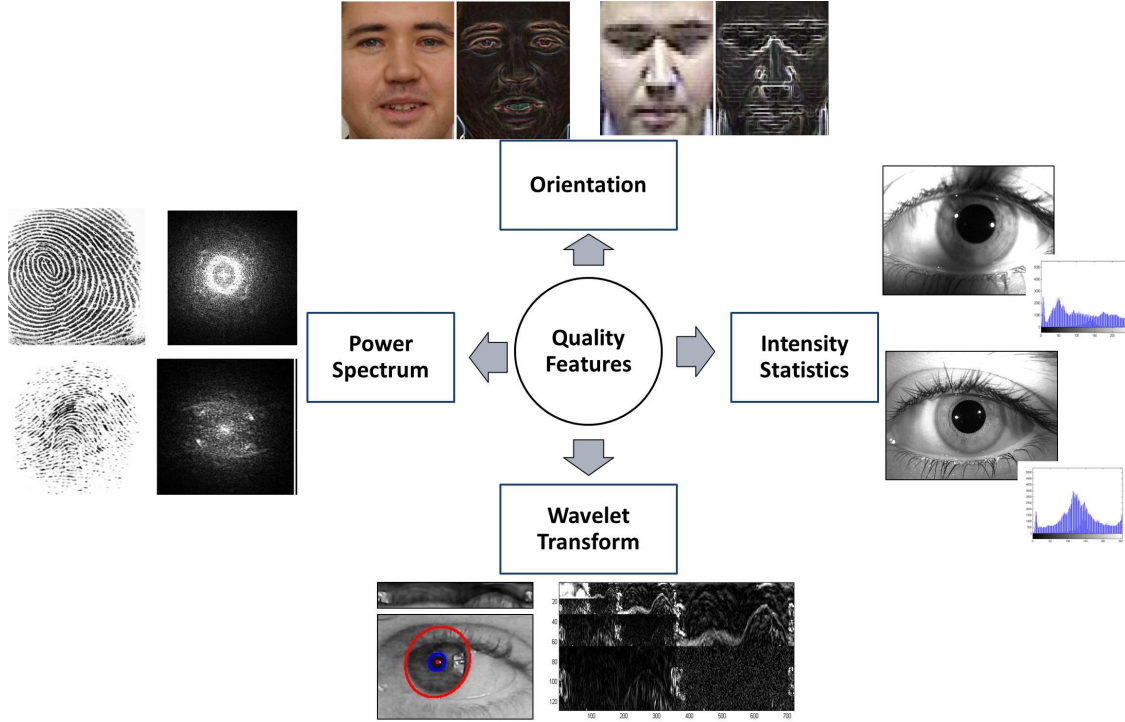


Figure 2.6: Four image features are primarily used for estimating quality of biometric images. Orientation, intensity statistics, power spectrum, and wavelet transform.

Hence, spectral energy is often computed for different image regions to obtain a local assessment of quality.

- *Intensity statistics* are direct statistical evaluation of intensities of pixels in the image. Typically, a statistical measure such as Kurtosis or Point Spread Function (PSF) estimation is used to estimate blurring or illumination degradation in the image. The measure can then be compared to the reference values obtained from ideal images to compute the extent of degradation.
- *Wavelet transform* provides both spatial and frequency understanding of the information content in each sub-band of the image. These are particularly suited to ascertain the presence of fine *micro edges* in the iris region and to obtain local analysis of quality in different regions of an image.

In addition to the four image features, the *shape* of the segmentation boundary of the biometric content of the image can also provide useful information of the quality of the sample. For instance, the circularity and pixel density of an iris segmentation are important quality measures and widely used in literature. However, we assert that the same degradations that affect recognition can also affect the segmentation performance.

Hence, the performance of shape as a quality feature deteriorates rapidly with nonideal images. In cases where color imagery is used for capture, multichannel information are also leveraged for QA. It has been reported in the literature that the discriminatory power of certain channels supersedes others. Therefore, quality metrics for each channel may also be considered separately. Finally, several QA techniques use multiple features to form a composite quality score via (statistical) fusion; they are referred to as *combined features*. Nonimage features such as image header information (EXIF), or cues obtained from sensor, may also be used as features for quality assessment. However, the subjective nature of these features leads to poor generalization.

2.2.4 Naturality, Fidelity, and Utility in Biometric Quality

Different QA algorithms in literature have some underlying similarities in their philosophy/approach. It might be helpful to classify existing algorithms based on these underlying principles for a thorough understanding of the current state of research and limitations of literature. Several attempts have been made at this classification; Kalka *et al.* [91] classified iris quality assessment algorithms into global and local algorithms. Beveridge *et al.* [22] classified techniques based on the properties of different covariates. Inspired by the visual quality model of Yendrikhovskij [211] (illustrated in Figure 2.7), this research presents three aspects of quality assessment in biometrics:

1. Biometric naturality: *the degree of apparent match of the biometric image with an internal reference of goodness*. Most of the no-reference quality assessment algorithms measure perceptual image quality, indicating the naturalness of that image. These methods [110, 119, 197] are based on unexpected changes in intensities or ratio of information in various spatial/temporal bands, effects that stand out in visual inspection of quality. Such metrics are adept at encoding image level degradations, such as illumination, compression artifacts, noise, and blurring. These metrics are computationally inexpensive and their performance is dependent on baseline parameters obtained from some knowledge of the intended application (internal reference of goodness).
2. Biometric fidelity: *the degree to which a biometric modality is correctly represented in the acquired image*. The quality or the extent to which the acquired image (from a sensor) successfully represents the biometric that is presented to a sensor is the measure of fidelity of a biometric sample. Measuring the fidelity is a challenging

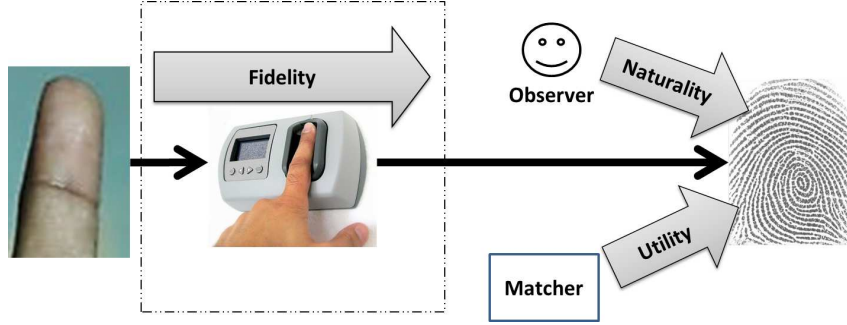


Figure 2.7: Three aspects of quality assessment: naturality, fidelity, and utility, in a typical biometric pipeline.

problem as there may not be additional information to verify the sample with respect to the source.

3. Biometric utility: *the degree of suitability of the sample for matching*. The utility of a biometric sample is based on its matching performance. While utility is surely dependent on the sample’s naturalness and fidelity, it has been shown that (face) biometric samples of the same person captured in similar settings can exhibit marked difference in matching performance. Further, the information, while correctly captured, may be useless to the particular matcher. Hence, the utility of a biometric is often independent of the other two aspects of biometric quality.

Alonso-Fernandez *et al.* [8, 10] also use similar nomenclature to describe quality assessment *viewpoints*, from which the authors conclude that for fingerprint biometrics, ‘utility’ is of primary focus. However, we contend that in order to obtain a complete understanding of the quality of a biometric sample, all three dimensions, naturality, fidelity, and utility must be evaluated. This is more pertinent for iris and face biometrics, where the features are not structured as compared to fingerprints.

2.3 Literature Review: Quality Assessment in Fingerprint, Iris, and Face

Several techniques have been proposed in literature to assess the quality of a biometric sample that is affected by aforementioned degradations. In this section, a literature review of quality assessment algorithms pertaining to three popular modalities, *viz.*, fingerprint, iris and face, are presented, along with the review of key techniques to evaluate quality assessment algorithms.

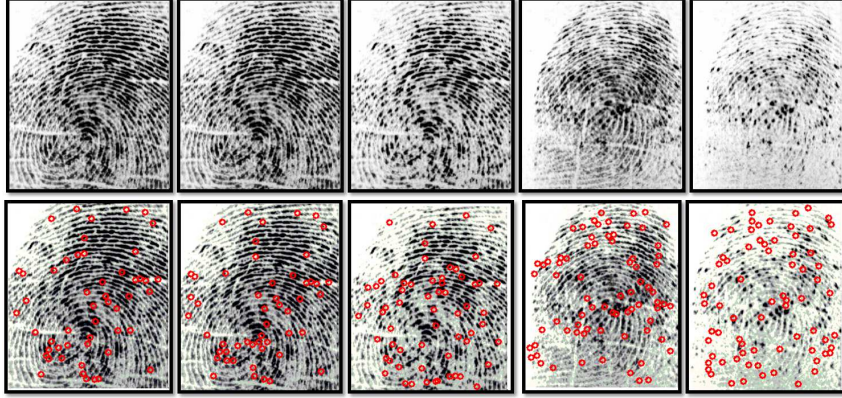


Figure 2.8: Poor quality fingerprint samples often lead to spurious minutia.

Table 2.3: A representative list of fingerprint quality assessment algorithms

Category	Algorithm	Description	Type
Pixel intensity	Chen <i>et al.</i> [38]	Grey level distributions of segmented ridges	Local
Wavelet transform	Vatsa <i>et al.</i> [186]	Combined response from RDWT for dominant edge information	Local
Power spectrum	Chen <i>et al.</i> [40]	In a ring-shaped region of the spectrum	Global
Combined features	NFIQ [177]	Amplitude, frequency, and variance of sinusoid to model valid ridges	Global
Orientation tensors	Fronthaler <i>et al.</i> [60]	Encode orientation with parabolic symmetry features	Global

2.3.1 Fingerprint Quality Assessment

Poor quality fingerprint images can lead to incorrect or spurious feature (minutia) detection (illustrated in Figure 2.8) and thereby degrading the performance of a fingerprint recognition system. Quality assessment of fingerprint ridge quality is essential for proper functioning of the recognition system. These metrics are primarily used in fingerprint sensors with active quality feedback for rejecting poor quality samples. Fingerprint quality is also used to evaluate local unrecoverable regions of the fingerprint, as enhancement of these regions for ridge information may be counter-productive. Further, region-wise assessment may also be useful in adaptive feature importance weighting schemes. Most fingerprint quality assessment metrics compute image properties in local regions and pool these metrics to present a single quality score. A detailed review of some seminal techniques is presented here along with a summary in Table 2.3.

Lim *et al.* [108] present a local-feature-based quality metric which computes *orienta-*

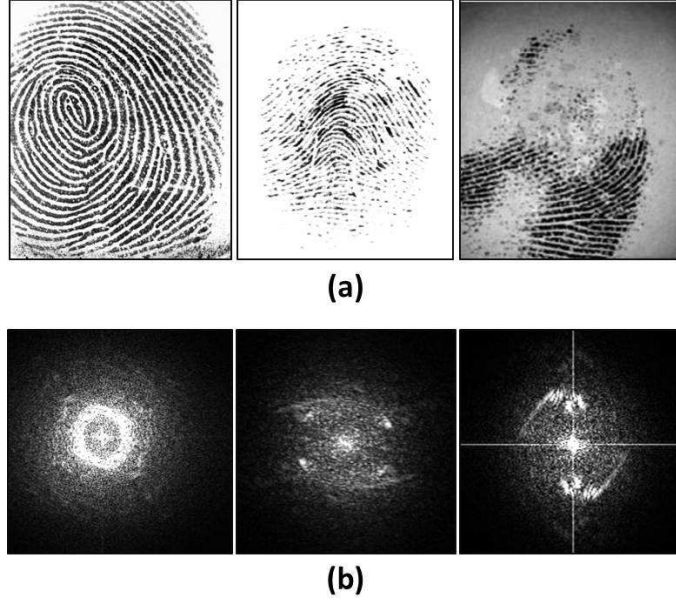


Figure 2.9: A fingerprint image (a) and corresponding Fourier transform (magnitude component after shifting) (b). The ridge information manifests as a bright band. Chen *et al.* [40] use the difference of two Butterworth filters to obtain a soft bandpass filter that captures the strength (and thereby quality) of the ridges.

tion certainty level (OCL), *ridge frequency*, *ridge thickness*, and *ridge-to-valley thickness ratio*. Shen *et al.* [169] use Gabor filters for quality assessment. Fingerprint image is tessellated into blocks, and Gabor filters with different orientations is applied on each block. For high-quality blocks, response from filters of some orientations is significantly higher than others, whereas for low-quality blocks, the difference in responses from the filters is generally low. The standard deviation of the responses thus indicates local quality for each block. The aggregated local quality is compared with scores from visual inspection. Similarly, Vatsa *et al.* [186] use Redundant Discrete Wavelet Transform (RDWT) to compute dominant ridge activity to measure fingerprint quality. The quality metric induced huge performance improvement when incorporated into a fingerprint feature level fusion framework on a large real-world database. Olsen *et al.* [140] also present a quality measure based on evaluating Gabor filter responses of a fingerprint image whose performance is more robust to its parameters.

In another approach, Chen *et al.* [40] measure the quality of ridge samples by energy spectral density concentration in particular frequency bands obtained by discrete Fourier transform (DFT). It is observed that good quality ridges manifest at a certain frequency band of the transformed fingerprint image as shown in Figure 2.9.

The most popular fingerprint quality assessment algorithm in literature is the National Institute of Standards and Technology (NIST) Fingerprint Image Quality (NFIQ) [177]. This approach also pioneers the use of quality metrics as performance predictor in fingerprints. A feature vector v consists of 11 quality features obtained on the basis of localized quality map per fingerprint image. The map is computed based on the local orientation, contrast, and curvature of each region of a rectangularly tessellated fingerprint image (blocks with size 3×3). Rather than using true labels based on human perception, normalized separation of genuine match score from the match score distribution obtained from an automatic fingerprint matcher is used to train a multilayered perceptron. Recently, NFIQ 2.0 [15] is introduced with a similar learning-based quality assessment framework in which several new image-based features are considered for inclusion, including Gabor filter responses.

The NFIQ quality metric has been extensively used in literature and tested across different datasets. However, the orientation estimated about the singularity points tends to fail for high curvature. Fronthaler *et al.* [60] present a solution based on characterizing orientation using parabolic symmetry features. The proposed technique first converts the image into orientation tensor representation. The orientation tensors in both horizontal and vertical direction are combined to encode the edge information obtained from the horizontal, vertical, or parabolic tensors. The information present in each local region is combined to obtain the final quality score. The chapter also discusses using the same technique with higher-order orientation tensors to encode information in face images. The results indicate that correlation of this quality score with NFIQ and with human annotations is high.

Alonso-Fernandez *et al.* [10] present a comparative study of several fingerprint quality metrics. These algorithms are segregated into global and local metrics depending on the nature of assessment. The study shows a high correlation of fingerprint quality metrics among themselves. This seems to indicate that the studied approaches encode similar information from the fingerprint image to predict quality. Recently, fingerprint quality computed using the ridge information in various sub-bands is shown to provide the best rejection criteria to improve performance [149]. The fingerprint ridge frequency and orientation were captured using short-time Fourier transform. The metric encodes the continuity of the ridge spectrum along the orientation of strong ridges in the image. In another research, self-organizing maps (SOM) are used to classify local regions of a fingerprint to different quality labels [139]. A SOM is trained to cluster blocks of fingerprints based on their spatial information to create a high-level representation of the

fingerprint. Further, a *Random Forest* is used to learn the relationship between the SOM representation and actual matching performance.

The fingerprint quality assessment techniques measure consistency and strength of the ridge patterns. A direct association is made between the properties of the ridge patterns and the recognition performance of the sample. The more challenging problem of *latent* fingerprint quality assessment is also being studied [76, 161, 212]. Background noise, smudging, and partial nature of these types of fingerprints, usually obtained from crime scenes, hinder a good fit to precomputed models of ridge flows or patterns. Fingerprint quality metrics are also important for effective compression techniques [74]. Finally, quality assessment of 3D fingerprints that are obtained either from a 3D sensor or reconstructed from multiple 2D views, is an open research problem.

2.3.2 Iris Quality Assessment

The performance of the iris as a biometric is highly dependent on the quality of the sample. Some major covariates in iris recognition include focus and motion blur (due to hand-held sensors), off-angle (pose), occlusion (eye lashes, hair, and spectacles), dilation/constriction, and resolution. Additionally, the presence of cosmetic contact lenses also affects the natural texture of the iris [207]. In order to compensate for these covariates, early iris capture systems were bulky and cumbersome to use. However, as newer and compact sensors with focus on usability emerge, there is greater need to measure the quality of the captured sample. Unlike fingerprints, iris patterns do not exhibit any *expected* behavior of the features, hence, quality is measured in terms of the impact of the covariate on the image. A brief description of some leading iris quality assessment methods is presented in Table 2.4.

Chen *et al.* [41] present a quality metric for iris based on the spectral energy in local regions. Firstly, iris is segmented using Canny edge detector and Hough transform. Next, occluded regions that may occur due to eyelashes are removed using intensity thresholding. The 2D Mexican hat wavelet decomposition is applied, and the product of responses from multiple scales (usually three) is used as the overall response. The iris region is partitioned into concentric bands with fixed width (8 pixels). The energy from concentric regions are separately computed and combined into a single quality score. Multiple overlapping filtering of the iris region approach is essential to encode the fine edges exhibited by the iris muscle tissue. The approach is also used for feature extraction. A similar approach is proposed by [2].

Table 2.4: A representative list of iris quality assessment algorithms

Category	Algorithm	Description	Type
Combined features	Daugman [50]	Focus estimate and off-angle measure by deformation function that maximizes circularity of pupil	Global
Power spectrum	Chen <i>et al.</i> [41]	Spectral energy in local regions of the iris	Local
Combined features	Zuo <i>et al.</i> [90]	Assessment of interlacing, illumination, focus, off-angle, area, blur pupil dilation	Local, global
Combined features	Kalka <i>et al.</i> [91]	Evaluation of seven quality parameters and fusing them statistically	Local, global
Combined features	Proenca [154]	Estimation of seven separate quality attributes that impact recognition	Local, global

In another approach, Kalka *et al.* [91] present quality assessment of iris images based on the evaluation of eight quality parameters (defocus, motion blur, off-angle, occlusion, specular reflectance, illumination, and pixel count). These individual quality scores are both image-based and biometric-specific in nature. Further, Dempster-Sheffer theory-based fusion is used to combine these individual scores to obtain a single quality value. The quality measure is evaluated on the iris dataset of the West Virginia University (WVU) multimodal biometric database [45], using the quality bins approach discussed previously.

Recent interest in nonideal iris imagery has sparked research on iris recognition in the visible spectrum. Proenca [154] presents a quality assessment algorithm for operation on visible iris imagery. Similar to Kalka *et al.* [91], seven quality attributes that impact recognition are identified and estimated. The algorithm is tested via improvement in recognition rate when the lowest quality images from the database are ignored. The author also presents a summary of existing quality assessment algorithms for iris. In another approach, Zuo *et al.* [216] present an iris quality assessment technique based on match score evaluation. By utilizing precomputed distributions of genuine and imposter scores, the quality of a sample is measured by statistical fusion of two quality metrics: (a) statistical error between the distribution of genuine and imposter scores and (b) normalized difference between the sample match score and some quantile points selected from the genuine and imposter distributions. The authors later improve the approach [217] using a multivariant prediction (feed-forward neural networks) to better map quality values with matching performance. Baig *et al.* [13] also discuss a score level quality assessment based on Mahalanobis distance. Du *et al.* [55] present a feature correlation approach

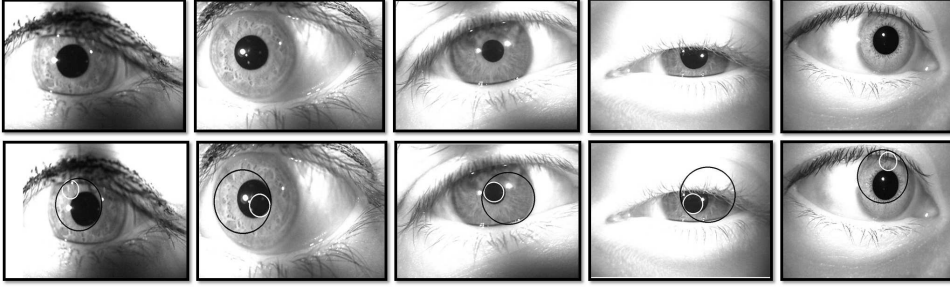


Figure 2.10: Samples of poor iris segmentation on images obtained from CASIA-V4 iris database.

to assess the quality of an iris template. The measure can discriminate between natural iris patterns from the artifacts that occur during compression. It is observed that the correlation between consecutive rows of an iris template increases with compression as the less significant features are lost. The metric uses this distance measure of randomness of features as a measure of biometric quality of an iris sample.

It must be observed that the quality metrics in current literature assume accurate segmentation of the iris region as a precursor to the assessment module. However, as illustrated in Figure 2.10, iris segmentation methods are also adversely affected by the above-mentioned covariates. Recently, it has been shown that local quality metrics are able to predict iris segmentation performance [7]. Further, there is a lack of a benchmark approach and test-bed evaluation for academic and commercial iris quality assessment techniques. Considering the low complexity of the prevalent Hamming distance matching function, it might be interesting to consider a predictive quality assessment method similar to NFIQ.

2.3.3 Face Quality Assessment

It is well established that quality measures are an important feature of modern face biometric systems due to the large degree of variations possible in face images (illustrated in Figure 2.11). However, quality assessment of faces has received comparatively less attention. Early research focuses on complete automation of essential capture guidelines in standards such as International Civil Aviation Organization (ICAO) and ISO. However, these guidelines are designed for manual recognition and provide minimal information about the quality of face biometric. More research focus must be directed towards this problem, since it has been observed in several empirical studies including the findings of biometric grand challenges that the covariates of face recognition (pose, illumination, expression, noise) affect the performance across different types of features or systems. A

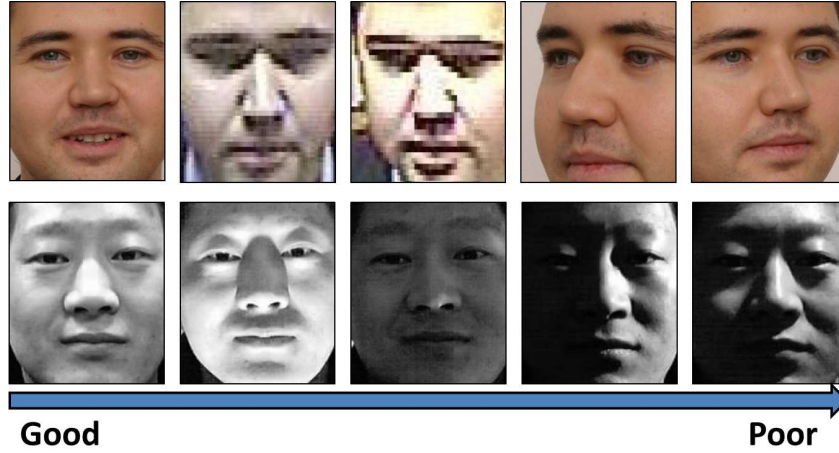


Figure 2.11: Face images illustrating different levels of biometric quality.

Table 2.5: A representative list of face quality assessment algorithms

Algorithm	Description
Subasic <i>et al.</i> [176]	Seventeen automatic tests in conjuncture with the <i>ICAO</i> face image presentation standards
Hsu <i>et al.</i> [78]	Automatic evaluator of the ISO/JEC 19794-5 face standards
Youmaran and Adler [213]	Biometric information defined from information theory
Gao <i>et al.</i> [65]	Asymmetry in LBP features [136] as a measure of the quality
Zhang <i>et al.</i> [214]	Asymmetry using SIFT features
Wong <i>et al.</i> [204]	Comparison of a facial image with <i>ideal</i> face models
Nasrollahi and Moeslund [126]	Geometrical pose estimation using face bounding box
Long <i>et al.</i> [112]	Assess sharpness, brightness, resolution, and pose in NIR videos
Yao <i>et al.</i> [209]	Sharpness measure from frame selection

discussion of the existing face quality metrics is presented here and a brief summary is also available in Table 2.5.

2.3.3.1 Still-face Images-based Techniques

Subasic *et al.* [176] present an evaluation scheme of a set of 17 automatic tests in conjunction with the ICAO face image presentation standards for automatic quality assessment. These tests are based on simple image processing techniques and semi-automatic annotation. The approach is tested on a set of 189 images. Further, the authors also mention some deficiencies in the ICAO standards such as lack of standard brightness, sharpness,

color balance, and tolerance of background. In a similar approach, Hsu *et al.* [78] present a more comprehensive evaluator for the ISO/JEC 19794-5 face standards. The approach combines several image quality metrics and face-specific metrics using facial feature detection. While a detailed description of the evaluation metrics is lacking, the authors evaluate several linear and nonlinear fusion schemes for match score prediction. Further, the authors use a nonlinear neural network, with the proposed set of quality metrics as feature vector, to predict the match score of a commercial face matching system.

Youmaran and Adler [213] discuss information content in biometric images termed as *Biometric Information* (BI). From the information theory perspective, BI is defined as the decrease in uncertainty of the identity of a person caused by the feature set. Assuming each feature to be a multivariate random variable, BI is modeled as the relative entropy $\Delta D(p||q)$ between the intra-person feature distribution $p(x)$ and the inter-person feature distribution $q(x)$.

$$\Delta D(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx \quad (2.1)$$

The approach is limited by the validity of the distribution q which is the model for all possible faces. While this research provides good insight into quality assessment, the algorithm is not practical to implement, since it requires a statistically valid number of samples for each subject and probe subject to estimate the distribution of subject's features. Klare and Jain [97] propose a perceived uniqueness measure of a given face sample and match scores from any face matcher. The measure computes the distance of a match score to a set of imposter scores, thus indicating face uniqueness. Gao *et al.* [65] proposed the use of asymmetry in LBP features [136] as a measure of the quality of face biometric. However, this approach is limited in applicability as the face image must first be normalized to scale for the measurement to be accurate. The authors attempt a laborious solution of training a model for each possible scale. Zhang and Wang [214] improve on this intuition using Scale Invariant Feature Transform (SIFT) features [113]. It is suggested that illumination variation primarily affects face recognition systems. The assessment of quality is based on the assumption that given a normalized frontal face image, the location of SIFT-based feature points will be symmetric with a vertical axis. Based on this observation, quality is estimated as the ratio of the number of available points on each side of the axis. The work does not discuss any guarantee that the SIFT features are symmetric over any axis in good quality images. Further, any natural asymmetry in face, any symmetric illumination, or other noise can lead to incorrect estimation.

Recently, quality assessment in face images has renewed interest attributed to insights from the Good, Bad, and Ugly (GBU) dataset [146]. The challenging dataset used in Face Recognition Vendor Test (FRVT) 2006 [148] consists of 9,307 frontal neutral expression face images taken in indoor or outdoor settings from 570 subjects. From this dataset, a subset of 2,170 images from 437 subjects is chosen and split into three sub-partitions (*Good*, *Bad* and *Ugly*) such that the fusion of the top three algorithms from FRVT 2006 results in GAR of 0.98, 0.80, and 0.15 at an FAR of 0.001. Further, no image appears in more than one subset and the subjects in all three partitions are the same. This unique partitioning of data enables researchers to focus on the hard matching problems of face recognition within the database. Also, this dataset can be used to better understand and model the change in *recognizability* of a subject in different environmental conditions. Phillips *et al.* [21, 144] show that simple image quality metrics can be combined to predict face recognition performance. Using a greedy pruning approach, ranking is predicted from a quality oracle. Aggarwal *et al.* [3] show that good, bad, ugly pairs can be predicted by using partial least square regression between image-based features (sharpness, hue, and intensity) and geometric attributes of a face (obtained using active appearance modeling). Hua *et al.* [81] use modulation transformation function to compute the sharpness in face images. Their results also indicate that sharpness is an important factor to improve face recognition results.

2.3.3.2 Video-based Techniques

An important application of quality assessment in face biometrics is in video face matching [168]. Here, face recognition is performed on a video stream rather than a single still image. Some approaches of this branch of research use quality assessment for *frame selection* in order to match the best possible frame from gallery and probe face video. Wong *et al.* [204] present a patch-based approach using the first d low-frequency components of the Discrete Cosine Transform (DCT) obtained from each facial patch. A multivariate probabilistic model is generated using a training set of frontal faces with acceptable illumination per patch, and the probe image is compared, patch-wise, to obtain the overall quality.

The general approach for video face quality assessment is based on comparing the input face image with face models developed from *ideal* example sets. In another approach, Nasrollahi and Moeslund [126] present a simple geometrical approach based on the dimensions of the bounding box of face detection algorithm in a video face recognition system. Since pose is a primary challenge in such systems, this approach can be considered as a simple pose assessment technique. A similar approach is also used recently by Long and

Li [112] for NIR video face recognition. Yao *et al.* [209] use a sharpness measure from frame selection for a recognition system designed for low-resolution face videos. It must be noted that while face quality assessment has received considerable attention in video face recognition research, the requirement in this particular application is for a *binary* decision (accept/reject) per video frame. Hence, such quality metrics may not sufficiently measure the quality of the face biometric sample.

The unique attribute of FRVT 2006 [148] is in providing several thought-provoking insights and directions to the problem of quality assessment in face recognition [65]. These findings are discussed by Beveridge *et al.* [20, 21, 22] with a detailed analysis of the *effect* of various *subjective* and *objective* covariates of face biometric. Current literature describes the quality of a face image as an intrinsic property of the image. Beveridge *et al.* [22] argue that if this intuition were true, a higher-quality sample would be consistently matched correctly. Likewise, a low-quality sample would consistently perform poorly. However, their experiments indicate that the confidence of match is dependent on the quality of both the images being matched, i.e., a considerable number of images that are hard to recognize as part of one match pair are easy to recognize as part of other match pairs. This indicates that verification can be correctly performed if both images lie in the same quality space. The NIST Multiple-Biometric Evaluation 2010 (MBE) [73] presents six state-of-the-art commercial face recognition systems on various demographic and covariate challenges which indicate that the performance of all algorithms is affected by various factors such as gender, age, and ethnicity, apart from known covariates of pose, illumination, and expression. Hence, it follows that a quantitative measure of quality of an input face image that provides an estimate of matching performance is critical. Recently, holistic descriptors extracted from the face region are shown to be good indicators of performance of face recognition systems [27]. The low computation time of these image descriptors make them ideal features for quality assessment. Further, pseudo-labels of quality obtained from matching performance provide a direct estimate of recognizability of a given face image. Therefore, the approach is more useful than separate estimation of different covariates. The large degree of freedom of face greatly increases variability in captured information compared to other biometric modalities, making quality assessment an essential prerequisite. For face recognition systems to have robust performance outside of studio-like conditions, quality assessment of face must encapsulate the aforementioned covariates effectively.

2.3.4 Evaluating Quality Assessment Approaches

An important aspect in the development of quality assessment algorithms is the way their performance is measured. Since the primary motivation of most image quality assessment techniques is in perceptual understanding of the image, human annotation of quality can be considered for comparison and testing of automatic algorithms. A set of volunteers is presented with images of different quality and their responses are aggregated to a mean operator score (MOS). A high correlation between the predicted quality and MOS from volunteers indicates high performance [43]. Based on the aforementioned discussion, MOS cannot be directly applied for biometric quality, as there is no conclusive evidence that human interpretation of quality correlates with the quality in terms of the performance of a recognition algorithm. In our observation, six prominent methods of evaluation of biometric quality metrics persist in literature apart from evaluation using MOS:

- Correlation analysis: As noted by [142], a biometric quality metric must be a good classifier performance predictor. With this view, a quality measure that is highly correlated (statistically) with match scores obtained from a classifier is the most desirable. Hence, several researchers discuss correlation with genuine match scores [10]. Since every match score can be associated to the quality of both gallery and probe sample, combining methods, such as $Q_{gallery} + Q_{probe}$ or $\sqrt{Q_{gallery} \times Q_{probe}}$ or $\min(Q_{gallery}, Q_{probe})$ are utilized.
- Modeling: Recently, quality metrics are utilized as predictors for dynamic processing and context switching. When correlation is established, the relationship between a series of quality scores (predictors) and associated match score (response) can be explicitly described by *modeling* using regression analysis, as shown subsequently in this research. Further, the goodness-to-fit can be evaluated by analysis of variance and inspection of residual error of fitting.
- Quality bins: In another approach, the impact of quality metrics is measured by segregating the entire dataset into a number of quality bins and performing individual recognition experiments on each of them. Further, the intuition that better quality data has better recognition accuracy is substantiated with recognition results on these quality bins [36, 40, 60, 91].
- Distance metric: Quality score is also used to alter the feature space to improve matching. Chen *et al.* [40] incorporate their proposed iris quality assessment metric

(computed for both gallery and probe) in the formulation of Hamming distance matcher to show improved results when compared to simple Hamming distance.

- Cross-correlation: Another possible method of evaluating quality metrics is by computing the cross-correlation between the given metric and various existing metrics [60]. In biometrics, this can be considered as a weak measure unless some additional benefits of the algorithms (in terms of computation time or better correlation with MOS) is described that differentiate from existing approaches.
- Computation time: The performance of a quality assessment algorithm in terms of computation time is an important aspect of its evaluation. In most use-cases, performing quality assessment is only meaningful when complexity is low. For instance, biometric quality assessment can only be a small overhead to the recognition pipeline. Reported computational time of a quality metric is dependent on the implementation platform and machine configuration in use. However, computational efficiency of techniques reported relative to computation time of PSNR allows for a machine-independent comparison [119].

2.4 Analysis of Quality Metrics

Quality metrics have been extensively used to improve the robustness and accuracy of biometric systems. Several fusion and context-switching approaches are proposed based on the intuition that quality can be indicative of the *utility* of a biometric sample. However, as discussed in Section 2.2, the role of a quality metric in improving the performance of a biometric system is not always implicit. Hence, an arbitrary quality metric ‘ q ,’ defined in abstraction in various formulations of multibiometrics, must be investigated more closely. In this section, a representative set of image and biometric quality metrics is evaluated to understand their relationship with each other and with match scores. For the evaluation, match scores obtained from commercial matchers are used on WVU multimodal biometric database.

2.4.1 Database and Evaluation Protocol

The evaluation is performed on the WVU multimodal database [45] that contains face, fingerprint, and iris modalities. For the experiment, two images pertaining to 250 subjects (per modality) are chosen for gallery and the remaining images are used as probe. To evaluate the performance of quality metrics, three uni-modal biometric matchers are used.

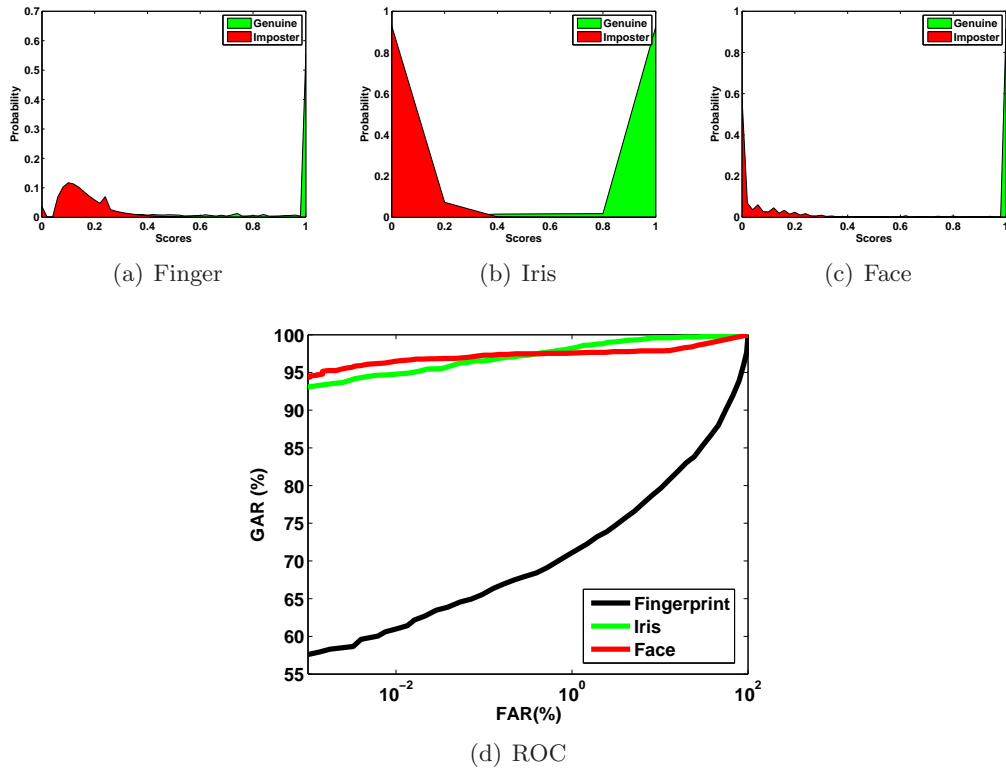


Figure 2.12: Genuine and imposter score distribution for (a) face, (b) fingerprint, and (c) iris matchers on the WVU multimodal dataset. (d) Receiver Operating Characteristic (ROC) curve illustrates the verification performance of the respective matchers indicating the overall quality of the database.

Fingerprint classifier used in this study is the NIST Biometric Image Software (NBIS) [132]. NBIS consists of a minutiae detector called MINDTCT and a fingerprint matching algorithm known as BOZORTH3. For face and iris biometrics, Neurotechnology feature extractors and matchers are used. The performance of the matchers is illustrated in Figure 2.12. The varied image quality result in a considerable overlap of genuine and imposter score distributions.

As discussed in previous sections, quality metrics can be either image-based or modality-specific. A representative set of quality metrics of both types are chosen for evaluation. Specifically, four image quality approaches and a biometric quality approach (that may each contain multiple measures) are considered for the evaluation. The abbreviations associated with each of the quality metrics are presented in Table 2.6 and a brief description is presented below. The techniques are all no-reference quality metrics and have low computational complexity when executed on a typical desktop machine. A detailed discussion of the computational complexity of each technique is available in the references:

- Spectral energy (SE) calculates the block-wise energy using Fourier transform components [131]. It describes abrupt changes in illumination and specular reflection. The image is tessellated into several nonoverlapping blocks, and the spectral energy is computed for each block. The value is computed as the magnitude of Fourier transform components in both horizontal and vertical directions that shows the amount of spectral energy per block.
- Marziliano *et al.* [118] have proposed edge spread (ES) as a measure to estimate irregularities based on edges and their adjacent regions. Specifically, it computes the effect of irregularity in an image based on the analysis of the difference in image intensity with respect to the local maxima and minima of pixel intensity at every row of the image. Edge spread can be computed in horizontal as well as vertical directions. However, the experiments in [118] show that either of the two directions suffices for quality assessment.
- A no-reference perceptual quality metric by Wang *et al.* [198] primarily measures compression artifacts. It is computed as the combination of blockiness and activity estimation in both horizontal and vertical directions, manifesting in three metrics: blockiness (B), activity (A), and zero-crossing rate (Z).
- A spatial domain no-reference quality assessment technique, termed *BRISQUE* (BR), proposed by Mittal *et al.* [119], provides a holistic assessment of *naturalness*. The quality metric is a deviation measure of a natural image from the regular statistics, indicating distortion.

Further, three modality-specific quality metrics are also used:

- Iris: Kalka *et al.* [91] evaluates defocus (DF), motion blur (MB), occlusion (O), illumination (I), specular reflectance (SR), and pixel count (PC). Further, a fused metric (Q) is obtained using DS-theory. The technique is discussed in Section 2.3.
- Fingerprint: As described in Section 2.3, Chen *et al.* [40] proposed ridge energy for fingerprint quality assessment. It is the Fourier spectrum energy computed on a frequency bandpass region where fingerprint ridges strongly manifest. In addition, a discrete quality value obtained from the NFIQ [177] tool is also utilized in this study.

Table 2.6: Various representative quality metrics considered in this study

Abbreviation	Quality metric
SE	Spectral energy
ES	Edge spread
B	Blockiness
A	Activity
ZC	Zero count
BR	BRISQUE
DF	Defocus
M	Motion
O	Occlusion
L	Lighting
S	Specular reflectance
PC	Pixel count
Q	Fused iris quality
RE	Ridge energy
NFIQ	NIST fingerprint image quality
P	Pose
F	Focus

- Face: For face quality assessment, geometric pose estimation (P) is computed. First, positions of eyes and mouth are estimated using corresponding Adaboost detectors [141]. Pose is estimated based on the deviation of geometric measures (inter-eye distance and eye-center to mouth distance) from mean values. Additionally, focus measure (F) reported in [20] is also utilized.

2.4.2 Experimental Analysis

Two key ideas are evaluated in this study: (i) the relationship between different quality metrics and (ii) the relationship of the quality of a pair of biometric samples with their

Table 2.7: Spearman correlation between face quality scores

	SE	ES	B	A	ZC	P	F	BR
SE	1.00	0.14	-0.02	-0.02	-0.03	-0.11	-0.07	0.12
ES		1.00	-0.06	0.09	-0.12	-0.04	0.08	-0.12
B			1.00	0.97	-0.15	-0.06	0.31	-0.57
A				1.00	-0.15	-0.07	0.29	-0.56
ZC					1.00	-0.10	-0.11	-0.33
P						1.00	0.08	0.08
F							1.00	-0.27
BR								1.00

Table 2.8: Spearman correlation between fingerprint quality scores

	SE	ES	B	A	ZC	RE	NFIQ	BR
SE	1.00	0.12	0.12	0.17	0.00	0.52	−0.02	0.12
ES		1.00	−0.21	−0.19	−0.18	0.05	0.03	−0.08
B			1.00	0.94	−0.40	−0.03	−0.30	0.61
A				1.00	−0.30	−0.04	−0.33	0.68
ZC					1.00	0.03	0.22	−0.37
RE						1.00	0.08	0.00
NFIQ							1.00	−0.27
BR								1.00

Table 2.9: Spearman correlation between iris quality scores

	SE	ES	B	A	ZC	DF	M	O	L	S	PC	Q	BR
SE	1.00	0.02	0.25	0.29	0.18	−0.09	0.03	−0.07	0.01	−0.11	−0.10	0.13	−0.05
ES		1.00	0.00	−0.01	−0.16	−0.06	−0.06	−0.03	0.01	0.13	0.02	0.09	−0.01
B			1.00	0.97	0.33	−0.53	−0.15	0.05	0.18	−0.08	−0.03	0.48	−0.23
A				1.00	0.38	−0.49	−0.12	−0.02	0.16	−0.15	−0.09	0.45	−0.19
ZC					1.00	−0.09	0.29	−0.05	−0.23	−0.43	−0.22	0.11	−0.02
DF						1.00	0.12	−0.12	−0.20	−0.15	−0.13	−0.57	0.14
M							1.00	−0.08	−0.15	−0.04	−0.10	−0.23	0.09
O								1.00	0.02	0.56	0.92	−0.27	0.03
L									1.00	0.07	0.06	−0.09	−0.03
S										1.00	0.74	−0.12	−0.06
PC											1.00	−0.30	0.05
Q												1.00	−0.24
BR													1.00

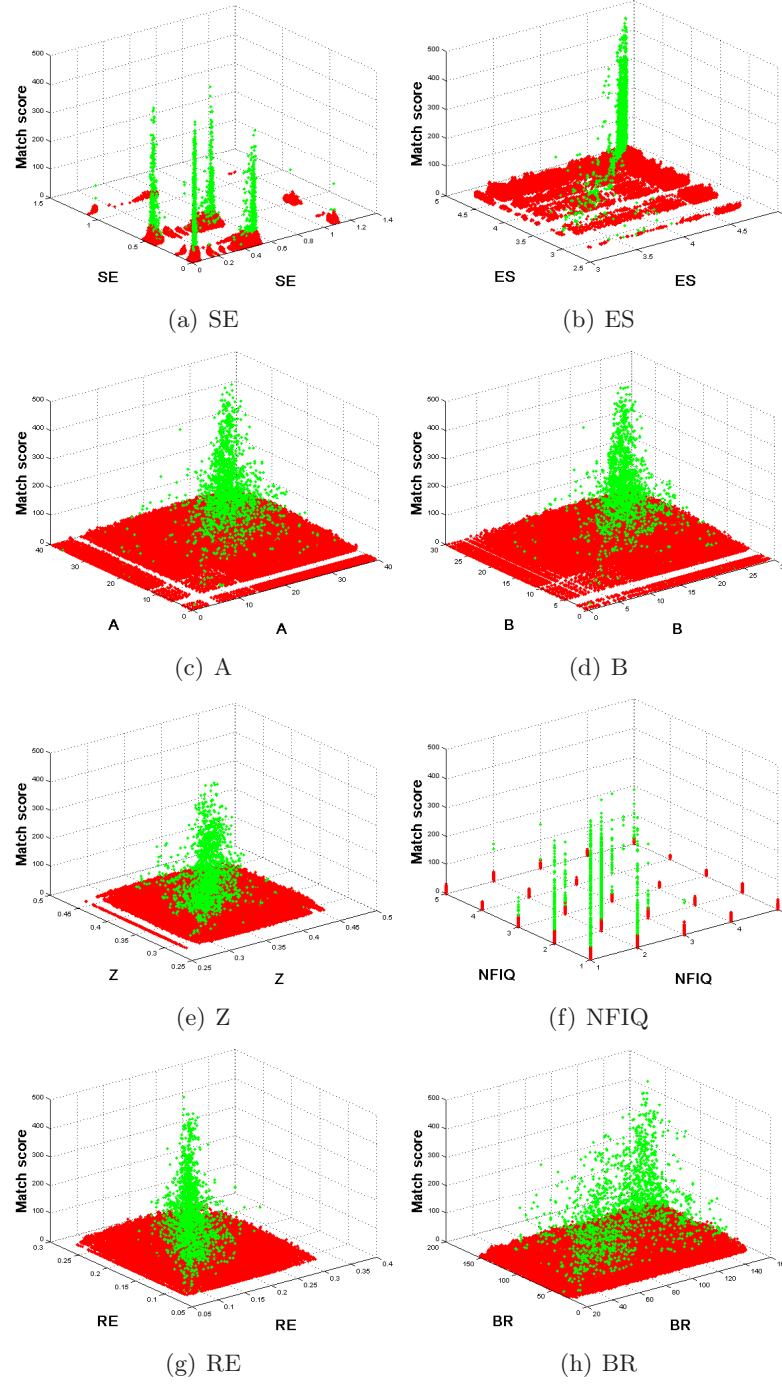


Figure 2.13: Relation between match scores obtained from NBIS fingerprint matcher and various quality metrics. Relation between match scores obtained from NBIS fingerprint matcher (z -axis) and various quality metrics (a) SE, (b) ES, (c) A, (d) B, (e) Z, (f) NFIQ, (g) RE, (h) BR for genuine (green) and imposter (red) match pairs. The x -axis pertains to gallery quality, while y -axis pertains to the probe quality. The scattering indicates that ES, A, B, Z, RE, and BR quality metrics can characterize genuine scores.

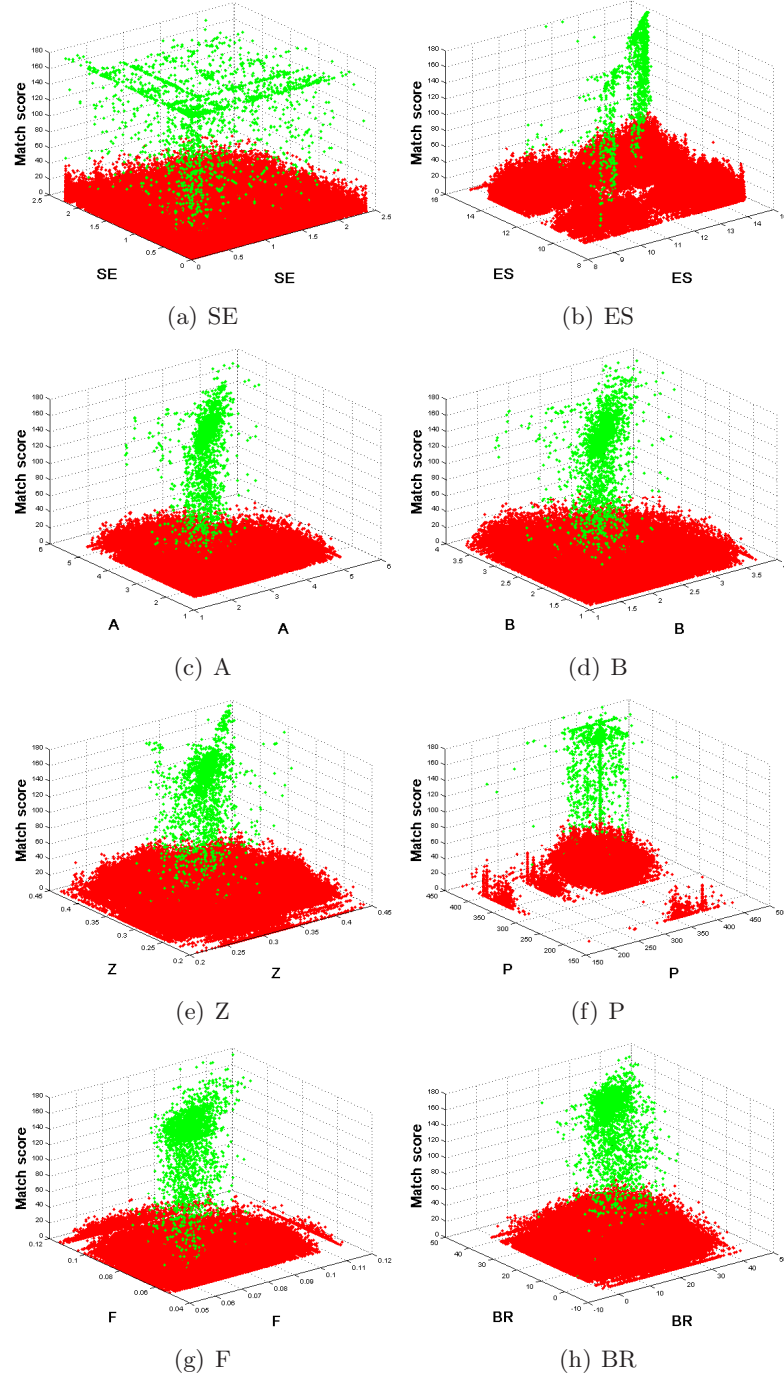


Figure 2.14: Relation between match scores obtained from a commercial face matcher and various quality metrics. Relation between match scores obtained from a commercial face matcher (z -axis) and various quality metrics [(a) SE, (b) ES, (c) A, (d) B, (e) Z, (f) P, (g) F, (h) BR] for genuine (green) and imposter (red) match pairs. The x -axis pertains to gallery quality, while y -axis pertains to probe quality. The scatterplot indicates that A, B, Z, F, and BR quality metrics can characterize genuine scores.

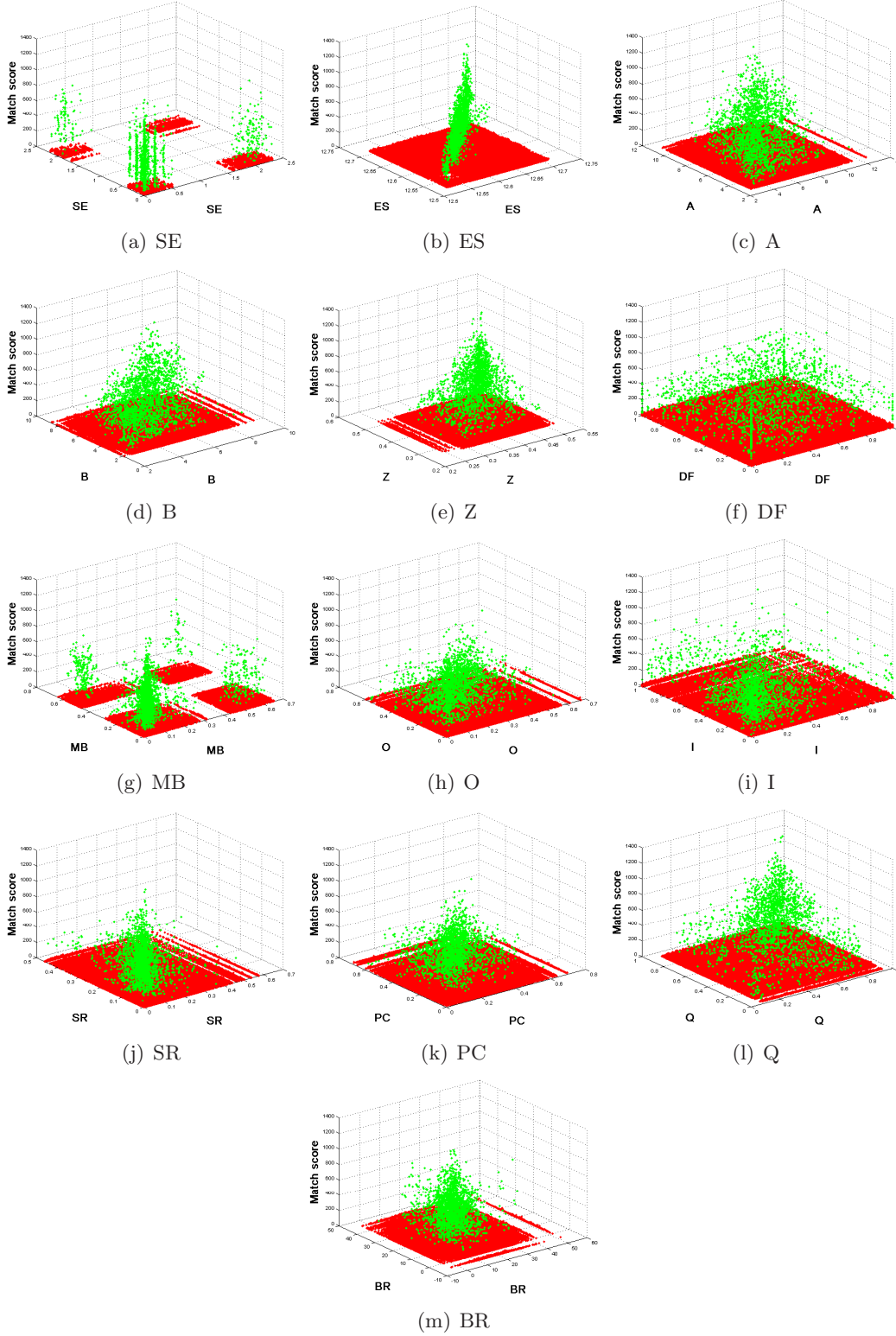


Figure 2.15: Relation between match scores obtained from a commercial iris matcher and various quality metrics. Relation between match scores obtained from a commercial iris matcher (z -axis) and various quality metrics [(a) SE, (b) ES, (c) A, (d) B, (e) Z, (f) DF, (g) MB, (h) O, (i) I, (j) SR, (k) PC, (l) Q, (m) BR] for genuine (green) and imposter (red) match pairs. The x -axis pertains to gallery quality while y -axis pertains to probe quality. The scatterplot indicates that ES, A, B, Z, SR, PC, and BR quality metrics can characterize genuine scores of match pairs. However, DF, O, I, and Q are unable to characterize genuine match scores.

match score. All match scores are converted to similarity measures for easy visualization. Some key insights can be drawn for both image-based and biometric-specific quality metrics as follows:

- Spearman correlation values for all quality metrics for face, fingerprint, and iris images are shown in Tables 2.7, 2.8, 2.9 respectively. The quality score from gallery and probe pair is combined as $Q = \sqrt{Q_{\text{gallery}} \times Q_{\text{probe}}}$. Low Spearman correlation is observed between the quality metrics in consideration indicating that they measure diverse aspects of quality. For instance, no-reference quality measures A in 8×8 blocks in the image. On the other hand, ES measures the gradient difference at edge boundaries, to measure blurring. Even though both are measures of blurring, the difference in approaches leads to low correlation between them.
- Scatter plot in Figures 2.13, 2.14, 2.15 illustrates genuine and imposter match scores against each quality metric in consideration. A three-dimensional plot of match scores versus quality of gallery and probe clearly illustrates the characteristic relation between them.
- For all three modalities, no relation is observed between quality scores and imposter match scores. A similar observation is made in the case of fingerprints in [10].
- In case of certain quality scores such as Activity, Zero-Cross rate, and Focus, genuine match scores are found only in specific quality bins. Hence, any pair exhibiting quality in this range during test phase induces more confidence in matching [107]. Such simple quality measures provide an additional information to improve classification. For example, in case of A of fingerprints, the values pertaining to genuine scores are observed in the range of 15 and 25.
- For face and iris modalities, quality metrics that measure prominence of edges better map to genuine scores. For instance, ES and RE provide more confidence to genuine score than other metrics such as DF. Further, spatial no-reference measure (BR) correlates with activity measures and also characterizes the genuine scores for face and fingerprint.
- In order to evaluate the relevance of quality scores in augmenting or predicting match scores, an illustration of the cumulative density function (CDF) is presented in Figure 2.16. The CDF of certain quality scores are more similar to the obtained match scores, such as RE, B, O, and I as compared to ES, BR, and Z.

- To test the relationship between the quality scores and match scores obtained from each modality, a linear regression analysis is performed between the genuine scores and quality scores. As discussed previously, the quality scores from gallery and probe are combined as $Q = \sqrt{Q_{\text{gallery}} \times Q_{\text{probe}}}$. Further, the data is randomly split into nonoverlapping train and test sets. The mean squared error (MSE) of each modality, over ten times random cross-validation, is shown in Figure 2.17. It is observed that even with 10% of the data as training samples, genuine scores from matchers can be predicted with quality metrics using a simple linear model. To analyze the quality of fit of the regression model, Analysis of Variance (ANOVA) is performed to assert the effect of each quality metric in consideration as match score predictors. The analysis indicates that ES, A, DF, MB, O, PC, Q, and BR are effective with p value less than 0.01 for iris modality. On the other hand, SE, ES, B, A, and Z are more effective in estimating match scores for fingerprints. We also observe that only P and ES are able to estimate match scores of the face.

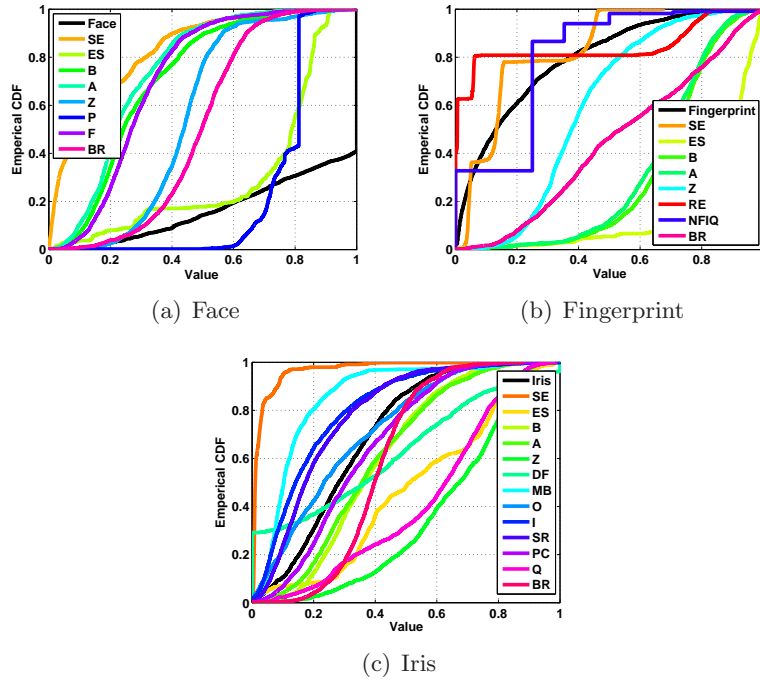


Figure 2.16: The cumulative density functions (CDF) between genuine score and quality metrics for (a) face, (b) fingerprint, and (c) iris modalities. The plots compare the distribution of each quality metric with the corresponding genuine score distribution.

In this study, it is empirically established that a direct relationship exists between certain quality metrics and match scores (which can also be viewed as classifier confidence).

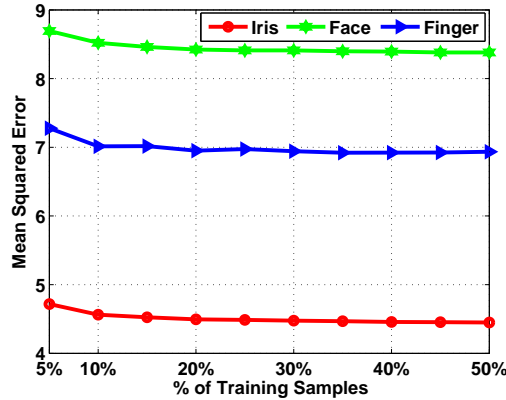


Figure 2.17: Results of the regression test. MSE of the regression test with genuine scores and quality metrics accumulated over 10 times cross-validation. Even with a small number of training samples, a linear model can predict match scores of genuine pairs showing that quality scores can be indicative of matching performance.

This encouraging result sanctions the use of quality metrics in multibiometric schemes such as quality-based fusion and context-switching. However, as observed from the scatter plots, the choice of quality metrics is an important factor.

2.5 Can Holistic Representations be used for Biometric Quality Assessment?

Biometric systems deployed in unconstrained environments, for example, large-scale identity projects such as Aadhaar and US-Visit, encounter varying quality of input face samples as shown in Figure 2.18. To improve the performance, usability, and robustness of such systems, recent research in face biometrics use *quality* of the sample not only to reject the poorly captured samples but also within the recognition process. The active involvement of quality scores beyond the capture stage encourages the formulation of more complex and accurate quality assessment techniques. Current research in face recognition generally uses simple image processing algorithms that are able to assess image degradations due to noise, compression or illumination. While the quality of a face image is susceptible to degradation during capture and storage, it may also have low quality by its very nature. For example, a high resolution face image with acute pose is of low biometric quality, irrespective of the high image quality. The complexity of the problem is further exacerbated by the lack of consensus in literature on facial (biometric) features.

Table 2.10 summarizes important approaches in face quality assessment. Early research in face quality [78, 176] focuses on complete automation of essential capture guidelines in



Figure 2.18: Face images of varying quality encountered by a face recognition systems.

face standards such as ICAO and ISO/JEC 19794–5 [79]. The effects of resolution and capture conditions, with an analysis of *subjective* and *objective* covariates of face biometric in Face Recognition Vendor Test (FRVT) 2006 is presented in [20, 22]. A probabilistic approach for performance prediction using background information (capture conditions, gender, race) is discussed [20]. Further, considerable research on leveraging simple quality metrics to improve multibiometrics recognition is summarized in [152]. Image quality assessment metrics, with focus on perceptual quality, are reviewed in [197].

This research explores a new direction in *face quality assessment using holistic representations*. Scene recognition techniques extensively use holistic features that are designed to encode commonalities in a large collection of scene images for image classification. As opposed to biometric features, that encode *unique* attributes of an image, scene recognition techniques effectively encode abstract and categorical features of an image such as vertical or horizontal structures in city images, and openness in landscape images. It is our assertion that these features can be used to *predict* the *usability* of a face image and segregate face images into abstract categories that are indicative of quality. The experiments on a heterogeneous database consisting of several covariates show that holistic image descriptors are able to successfully categorize biometric images (using a classifier) into quality bins ranging from *poor* to *excellent* quality, *that correlates with recognition performance*. Further, as a case study, improved face recognition performance is observed when the proposed approach is used to reject poor quality samples.

2.5.1 Quality Assessment of Face Biometric

Research in scene recognition has shown that holistic representation of an image is consistent with abstractly classifying images into broad categories such as *buildings*, *coastline* and *forests*. Inspired by this observation, a learning based approach to quality assessment is proposed in this research. The mapping between recognition performance based *quality labels* and holistic representation of images is learned in a supervised setting and utilized to predict quality.

Table 2.10: Summary of a representative list of existing approach in face quality.

Technique	Description
Subasic <i>et al.</i> [176]	17 automatic tests for <i>ICAO</i> standards.
Hsu <i>et al.</i> [78]	Automatic evaluator of ISO/JEC19794–5 face standards.
Youmaran and Adler [213]	Biometric information defined from information theory.
Gao <i>et al.</i> [65]	Asymmetry in LBP features as a measure of the quality.
Zhang <i>et al.</i> [214]	Asymmetry using SIFT features.
Wong <i>et al.</i> [204]	Comparison of a facial image with <i>ideal</i> face models.
Nasrollahi <i>et al.</i> [126]	Geometrical pose estimation using face bounding box.
Yao <i>et al.</i> [209]	Sharpness measure for frame selection.
Proposed	Use holistic descriptors with match score based pseudo-labels for quality prediction.

First, quality labels are generated based on match score distribution obtained from a powerful matcher. Next, these labels are assigned to a set of training images with different image and biometric degradations (illumination, low resolution, occlusion, and expression). A non-linear relation between these labels and multi-dimensional holistic descriptors is learned using a multi-class classifier.

2.5.1.1 Face Quality as Match Score Predictor

We develop the intuition of such an approach to quality assessment, from quality based match score prediction. As shown by Grother and Tabassi [142], there is a relationship between quality of a biometric sample and recognition accuracy. For a quality assessment algorithm (Q_1) that produces a scalar quality metric q , the match score $s_{d,d'}$ between the samples d and d' can be modeled using a predictor P as,

$$s_{d,d'} = P(Q_1(d), Q_1(d')) + \epsilon_{d,d'}, \quad (2.2)$$

The predictor P estimates the similarity score based on the quality of the templates ($\epsilon_{d,d'}$ is the error in that prediction). This problem of biometric match score prediction is challenging since Q produces a single quality value. However, Vatsa *et al.* [187] and Bhatt *et al.* [29] present evidence indicating that a comprehensive quality measure must be a vector rather than a scalar. Hence, Eq. 2.2 is redefined as,

$$s_{d,d'} = P(\vec{q} = Q_2(d), \vec{q}' = Q_2(d')) + \epsilon_{d,d'}, \quad (2.3)$$

where \vec{q} and \vec{q}' are the quality vectors of samples d and d' that may provide more information for P . In this research, \vec{q} is a multi-dimensional holistic representation of the probe image that preserves non-localized, categorical information of the image.

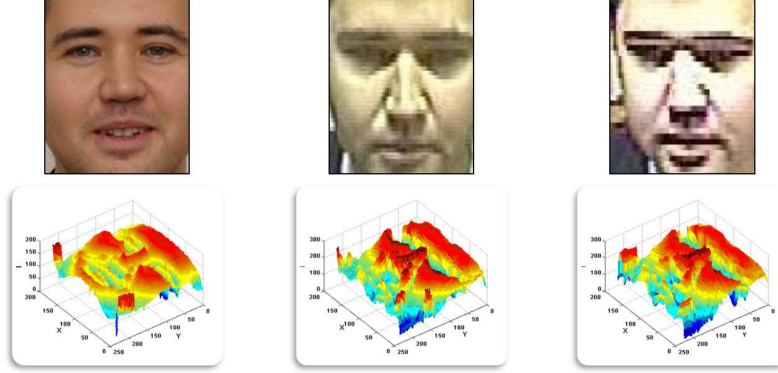


Figure 2.19: Face images with degradations exhibit more *roughness*, evident from the surface plots (z-axis is pixel intensity (I)). Roughness can be captured with holistic features and may be indicative of biometric quality.

2.5.1.2 Holistic Image Representations

In this research, two prominent holistic representations, Gist [138] and sparsely pooled HOG [47] are considered. As illustrated in Figure 2.19, poor quality face images have a typical *roughness* in intensity values as compared to a good quality image. The abstract and non-localized nature of Gist and HOG make them good candidates to assess biometric quality. Further, compared to local image descriptors, the feature length of Gist (512) and HOG (81) can ensure low computational time for quality assessment. A brief summary of Gist and HOG is presented below.

Gist: Olivia and Torralba [138] have proposed a holistic representation of the spatial envelope of a scene image. Rather than viewing an image as a configuration of objects, in this model they are viewed as an unitary model. The spatial properties of the image are well preserved in such a representation of the spatial envelope (referred to as GIST). These coarse features are extracted at highly abstract level by using windowed Fourier transform. To assess the utility of a face biometric sample, we propose to use low dimensional representations of the face images. Here a set of five perceptual dimensions, namely, naturalness, openness, roughness, expansion and ruggedness are used to compute low dimensional, holistic representation of the image. The nomenclatures is derived from scene recognition research and we assert that GIST [138] can be a good descriptor for biometric quality assessment for face.

1. Degree of Naturalness: This spatial property describes the distribution of edges in the horizontal and vertical orientations. It describes the presence of artificial elements such as spectacles.

2. Degree of Openness: The second major attribute describes the presence or lack of points of reference.
3. Degree of Roughness: This perceptual attribute refers to the size of the largest prominent object in the image. It evaluates the common global attributes of the image.
4. Degree of Expansion: This attribute describes the depth in the gradient of the space within the image.
5. Degree of Ruggedness: This attribute gives the deviation from horizontal by assessing the orientation of the contours of the image.

These perceptual properties are correlated with the second-order statistics and spatial arrangement of structured components in the image (for details of computing these properties, readers are referred to [138]). For a given face image I of size $M \times M$ with O number of orientations per scale, GIST is defined as a function f ,

$$GIST_{M,O}(I) = f(N, O, R, E, Rg) \quad (2.4)$$

where N = Naturalness, O = Openness, R = Roughness, E = Expansion, and Rg = Ruggedness. Once the GIST descriptors are calculated, they are classified using a RBF-kernel based multi-class SVM and a quality class label C is assigned.

$$C = mSVM(GIST_{M,O}(I)) \quad (2.5)$$

HOG: Dalal and Triggs [47] present a global descriptor for human (pedestrian) detection in street view images, known as histogram of gradient orientations (HOG). The approach has gained immense popularity in detection of humans, vehicles and animals in still imagery and videos. This is due to the low computation time yet surprisingly high accuracy and robustness across different postures. The algorithm is based on the intuition that the shape and position of the dominant object can be understood by the distribution of orientations in local regions of the image. The research extensively describes implementation of HOG and empirically analyzes the effects of different parameters on performance. A color image is first pre-processed using gamma-correction. Unidirectional gradient kernel is applied on the image to obtain orientations. Histograms of these orientation angles are collected and normalized based on the gradient magnitude. Depending on the variant of HOG to be used, the obtained histograms are pooled over densely overlapping windows. The histograms are often normalized by k -norm operation given by $\|v\|_k$ where

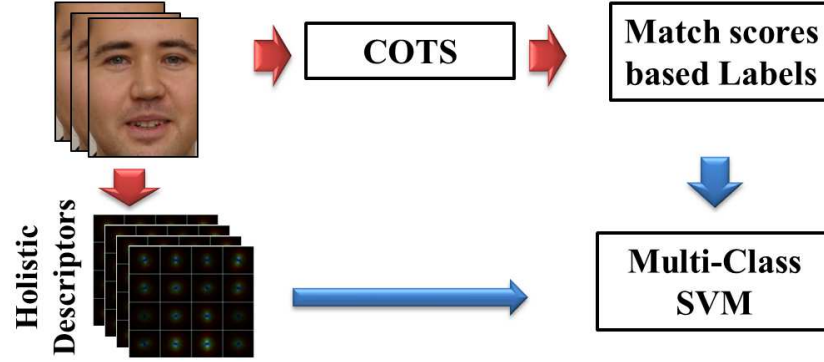


Figure 2.20: The training process of the proposed approach.

$k = 1, 2, 3$. The normalized descriptor is used as features for classification. In this research, this descriptor is used to classify pose in conjunction with SVM.

2.5.1.3 Quality Labels based on Face Matcher

As shown in Figure 2.20, the relationship between an image representation and quality label is learned from a training set using a non-linear classifier. The training samples are annotated based on the identification performance on the training set, inspired from [142]. The steps to obtain the quality label are as follows:

- A matching algorithm is used to obtain the match scores (s) on a training data that consists of a good quality (studio quality) image and several probe images of varying quality per subject. In order to minimize the misclassification rate, match scores obtained from two commercial off-the-shelf (COTS) face recognition system are fused using sum rule.
- The genuine match scores ($s_{d=d'}$) are Z -normalized and segregated into two sets. *Correct matches* refer to those genuine scores that result in Rank-1 matching. The remaining are referred to as *Incorrect matches*. Next, the empirical cumulative distribution function (ECDF) of both the sets are obtained (cdf_C, cdf_I), as illustrated in Figure 2.21. Further, the training probe samples are labeled, as *excellent*, *good*, *fair* and *poor* corresponding to the bins of match scores. The bin thresholds and number of bins can be varied according to the specific application scenarios.
- A one-vs-rest multi-class SVM with Gaussian kernel is trained for four bins of quality with the holistic descriptor as the input feature. The label corresponding to the most confident positive classification of SVM is selected in the testing phase.

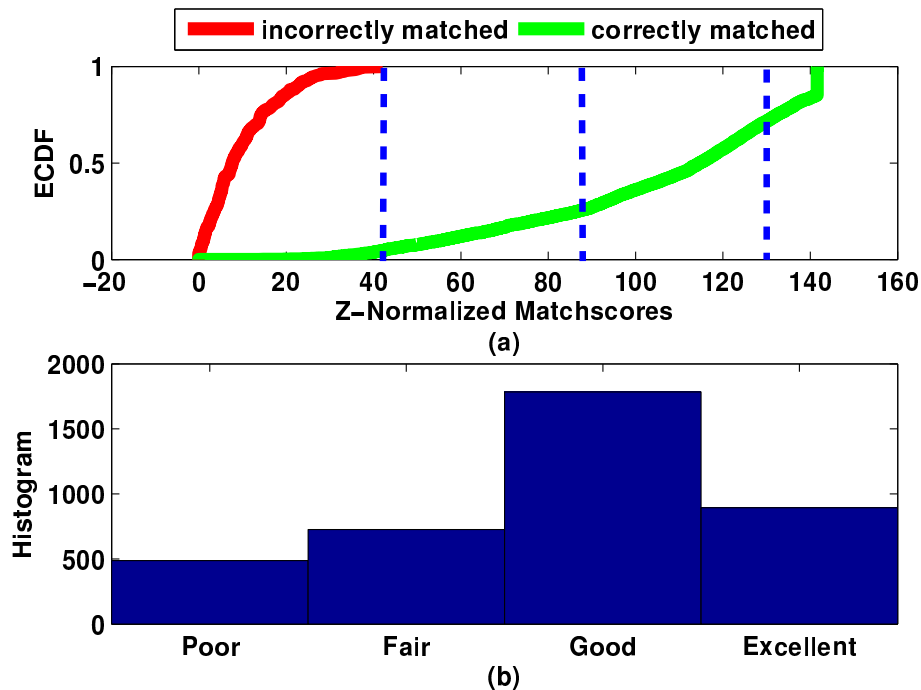


Figure 2.21: (a) The empirical cumulative density function (ECDF) of the z-normalized match scores (−20 to 160), (b) the number of samples per quality bin obtained for training.

Table 2.11: Summary of databases.

Database	Subjects (Train/Test)	Description
SCFace [71]	130 (39/91)	pose, low resolution
CAS-PEAL [64]	1040 (312/728)	pose, illumination, expression, accessories, background, distance
Combined	1170 (351/819)	all of the above

2.5.2 Experiment and Analysis

A face quality assessment technique must be aware of all the degradations that are encountered in face modality. A single face database is usually collected in similar settings and may lead to bias in quality assessment approaches that are based on training. Hence, in this research, a heterogeneous combination of two face databases, namely, the SCFace [71] and CAS-PEAL [64], with pose, illumination, expression, accessories, background, distance and resolution variations is used. Images corresponding to 30% of the subjects are used as training and the remaining as testing (summarized in Table 2.11). In both the training and testing phases, a single good quality image is used as gallery and the remaining images are used as probe. All the training samples corresponding to quality bins are used to train SVM and the parameters are obtained via grid search, with radial basis

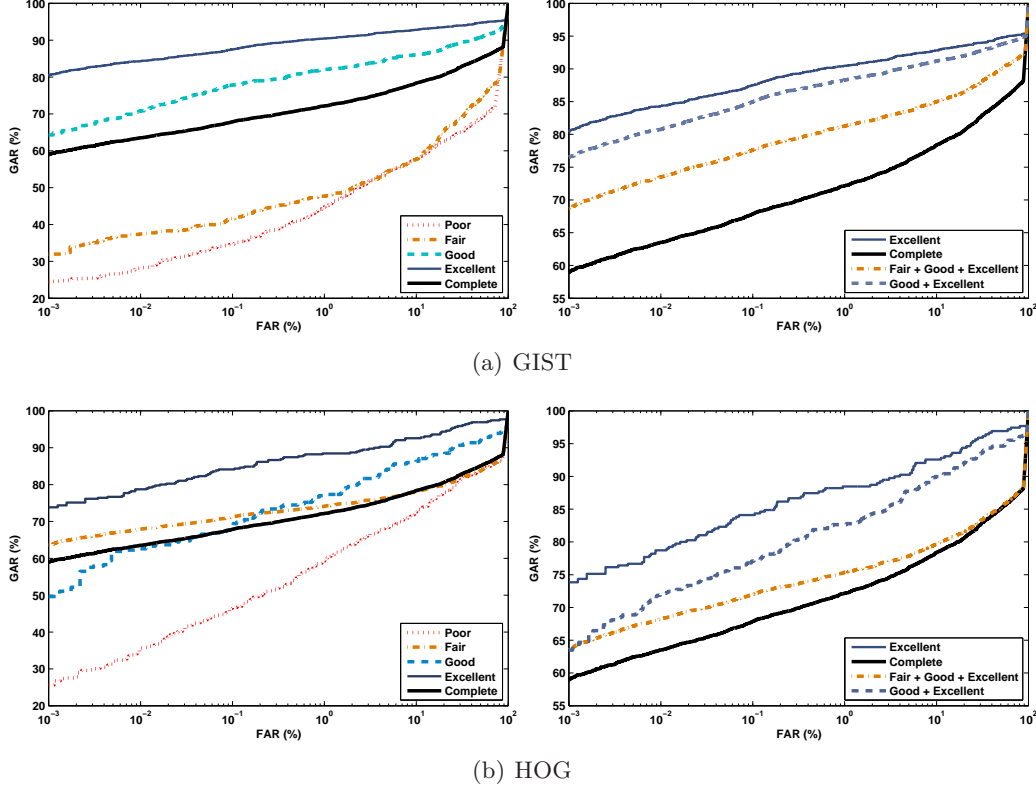


Figure 2.22: Verification performance of testing images when segregated into quality bins (left) and when lower quality bins are discarded (right) using a) Gist and b) HOG.

function as the kernel. To evaluate the correctness of quality labels, the identification and verification performance of each bin are computed separately using the better performing COTS, similar to the experimental procedure in [142].

- On the training database, the fusion of two COTS yields the rank-1 identification accuracy of 91.69%. Hence, all the *Incorrect matches* are marked as *poor* quality ($cdf_I^{-1}(1)$ ¹). Further, $cdf_I^{-1}(1)$ to $cdf_C^{-1}(0.25)$ are labeled *fair* quality, $cdf_C^{-1}(0.25)$ to $cdf_C^{-1}(0.75)$ as *good* and beyond $cdf_C^{-1}(0.75)$ as *excellent*. As mentioned, this configuration may be application dependent.
- Table 2.12 and Figure 2.22 show the performance of COTS on each of the quality bins obtained from both GIST and HOG. Better performance is observed for quality bins classified as *excellent* and *good* compared to *fair* and *poor*. Further, the percentage overlap for the genuine and imposter distributions is also increased for lower quality

¹Here, $cdf_X^{-1}(a)$ corresponds to the value of the random variable X where the cumulative density is less than or equal to a .

Table 2.12: Performance of COTS on each quality bin.

	Quality Bin	Count	% Hist. Overlap	Rank-1 %	EER
HOG	Excellent	390	2.65%	89.48%	7.58%
	Good	278	7.27%	75.89%	12.84%
	Fair	4198	15.06%	74.65%	20.15%
	Poor	639	16.18%	48.98%	21.50%
	Good + Excellent	668	17.43%	83.83%	10.03%
	Fair + Good + Excellent	4866	25.63%	75.91%	18.72%
Gist	Excellent	871	4.97%	91.10%	7.48%
	Good	2766	8.23%	82.43%	13.55%
	Fair	713	25.59%	45.86%	30.86%
	Poor	1155	32.43%	38.26%	34.11%
	Good + Excellent	3637	12.74%	89.03%	8.91%
	Fair + Good + Excellent	4350	19.68%	81.95%	14.28%
	Complete	5505	28.46%	72.78%	19.51%

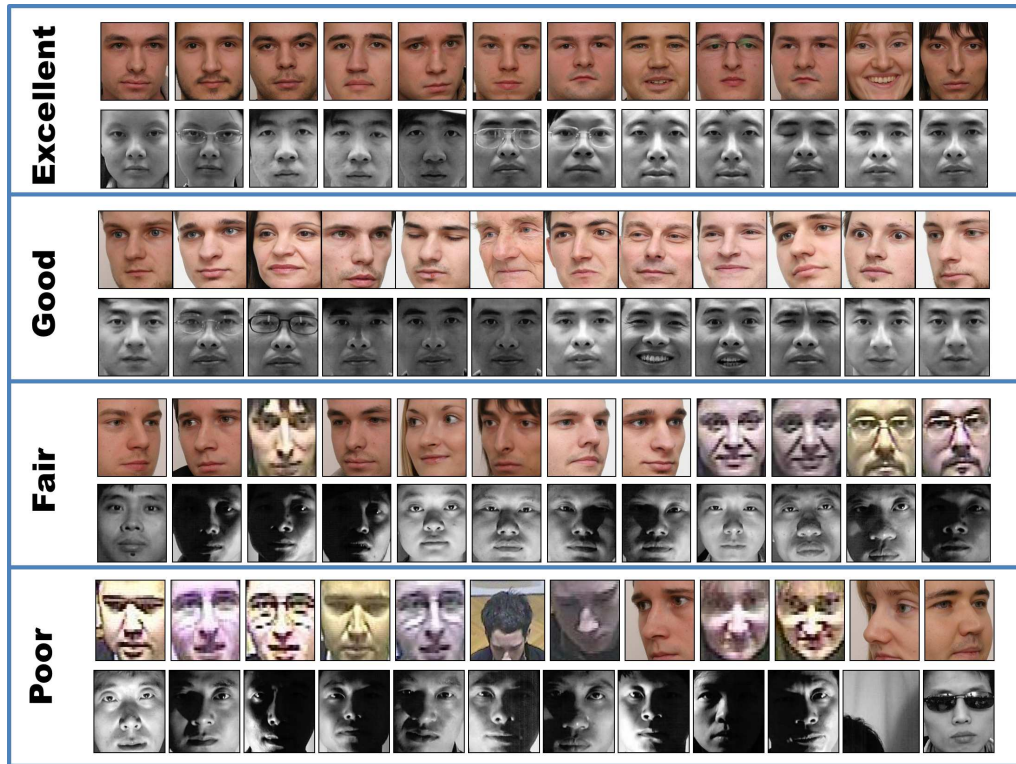


Figure 2.23: Sample images of four quality bins obtained from the proposed approach (common to both Gist and HOG).

images. The difference in performance of each bin indicates the validity of the assigned bin labels.

Table 2.13: Different noise and blur artifacts used in the experiments with their associated parameters.

Corruption	Parameters
White Noise	$\sigma = 0.05$
Local Variant White Noise	Dependent on local intensity
Poisson Noise	$\lambda=1$
Salt and Pepper Noise	$d = 0.05$ or 5%
Speckle	$v = 0.05$
Gaussian Blur	$\sigma = 1$
Motion Blur	$L= 1, r = 5$
Sharp	$\alpha = 0.1$

- In several applications of biometrics such as Aadhaar and US-Visit, low quality image samples are rejected to maintain the integrity of the database and to ensure high recognition accuracy. The proposed algorithm can be utilized to reject low quality samples. Figure 2.22 and Table 2.12 show improved performance compared to the complete database, when images classified as *poor* and/or *fair* are removed, indicating a direct relationship of the proposed metric with system performance.
- Figure 2.23 illustrates samples from the database classified into a particular quality bin. The illustrated instances are obtained from the set of images classified to a quality bin by both Gist and HOG. It can be observed that the classification correlates well with visual inspection.

2.5.2.1 Experiments with Noisy Images

In an attempt to circumvent the cumbersome feature extraction process, low dimensional representations of the face images (GIST) is used for noise assessment. The experiments are conducted on a subset of the AR face database [117] containing 400 frontal face images pertaining to 35 subjects.

A symmetrically corrupted database is prepared that consists of eight classes of artifacts and one class representing uncorrupt images. The parameters used are presented in Table 2.14. The experiment is conducted as described below.

- From the database, 50 images per artifact (corruption) class are chosen randomly for training the multi class SVM.
- GIST descriptors are computed using a bank of Gabor filters at eight orientations per scale. Image size is reduced to 128×128 and a block size of four is used for the windowed Fourier transform

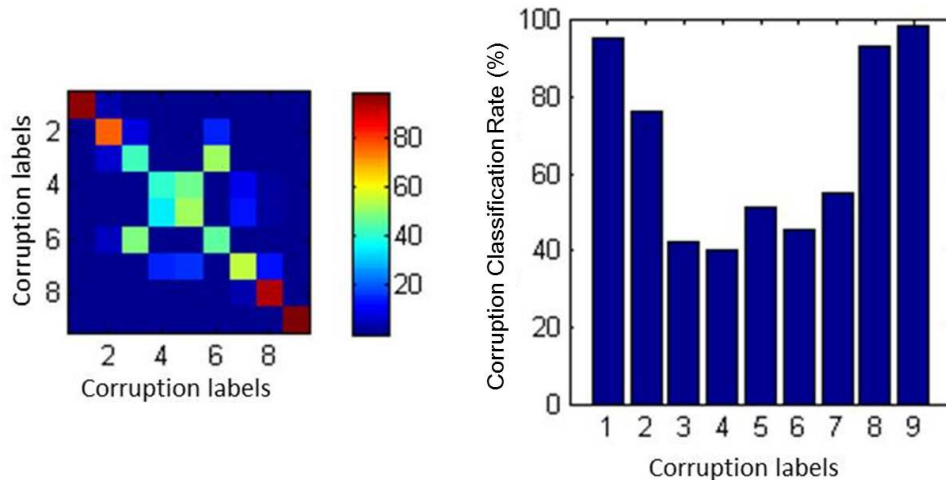


Figure 2.24: a) Confusion matrix of the performance of GIST descriptor b) Accuracy of SVM in classifying each of the eight noise/blur artifacts and 1 uncorrupt class labels. Labels are defined as 1- Gaussian blur, 2- Motion blur, 3- Uncorrupt (original), 4- White noise, 5- Localvar noise, 6- Poisson noise, 7-Salt & Pepper noise, 8- Speckle noise, and 9- Sharp.

- Since the Gabor bank is computed only once for a given image size and parameters, the descriptor is computed quickly. The average time is 0.18 seconds per image using Matlab on a standard desktop PC.
- A one-versus-all SVM is used for classifying images into the nine quality bins based on the GIST features. Figure 2.24 shows the confusion matrix and recognition accuracy per class of the multi class SVM classifier.
- These results indicate that the performance of the proposed method is suitable for identifying static and motion blur artifacts as well as distributed speckle noise. Further investigation on how the GIST descriptor is effected by noise can provide interesting insights towards assessment of quality.

2.5.2.2 Experiments with Pose

Pose estimation is a challenging problem in face recognition and several solutions have been proposed based on facial symmetry, orientation of nose region, shape of face, and 3D reconstruction. In this research, we present a simple learning based approach to pose estimation using the HOG descriptor. The experiment is conducted on the MultiPIE dataset [72]. We use a sub-sample of nine viewpoints with all 10 illumination conditions

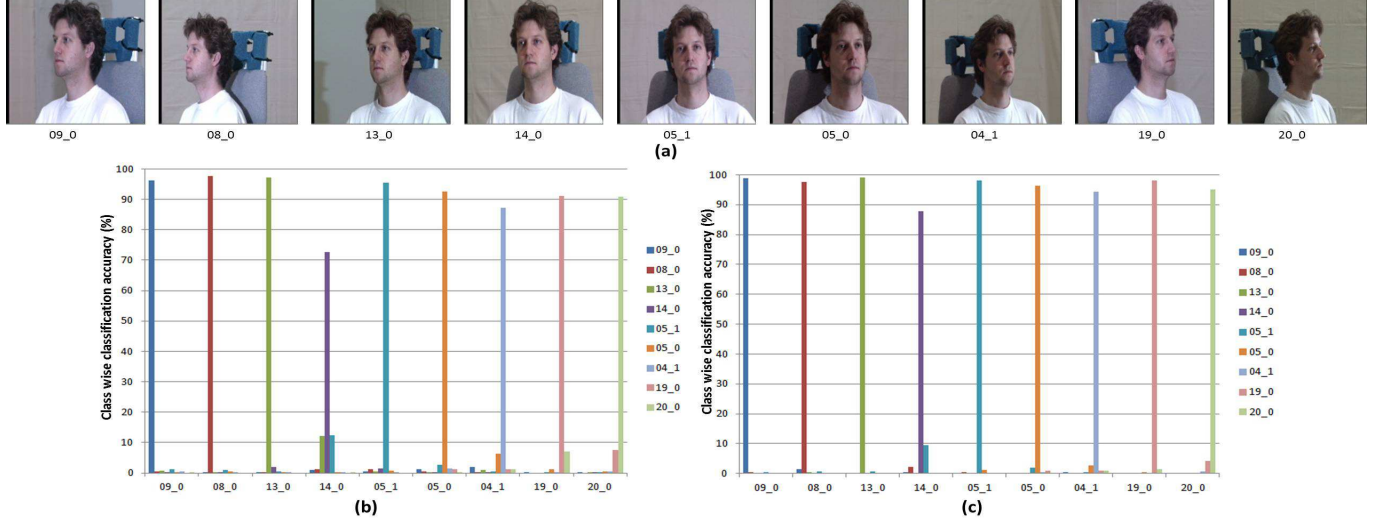


Figure 2.25: a) Frontal illumination images of a subject in all nine viewpoints from the MultiPIE dataset, b) a bar graph of the HOG and SVM pose estimator, and c) similar graph with PHOG descriptor.

and 4 sessions. The frontal illumination images of a subject in all nine viewpoints are shown in Figure 2.25.

- From the MultiPIE dataset, 30% of the total subjects from session 1 are chosen for training. Further, only two randomly selected images per user, to avoid over-fitting. Hence, 2772 images are used in the training phase. The remaining 313560 images are used for testing.
- HOG descriptor is computed for all images, with 9 histogram bins and 3×3 block size.
- A one-versus-all SVM is used for classifying images into one of the 9 classes based on the HOG features. The results are presented in Figure 2.25.
- The results show excellent classification performance when pose estimation is viewed as a supervised learning problem using simple descriptors. Further, the nature of the result prompt towards broader class labels for higher accuracy. A modified approach known as PHOG (HOG descriptor over three levels of Gaussian pyramid (down sampling) and concatenates the features) is also used. This approach used with SVM and same training and testing samples, yields improved results as shown in Figure 2.25.

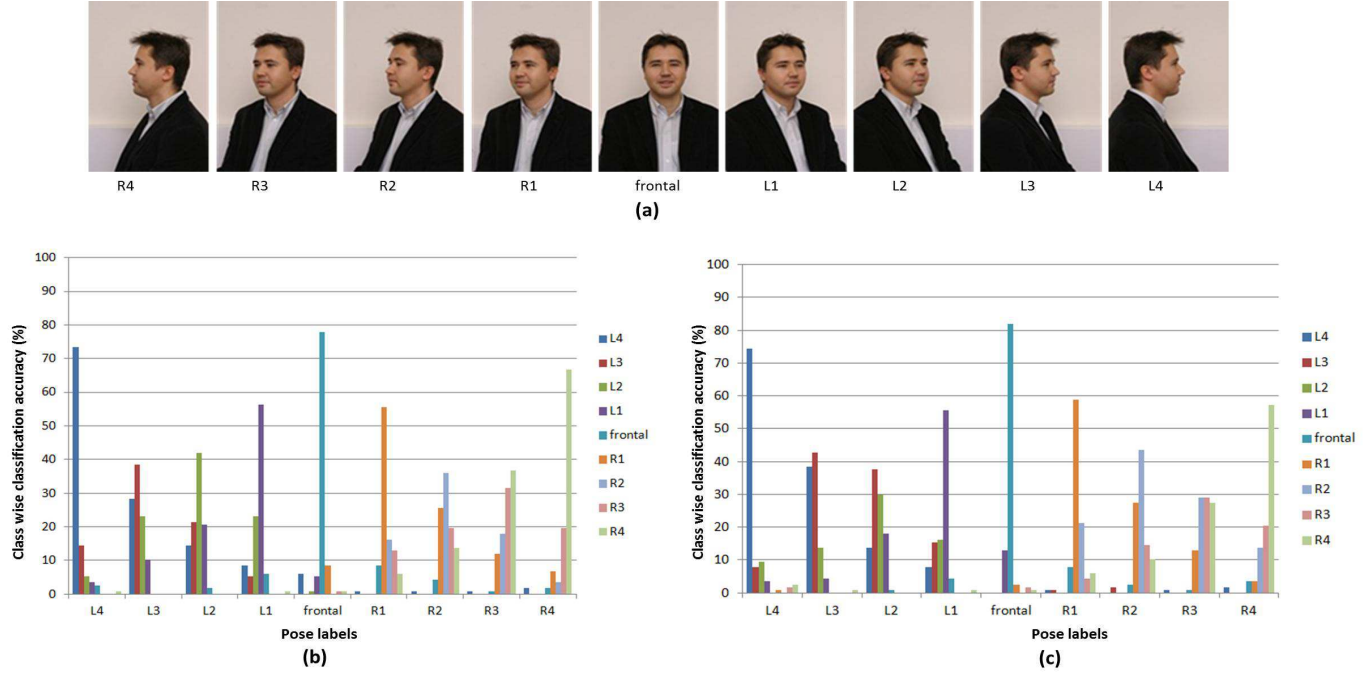


Figure 2.26: a) All nine poses from the SC-Face dataset for a single subject, b) a bar graph of the HOG and SVM pose estimator on the SC-Face dataset, and c) A similar bar graph with PHOG and SVM

Further, the pose estimation approach is evaluated with the smaller SC-Face dataset [71] in order to assess generality. This dataset also consists of high resolution face images in different pose angles.

- This database, consists of 130 subjects and 9 poses. All images of 13 subjects are used for training and the remaining 117 subjects for testing. All 9 poses from the dataset for a single subject are shown in Figure 2.26(a).
- Similar to the previous experiment, a one-versus-all SVM is used for classifying each image into one of the nine classes based on the HOG features. The results are shown in Figure 2.26(b). Further improvement is achieved with PHOG descriptors, as shown in Figure 2.26(c).

2.6 Quality Assessment based Denoising to Improve Recognition Performance

Quality of a biometric sample affects the performance of the recognition algorithm. In literature, several research papers exist on analyzing the effects of quality on the performance



Figure 2.27: Irregularities due to different types of noise degrade the quality of face images significantly. Sample images from the AR face dataset with synthetic noise.

of different biometric modalities such as iris and fingerprint [40, 60, 91]. Environmental corruption such as noise, blur, adverse illumination and compression rates (in JPEG and other compression techniques) influence the performance of state-of-art recognition algorithms. Several enhancement methods have been proposed in literature to handle these corruptions [35]. However, the parameters chosen for the enhancement algorithms have an adverse effect on the performance of automatic recognition algorithms.

Recent research in face biometrics, in an effort to address covariates such as pose, illumination and expression, have turned towards texture recognition algorithms. Texture algorithms such as Uniform Circular Local Binary Patterns (UCLBP) [136], are known to be more resilient towards these covariates compared to appearance based algorithms such as Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA). However, the experiments show that the texture algorithms are also susceptible to environmental noise. As shown in Figure 2.27, noise may be induced due to sensor error, transmission error or due to wrong capturing practices, affect face recognition performance. The experiment performed using data-driven noise and LBP demonstrates the effect of noise on face recognition. Cumulative Match Characteristic (CMC) curves in Figure 2.28 show a significant loss of performance in the identification accuracy due to synthetic addition of

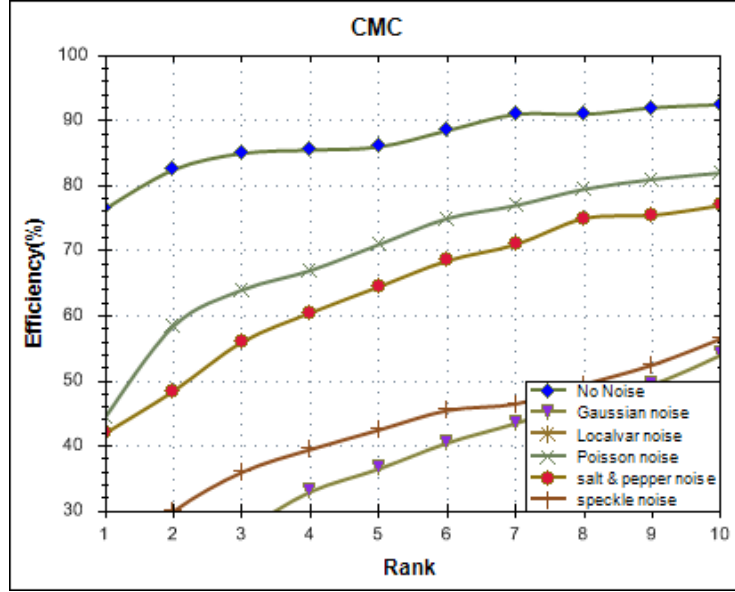


Figure 2.28: CMC curve of local binary pattern (LBP) + χ^2 matcher: Identification performance decreases when noise is added to the probe images

noise. Considerable advancements have been made in literature to denoise images, beginning from wavelet based hard thresholding to more elegant soft thresholding techniques. However, the performance of these approaches depend on parameters such as choice of mother wavelet and number of iterations, which are directly dependent on the amount of noise present in the image. It is our hypothesis that the full utility of denoising (or enhancement algorithms) can be realized with a framework that selects the best parameters for each of the given (probe) image.

This work presents a framework to select the image denoising parameters based on the quality assessment of individual images. The proposed framework utilizes Support Vector Machine (SVM) classification to learn the relationship between image quality assessment scores and the optimal parameters for denoising. Several image quality assessment techniques exist in literature that have shown high correlation with the assessment of human subjects [118, 198]. Also, biometric modality specific quality assessment techniques for fingerprint and iris exist in literature. This research focuses on computationally simple quality metrics that possess intuitive relevance and high correlation with face recognition accuracy are considered in this research, namely, *No-reference quality assessment* (Q1) [198] and *Edge spread measure* (Q2) [118]. In the proposed framework, both quality scores are provided to the SVM classifier as a 2D quality vector.

2.6.1 Image Denoising with BayesShrink

Intuitively, denoising a noisy face image improves the face recognition performance, provided the right set of parameters are used. BayesShrink [35], a wavelet based soft thresholding technique is used for denoising in the proposed quality assessment based denoising framework. BayesShrink [35] is an adaptive, data-driven wavelet thresholding approach for image denoising. The wavelet thresholds are derived from the Bayesian approach assuming that the data follows a generalized Gaussian distribution (GGD). This assumption is based on empirical findings that any natural image can be summarized by a GGD. From this assumption the mean square error (MSE) for each wavelet sub-band is modeled as a Bayesian squared error with known priors for each distribution applied independently and identically. Here the idea is to find soft-thresholds that minimize the Bayesian risk. Formally, given an uncorrupt image $f_{i,j}$ of size $M \times N$, the noisy image $g_{i,j}$ can be written as

$$g_{i,j} = f_{i,j} + \epsilon_{i,j} \quad (2.6)$$

where $i = 1, \dots, M, j = 1, \dots, N$, $\epsilon_{i,j}$ is independent and identically distributed (*iid*) noise assumed as normal $N(0, \sigma^2)$ and independent of image signal $f_{i,j}$. The purpose is to find an estimate $\hat{f}_{i,j}$ of the image $f_{i,j}$ that minimizes

$$MSE(\hat{\mathbf{f}}) = \frac{1}{N^2} \sum_{i,j=1}^N (\hat{f}_{i,j} - f_{i,j})^2 \quad (2.7)$$

Further, eq 2.6 in matrix form is given by $Y = X + V$, where X and V are independent of each other, hence

$$\sigma_Y^2 = \sigma_X^2 + \sigma^2 \quad (2.8)$$

Here σ^2 is the actual variance of noise distribution. From the detail sub-bands of wavelet transform, a threshold T is estimated as

$$\hat{\mathbf{T}}(\hat{\sigma}_X) = \hat{\sigma}^2 / \hat{\sigma}_X \quad (2.9)$$

where, $\hat{\sigma}^2$ is estimated variance of noise obtained from the wavelet transform of Y . Note that soft thresholding keeps the overall Bayesian risk small as compared to hard-thresholding techniques¹. The denoising algorithm is governed by two parameters: first is choice of mother wavelet and second is number of iterations required to denoise the image.

¹For further details of BayesShrink, refer to Chang *et al.*[35].

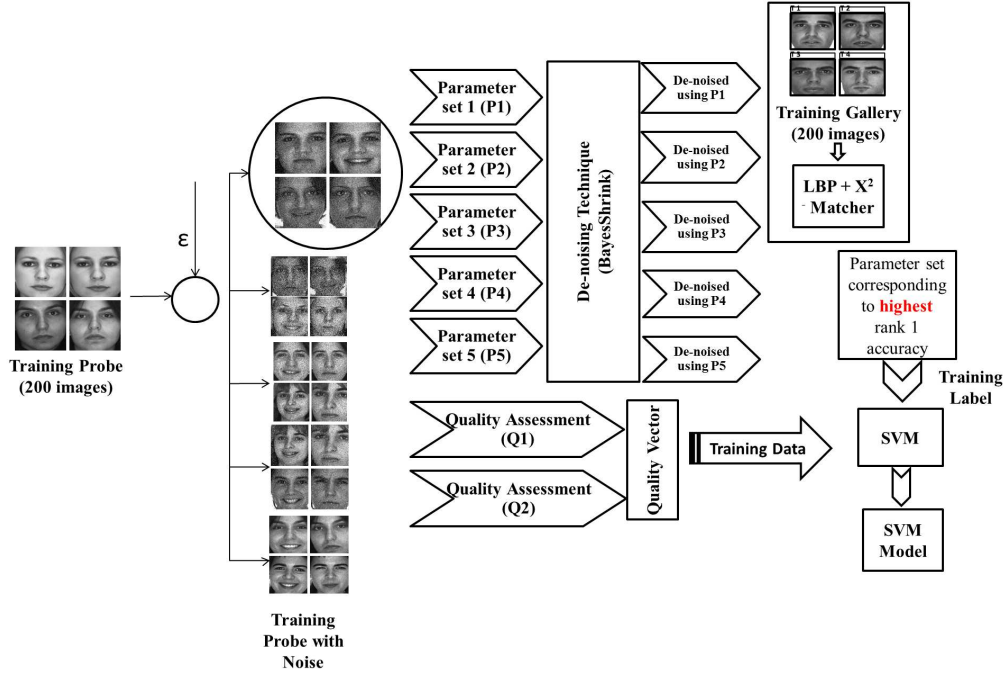


Figure 2.29: The training scheme of the proposed assessment based denoising framework. The process presented here is for a single noised training probe set; this process is repeated for all the five data driven noises.

2.6.1.1 Candidate Parameter Set

Several combination of parameters are possible for denoising using BayesShrink algorithm. For a given probe image, it is computationally expensive to sweep the entire parameter space, however it is more feasible to first assess the level of quality and select the most suitable parameter. In this research, it is observed that the performance of BayesShrink, in terms of improving face recognition accuracy, is related to the type of wavelet used and in some cases, the number of iterations. Hence the following subset of parameters are considered as candidate parameter set: Haar Wavelet (P1), Symmlet Wavelet (P2), Duabechies Wavelet (P3), Beylkin Wavelet (P4) (all single iteration) and Symmlet Wavelet with two iterations (P5) that is termed *Symmlet2*. As we will analyze the performance in Section 2.6.3.1, while P5 gives the best performance, P1 is computationally least expensive. Hence, correct use of these parameters can improve accuracy as well as computational time.

2.6.2 Proposed Quality Assessment based Denoising Framework

The proposed quality assessment based optimal parameter selection framework uses SVM classification. The training and testing phases of the framework are discussed.

2.6.2.1 Training

Training phase of the proposed framework is shown in Figure 2.29. The training labels for the parameter selection are generated using the training data which is partitioned into gallery and probe set.

- The images in the training and probe sets are each corrupted systematically by data driven noises, namely, Gaussian(white) noise, local variant white noise, Poisson noise, salt & pepper noise and speckle noise.
- Each of these corrupt training-probe-set are denoised with the wavelet based BayesShrink denoising algorithm[35] with each of the i candidate parameters $P_{1..i}$.
- The quality vector for each image with quality scores $[Q_1, Q_2]$ is computed and used as the training sample for a multi-class SVM classifier. The best parameters of the SVM classifier are converged upon by minimizing the training error via a 10-fold cross validation.
- The class label corresponding to each of the quality vector is the parameter $P_{1..i}$ which results in the best rank-1 efficiency with the training-gallery-set using local binary pattern (LBP) as the face recognition algorithm.
- While Figure 2.29 illustrates the process for a single noisy training probe set, the process is repeated for all the five data driven noises.

Figure 2.32 shows a scatter plot of the quality scores of the training data. The illustrated class labels correspond to the best parameters selected for denoising. The classes are well separated, confirming the initial hypothesis that images with a certain set of quality scores require a specific parameter for the best denoising.

2.6.2.2 Testing

The trained SVM model is used to select the parameters for denoising, as shown in Figure 2.30. Given an input (probe) image, the quality vector $[Q_1, Q_2]$ is calculated. Using this quality vector as input, the parameter class obtained from the trained SVM is used to denoise the input image.

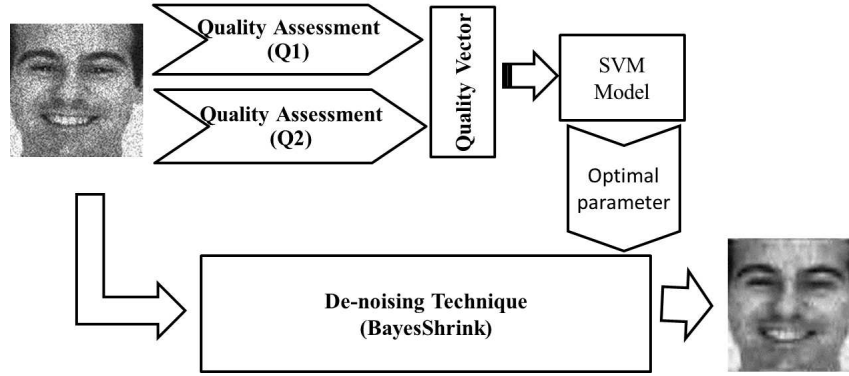


Figure 2.30: The testing scheme of the proposed assessment based denoising framework

Table 2.14: Different noise artifacts used in the experiments with their associated parameters.

Corruption	Parameters
Gaussian noise	$\sigma = 0.01$
Localvar noise	Dependent on local intensity
Poisson noise	$\lambda=1$
Salt and Pepper noise	$d = 0.05$ or 5%
Speckle noise	$v = 0.05$

2.6.3 Experimental Results

The experiments are conducted on the AR face database [117] containing 756 frontal face images pertaining to 126 subjects (i.e. six images per subject). From this dataset, images corresponding to 50 subjects are chosen for training and the remaining are in the testing set. For experimental purposes, different noise artifacts are synthesized for each image. The kernel parameters used to introduce these artifacts in the images are indicated in Table 2.14. Before evaluating the proposed framework, a correlation study is performed to establish that the combination of the two quality assessment scores utilized are indeed indicative of the performance of the face recognition algorithm.

2.6.3.1 Recognition Experiment

As discussed in Section 2.6.2, SVM model is learned using the training labels from the data driven approach on the training set of 50 individuals from the AR face dataset [117]. Figure 2.31 illustrates denoised output of the proposed algorithm. The scatter plot of the training set is shown in Figure 2.32. The best performance label corresponded to three of the five parameters. To evaluate the performance of the framework, random noise is added to probe images in the testing data set. As shown in Figure 2.33 (CMC curves)



Figure 2.31: Sample noisy (top) and denoised (bottom) images obtained using the proposed framework.

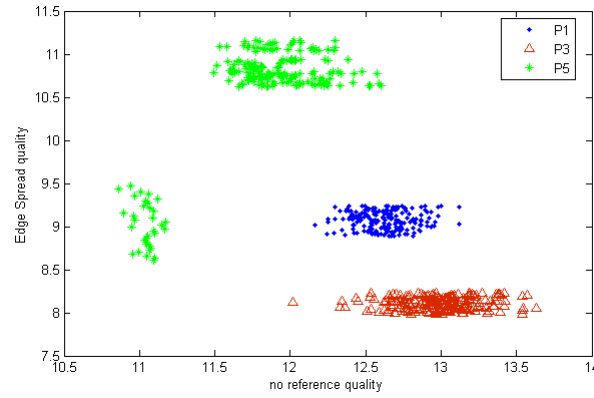


Figure 2.32: Scatter plot of the quality score of training data. The illustrated class labels correspond to the best parameters selected for denoising. This indicates that images with a certain set of quality scores require a specific parameter for best denoising.

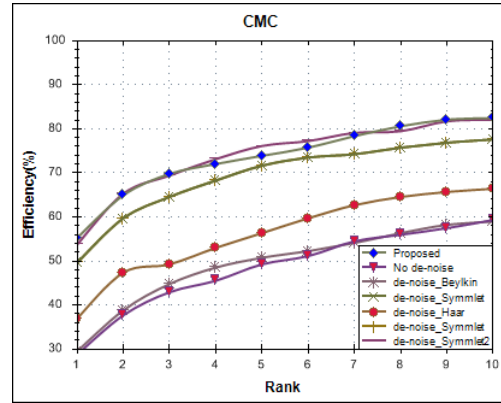


Figure 2.33: CMC of LBP based face recognition using proposed parameter selection framework and each parameters without selection. The selection framework slightly improves performance and reduces computation time.

and Table 2.15, the performance of the face recognition algorithm on the images denoised using the proposed approach is comparable to/better than the other denoising approaches. These results show that the proposed framework not only improves the accuracy but also

Table 2.15: Average computation time for denoising an image with each image enhancement parameter and Rank-1 efficiency with testing-gallery-set. All values computed in Matlab with a Dual-core CPU and 2 GB RAM

Parameters	Computation Time	Rank-1 Identification
Haar Wavelet	0.01sec	36.94%
Daubechies	2.77sec	50%
Symmlet	3.87sec	49.62%
Symmlet2	4.41sec	53.73%
Proposed	3.68sec	55.22%

reduces the computational time.

This research presents a quality assessment based denoising framework to improve the results of denoising by selecting optimal parameters. The results discussed in this work suggest that no-reference quality and edge spread quality assessment techniques correlate with the recognition performance of automated face recognition systems. Further, noisy images in different parts of the quality space require completely different parameters to result in the best possible denoising process. The proposed framework can easily be extended to a larger set of assessment and enhancement techniques. One limitation of the framework is the large computation time for training, however, we found that the generated prediction model is quite versatile. Large computational hardware will allow for a more comprehensive sweep of the parameter space.

2.7 Discussion

Traditional image quality metrics measures certain aspects of an image important for good visual perception. On the other hand, biometric quality assessment measures the potential of the sample for recognition. As shown in literature, such quality metrics not only help in improving data collection but also provide additional information at different stages of a biometric system. Based on the literature review and experimental analysis, here, we collate the important observations pertaining to biometric quality assessment:

- The prominent features used in quality assessment are orientation of edge features. While a strong case can be made for the performance of these features, research has shown potency of color-based and intensity-based features as well.
- There is a need for better evaluation framework for biometric quality assessment metrics. High correlation with match score performance along with statistical tests can help towards better evaluation. The good, bad, and ugly distribution of database

[146] is an interesting method for evaluating the performance of quality metrics for performance prediction.

- Researchers must emphasize on the computational cost in the development of quality assessment approaches, which must be lesser or comparable to the matching time.
- Quality metrics used for quality-based multibiometric fusion approaches must be carefully selected. As discussed in Section 3.4.3, not all quality metrics are useful for match score prediction. Quality metrics that measure different kinds of degradations, including modality-specific metrics, must be considered.
- In differential processing techniques such as context switching, quality metrics can be important cues for selection of recognition modules. Based on the modality in consideration, additional factors such as age and gender may also be considered as cues [84].
- It is our assertion that a better understanding of the behavior of biometric quality, in terms of *Naturalness*, *Fidelity* and *Utility*, can help in the development of more meaningful quality measures. Such quality metrics may also enhance the performance of quality-based multibiometric frameworks proposed in literature.
- Face quality is affected by pose, illumination, and expression apart from image degradations such as noise and blur. Other covariates such as aging, disguise, and occlusion degrade the performance relative to a reference sample.
- The quality of a match pair is a function of the quality of both gallery and probe images [22]. Further, high-resolution frontal face images do not directly imply high-quality biometric sample or confident match.
- Important findings from the results of the FRVT 2006 [148] and MBE 2010 [73] can help towards development of better quality assessment techniques. For instance, a slight gender bias is observed in the performance of the algorithms, with samples of female subjects performing better than male subjects in controlled environment. Also, the evaluations found that samples obtained from individuals of a certain race perform better than others, with East-Asian races performing the best.
- A strong correlation has been observed between simple image quality measures and performance of the top three algorithms of the vendor test [20]. Precisely, a high

correlation has been observed between the recognition rates and a simple gradient energy-based focus measure.

- The performance of samples captured in indoor studio-like conditions is better than the performance of samples taken in uncontrolled outdoor conditions. While this result is expected, it is interesting to note that this penalty in performance decreases with relaxed false acceptance rates.
- The quality of a fingerprint sample is largely governed by the sensor in deployment. It is observed that the common factors include scars, burns, dryness, and temperature. *Auto capture* is a common feature in modern fingerprint sensors, requiring real-time quality assessment of the presented sample. Therefore, most quality metrics evaluate ridge clarity and number of detected minutia.
- The performance of iris as a biometric is hugely dependent on the quality of captured sample. The micro-features of iris texture are easily contaminated by adverse illumination, lenses, glasses, or disease. The most prevailing approach for iris quality measurement continues to be the fusion of assessment of several known quality factors.
- Due to the requirement of low computational time, *auto capture* in iris sensors is usually based on confidence of segmentation. A major drawback of existing approaches is in the assumption of good quality segmentation before quality assessment. However, same factors that affect biometric quality are also known to effect iris segmentation.
- Current research uses typical image processing algorithms that evaluate image degradations due to noise, compression, or illumination. However, a quality metric that entails a greater insight of the *usefulness* of the biometric sample in consideration can improve the performance of these systems by providing more discernible quality *cohorts*.
- Using quality assessment metric cannot, however, be a panacea for the recognition of poor quality images. Beveridge *et al.* [145] place a bound on the extent to which quality metrics can improve the performance of matching systems when they are used as performance predictors.

Quality assessment of biometric samples is an important challenge for the biometrics research community. A clear distinction is made between the image quality and biometric

quality of a biometric sample to capture modality-specific intuitions of quality assessment. It is our assertion that quality metrics are an important ingredient in improving the robustness of large real-world biometric systems. In an attempt to demystify the definition and work of biometric quality, several factors that affect a biometric sample are presented. Different image features utilized in literature for quality assessment, evaluation processes, and match score predictability are discussed. Further, a literature survey of the quality assessment techniques in three biometric modalities reveals that techniques often focus on naturalness alone. It is imperative that quality assessment entails a notion of fidelity of capture and modality-specific utility as well. Further, the performance of a biometric quality assessment metric in terms of computational complexity must also be discussed more actively in research. The development of quality assessment algorithms of biometric samples that are computationally inexpensive to compute yet correctly encode quality will be the *sine qua non* of real-world large-scale deployments.

Quality metrics are an important ingredient to improve the robustness of large scale real-world face biometric systems. This research also investigates the possibility of using holistic representation of an image for quality assessment. The results with Gist and HOG show promise towards a robust solution to the important problem of quality assessment in face biometrics. By further evaluating the effects of each quality class on recognition accuracy, the techniques described in this research can also be used for classifier performance prediction. This research also presents a quality assessment based denoising framework to improve the results of denoising by selecting optimal parameters. The results also suggest that noisy images in different parts of the quality space require completely different parameters to result in the best possible denoising process. Quality metrics can be used to infer the optimal parameters for a quality enhancement technique in a data driven way.

Chapter 3

QFuse: Online Learning Framework for an Adaptive Biometric System

Existing biometrics techniques are unable to provide significant levels of accuracy in uncontrolled noisy environments in several applications such as assisting law enforcement agencies to control crime and fraud. Further, scalability is another challenge due to variations in data distribution with changing conditions. This chapter presents a novel adaptive context switching algorithm coupled with online learning to address both these challenges. The proposed framework, termed as QFuse, uses the quality of input images to dynamically select the best biometric matcher or fusion algorithm to verify the identity of an individual. The proposed algorithm continuously updates the selection process using online learning to address the scalability and accommodate the variations in data distribution. The results on the WVU multimodal database and a large real world multimodal database obtained from a law enforcement agency show the efficacy of the proposed framework.

3.1 Introduction

A biometric system classifies an individual as genuine or impostor based on modalities such as face, fingerprint and iris. A traditional unimodal biometric system extracts features for the given biometric modality and compares it with stored database templates and computes a match score [159]. For verification settings (1 : 1 matching), the match score is classified as genuine or impostor. Unification of different biometric samples or evidence (such as face, fingerprints, and iris) to verify the identity of an individual is referred to as multicentric. Such multimodal systems offer additional benefits over unimodal systems such as resiliency to noise and malfunction, universality, and improved accuracy.

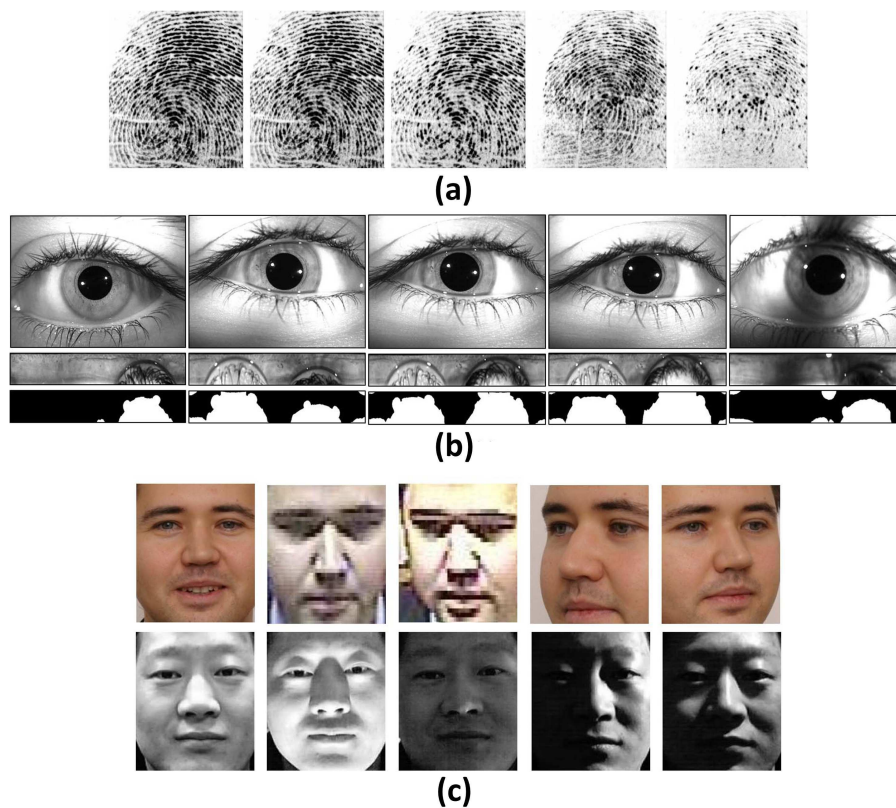


Figure 3.1: Illustrating examples of multimodal images from the same subject with varying quality (a) fingerprint, (b) iris (the last two rows demonstrate the unwrapped iris images and occlusion mask that indicate iris feature occlusion), and (c) face images.

However, the performance of a biometric system degrades when probe (query) images are of lower quality compared with the images that the system has encountered during training. As shown in Figure 3.1, the quality of probe images may degrade because of several reasons, such as improper illumination, improper interaction with the sensor (e.g. pose variations), and different kinds of noise or blur introduced in the probe image when the image is captured in an uncontrolled environment (applicable in many real world applications). Fingerprints can suffer from dryness, iris images can have occlusion or improper illumination, while face images can be of low resolution or have pose variations.

Biometric systems generally do not facilitate case-based switching for selecting an appropriate classifier or combination of classifiers. Moreover, any biometric system or case-based switching criteria that is learned on limited (in terms of availability and variety) training data performs adequately only if the test data distribution is similar to the training

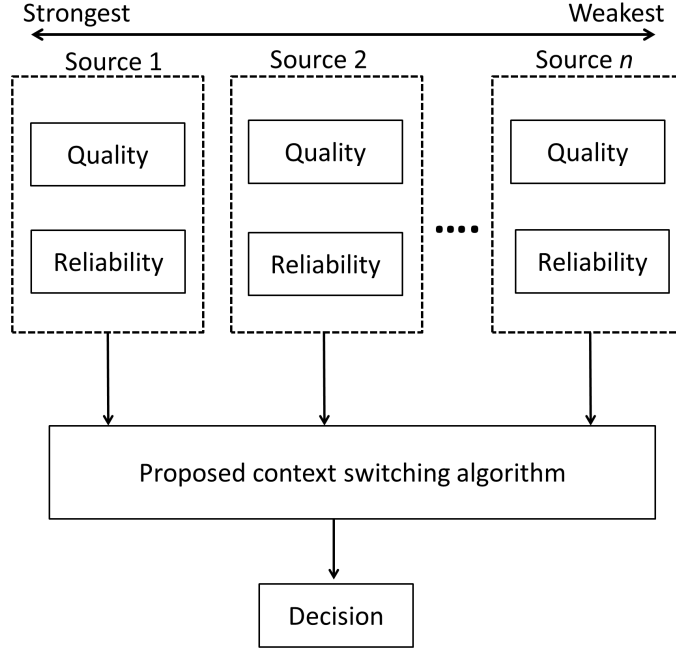


Figure 3.2: General concept of the proposed QFuse algorithm that selects a single information source or fusion of multiple sources based on its reliability and the quality of information.

data distribution. Since new users continue to be enrolled into the system, a biometric system needs to be re-trained with new enrolments to accommodate the variations caused due to incremental data and maintain the accuracy levels. However, in many real world applications, re-training biometric classifiers may not be pragmatic as it requires all the training data in batch mode.

It is well understood that in different operating scenarios, information from some sources may be more useful than others. Hence, a mechanism is required to efficiently combine evidences based on situational cues. To incorporate this facility, the thesis proposes *QFuse*, an online learning algorithm for adaptive biometric fusion that incorporates image quality in the dynamic selection of unimodal classifiers and their fusion. Figure 3.2 shows the generalized concept of the proposed algorithm that selects information source(s) based on the reliability of each source and certain discriminatory cues (quality) of the information. To accommodate the variations in data distributions and sustain the performance with increasing number of users, the thesis also proposes to update the classifiers used in the proposed algorithm, QFuse, in an online manner. Specifically, to address the first challenge, the proposed algorithm uses different unimodal classifiers arranged serially in decreasing order of their reliability (accuracy) to process gallery-probe pairs one modality at a time. The serial arrangement of classifiers is based on our assertion that a unimodal

classifier can efficiently match a good quality gallery-probe pair. However, unimodal classifiers yield conflicting results when the quality of gallery-probe pair degrades. In such cases, complementary information from multiple unimodal classifiers can be efficiently combined to yield correct results. Secondly, the research proposes to update QFuse in an online manner with only the new enrolment data. It provides a huge benefit in terms of computational time as well as improved accuracy (later validated in experimental results). The major contributions of this research are summarized below:

1. A serial context switching algorithm is proposed to select the most appropriate constituent unimodal classifier or fusion algorithm for a given set of gallery-probe pair based on its quality.
2. An online learning algorithm is proposed to update the context switching rule with the new incremental enrolment data to address the variations in data distribution.

The experiments are performed on two databases: WVU multimodal database [45] and a large scale multimodal database obtained from law enforcement agencies. The experimental analysis shows that the proposed framework not only improves the verification accuracy but also significantly reduces the computation time required for updating the framework.

3.2 Literature Review

Unification of multiple biometric information can be performed via two approaches: (1) matcher fusion and (2) dynamic matcher selection [205]. In matcher fusion, all the constituent matchers are used and their evidences are combined using fusion rules [96], [158], [172], [125]. On the other hand, dynamic matcher approaches include selecting the most appropriate matcher or a subset of specific matchers [69], [187], [152]. In the biometric literature, matcher fusion approaches have received significant attention [159], [158]; however, dynamic matcher selection has not been extensively explored. Traditionally, in multi-modal fusion approaches, designing a fusion scheme and performance evaluation are considered as two different stages. However, Toh *et al.* [181] proposed to simultaneously optimize the target performance and design of classifier fusion algorithm based on an approximation of the total error rate. Marcialis *et al.* [115] proposed serial fusion of face and fingerprint matchers where significant reduction in verification time was achieved. However, their approach did not consider the quality of the gallery-probe pair and was based only on the match score distribution of genuine and impostor scores. Traditional

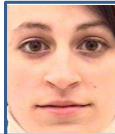
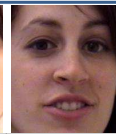

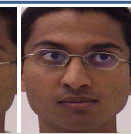
multi-biometric systems work on static fusion rules which may not adapt itself to the dynamically changing environment and thus degrade the performance as the environment changes. Geng *et al.* [68] proposed a context aware fusion scheme that takes into account the viewing angle and distance of the subject from the camera to select an appropriate fusion scheme for improved performance. Abaza and Ross [1] proposed including image quality in the fusion scheme to enhance the performance in presence of weak matchers or low quality input images. Alonso-Fernandez *et al.* [9] proposed a method for efficient combination of match scores from different devices (sensors) depending on the quality of data source. Different modalities may yield heterogeneous scores; therefore, score normalization is required to transform these scores into a common domain. Jain *et al.* [85] analyzed different normalization and fusion techniques in the context of a multimodal biometric system. Veeramachaneni *et al.* [191] proposed using particle swarm optimization to switch between different fusion rules for combining decisions received from multiple biometric sensors. Kumar *et al.* [103] proposed a hybrid particle swarm optimization based approach for adaptive combination of multiple biometric modalities. Raghavendra *et al.* [157] proposed an efficient fusion scheme to combine complementary information from different biometric modalities at match score level. They also proposed a particle swarm optimization (PSO) procedure for reducing the dimensionality of feature space by identifying a subspace of the large dimension features. For evaluating, comparing, and bench-marking quality-dependent, client-specific, cost-sensitive score-level fusion algorithms, Poh *et al.* [150] prepared a score and quality database. They also reported the baseline experimental results for evaluating the above three types of fusion scenarios. Faundez-Zanuy [57] analyzed different types of data fusion and stages in a biometric system where fusion can be applied. Poh *et al.* [153] proposed a user-specific and selective fusion strategy to combine multiple biometric modalities. Their algorithm assigned a different set of fusion parameters to a given enrolled user and selected a subset of modalities for fusion. Their approach achieved better performance at a reduced computational cost based on a criterion called B-ratio that ranked subjects based on their match score statistics. Recently, Huang *et al.* [82] proposed general multimodal recognition framework which is termed as adaptive bimodal sparse representation-based classification. Utilizing a two-phase sparse coding strategy for precise quality assessment of face and ear images, the framework combined multimodal features for improved performance. Kanhangad *et al.* [92] propose a dynamic match score combination approach for palm print and 3D hand geometry techniques.

Vatsa *et al.* [185] proposed a parallel algorithm to select an appropriate constituent unimodal matcher or the fusion algorithm. Their algorithm supported biometric image quality based and case-based switching for improved recognition performance. However, their approach performed switching in a parallel manner and required all the biometric modalities to be processed upfront; therefore, it was computationally more expensive. Recently, Nair *et al.* [122] proposed multinomial and geometric models for multibiometric systems in which the framework predicts the matching subjects in a multi-view/multimodal biometric environment via a case based switching approach. In a preliminary version of this manuscript, Bhatt *et al.* [29] proposed a serial context switching algorithm to address the limitations of a parallel framework and achieve better performance.





Updating a classifier’s knowledge using online learning has been actively studied in the machine learning community [34]; however its applicability has recently been realized in the biometrics community. In the biometrics literature, incremental learning approaches with principal component analysis [156] and linear discriminant analysis [183] have been shown to be robust against the variations introduced due to incremental data. Singh *et al.* [171] introduced an online learning approach for updating a face matcher. Later, Kim *et al.* [95] proposed an online learning algorithm for biometric score fusion. Recently, Bhatt *et al.* [30] proposed to use labeled as well as unlabeled information for updating biometric matchers using an online co-training approach. While most of the existing approaches in biometrics literature have focused on updating a unimodal matcher, this work, to the best of our knowledge, is the first to incorporate online learning in a quality based modality switching algorithm.

3.3 QFuse: Quality Based Context Switching with Online Learning

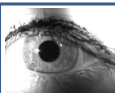
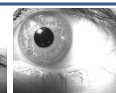
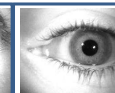
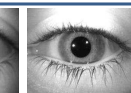
The proposed quality based context switching algorithm recognizes individuals captured in uncontrolled environment where the quality of probe images is low. In this research, “*context*” refers to a biometric modality, i.e. face, fingerprint, and iris or their multi-modal fusion. Therefore, context switching refers to switching from one modality to another depending on certain cues (the quality) obtained from a given gallery-probe image pair. The proposed algorithm efficiently matches individuals using one of the unimodal matchers when the gallery-probe pair is of good quality and dynamically switches to fusion of multiple matchers across different modalities when the gallery-probe pair is of poor quality. The algorithm is further trained in an online manner to adapt the variations introduced

				
Energy spectrum	0.01	0.70	0.08	0.06
Edge spread	0.49	0.31	0.77	0.77
Blockiness	0.41	0.24	0.22	0.58
Activity	0.36	0.19	0.26	0.46
ZC-rate	0.19	0.41	0.48	0.43
Pose	0.79	0.84	0.67	0.69

(a)

				
Energy spectrum	0.10	0.01	0.08	0.00
Edge spread	0.33	0.40	0.50	0.52
Blockiness	0.87	0.78	0.62	0.67
Activity	0.47	0.75	0.60	0.66
ZC-rate	0.37	0.67	0.90	0.63
Global Entropy	0.54	0.72	0.77	0.80

(b)

				
Energy spectrum	0.05	0.04	0.04	0.05
Edge spread	0.83	0.89	0.88	0.86
Blockiness	0.72	0.18	0.24	0.24
Activity	0.63	0.39	0.26	0.44
ZC-rate	0.36	0.30	0.42	0.04
De-focus	0.04	0.06	0.30	0.11
Motion	0.32	0.01	0.09	0.26
Occlusion	0.48	0.37	0.35	0.17
Specular Reflectance	0.17	0.12	0.05	0.12
Lighting	0.48	0.15	0.12	0.29
Pixel count	0.58	0.80	0.75	0.76

(c)

Figure 3.3: Sample images and their corresponding quality assessment scores for (a) face, (b) fingerprint, and (c) iris images. These quality scores are obtained after segmentation and utilized by support vector machines for context switching.

due to new enrolments. This section first presents the quality assessment algorithms used in the context switching algorithm followed by the proposed algorithm.

3.3.1 Quality Assessment

It has been shown in literature that quality assessment scores of a biometric sample can be indicative of its recognition performance [10, 27, 28]. The proposed algorithm computes a quality vector for a given image using computationally inexpensive quality assessment techniques. The quality vector comprises of both image quality metrics and modality specific quality metrics. Each of the quality metric is briefly discussed below. Further details regarding the quality metrics used in this research are described in 6.

- **Image quality metrics:** The first set of quality metrics is related to the quality of input gallery and probe images. We have used three algorithms for quality assessment:

No-reference quality: Quality degradation due to compression artifacts can be computed by estimating the blockiness and activity of an image, as proposed by Wang *et al.* [199]. To effectively utilize the quality metric, three separate estimations of degradation in the image, namely blockiness (B), activity (A) and zero-crossing rate (ZC) are computed and combined in both horizontal and vertical directions.

Edge spread: Motion and off-focus blur are measured using edges and adjacent regions [59]. It is estimated as the difference in image intensity with respect to the local maxima and minima of pixel intensity at every row of the image.

Spectral energy: Block-wise spectral energy is calculated using Fourier transform components [131] which represent sudden changes in illumination and specular reflection. The image is divided into non-overlapping blocks and the spectral energy is computed as the magnitude of Fourier transform components within each block.

- **Modality specific image quality:** The next set of quality parameters is related to biometric information. For each modality, a specific set of parameters (details in Appendix A) is calculated which represents the usability of the image (here we focus only on three modalities, viz. face, fingerprint, and iris).

Face quality: Pose variations in face degrade the performance as some of the facial features may not be visible. Such variations reduce the usability of the face image and a good quality image may not be useful for recognition. In this research, pose is estimated geometrically based on positions of eyes and mouth.

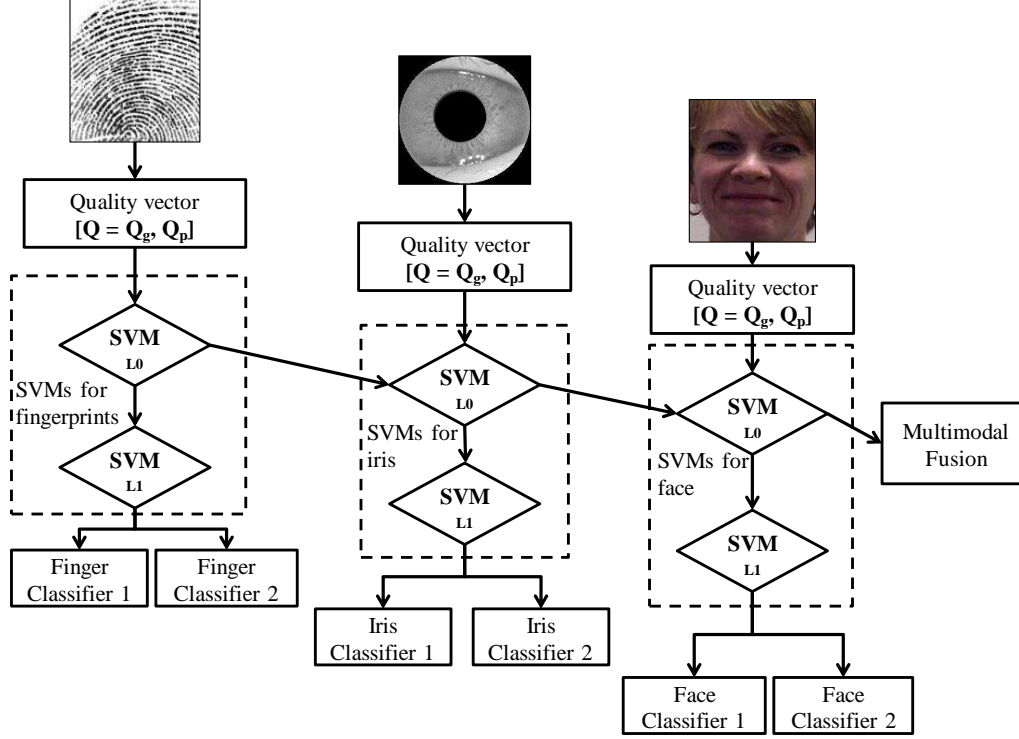


Figure 3.4: Illustrating the proposed quality based serial context switching algorithm.

Fingerprint quality: For fingerprint images, Chen *et al.* [40] proposed to measure the quality of ridge samples by Fourier energy spectral density concentration in particular frequency bands where strong ridges manifest.

Iris quality: Kalka *et al.* [91] presented quality assessment of iris images based on the evaluation of six separate quality factors (defocus, motion blur, occlusion, specular reflectance, illumination, and pixel count).

Using the above mentioned quality assessment algorithms, a quality vector is computed for a given image. Figure 3.3 shows quality metrics obtained from sample images of each modality. For a given gallery-probe pair, the quality vector of both gallery and probe images are concatenated to form the quality vector Q , represented as $Q = [Q_g, Q_p]$, where Q_g and Q_p are the quality vectors of gallery and probe images respectively. This quality vector is then used in the proposed context switching algorithm to dynamically select a biometric matcher/fusion algorithm. These quality metrics are based on using computationally inexpensive cues to determine the work flow of the system as they are known to be indicative of the ‘ability’ of a given sample to be identified by a given biometric modality. However, these may be replaced by other metrics [210] or meta-data depending

on the specific use-case scenario. As shown in Figure 3.4, the quality vector for each modality is computed sequentially one at a time when the modality is requested. Unlike existing approaches that process all modalities simultaneously, the proposed approach enhances the computational ease by processing each modality serially. The quality vector of the next modality is computed only when the previous modality either fails to efficiently classify the given gallery-probe pair or is not available for processing.

3.3.2 Context Switching Algorithm

As mentioned before, context switching refers to switching from one modality to another depending on the quality of a given gallery-probe image pair. It is observed that when the quality of a given gallery-probe pair is good, uni-modal matchers are sufficient to classify the given pair as genuine or impostor; however, when the quality is poor, fusion of different modalities is required for classification. The proposed context switching algorithm is hierarchical in nature (as shown in Figure 3.4) and the selection of biometric modality/unimodal matcher is posed as a classification problem. In this research, Support Vector Machine (SVM) is used because (i) it has been shown to provide better performance for higher dimensional classification tasks [166], and (ii) the learning is dependent on representative samples rather than the numbers of training samples. At level-0, the quality vectors of the gallery-probe pairs are processed by the first SVM to determine the biometric modality to be used and at level-1, the matcher for the biometric modality is selected using the second SVM. The algorithm allows each modality to be processed one at a time, and the control is passed to other modality only if the first modality is not sufficient to match the given gallery-probe pair. The context switching algorithm is divided into three stages: training the SVMs, online learning during new enrolments, and dynamic matcher selection during probe verification. Each of the three steps is explained below in detail.

3.3.2.1 Training the SVMs

The SVMs for each biometric modality are trained independently using the labeled training data. For each modality, SVM classifier at level-0 and level-1 are referred to as SVM_{L0} and SVM_{L1} respectively. Further details about training the SVM at each hierarchical level are elaborated below:

1. At level-0, a binary SVM is trained for the j^{th} biometric modality using the labeled training data $\{x_{ji}, y_{ji}\}$. Here, input $x_{ji} = [Q_g, Q_p]_{ji}$ is the quality vector of the i^{th}

gallery-probe pair in the training set. $y_{ji} = \{-1, +1\}$ is the label such that $\{-1\}$ is assigned when the gallery-probe pair can be correctly classified using the matchers in the given modality, otherwise, $\{+1\}$ is assigned (i.e. control is switched to other modality). Using the labeled training data, three SVM_{L0} are trained: one for each modality.

2. Similarly, SVMs at level-1 are also binary classifiers trained for selecting one of the unimodal matchers for a given modality. For the j^{th} biometric modality, input to SVM_{L1} is also the quality vector $[Q_g, Q_p]_{ij}$ of the i^{th} gallery-probe image pair in the training set. As shown in Figure 3.5, the labels are assigned based on the distribution of the genuine-impostor scores and the likelihood ratios [123]. For a matcher, if the score corresponding to a gallery-probe pair is greater than the maximum genuine score (i.e. confidently matched as impostor) or is less than the minimum impostor score (i.e. confidently matched as genuine), then this matcher can efficiently match the pair as genuine or impostor. Label $\{-1\}$ is assigned when matcher1 can confidently match the gallery-probe pair, otherwise, label $\{+1\}$ is assigned to select matcher2. If both the matchers correctly match the given gallery-probe pair then likelihood ratio is used to break the tie. The gallery-probe pair is assigned the label corresponding to the matcher that is more likely to match the gallery-probe pair as genuine (likelihood > 1) or impostor (likelihood < 1). The scores from the matchers are converted to distance scores (wherever required) before assigning the labels. For each modality, the labeled data is then used to train SVM_{L1} .

3.3.2.2 Online Learning

As mentioned previously, large scale programs such as US Visit (now OBIM: Office of Biometric Identity Management)¹ and Aadhaar² continuously enroll new subjects on a regular basis while performing probe verification. With increase in the number of enrolment, the quality and match-score distributions tend to drift. The classifiers trained with small training samples are unable to generalize well to this concept drift and require re-training to accommodate the variations in data distribution. Re-training the classifiers in batch mode with all the existing data is not pragmatic as it requires large amount of time. Online learning provides an efficient way to sustain the performance by addressing the variations in data (match score and quality score) distribution introduced by the newly

¹<http://www.dhs.gov/obim>

²<http://uidai.gov.in/>

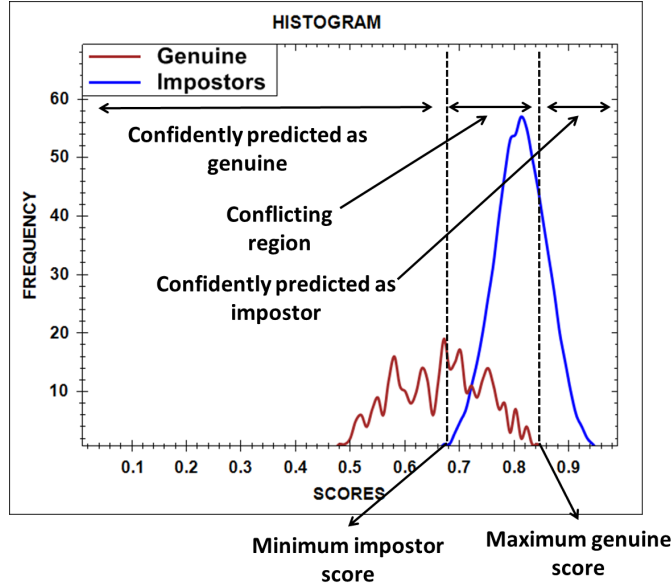


Figure 3.5: Illustrating the process of assigning labels: genuine and impostor match score distributions are used to assign labels to the input gallery-probe quality vector $Q = [Q_g, Q_p]$ during SVM training.

enrolled individuals. The idea of context switching algorithm that can evolve with increasing number of new users is novel and to the best of our knowledge, this work presents the first approach for online context switching in biometric literature. Labeled information from only “newly enrolled individuals” is used to update the proposed context switching algorithm (i.e. decision boundary of SVM classifiers) in an online learning (incremental+decremental) manner. In this research, we add or remove one sample at a time to update SVM using the incremental+decremental method proposed by Cauwenberghs and Poggio [34]. They proposed a solution for $N \pm 1$ samples that can be obtained using the N old support vectors and the sample to be added or removed. SVMs are first trained using an initial training database and a decision hyperplane is obtained, as illustrated in Section 3.3.2.1. Online learning algorithm for updating the SVM with additional labeled instances from new enrolments and m support vectors learned on an initial labeled training data D_L is described in Algorithm 1.

During new enrolment, a unique identification is assigned to every user. Impostor scores are computed by comparing a new enrollee with the stored gallery. For genuine scores, the enrollee is compared with its own multiple samples captured during enrolment. The labels (ground truth) corresponding to the enrollee are known during enrolment and are compared with the predictions of the SVM classifiers. The decision boundary of SVM

Algorithm 1 Online learning with new enrolments

Input: A SVM model, N additional labeled instances $\{x_i, z_i\}$.

Process: Online training of SVM model

for $i = 1$ **to** N **do**

 Predict labels: $SVM(x_i) \rightarrow y_i$.

if $y_i \neq z_i$ **then**

 Update SVM decision boundary with labeled instance $\{x_i, z_i\}$ & m support vectors.

end if.

end for.

Output: Updated SVM model.

classifier is updated in online fashion only for the instances for which the classifier makes incorrect predictions. The process of updating SVM decision boundary using the new available instances and the previous support vectors is elaborated below:

1. Let x_i be the instance for which SVM needs to be updated and z_i is the associated label.
2. SVM decision hyperplane is recomputed using the m trained support vectors and the new training instance $\{x_i, z_i\}$ using standard batch mode, as explained in Section [3.3.2.1](#).
3. The number of support vectors may increase on recomputing the hyperplane. To avoid the number of support vectors from growing in an uncontrolled manner, a threshold λ is introduced that controls the number of support vectors. If the number of support vectors is more than $m \pm \lambda$, then the farthest support vector from the current decision hyperplane is selected.
4. The farthest support vector is then removed from the list of support vectors and is added to a separate list, l . The classifier with remaining $m + \lambda - 1$ support vectors is the updated classifier.
5. The support vectors in the list l are used to test the updated classifier. If there is any misclassification, Step 2 is repeated to minimize the classification error on the removed support vectors.

Online learning is used to update both the classifiers (SVM_{L0} and SVM_{L1}) in each modality. It facilitates to update the context switching algorithms with the varying quality and match score distribution. In this work, SVM with radial basis function (RBF) kernel is used where $\gamma = 6$.

3.3.2.3 Context Switching during Verification

For verification, the trained SVMs are used to select the modality and the most appropriate matcher for matching an individual using the quality of gallery-probe pairs. Each biometric modality is used one at a time and the second modality is invoked only when the unimodal matchers in the first modality are unable to classify the gallery-probe pair. The proposed online context switching algorithm processes a given gallery-probe pair as explained below:

1. The quality scores of the gallery-probe pair $Q = [Q_g, Q_p]$ corresponding to the first modality are computed and provided as input to the trained SVM_{L0} of the first modality. Based on the quality vector, SVM_{L0} predicts if the matchers in this modality can be used to correctly classify the pair or not.
2. If it predicts that unimodal matchers in the first modality can be used to correctly classify the given gallery-probe pair, then SVM_{L1} in first modality is used to predict the unimodal matcher that should be selected.
3. Otherwise, if SVM_{L0} predicts that the matchers in the first modality cannot efficiently classify the gallery-probe pair, quality vector corresponding to the second modality is computed and provided as input to the corresponding SVM_{L0} of that modality. SVM_{L0} pertaining to the second modality predicts whether the matchers in this modality can be used to correctly classify the given gallery-probe pair or not.
4. It is done serially for all the modalities until an appropriate unimodal matcher is selected.
5. If none of the unimodal matchers can efficiently classify the given gallery-probe pair, normalized score level fusion [158] of unimodal matchers across all modalities is selected to process the gallery-probe pair.

It should be noted that for probe verification, the algorithm does not require computing the match score between the gallery-probe pair to select the most appropriate matcher and SVM prediction is based only on the quality vector ($Q = [Q_g, Q_p]$) of the gallery-probe pair. Further, the proposed context switching algorithm is generic and can be modified to add or remove biometric modalities and matchers within a modality.

3.4 Experimental Results

To evaluate the effectiveness of the proposed context switching algorithm, the performance is evaluated on two multimodal biometric databases captured in significantly different conditions. The performance of the algorithm is also compared with individual unimodal matchers and normalized score level fusion. The effectiveness of the proposed online learning in updating the context switching algorithm is also compared with batch training. Additionally, a comparison is made with the parallel quality based context switching approach by Vatsa *et al.* [187]. Section 3.4.1 presents the unimodal matchers used in the algorithm, Section 3.4.2 illustrates the database characteristics and experimental protocol and Section 3.4.3 presents the results and key observations.

3.4.1 Unimodal Matchers

Three biometric modalities namely, face, fingerprint, and iris are used in the proposed context switching algorithm. For face, two matchers are used: Uniform Circular Local Binary Pattern (UCLBP) [6] as face matcher1 and Speeded Up Robust Features (SURF) [53] as face matcher2. UCLBP is computed with circular encoding of eight neighboring pixels evenly positioned on a circle of radius two. SURF is a scale and rotation invariant descriptor [53] that computes the descriptor from the spatial distribution of gradient information around the interest points. To match two corresponding UCLBP features or SURF descriptors, χ^2 distance measure is used. Fingerprint matcher1 (NBIS)¹ uses a minutiae based fingerprint matching algorithm. A commercial fingerprint matching software (Neurotechnology Veri-Finger) is used as fingerprint matcher2. Iris matcher1 is an implementation of the algorithm proposed by Vatsa *et al.* [184] which uses curve evolution based segmentation and 1-D log polar Gabor filters. Commercial iris matching software (Neurotechnology Veri-Eye) is used as iris matcher2. In each case, standard parameters from the corresponding cited work are used. Further, match scores are normalized using min-max normalization and sum rule is used for score level fusion [159].

3.4.2 Database and Experimental Protocol

The performance of the proposed algorithm is evaluated on two databases: WVU multimodal database [45] and the database provided by a Law Enforcement Agency (referred to as the “LEA” database).

¹<http://www.nist.gov/itl/iad/ig/nbis.cfm>

- **WVU Multi-modal database** [45] comprises fingerprint, face, and iris images corresponding to 270 subjects with multiple samples per subject. The database is divided into three parts: 1) initial training, 2) online learning, and 3) testing. The training database comprises images pertaining to 108 subjects (40% of the total database) and the remaining images pertaining to 162 subjects are used for online learning and performance evaluation. The gallery comprises two images per subject and the remaining images are used as probe.
- **LEA database** is a multimodal database captured in unconstrained real world conditions with uncooperative users and comprises fingerprint, face, and iris images. The database is noisy, has both good and poor quality images, and has missing data as well; for instance, information from all the modalities are not available for every individual. The images are divided in two sets, set A and set B, and each set consists of 18,000 samples. LEA database is divided into initial training, online learning, and testing. The initial training is performed on images corresponding to 4,500 individuals, the online learning of the context switching algorithm is performed on the next 4,500 individuals, and the performance is evaluated on images corresponding to 9,000 individuals. In cases where one or more biometric modality is missing, the algorithm uses samples from other biometric modalities to perform context switching.

Face images in both the databases are normalized and the size of each detected image is 196×224 pixels. The proposed context switching algorithm selects the most appropriate matcher to process the gallery-probe pair based on the quality. QFuse also evolves with new enrolments to accommodate the variations in data distribution in online manner. During training, the parameters of feature extractors are learned and SVMs are trained using the gallery-probe quality vector and match score as explained in Section 3.3.2.1. During online learning, the decision boundary of SVM classifiers is modified to update the context switching rule as described in Section 3.3.2.2.

3.4.3 Results and Analysis

Figures 3.6 and 3.7 show the ROC curves comparing the performance of the proposed context switching algorithm with different unimodal matchers and sum-rule fusion. QFuse is also compared with a parallel quality based context switching framework of Vatsa *et al.* [187]. The key results and analysis are listed below.

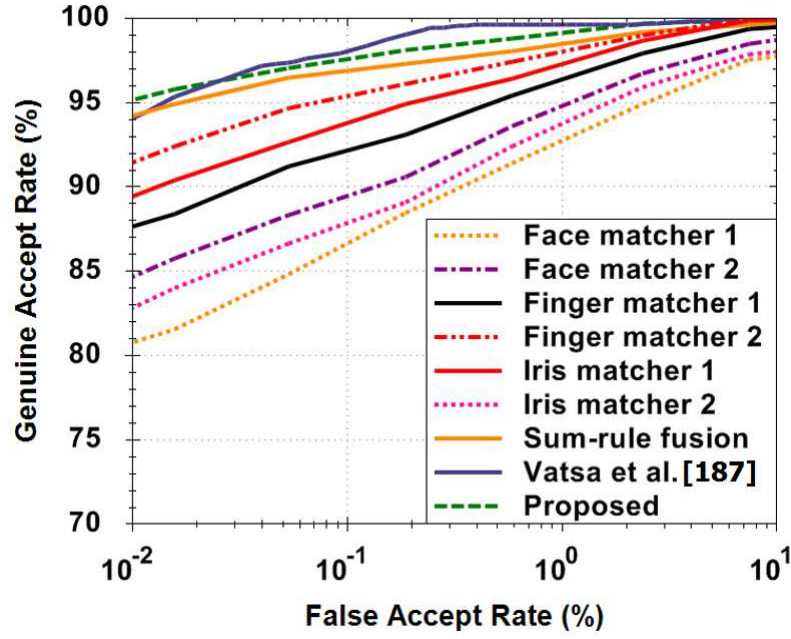


Figure 3.6: ROC curves of the proposed quality based context switching algorithm QFuse and comparison with different unimodal matchers, sum-rule fusion and Vatsa *et al.* [187] on the WVU database [45].

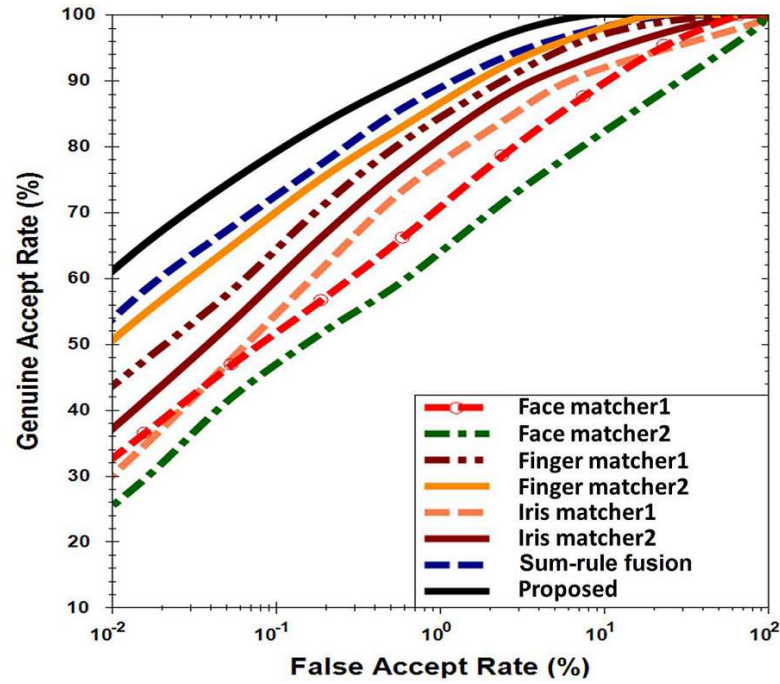


Figure 3.7: ROC curves of the proposed quality based context switching algorithm and comparison with different unimodal matchers and sum-rule fusion on the LEA database.

- Figures 3.6 and 3.7 show that the proposed quality based context switching algorithm outperforms the unimodal matchers by at least 3.8% and 10.6% on the WVU multimodal and LEA databases respectively. It also outperforms sum-rule fusion of unimodal matchers by at least 0.9% and 7.4% on the two databases respectively. The improvement is attributed to the fact that when images in a given modality are of bad quality, the context switching algorithm selects another modality or uses sum-rule [158] fusion of different unimodal matchers to process the gallery probe pair.
- At lower FARs, the proposed approach yields improved verification performance compared to the parallel quality based context switching approach of Vatsa *et al.* [187]. The results show that comparable performance is achieved despite an additional evidence theory based fusion approach employed in the context switching framework of Vatsa *et al.* [187] along with the sum rule fusion. It is also observed that the proposed algorithm is computationally less expensive (at least 1.5 times faster) due to its serial nature, avoiding computation when strong evidence is obtained from a unimodal matcher. Since the parallel nature of the existing approach requires that all the evidences of the identity (i.e., all captured biometric modalities) should be available before processing and the LEA database contains instances of missing data, Vatsa *et al.*'s algorithm fails to process majority of the samples from the LEA database. Therefore, in this research, the results of the proposed and existing algorithms are compared only on the WVU database.
- The results suggest that the performance of unimodal matchers reduces significantly on the real world challenging LEA database due to noisy images. The proposed quality based context switching algorithm sustains its performance across noisy images and even when images from some modalities are not available. Since the size of the LEA database is large, the results also suggest that online learning makes it scalable as it adapts to changes in data distributions.
- To evaluate the effectiveness of QFuse, the total number of individuals available for online learning are divided into 10 equal size batches and the average performance is reported by considering each batch sequentially. As shown in Table 3.2 and Figures 3.8a and 3.8b, online learning provides a minor improvement in verification accuracy over batch/offline training. However, Table 3.2 reports the overall training time for online learning and batch/offline training. The combined results (accuracy and time)

validate our assertion that the proposed online context switching algorithm provides significant reduction in training time (at least 1/4 of the batch/offline training time) and sustains the performance with increasing number of users.

- In the proposed algorithm, quality scores of a particular biometric modality are computed only when that modality is requested in the serial framework. Moreover, the algorithm can skip a modality, if the images for that modality are not successfully captured. Computationally, on an Intel i5 processor with 4GB RAM, the proposed algorithm requires an average of 3.1 seconds for quality assessment, feature extraction, context switching, and matching.
- Figures 3.9 and 3.10 illustrate few examples where the proposed context switching algorithm selects different unimodal matchers or their fusion for different quality gallery-probe image pairs. Since the images in the LEA database are captured in extremely harsh unconstrained real world environment with uncooperative users and are of poor quality, as shown in Table 3.3, 36.3% instances are processed using multimodal fusion. On the other hand, in the WVU multimodal database, which is prepared in controlled lab conditions with only selected irregularities; fusion is selected for only 14.7% instances. In both the cases, the results show that our assertion (i.e. for a good quality gallery-probe pair, unimodal matchers are sufficient and fusion should be used only for poor quality images) holds true and the proposed algorithm improves both accuracy and time.
- In this research, different modalities are arranged serially and each gallery-probe pair is processed starting from the strongest biometric modality. However, the hierarchy of the classifiers can also be decided based on the particular application scenario, based on performance, user convenience or other domain specific information.

The proposed quality based serial context switching algorithm can be easily extended to include other biometric modalities, unimodal matchers, and fusion rules. Since the LEA database is collected in-field by law enforcement officials, improved results (in terms of accuracy, computational time, and scalability) suggest that such a algorithm is very useful in real world large scale applications.

To evaluate the effectiveness of the proposed context switching algorithm, the performance is evaluated on two multimodal biometric databases captured in significantly different conditions. The performance of the algorithm is also compared with individual unimodal matchers and normalized score level fusion. The effectiveness of the proposed

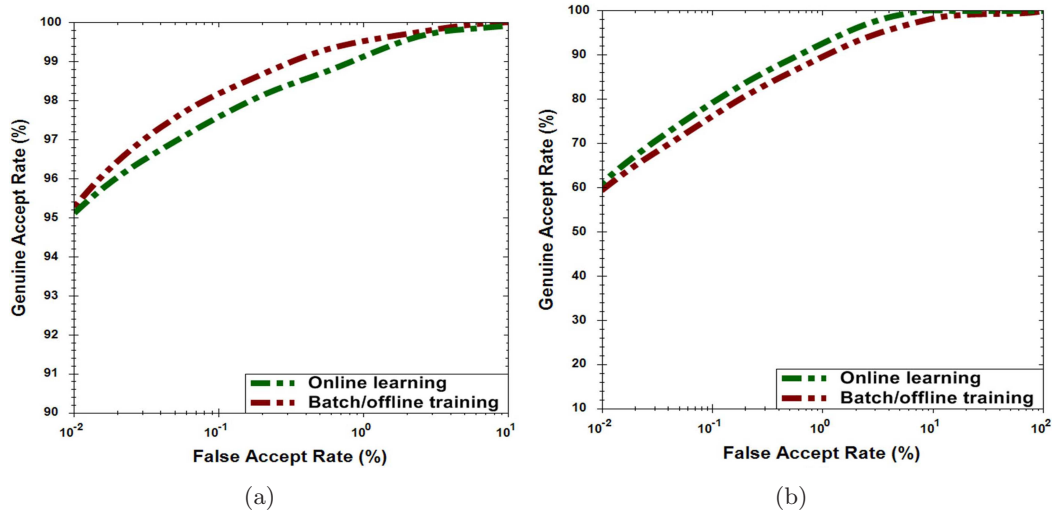


Figure 3.8: Improvement with online learning over batch/offline learning on the a) WVU multimodal database [45] and b) LEA database.

Table 3.1: Verification accuracy of individual matchers, fusion algorithms, and QFuse at 0.01% false accept rate (FAR).

Algorithm	Accuracy(%)	
	WVU	LEA
Face matcher 1	84.6	32.1
Face matcher 2	80.7	25.6
Finger matcher 1	87.4	43.7
Finger matcher 2	91.3	50.3
Iris matcher 1	89.1	30.1
Iris matcher 2	82.8	36.7
Sum-rule Fusion	94.2	53.5
Vatsa <i>et al.</i> [187]	94.7	–
Proposed	95.1	60.9

Table 3.2: Comparing the verification accuracy and computational time of QFuse when the training is performed in online and offline manner.

Database	Training	Accuracy (%)	Training time (min)
WVU	Online learning	95.1	13.2
	Batch/offline training	95.3	82.5
LEA	Online learning	60.9	95.4
	Batch/offline training	59.7	393.7

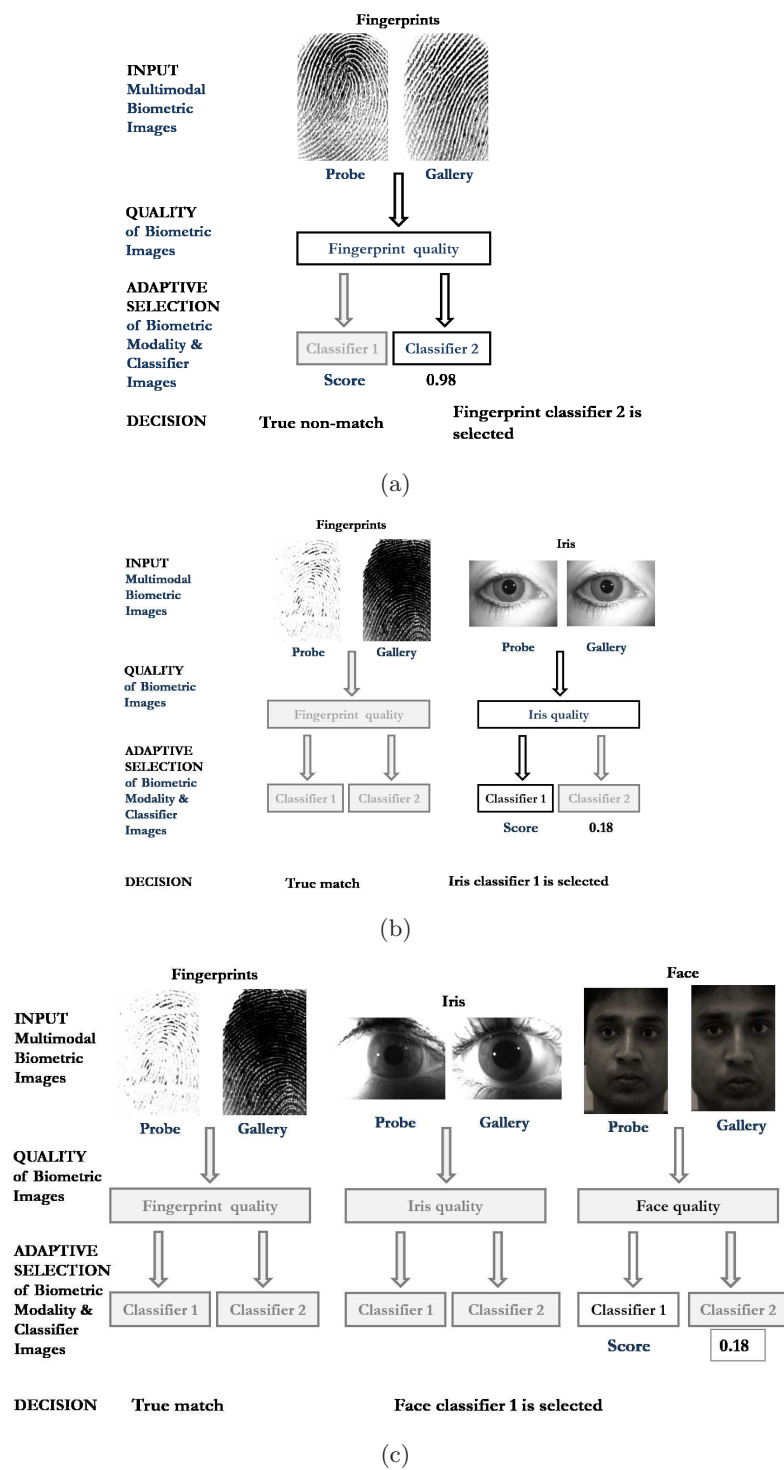


Figure 3.9: Illustrating sample cases where unimodal a) fingerprint, b) iris, and c) face matchers is selected.

Table 3.3: Illustrating percentage of instances processed by individual components involved in the proposed context switching algorithm.

Matcher	% instances	
	WVU	LEA
Face matcher 1	10.6	12.6
Face matcher 2	8.9	11.3
Finger matcher 1	19.5	17.5
Finger matcher 2	17.3	13.3
Iris matcher 1	15.6	5.6
Iris matcher 2	13.4	3.4
Sum-rule fusion	14.7	36.3

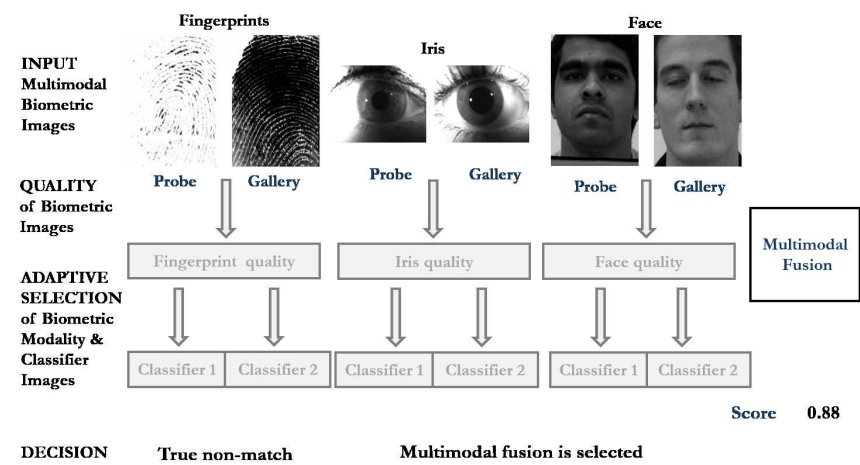


Figure 3.10: Illustrating a sample case where multimodal fusion is selected.

online learning in updating the context switching algorithm is also compared with batch training. Additionally, a comparison is made with the parallel quality based context switching approach by Vatsa *et al.* [187].

3.5 Summary

In biometrics, *evidence fusion* paradigm has been widely used to establish the identity of an individual with greater confidence. Extensive research has been performed for controlled environment with cooperative users and higher accuracy has been achieved. However, there is a need to enhance the capabilities when operating under uncontrolled environment with noisy data. This thesis presents QFuse, a context switching algorithm coupled with online learning that fills the gap in the current state-of-the-art. The proposed algorithm analyzes the biometric samples that may be from diverse sensors with varying *quality*. It adaptively makes a decision if a unimodal biometric matcher can reliably verify the individual or a fusion rule is required. This research also updates the context switching algorithm using online learning approach in order to (1) make it scalable and (2) adapt to drift in data distribution due to new enrolments. The experimental results show that the proposed algorithm optimizes the accuracy and computation time for challenging large scale applications. Further, it is our assertion that this algorithm can be easily extended for other multi-matcher problems in pattern recognition and machine learning applications.

Chapter 4

Aiding Face Recognition with Social Context Association Rule based Re-Ranking

Humans are very efficient at recognizing familiar face images even in challenging conditions. One reason for such capabilities is the ability to understand social context between individuals. Sometimes the identity of the person in a photo can be inferred based on the identity of other persons in the same photo, when some social context between them is known. This chapter presents an algorithm to utilize the co-occurrence of individuals as the social context to improve face recognition. Association rule mining is utilized to infer multi-level social context among subjects from a large repository of social transactions. The results are demonstrated on the G-album and on the SN-collection pertaining to 4675 identities prepared by the authors from a social networking website. The results show that association rules extracted from social context can be used to augment face recognition and improve the identification performance.

4.1 Introduction

Face recognition capabilities of humans have inspired several researchers to understand the *science* behind it and use it in developing automated algorithms. Recently, it is also argued that encoding *social context* among individuals can be leveraged for improved automatic face recognition [175]. As shown in Figure 4.1, often times a person's identity can be *inferred* based on the identity of other persons in the same photo, when some social context between them is known. A subject's face in consumer photos generally co-occur along with their socially relevant people. With the advent of social networking services, the social context between individuals is readily available. Face recognition performance

in such photos can be improved by considering this contextual information and can have interesting applications in image sharing, forensics, and intelligent surveillance systems.

The term *context* has been used in object recognition [182], person detection and also in face recognition research [105, 164] to imply acceptable co-occurrence of various parts or attributes of an object or face. Researchers compute context from scene recognition and/or spatiotemporal information obtained from image headers, camera devices, user labels for events, locations, people, and other secondary sources. In this research, *social context* refers to the possibility of co-occurrence of two or more individuals in a single photo. Since consumer photography is generally a social exercise, involving people of social importance, an improved understanding of *social context* can augment the quality of face recognition and tagging.

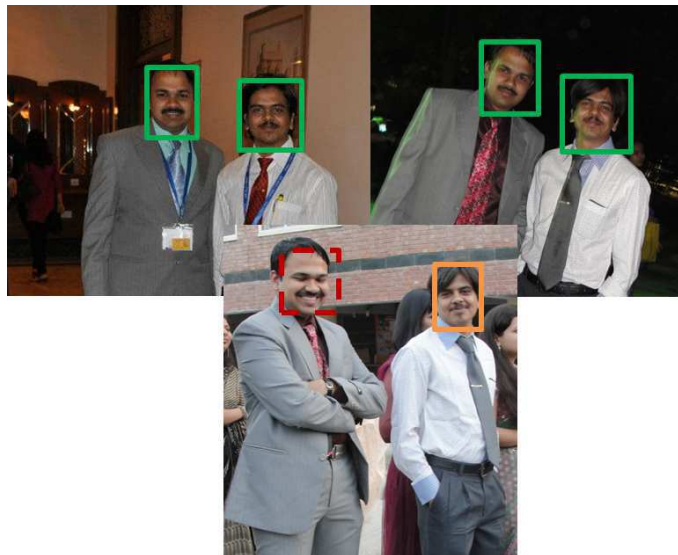


Figure 4.1: In this illustration, context derived from two images help confirm identity of a face in third. *A professor meets a couple of PhD students, A and T, after an invited talk at a university. Later at a conference, he is unable to recognize A but infers his identity when he sees A and T seated together.*

4.1.1 Related Work

We briefly review literature of various problem domains that use social context to aid face recognition performance. As illustrated in Figure 4.2, augmenting the performance of face recognition using various cues have been studied in literature with two application scenarios:

- i) *Organizing photo album collections*: Personal photo collections stored in computers and smart phones are generally clustered based on events, people, locations. Several research

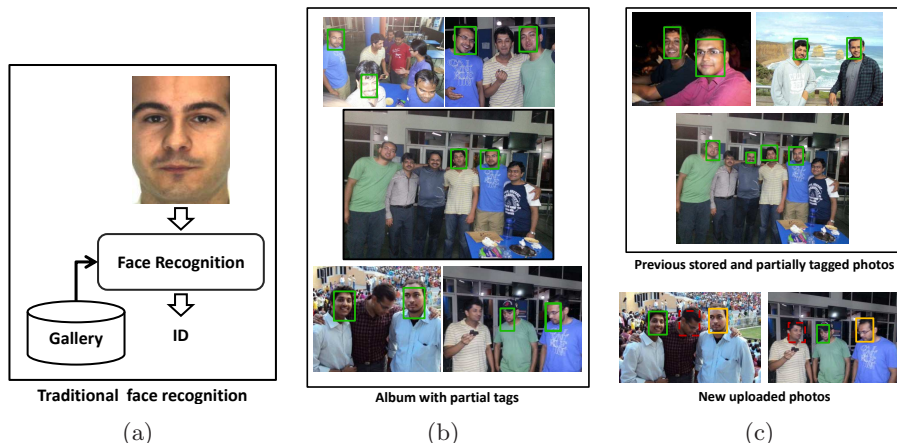


Figure 4.2: Three problem domains: (a) traditional face recognition, (b) photo album organization (based on events, locations and people) with imperfect (manual) annotations and (c) reliable face tagging approaches of large number of uploaded photos, which is the focus of this work.

directions have been undertaken to improve automatic photo organizer by attaching meaningful labels pertaining to identities, relationships, and other demographics including the use of meta-information from capture devices (cellID, GPS, time) [51, 135]. The popularity of digital photos has also lead towards commercial photo management tools such as Google Picasa, Microsoft EasyAlbum [46] and Apple iPhoto.

Kapoor *et al.* [93] used active learning to minimize manual labor by inferring probabilistic models to simultaneously tag people, events, and locations by deriving cross-domain relationships in semi-supervised settings. Other researchers have used heuristically generated priors based on rules such as *height of husband is greater than height of wife* [195]. O’Hare *et al.* [135] combined several context cues derived from text, event detection, image color descriptors and body part analysis to improve person identification in photo collections. Gallagher and Chen [63] improved ambiguous person labels in image collection by learning group priors to classify unlabeled persons. In another approach [62], they evaluated relative position of people in group photos to improve gender and age classification. Further, clothing [61] is also used to cluster images from the same event. Face clustering approaches incorporate contextual constraints such as *same-day* (a person on a given day is wearing the same clothing throughout), *Person-Exclusive* constraint (a person can occur in a photo only once), and co-occurrence [215] to improve recall. Manyam *et al.* [114] questioned the independence assumption between features extracted from two faces in the same image. Exploiting the consistent environmental settings, relative features are extracted to augment joint face recognition with a conditional probabilistic model. Chen

et al. [42] infer pair-wise relationships and social-subgroups within images (such as siblings, couple etc.) based on term frequency of low level visual features. Satish *et al.* [163] explore context discovery from multiple information sources. Recently, Barr *et al.* [14] use active learning to generate social constraints to improve face clustering and Hochreiter *et al.* [77] model co-occurrence of individuals in photo albums.

ii) *Reliable face tagging*: It has extensive utility, particularly for cloud based storage and online social networking services. A large number of consumer images are stored for easy access, sharing, and reliable storage (with over 300 million photos per day uploaded on Facebook alone [80]). The shared and tagged photos provide extensive social context to predict possible labels of an input photo without manual intervention. Becker *et al.* [17] perform preliminary evaluation of several established appearance based approaches on a set of images mined from Facebook. In another study, Stone *et al.* [174, 175] perform a large scale face recognition study on photos from Facebook. The match scores obtained between two subjects are augmented with a social context score, that measures whether both subjects are *friends* (acquaintance of one another on the social network site) or not. Hence, the social cue corrects the face match score and is suitably weighted within a conditional random field framework. The approach is also implemented on mobile phones [49], showcasing the applicability of social context augmentation in face recognition in realtime applications. Lin *et al.* [109] present a framework of tagging face images in consumer photos. Recently, Wu *et al.* [206] use belief propagation for identity discovery in a synthetically generated social network. Sapkota *et al.* [162] use an ensemble of context and facial feature classifiers to improve face recognition performance.

4.1.2 Research Contributions

As discussed, several techniques have been proposed in literature to augment face recognition with a social context. Social context such as clothing color, relative location of faces, and age/gender estimation have shown promising results on consumer photo datasets of relatively small size. In such cases, the number of subjects are limited to 10-15 (for example, family members). In order to robustly use context information in large, more uncertain consumer photo collections, such as social networking sites or surveillance systems, the applicability of context must be more closely re-examined.

This research aims to broaden the scope of face recognition in consumer photos using social context. Rather than binary cues that have been explored in literature, such as $\{friend, no-friend\}$, the proposed approach infers association between groups of individuals from multi-level social cues such as co-occurrence of people in consumer photos,

to improve face identification. These context cues are used to re-rank face recognition results to improve the overall performance. Secondly, to evaluate the performance of the proposed algorithm, a large dataset is mined from a leading social networking site consisting of 160,264 images from 4675 connected users. An anonymized subset of the social graph, together with rank orders of identities obtained from a powerful face recognition system on the uploaded photos are made available to the research community.

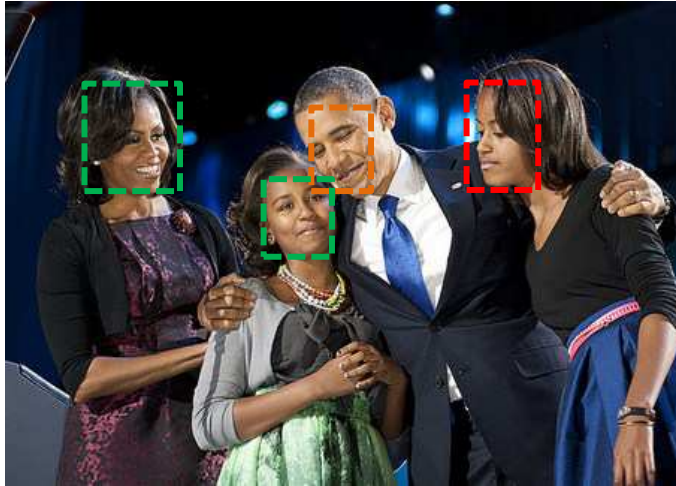


Figure 4.3: In case of challenging imaging condition, prior knowledge of association of subjects may induce additional information to improve face recognition. The proposed approach generates context association rules, $\{Michelle, Sasha\} \rightarrow Barack$ and $\{Michelle, Barack, Sasha\} \rightarrow Malia$, based on prior co-occurrence to aid face recognition.

4.2 Social Context Aided Face Recognition

The key concept of the proposed algorithm is illustrated in Figure 4.3. The algorithm first learns the social context from a set of training images. The contextual information is used to improve the face recognition ranking of a probe image obtained from any face recognition system. To replicate a practical scenario where a user uploads several consumer photos to a photo sharing service, a non-overlapping face identification scenario is considered in this research. The social context obtained is used to augment face recognition in unseen faces. The details of the proposed approach are as follows.

4.2.1 Building Context from Tagged Faces

In this research, co-occurrence of people in a consumer photo is used as transactions to improve face recognition performance. By viewing a photo consisting of more than one individual as a single *social transaction*, it is possible to infer contextual association rules (AR)

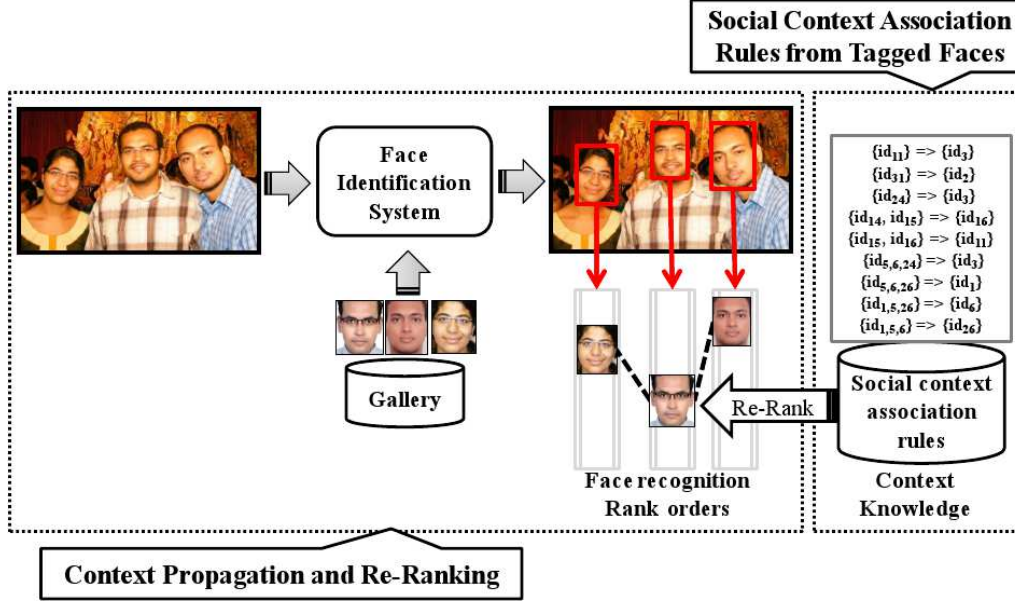


Figure 4.4: An illustration of the proposed multi-level association rule mining to derive social context to improve face identification.

that capture social context. For instance, a rule $\{Michelle, Barack, Sasha\} \rightarrow Malia$, can assist in recognition of off-angle faces in challenging imaging conditions such as in Figure 4.3. Association rules of the form $X \rightarrow Y$ are mined from *tagged* faces in group photos. Since there is no restriction on the cardinality of the antecedent of rule (X), the social cue inherently provides richer information regarding the co-occurrence of subjects. While tagging may not be reliable for accurate face labeling, useful contextual information may be derived from each image, based on which users co-occur in tags.

During the training phase, a large number of *social transactions* are utilized to create social context association rules. Higher frequency of occurrence of the same set of individuals in a large collection of social transactions is inferred as *a greater chance of their co-occurrence in the probe images*. Association rule mining is used to derive inferences from the training samples. Several such association rules encode the hierarchical (level-wise) co-occurrence between different clusters of individuals. In this work, we associate photos that contain two or more *tagged* individuals as a single social transaction. Further, we hypothesize that association rules thus derived can be used to augment face recognition performance in challenging settings.

In the testing phase illustrated in Figure 4.4, face detection and recognition are first applied to each probe image and rank ordering of gallery for each of the detected faces is obtained. Next, based on the confidence of detection and recognition, social context

rules involving the most confident match are used to *influence* the rank ordering of the remaining faces. Finally, the set of rank orders obtained via face recognition and different combinations of social context are fused to obtain a single rank order. The proposed approach is presented in Algorithm 2. A two stage augmentation is applied to each probe face, described as follows:

- **Unique Identity Pruning:** It can be safely assumed that the same person does not occur in a single photo more than once. Hence, if a label is assigned confidently to a face, the probability of the same label being assigned to another face in the same photo is suitably reduced.
- **Social Co-occurrence:** Based on the tags from previous photos in the collections, multi-level social context is inferred and used to improve face recognition by adjusting rank order. The mining of association rule and fusion scheme are discussed in detail next.

4.2.2 Social Context Association Rules from Tagged Faces

Association rule mining is a data mining technique widely used in literature to extract rules from transactional data for applications such as market-basket analysis. These rules provide insights into the behavior of customers by inferring relationships between frequent items that are in consideration. AR mining is also used in web user behavior analysis, intrusion detection, DNA sequencing, various web recommendation services and sub-group discovery. Liu *et al.* [111] use the Apriori algorithm [4] to post-filter semantic concepts that are detected in videos using association rules between known semantic concepts. An association rule of the concept *sky* \rightarrow *outdoor* has higher confidence than with *indoor*, hence post-processing improved performance of concept detection in videos. Given a set of social transactions, we define the following:

- **Social Context Association Rule:** An association rule of the form $X \rightarrow Y$ represents a social context between two *sets* of subject labels X and Y . Here, $X \cap Y = \emptyset$ and both X and Y are subsets of U , the set of all face labels. For multi-level context mining, only those association rules that have one item as consequent (righthand side) are considered, i.e., $\{I_1, I_2, \dots, I_{k-1}\} \rightarrow I_k$ for a set of tagged users, I_1, I_2, \dots, I_k . These type of rules are referred to as Class Association Rules (CAR).
- **Support** of a set X denoted by $\text{Supp}(X)$, is the probability of occurrence of the members of X in the given set of transactions. The support of an association rule of

the form $X \rightarrow Y$ represents the probability that members of X and Y have co-occurred in the transactions.

- **Confidence** of an association rule represents the probability that members of X and Y have co-occurred given the occurrence of X . Confidence may also be interpreted as the conditional probability $P(Y|X)$.

We consider the extensively used association rule mining algorithm, Apriori [4], that uses breadth-first (level-wise) search to determine rules based on the downward closure property of support. The iterative approach effectively finds frequent sets of k subjects that co-occur in the social transactions. Further, k subject set is used to generate $(k + 1)$ frequent subject set, focusing only on subjects that frequently occur in the transactions. Further, a pre-defined minimum threshold, *min_suppt*, is maintained to make the breadth-first search computationally tractable. The association rules are stored in lookup table (*ApR*) with the rule confidence (C) as *data* and rule items ($\{X \cup Y\}$) as the *key*. For large datasets, rules with higher support can be cached in memory, since they are more expensive to re-compute. In comparison to existing graphical modeling approaches in literature, association rules inherently capture a higher cardinality of associations. Figure 4.5 shows a set of 100 rules connected based on confidence. As illustrated, the rules capture dense information pertaining to n-tuple co-occurrences.

4.2.3 Context Propagation and Re-Ranking

Face recognition based rank order is first evaluated for confidence. The confident face match is used as evidence to infer contextual information. To compute the confidence of a rank order, the likelihood ratio $\frac{P(X=l_1)}{P(X=l_2)}$ of scores between top match and the next closet match is considered. A high likelihood ratio indicates a confident match that forms the basis for deriving social context of other faces in the same photo. The proposed approach builds social context in a photo starting from the *most confident* face match. Given a confident match x_{conf} , unique identity pruning is first performed on the remaining faces, as shown in Eq. 4.1.

$$\hat{P}(x) = (1 - P_{face}(x_{conf})) \times P_{face}(x) \quad (4.1)$$

Next, the rank ordering obtained from a face recognition system is re-ranked based on the social association rules obtained in the training phase. The updated score for a particular identity ($R(x)$) after re-ranking is obtained by a weighted aggregation of the

Algorithm 2 Proposed context association rule based re-ranking approach

input: Given social transactions for training T , gallery set G recognize faces in probe P

face detection: detect faces F in P

training phase:

create look-up table ApR of association rules from T

testing phase:

compute C_i , the confidence of match for F_i in P

sort: F by C

apply unique identity pruning on rankorders in F (as in Eq. 4.1)

compute context scores $S = \text{Context}(F, ApR)$

update scores by re-ranking F (as in Eq. 4.2)

output: social context re-ranked faces for probe P

procedure: $\text{Context}(F, ApR)$

compute *Rankscores* of F_n with social context from all subsets of $F_1, F_2 \dots F_{n-1}$

return: *Rankscores*

end procedure:

4.3 Database and Protocol

The proposed approach is evaluated on a publicly available database (G-album) [61] along with a large database, SN-collection, collected by the authors from a social networking site. These face databases are selected instead of standard large test databases in order to demonstrate the advantage of derived social context in augmenting face recognition performance on challenging consumer photos. The details of the databases are summarized in Table 4.1.

4.3.1 The G-album

It consists of 589 images pertaining to 32 subjects from various family events. The photographs are taken in unconstrained settings with illumination, expression and pose variations. Most of the photographs consist of more than one subject with high co-occurrence of several subjects. The experimental protocol is designed to replicate a typical scenario in which a collection of existing photographs are available with tagging. For the G-album database, 50% images from the entire database are set aside as probe (depicting uploaded images) and the remaining are used as training samples. This forms a seen experiment, with different images of the same subjects in training and testing.

4.3.2 SN-collection

A major bottleneck in the progression of this research area has been the lack of a large database for comparison and evaluation. Photos from social networking services and cor-

responding social connections of the users are difficult to share without violating privacy concerns. Hence, for the benefit of the research community, we prepared the SN-collection database. The dataset is prepared using 160,264 images that are obtained from 65 active users (undergraduate students) and their self-declared, bi-directional *friends*, a total of 4,675 users. The experimental results are demonstrated on photos that are all tagged. A tag is a manually annotated identity label assigned to each face in a photo. One random face tag that overlaps with face detection rectangle of FaceVacs is assumed to be correctly labeled and used in gallery. In our experiments, 1455 images form the gallery set (containing one face instance per user) and 2893 images form the probe (note that multiple faces co-occur in a single image). From the remaining images, 10433 social transactions are obtained and used to extract social context association rules, at a $min_suppt = 0.001$. While tags may not be reliable to supplement face recognition or detection, useful contextual information may be derived from each image, based on which users co-occur in tags. The images obtained are extremely challenging and include illumination, expression, and pose variations along with low image resolution images. Another challenge that makes the with social networking data limiting [174] is that there are several subjects that are not under consideration making the problem an open-set experiment.

An anonymized social graph of this collection will be made available to the research community¹. As part of the distribution, we provide tagged face IDs of individuals (in anonymous form) within the social group in consideration. Further, we provide match scores obtained from a commercial face recognition system, FaceVacs. Due to privacy and legal constraints, face images or any other identifiable information are not provided. It is our assertion that this dataset will allow researchers to examine the social context in the photos and improve automatic tagging that is assisted with face recognition technology.

Table 4.1: Experimental protocol of both the databases.

	The G-album	SN-collection
Probe samples	524	2893
No. of subjects	32	4675
No. of detected faces	971	8316
No. of transactions	65	10433
No. of association rules	223312	72455

¹The database will be available at <http://iab-rubric.org/resources.html>

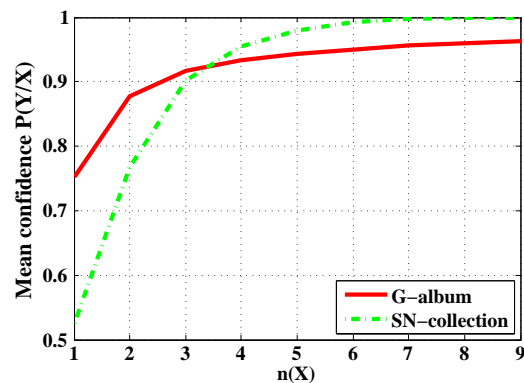


Figure 4.6: Mean confidence vs. cardinality of X of a social association rule $(X \rightarrow Y)$ shows more confidence as $n(X)$ increases.

4.4 Experimental Results and Analysis

The proposed social context based face recognition approach is presented in identification ($1 : N$) settings, as it is a more realistic application scenario. Any vanilla face recognition that provides a rank ordered gallery set can be augmented with social context extracted using the proposed approach. Here, FaceVacs face recognition system is used for face detection and matching. The system provides similarity scores between pairs of face images that are used to generate rank-ordering of the gallery set.

- The context association rules are mined from the social transactions from the G-album with *min_suppt* of 0.01 and 0.001 for SN-collection owing to the size of each dataset. A low support is indicative of low chance of occurrence of a set of individuals in the entire dataset. Further, for both datasets, only rules with confidence greater than 0.5 are chosen. A high confidence indicates high chance of their co-occurrence. On Intel Quad Core (2.5 GHz) processor and 4GB RAM, the training time on G-album is 23.4 seconds and SN-collection is 160.3 seconds.
- With sufficient number of transactions, rules with high confidence can be generated. As shown in Figure 4.6, the mean confidence of association rules obtained at higher cardinality are more confident. It implies that association rules obtained due to frequent co-occurrence of a large number of individuals are more confident.
- It has been reported in literature that annotation of faces in social networking sites are not reliable enough to be used as face detection [174]. In the proposed approach the tagged faces are used only to derive social context rather than to obtain facial

	The G-album (Rank-1)	SN-collection (Rank-25)
Face Only	79.39%	50.65%
Stone <i>et al.</i> [174]	79.41%	50.65%
Proposed	81.33%	55.72%

Table 4.2: Identification accuracy (%) on two databases.

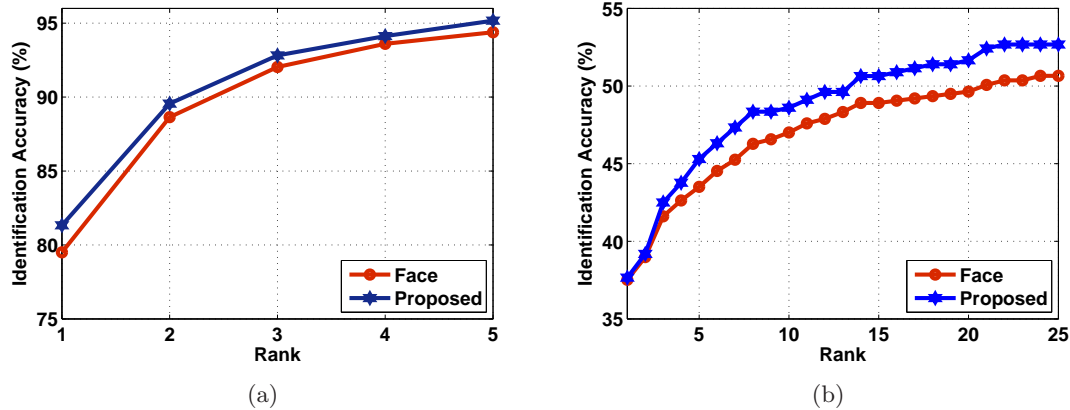


Figure 4.7: Identification accuracy obtained on a) G-Album database and b) SN-collection database.

features for matching. We affirm that the social context of co-occurrence of individuals is more likely to be maintained even if the tagging of facial regions is not positionally accurate.

- Figure 4.7 shows the identification accuracy (using cumulative match characteristics (CMC) curve) of only the face recognition system and the improvement achieved by multi-level social context. The improvement in performance is attributed to the re-ranking based on association rules obtained from the training data. Moreover, it is observed that, on both the datasets, retrieving context information during test/probe has marginal effect on overall computational time. A Kendall tau rank correlation test (social context vs face recognition) shows low correlation: 0.14 with G-album and 0.18 with SN-collection. This indicates that social-context provides additional evidence of identity that improves recognition performance.
- As shown in Table 4.2, the performance of the proposed approach improves when compared to face recognition alone. Further, the proposed approach is compared with pair-wise conditional random field [174], where pair-wise constraints are obtained by combining co-occurrence and friendship flags. The proposed approach is



Figure 4.8: Illustrating three cases from the G-album dataset where initially a face is recognized incorrectly at rank-1. With the application of the proposed context association rule based re-ranking, using the correctly recognized faces and social context, rank-1 performance is achieved for these samples.

at an advantage as higher order co-occurrences are utilized which lead to improved accuracy.

- In order to investigate the performance of the re-ranking approach, it is compared with Borda count rank aggregation [56]. This approach performs weighted fusion of rank orders obtained from face recognition and context rather than relevance probabilities. However, the performance of rank fusion approach is found to be lower than the proposed approach. On the G-album, it provides the rank-1 accuracy of 80.33% and on the SN-collection, rank-25 of 50.65% is achieved.
- The performance of the proposed context propagation approach is dependent on the correctness of the most confident face recognition aided tag. In album organization applications, it is often assumed that some faces are manually tagged by users. To explore the performance of context in face recognition for scenario illustrated in Figure 4.1, rank accuracy with single correct label is computed. The performance of the proposed approach with the assumption of a single correct label improves rank-1 accuracy to 92.5% on G-album and 76.71% on SN-collection. This also indicates that the proposed approach benefits from precise face matching.
- We re-emphasize that the experimental setup is motivated from social networking and cloud photo storage services (illustrated in Figure 4.2c). The social context is derived from a partial subset of the available datasets and used to predict face labels in *unseen* photos. This is a deviation from existing work (discussed earlier) that focuses on album organization application where context from the entire dataset is used from face clustering.

- Figure 4.8 illustrates sample instances where one face (right most) is incorrectly matched at rank-1 using face recognition alone. However, the proposed context re-ranking approach uses social context derived from the other images to achieve rank-1 accuracy.

4.5 Summary

This research presents an algorithm to augment automatic face recognition with multi-level social context. The proposed approach utilizes association rule mining techniques to extract social cues from co-occurrence of individuals in consumer photos. It is evaluated on G-album, a small publicly available database, and SN-collection, a large database collected by the authors from a social networking site. The results show that multi-level social context helps in improving face identification performance with marginal computational overhead. Social context can be improved with online rule generation and combined with other cues [28] to further improve performance.

Chapter 5

Learning Based Encoding and Distance Metric Approach to Newborn Face Recognition

Biometric recognition of newborn babies is an opportunity for the realization of several useful applications such as improved security against swapping and abduction, accurate census and effective drug delivery. However, the uncooperative nature, elastic faces, and lack of a publicly labelled database makes face recognition a challenging problem. This chapter explores the possibility of enhancing face recognition of newborns by harnessing large number of domain-specific adult face images and unlabeled problem-specific newborn face images obtained from the web towards an affordable and friendly biometric modality for newborns. The largest publicly available database of 96 newborns collected from various hospitals to study face recognition, is first introduced and evaluated on a common benchmark test. This research then proposes a learning based encoding method based on deep neural networks to obtain an effective representation. Further, the representation is coupled with a learning based distance metric, an online SVM formulation of the one shot similarity, to match extracted features with low semantic gap. In a multiple gallery recognition settings, an identification performance at rank-1 of 78.5%, and a verification performance of 62.9% at 0.1% FAR is achieved.

5.1 Introduction

Automatic recognition of adult faces has received significant attention in literature [31, 54]. However, automatic recognition of newborns¹ is a challenging problem with various applications to humanity. Current technologies that utilize RFID bracelets, color-coded tags,

¹The terms newborns, babies and infants are used interchangeably.

Table 5.1: Newborn recognition using different biometrics. *For newborns and toddler recognition.

	Modality	Type	# Subjects
Shepard <i>et al.</i> [170]	Footprint	Manual	51
Pela <i>et al.</i> [143]	Footprint	Manual	1917
Jai <i>et al.</i> [89]	Footprint	Automatic	101
Weingaertner <i>et al.</i> [201]	Palmprint	Manual	106
Lemes <i>et al.</i> [106]	Palmprint	Automatic	20
Fields <i>et al.</i> [58]	Ear	Manual	206
Tiwari <i>et al.</i> [179]	Ear	Automatic	210
Jain <i>et al.</i> [86]*	Fingerprint	Automatic	20&70
Bharadwaj <i>et al.</i> [25]	Face	Automatic	34
This research	Face	Automatic	96

and footprinting to account for babies are shown to have limited reliability. It is reported [48] that during hospital stays, babies are transferred to and away from the mother about 23 million times a year, with an estimated 10% of the transfers being erroneous but most cases are corrected before discharge. Another study performed in the United States by Gray *et al.* [70] concluded that, in the 34 newborns that are admitted to a neonatal intensive care unit at any given day, there is 50% chance of incorrect identification. Several cases of newborns being delivered to wrong parents and corrected only after intervention have been reported throughout the world [12, 32, 127, 128, 129, 130]. Abduction of newborns from hospitals is another social challenge. In the United States, the National Center for Missing and Exploited Children (NCMEC) has reported 299 infant abductions since 1983 (including 2 in 2015). Of these cases, 46% abductions were from hospitals or other health care facilities, with more than half from the mother’s hospital room [127]. Methodical documentation of events has brought this problem to light in developed countries, however, infant swapping and abduction are major problem for developing countries as well such as India. With no/minimal security measures and non-availability of any cost effective solution, it is very difficult to ensure that babies are not accidentally or intentionally swapped. Though medical science techniques such as Deoxyribonucleic Acid (DNA) typing and Human Leukocyte Antigen (HLA) typing are accurate, they are costly and time consuming. Therefore, institutions such as hospitals and neonatal care centers can benefit from a user friendly and reliable biometric system for newborns. It is our assertion that face biometrics can be useful; however, domain/problem specific knowledge will have to be incorporated. As shown in Figure 5.1, newborns can pose multiple challenges which may not be applied in regular (adult) face recognition. Nevertheless, if developed carefully, such a recognition system may also be utilized for efficient vaccine delivery [86] and to uphold child’s right to be with biological parents.



Figure 5.1: Sample images pertaining to three newborns from the Newborn Face Database illustrating large facial variations during capture. We term newborns as *unintentionally un-cooperative* users of face recognition.

5.1.1 Related Work

Despite enormous advancements, use of *biometrics* for newborn identification is very limited. A brief account of existing work is presented in Table 5.1. Shepard *et al.* [170] presented an analysis of newborn footprints from 51 subjects that are examined by fingerprint experts. The experts were able to correctly identify only 10 babies using the footprints. The poor performance is attributed to incorrect capture practices in hospitals. Most hospitals in the United States perform foot printing of the babies within 2 hours of their birth using the newborn identification form (shown in Figure 5.2) recommended by the Federal Bureau of Investigation [173]. The footprint of the child and fingerprint of the mother are collected using ink based methods. However, Pela *et al.* [143] reported that an analysis of 1917 footprints had insufficient information for accurate manual identification. Further, the American Academy of Pediatrics and the American College of Obstetricians and Gynecologists stated that [33]: “*individual hospitals may want to continue the practice of foot printing or fingerprinting, but universal use of this practice is no longer recommended*”.

Later, Weingaertner *et al.* [201] and Lemes *et al.* [106] carefully captured footprints and palmprints of 106 newborns at high resolution (≥ 1500 ppi) using specialized sensors to improve useful biometric information content. Images captured from only 20 newborns

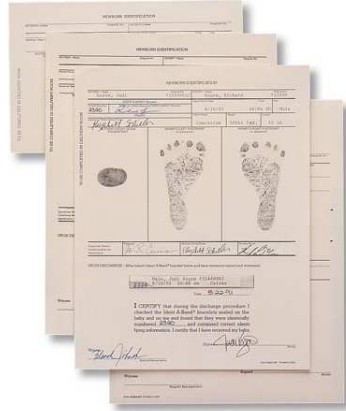


Figure 5.2: Identification form used to capture the footprints of the newborn and fingerprint of the mother. Image is taken from <http://www.amazon.com/Briggs-Newborn-Attachment-Signatures-Verification/dp/B002C18XRW>

(about 5% of the entire captured dataset) deemed to be of useful quality. More large scale experimentation may provide greater confidence towards the use of palmprints and footprints as a biometrics for newborns. However, such biometric modalities, are by their nature, contact based and introduce capture challenges for uncooperative subjects such as newborns. Fields *et al.* [58] and Tiwari *et al.* [179] studied the feasibility of ear recognition for newborns and indicated that ears might provide distinguishable features for newborn recognition. In a preliminary study, Bharadwaj *et al.* [25] suggested proposed an algorithm for recognizing newborns using face images. The algorithm comprised multi-resolution texture representations from three different scales for effective matching in identification problem settings. Recently, Jain *et al.* [86] showed the effectiveness of fingerprints for a larger age group encompassing both newborns and toddlers.

5.1.2 Research Contributions

There are three important contributions of this research:

1. The unique nature and behavior of newborn babies leads to interesting challenges towards a newborn biometric system. Considering the non-intrusive nature of face biometrics, this research explores the possibility of using face recognition for determining the identity of newborns. Inspired from research across multiple domains, we discuss the challenges and opportunities of a newborn face recognition system. Further, a novel learning based representation and matching scheme is proposed that is tailored towards newborns and hence provides state-of-the-art results.

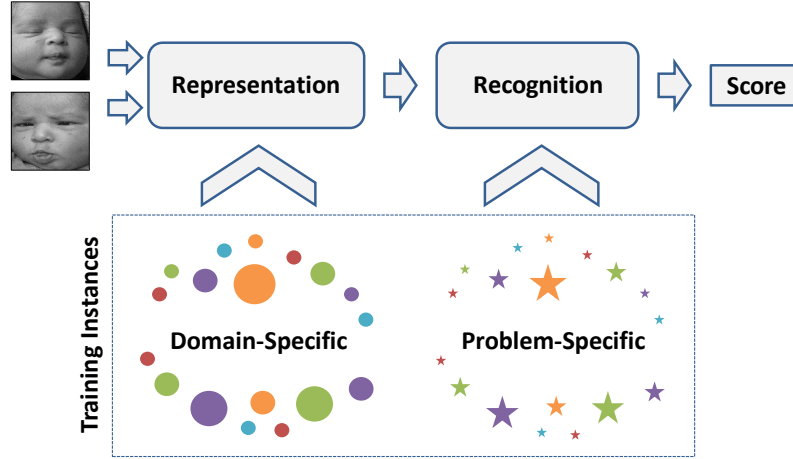


Figure 5.3: This research presents a two-stage learning framework using domain-specific and problem-specific unlabeled instances to perform face recognition in newborns.

2. The asymmetric development of newborn faces results in a unique craniofacial structure that is not proportionally equivalent to a miniature adult face. Further, a newborn baby being a relatively unconstrained user of face recognition presents large variations in pose and expression. In this research we present a two-stage learning framework, illustrated in Figure 5.3.

- The first stage learns a *domain-specific* representation of the human face with a deep learning architecture. Specifically a patch-based encoding scheme constructed from a stacked denoising autoencoder is utilized.
- Next, the learned representation of two input faces are matched with a one-shot similarity metric. Matching visual features obtained from two newborn face images using traditional distance metrics such as χ^2 or $l^2 - norm$ does not account for the *semantic gap* associated with matching, especially in case of the highly *elastic* faces of newborns. This research explores learning based distance metric that learns the *problem-specific* semantic understanding using 1-class-online-SVM, to improve the recognition performance.

3. Newborn face database of 96 newborns is first introduced to the research community. To the best of our knowledge, this is the largest newborns dataset publicly available for research on face recognition in newborns. Further, the performance of both academic algorithms and commercial biometric systems are evaluated across different gallery sizes on a standard benchmark.

¹<http://www.fgnet.rsunit.com/>



Figure 5.4: The craniofacial characteristics of an infant face is not proportionally equivalent to a miniature adult face as observed in the above illustrations. However, several studies indicate that newborn faces do possess discriminative facial features that enable newborn-newborn face recognition, mitigating a common notion that all newborns look alike. Images are taken from the FG-Net database¹.

5.2 Newborn Faces: Characteristics and Database

We first discuss our observations on the characteristics of newborn faces and present the details of newborn face databases along with the challenges it encompasses.

5.2.1 Characteristics of Newborn Face

It is our assertion that face recognition can be a friendly and cost effective solution for identifying newborns if the performance of automatic matching algorithms is satisfactory. As shown in Figure 5.4, newborn faces are structurally different from adult faces, however, the following studies provide evidence to support that they do possess sufficient discriminating characteristics.

- It is a seemingly intuitive notion that newborns look alike and are difficult to differentiate. Kuefner *et al.* [100] conducted a study on 31 adults and demonstrated a statistically significant decrease in recognition accuracy when subjects were presented with newborn faces as compared to adult faces. Interestingly, improved performance was observed when the same experiment was conducted on 18 pre-school teachers who were in contact with children for at least 30 hours per week. The analysis suggests that inexperienced adults may be unable to extract biometric features from newborn faces due to the *other age effect*. Anastasi and Rhodes observed face recognition performance of children and older adults and analyzed that both are good at recognizing own age faces than other age faces [11].

- Every face possess some unique facial traits and subtle differences in shape, proportions of hard and soft tissues, and topographical contours. Human beings are able to perceive these differences with exceptional efficiency. To understand the applicability of face recognition for newborns, it is important to identify those facial characteristics that lead to unique and discriminative features that enable recognition. The facial structure of infants begins to develop prenatally at about 10 to 14 weeks and continues to develop proportionally for a longer duration the farther they are from the *neurocranium* (skull region). For example, the growth of the *mandible* (lower jaw) and chin is slower and continues longer than mid-facial development resulting in newborns often being characterized by large foreheads. Hence, different regions of newborn face provides evidence of identity with varying levels of confidence.
- The nasal region of a face is generally an important point in the facial architecture as the surrounding arches rely on it for support. The region is extensively studied to improve the aesthetics of plastic surgery procedures. The nasal region of newborns is shallow as compared to adults. Hence, they do not possess bold topographical features as compared to adults, shown in Figure 5.4. Further, the craniofacial structure of newborn faces are characterized by prominent eyes, small jaws, puffy cheeks and a high forehead. The eyebrow ridge is fine and the overall proportions are wide and short. These observations indicate that the shape and structure of infants are not miniature adult faces.
- Studies related to facial reconstruction of infant patients (particularly cleft lip patients) [155] show a significant influence of race, gender and age in planning the reconstructive surgery. These observations indicate that *newborn faces also possess unique and characteristic facial features, that can be used for recognition.*

5.2.2 Newborn Face Database and Challenges

Research in newborn face recognition is constrained by the lack of a large database for experimentation. In this research, we have prepared the **IIITD Newborns Face Database**¹ is prepared for experimental evaluation, that consists of 855 face images captured from 96 babies, each having 3 – 10 images taken in multiple sessions from various hospitals. The time of capture varies from one hour to a few weeks after birth. A high resolution camera is used during the capture procedure. Although no constraint is enforced on the babies, an effort is made to capture near frontal faces.

¹The database will be available to researchers at <http://iab-rubric.org/resources.html>



Figure 5.5: Some challenging images from the IIITD Newborns face database indicate the non-cooperative nature of newborns that hinder traditional face recognition approaches.



Figure 5.6: Images from the newborns face database pertaining to two subjects. The figure illustrates the high intra-class variation in the images captured in two sessions posing challenges in detection and recognition.

While capturing the database, we observed some unique challenges with newborn faces that can deter a traditional approaches of face recognition. The primary challenges that make the problem of newborn face recognition unique are discussed here:

- It is not possible to ensure that the subjects present a neutral expression or frontal face due to their unpredictable behavior and elastic faces. Hence, as shown in Figure 5.5, head pose and acute expression variations manifest strongly in the newborn face database. Further, excessive movement during capture occasionally causes motion blurring in the images. Due to these variations, we term newborns as *unintentionally non-cooperative users* of face recognition. Figure 5.6 demonstrates the large variation in images of two newborn babies.
- Another interesting challenge in newborn face recognition is the recognition of new-

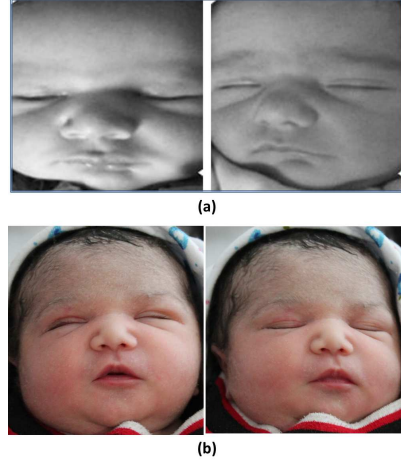


Figure 5.7: Sample images of (a) twins and (b) newborns with *lanugo*, from the Newborn face database.

born identical twins. As shown in Figure 5.7(a), newborn twins (that occur in the prepared database) do not possess the differentiating characteristics that adult twins usually develop with age. However, as with adult twin face recognition [147], research effort in the direction may provide new insights into invariant face features in newborns. In certain cases, shown in Figure 5.7(b), some newborns may be covered with soft hairs, known as *lanugo*, especially on cheek and forehead. These hairs are shed and replaced as the baby ages resulting in variation in image texture.

5.3 Learning based Encoding and Distance Metric Approach

As discussed in the previous section, newborn face recognition has several challenges that arise from their uncooperative nature. Hence, existing hand crafted feature extraction and matching techniques may be unable to perform recognition on par with performance obtained with cooperative adult face images. In such non-ideal conditions, learning based feature extraction and matching techniques are required that explicitly encode the properties of the feature space and hence show improved recognition performance.

5.3.1 Preliminaries

In this subsection, existing learning based representation, stacked autoencoder and one-shot LDA based distance metric are first described. The proposed approach is motivated from these techniques and is presented later.

5.3.1.1 Stacked Denoised Autoencoder

Recent literature has seen several efforts towards learning based rather than hand crafted feature representations to improve the recognition performance. A large number of unlabeled instances that are readily available to train computationally intensive generic representations of face images are utilized to train a variant of the stacked autoencoder [19] with a denoising component, known as a Stacked Denoising Autoencoder (SDAE) [193]. A stacked autoencoder is a neural network where each layer is independently trained (using the backpropagation algorithm) in a greedy manner to reconstruct the output of the previous layer. Specifically, the non-linear mapping function ($f_{W,b}$) of the vectorized input image x is given as follows,

$$y = f_{W,b}(x) = \mathcal{S}(Wx + b) \quad (5.1)$$

where $\mathcal{S}(\cdot)$ presents the sigmoid activation function and the parameters, W and b for the linear component of the mapping function. Next, in a similar manner, a reconstruction step ($g_{W',b'}$) is performed on the lower dimensional mapping (y) as follows,

$$\hat{x} = g_{W',b'}(y) = \mathcal{S}(W'y + b') \quad (5.2)$$

The unsupervised training of such an architecture is performed one layer at a time. Each layer is trained as a denoising auto-encoder by minimizing the reconstruction of its input (which is the output code of the previous layer). The parameters of these mapping functions are computed via back propagation by greedily minimizing the loss function (J) given by,

$$J(W, b) = \|x - \hat{x}\|_F^2 \quad (5.3)$$

Additional constraints may be imposed to ensure that the encoded representation is robust. A sparsity constraint is typically added to the cost function as follows.

$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^s \text{KL}(\rho || \hat{\rho}_j) \quad (5.4)$$

where, ρ is the sparsity parameter, s is the size of the hidden layer, and $\text{KL}(\cdot)$ is the KL-Divergence metric that penalizes the response of a neuron that deviates from the sparsity constraint, i.e.,

$$\text{KL}(\rho || \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (5.5)$$

Further to enhance the robustness of the representation, at each stage the input data is corrupted before reconstruction step. In this case, by obfuscating random features ensures that the representation is generalizable and this architecture is known as denoising autoencoders. In literature, the unsupervised feed-forward learning of the stacked denoising autoencoders is generally followed by a fine tuning step or a classification step, where a multilayer perceptron is trained in a supervised fashion on specific instances to minimize the classification accuracy.

5.3.1.2 Learning based Distance Metrics: One Shot Similarity using LDA

A metric function is defined as a positive definite distance measure between two elements in a set. Formally, for a set ω , a function $d : \omega \times \omega \rightarrow [0, 1]$ represents a distance function that can be used to measure dissimilarity between two elements in ω . Unlike traditional distance metric functions such as $l^2 - norm$ and χ^2 , learning based metrics argue that given a set of discrete labeled (as 0 or 1) pairs of instances from ω , the distance function can be learned by posing it as a classification task. That is, given a set of pairs of points labeled as similar or dissimilar, a classifier margin can be learned and used as a distance metric. Hence, the semantic knowledge obtained from this annotation can be preserved from within the distance metric. Several learning based distance metrics have been proposed in literature [208]. In this research, we consider One-Shot Similarity (OSS), a semi-supervised technique that uses unlabeled training data as a set of negative constraints against which two input samples are matched.

Wolf *et al.* [202] describe the use of a semi-supervised distance metric learning over a variant of local binary pattern to improve face recognition performance in unconstrained face image data. One-shot similarity measures the dissimilarity between a given instance and a separate class of negative instances available during matching, as illustrated in Figure 5.8. Specifically, given a pair of instances to match, say A and B , an instance specific classifier is trained with A as the single positive sample and a separate set I (which is (not A) and (not B) in Figure 5.8) of negative samples. A similarity index between A and set I is computed (S_1). Similarly, a similarity index is trained and computed between B and the same set I (S_2). The OSS similarity measure is defined as a combination of S_1 and S_2 (by summation). Here, the formulation based on Linear Discriminant Analysis is used as the classifier.

The availability of only one element in the positive class restricts the number of possible classifiers that can be used. Wolf *et al.* [202] used Linear Discriminant Analysis (LDA) as it provides the advantage of pre-computing the within-class covariance matrix. Let $P_i \in \mathbb{R}^d$,

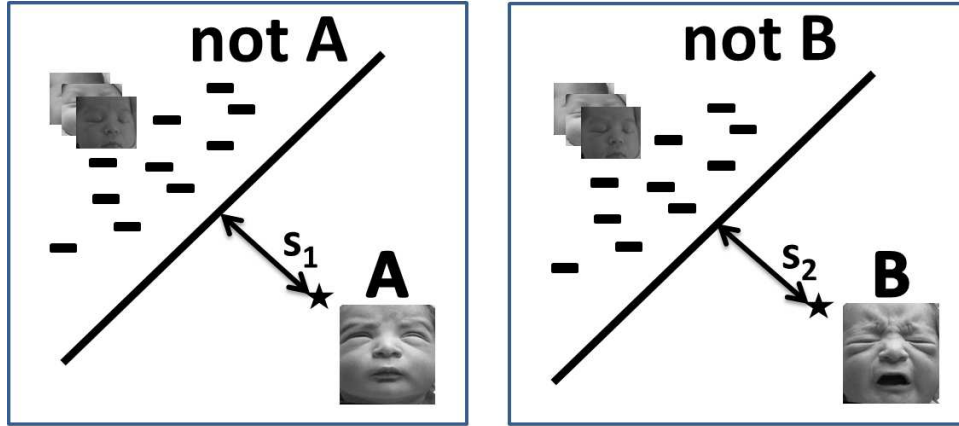


Figure 5.8: One Shot Similarly: A and B are the two instances to be matched. The distance between them is computed using the classifier output obtained using each instance and the same negative background set I .

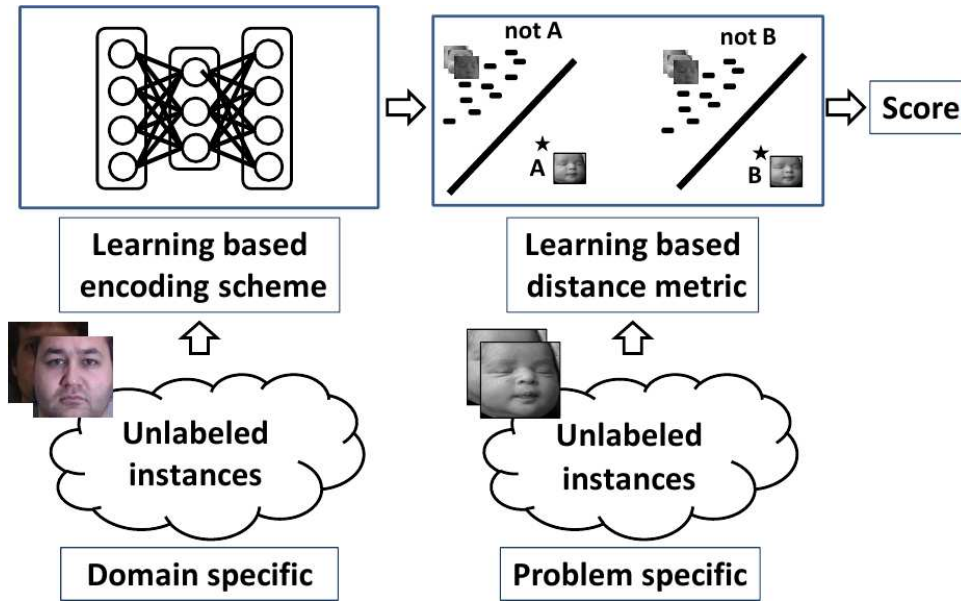


Figure 5.9: The training phase, is performed by effectively utilizing abundant domain-specific instances and limited problem-specific instances to learn a representation and distance metric.

where $i = 1, 2 \dots m_1$ be the set of positive samples, and $N_i \in \mathbb{R}^d$, where $i = 1, 2 \dots m_2$ be the set of negative samples with means μ_p and μ_n respectively. The mean of all the points is given by μ . The between-class scatter matrix (S_B) and within-class scatter matrix (S_W)

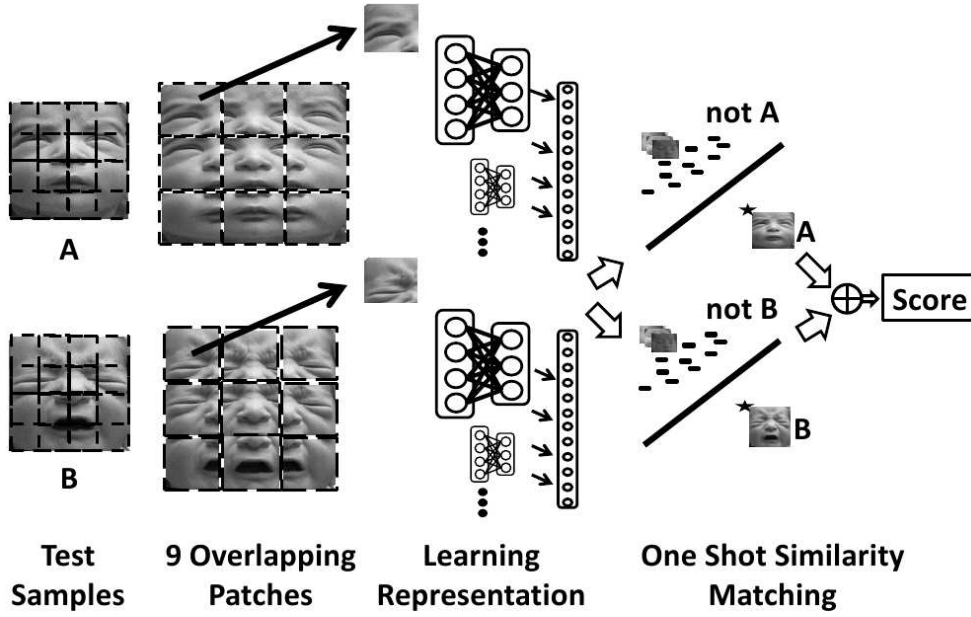


Figure 5.10: The testing phase of the proposed approach where each input pair to be matched is converted into previously learned representation with the separate encoder learnt for each patch. Next the representations are matched with the online SVM based one-shot similarity metric approach.

are defined as,

$$S_B = (\mu_p - \mu_n)(\mu_p - \mu_n)^\top \quad (5.6)$$

$$S_W = \frac{1}{m_1 + m_2} \sum_{i=1}^{m_1} (P_i - \mu_p)(P_i - \mu_p)^\top + \frac{1}{m_1 + m_2} \sum_{i=1}^{m_2} (N_i - \mu_n)(N_i - \mu_n)^\top \quad (5.7)$$

LDA finds a projection v that maximizes the Rayleigh equation given as,

$$v = \arg \max_v \left(\frac{v^\top S_B v}{v^\top S_W v} \right) \quad (5.8)$$

In this case, P contains a single sample (either A or B) and N is a separate set of background examples (I in the example). Thus, the positive samples do not contribute to S_W and S_B . Further, S_W is constant for the background set (I) and can be pre-computed. For a single positive sample x , a signed score is computed as $v^\top x - v_0$, where v_0 is the bias defined as $v_0 = v^\top \frac{x + \mu_n}{2}$. While it is difficult to obtain a large labeled training set of

newborn babies, unlabeled negative instances are easier to obtain. The similarity measure computed from these *background* samples, via OSS metric, can improve the matching performance.

5.3.2 Proposed Approach

The proposed two-stage learning framework for newborn face recognition, illustrated in Figure 5.9, first learns a *domain-specific* representation of the face with variable expressions using a stacked denoising autoencoder. A large number of unlabeled face samples can be leveraged to learn the representation. The advantage of utilizing SDAE for representing a face is the data driven nature of the encoding scheme. The denoising and sparsity constraint enable the SDAE to be robust to changes in the intensities of an image due to pose, expression and illumination variations. Further, the learned representation is coupled with a *problem-specific* learned distance metric that benefits from the availability of additional unlabeled problem-specific (background) samples to reduce the semantic gap during matching thereby improving newborn face recognition performance.

5.3.2.1 Domain Specific Representation via SDAE

As illustrated in Figure 5.10, the input face image is first tessellated into 9 overlapping patches to capture the uneven development of the craniofacial structure of the newborns. Overlapping blocks were chosen to offset the effects of pose variation, a dominant covariate in the database. Separate sets of multi-layer encoders are learned for each overlapping patch of a face image, that helps enhance the depth of the encoders. The patch encoding ensures spatial coherence of the resultant representation. Each encoder provides a representation of a component of the face image, such as forehead, periocular, mouth, and chin regions. The representation thus obtained is concatenated into a single feature vector. The optimal configuration of the encoder (SDAE) is determined experimentally as [2500|1000|500]. SDAE is trained on a large number of domain-specific samples, i.e., face images with varying illumination and expression. We hypothesize that the SDAE learns patterns that are adequate representation of face images. However, the *special* constraints of newborn faces must also be augmented in the recognition framework. Next, this research shows that problem-specific fine-tuning (classification) step can be performed with a learning based distance metric to match the aforementioned learned representations of two input face images in a meaningful distance space.

5.3.2.2 Problem Specific Distance Metric Learning via One Shot Similarity with 1-class-online-SVM

Low-level feature descriptors of images suffer from a *semantic gap effect* when dealing with matching for a specific application. This gap arises due to the difficulty in mapping features to meaningful interpretations, such as identity of a subject. Recent approaches to face recognition propose learning feature relevance using explicit training samples. In this research, distance metric learning technique is explored to reduce the semantic gap effect in newborn face representations. Using training samples from unseen classes (referred to as *problem-specific* samples), the structure of the newborns face manifold can be analyzed. Further, the best metric that is suitable to the classification task for faces of that particular manifold can be learned and then utilized for matching. Specifically, for newborn face recognition, a distance metric that learns the mapping space in which two newborn faces can be best matched, can improve the recognition performance.

- Rather than viewing the feature manifold as a linear space, we propose to discriminate between two samples in the support vector space (kernelized feature space). This ensures a robust non-linear and high dimensional representation of the distance space and may provide a generalizable solution. A soft margin Support Vector Machine (SVM) learned from training samples $\{(x_i, y_i), x_i \in IR^m, y_i \in \{-1, 1\}, \forall i \in \{1, \dots, N\}\}$ that is of the form $f(x) = w \cdot \Phi_k(x_i) + b$ can be summarized as follows,

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (5.9)$$

subject to the constraints of correct classification given by,

$$y_i(w \cdot \Phi_k(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i \in \{1, \dots, N\} \quad (5.10)$$

The learning formulation of SVM, popularly solved with quadratic programming, is expressed in its dual form.

$$\min_{0 \leq \alpha_i \leq C} \mathcal{W} = \frac{1}{2} \sum_{i,j=1}^N \alpha_i Q_{ij} \alpha_j - \sum_{i=1}^N \alpha_i + b \sum_{i=1}^N y_i \alpha_i \quad (5.11)$$

where b is the offset and $Q_{ij} = y_i y_j \phi(x_i) \cdot \phi(x_j)$ is the Lagrange multiplier that results in the dual form solution $f(x) = \sum_{i=1}^N y_i \alpha_i \phi(x_i) \cdot \phi(x) + b$. The *Karush-Kuhn-Tucker* (KKT) conditions uniquely define the solution of the parameters (α, b) by minimizing

the dual form and determining support vectors from the training sample set. An incremental formulation and parameter perturbation of the SVM proposed by Diehl and Cauwenberghs [52] partitions every training sample encountered, based on the partial derivative $g_i = \frac{\partial \mathcal{W}}{\partial \alpha_i}$, into 3 groups: margin support vectors ($g_i = 0$), error vectors ($g_i < 0$) and reserve vectors ($g_i > 0$). The perturbation and incremental update process involves the addition of new training examples into the three groups while simultaneously maintaining the KKT conditions for all the previously viewed samples. This is achieved by varying the α coefficients in a sequence of *adiabatic* steps of small incremental values determined based on the membership of the groups.

- In this research, we first extend the incremental/decremental (online) learning paradigm to a 1-class Support Vector Machine ($SVM_{online}^{1-class}$) using the Schölkopf *et al.* [165] formulation of one-class SVM. The training samples from the background set (problem-specific data) are labeled as one class and the maximally separating hyperplane from the origin is determined. This results in a non-linear mapping function that captures the probability density of the training data in the input feature space. The 1-class Online Support Vector Machine ($SVM_{online}^{1-class}$) is then utilized to formulate the one-shot similarity metric. Similar to the LDA formulation described previously, the proposed approach can leverage the non-linearity of a kernel SVM function that improves the performance of the OSS distance metric.
- During the training phase, a grid search is performed to estimate the model parameters C and kernel parameter k , using a K-fold cross validation of the training set (problem specific samples, I). The distance metric is defined as the measure of the distance from hyperplane for each point in consideration. To maintain the computational complexity of the proposed approach, the SVM must be pre-trained on the background samples before hand. The pre-training is modeled as a one-class SVM learning problem where samples are labeled as a single class. Next, the one-class SVM is converted to a two-class SVM by adiabatically adjusting the boundary based on a newly introduced sample (say ‘A’) with a positive label. The SVM function then converts to a binary classifier with the distance from the hyperplane indicating the confidence of the classification. The one-shot matching approach using online SVM for newborn verification is presented in Algorithm 3.

Algorithm 3 Matching two newborn faces with SVM_{online} based one-shot similarity.

Input: Two newborn face image representations A and B , trained model of $SVM_{online}^{1-class}$ with problem-specific background samples I , and decision threshold T .

$SVM_{online}^{A,I}$ = Online update SVM model with A as positive sample

S_1 = distance from decision boundary of B from $SVM_{online}^{A,I}$

$SVM_{online}^{B,I}$ = Online update SVM model with B as positive sample

S_2 = distance from decision boundary of A from $SVM_{online}^{B,I}$

$S = S_1 + S_2$

Output: if $(S > thr)$ report “genuine” else “imposter”

5.4 Performance Evaluation and Analysis

The first step in recognizing a face image is face detection. However, due to the challenges discussed earlier, detecting face regions from captured images of newborns is an error prone task. Existing techniques such as Viola-Jones [194] and Active Appearance Models (AAMs) [44] have failed in reliably segmenting face regions despite explicit training. Therefore, all the images used in this research are cropped using manually annotated eye and mouth locations. The cropped images are corrected for in-plane rotation, resized to 200×200 , and pre-processed by histogram normalization. Owing to the challenging nature of the problem, the experiment protocol is divided into two parts, domain-specific and problem-specific.

- **Domain-specific information:** In order to train the SDAE to learn the encoding of facial features, 50,282 frontal face images pertaining to 346 subjects from the Multi-PIE database [72] are utilized. The face region is detected from each image using a commercial face detector, cropped and resized to 200×200 .
- **Problem-specific information:** The problem specific database is the newborns face database which consists of multiple images of 96 newborns. Additionally, an *extended dataset* of 358 newborns with single sample per subject is collected from various sources while ensuring similar image resolution. A majority of the images (298) are obtained from the Face Tracer dataset [104], a publicly available database collected by manually labeling images downloaded from the internet. Only those samples that are manually marked as ‘babies’ are utilized. The remaining 60 samples are collected from various sources by the authors. The extended dataset are unlabeled single sample images that are described as background images for problem specific learning, to enable matching.

5.4.1 Experimental Protocol

Along with evaluating the effectiveness of the proposed algorithm, we have also evaluated the results of several hand-crafted features, learnt representations, and commercial systems. The hand-crafted features used for comparison are:

- Local Binary Patterns (LBP) [137],
- Dense Scale Invariant Feature Transform (SIFT) [113], and
- Speeded Up Robust Features (SURF) [16].

The learnt representations are,

- Principal Component Analysis (PCA) [18],
- Linear Discriminant Analysis (LDA) [18], and
- Vector of Locally Aggregated Descriptors (VLAD) [193].

The commercial systems used for comparison are:

- Verilook (termed as COTS-1) and
- FaceVACS (termed as COTS-2).

Three different experiments are performed to evaluate the effectiveness of the proposed algorithm and compare with existing algorithms in multiple real world settings.

1. *Experiment 1 - Benchmarking Existing Feature Extractors:* We first benchmark existing face recognition techniques on the newborn face database. 10 subjects ($\approx 10\%$ of the 96) from the database are randomly selected for training and the remaining images (corresponding to 86 subjects) are used for testing. Five times random sub-sampling is performed for all experiments to seek generality.
2. *Experiment 2 - Proposed Algorithm:* The experimental protocol used to evaluate the effectiveness of the proposed algorithm in identification and verification settings is same as experiment 1. Images of 10 newborns are randomly selected for training and the remaining images corresponding to 86 newborns are used for testing with 1, 2, 3, and 4 images per subject in the gallery. The least distance score obtained per subject is used as the match score. Further, all the results are reported with five times random sub-sampling. The proposed learning based algorithm requires

additional background samples for training. Hence, the additional 358 samples of newborns are utilized for training as background samples. In addition to variations of the proposed approach, comparison with SIFT and VLAD [193] representations is also performed.

3. *Experiment 3 - Effect of Incorporating Problem Specific Information:* This experiment showcases the impact of using problem-specific information for training. The experiment evaluates the proposed online SVM based OSS distance metric by substituting the unlabeled newborn samples utilized for training the SVM, with a set of 1000 samples randomly chosen from the cropped multi-PIE dataset [72].

Table 5.2: Rank-1 identification accuracies (%) for experiment 1. The results are reported in terms of average accuracy with standard deviation over times cross validation. *Commercial systems rejected a portion of the samples based on quality.

Algorithms		Number of images per person in gallery			
		1	2	3	4
Existing handcrafted features	LBP + χ^2	21.1 (4.6)	32.5 (5.6)	39.5 (5.2)	44.7 (2.2)
	SIFT + χ^2	31.3 (4.8)	41.9 (3.6)	48.0 (3.07)	53.4 (2.7)
	SURF + χ^2	66.5 (1.8)	79.2 (1.1)	84.6 (1.4)	87.3 (1.5)
Appearance based features	LDA + L_2	–	5.1 (3.0)	6.2 (3.8)	8.9 (2.7)
Existing learnt features	VLAD+OSS (LDA)	13.7 (1.4)	21.0 (2.1)	26.6 (1.7)	31.4 (1.7)
	VLAD+OSS ($SV M_{online}$)	13.7 (0.9)	22.3 (2.5)	29.3 (2.6)	34.2 (2.0)
Commercial systems	COTS-1*	24.0 (2.1)	32.1 (1.6)	33.5 (1.3)	34.3 (1.4)
	COTS-2*	41.0 (1.6)	53.6 (2.3)	60.1 (2.2)	64.6 (2.2)
Proposed	SDAE+OSS (LDA)	43.7 (3.8)	57.7 (3.2)	66.0 (2.3)	72.6 (1.6)
	SDAE+OSS ($SV M_{online}$)	51.1 (2.5)	66.0 (2.4)	73.1 (1.7)	78.5 (1.7)

5.4.2 Results of Experiment 1: Benchmarking Existing Feature Extractors

This experiment focuses on benchmarking some of the existing handcrafted and learnt features along with the two commercial systems. Since a newborn face recognition system can be used in both identification and verification applications, all the experiments are performed in both the scenarios. The identification performance is presented in Table 5.2 in terms of rank-1 identification accuracy and the verification performance in terms of genuine accept rate (GAR) at 0.1% false accept rate (FAR) is presented in Table 5.3.

- We observe that the performance of texture based approaches (LBP and SIFT) is superior than appearance based techniques (PCA and LDA). For instance, LBP yields the rank-1 accuracy of 21.1% and GAR of 16.7% at 0.1% FAR, compared to PCA which yields 4.9% and 2.4% with a single gallery images per subject. We have also evaluated the results with multiple images in the gallery. For every test image,

Table 5.3: Verification accuracy (at 0.1%FAR) and standard deviation on the newborn face database for experiment 1. *Commercial systems rejected a portion of the samples based on quality.

Algorithms		Number of images per person in gallery			
		1	2	3	4
Existing handcrafted features	LBP + χ^2	16.7 (0.8)	20.4 (0.9)	22.8 (1.4)	24.0 (1.1)
	SIFT + χ^2	15.7 (2.6)	22.8 (2.8)	26.1 (1.4)	32.3 (1.6)
	SURF + χ^2	0.2 (0.2)	0.4 (0.1)	0.9 (0.3)	0.2 (0.2)
Appearance based features	LDA + L_2	–	15.9 (2.7)	16.9 (2.5)	25.6 (2.9)
Existing learnt features	VLAD+OSS (LDA)	6.1 (3.2)	9.9 (4.5)	13.5 (3.4)	16.1 (4.0)
	VLAD+OSS (SVM_{online})	10.0 (1.8)	14.9 (2.9)	21.5 (2.8)	25.9 (1.3)
Commercial systems	COTS-1*	7.1 (2.1)	11.6 (1.3)	14.3 (1.7)	18.2 (1.4)
	COTS-2*	10.4 (0.9)	16.1 (1.6)	23.1 (1.7)	31.6 (2.2)
Proposed	SDAE+OSS (LDA)	25.3 (4.4)	38.3 (5.5)	46.8 (2.9)	53.0 (2.7)
	SDAE+OSS (SVM_{online})	31.8 (5.7)	45.8 (4.7)	55.8 (2.6)	63.4 (2.0)

the least distance score for every subject is utilized for evaluation. The results show that even with multiple gallery images, the accuracies improve but the performance trend remains the same. The superior performance of texture based approaches can be attributed to spatial collation in regional blocks that is able to better deal with the covariates such as pose, expression and illumination.

- It is interesting to note that SURF yields the highest accuracies in identification and the lowest accuracies in verification. Since SURF is a key point based feature extractor, we have analyzed the key-points detected by SURF. As shown in Figure 5.11, we have observed that the key-points are unreliable both in genuine and impostor pairs. The unstable edge information in newborns faces leads to poor spatial coherence and thereby inconsistent key-point detection. Similar observation is made for key-point detection based SIFT, hence, fixed key-points in a grid of size 8×8 are used for SIFT experiments. With four gallery images, SIFT yields rank-1 identification accuracy of 53.4% and a GAR of 32.3% at 0.1% FAR.
- VLAD is a learning based descriptor which extracts visual words based on clustering the training data in an unsupervised manner. The large variations in pose and expression of newborns may cause inconsistency in the word dictionary. As shown in Table 5.2, the best recognition rates for this representation is obtained with OSS (SVM_{online}) when the gallery consists of four images per subject, with rank-1 accuracy of 34.2% and verification of 25.9% at 0.1% FAR.
- The performance of both the commercial systems (COTS-1 and COTS-2) is low, with a rank-1 identification accuracies of 34.3% and 64.6%, and verification accuracies of 18.2% and 31.6% at 0.1% FAR respectively. The systems are presumably trained



Figure 5.11: The performance of key-point detection based techniques such SURF is misleading. The detected keypoints are unreliable both in genuine and imposter pairs. This may be attributed to unstable edge information presented due to face wrinkles.

for adult face recognition and do not adapt to newborn faces accurately. Further, a portion of the images are also rejected by the system due to minimum quality thresholds imposed by COTS.

- While handcrafted features show better performance than learnt features, handcrafted features have limited representation capacity as they are not designed for the specific problem domain. Based on the observations made and the challenges of newborn faces, we assert that automatic face recognition techniques that are tailored specifically for newborns, via explicit training, may be able to identify newborns more effectively and can provide a friendly as well as cost effective solution. Therefore, it is our assertion that the proposed learning based encoding and distance metric algorithm that *learns* a suitable feature representation and a distance metric that captures the semantic understanding of the encoding scheme should improve the recognition performance.

5.4.3 Experiment 2: Results of the Proposed Algorithm

After demonstrating the results of existing features for recognizing newborn face images, we now present the results of the proposed algorithm represented as $SDAE + OSS(SVM_{Online})$.

- The results in both identification and verification scenarios are summarized in Tables 5.2 and 5.3 respectively. Compared to existing features and commercial systems, the proposed algorithm yields improved verification and identification accuracies.

- The proposed algorithm has two components: learning a robust feature representation and learning the distance metric. To evaluate the effectiveness of both the components, we strategically replaced one component at a time with existing descriptors or matchers and compared the results with four gallery images per person. For feature representation, VLAD and DSIFT are used. For the matching component, VLAD and DSIFT are used. For the matching component,

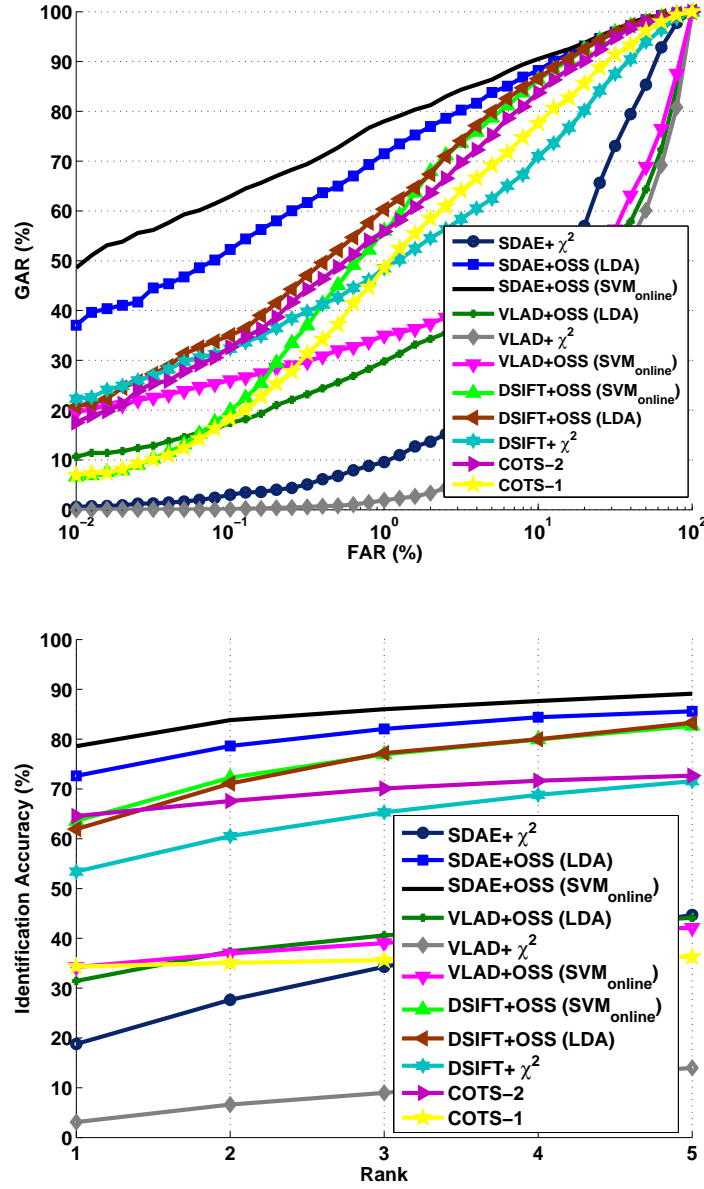


Figure 5.12: Experiment 2: ROC and CMC curves to evaluate the effectiveness of individual components of the proposed algorithms. The experiments are performed with four gallery images per subject. The graphs are best viewed in color.

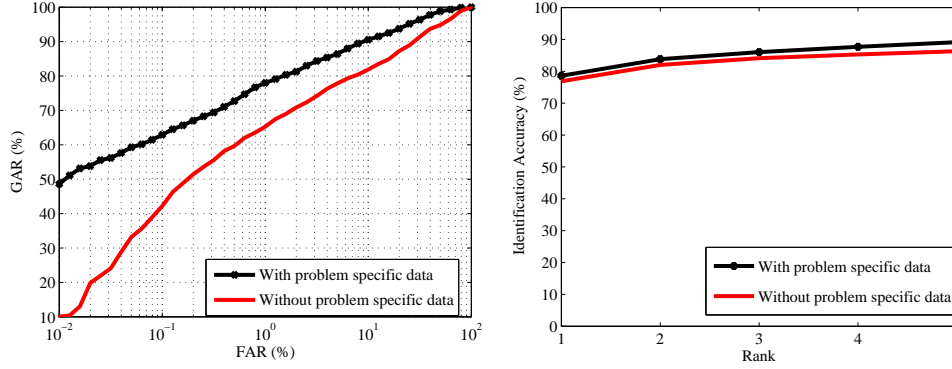


Figure 5.13: Experiment 3: ROC and CMC curves of the proposed algorithm ($SDAE + OSS(SVM_{online})$) to evaluate the effectiveness of using problem specific data. The results are computed with 1000 adult face images as background samples. The graphs are best viewed in color.

OSS with LDA and χ^2 distance metrics were used. The results of this evaluation are illustrated in Figure 5.12.

- With SDAE, OSS (LDA) metric achieves the rank-1 identification accuracy of 72.6% and verification accuracy of 53.0% at 0.1% FAR. Further, with the proposed SVM_{online} based OSS formulation, SDAE yields 78.5% rank-1 identification and 63.4% GAR at 0.1% FAR; significantly surpassing other approaches in both forms of matching. The results for different gallery sizes for each of the mentioned techniques are presented in Table 5.2. It can be observed that with increasing the number of gallery images per subject from one to four, the verification performance of the proposed SDAE+OSS (SVM_{online}) approach improves from $31.8 \pm 5.7\%$ to $63.4 \pm 2.0\%$ and for identification setting from $51.1 \pm 2.5\%$ to $78.5 \pm 1.7\%$. Further, the variance in accuracy across the cross-validation folds also reduces.
- Figure 5.12 and Figure 5.13 show the effectiveness of using the deep learning neural architecture over the clustering based VLAD representation [193] with χ^2 distance, LDA and SVM_{online} based learning distance metrics, in both verification and identification settings. Similarly, the proposed approach outperforms the handcrafted DSIFT feature as well.
- The average computational time of the proposed approach for feature extraction and matching a pair of images is 1.87 seconds with a MATLAB implementation on a standard i5 processor based desktop with 8 GB RAM. The training time of the proposed algorithm is initially high, due to the training of the encoding scheme and

the distance metric learning. However, the online nature of distance metric results in lower computational time of the matching stage.

5.4.4 Experiment 3: Effectiveness of Problem Specific Information

- Figure 5.13 showcases the effect of incorporating problem-specific information in the proposed approach. For this experiment, the background samples of the one-shot SVM similarity used for training are substituted with a subset of 1000 samples, randomly chosen from the multi-PIE dataset [72]. On removing the problem specific data, the verification accuracy reduces from 63.4% to 41.7% and the rank-1 identification accuracy reduces from 78.5% to 76.8%. This shows that problem-specific background set improves the performance of the proposed approach for both identification and verification scenarios.
- The decrease in performance when a random set of adult faces is utilized to train the learned distance metric in comparison with the problem-specific newborn faces, can be attributed to the corresponding distance subspace being learned. The distance subspace obtained from problem-specific samples provides a more relevant feature weighting that minimizes the intra-class variation and maximizes the inter-class variation of the feature representation. It must be noted that the impact of the problem-specific background training samples is higher for the verification experiment.

5.5 Summary

We present a learning based encoding and distance metric approach to the novel problem of newborn face recognition. The proposed approach combines deep learning based feature encoding scheme with a learning based distance metric to improve the performance of face recognition. The deep learning encoding approach learns a domain-specific representation of face image utilizing the large number of unlabeled samples available. Next, a one-shot similarity distance metric is learned using relatively small amount of problem-specific information (newborn face images) for effective recognition. On a newborn face database collected from various hospitals under challenging settings, the results show that the proposed face recognition approach yields improved performance compared to existing algorithms and commercial systems.

Further improvement in recognition performance can be achieved by combining newborn face recognition with auxiliary soft information pertaining to newborns such as

height, weight, blood type and gender [180]. However, availability of such information outside health facility or hospital settings may not be as straight forward. In future, we plan to extend the database size to perform a more large scale evaluation of both existing and novel techniques in face recognition. The proposed technique has scope for development on two accounts. Firstly, the accuracy of the system does not still match commercial face recognition standards for large adult face datasets (even though it is rare to have more than 50 babies in a single hospital unit). Secondly, segmentation of region of interest in newborn face images is a challenge. Moreover, like any recognition system, there is scope for tampering at enrollment phase. Given the high motivation levels of perpetrators, this challenge is important and is still an open research problem.

Chapter 6

Conclusions and Future Work

In biometrics, evidence fusion paradigms has been widely used to establish the identity of an individual with greater confidence, particularly when the system is intended to be deployed in adverse environmental conditions and noisy data. In order to further enhance the versatility of a biometric system, additional information that is readily available in various deployment scenarios can act as ‘situational cues’ to dynamically adapt or fine-tune recognition. This thesis introduces the advantages of leveraging auxiliary sources of information to enhance the performance of biometric systems. As illustrated in Figure 6.1, this thesis considers four auxiliary information sources: *biometric* quality, reliability and context, social networks, and unlabeled images obtained from the web, to enhance the capabilities of biometric systems that are intended for deployment in three non-ideal conditions: biometric recognition of uncooperative users, face identity tagging in social network photos, and newborns face recognition.

The first source of auxiliary information, the biometric quality of the sample measures the potential of the sample for good recognition performance, unlike traditional image quality metrics which measure aspects of an image important for good visual perception. This thesis first presents an understanding of *quality assessment* in biometrics. It is our assertion that biometric quality metrics are an important ingredient to improve the robustness of large scale real-world biometric systems. A novel learning based face quality assessment metric that utilizes holistic super-ordinate representations of the face is also presented. Further, this research couples the quality assessment metrics with the pre-processing stage of the biometric pipeline to improve the effectiveness of denoising techniques in a quality assessment based enhancement framework.

The reliability and context of every biometric samples encountered by a multibiometric system is the second source of auxiliary information that is leveraged to enhance the

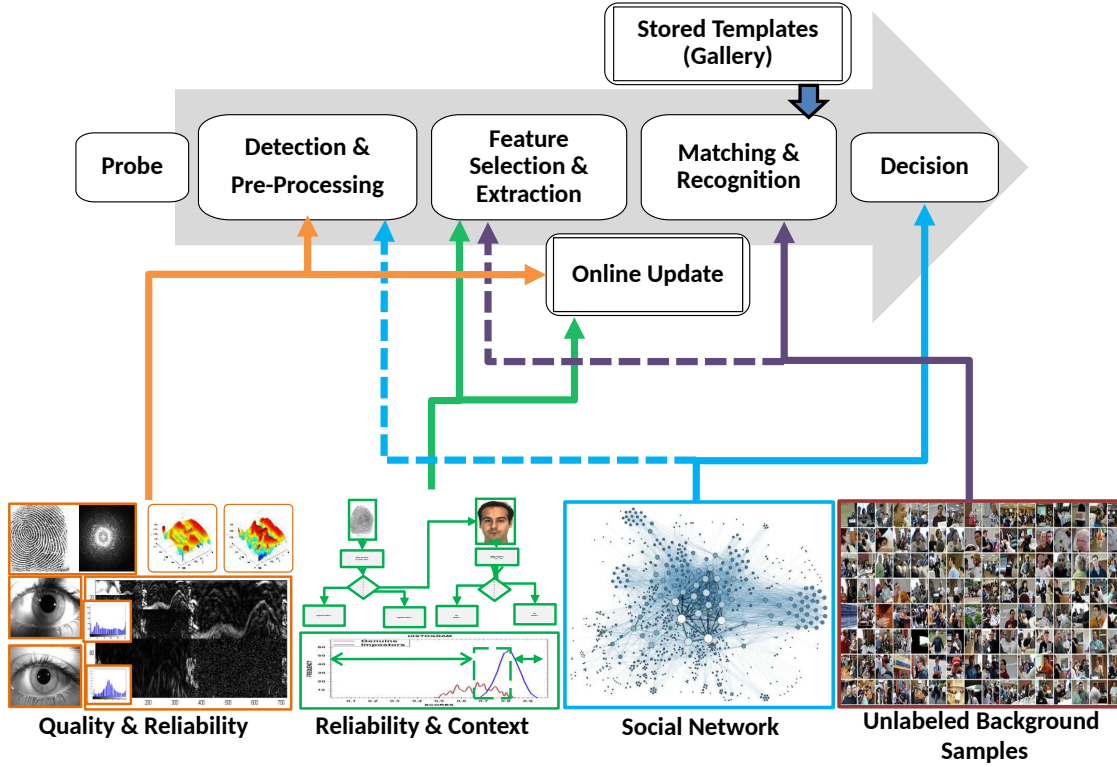


Figure 6.1: This thesis establishes the advantages of leveraging auxiliary sources of information to enhance the performance and reliability of a biometric system. This figure illustrates the connections made in this work between the biometric recognition pipeline with four auxiliary sources, namely, biometric quality, reliability and context, social networks, and unlabeled background images obtained from the web. The jagged lines show future scope of utilizing an auxiliary source.

performance and scalability in challenging conditions and variations in data distribution. The proposed framework, termed as *QFuse*, uses the quality of input images to evaluate the sample’s reliability. The framework dynamically changes context to the best biometric matcher or fusion algorithm to verify the identity of an individual. The experimental results show that the proposed algorithm optimizes the accuracy and computation time for large scale challenging applications.

Often the identity of a person in a photo can be inferred based on the identity of other persons in the same photo, when some social-context between them is known. This research leverages social context obtained from popular social networking platforms as the third source of auxiliary information to improve face recognition and counter the challenging imaging conditions. A framework for person identification is presented, that can

exploit co-occurrence of individuals to improve face recognition. Association rule mining is used to infer multi-level social-context between subjects from a large repository of social transactions.

As the final contribution, this research presents a learning based encoding and distance metric approach to the problem of newborn face recognition. The proposed approach first utilizes deep learning based encoding and learns a domain-specific representation of face image utilizing the large number of unlabeled background samples available from the web. Next, a one-shot similarity distance metric is learned using relatively small amount of problem-specific background information (unlabeled newborn face images) for effective recognition. On a newborn face database collected from various hospitals under challenging settings, the results show that the proposed face recognition approach yields improved performance compared to existing algorithms and commercial systems.

The key contributions of this thesis showcase algorithms that improve biometric systems performance and reliability by effectively utilizing auxiliary information present within unstructured data. While the thesis showcases substantial improvement to performance of biometric systems, there are several open possibilities to more tightly couple auxiliary data to the biometrics pipeline and also to utilize other existing sources of auxiliary information. Some important extensions are described next as future work.

- Image based biometric systems capture multiple images (or a video) of the presented biometric modality. In traditional biometric system, each frame is evaluated by the confidence of segmentation, and/or by simple quality metrics. Once a desired frame is captured, the other images are discarded. We assert that such (discarded) images are another auxiliary source of information that can be effectively utilized in unconstrained deployment of biometric systems. For example, our research on periocular biometrics [24] can be utilized to leverage the images of periocular region captured by iris scanners while attempting segmentation in a context switching approach. Further, the consecutive video frames obtained from the camera that are otherwise discarded, can be utilized for anti-spoofing detection by encoding distinct flow and texture signature of real frames [26].
- The development of quality assessment algorithms of biometric samples that are computationally inexpensive to compute yet correctly encode quality will be the *sine qua non* of real-world large-scale deployments. Using quality assessment metric cannot, however, be a panacea for the recognition of poor quality images. Beveridge

et al. [145] place a bound on the extent to which quality metrics can improve the performance of matching systems when they are used as performance predictors.

- This research presents a framework to augment face recognition with multi-level social context using association rule mining techniques in a supervised settings. An important constraint is the large computational complexity and exponential growth in processing requirement with large number of social transactions. To ensure a tractable solution, the approach prunes low support and confidence associations during rule generation phase. Rule generation can be improved further by online rule generation and caching [192]. The performance of association rule mining can also be improved with several engineering enhancements, e.g., the algorithm is also shown to be parallelizable [5]. Our current work incorporates some of these derivatives and relates to extending this notion of context to other interesting applications including video surveillance. As suggested in Figure 6.1, cues obtained from social media that are indicative of quality of the photo such as time of day, comments on image content, tagged users and preferences can also be leveraged in context based enhancement of the face region, and online update.
- Finally, the accuracy of the newborns face recognition system still does not match the performance of commercial adults face recognition systems. Further, segmentation of region of interest in newborn face images is a challenging problem. Newborn face recognition systems that are typically deployed in neonatal care centers of medical hospital collect auxiliary soft information pertaining to newborns such as height, weight, blood type and gender that can be leveraged to improve recognition performance [180].

Appendix

Appendix A

Image Quality Assessment

Quality is an attribute or a property of an item that quantitatively measures specific traits. The word has several connotations in business, science and philosophy. This work aims to define and demystify the meaning and understanding of quality in the field of biometrics. Further, we investigate its applicability in face recognition, an area that is yet to receive proportionate attention from the research community. First, a review of research in the image quality is first presented.

The assessment of the quality of an *image* is important to measure and control its degradation during acquisition, compression, transmission, processing and reproduction [197]. Several quality assessment algorithms exist in image processing literature, which pursue different philosophies, performance, and applications. A majority of these methods are motivated towards accurate *perceptual* image quality i.e. quality as perceived by the sophisticated human visual system (HVS). Two distinct approaches exist in literature to model the HVS, a *bottom-up* and a *top-down* approach [197]. The first approach is based on the replication of various mechanisms of the HVS which entails a deep understanding of its anatomy and psychophysical features. Many are categorized and summarized by Wang and Bovik [197]. The second approach treats the performance of the HVS as a black box, dealing with only the input to and output from the HVS. Both approaches are important; however optimized solutions often lie in a middle ground of both approaches to this problem.

Depending on the amount of information required, quality assessment algorithms can be segregated as *full-reference* (FR), *no-reference* (NR), and *reduced-reference* (RR) quality assessment. A detailed discussion of each of these categories is presented next.

1. **Full-Reference** (FR): This category of algorithms require a distortion free or perfect quality version of the same image, the ‘original image’, in order to assess the quality of the input images. These approaches perhaps have received most interest from the community due to wide applicability in areas of quality of service (QoS) in delivery of image based content. Most FR *bottom-up* quality assessment methods share a similar framework known as the *error-visibility* paradigm [197]. The strength

of error computed between the given image and the original (reference) image are weighted based on known features of the HVS. This ensures that the quality metric validates those errors which have the maximum affect on human perception. A generic error-visibility based quality assessment framework consists of four phases discussed below (as shown in Figure A.1).

- (a) *Preprocessing*: The input reference and distorted image undergo a preprocessing stage, usually comprising of spatial registration, color space transform (to YCbCr), and filtering. It is assumed that reference and given images become properly aligned. Even small errors in registration can lead to largely incorrect prediction of quality. Sometimes, some point-wise non-linear transformations can be applied to reduce the dynamic range of the luminance. These preprocessing techniques are also often have channel specific parameters, as different channels have different characteristics.
- (b) *Channel Decomposition*: Motivated by the frequency and orientation specific neurons in the visual cortex, the image is usually decomposed into multiple channels using decomposition techniques such as Fourier decomposition, Gabor decomposition, DCT transform, or separable wavelet transform. Each of these decomposition techniques differ in their mathematics, implementation details, and suitability to task, however there is no clear consensus on which decomposition is better than the rest.
- (c) *Error Normalization and Pooling*: After decomposition of both reference and given image, the error is calculated as the (weighted) difference between both sets of coefficients. These errors are often normalized in a *perceptually meaningful way* [197]. Most methods use the Minkowski form of pooling errors given as:

$$E = \left(\sum_m \sum_n |e(m, n)|^\beta \right)^{1/\beta}, \quad (1)$$

where $e(m, n)$ is the normalized error of the n^{th} coefficient in the m^{th} channel of the images and β is a constant ranging from 1 to 4.

Watson’s wavelet model [200] is based on the error visibility model. This model evaluates the subjective sensitivity of each band of the linear-phase 9/7 bi-orthogonal filters (this widely used filter is now adapted by the JPEG2000 standards [197]). These sensitivity values are not only used for quality assessment but also in image compression.

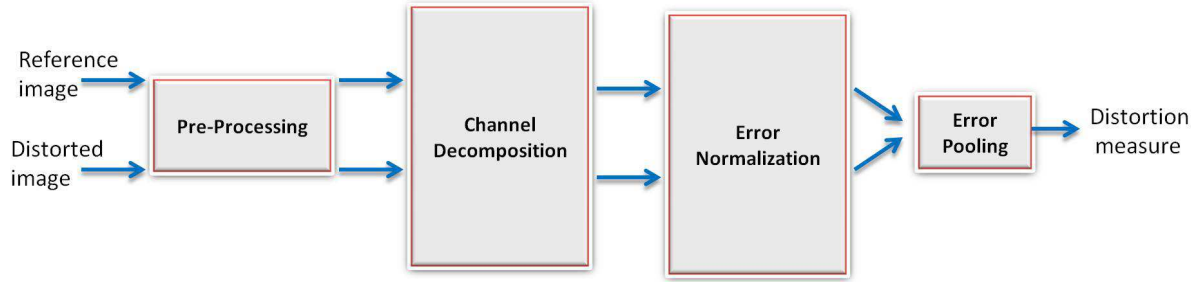


Figure A.1: A typical Full Referenced (FR) bottom-up image quality assessment system based on error visibility.

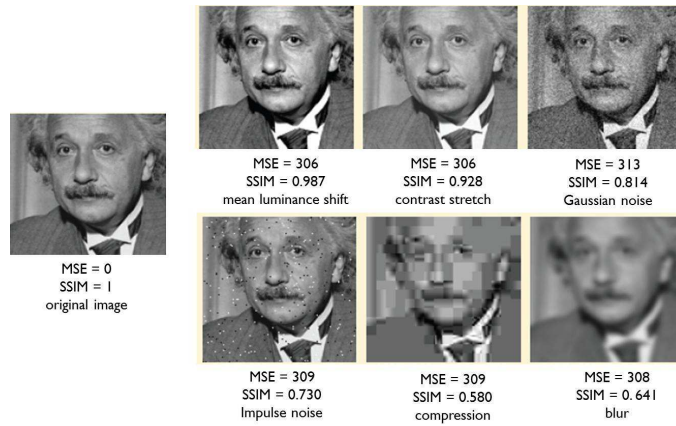


Figure A.2: Portrait image altered by various distortions. a) original image b) mean luminance shift c) contrast stretch d) impulse noise e) Gaussian noise f) blur g) compression h) spatial shift i) scale j) rotation. While most images have the same mean square error (MSE), there is a drastic difference in the visual quality. A motivating example towards SSIM quality index obtained from [198].

The FR *top-down* quality assessment algorithms have been very successful in a wide range of applications primarily due to their simplicity in design. A popular approach in literature is the *structural similarity*. This quality assessment paradigm utilizes the fact that natural images are highly structured. Hence, any unstructured information in the image is quality degradation. A spatial domain implementation of this idea is the *structural similarity index metrics* (SSIM) [198]. Given a distorted image (\mathbf{x}) and reference image (\mathbf{y}), the SSIM index of quality depends on the comparison of \mathbf{x} and \mathbf{y} by three measures: luminance, contrast, and structure.

The luminance is compared as the function $l(\mathbf{x}, \mathbf{y})$, given by

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (2)$$

where μ_x and μ_y are the mean intensities of the local luminance of \mathbf{x} and \mathbf{y} respec-

tively.

The contrast, $c(\mathbf{x}, \mathbf{y})$, is given by

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (3)$$

where σ_x and σ_y are the variance in intensities of the local luminance of \mathbf{x} and \mathbf{y} respectively.

The structure, $s(\mathbf{x}, \mathbf{y})$, is given by

$$s(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (4)$$

where σ_{xy} is the covariance of intensities of the local luminance of \mathbf{x} and \mathbf{y} .

The structural similarity (SSIM) index is given as

$$S(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5)$$

Here C_1, C_2, C_3 are mathematical constants. These equations are obtained from several observations of the HVS, such as *relative* sensitivity to luminance change (Weber's law) and some reasonable constraints on the similarity measure. Further, Equation 5 reduces to the Wang-Bovik index [196] at $C_1 = C_2 = 0$. The SSIM index parameters, σ_{xy} , μ_x and μ_y , are computed in a local region with a sliding window, with Gaussian smoothed weights to reduce boundary effects. The performance of this algorithm far exceeds traditional metrics such as mean-squared error (MSE), as shown in Figure A.1. One major drawback of spatial domain SSIM described here is the sensitivity to distortion due to translation, rotation and scaling. One solution is to use the SSIM index formulated in the complex wavelet transform domain.

2. *No-Reference* (NR): Blind or no-reference quality assessment is a more difficult problem as there is no reference image for comparison. Human visual system is able to perform blind assessment primarily due to immense prior knowledge and superior understanding of what an image is. Some distortions in an image can be assessed effectively without reference, for example, blurring and blockiness during image compression. In general, for NR quality assessment, it helps to have prior knowledge of the expected degradation process on the image. A NR perceptual quality assessment algorithm for JPEG compression is proposed by Wang et al. [198]. This method primarily measures distortions in an image due to compression (such as blockiness and blurring). It is a combination of blockiness and activity estimation in both horizontal and vertical directions.

- (a) Blockiness is estimated by the average intensity difference between block boundaries of the image x . For an image of size $M \times N$, the blockiness in horizontal direction (B_h) is given by Equation 6:

$$B_h = \frac{1}{M([N/8] - 1)} \sum_{i=1}^M \sum_{j=1}^{[N/8]-1} |d_h(i, 8j)| \quad (6)$$

where d_h is the differentiating signal in horizontal direction $d_h(m, n) = x(m, n+1) - x(m, n)$ for $n \in [1, N-1]$.

- (b) Activity of the image provides insight on the effects of compression and blur in the image. Activity (A_h) of an image of size $M \times N$ is given by:

$$A_h = \frac{1}{7} \left\{ \frac{8}{M(N-1)} \sum_{i=1}^M \sum_{j=1}^{N-1} |d_h(i, j)| - B_h \right\} \quad (7)$$

Activity of an image may also be measured via zero-crossing rate of the image of size $M \times N$ and it is given by:

$$Z_h = \frac{1}{M(N-2)} \sum_{i=1}^M \sum_{j=1}^{N-2} z_h(m, n) \quad (8)$$

where,

$$z_h(m, n) = \begin{cases} 1 & \text{if horizontal ZC at } d_h(m, n) \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

and $n \in [1, N-2]$.

Similarly, blockiness, activity and zero crossing rate is measured in vertical directions as B_v , A_v and Z_v . The overall estimation of values B, A, and Z are given by:

$$B = \frac{B_h + B_v}{2}, \quad A = \frac{A_h + A_v}{2}, \quad Z = \frac{Z_h + Z_v}{2} \quad (10)$$

Finally, the blockiness, activity and zero-crossing rate are combined to obtain quality score S,

$$S = \alpha + \beta B^{\gamma_1} A^{\gamma_2} Z^{\gamma_3} \quad (11)$$

where the parameters α , β , γ_1 , γ_2 and γ_3 are the model parameters that must be estimated for a given data set.

In another approach, Marziliano et al. [118] have proposed edge spread as a measure to estimate irregularities based on edges and their adjacent regions. Specifically, it computes the effect of irregularity in an image based on the analysis of the difference

in image intensity with respect to the local maxima and minima of pixel intensity at every row of the image. Edge spread can be computed in horizontal as well as vertical directions. However, the experiments in [118] show that either of the two directions suffices for quality assessment.

3. *Reduced-Reference* (RR): Quality assessment with reduced references is a relatively newer aspect of image quality assessment research. Here, the ancillary channel (usually noise-free, but not necessarily) transmits features of the original image that can be used to determine quality of the image at the receiver end. This quality assessment paradigm is developed to monitor the quality of video streams transmitted through various noisy channels. An early technique in literature, computes reference information from a random set of pre-selected pixel values. At the receiver end, the mean-squared error (MSE) of pixel values of original and distorted image is computed to obtain quality. Gao et al. [67] propose using multiscale geometrical analysis and compute a concise feature set that is normalized to improve HVS consistency. This feature vector (used as reference) encodes structural information that is perceived by HVS.

The primary method of representing biometric information of an individual is by an image. As noted above, most image quality assessment research is motivated towards *perceptual* quality of an image. Nevertheless, several important insights can be drawn from this matured research area towards a quality metric relevant to biometrics. An important difference being, that biometric quality *relates* to the performance of automatic biometric systems rather than the human visual system. In fact, this constraint can have several advantages such as ease of evaluation, algorithms can be easily tested when compared to testing with human subjects; also, most recognition algorithms are better understood internally than the human visual system, hence there is no need to account for various cognitive anomalies.

Appendix B

Biometric Standards

A large number of commercial and public biometric systems/solutions have lead to the standardization of several processes. This ensures inter-operability among different vendors and ensures easy integration. Here, some leading biometric standards are presented [83, 178]:

1. *CBEFF*: The Common Biometrics Exchange File Format (CBEFF) [83], developed in 2001, facilitates exchange of biometric data including raw and processed biometric sample. The standardization is achieved through three major sections: Standard biometric header (SBH), Biometric Data Block (BDB), and Signature Block (SB). Further, this standard presents a nested structure with same or different modalities. This ensures a single block structure per template in multimodal or multisample systems. Within the BDB block, there is an optional field called Biometric Data Quality. The block provisions for a single scalar quantity (0 to 100) based on the ANSI/INCITS-358 standards of 2002 (discussed next). Additionally, the field also notes if the quality value is of a nonstandard variety.
2. *BioAPI*: This standard describes the specifications of an Application Programming Interface (API) in order to accommodate for a large number of biometric systems, sensors, and applications. This API is designed for system integration and application development in biometrics. The bioAPI 1.1 standard describes in Section 2.1.46 [178], a structure called `bioapi_quality` that indicates the quality of the biometric sample in the biometric identification record [178]. Since there is no ‘universally accepted’ definition of quality, bioAPI has elected to provide this structure with the goal of framing the effect of quality on usage of the vendors. The scores are based on the purpose (another structure in bioAPI called `bioapi_purpose`) indicted by the application (e.g., capture for enrollment/verify, capture for enrollment/identify, and capture for verify). Additionally, the demands upon the biometric vary based on the actual customer application and/or environment (i.e., a particular application usage may require higher quality samples than would normally be required by less demanding applications). Quality measurements are reported as an integral value in the range of 0 to 100. These quality scores have the following interpretation:

- 0 to 25: Unacceptable - the biometric data cannot be used for the purpose specified by the application (bioapi_purpose). The biometric data must be replaced with a new sample.
- 26 to 50: Marginal - the biometric data will provide poor performance for the purpose specified by the application and in most application environments will compromise the intent of the application. the biometric data should be replaced with a new sample.
- 51 to 75: Adequate - the biometric data will provide good performance in most application environments based on the purpose specified by the application. The application should attempt to obtain higher quality data if the application developer anticipates demanding usage.
- 76 to 100: Excellent - the biometric data will provide good performance for the purpose specified by the application. The application may want to attempt to obtain better samples if the sample quality (bioapi_quality) is in the lower portion of the range (e.g., 76, 77,...) when convenient (e.g., during enrollment).

BioAPI states that the primary objective to include quality is to provide information on the suitability of the sample, i.e., the quality metric is used simply to decide to neglect a particular sample.

3. *e-Governance standards*: The Government of India has established biometric standards for identification and verification in various e-Governance applications [134]. These standards are largely based on the ISO /IEC 19794-5:2005 international best practices. While they are primarily designed for visual inspection, they can be improvised for future use as input to automatic systems. Further, these standards are being implemented for *Aadhaar* project by the Unique Identification Authority of India (UIDAI) [133].

Biometric standardization is much needed in the community to ensure easy exchange of ideas and information, with the community still struggling with problems of interpretability. One reason could be that most standardization committees are closed grouped and are not available publicly.

Bibliography

- [1] A. Abaza and A. Ross, “Quality based rank-level fusion in multibiometric systems,” in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1–6. [27](#), [85](#)
- [2] A. Abhyankar and S. Schuckers, “Iris quality assessment and bi-orthogonal wavelet based encoding for recognition,” *Pattern Recognition*, vol. 42, no. 9, pp. 1878–1894, 2009. [39](#)
- [3] G. Aggarwal, S. Biswas, P. J. Flynn, and K. W. Bowyer, “Predicting good, bad and ugly match pairs,” in *Proceedings of IEEE Workshop on Applications of Computer Vision*. IEEE, 2012, pp. 153–160. [44](#)
- [4] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in *Proceedings of International conference on Very Large databases*, vol. 1215, 1994, pp. 487–499. [111](#), [112](#)
- [5] R. Agrawal and J. C. Shafer, “Parallel mining of association rules,” *IEEE Transactions of Knowledge and Data Engineering*, vol. 8, no. 6, pp. 962–969, 1996. [150](#)
- [6] T. Ahonen, A. Hadid, and M. Pietikinen, “Face description with local binary patterns: application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006. [95](#)
- [7] F. Alonso-Fernandez and J. Bigun, “Quality factors affecting iris segmentation and matching,” in *Proceedings of IAPR International Conference of Biometrics*, 2013, pp. 1–7. [41](#)
- [8] F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, “Quality measures in biometric systems,” *IEEE Security and Privacy*, vol. 10, no. 6, pp. 52–62, 2012. [35](#)

- [9] F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Gonzalez-Rodriguez, "Quality-based conditional processing in multi-biometrics: Application to sensor interoperability," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 40, no. 6, pp. 1168–1179, 2010. [27](#), [85](#)
- [10] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun, "A comparative study of fingerprint image-quality estimation methods," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 734–743, 2007. [29](#), [35](#), [38](#), [46](#), [55](#), [88](#)
- [11] J. S. Anastasi and M. G. Rhodes, "An own-age bias in face recognition for children and older adults," *Psychonomic Bulletin & Review*, vol. 12, no. 6, pp. 1043–1047, 2005. [126](#)
- [12] W. article, "Babies switched at birth," http://en.wikipedia.org/wiki/Babies_switched_at_birth, Last accessed on April 15, 2015. [122](#)
- [13] A. Baig, A. Bouridane, and F. Kurugollu, "A novel modality independent score-level quality measure," in *Proceedings of International Symposium on Communication Systems Networks and Digital Signal Processing*, 2011, pp. 732–735. [40](#)
- [14] J. R. Barr, L. A. Cament, K. W. Bowyer, and P. J. Flynn, "Active clustering with ensembles for social structure extraction," in *Proceedings of Winter Conference on Applications of Computer Vision*, Mar. 2014, pp. 969–976. [108](#)
- [15] O. Bausinger and E. Tabassi, "Fingerprint sample quality metric NFIQ 2.0," in *Proceedings of International Conference of the Special Interest Group on Biometrics*, 2011, pp. 167–171. [38](#)
- [16] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, "Surf: Speeded up robust features," *Elsevier Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008. [138](#)
- [17] B. Becker and E. Ortiz, "Evaluation of face recognition techniques for application to facebook," in *Proceedings of IEEE International Conference on Face and Gesture Recognition*, 2008, pp. 1–6. [108](#)
- [18] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997. [138](#)

- [19] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks,” *Advances in neural information processing systems*, vol. 19, p. 153, 2007. [130](#)
- [20] J. Beveridge, G. Givens, P. Phillips, B. Draper, D. Bolme, and Y.M.Lui, “Focus on quality, predicting FRVT 2006 performance,” in *Proceedings of IEEE International Conference on Automatic Face Gesture Recognition*, 2008, pp. 1–8. [45](#), [50](#), [58](#), [78](#)
- [21] —, “FRVT 2006: Quo vadis face quality,” *Elsevier Image and Vision Computing*, vol. 28, no. 5, pp. 732–743, 2010. [23](#), [24](#), [32](#), [44](#), [45](#)
- [22] J. Beveridge, P. Phillips, G. Givens, B. Draper, M. Teli, and D. Bolme, “When high-quality face images match poorly,” in *Proceedings of IEEE Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 572–578. [34](#), [45](#), [58](#), [78](#)
- [23] S. Bharadwaj, H. Bhatt, M. Vatsa, R. Singh, and A. Noore, “Quality assessment based denoising to improve face recognition performance,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, 2011, pp. 140–145. [3](#), [25](#), [26](#)
- [24] S. Bharadwaj, H. Bhatt, R. Singh, and M. Vatsa, “Periocular biometrics: When iris recognition fails,” in *Proceedings of IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–6. [149](#)
- [25] S. Bharadwaj, H. Bhatt, R. Singh, M. Vatsa, and S. Singh, “Face recognition for newborns: A preliminary study,” in *Proceedings of IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–6. [122](#), [124](#)
- [26] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, “Computationally efficient face spoofing detection with motion magnification,” in *Proceedings of IEEE Proceedings of Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 1–6. [149](#)
- [27] S. Bharadwaj, M. Vatsa, and R. Singh, “Can holistic representations be used for face biometric quality assessment?” in *Proceedings of IEEE International Conference on Image Processing*, 2013, pp. 1–7. [45](#), [88](#)
- [28] —, “Biometric quality: a review of fingerprint, iris, and face,” *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 34, 2014. [88](#), [119](#)

- [29] H. Bhatt, S. Bharadwaj, M. Vatsa, R. Singh, A. Ross, and A. Noore, “A framework of quality-based biometric classifier selection,” in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics*, 2011, pp. 1–7. [3](#), [17](#), [26](#), [27](#), [59](#), [86](#)
- [30] H. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, A. Noore, and A. Ross, “On co-training online biometric classifiers,” in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics*, 2011, pp. 1–7. [27](#), [86](#)
- [31] H. Bhatt, R. Singh, M. Vatsa, and N. Ratha, “Improving cross-resolution face matching using ensemble-based co-transfer learning,” *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5654–5669, 2014. [121](#)
- [32] A. Burgess, K. Carr, C. Nahirny, and J. Rabun, “Nonfamily infant abductions,” *American Journal of Nursing*, vol. 108, no. 9, pp. 32–38, 2008. [122](#)
- [33] A. Burtz, F. Oski, J. Repke, and B. Rosenstein, “Newborn identification: Compliance with aap guidelines for perinatal care,” *Clinical Pediatrics*, vol. 32, no. 2, pp. 111–113, 1993. [123](#)
- [34] G. Cauwenberghs and T. Poggio, “Incremental and Decremental Support Vector Machine Learning,” in *Proceedings of Advances in Neural Information Processing Systems*, 2000, pp. 409–415. [86](#), [92](#)
- [35] S. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000. [70](#), [72](#), [74](#)
- [36] J. Changlong, K. Hakil, X. Cui, E. Park, J. Kim, J. Hwang, and S. Elliott, “Comparative assessment of fingerprint sample quality measures based on minutiae-based matching performance,” in *International Symposium on Electronic Commerce and Security*, 2009, pp. 309–313. [46](#)
- [37] M.-J. Chen and A. C. Bovik, “No-reference image blur assessment using multiscale gradient,” *EURASIP Journal on Image and Video Processing*, vol. 2011, no. 1, pp. 1–11, 2011. [23](#)
- [38] T. Chen, X. Jiang, and W. Yau, “Fingerprint image quality analysis,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, 2004, pp. 1253–1256. [36](#)

- [39] X. Chen, P. Flynn, and K. Bowyer, “IR and visible light face recognition,” *Elsevier Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 332–358, 2005. [15](#)
- [40] Y. Chen, S. Dass, and A. Jain, “Fingerprint quality indices for predicting authentication performance,” in *Proceedings of Audio and Video-Based Biometric Person Authentication*, 2005, pp. 160–170. [4](#), [17](#), [23](#), [24](#), [36](#), [37](#), [46](#), [49](#), [70](#), [89](#)
- [41] —, “Localized iris image quality using 2-D wavelets,” in *Proceedings of Advances in Biometrics*, 2005, vol. 3832, pp. 373–381. [39](#), [40](#)
- [42] Y. Y. Chen, W. H. Hsu, and H.-Y. M. Liao, “Discovering informative social subgraphs and predicting pairwise relationships from group photos,” in *ACM Multimedia*, 2012, pp. 669–678. [108](#)
- [43] H. Choi and C. Lee, “No-reference image quality metric based on image classification,” *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–11, 2011. [46](#)
- [44] T. Cootes, G. Edwards, and C. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001. [137](#)
- [45] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak, “A protocol for multibiometric data acquisition, a protocol for multibiometric data acquisition, storage and dissemination,” West Virginia University, Tech. Rep., 2007. [7](#), [40](#), [47](#), [84](#), [95](#), [96](#), [97](#), [100](#)
- [46] J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang, “Easyalbum: an interactive photo annotation system based on face clustering and re-ranking,” in *CHI*, 2007, pp. 367–376. [107](#)
- [47] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893. [60](#), [61](#)
- [48] J. Dalton, I. Kim, and B. Lim, “RFID technologies in neonatal care,” Intel Corporation, LG CNS, ECO Inc. and WonJu Christian Hospital, White Paper, 2005. [122](#)

- [49] M. Dantone, L. Bossard, T. Quack, and L. van Gool, “Augmented faces,” in *ICCV Workshops*, 2011, pp. 24–31. 108
- [50] J. Daugman, “Combining multiple biometrics,” <http://www.cl.cam.ac.uk/~jgd1000/combine/combine.html>, 2010. 27, 40
- [51] M. Davis, M. Smith, J. Canny, N. Good, S. King, and R. Janakiraman, “Towards context-aware face recognition,” in *ACM Multimedia*, 2005, pp. 483–486. 107
- [52] C. P. Diehl and G. Cauwenberghs, “SVM incremental learning, adaptation and optimization,” in *Proceedings of the International Joint Conference on Neural Networks*, vol. 4. IEEE, 2003, pp. 2685–2690. 136
- [53] P. Dreuw, P. Steingrube, H. Hanselmann, and H. Ney, “Surf-face: Face recognition under viewpoint consistency constraints,” in *Proceedings of British Machine Vision Conference*, 2009, pp. 1–11. 95
- [54] M. Du, A. Sankaranarayanan, and R. Chellappa, “Robust face recognition from multi-view videos,” *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1105–1117, 2014. 121
- [55] Y. Du, C. Belcher, Z. Zhou, and R. Ives, “Feature correlation evaluation approach for iris feature quality measure,” *Signal Processing*, vol. 90, no. 4, pp. 1176–1187, 2010. 40
- [56] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, “Rank aggregation methods for the web,” in *Proceedings of International World Wide Web Conference*, 2001, pp. 613–622. 118
- [57] M. Faundez-Zanuy, “Data fusion in biometrics,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 1, pp. 34–38, 2005. 85
- [58] C. Fields, C. Hugh, C. Warren, and M. Zimmeroff, “The ear of the newborn as an identification constant,” *Journal of Obstetrics and Gynecology*, vol. 16, pp. 98–101, 1960. 122, 124
- [59] P. M. Frederic, F. Dufaux, S. Winkler, T. Ebrahimi, and G. Sa, “A no-reference perceptual blur metric,” in *Proceedings of International Conference on Image Processing*, 2002, pp. 57–60. 88

- [60] H. Fronthaler, K. Kollreider, and J. Bigun, “Automatic image quality assessment with application in biometrics,” in *Proceedings of IEEE Computer Vision and Pattern Recognition Workshops*, 2006, p. 30. [17](#), [36](#), [38](#), [46](#), [47](#), [70](#)
- [61] A. Gallagher and T. Chen, “Clothing cosegmentation for recognizing people,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. [8](#), [107](#), [113](#), [114](#)
- [62] —, “Understanding images of groups of people,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2009, pp. 256–263. [107](#)
- [63] A. C. Gallagher and T. Chen, “Using group prior to identify people in consumer images,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8. [107](#)
- [64] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, “The CAS-PEAL large-scale chinese face database and baseline evaluations,” *IEEE Transactions on System Man, and Cybernetics-A*, vol. 38, no. 1, pp. 149–161, 2008. [63](#)
- [65] X. Gao, S. Li, R. Liu, and P. Zhang, “Standardization of face image sample quality,” in *Proceedings of Advances in Biometrics*, 2007, vol. 4642, pp. 242–251. [42](#), [43](#), [45](#), [59](#)
- [66] P. Gastaldo, R. Zunino, and J. Redi, “Supporting visual quality assessment with machine learning,” *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 1–15, 2013. [23](#)
- [67] C. Geng and X. Jiang, “Face recognition using sift features,” in *Proceedings of International Conference on Image Processing*, 2009, pp. 3313–3316. [156](#)
- [68] X. Geng, K. Smith-Miles, L. Wang, M. Li, and Q. Wu, “Context-aware fusion: A case study on fusion of gait and face for human identification in video,” *Pattern Recognition*, vol. 43, no. 10, pp. 3660–3673, 2010. [85](#)
- [69] G. Giacinto and F. Roli, “Methods for dynamic classifier selection,” in *Proceedings of International Conference on Image Analysis and Processing*, 1999, pp. 659–664. [84](#)

- [70] J. Gray, G. Suresh, R. Ursprung, W. Edwards, J. Nickerson, and P. Shinno, “Patient misidentification in the neonatal intensive care unit: Quantification of risk,” *Pediatrics*, vol. 117, pp. 46–47, 2006. [122](#)
- [71] M. Grgic, K. Delac, and S. Grgic, “SCface surveillance cameras face database,” *Multimedia Tools and Applications*, vol. 51, no. 3, pp. 863–879, 2011. [63](#), [69](#)
- [72] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE,” *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010. [67](#), [137](#), [139](#), [144](#)
- [73] P. J. Grother, G. W. Quinn, and P. J. Phillips, “Report on the evaluation of 2D still-image face recognition algorithms,” 2010. [45](#), [78](#)
- [74] Z. Haddad, A. Beghdadi, A. Serir, and A. Mokraoui, “Wave atoms based compression method for fingerprint images,” *Pattern Recognition*, vol. 46, no. 9, pp. 2450–2464, 2013. [39](#)
- [75] M. Hahsler, S. Chelluboina, K. Hornik, and C. Buchta, “The arules r-package ecosystem: Analyzing interesting patterns from large transaction data sets,” *Journal on Machine Learning Research*, vol. 12, pp. 2021–2025, 2011. [8](#), [113](#)
- [76] R. A. Hicklin, J. Buscaglia, and M. A. Roberts, “Assessing the clarity of friction ridge impressions,” *Forensic Science International*, vol. 226, no. 1, pp. 106 – 117, 2013. [39](#)
- [77] J. Hochreiter, Z. Han, S. Z. Masood, S. Fonte, and M. Tappen, “Exploring album structure for face recognition in online social networks,” *Image and Vision Computing*, 2014. [108](#)
- [78] R. Hsu, J. Shah, and B. Martin, “Quality assessment of facial images,” in *Proceedings of Biometric Consortium Conference*, 2006, pp. 1–6. [24](#), [42](#), [43](#), [57](#), [59](#)
- [79] <http://isotc.iso.org/isotcport1>, “ISO/IEC 19794 Biometric Data Interchange Formats,” *ISO/IEC JTC1/SC37/ Working Group 3*, 2005. [58](#)
- [80] <http://www.facebook.com/press/info.php?statistics>, “Facebook statistics[online].” [108](#)
- [81] F. Hua, P. Johnson, N. Sazonova, P. Lopez-Meyer, and S. Schuckers, “Impact of out-of-focus blur on face recognition performance based on modular transfer function,”

- in *Proceedings of IEEE/IAPR International Conference on Biometrics*, 2012, pp. 85–90. 44
- [82] Z. Huang, Y. Liu, X. Li, and J. Li, “An adaptive bimodal recognition framework using sparse coding for face and ear,” *Pattern Recognition Letters*, vol. 53, pp. 69–76, 2015. 85
- [83] IBIA, “What is CBEFF (common biometric exchange formats framework)?” http://www.ibia.org/cbeff/#what_is_cbeff, 2004. 157
- [84] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, “The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1511–1521, Oct 2012. 78
- [85] A. Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005. 85
- [86] A. K. Jain, K. Cao, and S. S. Arora, “Recognizing infants and toddlers using fingerprints: Increasing the vaccination coverage,” in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics*, 2014, pp. 1–8. 122, 124
- [87] A. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*. Springer, 2008. 15
- [88] A. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004. 14
- [89] W. Jia, H.-Y. Cai, J. Gui, R.-X. Hu, Y.-K. Lei, and X.-F. Wang, “Newborn footprint recognition using orientation feature,” *Neural Computing & Applications*, pp. 1–9, 2011, 10.1007/s00521-011-0530-9. 122
- [90] Z. Jinyu and N. Schmid, “Global and local quality measures for nir iris video,” in *Proceedings of IEEE Computer Vision and Pattern Recognition Workshop*, vol. 0, 2009, pp. 120–125. 40
- [91] N. D. Kalka, J. Zuo, N. Schmid, and B. Cukic, “Estimating and Fusing Quality Factors for Iris Biometric Images,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, no. 3, pp. 509–524, 2010. 17, 23, 24, 34, 40, 46, 49, 70, 89

- [92] V. Kanhangad, A. Kumar, and D. Zhang, “Contactless and pose invariant biometric identification using hand surface,” *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1415–1424, 2011. [85](#)
- [93] A. Kapoor, G. Hua, A. Akbarzadeh, and S. Baker, “Which faces to tag: Adding prior constraints into active learning,” in *Proceedings of IEEE International Conference on Computer Vision*, 2009, pp. 1058–1065. [107](#)
- [94] B. Keelan, *Handbook of Image Quality: Characterization and Prediction*. CRC Press, 2002. [28](#)
- [95] Y. Kim, K. Toh, and A. B. J. Teoh, “An online learning algorithm for biometric scores fusion,” in *Proceedings of International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–6. [86](#)
- [96] J. Kittler, M. Hatef, R. Duin, and J. Matas, “On combining classifiers,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998. [84](#)
- [97] B. Klare and A. K. Jain, “Face recognition: Impostor-based measures of uniqueness and quality,” in *Proceedings of IEEE/IAPR International Conference on Biometrics: Theory, Applications and Systems*, 2012, pp. 23–26. [43](#)
- [98] K. Kryszczuk and A. Drygajlo, “Credence estimation and error prediction in biometric identity verification,” *Signal Processing*, vol. 88, no. 4, pp. 916–925, 2008. [26](#)
- [99] K. Kryszczuk, J. Richiardi, and A. Drygajlo, “Impact of combining quality measures on biometric sample matching,” in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1–6. [23](#), [24](#), [26](#)
- [100] D. Kuefner, V. Cassia, M. Picozzi, and E. Bricolo, “Do all kids look alike? evidence for an other-age effect in adults,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 34, no. 4, pp. 811–817, 2008. [126](#)
- [101] A. Kumar and S. Shekhar, “Personal identification using multibiometrics rank-level fusion,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 41, no. 5, pp. 743–752, 2011. [27](#)

- [102] A. Kumar and D. Zhang, “Improving biometric authentication performance from the user quality,” *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 3, pp. 730–735, 2010. [17](#), [23](#), [24](#)
- [103] A. Kumar, V. Kanhangad, and D. Zhang, “A new framework for adaptive multi-modal biometrics management,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 92–102, 2010. [85](#)
- [104] N. Kumar, P. N. Belhumeur, and S. K. Nayar, “FaceTracer: A Search Engine for Large Collections of Images with Faces,” in *Proceedings of European Conference on Computer Vision*, Oct 2008, pp. 340–353. [137](#)
- [105] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar, “Attribute and simile classifiers for face verification,” in *Proceedings of International conference on computer Vision*, 2009, pp. 365–372. [106](#)
- [106] R. Lemes, O. Bellon, L. Silva, and A. Jain, “Biometric recognition of newborns: Identification using palmprints,” in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–6. [122](#), [123](#)
- [107] X. Li, Z. Sun, and T. Tan, “Predict and improve iris recognition performance based on pairwise image quality assessment,” in *Proceedings of International Conference of Biometrics*, 2013, pp. 1–8. [55](#)
- [108] E. Lim, X. Jiang, and W. Yau, “Fingerprint quality and validity analysis,” in *Proceedings of IEEE/IAPR International Conference on Image Processing*, 2002, pp. 469–472. [36](#)
- [109] D. Lin, A. Kapoor, G. Hua, and S. Baker, “Joint people, event, and location recognition in personal photo collections using cross-domain context,” in *Proceedings of European Conference on Computer Vision*, 2010, pp. 243–256. [108](#)
- [110] W. Lin and C. C. J. Kuo, “Perceptual visual quality metrics: A survey,” *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011. [23](#), [34](#)
- [111] K. Liu, M. Weng, C. Tseng, Y. Chuang, and M. Chen, “Association and temporal rule mining for post-filtering of semantic concept detection in video,” *IEEE transactions on Multimedia*, vol. 10, no. 2, pp. 240–251, 2008. [111](#)

- [112] J. Long and S. Li, “Near infrared face image quality assessment system of video sequences,” in *Proceedings of International Conference on Image and Graphics*, 2011, pp. 275–279. [42](#), [45](#)
- [113] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004. [43](#), [138](#)
- [114] O. Manyam, N. Kumar, P. Belhumeur, and D. Kriegman, “Two faces are better than one: face recognition in group photographs,” in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–8. [107](#)
- [115] G. Marcialis, F. Roli, and L. Didaci, “Personal identity verification by serial fusion of fingerprint and face matchers,” *Pattern Recognition*, vol. 42, no. 11, pp. 2807–2817, 2009. [17](#), [84](#)
- [116] G. Marcialis and F. Roli, “Fingerprint verification by fusion of optical and capacitive sensors,” *Pattern Recognition Letters*, vol. 25, no. 8, pp. 1315–1322, 2004. [15](#)
- [117] A. Martinez and R. Benavente, “The AR face database,” 1998, Computer Vision Center, Technical Report. [66](#), [75](#)
- [118] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “A no-reference perceptual blur metric,” in *Proceedings of IEEE International Conference on Image Processing*, 2004, pp. 57–60. [49](#), [71](#), [155](#), [156](#)
- [119] A. Mittal, A. Moorthy, and A. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012. [34](#), [47](#), [49](#)
- [120] S. K. Modi and S. J. Elliott, “Impact of image quality on performance: comparison of young and elderly fingerprints,” in *Proceedings of International Conference on Recent Advances in Soft Computing*, 2006, pp. 449–45. [29](#)
- [121] E. Murphy-Chutorian and M. Trivedi, “Head pose estimation in computer vision: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607–626, 2009. [32](#)
- [122] S. K. R. Nair, B. Bhanu, S. Ghosh, and N. S. Thakoor, “Predictive models for multibiometric systems,” *Pattern Recognition*, vol. 47, no. 12, pp. 3779 – 3792, 2014. [86](#)

- [123] K. Nandakumar, Y. Chen, S. Dass, and A. Jain, “Likelihood ratio-based biometric score fusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342–347, 2008. 91
- [124] K. Nandakumar, Y. Chen, A. Jain, and S. Dass, “Quality-based score level fusion in multibiometric systems,” in *Proceedings of IEEE/IAPR International Conference on Pattern Recognition*, vol. 4, 2006, pp. 473–476. 26
- [125] L. Nanni, A. Lumini, M. Ferrara, and R. Cappelli, “Combining biometric matchers by means of machine learning and statistical approaches,” *Neurocomputing*, vol. 149: Part B, pp. 526–535, 2015. 84
- [126] K. Nasrollahi and T. Moeslund, *Face Quality Assessment System in Video Sequences*, ser. Lecture Notes in Computer Science. Springer, 2008, vol. 5372. 42, 44, 59
- [127] NCMEC, “Newborn/infant abductions,” http://www.missingkids.com/en_US/documents/InfantAbductionStats.pdf, Last accessed on April 14, 2015. 122
- [128] newspaper article, “Hospital Newborn Swap Outrages Parents,” <http://abcnews.go.com/GMA/story?id=128144>, Last accessed on April 15, 2015. 122
- [129] —, “Parents horrified after hospital mixup allows their baby to be breastfed by stranger,” http://www.huffingtonpost.com/2014/01/11/baby-breastfed-by-stranger_n_4578108.html?ir=India, Last accessed on April 15, 2015. 122
- [130] —, “Switched at birth, then meeting aged 12,” <http://www.bbc.com/news/magazine-15432846>, Last accessed on April 15, 2015. 122
- [131] N. Nill and B. Bouzas, “Objective image quality measure derived from digital image power spectra,” *Optical Engineering*, vol. 31, no. 4, pp. 813–825, 1992. 49, 88
- [132] NIST, “<http://www.nist.gov/itl/iad/ig/nbis.cfm>,” Last accessed on April 14, 2015. 48
- [133] G. of India, “Adhaar project,” <http://uidai.gov.in/>. 158

- [134] D. of Information Technology, “Face image data standard for e-governance applications in india,” in *Government of India*, 2005. [158](#)
- [135] N. O’Hare and A. Smeaton, “Context-aware person identification in personal photo collections,” *IEEE Transactions on Multimedia*, vol. 11, no. 2, pp. 220–228, 2009. [107](#)
- [136] T. Ojala and M. P. D. Harwood, “A comparative study of texture measures with classification based on feature distributions,” *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, January 1996. [42](#), [43](#), [70](#)
- [137] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002. [138](#)
- [138] A. Oliva and A. Torralba, “Modeling the shape of the scene: A holistic representation of the spatial envelope,” *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001. [60](#), [61](#)
- [139] M. A. Olsen, E. Tabassi, A. Makarov, and C. Busch, “Self-organizing maps for fingerprint image quality assessment,” in *Proceedings of IEEE International Conference of Computer Vision and Pattern Recognition Workshops*, 2013, pp. 1–7. [38](#)
- [140] M. A. Olsen, H. Xu, and C. Busch, “Gabor filters as candidate quality measure for NFIQ 2.0,” in *Proceedings of IAPR International Conference on Biometrics*, 2012, pp. 158–163. [37](#)
- [141] OpenCV Library, “<http://opencv.org/>.” [50](#)
- [142] G. Patrick and T. Elham, “Performance of biometric quality measures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 531–524, 2007. [23](#), [24](#), [26](#), [46](#), [59](#), [62](#), [64](#)
- [143] N. Pela, M. Mamede, and M. Tavares, “An´alise cr´ytica de impress˜oes plantares de rec´em-nascidos,” *Revista Brasileira de Enfermagem*, vol. 29, pp. 100–105, 1975. [122](#), [123](#)
- [144] P. J. Phillips, J. R. Beveridge, D. S. Bolme, B. A. Draper, G. H. Given, Y. M. Lui, S. Cheng, M. N. Teli, and H. Zhang, “On the existence of face quality measures,” in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2013, pp. 1–8. [44](#)

- [145] P. Phillips and J. Beveridge, “An introduction to biometric-completeness: the equivalence of matching and quality,” in *Proceedings of IEEE international conference on Biometrics: Theory, applications and systems*, 2009, pp. 414–418. [79](#), [150](#)
- [146] P. Phillips, J. Beveridge, B. Draper, G. Givens, A. O’Toole, D. Bolme, J. Dunlop, Y. Lui, H. Sahibzada, and S. Weimer, “An introduction to the good, the bad, and the ugly face recognition challenge problem,” in *Proceedings of IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 346–353. [44](#), [78](#)
- [147] P. Phillips, P. Flynn, K. Bowyer, R. Bruegge, P. Grother, G. Quinn, and M. Pruitt, “Distinguishing identical twins by face recognition,” in *Proceedings of IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 185–192. [129](#)
- [148] P. Phillips, W. Scruggs, A. O’Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, “FRVT 2006 and ICE 2006 large-scale results,” 1998, NIST Technical Report NISTIR 7408. [44](#), [45](#), [78](#)
- [149] K. Phromsuthirak and V. Areekul, “Fingerprint quality assessment using frequency and orientation subbands of block-based fourier transform,” in *Proceedings of IAPR International Conference of Biometrics*, 2013, pp. 1–7. [38](#)
- [150] N. Poh, T. Bourlai, and J. Kittler, “A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms,” *Pattern Recognition*, vol. 43, pp. 1094–1105, 2010. [85](#)
- [151] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-garcia, D. Maurer, A. Salah, T. Scheidat, and C. Vielhauer, “Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms,” *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 849–866, 2009. [26](#)
- [152] N. Poh and J. Kittler, “A unified framework for biometric expert fusion incorporating quality measures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 3–18, 2012. [23](#), [24](#), [26](#), [27](#), [58](#), [84](#)

- [153] N. Poh, A. Ross, W. Lee, and J. Kittler, “A user-specific and selective multimodal biometric fusion strategy by ranking subjects,” *Pattern Recognition*, vol. 46, no. 12, pp. 3341–3357, 2013. [85](#)
- [154] H. Proenca, “Quality assessment of degraded iris images acquired in the visible wavelength,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 82–95, 2011. [40](#)
- [155] V. Radhakrishnan, V. Sabarinath, P. Thombare, P. Hazarey, R. Bonde, and A. Sheorain, “Presurgical nasolabial molding assisted primary reconstruction in complete unilateral cleft lip palate infants,” *Journal of Clinical Pediatric Dentistry*, vol. 34, no. 3, pp. 267–274, 2010. [127](#)
- [156] B. Raducanu and J. Vitri, “Online nonparametric discriminant analysis for incremental subspace learning and recognition,” *Pattern Analysis and Applications*, vol. 11, pp. 259–268, 2008. [86](#)
- [157] R. Raghavendra, B. Dorizzi, A. Rao, and G. H. Kumar, “Designing efficient fusion schemes for multimodal biometric systems using face and palmprint,” *Pattern Recognition*, vol. 44, no. 5, pp. 1076–1088, 2011. [85](#)
- [158] A. Ross and A. Jain, “Information fusion in biometrics,” *Pattern Recognition Letters*, vol. 24, pp. 2115–2125, 2003. [84](#), [94](#), [98](#)
- [159] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics*. Springer, 2006. [15](#), [81](#), [84](#), [95](#)
- [160] A. Ross and R. Nadgir, “A thin-plate spline calibration model for fingerprint sensor interoperability,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1097–1110, 2008. [31](#)
- [161] A. Sankaran, M. Vatsa, and R. Singh, “Automated clarity and quality assessment for latent fingerprints: A preliminary study,” in *Proceedings of IEEE/IAPR International Conference on Biometrics: Theory, Applications and Systems*, 2013, pp. 1–7. [39](#)
- [162] A. Sapkota, R. Gopalan, E. Zavesky, and T. Boulton, “Propagation of facial identities in a social network,” in *Proceedings of IEEE/IAPR International Conference on Biometrics: Theory, Applications and Systems*, 2013, pp. 1–7. [108](#)

- [163] A. Satish, R. Jain, and A. Gupta, “Context networks for annotating personal media,” UCI, Tech. Rep. ESL.UCI.EDU-TR, 2013. [108](#)
- [164] W. Scheirer, N. Kumar, K. Ricanek, P. Belhumeur, and T. Boulton, “Fusing with context: A bayesian approach to combining descriptive attributes,” in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–8. [106](#)
- [165] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, “Support vector method for novelty detection,” in *Proceedings of Neural Information Processing Systems*, vol. 12, 1999, pp. 582–588. [136](#)
- [166] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001. [90](#)
- [167] H. Sellaheewa and S. Jassim, “Image-quality-based adaptive face recognition,” *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 805–813, 2010. [27](#)
- [168] C. Shaokang, M. Sandra, H. Mehrtash T, S. Conrad, B. Abbas, and L. Brian C, “Face recognition from still images to video sequences: A local-feature-based framework,” *EURASIP journal on image and video processing*, vol. 2011, 2011. [44](#)
- [169] L. Shen, A. Kot, and W. Koo, “Quality measures of fingerprint images,” in *Proceedings of Audio and Video Based Biometric Person Authentication*, 2001, vol. 2091, pp. 266–271. [37](#)
- [170] K. Shepard, T. Erickson, and H. Fromm, “Limitations of footprinting as a means of infant identification,” *Pediatrics*, vol. 37, no. 1, pp. 107–108, 1966. [122](#), [123](#)
- [171] R. Singh, M. Vatsa, A. Ross, and A. Noore, “Biometric classifier update using online learning: A case study in near infrared face verification,” *Image and Vision Computing*, vol. 28, no. 7, pp. 1098–1105, 2010. [27](#), [86](#)
- [172] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, “Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 450–455, 2005. [84](#)
- [173] M. Stapleton, “Best foot forward: Infant footprints for personal identification,” FBI, Law Enforcement Bulletin 63, 1999. [123](#)

- [174] Z. Stone, T. Zickler, and T. Darrell, “Autotagging facebook: social network context improves photo annotation,” in *CVPR Workshops*, 2008, pp. 1–8. [108](#), [115](#), [116](#), [117](#)
- [175] ———, “Toward large-scale face recognition using social network context,” *Proceedings of IEEE*, vol. 98, no. 8, pp. 1408–1415, 2010. [105](#), [108](#)
- [176] M. Subasic, S. Loncaric, T. Petkovic, H. Bogunovic, and V. Krivec, “Face image validation system,” in *Proceedings of International Symposium on Image and Signal Processing and Analysis*, 2005, pp. 30–33. [42](#), [57](#), [59](#)
- [177] E. Tabassi, C. Wilson, and C. Watson, “Fingerprint image quality (NISTIR 7151),” ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7151/ir_7151.pdf, 2004. [36](#), [38](#), [49](#)
- [178] C. Tilton, “The BioAPI specification, 2002,” *National Standards Institute*, 2002. [24](#), [157](#)
- [179] S. Tiwari, A. Singh, and S. Singh, “Newborn’s ear recognition: Can it be done?” in *Proceedings of International Conference on Image Information Processing*, 2011, pp. 1–6. [122](#), [124](#)
- [180] ———, “Can face and soft-biometric traits assist in recognition of newborn?” in *Proceedings of International Conference on Recent Advances in Information Technology*, 2012, pp. 74–79. [145](#), [150](#)
- [181] K. Toh, J. Kim, and S. Lee, “Biometric scores fusion based on total error rate minimization,” *Pattern Recognition*, vol. 41, no. 3, pp. 1066–1082, 2008. [84](#)
- [182] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, “Context-based vision system for place and object recognition,” in *Proceedings of International Conference on Computer Vision*, 2003, pp. 273–280. [106](#)
- [183] M. Uray, D. Skocaj, P. M. Roth, and H. Bischof, “Incremental LDA learning by combining reconstructive and discriminative approaches,” in *Proceedings of British Machine Vision Conference*, 2007, pp. 272–281. [86](#)
- [184] M. Vatsa, R. Singh, and A. Noore, “Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 38, no. 4, pp. 1021–1035, 2008. [95](#)

- [185] —, “Context switching algorithm for selective multibiometric fusion,” in *Proceedings of International Conference on Pattern Recognition and Machine Intelligence*, 2009, pp. 452–457. [86](#)
- [186] M. Vatsa, R. Singh, A. Noore, and M. Houck, “Quality-augmented fusion of level-2 and level-3 fingerprint information using DS_m theory,” *International Journal of Approximate Reasoning*, vol. 50, no. 1, pp. 51 – 61, 2009. [36](#), [37](#)
- [187] M. Vatsa, R. Singh, A. Noore, and A. Ross, “On the dynamic selection of biometric fusion algorithms,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 470–479, 2010. [7](#), [17](#), [27](#), [59](#), [84](#), [95](#), [96](#), [97](#), [98](#), [100](#), [103](#)
- [188] M. Vatsa, R. Singh, A. Ross, and A. Noore, “Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition,” *Pattern Recognition - Special Issue on Multimodal Biometrics*, vol. 41, no. 3, pp. 880–893, 2008. [26](#)
- [189] —, “Quality-based fusion for multichannel iris recognition,” in *Proceedings of International Conference on Pattern Recognition*, 2010, pp. 1314–1317. [26](#)
- [190] M. Vatsa, R. Singh, A. Tiwari, S. Bharadwaj, and H. Bhatt, “Analyzing fingerprint of Indian population using image quality: A UIDAI case study,” in *Proceedings of Workshop on Emerging Trends and Challenges in Hand based Biometrics, International Conference on Pattern Recognition*, 2010, pp. 1–5. [25](#)
- [191] K. Veeramachaneni, L. Osadciw, and P. Varshney, “An adaptive multimodal biometric management algorithm,” *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, vol. 35, no. 3, pp. 344–356, 2005. [85](#)
- [192] A. A. Veloso, H. M. Almeida, M. A. Gonçalves, and W. Meira Jr, “Learning to rank at query-time using association rules,” in *SIGIR*, 2008, pp. 267–274. [150](#)
- [193] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103. [130](#), [138](#), [139](#), [143](#)
- [194] P. Viola and M. Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004. [137](#)

- [195] G. Wang, A. Gallagher, J. Luo, and D. Forsyth, “Seeing people in social context: Recognizing people and social relationships,” in *Proceedings of European Conference on Computer Vision*, vol. 6315, 2010, pp. 169–182. [107](#)
- [196] Z. Wang and A. Bovik, “A universal image quality index,” *Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002. [154](#)
- [197] —, *Modern Image Quality Assessment*. Synthesis Lectures on Image, Video, and Multimedia Processing, 2006, vol. 2. [23](#), [34](#), [58](#), [151](#), [152](#)
- [198] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. [10](#), [49](#), [71](#), [153](#), [154](#)
- [199] Z. Wang, H. Sheikh, and A. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” in *Proceedings of International Conferencing on Image Processing*, 2002, pp. 477–480. [88](#)
- [200] A. Watson, G. Yang, J. Solomon, and J. Villasenor, “Visibility of wavelet quantization noise,” *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1164–1175, 1997. [152](#)
- [201] D. Weingaertner, O. Bellon, M. Cat, and L. Silva, “Newborn’s biometric identification: Can it be done?” in *Proceedings of International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2008, pp. 200–205. [122](#), [123](#)
- [202] L. Wolf, T. Hassner, and Y. Taigman, “Effective unconstrained face recognition by combining multiple descriptors and learned background statistics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1978–1990, 2011. [131](#)
- [203] R. Wong, N. Poh, J. Kittler, and D. Frohlich, “Interactive quality-driven feedback for biometric systems,” in *IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–7. [24](#)
- [204] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. Lovell, “Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, 2011, pp. 74–81. [42](#), [44](#), [59](#)

- [205] K. Woods, W. P. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 405–410, 1997. [84](#)
- [206] T. Wu, J. Phillips, and R. Chellappa, "Propagation of facial identities in a social network," in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2013, pp. 1–7. [108](#)
- [207] D. Yadav, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, and K. W. Bowyer, "Unraveling the effect of textured contact lenses on iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 851–862, 2014. [39](#)
- [208] L. Yang and R. Jin, "Distance metric learning: a comprehensive survey [online]," Michigan State University, Tech. Rep., 2006. [Online]. Available: http://www.cs.cmu.edu/~liuy/frame_survey_v2.pdf [131](#)
- [209] Y. Yao, B. Abidi, and M. Abidi, "Quality assessment and restoration of face images in long range/high zoom video," in *Face Biometrics for Personal Identification*, ser. Signals and Communication Technology. Springer Berlin Heidelberg, 2007, pp. 43–60. [42](#), [45](#), [59](#)
- [210] Z. Yao, J. Le Bars, C. Charrier, and C. Rosenberger, "Fingerprint quality assessment combining blind image quality, texture and minutiae features," in *Proceedings of International Conference on Information Systems Security and Privacy*, 2015. [89](#)
- [211] S. Yendrikhovskij, "Image quality: Between science and fiction," in *Proceedings of Image Processing, Image Quality, Image Capture, Systems Conference*, 1999, pp. 173–178. [34](#)
- [212] S. Yoon, K. Cao, E. Liu, and A. Jain, "LFIQ: Latent fingerprint image quality," in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2013, p. 1. [39](#)
- [213] R. Youmaran and A. Adler, "Measuring biometric sample quality in terms of biometric information," in *Proceedings of Biometric Consortium*, 2006, pp. 1–6. [23](#), [24](#), [42](#), [43](#), [59](#)
- [214] G. Zhang and Y. Wang, "Asymmetry-based quality assessment of face images," in *Advances in Visual Computing*, vol. 5876, 2009, pp. 499–508. [42](#), [43](#), [59](#)

- [215] L. Zhang, D. V. Kalashnikov, and S. Mehrotra, “A unified framework for context assisted face clustering,” in *Proceedings of International Conference Multimedia Retrieval*, 2013, pp. 9–16. [107](#)
- [216] J. Zuo, F. Nicolo, N. Schmid, and H. Wechsler, “Adaptive biometric authentication using nonlinear mappings on quality measures and verification scores,” in *Proceedings of IEEE International Conference on Proceedings of Image Processing*, 2010, pp. 4077–4080. [40](#)
- [217] J. Zuo and N. Schmid, “Adaptive quality-based performance prediction and boosting for iris authentication: Methodology and its illustration,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 1051–1060, 2013. [40](#)