

Graduate Theses, Dissertations, and Problem Reports

2019

Automatic Kinship Verification in Unconstrained Faces using Deep Learning

Naman Kohli West Virginia University, nakohli@mix.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation

Kohli, Naman, "Automatic Kinship Verification in Unconstrained Faces using Deep Learning" (2019). *Graduate Theses, Dissertations, and Problem Reports.* 3938. https://researchrepository.wvu.edu/etd/3938

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Automatic Kinship Verification in Unconstrained Faces using Deep Learning

NAMAN KOHLI

Dissertation submitted to the Benjamin M. Statler College of Engineering and Mineral Resources at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computer Science

Afzel Noore, Ph.D., Chair Mayank Vatsa, Ph.D. Richa Singh, Ph.D. Victor Fragoso, Ph.D. Keith B. Morris, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia 2018

Keywords: Kinship Verification, Face Recognition, Deep Learning

Copyright 2018 ${\ensuremath{\mathbb C}}$ Naman Kohli

Abstract

Automatic Kinship Verification in Unconstrained Faces using Deep Learning by Naman Kohli

Kinship verification has a number of applications such as organizing large collections of images and recognizing resemblances among humans. Identifying kinship relations has also garnered interest due to several potential applications in security and surveillance and organizing and tagging the enormous number of videos being uploaded on the Internet. This dissertation has a five-fold contribution where first, a study is conducted to gain insight into the kinship verification process used by humans. Besides this, two separate deep learning based methods are proposed to solve kinship verification in images and videos. Other contributions of this research include interlinking face verification with kinship verification and creation of two kinship databases to facilitate research in this field. WVU Kinship Database is created which consists of multiple images per subject to facilitate kinship verification research. Next, kinship video (KIVI) database of more than 500 individuals with variations due to illumination, pose, occlusion, ethnicity, and expression is collected for this research. It comprises a total of 355 true kin video pairs with over 250,000 still frames.

In this dissertation, a human study is conducted to understand the capabilities of human mind and to identify the discriminatory areas of a face that facilitate kinship-cues. The visual stimuli presented to the participants determines their ability to recognize kin relationship using the whole face as well as specific facial regions. The effect of participant gender, age, and kin-relation pair of the stimulus is analyzed using quantitative measures such as accuracy, discriminability index d', and perceptual information entropy. Next, utilizing the information obtained from the human study, a hierarchical Kinship Verification via Representation Learning (KVRL) framework is utilized to learn the representation of different face regions in an unsupervised manner. We propose a novel approach for feature representation termed as filtered contractive deep belief networks (*fcDBN*). The proposed feature representation encodes relational information present in images using filters and contractive regularization penalty. A compact representation of facial images of kin is extracted as the output from the learned model and a multi-layer neural network is utilized to verify the kin accurately. The results show that the proposed deep learning framework (KVRL-fcDBN) yields state-of-the-art kinship verification accuracy on the WVU Kinship database and on four existing benchmark datasets.

Additionally, we propose a new deep learning framework for kinship verification in unconstrained videos using a novel Supervised Mixed Norm regularization Autoencoder (SMNAE). This new autoencoder formulation introduces class-specific sparsity in the weight matrix. The proposed three-stage SMNAE based kinship verification framework utilizes the learned spatio-temporal representation in the video frames for verifying kinship in a pair of videos. The effectiveness of the proposed framework is demonstrated on the KIVI database and six existing kinship databases. On the KIVI database, SMNAE yields videobased kinship verification accuracy of 83.18% which is at least 3.2% better than existing algorithms. The algorithm is also evaluated on six publicly available kinship databases and compared with best reported results. It is observed that the proposed SMNAE consistently yields best results on all the databases.

Finally, we end by discussing the connections between face verification and kinship verification research. We explore the area of self-kinship which is age-invariant face recognition. Further, kinship information is used as a soft biometric modality to boost the performance of face verification via product of likelihood ratio and support vector machine based approaches. Using the proposed KVRL-fcDBN framework, an improvement of over 20% is observed in the performance of face verification. By addressing several problems of limited samples per kinship dataset, introducing real-world variations in unconstrained databases and designing two deep learning frameworks, this dissertation improves the understanding of kinship verification across humans and the performance of automated systems. The algorithms proposed in this research have been shown to outperform existing algorithms across six different kinship databases and has till date the best reported results in this field.

Dedication

|| जय श्री राम ||

Dedicated to, my parents Kavita and Adarsh, my sister and brother-in-law Jalaj and Gaurav, my loving nephew Nevaan, my paternal grand-parents, Usha and Onkar, and my maternal grand-parents, Krishna and Satya Pal

Acknowledgements

I would like to take the opportunity to thank the people who I have met during this phase of my career and who have helped me in achieving this doctorate degree. There are many whom I will be unable to mention here but do know, that each and every interaction has helped me in maturing and becoming the person who I am now.

Foremost, I will begin by thanking my advisor **Dr. Afzel Noore**. I have been blessed to have a Ph.D advisor who has not only provided me with a great atmosphere for conducing research, but who has also shown me how to be a good human being. From the beginning, he considered this a symbiotic relationship where he also wanted to learn alongside me. His support at every step of my Ph.D life including research discussions, freedom to call at anytime and guidance towards my career goals is genuinely appreciated. I would also like to deeply thank my mentors **Dr. Richa Singh** and **Dr. Mayank Vatsa** as this Ph.D. degree would not have been possible without them. My interaction with them since my undergraduate days has instilled in me passion for research and I know that I can always count on them for their guidance and support. Their vision and mentor-ship has shaped my life and inculcated in me a deep appreciation for the role of an advisor. I would also like to thank **Dr. Keith Morris** and **Dr. Victor Fragoso** for serving in my committee and for their willingness to accommodate me. My summer internship with **Dr. Keith Morris** was a very memorable part of my stay here in Morgantown.

Over the past six years, I have met many people who have provided support in various forms. I want to give my special thanks to my co-authors Maneet, Shruti, **Prateekshit**, **Akshay**, **Ekampreet** and **Shivangi** who have helped me in growing as a person. I also want to thank **Anush**, **Tejas**, **Gaurav**, **Himanshu**, **Samarth**, **Rohit**, **Brian** and **Soumyadeep** whom I have interacted with which has resulted

in broadening my perspective in life. My friends **Gaurav** and **Bisleshana** provided a safe and friendly environment for living in Morgantown and I would like to thank them for it. I want to specially thank **Daksha** who has proved to be a safety belt along this roller coaster ride over the past few years. I have shared my entire Ph.D. journey with her as a lab mate and a friend and her support does not deserve a simple thanks.

Finally, I thank my **parents**, **sister**, and **brother-in-law** whose sacrifices and prayers deserve a special mention. My mother, **Kavita**, has always backed me to go for what I dream and has always kept a brave face on for me. My father, **Adarsh**, whose resolute confidence in me has always guided me back to my path. My sister, **Jalaj**, has been a strength of pillar for me. I could not have reached this stage without their faith in me and their unwavering support. I also want to thank my relatives back in India who have always been considerate and ever ready to provide their guidance.

Contents

D	Dedication				
A	cknov	wledgements	\mathbf{v}		
Li	st of	Figures	x		
Li	st of	Tables	xiii		
P۱	ublica	ations	xv		
1	Intr	roduction	1		
	1.1	Motivation	4		
	1.2	Relationship with Face Verification	5		
	1.3	Characteristics and Covariates of Kinship Verification	7		
		1.3.1 Images in the Wild	8		
		1.3.2 Age Gap between Kin Pairs	10		
		1.3.3 Gender Differences	10		
		1.3.4 Hierarchy in Kinship Verification	10		
	1.4	Contributions and Dissertation Structure	12		
2	Lite	erature and Databases	13		
	2.1	Image-Based Kinship Verification Research	13		
		2.1.1 Non-Deep Learning Research	13		
		2.1.2 Deep Learning Research	18		
	2.2	Video-Based Kinship Verification Research	19		
	2.3	Kinship Verification Databases	20		

3	Uno	derstanding Human Performance for Kinship	Verification	24
	3.1	Introduction		24
	3.2	Experimental Design of the Study		25
	3.3	Results and Analysis		27
	3.4	Conclusion		32
4	Ima	age-Based Kinship Verification		34
	4.1	Introduction		34
	4.2	Representation Learning Framework		35
		4.2.1 Stacked Denoising Autoencoder (SDAE) .		35
		4.2.2 Deep Belief Network (DBN)		36
		4.2.3 Proposed Filtered Contractive DBN (fc DB	N)	37
		4.2.4 Kinship Verification via Representation Lear	ning (KVRL) Frame-	
		work		39
	4.3	Experimental Evaluation		42
		4.3.1 Datasets		42
		4.3.2 Implementation Details		43
		4.3.3 Experimental Protocol		45
		4.3.4 Results of Kinship Verification		46
	4.4	Conclusion		52
5	Vid	leo-Based Kinship Verification		54
	5.1	Introduction		54
	5.2	Kinship Verification using SMNAE \ldots		55
		5.2.1 Supervised Mixed Norm Autoencoder (SMN	NAE)	55
		5.2.2 SMNAE for Kinship Verification in Unconst	trained Videos	60
		5.2.3 Implementation Details		63
	5.3	Kinship Video (KIVI) Database		63
	5.4	Experimental Evaluation		65
		5.4.1 Video-based Kinship Verification Results .		66
		5.4.2 Image-based Kinship Verification Results .		73
		5.4.3 Performance on Existing Databases		75
	5.5	Conclusion		75
6	Rel	ationship between Kinship Verification and Fa	ace Verification	77
	6.1	Introduction		77
	6.2	Self-Kinship Problem (Age-Invariant Face Verification	ion) \ldots \ldots \ldots	78

		6.2.1	Self-Kinship Experimental Results	78	
	6.3	Boosti	ng Face Verification using Kinship	80	
		6.3.1	Experimental Protocol	83	
		6.3.2	Results of Boosting Face Verification using Kinship as Context	83	
	6.4	Conclu	sion	83	
7	Con	clusior	and Future Work	85	
	7.1	Conclu	sion	85	
	7.2	Future	Research	86	
8	App	endix		88	
Bi	bliography 89				

List of Figures

1.1	Common examples of lineal kin relationships	1
1.2	Diverse research areas where kinship has been explored	2
1.3	Illustrating the problem of kinship verification in facial images	3
1.4	Similarities in stages involved in kinship verification and face verification.	6
1.5	Images captured in real-world unconstrained environment. \ldots . \ldots .	7
1.6	Kin pairs captured in unconstrained environment	8
1.7	Kin pairs can have considerable age gap (a) image of younger child, (b)	
	image of parent at similar age as child, (c) image of parent at older age	9
1.8	Kin pairs can have considerable gender variations due to different kin	
	relationships.	11
2.1	Progression of research in automatic kinship verification: (a) Cornell Kin-	
	Face $[1]$ database (2010), (b) Family 101 database $[2]$ (2013), and (c)	
	proposed KIVI database (2018)	22
3.1	Sample images from databases considered in this research: (a) Vadana	
	DB, (b) KinFace DB, and (c) UB Kin DB	26
3.2	Sample images demonstrating face regions considered in this research	27
4.1	Two-stage kinship verification via representation learning (KVRL) with	
	three-level Stacked Denoising Autoencoder (SDAE) approach [3]	40
4.2	Challenges of pose, illumination, and occlusion in multiple images of the	
	same kin-pair	44
4.3	ROC Curve for Cornell KinFace Database	46
4.4	ROC Curve for KinFaceW-I Database.	47

4.5	ROC Curve for KinFaceW-II Database	48
4.6	ROC Curve for UB-KinFace Database.	48
4.7	ROC Curve for WVUKin Database.	49
4.8	Variations in the performance of KVRL- fc DBN with respect to number	
	of filters and type of facial regions on the WVU kinship database	51
5.1	Proposed kinship verification framework where two input videos are di- vided into non-overlapping group of frames (vidlets). For every vidlet, features are learned using the proposed SMNAEs and the final video clas-	
	sification is performed by fusion of all the vidlet pair scores	56
5.2	Proposed three-stage kinship verification in unconstrained videos frame- work by utilizing mixed norm supervised autoencoder (SMNAE). In the first stage, a pair of videos is split into vidlets which are provided as in-	
	put to stacked SMNAE. The spatial representations learned from the first	
	stage are concatenated in a pairwise fashion and provided to the second	
	stage stacked SMNAE. The third stage learns the kinship-specific represen-	
	tation of the vidlets using stacked SMNAE to encode the spatio-temporal	
	information which is employed as the input to SVM for kin vs. non-kin	
	video classification.	59
5.3	Sample kin-pair subjects from Kinship Video (KIVI) face database com-	
	prising of seven kin-relations. The number of respective kin pairs present	
	in each kin-relations category in KIVI database is also indicated	64
5.4	Video-based kinship verification performance of (a) existing image-based	
	algorithms using frame level aggregation and (b) proposed framework us-	
	ing 3-Stage existing SDAE, 3-Stage $\ell_{2,p}$ AE with mixed norm regulariza-	
	tion, and the proposed 3-Stage SMNAEs on the KIVI database	68
5.5	Sample video pairs from kin and non-kin classes which are correctly and	
	incorrectly classified by the proposed SMNAE framework	69
5.6	Variations across the pivot frame with respect to number of neighbors (z) .	70
5.7	Image-based kinship verification performance of (a) existing image-based algorithms (b) first stage of SDAE, $L_{2,p}$ AE, and proposed SMNAE on	
	still frames of KIVI database.	72
6.1	An example of age-progressed pair of images of an individual. This is also an example of self-kinship as these age-separated face images can be	70
	vermed using kinsnip verification algorithms	18

6.2	ROC curves demonstrating the performance of Self-Kinship (age-invariant	
	face verification) using the KVRL algorithm	79
6.3	Humans utilize kinship as a context to identify siblings of famous person-	
	alities	80
6.4	Illustrating the steps involved in the proposed context boosting algorithm	
	where kinship verification scores generated from the KVRL framework are	
	used to improve the face verification performance	81
6.5	A probe image can have a match score (s) with an image in the gallery	
	and a kin score (k) with the associated kin in the gallery to boost the face	
	verification performance	82
6.6	ROC curves summarizing the results of Kinship aided Face Verification	
	using PLR and SVM	84

List of Tables

2.1	Non deep-learning image-based kinship verification algorithms and databases	
	used in their corresponding research published in the literature. The	
	databases in order are A: Cornell Kin, B: UB KinFace, C: KinFaceW-I	
	and KinFaceW-II and D: Other databases	14
2.2	Deep learning image-based kinship verification algorithms and databases	
	used in their corresponding research published in the literature. The	
	databases in order are A:Cornell Kin, B: UB KinFace, C: KinFaceW-I	
	and KinFaceW-II and D: Other databases	17
2.3	Video-based kinship verification research published in literature	19
2.4	Summary of kinship databases in the literature. Databases highlighted	
	have been created as part of this research. \ldots \ldots \ldots \ldots \ldots \ldots	21
3.1	Quantitative analysis of human performance on kinship verification	29
4.1	Characteristics of the five databases used in this research.	43
4.2	Kinship verification performance of the proposed KVRL framework on 5	
	different datasets	46
4.3	Comparing the kinship verification performance (%) of the proposed KVRL	
	framework with existing kinship verification algorithms on multiple datasets.	50
5.1	Characteristics of the unconstrained KIVI database	65
5.2	Video-based kinship verification accuracy $(\%)$ of existing image-based and	
	proposed kinship verification algorithms on KIVI face database	67
5.3	Video-based kinship verification performance on the seven kin-relations in	
	the KIVI database using the proposed SMNAE based framework	73

5.4	Image-based kinship verification accuracy of existing image-based and	
	proposed kinship verification algorithms on still frames of the KIVI face	
	database.	74
5.5	Kinship verification accuracy $(\%)$ of existing and proposed kinship verifi-	
	cation algorithms on existing kinship datasets denotes that the results	
	are not reported in the original paper.	74
8.1	Performance of the proposed SMNAE, KLD, and GSAE autoencoders on	
	MNIST and CIFAR-10 databases	88

Publications

1. Journal Publications

- a) Naman Kohli, Daksha Yadav, Mayank Vatsa, Richa Singh, and Afzel Noore. Supervised mixed norm autoencoder for kinship verification in unconstrained videos. *IEEE Transactions on Image Processing*, 2018.
- b) Naman Kohli, Mayank Vatsa, Richa Singh, Afzel Noore, and Angshul Majumdar. Hierarchical representation learning for kinship verification. *IEEE Transactions on Image Processing*, 26(1):289–302, 2017.
- c) Naman Kohli, Daksha Yadav, and Afzel Noore. Multiple projective dictionary learning to detect plastic surgery for face verification. *IEEE* Access, 3:2572–2580, 2015.

2. Book Chapters

- a) David Yambay, Adam Czajka, Kevin Bowyer, Mayank Vatsa, Richa Singh, Afzel Noore, Naman Kohli, Daksha Yadav, and Stephanie Schuckers. Review of iris presentation attack detection competitions. In Handbook of Biometric Anti-Spoofing, pp. 169-183. Springer, Cham, 2019.
- b) Naman Kohli, Daksha Yadav, Mayank Vatsa, Richa Singh, and Afzel Noore. Deep face-representation learning for kinship verification. In Mayank Vatsa, Richa Singh, and Angshul Majumdar, editors, Deep learning in biometrics. CRC Press, 2018.

3. Conference Publications

a) **Naman Kohli**, Daksha Yadav, Mayank Vatsa, Richa Singh, and Afzel Noore. Face verification with disguise variations via deep disguise recog-

nizer. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1–8, 2018.

- b) Daksha Yadav, Naman Kohli, Agarwal Akshay, Mayank Vatsa, Richa Singh, and Afzel Noore. Fusion of handcrafted and deep learning features for large-scale multiple iris presentation attack detection. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1–6, 2018.
- c) Daksha Yadav, Naman Kohli, Ekampreet Kalsi, Mayank Vatsa, Richa Singh, and Afzel Noore. Unraveling human perception of facial aging using eye gaze. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1–6, 2018.
- d) Daksha Yadav, Naman Kohli, Shivangi Yadav, Mayank Vatsa, Richa Singh, and Afzel Noore. Iris presentation attack via textured contact lens in unconstrained environment. In IEEE Winter Conference on Applications of Computer Vision, pages 1–8, 2018.
- e) Akshay Agarwal, Daksha Yadav, Naman Kohli, Richa Singh, Mayank Vatsa, and Afzel Noore. Face presentation attack with latex masks in multispectral videos. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 275–283, 2017.
- f) Daksha Yadav, Naman Kohli, Shruti Nagpal, Maneet Singh, Prateekshit Pandey, Mayank Vatsa, Richa Singh, Afzel Noore, Gokul Prabhakaran, and Harsh Mahajan. Region-specific fMRI dictionary for decoding face verification in humans. In International Joint Conference on Neural Networks, pages 3814–3821, 2017.
- g) Naman Kohli, Daksha Yadav, Mayank Vatsa, Richa Singh, and Afzel Noore. Synthetic iris presentation attack using iDCGAN. In International Joint Conference on Biometrics, pages 1–7, 2017.
- h) Daksha Yadav, Naman Kohli, Mayank Vatsa, Richa Singh, and Afzel Noore. Unconstrained visible spectrum iris with textured contact lens variations: Database and Benchmarking. In International Joint Conference on Biometrics, pages 1–7, 2017.
- i) David Yambay, Benedict Becker, Naman Kohli, Daksha Yadav, A. Czajka, K. W. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore, D. Gragnaniello, C. Sansone, L. Verdoliva, L. He, Y. Ru, H. Li, N. Liu, Z. Sun,

and T. Tan. Livdet iris 2017 - Iris liveness detection competition. In IEEE International Joint Conference on Biometrics, pages 733–741, 2017.

- j) Naman Kohli, Daksha Yadav, Mayank Vatsa, Richa Singh, and Afzel Noore. Detecting medley of iris spoofing attacks using DESIST. In IEEE International Conference on Biometrics Theory, Applications and Systems, pages 1–6, 2016.
- k) Daksha Yadav, Naman Kohli, Prateekshit Pandey, Richa Singh, Mayank Vatsa, and Afzel Noore. Effect of illicit drug abuse on face recognition. In IEEE Winter Conference on Applications of Computer Vision, pages 1–8, 2016.

CHAPTER -

Introduction

Kinship refers to sharing of selected genetic characteristics and features between family members. Some examples of kin relations are shown in Figure 1.1. Kinship verification is the task of judging if two individuals are kin or not. The traditional way to determine kinship in humans uses DNA (Deoxyribonucleic Acid) testing which reveals the similarity of genetic material between two individuals. The result of this technique is highly accurate and is also utilized for locating genetic relatives and gathering family genealogical details. However, this reliable process is intrusive and currently may require several hours for processing. Thus, techniques utilizing DNA for kinship verification cannot be used in situations which require real-time processing or applications where we have non-cooperative users.

Faces are valuable visual stimuli and it has been observed that humans have excellent face processing and cognition skills [4]. Faces can convey ethnicity, gender,



Brother-Brother



Brother-Sister



Sister-Sister



Mother-Daughter



Kother-Son

Figure 1.1: Common examples of lineal kin relationships.



Figure 1.2: Diverse research areas where kinship has been explored.

age, and emotions of an individual. In the real world, kinship can often be estimated by measuring the similarity between facial regions between kin. Thus, it is intuitive to hypothesize if faces can be utilized to automatically determine kinship. In this research, kinship refers to at most one generation direct blood descendants. It has been demonstrated that these descendants share similarities in face structure as well. These similarities have been called familial traits by Hogben [5].

Kinship and its characteristics have been widely studied in diverse scientific disciplines such as anthropology, psychology, neuroscience, and computer vision (as displayed in Figure 1.2). The study of kinship is valuable from various perspectives. Kinship was the central idea of anthropology research for over a century [6]. The research on kinship systems and their anthropological aspect was pioneered by Morgan [7] and has gathered significant attention over the years. Looking at kinship from a social perspective, organisms tend to form groups based on kinship relations. There have been several studies that aim to unravel the process of kinship recognition among organisms. This ability is observed across all primates in terms of living in groups, forming bonds with offspring, and recognition of kin [8].

In the domain of psychology, the hypothesis that similarity among human faces can be a cue for kinship was first formulated by Daly and Wilson [9]. Since then, facial similarity/resemblance has been utilized to determine kinship in various research experiments [10, 11, 12, 13, 14]. In these experiments, participants were presented with the face images and were asked to judge if a kin relationship existed. Maloney and Martello [15] examined the relation between *similarity* and *kinship detection* among siblings and concluded that observers look for similarity in judging kinship among children. Martello and Maloney [16] concluded that the upper portion of the face has



Figure 1.3: Illustrating the problem of kinship verification in facial images.

more discriminating power as compared to the lower half in kinship recognition. In a different study, to determine the effect of lateralization on allocentric kin recognition, they concluded that the right half portion of the face is equal to the left half portion of the face [17]. Kaminsky et al. [18] demonstrated that humans can identify kin generations apart with varying success ratio. They also deduced that children resemble their mothers more as compared to their fathers, a result which is also presented in [19] and [20].

Various neurological studies have been conducted to examine the reasons behind the ability of humans to detect genetic relatedness. Lieberman et al. [21] proposed the existence of Kinship Model Index (KI_i) which allows humans to detect kin. Kaminsky et al. [22] established that later-born children are better at kin recognition as compared to first-born which further substantiated the effect of MPA given by [21]. Platek and Kemp [23] performed a fMRI study to investigate the differences among humans between viewing kin-images as compared to other classes of the face. Their findings suggest that the pre-existing facial network learned by humans from their birth is used to discriminate between kin and non-kin with the presence of an established neuro-cognitive system.

In the field of computer vision, kinship verification research started in 2011 and has gathered enthusiasm from both computer vision and machine learning communities. The focus of the research in this field has been on automatically identifying kinship characteristics from images and more recently, videos. In this dissertation, the problem of automatic kinship verification is defined as determining if the individuals in the given pair of media are kin or not. This research area is particularly challenging due to the large intra-class variations among different kin pairs and different kin relations. Figure 1.3 showcases the problem of kinship verification where given a pair of images/videos, the objective is to classify it as kin or non-kin. A detailed literature review of the algorithms proposed for dealing with this challenging problem is conducted in Chapter 2.

1.1 Motivation

Automatic kinship verification using facial images has several applications such as locating relatives in public databases, determining the kin of a victim or suspect by law enforcement agencies, screening asylum applications where kinship relationships are to be determined, organizing and resolving identities in photo albums. There are several security aspects of kinship verification where kin of people identified as a security threat can be identified by using an automatic kinship verification framework. Automatically determining kinship information can also be used to boost automatic face recognition capabilities by utilizing the kinship characteristics as soft biometrics.

On the other hand, automatic kinship verification in videos is a relatively unexplored research area and can be highly valuable in diverse settings such as security, surveillance, and immigration control. For instance, during the investigation of surveillance footage in the Boston Marathon bombing, two male suspects were determined as the bombers. Later, it was established that the two males were brothers which led to their identification. An automatic kinship verification system that determines kin in a video could have expedited this investigation. Another application of kinship verification using videos is for border control where surveillance videos can be applied to validate the relationship between an adult and the child, thereby preventing illegal child trafficking.

Additionally, video-based kinship verification can validate or disprove kinship claims of refugees and asylum seekers. Currently, as part of its reunification program, the U.S. State Department conducts DNA testing to allow people who have relatives in the U.S. to enter as refugees [24]. Rapid-DNA is being used for this purpose; however, an automatic kinship verification algorithm can produce cost-effective results in real-time. In these instances, automatic kinship verification can be beneficial in screening possible fraudulent cases in a non-intrusive manner.

Kinship information can also be used for managing multimedia on social media

websites such as Facebook and Youtube. According to Wagner [25], in 2016, 500 million people watched Facebook videos every day. In many cases, family members have different Youtube channels where they upload daily videos. Kinship information can be applied for automatically tagging such videos and identifying the kin present in them. Kinship context in videos can also be used for automatic indexing and organization of videos, making them easily searchable.

1.2 Relationship with Face Verification

The problem of kinship verification shares commonalities with face verification. As can be seen from Figure 1.4, the stages of both the frameworks involve similar steps. Both the problems involve binary classification and require an input of two face pair media. A face detection and normalization module are typically involved to detect the face in the images which are used by the feature extraction module. A classifier at the end is used to classify if a face pair is the same subject in the case of face verification, or if it is a kin/non-kin in the case of kinship verification.

Both the problems rely on the presence of similarities in facial regions between the positive pairs. This is easily evident in the problem of face verification where the person in the positive pair is same. It is also observed that kin pairs exhibit similarities in facial regions such as similar jawline or similar eye shape. However, as is evident from Figure 1.4, the positive pairs in the problem of face recognition belong to the same individual. In the scenario of kinship verification, the positive pairs belong to two different individuals that are kin. Thus, there is an implicit challenge in the problem of kinship verification where the individuals present in the face images can have a lot of variations.

Kinship verification can also be considered as a superset of face verification. Face images belonging to the same individual but separated by age-gap would also be considered as kin by an ideal kinship verification algorithm. The problem of ageinvariant face verification is called the problem of self-kinship and has been given attention. Additionally, kinship verification framework can also be used in a face verification system by using it as a soft biometric. Corresponding kin of the gallery image can be matched with the probe image to retrieve a kinship verification score, which should be high if the probe image and the gallery image belong to the same individual. The above two scenarios are studied in Chapter 6.



Figure 1.4: Similarities in stages involved in kinship verification and face verification.



Figure 1.5: Images captured in real-world unconstrained environment.

1.3 Characteristics and Covariates of Kinship Verification

Kinship verification is a challenging research problem due to the lack of standardized experimental protocol in classifying kin in the published literature. At the same time, a small number of images in publicly available datasets lead to difficulties in training deep learning based kinship verification models. Covariates can also affect the performance of any machine learning algorithm and it is important to identify and deal with them. For instance, covariates for face recognition such as unconstrained pose, illumination, and expression can lead to a large increase in intra-class variation and decrease the inter-class separation between different subjects. In this section, we identify four different covariates of kinship verification which makes the problem more challenging.



Figure 1.6: Kin pairs captured in unconstrained environment.

1.3.1 Images in the Wild

As can be seen from Figure 1.5, real-world applications do not have any constraints with respect to the acquisition of the image. The type of face images depends a lot on the acquisition location and user co-operation. In locations such as border control, visa stations can ensure that the acquisition environment of the images is within the measures required for the obligatory applications. However, locations like airports and railways, cannot ensure that the subject image is always captured in the frontal pose with limited illumination and expression variations. Similarly, images acquired of uncooperative users can be difficult to identify as there are no constraints. For law enforcement applications and surveillance applications, images can be acquired in a non-ideal environment with faces captured at long distances, at lower resolution or occluded in certain facial regions.

Face images acquired in controlled conditions are considerably easier to verify as compared to those in uncontrolled conditions. The ongoing Face Recognition Vendor Test by NIST [26] has evaluated many face verification algorithms on controlled visa database and uncontrolled selfie and child exploitation image databases. The current results showcase that systems have higher failure to enroll rate as well as higher non-match rate in uncontrolled databases as compared to controlled databases.

An automated kinship verification algorithm should be able to deal with images that have such wild variations of pose, illumination, and expression. Figure 1.6 showcases a kin pair where both the subjects have a lot of variation with respect to pose,



Figure 1.7: Kin pairs can have considerable age gap (a) image of younger child, (b) image of parent at similar age as child, (c) image of parent at older age.

illumination and occlusion in terms of presence of glasses and beard. Majority of the algorithms initially proposed for this application considered only controlled face images, acquired in ideal conditions. Therefore, it is important to utilize kinship databases that have such wild variations.

While traditional algorithms have faced difficulty in scaling with such variations, many deep learning algorithms have been proposed in the field of face recognition that deal with such covariates [27, 28, 29]. Similarly, newer techniques based on deep learning have been proposed in the area of kinship verification that focuses on learning representations invariant to such variations [30, 31].

1.3.2 Age Gap between Kin Pairs

There can be a considerable age gap between the kin pairs, particularly when verifying individuals from different generations. Figure 1.7 showcases kin images with limited age variation between parent-child[(a)-(b)] and high age variation between parent-child [(a)-(c)]. Parent images utilized in kinship databases are considerably older than their corresponding child images. This can lead to significant textural differences between the two faces, particularly due to aging.

In the literature, it has been shown that faces of parents when they were at the same age as their child share more similarities [32]. Several papers have been proposed which utilize the distribution of young parents to minimize the differences between older parents and younger child. An automatic kinship verification algorithm should be robust against such aging variations and needs to account for such naturally existing contrasts.

1.3.3 Gender Differences

Another covariate of kinship verification is gender as there can be considerable gender differences between the kin pair. This covariate is not present in face verification and thus makes the problem of kinship verification more challenging. As shown in Figure 1.8, mother-son, father-daughter, and brother-sister have different gender variations. These variations increase the intra-class distance between the samples of the positive class.

1.3.4 Hierarchy in Kinship Verification

In this dissertation, kinship has been considered only at a single level, i.e. mother, father and their children. However, facial similarities can also exist if we consider the height of the family tree (such as grandchild-grandfather) or breadth of the family tree (such as cousins and nieces). The problem becomes more difficult as we go away from direct descendants. Thus, the hierarchy of the family tree that we are considering is an important covariate that can determine the performance of the automated kinship verification algorithm.



Figure 1.8: Kin pairs can have considerable gender variations due to different kin relationships.

1.4 Contributions and Dissertation Structure

In this dissertation, we first systematically understand the kinship cues utilized by humans during the task of kinship verification in Chapter 2. Responses are recorded from over 450 individuals and detailed analysis is done to understand the mechanism behind kinship verification. It is observed that females might be better than males in the task of kinship verification. It is also observed that results for kinship verification differ based on the presence of a female in the stimulus shown. The effect of local and global face features is also analyzed and the binocular region is shown to have less significance in kinship verification, unlike face verification.

In Chapter 3, we introduce a novel Kinship Verification via Representation Learning (KVRL) framework, that utilizes filtered contractive Deep Belief Networks to verify kin across facial images. One of the major challenges of kinship verification is limited labeled datasets. Therefore to deal with this issue, we first learn a compact representation of a face in an unsupervised manner utilizing over 600,000 images. Then, a multi-layer neural network is trained using the reduced representations for supervised classification of kin and non-kin.

Kinship verification across videos is explored in Chapter 4. A novel supervised mixed norm autoencoder is proposed to learn kinship-specific representation in unconstrained videos. A pairwise constraint is enforced for better discrimination between kin and non-kin class representation. Spatio-temporal information present in video frames is utilized by creating vidlets and changes with respect to a pivot frame are learned for the final classification. To facilitate research in this area, a large video database termed as KIVI database is created and released to the public for research purposes. It is observed that the proposed algorithm outperforms existing imagebased and video-based kinship verification algorithms.

Finally, in Chapter 5, the connection between kinship verification and face verification is explored in two case studies. In the first study, we observe that kinship verification is a superset of face verification and analyze the problem of self-kinship. An ideal kinship verification algorithm will give perfect kinship verification accuracy if the input pair of face images belong to the same individual. Thus, the problem of self-kinship is similar to face verification across aging and experiments are conducted to see the performance of KVRL framework in this area. In the second study, we analyze how kinship scores can be utilized as a soft biometric for improving the performance of face verification. An experiment is designed where KVRL framework is used as a context boosting algorithm for improving existing face verification procedures.

CHAPTER 2

Literature and Databases

The research in automatic kinship verification started in 2010 and has come a long way since then. Progress has been made in developing databases of larger size and introducing novel deep learning architectures for better kinship verification performance. The next section will discuss the different research papers in this field and the following will showcase the databases typically utilized for this research.

2.1 Image-Based Kinship Verification Research

The research in image-based kinship verification can be divided into two groups: utilizing handcrafted features and utilizing deep learning based features for classification. Table 2.1 and 2.2 showcase the research progression in the area of image-based kinship verification. The research in kinship verification utilizing facial images was introduced in 2010 by Fang et al. [1]. They collected the first kin face pair dataset, Cornell Kin-Face database. They also proposed an algorithm for facial feature extraction and feature selection for verifying kin pairs. A pictorial structure model with spring-like connections was utilized along with color features, the length between facial parts, and gradient histogram for verifying kin. They demonstrated a kinship verification performance of 70.67% on the Cornell KinFace database. Since then, several algorithms have been proposed for detecting kin utilizing different machine learning techniques.

2.1.1 Non-Deep Learning Research

Xia et al. [32] introduced a new publicly available UB KinFace database which consists of 200 groups, each consisting of one image of child and parents when they were young and old. An experimental protocol where the negative kin pairs were randomly created was discussed. They proposed a transfer subspace method for kinship verification where the images of parents when they were young were utilized as an intermediate distribution towards which the images of old parents and images of young children can be bridged. They showed kinship verification accuracy of 60% on the UB KinFace Database. In another paper, Shao et al. [33] used Gabor filters alongside metric learning and transfer subspace learning on the same database. They minimized the distance between true kin pairs while ensuring that the distance between non-kin is greater than a threshold. They reported an accuracy of 69.67% for kinship verification. Xia et al. [34] also employed the above algorithm alongside understanding semantic relevance in the associated metadata to identify kin pairs in images.

Table 2.1: Non deep-learning image-based kinship verification algorithms and databases used in their corresponding research published in the literature. The databases in order are A: Cornell Kin, B: UB KinFace, C: KinFaceW-I and KinFaceW-II and D: Other databases

Authors	Kinship Algorithm	A	В	C	D
Fang et al. [1]	Gabor-based gradient orientation pyramid	1			
Siyu et al. [32]	Transfer learning		1		
Zhou et al. $[35]$	Spatial pyramid based learning				1
Shao et al. [33]	Gabor filters with metric learning		1		
Zhou et al. [36]	Gabor-based gradient orientation pyramid				1
Xia et al. [34]	Transfer subspace learning based algorithm		1		1
Kohli et al. [37]	Self-similarity representation of Weber faces		1		1
Fang et al. [2]	Reconstruction using parts from a set of families				1
Lu et al. [38]	Multi-view NRML			1	
Hu et al. [39]	Large margin multi metric learning			1	

Van at al [40]	Discriminative multimetric				
$\begin{bmatrix} ran et al. [40] \end{bmatrix}$	learning	~	v	~	
Guo et al. [41]	Graph-based approach				1
	Prototype discriminative feature				
Yan et al. [42]	learning				
I:	Inheritable Fisher vector feature				
Liu et al. [45]	based kinship			v	
Alirezazadeh et al.	Genetic algorithm for feature			1	
[44]	selection			~	
Wu et al. [45]	Utilized color-texture features			1	1
Ven [46]	Neighborhood repulsed			1	
	correlation metric learning			~	~
Lopez et al. [47]	Chromaticity and color features			1	
Xu and Shang [48]	Used structured similarity fusion			1	
Puthenputhussery et	SIFT flow based genetic Fisher				
al. [49]	vector feature			~	
	Relative symmetric bilinear				
Qin et al. $[50]$	model and spatially voted			1	1
	feature selection method				
Zhou et al. [51]	Ensemble similarity learning			1	
Liu et al. [52]	Status-aware projection learning			1	
Mahad at al [52]	Multi-view hybrid distance	1	1		
[manpod et al. [55]	learning		•		

Metric learning approaches involve learning a distance function for a task. Lu et al. [38] proposed a multi-view neighborhood repulsed metric learning approach, where a metric was learned such that kin pairs are closer to each other as compared to non-kin pairs. They also introduced two new databases, KinFaceW-I, and KinFaceW-II for promoting research in automatic kinship verification and developed fixed protocols for these databases. Each pair of images in KinFaceW-I is acquired from a different photo while for KinFaceW-II, is acquired from the same photo.

Yan et al. [40] jointly learned multiple distance metrics on different features extracted from a pair of images. The correlation of different features belonging to the same sample is maximized alongside the probability of kin images belonging together. They showed the performance of their database on four different kinship verification databases.

Yan et al. [42] also proposed a prototype discriminative feature learning algorithm, where mid-level features representing decision values from SVM hyperplane are utilized and a metric is learned that minimizes the distances between kin and maximizes neighboring non-kin samples. Experimental results showcase that learning mid-level features achieves better verification performance as compared to the original low-level features. A new ensemble similarity learning metric was proposed by Zhou et al. [51] where a sparse bilinear similarity function was utilized to model the relative characteristics encoded in kin images. An ensemble of similarity models was employed to achieve strong generalization ability by learning an ensemble of similarity models from the single feature space and enforcing the diversity on the fused measure space to enhance the robustness of the similarity model to the large intra-class variance of real-world kin data.

Xu and Shang [48] extracted multiple feature descriptor of the input pair of images and learned a sparse bilinear similarity model for kinship verification. They utilized sparsity-inducing norms to ensure that the similarity score of kin is consistently higher than that of non-kin. Interactions and correlations between multiview data were leveraged by using l_2 norm on each view and l_1 norm between views to obtain the fused and higher-level information. Liu et al. [52] learned two status-specific projections for parent and child respectively ensuring that there is a common component that connects the two projections. These projections were linear in nature and then the objective changes to making the positive kin pair vectors parallel as compared to negative ones. They showed an improved performance on the KinFaceW-I and KinFaceW-II database.

Mahpod et al. [53] introduced a multi-view hybrid distance learning framework for learning a margin maximizing the multi-view image representations. Different feature descriptors were extracted from a single image to create this multi-view representation. Dual discriminative representations were learned for the parents and the children using a margin maximization learning scheme. They utilized Stochastic Gradient Descent in their formulation and showed improved results on three databases: Cornell KinFace, KinFaceW-I, and KinFaceW-II.

Several feature descriptors have been proposed for the problem of kinship verification. Zhou et al. [35] presented a new Spatial Pyramid Learning based feature descriptor (SPLE) for the purpose of kinship verification. They utilized normalized absolute histogram distance and reported a performance accuracy of 67.75% for kinship verification on their in-house database. Kohli et al. [37] applied Weber faces

Table 2.2: Deep learning image-based kinship verification algorithms and databases used in their corresponding research published in the literature. The databases in order are A:Cornell Kin, B: UB KinFace, C: KinFaceW-I and KinFaceW-II and D: Other databases.

Authors	Kinship Algorithm	А	В	С	D
Dehrhan et al [57]	Discrimination via gated			1	
Denghan et al. [37]	autoencoders				
Robinson et al. [58]	Fine-tuning VGG network				\checkmark
Zhang et al [20]	Deep Convolutional Neural				
Zhang et al. [50]	Networks			~	
Li et al. [59]	Siamese convolutional neural net			1	
Li et al. [31]	KinNet				✓
Wang at al [60]	Denoising auto-encoder based				1
wang et al. [00]	metric learning				~
In at al [61]	Discriminative deep multi-metric			1	
	learning			~	
Duan et al. $[62]$	Coarse to Fine Transfer	1	1	1	
Kahli at al [2]	Kinship verification via	1	1	1	1
	representation learning		•	•	~

for illumination correction and proposed self-similarity based approach for handling kinship variations in facial images. Self similarity descriptor computes the local self similarity cues in an image which is invariant to scale, translation and rotation and thus was applied to measure the similarity between kin pairs vs non-kin pairs.

Zhou et al. [36] also proposed Gabor based gradient oriented pyramid for the problem of kinship verification. Their hypothesis was that gradient-based feature representation methods may be more robust to practical variations existing in real-world images. They utilized SVM classifier to classify if a given pair of images was kin or not. Guo et al. [54] utilized DAISY descriptors along with the product of likelihood ratio for verifying kin. A face image was divided into multiple salient parts and the top three scores across different salient parts between a pair of images were taken into consideration to determine if they are kin or non-kin. Liu et al. [55] proposed a novel inheritable Fisher vector feature which maximizes the similarity between kin pairs while minimizing the similarity between non-kin pairs. They utilized the opponent color space [56] to extract the features and used a novel fractional power cosine similarity measure for classification.

2.1.2 Deep Learning Research

Recently, there has been a shift in machine learning algorithms to employ deep learning frameworks if there are large amounts of training data available. Table 2.2 showcases deep learning algorithms proposed in the area of kinship verification. Dehgan et al. [57] utilized gated autoencoders to fuse generated features with a discriminative neural layer at the end to delineate parent-offspring relationships. Firstly, a generative model utilizing gated autoencoders is used to learn the relation between the input pair of images. Discriminative training follows the generative training to ensure that the model can predict whether the input pair images are kin or not. The results showcase that mother-daughter pair has high facial similarity as compared to father-daughter pair. Zhang et al. [30] utilized facial keypoints to extract 10 image patches and created multiple convolutional neural networks. The CNN-Points structure was combined together in the last layer with a fully connected network and softmax activation was utilized at the end to predict kinship. Experimental results showcased the efficacy of using end-to-end deep learning framework as compared to handcrafted features for this research.

A siamese convolutional network approach was proposed by Li et al. [59] to verify kin pair. Separate four convolutional and three pooling layers were used for each input image but the weights were shared between the two networks. A distance metric loss function was utilized to ensure that the distance between kin pairs is smaller as compared to non-kin pairs. Lu et al. [61] developed a discriminative deep metric which maximizes the correlation of different features for face verification and also showed its application on kinship verification. Utilizing k different neural networks, the objective was to exploit the discriminative information from each feature while minimizing the difference between representations learned of one sample across multiple networks.

Robinson et al. [58] utilized the learned VGG convolutional neural network for the problem of kinship verification. Additionally, they also released the largest kinship database, FIW database, that contains images of over 1000 families. They showed that the combination of VGG features alongside PCA gives the best performance. The authors also report the results of family recognition where the objective is to identify the family, input face image belongs to. Li et al. [31] introduced a transfer learning approach where initially a network was trained for face classification and then fine-tuned for kinship verification. Four different residual networks were trained on MS-Celeb database for face recognition and the ensemble was retrained for kinship verification using a soft-triplet loss. They showcased improved results on the Families
Authors	Kinship Verification	Databagog		
Authors	Algorithm	Databases		
Dipolition of al [63]	Spatio-temporal features	Und NEMO Smile		
Dibeknogiu et al. [03]	utilizing facial dynamics	OVA-NEWO Sinne		
Dibakliaght [64]	Visual transformation	Und NEMO Smile		
Dibeknogiu [04]	aided contrastive learning	OVA-NEWO Sinne		
	Framework utilizing			
Kohli et al. $[65]$	Supervised Mixed Norm	KIVI, UvA-NEMO Smile		
	Autoencoder			

Table 2.3: Video-based kinship verification research published in literature.

in the Wild dataset.

Duan et al. [62] introduced a coarse to fine transfer learning framework for kinship verification and showed improvement on three different kinship databases. The initial coarse network is trained on the CASIA-WebFace database for face recognition with 10,575 classes. Importantly, the trained network is finetuned for kinship classification with every positive pair labeled as a separate class on the smaller kinship datasets. Wang et al. [60] introduced a new denoising auto-encoder based metric learning (DML) method for kinship verification in the wild. They utilized an autoencoder to learn a hidden representation by ensuring good reconstruction loss. Additionally, they enforced the constraint that the positive kin images in the hidden space are closer together and tested the framework on Family in the Wild database.

2.2 Video-Based Kinship Verification Research

Video-based kinship verification has a lot of applications particularly in surveillance and border control. Videos can provide more information as they have additional sources of information as compared to single image verification. At the same time, it is challenging to process videos since the faces presented in multiple frames do not have to be aligned to their respective kin frame. Table 2.3 shows the algorithms published in this domain. As seen in the table, research in this field has started slowly due to the lack of publicly available databases. The only database used in this research is by Dibeklioglu; the UvA-NEMO Smile database which was originally created as part of facial expression study.

Dibeklioglu et al. [63] used features that describe facial dynamics and spatio-

temporal appearance over smile expression to verify kin in videos. Using specific tracked landmarks, displacement signals of eyebrows, eyelids, cheeks, and lip corners are computed and facial expression dynamics are extracted. Alongside these dynamic features, temporal Completed Local Binary Patterns are computed from eyes, cheeks and lip regions. The combination of these features is given to a Support Vector Machine for classification. They achieved an accuracy of 72.89 % on spontaneous smiles.

In another paper, Dibeklioglu [64] introduced a method for learning contrastive mapping of facial appearance of kin pairs. A siamese convolutional encoder-decoder network was trained to reveal resemblance patterns between kin and a contrastive loss is used to enforce kin pairs to be close to each other. Two networks learn transformation from kin1 to kin2 and vice versa. Additionally, the frames of the videos are matched to have a similar expression. The authors report that an accuracy of 93.65% was observed on the controlled UvA-NEMO database.

2.3 Kinship Verification Databases

An important factor in stimulating research in kinship verification is the availability of databases. Table 2.4 showcases the existing kinship image and video-based datasets available publicly. As shown in Table 2.4, there are several image-based databases with Family in the Wild (FIW) [58] being the largest. Except for FIW and WVUkin [3], other databases focus on 4 key relationships. Figure 2.1 showcases the progression of kinship research with research initially focused on controlled frontal faces, followed by uncontrolled facial images and unconstrained face videos.

The details of the databases are presented below:

- (i) Cornell Kinship [1]: It consists of 286 images pertaining to 143 subject pairs. The facial images in this database are frontal pose and have a neutral expression.
- (ii) KinFaceW-I [38]: This database consists of 1066 images corresponding to 533 kin pairs. It has 156 Father-Son, 134 Father-Daughter, 116 Mother-Son, and 127 Mother-Daughter kin pair images.
- (iii) KinFaceW-II [38]: This database has been created such that images belonging to the kin pair subjects are acquired from the same photograph. It consists of 1000 kin pair images with an equal number of images belonging to the four kinship relationships: Father-Son, Father-Daughter, Mother-Son, and Mother-Daughter.

	Dataset	No. of Subjects	No. of Samples	Relations
	CornellKin [1]	300	300	4
	UB KinFace [32]	400	600	4
KFW-I [38] KFW-II [38] WVUKin [3] Family101 [2] TSKinFace [5]	KFW-I [38]	1,066	1,066	4
	KFW-II [38]	2,000	2,000	4
	WVUKin [3]	226	904	7
	Family101 [2]	607	14,816	4
	TSKinFace [50]	2,589	2,000	4
	FIW [58]	10,700	31,000	11
Video	UvA-Nemo [63]	152	514	7
v ideo	KIVI	503	503	7

Table 2.4: Summary of kinship databases in the literature. Databases highlighted have been created as part of this research.

- (iv) UB KinFace [32]: This database consists of 200 groups consisting of 600 images. Each group has one image of the child and one image belonging to the corresponding parent when they were young and when they were old. The database has 91 Father-Son, 79 Father-Daughter, 15 Mother-Son, and 21 Mother-Daughter kin pair images.
- (v) WVU Kinship [3]: All the above kinship datasets contain only one image per kin-pair and hence are not suitable for face recognition experiments. The WVU kinship database [3] was developed as part of this research. It consists of 113 pairs of individuals from Caucasian and Indian ethnicity. The dataset has four images per person, which introduces intra-class variations for a specific kin-pair. These multiple images per kin-pair include variations in pose, illumination, and occlusion. It also consists of seven kin-relations:
 - (1) Brother-Brother (BB): 22 pairs,
 - (2) Brother-Sister (BS): 9 pairs,
 - (3) Sister-Sister (SS): 13 pairs,
 - (4) Mother-Daughter (MD): 13 pairs,
 - (5) Mother-Son (MS): 8 pairs,
 - (6) Father-Son (FS): 34 pairs, and
 - (7) Father-Daughter (FD): 14 pairs.



Figure 2.1: Progression of research in automatic kinship verification: (a) Cornell KinFace [1] database (2010), (b) Family 101 database [2] (2013), and (c) proposed KIVI database (2018).

- (vi) Family 101 [2]: The Family 101 dataset contains 101 different families with distinct family names, including 206 nuclear families, 607 individuals, with 14,816 images. The database is particularly useful for family classification where the problem is identifying which family a particular image belongs to.
- (vii) TSKinFace [50] : The Tri-subject kinship face database consist of images belonging to the child, mother, and father. The database consists of 513 images of Father, Mother and Son group and 502 images of Father, Mother and Daughter

group.

(viii) FIW [58]: The families in the wild dataset consist of 10,700 images belonging to 1000 families.

Research in kinship verification across videos is new and only Dibeklioglu et al. [66] have published papers in that field. One of the reasons for that is lack of availability of a video-based kinship verification database. Previously, there has been one kinship video database, UvA-Nemo [63], which has over 500 videos encompassing 152 individuals and 7 kin-types. The database used was collected for expression-based studies and thus is not suitable for extensive kinship verification in the wild research. As part of this research, KIVI database has also been released consisting of 503 individuals and 355 positive kin pairs. The number of video frames is over 250,000.

Research in any machine learning task requires having good, well-labeled and balanced datasets. Dawson et al. [67] recently concluded that some of the kinship databases consist of kin images coming from the same photograph. This can create issues as the final classifier trained on these images may not be learning kinship characteristics and can learn "same-photo" similarities. Thus, it is our view that kinship verification algorithms can be improved by utilizing large, and difficult kinship verification datasets such as Families in the Wild and KIVI database.

Chapter 3

Understanding Human Performance for Kinship Verification

3.1 Introduction

In nature, kinship affinities have been observed in many organisms. Taylor and Sussman [68] studied kinship affinities among Lemur catta and established that several factors such as proximity, grooming, and mating patterns are dependent on kinship among the lemurs. It has been suggested that kin-group characteristics affect the co-operation among the kin through residence and mating patterns [69].

This phenomenon has also been observed in humans where the cognitive process of humans to identify kinship relations and exploration of forming kin-based bonds has been a popular topic of interest. It has been established that humans recognize and rely on kinship affinities that trigger cognitive and social interactions, and family support networks, including an understanding of common genetic diseases for possible medical interventions. Likewise, the bonds between children and mother have been an area of huge interest in the field of psychology and neuroscience [70, 71, 72]. Moreover, researchers have evaluated the effect of nature vs. nurture [73] and have observed social bonds arising in a specific cultural context. The entire area of *new genetics* has brought kinship to the forefront due to the prevailing understanding of disease etiology [74].

In this chapter, a human study is conducted to understand the ability of humans in identifying kin and a thorough evaluation is performed. Several studies have been performed in a similar manner in face recognition literature, to understand the recognition capabilities of the human mind. This would help in (a) understanding the underlying cues that humans use to identify kin, and (b) integrating these findings in an automatic algorithm to achieve better kinship verification accuracy. Lu et al. [38] have performed a similar human study based on kinship verification. They have focused on the overall kinship verification accuracy and concluded that using context information such as hair and background improves kinship verification.

3.2 Experimental Design of the Study

Amazon MTurk is an online platform specifically designed for aiding research by organizing surveys and collecting results in a comprehensive manner. MTurk allows crowdsourcing and enables researchers to include participants across diverse demographics. It has been shown to provide reliable data as compared to data provided through the traditional means of survey collection and offers a rich pool of participants [75]. It allows the creation of Human Intelligence Tasks (HITs) for surveys, studies, and experiments which are in turn completed by participants. The participants receive a reward for completing a HIT if their results are approved by the requester. In this study conducted on Amazon MTurk, a total of 479 volunteers (200 male, 279 female) participated. Among all the participants, 366 were Indians (Mean Age (M) = 33.45 years, Standard Deviation in Age (SD) = 11.67 years), 81 were Caucasians (M = 35.39 years, SD = 10.74 years), 29 were Asians (non-Indians) (M = 28.13 years, SD = 6.93 years), and 3 were African-Americans (M = 30.33 years, SD = 8.17 years).

The database used for this human study contains images taken from three databases: Vadana [76, 77], Kinship Verification database [1], and UB Kin database [34, 32, 33]. Figure 3.1 showcases images belonging from these three databases. The database consists of 150 kin pairs and 150 non-kin pairs with 39 Sister-Sister (SS) combinations, 36 Brother-Sister (BS) combinations, 35 Brother-Brother (BB) combinations, 50 Father-Son (FS) combinations, 40 Father-Daughter (FD) combinations, 41 Mother-Daughter (MD) combinations, and 59 Mother-Son (MS) combinations. Each participant is shown five pairs of images that are assigned in a random order. The participant has to answer if the pair of images belonging to two different subjects appears to be kin to each other. Additionally, the participants are also asked if they have seen the subjects in the images before. This allows us to evaluate the differences in the responses when the participants are familiar with the stimuli.

Generally, the experiments are performed with full faces, however, it is not necessary that the whole face contributes in determining kinship. Therefore, we also



(a)



(b)



(c)

Figure 3.1: Sample images from databases considered in this research: (a) Vadana DB, (b) KinFace DB, and (c) UB Kin DB.

perform the experiments with individual facial regions and the performance of the participants is determined for the following visual stimulus:

- 1. full face,
- 2. T region (containing nose and eyes),

- 3. not-T region (containing the face with the eye and nose regions obfuscated),
- 4. lower part of facial image, and
- 5. binocular region (eye strip).

Figure 3.2 illustrates different facial regions extracted from faces of subjects. The binocular region is chosen to observe the effect of eyes on kinship verification. The T region represents features in the region of the face around the eyes and nose. Furthermore, to observe the effect of outer facial regions, not-T region is chosen (which does not have regions that are included in the T region). The lower facial region is included to evaluate a hypothesis stated in an earlier research study [16] which claims that kinship cues are not present in this region.

3.3 Results and Analysis

In this study, we analyze (a) the gender and age demographics of participants, (b) the kinship relation between stimuli, and (c) the discriminative local and global face features that humans rely on to correctly detect kinship.

A quantitative analysis of the data is performed using three independent measures: accuracy of correct kinship detection based on responses from participants, discriminability or sensitivity index (d'), and using information theory to compute the kin entropy and non-kin entropy. Discriminability or sensitivity index (d') is used in signal detection theory to quantify the difference between the mean signal and noise distributions in a given stimulus as perceived by participants.

There is an inherent uncertainty in determining the relationship between stimuli. This uncertainty can be attributed to noise and higher response categories. The stimulus information entropy H(S) and noise in the signal H(S|r) are computed from the confusion matrix using Eq. 3.1 and 3.2 respectively.



Figure 3.2: Sample images demonstrating face regions considered in this research.

$$H(S) = -\sum_{i=1}^{n} p(S_i) log(p(S_i))$$
(3.1)

$$H(S|r) = -\sum_{i=1}^{n} \sum_{j=1}^{n} p(S_i, r_j) log(p(S_i|r_j))$$
(3.2)

$$I(S|r) = H(S) - H(S|r)$$
 (3.3)

Here, r refers to the response of participants and S refers to the stimulus. The information entropy I(S|r) is calculated by subtracting the noise in the signal from the stimulus entropy as shown in Eq. 3.3. The information entropy is divided by log 2 to represent in bits and larger values of the bits determine higher perceptual judgment of the participants. Higher values of accuracy or d' or total entropy indicate that the signals can be more readily detected compared to other visual artifacts that do not contribute to the kinship detection.

The results are analyzed to understand the effect of four different attributes on the performance of participants on kinship detection: gender and age of participants, stimuli kin-pairs, and facial regions. The results are summarized in Table 3.1.

Effect of Participant's Gender on Kinship Verification

In face recognition, several studies have demonstrated that women outperform men in the ability to recognize faces [78, 79]. In a meta-analysis study of over 140 face recognition studies, Herlitz and Loven [80] have found that females consistently outperform males in recognizing faces. This fact is also supported in [81], where females performed better than males in the face recognition task. The effect of gender in kinship verification is tested to see if there exists any kind of prevalence in the skills of kinship verification among males and females. As shown in Table 3.1, it is observed that there is only 1% increase in the overall accuracy of females as compared to males. Overall accuracy is defined as the proportion of correct kin and correct non-kin responses as compared to the total responses.

However, from Table 3.1, higher d' value for females as compared to males indicates a higher sensitivity of females over males in detecting the kin signal across images. This observation is also supported by the information entropy based on responses from females and males. A z-test of proportion [82] conducted at 95% confidence level also validates this claim. These quantitative measures give us an intuition that females may have the ability to verify kin better than males. One reason for this could

S. No.	Experiments	d'	Kin Entropy	Non-Kin Entropy	Total Entropy	Overall Accuracy (in %)		
Parti	cipant's Demogra	aphic - Ge	nder					
1.	Female	0.3703	0.0063	0.0069	0.0132	56.00		
2.	Male	0.2982	0.0045	0.0048	0.0093	55.00		
Partic	Participant's Demographic - Age							
1.	<30	0.3498	0.0056	0.0061	0.0117	55.60		
2.	30 - 50	0.3119	0.0050	0.0053	0.0102	55.51		
3.	>50	0.3986	0.0077	0.0082	0.0159	56.95		
Ethnicity of Participant								
1.	Caucasian	0.4948	0.0132	0.0136	0.0268	59.45		
2.	Indian	0.4266	0.0099	0.0065	0.0164	54.89		
3.	Asian	0.3914	0.0064	0.0074	0.0137	55.22		
Stimul	us Kin Relations	ship						
1.	Mother-Son	0.8211	0.0162	0.0383	0.0545	55.39		
2.	Sister-Sister	0.5505	0.0181	0.0059	0.0240	66.23		
3.	Father- Daughter	0.3762	0.0065	0.0072	0.0137	56.01		
4.	Mother- Daughter	0.3088	0.0046	0.0051	0.0097	54.74		
5.	Brother-Sister	0.2482	0.0024	0.0035	0.0059	50.11		
6.	Father-Son	0.2092	0.0021	0.0023	0.0044	53.48		
7.	Brother-Brother	0.0560	0.0002	0.0001	0.0003	54.10		
Local of	and Global Regio	ns of Face	2					
1.	Face	0.4531	0.0107	0.0115	0.0221	58.36		
2.	Not-T	0.4212	0.0084	0.0093	0.0177	57.02		
3.	Т	0.3466	0.0059	0.0064	0.0123	55.92		
4.	Chin	0.2772	0.0037	0.0040	0.0077	54.57		
5.	Binocular	0.1656	0.0013	0.0013	0.0026	52.58		

Table 3.1	Quantitative	analysis	of human	performance	on kinshir	verification
10010 0.1.	guannaure	anarysis	or munian	performance	on kinsin	vermeauon.

be that the measure being employed for testing kinship is facial similarity analogous to facial recognition; however, this needs to be tested in future studies.

The accuracy for kinship verification increases drastically when the faces are known to the subjects; the female participants achieve an accuracy of 64.54% while the male participants achieve an accuracy of 61.95%. Also, the accuracy of non-kin detection increases to 72.47% for females whereas it is only 52.34% for males. This is in accordance with the belief that women perform better in episodic memory tasks [83]. For unfamiliar faces, the trend follows the overall accuracy with females outperforming males in kinship verification.

Effect of Participant's Age on Kinship Verification

The effect of the age of participants is studied to determine whether people of a particular age group are significantly better than others in verifying kin and non-kin. Due to a limited number of participants in the younger and older age groups, the age categories have been combined into 3 different groups: <30 years, 30-50 years, and >50 years. As shown in Table 3.1, an overall accuracy of 56.95% is observed by the participants of age-group >50 years while the second highest accuracy is observed to be 55.6% in the age-group of <30 years. At the same time, a higher d' value of 0.3986 and a higher total entropy of 0.0159 as shown in Table 3.1, indicates that older age group may better distinguish between kin and non-kin. However, z-test of proportion at 95% confidence level does not indicate statistical difference among these groups which suggests that participant's age may not have a significant effect on kinship verification.

Effect of Participant Ethnicity on Kinship Detection

The dependence between the ethnicity of the participant and their kinship detection accuracy is evaluated. It is observed that Caucasian participants are more adept at recognizing kin/non-kin as observed by the overall accuracy of 59.45% shown in Table 3.1. The fact is further supported by a higher d' value of 0.4948 and a higher total entropy of 0.0268 bits as shown in Table 3.1. This result may denote higher discriminability characteristics of Caucasian participants for detecting kin or it could be a result of the composition of participants involved in the study.

The results for other-race effect are also tested and no evidence of the presence of the other-race effect in kinship detection is observed. The *other-race effect* is defined as the ability of people to recognize the common characteristics of members of their own group more accurately as compared to others. In fact, it is observed that participants belonging to the Caucasian group are better in recognizing kinship in Indian subjects than participants of Indian ethnicity. In the case of images belonging to people of Caucasian ethnicity, both Indian and Caucasian participants performed equally well. These findings are also consistent across the images of another ethnicity. Our study further substantiates the evidence presented by Alvergne *et al.* [84] in which they disprove the notion of other-race effect in kinship recognition which is in direct contrast to the face recognition skills.

Effect of Stimuli Kin-Pairs on Kinship Verification

In a number of experiments, females have outperformed males in identifying female faces [85, 86]. Therefore, it is interesting to examine if the type of kin-pair affects the decision-making process of the participants in evaluating whether the stimuli pair is related or not. As shown in Table 3.1, the **sister-sister** kin pair has the highest overall accuracy of 66.23%. However, using the d' test of significance, it is observed that the **mother-son** pair has the highest d' value of 0.8211 and the highest total entropy bits of 0.0545 as shown in Table 3.1.

We also analyzed the verification results separately for familiar and unfamiliar faces. For familiar faces, we observe that the accuracy of father-son pair increases from 53.49% to 65.98% and the sister-sister kin pair goes up to 82.2% when people are familiar with the faces. This trend is seen in all the pairs and is reflective of the memory-cognitive ability of humans. The trend for unfamiliar faces is as expected, lower than familiar and exactly similar to the overall trend i.e. the sister-sister kin pair is the easiest to detect as kin with an accuracy of 46.0%.

Using the d' values, it is observed that pairs having a female stimulus are more accurately detected as kin. The order of the pairs is Mother-Son > Sister-Sister > Father-Daughter > Mother-Daughter > Brother-Sister > Father-Son > Brother-Brother. The results are in accordance with the study conducted by Kaminsky *et al.* [18] wherein they mentioned that the presence of a female boost the kinship accuracy. This can be attributed to partial occlusion of facial features such as beard and mustache in men as compared to women. Another reason could be the higher facial recognition capability of female participants in focusing more on female faces than male faces [80]. Therefore, the results of kin with at least one female stimuli are more significant.

The results obtained for effect of participants' gender and age, as well as the kin relationship between the stimuli, are used to validate our multi-approach quantitative analysis with conclusions arrived by other researchers who may not have used the same measures as we have. With this validation, we analyze the results obtained for the effect of discriminative local and global face features on kinship verification. Our motivation is to identify the top three regions from the human study to be integrated with the automatic kinship verification.

Effect of Facial Regions on Kinship Verification

Many studies in psychology have analyzed the effect of global facial features vs local features for face recognition abilities of humans [87]. Keil [88] has emphasized the role of internal features in face recognition by concluding that eves determine the optimal resolution for face recognition. These local features have been used as parts of a descriptor in computational methods to verify kinship [54]. However, to the best of our knowledge, no study has been conducted to analyze the effect of individual facial regions in kinship verification in a human study with statistical analysis to determine their individual effects. The two above mentioned studies have focused on larger facial regions by dividing the face into two halves (laterally and horizontally). Intuitively, the subjects should perform better when the whole face is shown. However, the results in Table 3.1 show that even though the whole face yields an accuracy of 58.36%, it is not very much different compared to local regions. The local features such as **not-T region** and **T region** show an accuracy of 57.02% and 55.92% respectively. The trend remains the same even when unfamiliar image responses are taken into account. The accuracy of the T region increases to 63.45% when the image subjects are known to humans indicating that the eye features along with the nose play an important role in kinship verification.

These results are supported by the d' test of perception and total information entropy values from the stimulus and response of participants. The complete face region has the highest d' value of 0.4531 and total entropy value of 0.0221 as shown in Table 3.1, followed by the not-T region and the T region. A z-test of proportion at 95% also validates the above pattern. The results are consistent with the face recognition studies where it has been observed that face outline, eyes, and upper face are important areas for perceiving faces [87].

3.4 Conclusion

The results obtained for effect of participants' gender and age, as well as the kin relationship between the stimuli, are used to validate our multi-approach quantitative analysis with conclusions arrived by other researchers who may not have used the same measures as we have. With this validation, we analyze the results obtained for the effect of discriminative local and global face features on kinship verification. Our motivation is to identify the top regions from the human study which contain kinship cues to be integrated into an automatic kinship verification framework. In the next chapter, an automatic kinship verification algorithm is designed which will utilize the insights gained from this human study.

CHAPTER **Z**

Image-Based Kinship Verification

4.1 Introduction

The research in image-based kinship verification has seen tremendous growth with an increase in the number of kinship specific databases released in the public. As mentioned previously, research in this field started with images having frontal pose faces with limited variations. The focus of the research was on estimating handcrafted structural and textural features for verifying kin in images. Since then, several databases have been created which consist of faces with real-world variations of illumination, aging, and pose. However, one of the major challenges in image-based kinship verification research is the limited number of samples present across publicly available datasets. This poses a big challenge for developing accurate deep learning based models for this task.

Deep learning has been successfully applied to model representations in natural language processing [89, 90, 91, 92], speech recognition [93], image segmentation [94], and object recognition [95, 96]. These algorithms learn the deep features in an abstract manner by utilizing a network of non-linear transformations in a greedy layer-wise mechanism [97, 98]. Among many deep learning algorithms, stacked denoising autoencoders (SDAE) and deep belief networks (DBN) are two very popular techniques.

Using the cues from the human study conducted in Chapter 2, this research presents a deep learning based kinship verification framework that relies on learning face representations in an unsupervised manner. A new approach using the proposed filtered contractive deep belief networks (fcDBN) is presented where the formulation of RBMs is extended through a filtering approach and a contractive penalty. The

idea of this approach stems from the fact that facial images have an inherent structure which can be emphasized using filters. By simultaneously learning filters and weights, an invariant representation of the faces is learned which is utilized in kinship verification. Using contractive penalty, we learn robust features that are invariant to local variations in the images. The proposed approach shows state-of-the-art results on multiple datasets used in kinship verification research.

4.2 Representation Learning Framework

The next section describes two popular deep learning techniques, including how they are hierarchically structured in a **Kinship Verification via Representation Learning** (KVRL) framework to learn the representations of faces for kinship verification.

4.2.1 Stacked Denoising Autoencoder (SDAE)

Let $\mathbf{x} \in \mathbb{R}^{\alpha}$ represent the input feature vector. An autoencoder[99] maps the input vector to a reduced feature \mathbf{y} using a deterministic encoder function f_{θ} as given in Equation (4.1).

$$\mathbf{y} = f_{\theta}(\mathbf{x}) = \sigma(W\mathbf{x} + b) \tag{4.1}$$

where $\theta = \{W, b\}$ represents the weight and bias to be learned and σ represents the activation function used in the multi-layer neural network. This feature **y** is then used by a decoder function to reconstruct the input as shown in Equation 4.2.

$$\hat{\mathbf{x}} = f_{\theta'}(\mathbf{y}) = \sigma(W'\mathbf{y} + b') \tag{4.2}$$

where $\theta' = \{W', b'\}$ is the approximate weight and bias, σ is the sigmoid activation function and $\hat{\mathbf{x}}$ represents the probabilistic approximation of \mathbf{x} obtained from \mathbf{y} . The autoencoder minimizes the reconstruction error between \mathbf{x} and $\hat{\mathbf{x}}$.

$$\arg\min_{\theta} ||\mathbf{x} - \hat{\mathbf{x}}||^2 \tag{4.3}$$

The autoencoders are further optimized by introducing sparsity to activate a few hidden units during training. A sparsity constraint [100] on the hidden units of the autoencoder ensures that a sparse representation is obtained according to the optimization objective function given in Equation 4.4.

$$min\left(\sum_{i=1}^{m} ||\mathbf{x}_{i} - \mathbf{\hat{x}}_{i}||^{2} + \beta \sum_{j=1}^{n} KL(\rho || \hat{\rho}_{j})\right)$$
(4.4)

where *m* is the input size of the data, *n* is the number of hidden nodes, ρ is the sparsity constant, β is the weight for sparsity penalty term, and KL(.) is the Kullback-Leibler divergence metric given as,

$$KL(\rho||\hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \left(\frac{1-\rho}{1-\hat{\rho}_j}\right)$$
(4.5)

Here, $\hat{\rho}_{\mathbf{j}}$ is the average activation of the hidden units in the autoencoder during the training phase. The algorithm aims to optimize the value of $\hat{\rho}_j$ close to ρ which will make the penalty term, $KL(\rho||\hat{\rho}_j) = 0$ in the objective function. For learning robust and useful higher-level representations, denoising autoencoders are used. Denoising autoencoders are introduced to circumvent learning of identity mapping where the input vector \mathbf{x} is first corrupted to $\bar{\mathbf{x}}$ and trained to get the reconstructed vector $\hat{\mathbf{x}}$. In effect, these autoencoders are called sparse denoising autoencoders (SDAE).

4.2.2 Deep Belief Network (DBN)

A Deep Belief Network (DBN) is a graphical model that consists of stacked Restricted Boltzmann Machine (RBM) which are trained greedily layer by layer [95]. A RBM represents a bipartite graph where one set of nodes is the visible layer and the other set of nodes is the hidden layer. A DBN models the joint distribution between the observed vector \mathbf{x} and \mathbf{n} hidden layers (\mathbf{h}) as follows:

$$P(\mathbf{x}, h^1, \dots, h^n) = \left(\prod_{j=0}^{n-2} P(h^j | h^{j+1})\right) P(h^{n-1}, h^n)$$
(4.6)

where $\mathbf{x} = \mathbf{h}^{\mathbf{0}}$, $P(h^k|h^{k+1})$ is a conditional distribution for the visible units conditioned on the hidden units of the RBM at level k + 1, and $P(h^{n-1}, h^n)$ is the visible-hidden joint distribution in the top-level RBM.

In their recent work, Hinton et al. [101], proposed *dropout* training as a successful way to prevent overfitting and an alternate method for regularization in the network. This inhibits the complex co-adaptation between the hidden nodes by randomly dropping out a few neurons during the training phase. It can be thought of as a sampling process from a larger network to create random sub-networks with the aim of achieving good generalization capability. Let f denote the activation function for the nth layer, and \mathbf{W} , \mathbf{b} be the weights and biases for the layer, * denotes the element-wise

multiplication, and **m** is a binary mask with entries drawn *i.i.d.* from *Bernoulli* (1-r) indicating which activations are not dropped out. Then the forward propagation to compute the activation $\mathbf{y_n}$ of *n*th layer of the architecture can be calculated as,

$$y_n = f\left(\frac{1}{1-r}y_{n-1} * \mathbf{mW} + \mathbf{b}\right)$$
(4.7)

4.2.3 Proposed Filtered Contractive DBN (fcDBN)

The energy function of a RBM is defined as:

$$E(v,h;\theta) = -\sum_{i=1}^{D} \sum_{j=1}^{F} v_i W_{ij} h_j - \sum_{i=1}^{D} b_i v_i - \sum_{j=1}^{F} a_j h_j$$
(4.8)

or

$$E(v,h;\theta) = -\mathbf{v}^T \mathbf{W} \mathbf{h} - \mathbf{b}^T \mathbf{v} - \mathbf{a}^T \mathbf{h}$$
(4.9)

where, $\mathbf{v} \in \{0, 1\}^D$ denotes the visible variables and $\mathbf{h} \in \{0, 1\}^F$ denotes the hidden variables. The model parameters are denoted by $\theta = \{\mathbf{a}, \mathbf{b}, \mathbf{W}\}$ and W_{ij} denotes the weight of the connection between the i^{th} visible unit and j^{th} hidden unit and b_i and a_j denote the bias terms of the model. For handling real valued visible variables such as image pixel intensities, Gaussian-Bernoulli RBMs are one of the popular formulations and the energy function is defined as:

$$E_r(v,h;\theta) = -\sum_{i=1}^{D} \sum_{j=1}^{F} \frac{v_i}{\sigma_i} W_{ij} h_j - \sum_{i=1}^{D} \frac{(v_i - b_i)^2}{2\sigma^2} - \sum_{j=1}^{F} a_j h_j$$
(4.10)

Here, $\mathbf{v} \in \mathbb{R}^{D}$ denotes the real-valued visible vector and $\theta = {\mathbf{a}, \mathbf{b}, \mathbf{W}, \sigma}$ are the model parameters. The joint distribution over \mathbf{v} and \mathbf{h} , and the marginal distribution over \mathbf{v} is defined as:

$$P(v,h) = \frac{1}{Z} exp(-E(v,h;\theta))$$
(4.11)

and

$$P(v) = \sum_{h} P(v,h) \tag{4.12}$$

where, $Z = \sum_{v,h} exp(-E(v,h))$ is a partition function.

Let L_{RBM} be the loss function of RBM with the energy function defined in Eq. 4.9. It can be defined as

$$L_{RBM} = -\sum_{i=1}^{n} \log P(v_i)$$
 (4.13)

The idea of introducing contractive penalty stems from Rifai et. al [102] where they introduce contractive autoencoders. A regularization term is added in the autoencoder loss function for learning robust features as shown in Eq. 4.14.

$$L_{AE} = \arg \min_{\theta} \| \mathbf{v} - \phi(\mathcal{W}'(\phi(\mathcal{W}\mathbf{v}+b)) + b') \|^2 + \lambda \| \mathcal{J} \|_E^2$$
(4.14)

where, $\theta = \{\mathcal{W}, b\}$ represents the weight and the bias of the autoencoder to be learned, ϕ represents the activation function, λ represents the regularization parameter, and

$$\parallel \mathcal{J} \parallel_F^2 = \parallel (J(\phi(\mathcal{W}\mathbf{v}))) \parallel_F^2$$

represents the Jacobian of the encoder function of the autoencoder with respect to the input. For a linear activation function, the contractive penalty boils down to a simple weight decay term (Tikhonov type regularization). For a sigmoid the penalty is smooth and is given by:

$$\| \mathcal{J} \|_{F}^{2} = \| J(\phi(\mathcal{W}\mathbf{v})) \|_{F}^{2}$$
$$= \sum_{i} \left(\phi(\mathcal{W}\mathbf{v})_{i}(1 - \phi(\mathcal{W}\mathbf{v})_{i}) \right)^{2} \sum_{j} \mathcal{W}_{ij}^{2}$$
(4.15)

Our work is motivated by the analytic insight and practical success of contractive autoencoders. We propose to apply the contractive penalty term to the RBM formulation. Thus, the modified loss function for contractive RBMs (**c-RBM**) can be expressed as:

$$\mathcal{L}_{cRBM} = L_{RBM} + \alpha \parallel \mathcal{J} \parallel_F^2 \tag{4.16}$$

where, $\parallel \mathcal{J} \parallel_F^2$ represents Frobenius norm of the Jacobian matrix (i.e. it is l_2 -norm of the second order differential) as shown in Eq. 4.15. Penalizing the Frobenius norm of the Jacobian matrix leads to penalization of the sensitivity; which encourages robustness of the representation. The contractive penalty encourages the mapping to the feature space to be contractive to the neighborhood of the training data. The flatness induced by having low valued first derivatives will lead to the invariance of the representation for small variations in the input. We further introduce a filtering approach in the RBM. Facial images have an inherent structure in them which is not found in other types of images. Using filters, we can extract this structural information and only use the relevant filtered information for training the network. Therefore, we propose extending Eq. 4.9 (and in a similar fashion, Eq. 4.10) with a filtering approach that can incorporate the structural and relational information in the image using filters. Let $\mathcal{V}_k = (\mathbf{f}_k \cdot \mathbf{v})$,

$$E_f(\mathcal{V}_k, h; \theta_f) = -\mathcal{V}_k^T \mathbf{W} \mathbf{h} - \mathbf{b}^T \mathcal{V}_k - \mathbf{a}^T \mathbf{h}$$
(4.17)

where, "·" is the convolution operation. \mathbf{f}_k is the k^{th} learned filter of size mn and therefore, θ_f includes \mathbf{f}_k and other weight parameters. Here, the filters \mathbf{f}_k transform the input image \mathbf{v} , emphasizing relevant structural information which is used to train the RBMs. Utilizing the above energy function, the loss function of the filtered RBM, L_{fRBM} is defined similar to Eq. 4.13. Note that, the proposed formulation is different from convolutional RBMs [103]. In convolutional RBMs, the weights are shared among all locations in the image and thus, a pooling step is required to learn high level representations. In the proposed formulation, we have introduced separate filters that will account for the structure of the image and learn these filters and weight matrix simultaneously.

Combining the above two components, we define filtered contractive RBMs (fcRBM) and the loss function is modeled as:

$$\mathcal{L}_{fcRBM} = L_{fRBM} + \alpha \parallel \mathcal{J} \parallel_F^2 + \beta \parallel \mathbf{f} \parallel_2^2$$
(4.18)

where, α and β are the regularization parameters. l_2 -norm applied over the filters prevents large deviation of values that could potentially have an unwarranted filtering effect on the images. Both the components of the proposed formulation are smooth and hence differentiable; and can be solved iteratively using contrastive divergence based approach. Multiple *fc*RBMs are then stacked together to form *fc*DBN.

4.2.4 Kinship Verification via Representation Learning (KVRL) Framework

Human analysis conducted in the previous chapter suggests that out of the five facial regions, full face, T-region, and not T-region yield the best performance for kinship verification. Inspired by this observation, we design a kinship verification framework that classifies a pair of images as kin or not-kin using these three regions as input. As discussed earlier, it is challenging to define the similarities and differences in kin



a Local and global facial representation rraining using the SDAE framework.



b Supervised kin vs. non-kin classification using learned representations.

Figure 4.1: Two-stage kinship verification via representation learning (KVRL) with three-level Stacked Denoising Autoencoder (SDAE) approach [3].

and non-kin image pairs. Therefore in this research, we propose a 2-stage **Kinship Verification via Representation Learning** framework to learn the representations of faces for kinship verification using deep learning paradigm. Figure 4.1 shows the steps involved in the proposed formulation.

In the first stage, the representations of each facial region are learned. These individually learned representations are combined to form a compact representation of the whole face in the second stage. Finally, a multi-layer neural network is trained using these reduced feature representations of the whole face for supervised classification of kin and non-kin. The two stages of the KVRL framework - with stacked denoising autoencoders (SDAE) and with deep belief networks (DBN) are described below.

KVRL-SDAE: A representation of a face image is learned by stacking sparse denoising autoencoders (SDAE) in two stages and learning in a greedy layer-wise manner. Stacking autoencoders reduces the classification error [104] and thus, learns better representations than a single autoencoder. The complete face image is taken to model the global features while T-region and not-T region are chosen to model the local features as reported in [3]. These regions are also shown in Figure 4.1. Therefore, in the first stage of the KVRL framework, one SDAE is trained for each facial region so that each stacked autoencoder can learn a representation of the given face region. Each region is resized to a standard $M \times N$ image and converted to $1 \times MN$ vector which is provided as an input to a three-layer stacked sparse denoising autoencoder. Each SDAE is trained using a stochastic gradient descent followed by fine-tuning using backpropagation.

 AE_j^i represents an autoencoder that maps the input vector to a reduced feature vector. Here **i** denotes the layer to which the autoencoder belongs and **j** represents the face region on which the autoencoder is trained. Therefore, a stacked autoencoder can be represented as $[AE_1^1, AE_1^2, AE_1^3]$ indicating that the stacked autoencoders are trained on face region 1 using three individual autoencoders. Once the stacked autoencoders are toencoders are trained for each face region, the output from the last layer of each SDAE $[AE_1^3, AE_2^3, AE_3^3]$ are concatenated and given to the second stage stacked autoencoder. This represents the combination of higher level features learned from local and global facial regions. The SDAE in the second stage represents the dimensionality reduction of the input feature vector and is used for training the classifier.

KVRL-DBN: Similar to KVRL-SDAE approach, a representation of a face image is also learned by stacking RBMs and learning greedily layer by layer to form a deep belief network. Each facial region is resized to a standard M × N image and converted to 1 × MN vector which is provided as an input to a three-layer restricted Boltzmann machine. Let **RBM**ⁱ_j represent a RBM where **i** denotes the layer to which the RBM belongs and **j** represents the face region on which the RBM is trained. In the first stage, a stacked RBM is trained and the output from $[RBM_1^3, RBM_2^3, RBM_3^3]$ is concatenated and given to the second stage DBN.

KVRL-*fc***DBN**: The representation of face image is learned by stacking *fc*RBMs and learning the weights in a greedy layer by layer fashion to form a filtered contractive deep belief network (*fc*DBN). As shown in Figure 4.1, we extract three regions from

the input face image to learn both global and local features. These regions are selected based on the results of the human study that indicates the complete face, T region, and not-T region are more significant than other face regions.

Dropout is introduced in the second stage to learn the reduced features. The advantage of applying dropout is that different configurations of neural networks are trained based on the larger architecture and by employing shared weights for the nodes that are not dropped out. After learning the individual features in the first stage, the objective is to create a combined vector that encodes this information in a reduced manner. In kinship verification, the positive class has to encode large variation among the two pairs unlike face recognition (where the subjects are same). Therefore, the aim is to create a representation of faces that generalizes well. By introducing *dropout* in this approach, efficient generalization is obtained that emulates sparse representations to mitigate any possible overfitting.

The number of images in currently available kinship datasets are limited and cannot be used directly to train the deep learning algorithms. Therefore, a separate database is required to train the model employed in the KVRL framework. Once the KVRL-SDAE and KVRL-DBN algorithms are trained, representations of face images are extracted from it. This model provides a compact representation of the combined local and global features representing the whole face image. For a pair of kin images, the features are concatenated to form an input vector for the supervised classification. A three-layer feed-forward neural network is trained for binary classification of kinship.

4.3 Experimental Evaluation

This section describes the datasets, implementation details, and experimental protocols used for evaluating the effectiveness of the proposed representation learning for kinship using hierarchical multi-stage filtered contractive deep belief network (KVRLfcDBN) along with the PRL and SVM based face verification score boosting algorithms.

4.3.1 Datasets

The efficacy of the proposed kinship verification algorithm is evaluated on the following four publicly available databases.

• UB KinFace Dataset[34],

Database	No. of	Total Images	Kin Belations	Multiple
Database	Subjects	10tai images		Images
Cornell Kin [1]	286	286	4	No
UB KinFace [34]	400	600	4	No
KinFaceW-I [38]	1066	1066	4	No
KinFaceW-II [38]	2000	2000	4	No
WVU Kinship	226	904	7	Yes

Table 4.1: Characteristics of the five databases used in this research.

- Cornell Kinship Dataset [1],
- KinFace-I [38], and
- KinFace-II [38].

Along with these four, we have also prepared a new kinship database, known as the WVU Kinship Database, containing multiple images of every person¹. The WVU Kinship dataset consists of 113 pairs of individuals. The dataset has four images per person, which allows us to have intra-class variations for a specific kin-pair along with the inter-class variations generally available with all other databases. It consists of seven kin-relations: Brother-Brother (BB), Brother-Sister (BS), Sister-Sister (SS), Mother-Daughter (MD), Mother-Son (MS), Father-Son (FS), and Father-Daughter (FD). The database has 22 pairs of BB, 9 pairs of BS, 13 pairs of SS, 14 pairs of FD, 34 pairs of FS, 13 pairs of MD and 8 pairs of MS where every pair has eight images each. As shown in Figure 4.2, the multiple images per kin-pair also include variations in pose, illumination, and occlusion. Table 4.1 summarizes the characteristics of all five databases.

Kinship verification results are shown on all five databases. However, the results of face score boosting are shown only on the WVU Kinship database because the other four databases only contain a single image per person.

4.3.2 Implementation Details

Training the fcDBN algorithm to learn the face representation for kinship requires a large number of face images. For this purpose, about 600,000 face images are used. These images are obtained from various sources including CMU-MultiPIE and Youtube faces databases [105], [106]. Note that, existing algorithms do not use outside

¹The chrominance based algorithm, given by Bordallo et al. [47] performs poorly on the WVU Kinship database which validates the correctness of the database.



Figure 4.2: Challenges of pose, illumination, and occlusion in multiple images of the same kin-pair.

data; however, as mentioned previously, due to the nature of deep learning paradigm, the proposed algorithm requires large data to learn face representation useful for kinship verification.

For face detection, all the images are aligned using affine transformation and Viola-Jones face detection algorithm [107]. Facial regions are extracted from each image and resized to 32×32 . The resized regions are converted to a vector of size 1024 and given as input to individual fcDBN deep learning algorithm in the first stage of Fig 4.1(a). For every individual *fcDBN*, three filtered contractive RBMs are stacked together and all of them are learned in a greedy layer-wise fashion where each layer receives the representation of the output from the previous layer. In the first stage, the number of nodes is 1024, 512, and 512 respectively. An output vector of size 512 is obtained from each deep belief network and is concatenated to form a vector of size 1536. A compact representation is learned from the *fc*DBN in the second stage and is used for training the classifier. In the second stage of the deep belief network, the size of the three layers is 1536, 1024, and 512, respectively. The *dropout* is applied with probability 0.5 on the hidden nodes and 0.2 on the input vectors. The performance of the proposed KVRL-*fc*DBN algorithm is also evaluated when only face is used or all the five facial regions (shown in Figure 3.2) are used. Since the architecture is flexible in nature, we also compare the results using both Sparse Denoising Autoencoders (SDAE) and Deep Belief Network (DBN) in the KVRL framework. We term these approaches of KVRL framework as KVRL-SDAE and KVRL-DBN. A similar process is followed for KVRL-DBN and KVRL-SDAE. For every individual SDAE, three different autoencoders

are stacked together and all of them are learned For all three facial regions, the number of nodes in the SDAE learned in the first stage is same $\left[AE_j^1, AE_j^2, AE_j^3\right] = [1024, 512, 256]$. An output vector of size 256 is obtained from the third level of each stacked denoising autoencoder. These outputs are the representations learned from the global and local face regions and are concatenated to form a vector of size 768. This vector is input to the stacked denoising autoencoder in the second stage of the proposed KVRL architecture which yields the final vector of size 512. This is the final learned representation of the *face* image and used for classification with the neural network.

4.3.3 Experimental Protocol

The performance of the proposed KVRL-fcDBN framework is evaluated on the same experimental protocol as described by Yan et al. [40], where five-fold cross-validation for kin verification is performed by keeping the images in all relations to be roughly equal in all folds. This protocol is followed to ensure that the experimental results are directly comparable even though the list of negative pairs may vary. In this algorithm, a random negative pair for kinship is generated such that each image is used only once in the training phase. The performance of the proposed algorithm (KVRL-fcDBN) is compared with the baseline evaluations of KVRL framework along with three state-of-the-art algorithms.

- Multiview neighborhood repulsed metric learning (MNRML)[38][†],
- Discriminative multi-metric learning $(DMML)[40]^{\dagger}$, and
- Discriminative model $[57]^{\dagger}$.

Since the proposed architecture is flexible in nature, we also utilize Sparse Denoising Autoencoders (SDAE) and Deep Belief Network (DBN) in the KVRL framework. We term these approaches of KVRL framework as KVRL-SDAE and KVRL-DBN. The proposed approach (KVRL-fcDBN) is compared with KVRL-SDAE, KVRL-DBN, and KVRL-cDBN (where contractive RBMs are utilized in the KVRL framework). We also analyze the effect of regions is observed where different combinations of facial regions are given as input to the KVRL-fcDBN framework.

Algorithm	Cornell	UB	KinFace-I	KinFace-II	WVU
KVRL-SDAE	82.0	85.9	92.3	92.7	78.7
KVRL-DBN	83.6	88.3	93.0	93.9	83.5
KVRL-	80.5	01.9	06.1	06.2	00.8
fcDBN	69.0	91.0	50.1	50.2	90.0

Table 4.2: Kinship verification performance of the proposed KVRL framework on 5 different datasets.



Figure 4.3: ROC Curve for Cornell KinFace Database.

4.3.4 Results of Kinship Verification

Table 4.2 and Figure 4.3, 4.4, 4.5, 4.6, 4.7 shows the results obtained using the experiments conducted on multiple databases. It is observed that KVRL-*fc*DBN consistently performs better than the KVRL-SDAE and KVRL-DBN approach on all the datasets. The transformation of original input through the filters improves learning of the underlying representations.

For Cornell Kinship database, kin pairs belonging to Mother-Son relation are correctly detected with the highest accuracy of 95.2% by the KVRL-*fc*DBN framework. Similarly, the KVRL-SDAE framework detects Mother-Son kin pairs with 85.0% kinship verification accuracy. The KVRL-*fc*DBN and KVRL-DBN framework outperform the existing MRNML[38], DML[40], and Discriminative Model[57] techniques for

[†]Since the experimental protocol is same, results are directly reported from the papers.

kinship verification on KinFaceW-I and KinFaceW-II databases. KVRL-*fc*DBN yields the highest kinship verification of 98.4% for Mother-Daughter relation on KinFaceW-I database. For KinFaceW-II database, the KVRL-*fc*DBN achieves 96.8% accuracy for Father-Son and Mother-Daughter relations. A general trend appears for KinFaceW-I and KinFaceW-II databases where the images of same-gender kin perform better than different-gender kin images. Thus, Father-Son and Mother-Daughter kinship relations have a higher kinship verification accuracy than Father-Daughter and Mother-Son.



Figure 4.4: ROC Curve for KinFaceW-I Database.



Figure 4.5: ROC Curve for KinFaceW-II Database.



Figure 4.6: ROC Curve for UB-KinFace Database.



Figure 4.7: ROC Curve for WVUKin Database.

UB KinFace database [32] consists of groups of images, which includes images of children, young parents, and old parents. The database is built on the hypothesis that images of parents when they were young are more similar to images of children, as compared to images of parents when they are older. Results of comparative analysis of kinship verification performance on the two sets of this database are reported in Table 4.2 and Figure 4.6. It is observed that the kinship verification performance is better in child-young parent kin pair (Set 1) as compared to child-old parent kin pair (Set 2) where there is a significant age gap between the kin pairs. The performance of the KVRL-fcDBN framework is higher than KVRL-DBN and KVRL-SDAE framework for all seven kin relations on the WVUKin database. Similar to KinFaceW-I and KinFaceW-II databases, kin pairs belonging to the same gender have greater kinship verification accuracy as compared to different gender kin pairs.

Table 4.3 also shows the results for different kin-relations obtained using the proposed deep learning algorithms. Compared to existing algorithms, KVRL-fcDBN framework consistently yields state-of-the-art results and shows improvement of up to 21% for all kin relations. It is observed that for UB database, the algorithm performs better when the images belong to children and young parents (Set 1) as compared to when there is a considerable gap between the ages of the kin (Set 2). A general trend appears for KinFace-I, KinFace-II and WVU Kinship database, where Table 4.3: Comparing the kinship verification performance (%) of the proposed KVRL framework with existing kinship verification algorithms on multiple datasets.

Algorithm	\mathbf{FS}	FD	MS	MD
MNRML[38]	74.5	68.8	77.2	65.8
DMML[40]	76.0	70.5	77.5	71.0
KVRL using SDAE	85.0	80.0	85.0	75.0
KVRL using DBN	88.3	80.0	90.0	72.5
KVRL using c-DBN	90.0	84.8	90.0	78.9
KVRL using <i>fc</i> DBN	91.7	87.9	95.2	84.2

(a) Cornell Kinship Dataset

Algorithm	Child-Young Parents	Child-Old Parents
MNRML[38]	66.5	65.5
DMML[40]	74.5	70.0
KVRL using SDAE	85.9	84.8
KVRL using DBN	88.5	88.0
KVRL using c-DBN	90.0	89.5
KVRL using fc DBN	92.0	91.5

(b) UB Kinship Dataset

Algorithm	FS	FD	MS	MD
MRNML[38]	72.5	66.5	66.2	72.0
DML[40]	74.5	69.5	69.5	75.5
Discriminative ^[57]	76.4	72.5	71.9	77.3
KVRL using SDAE	95.5	88.8	87.1	96.9
KVRL using DBN	96.2	89.6	87.9	97.6
KVRL using c-DBN	97.4	93.3	90.5	98.4
KVRL using <i>fc</i> DBN	98.1	96.3	90.5	98.4

(c) KinFaceW-I Dataset

(d) KinFaceW-II Dataset

Algorithm	\mathbf{FS}	FD	MS	MD
MNRML[38]	76.9	74.3	77.4	77.6
DML[40]	78.5	76.5	78.5	79.5
Discriminative ^[57]	83.9	76.7	83.4	84.8
KVRL using SDAE	94.0	89.2	93.6	94.0
KVRL using DBN	94.8	90.8	94.8	95.6
KVRL using c-DBN	96.0	92.4	96.4	96.8
KVRL using <i>fc</i> DBN	96.8	94.0	97.2	96.8

(e) WVU Kinship Dataset

Algorithm	FS	FD	MS	MD	BB	BS	SS
KVRL using SDAE	80.9	76.1	74.2	80.7	81.6	76.5	80.3
KVRL using DBN	85.9	79.3	76.0	84.8	85.0	79.9	85.7
KVRL using c-DBN	87.9	79.9	83.6	91.3	86.9	82.6	91.8
KVRL using fc DBN	90.8	84.4	90.6	95.2	90.9	87.5	95.7



a Verification performance with changing the number of filters.



b Kinship verification performance with respect to regions taken in the KVRL- $fc{\rm DBN}$ framework.

Figure 4.8: Variations in the performance of KVRL-fcDBN with respect to number of filters and type of facial regions on the WVU kinship database.

the images of kin of the same gender perform better than images belonging to a different gender. Specifically, Father-Son and Mother-Daughter kinship relations have a higher accuracy than Father-Daughter and Mother-Son. This relationship is also observed for the brothers and sisters as compared to Brother-Sister pair in the WVU Kinship database.

The performance of the KVRL-fcDBN approach is also computed with respect to the number of filters as shown in Figure 4.8(a). It is observed that the accuracy increases as the number of filters increases but no noticeable improvement is observed after six filters. From the human study, as mentioned previously, it is observed that the full face, T and Not-T regions are more discriminatory and thus are utilized in the KVRL-fcDBN framework. For validation, experiments are performed by providing different regions as input to the KVRL framework and the results are shown in Figure 4.8 (b). It is observed experimentally that the combination of face, T and Not-T regions perform the best in the proposed KVRL-fcDBN framework. This approach is also computationally less intensive than using all the regions in the framework.

We also compare the performance of neural network classifier with SVM classifier for kinship verification. Using SVM with RBF kernel, across all the databases yields slightly lower performance compared to the neural network and the difference is 0.2-0.5%. Computationally, on a six-core Xeon Processor with 64GB RAM, the proposed framework requires 1 second for feature extraction and kinship verification.

4.4 Conclusion

In this chapter, we develop an automatic deep learning framework for kinship verification using images. Kinship verification research using facial images has real-world applications such as locating missing relatives and validating kinship claims during the asylum-seeking process. Different approaches have been proposed for this task such as metric learning, feature-based solutions, and deep learning algorithms. Due to the limited number of labeled samples in this task, we develop an unsupervised Kinship Verification via Representation Learning (KVRL) to extract a compact representation of a face. This representation utilizes the important regions identified by the human study in Chapter 2 for kinship verification.

KVRL is a two-stage hierarchical representation learning framework that utilizes the trained deep learning representations of faces to calculate a kinship similarity score. We propose a filtered contractive deep belief network as a module for learning features that are robust to minute neighborhood changes of the input. Additionally, we also examine the efficacy of two deep learning algorithms: deep belief networks and stacked denoising autoencoders in Kinship Verification via Representation Learning (KVRL) framework [3]. We demonstrate that these kinship verification approaches outperform recently reported results on five different kinship datasets publicly available.

Chapter **C**

Video-Based Kinship Verification

5.1 Introduction

It is interesting to observe that the majority of the research on kinship classification has focused on small-scale image databases. However, their applicability also lies in large-scale real-world unconstrained scenarios. It is our assertion that existing algorithms may not be able to scale well to unconstrained scenarios due to challenges such as variations in pose, illumination, and expression. As these covariates are inherently present in videos, they are a natural extension of the image-based kinship verification problem. Videos provide larger sources of information and the spatiotemporal information across different frames can be utilized for learning more complex representations as compared to still images.

In 2013, Dibeklioglu et al. [63] introduced the problem of kinship verification in videos. They computed spatio-temporal facial dynamics features of high-quality videos to model kinship characteristics between two individuals. They used UvA-NEMO Smile database for experimental purposes where smile based dynamic spatiotemporal features and complete local binary patterns on three orthogonal planes features [108] were extracted and used for kinship verification. This database was initially developed for expression variations and was captured in controlled settings. Recently, Dibeklioglu [64] introduced visual transformation between the facial appearance of kin for the task of video-based kinship verification.

In this research, we utilize a deep learning framework to learn the spatio-temporal kinship information in videos. We also introduce a new unconstrained video-based kinship face database. The key contributions of this research are:

1. A novel Supervised Mixed Norm autoencoder (SMNAE) using $\ell_{2,p}$ norm and
class-based reconstruction penalty is proposed to learn kinship-specific representation. This new autoencoder formulation introduces class-specific sparsity in the weight matrix through pairwise constraints.

- 2. A deep learning framework utilizing SMNAEs is proposed for kinship verification in unconstrained videos. The proposed multi-stage framework utilizes the spatio-temporal information present in the video frames to verify kinship.
- 3. A new kinship video (KIVI) face database of 503 individuals with *wild* variations due to pose, illumination, occlusion, ethnicity, and expression is introduced. The database consists of 252,804 still frames corresponding to seven kin-relations.

5.2 Kinship Verification using SMNAE

In this section, we elaborate on the proposed deep learning framework for kinship verification using unconstrained videos. This framework utilizes a novel formulation of autoencoder using $l_{2,p}$ norm and class-based reconstruction error to promote class-based sparsity in the learned weight matrix. The details of the proposed SMNAE and kinship verification framework are described in the following subsections.

5.2.1 Supervised Mixed Norm Autoencoder (SMNAE)

An autoencoder consists of an encoder that maps the input data (X) to the latent space and a decoder that maps the learned representation to its reconstruction by minimizing the following loss function:

$$\underset{\mathbf{W},\mathbf{W}'}{\arg\min} \| \mathbf{X} - \mathbf{W}'\phi(\mathbf{W}\mathbf{X}) \|_F^2 + \lambda R$$
(5.1)

where, **W** is the weight matrix to be learned, ϕ is the activation function, λ is the regularization constant of the regularizer R, and $\|\cdot\|_F^2$ denotes the Frobenius norm. Introducing sparsity constraint on the autoencoders forces autoencoders to learn underlying patterns by utilizing only a few hidden nodes. A variety of regularization schemes such as ℓ_1 -norm or ℓ_2 - norm can be employed on the weights of the autoencoder to promote sparsity. Sparse autoencoders [100] have been proposed which utilize Kullback-Leibler divergence to ensure that the average activation of the nodes stays below the pre-defined sparsity parameter.



Figure 5.1: Proposed kinship verification framework where two input videos are divided into non-overlapping group of frames (vidlets). For every vidlet, features are learned using the proposed SMNAEs and the final video classification is performed by fusion of all the vidlet pair scores.

Recently, new formulations of supervised autoencoders have been introduced that learn representative features for each class separately. They utilize the class labels to enforce class-wise constraints on the weights of the autoencoders, thus, enforcing discrimination. For learning similarity between faces, Gao et al. [109] introduced supervised autoencoders using label information. Similarly, Majumdar et al. [110] proposed class-wise sparsity by applying $\ell_{2,1}$ norm on the weight matrix of the autoencoders. However, it has been established that utilizing ℓ_p norm with 0 $can find sparser solutions as compared to using <math>l_1$ norm [111, 112]. Therefore, $\ell_{2,p}$ norm with $0 may enforce better sparsity on the weights than <math>\ell_{2,1}$ norm. Yan et al. [113] demonstrated the effectiveness of $\ell_{2,p}$ norm for learning local structures of the data distribution for clustering. In this research, we utilize $\ell_{2,p}$ mixed norm for introducing class-wise sparsity on the weights. The intuition behind this approach is to learn sparse representative features for each class separately.

The loss function J_{L2P} for this autoencoder (referred to as $L_{2,p}$ autoencoder) is defined as:

$$J_{L2P} = \underset{\mathbf{W},\mathbf{W}'}{\operatorname{arg\,min}} \| \mathbf{X} - \phi(\mathbf{W}'\phi(\mathbf{W}\mathbf{X})) \|_{F}^{2} + \lambda \sum_{c=1}^{C} \| \mathbf{W}\mathbf{X}_{c} \|_{2,p}$$

$$(5.2)$$

where, $0 , <math>X_c$ denotes the input data belonging to class c, and C is the number of classes of the input data. For a vector \mathbf{W} , $\ell_{2,p}$ norm is computed by taking ℓ_p norm across the samples in rows of \mathbf{W} , w^i .

$$\|\mathbf{W}\|_{2,p} = \left(\sum_{i=1}^{d} \|w^{i}\|^{p}\right)^{1/p}$$
(5.3)

Additionally, we introduce a class-based reconstruction penalty on the autoencoder to encode discriminatory information between classes. The motivation of using this pairwise constraint is to add more discriminating power to the autoencoder. Therefore, we combine $L_{2,p}$ norm and pairwise class-based sparsity penalty to propose a new Supervised Mixed Norm autoencoder (SMNAE).

Given an input **X**, containing N pairs of samples for a binary classification problem, the loss function for the proposed SMNAE, J_{SMNAE} is described below:

$$J_{SMNAE} = \underset{\mathbf{W},\mathbf{W}'}{\operatorname{arg\,min}} \| \mathbf{X} - \phi(\mathbf{W}'\mathbf{H}) \|_{F}^{2} + \lambda \sum_{c=1}^{C} \| \mathbf{W}\mathbf{X}_{c} \|_{2,p} + \beta(\operatorname{Tr}(\mathbf{H}^{T}\mathbf{HL}))$$
(5.4)

where, λ and β are the regularization constants of respective cost functions, Tr denotes the trace of the matrix, Laplacian matrix **L** is constructed as $\mathbf{L} = \mathbf{D} - \mathbf{M}$, **D** is the diagonal matrix formed as $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_N)$, $d_i = \sum_{j=1}^C M_{i,j}$, $\mathbf{H} = \phi(\mathbf{W}\mathbf{X})$, and **M** is a matrix such that

 $M_{i,j} = \begin{cases} +1, & \text{if } x_i \text{ and } x_j \text{ belong to same class} \\ -1, & \text{if } x_i \text{ and } x_j \text{ belong to different classes} \\ 0, & \text{otherwise} \end{cases}$

The loss function in Eq. 5.4 can be rewritten as:

$$J_{SMNAE} = J_1 + \lambda J_2 + \beta J_3 \tag{5.5}$$

where, J_1 is the first part denoting the traditional autoencoder loss function, J_2 is the middle term belonging to the $\ell_{2,p}$ norm, and J_3 is the last term denoting the pairwise class-based discrimination term. Here, J_1 and J_3 are convex and smooth, whereas J_2 is non-convex. This equation can be converted to:

$$\underset{\mathbf{W}}{\operatorname{arg\,min}} J_{SMNAE}(\mathbf{W}) = f(\mathbf{W}; \mathbf{X}) + \lambda \sum_{c=1}^{C} \| \mathbf{W} \mathbf{X}_{\mathbf{c}} \|_{2, p}$$
(5.6)

where, $f(\mathbf{W}; \mathbf{X})$ is a convex function consisting of J_1 and J_3 and whose gradient is Lipschitz continuous. The proximal gradient algorithm [114, 115] is applied to solve the optimization problem shown in Eq. 5.6. Beginning with an initial value of \mathbf{W}_0 , \mathbf{W} is updated iteratively using $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_t$. Thus, the proximal operator equation becomes the following and can then be solved using the Newton's method [116]:

$$\mathbf{W}_{t+1} = \operatorname*{arg\,min}_{\mathbf{W}} \frac{1}{2\eta} \| \mathbf{W} - \mathbf{A} \|_{F}^{2} + \lambda \sum_{c=1}^{C} \| \mathbf{W} \mathbf{X}_{c} \|_{2,p}$$
(5.7)

where, $\mathbf{A} = \mathbf{W}_t - \eta \nabla f(\mathbf{W}_t)$ represents the update for \mathbf{W}_t . Considering linearity, $\nabla f(\mathbf{W}_t)$ can be represented as

$$\nabla f(\mathbf{W}_{t}) = \frac{\partial J_{1}}{\partial \mathbf{W}_{t}} + \frac{\partial J_{3}}{\partial \mathbf{W}_{t}}$$

$$= -2\mathbf{X}[\mathbf{X} - \mathbf{W}'\mathbf{W}\mathbf{X}] + 2\mathbf{X}\beta\mathbf{W}\mathbf{X}\mathbf{L}_{i}$$
(5.8)



Figure 5.2: Proposed three-stage kinship verification in unconstrained videos framework by utilizing mixed norm supervised autoencoder (SMNAE). In the first stage, a pair of videos is split into vidlets which are provided as input to stacked SMNAE. The spatial representations learned from the first stage are concatenated in a pairwise fashion and provided to the second stage stacked SMNAE. The third stage learns the kinship-specific representation of the vidlets using stacked SMNAE to encode the spatio-temporal information which is employed as the input to SVM for kin vs. nonkin video classification.

Next, we utilize this proposed formulation of SMNAE to develop a framework for unconstrained video-based kinship verification. The $\ell_{2,p}$ norm on the weights of the autoencoders introduces sparsity for kin and non-kin class individually where the value of p can be tuned to obtain the optimum amount of sparsity in the features. The pairwise constraint terms enforce better discrimination between the kin and non-kin class representations.

5.2.2 SMNAE for Kinship Verification in Unconstrained Videos

Using SMNAE as the integral representation learning unit, we propose a framework for verifying kinship relation in a pair of videos as shown in Figure 5.1. Each video in the input pair of videos is decomposed into *vidlets*. A vidlet is defined as a nonoverlapping contiguous group of frames with z number of neighboring frames before and after the pivot frame. The center frame of the vidlet is the pivot frame and the variations across the small temporal neighborhood of the pivot frame are learned. In addition, the kinship-specific spatial information in every frame is encoded. A pair of vidlets is passed as input to the proposed SMNAE based framework. The framework classifies the vidlet pair as kin or non-kin. This process is repeated for every vidlet pair. Fusion of these vidlet scores is performed using sum rule fusion to produce the final classification of the input pair of videos.

Figure 5.2 illustrates the detailed architecture of vidlet feature learning and classification in the proposed kinship verification framework. The training of the proposed three-stage SMNAE framework is performed in a modular fashion where the first stage learns the spatial kinship representation of images in each frame. The second stage training learns the pairwise spatio-temporal kinship representation of frames in a vidlet, while the third stage training learns the cumulative vidlet kinship representation, aggregating both the spatial and temporal features. Let Vid_i and Vid_j be the videos of subjects i and j respectively. The individual frames of the videos of subject i are denoted as $Vid_i = [Fr_i^1, \ldots, Fr_i^n]$. The input to the proposed framework is a pair of vidlets from both the videos. The three stages of the proposed framework are discussed in detail below.

First Stage - Learning Spatial Kinship Representations

In the first stage of the proposed framework, compact kinship-specific representations are learned from frame-level spatial information. The input pair of videos is treated as a stream of vidlets. Each vidlet comprises frames $[Fr_i^{(piv-z)}, \ldots, Fr_i^{piv}, \ldots, Fr_i^{(piv+z)}]$ and $[Fr_j^{(piv-z)}, \ldots, Fr_j^{piv}, \ldots, Fr_j^{(piv+z)}]$, where Fr_i^{piv} is the pivot frame. The corresponding frames of the pair of vidlets are concatenated and presented as input to stacked SMNAE to learn the spatial representation of kinship in the images. In this first stage,

- Input: $\mathbb{X} = [Fr_i^k, Fr_j^k] \forall k \text{ in } \{1, \dots, (2 * z + 1)\}$, and the regularization parameters λ and β .
- Output: Learned weight matrix **W** and $\mathbf{H} = \phi(\mathbf{W}X)$
- Loss function (J_F) :

$$J_F = \underset{\mathbf{W},\mathbf{W}'}{\operatorname{arg\,min}} \| \mathbb{X} - \phi(\mathbf{W}'\mathbf{H}) \|_F^2 + \lambda \sum_{c=1}^C \| \mathbf{W} \mathbb{X}_{\mathbf{c}} \|_{2,p} + \beta(\operatorname{Tr}(\mathbf{H}^T\mathbf{HL}))$$
(5.9)

For all three stages, $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_N), d_i = \sum_{j=1}^N M_{i,j}, \mathbf{L} = \mathbf{D} - \mathbf{M}$, and **M** is created such that

$$M_{i,j} = \begin{cases} +1, & \text{if Vid}_i \text{ and Vid}_j \text{ belong to kin} \\ -1, & \text{if Vid}_i \text{ and Vid}_j \text{ belong to non-kin} \\ 0, & \text{otherwise} \end{cases}$$

Second Stage - Learning Pairwise Spatio-temporal Kinship Representations in Temporal Neighborhood

To account for spatial changes across a small temporal neighborhood, a second stacked SMNAE is trained, greedily layer by layer. The second stage SMNAE is trained to learn and encode variations between a pivot frame and its neighboring frames. These encodings facilitate learning a robust representation between kin pairs across variations due to occlusion, illumination, pose, and expression changes.

The output of first stage is the learned representation of the pair of (2 * z + 1)frames of the vidlet and is denoted by H^k , $\forall k$ in $\{1, \ldots, (2*z+1)\}$. In this stage, the pivot index (piv) is chosen and 2 * z pairwise combinations (\mathcal{X}) are formed between the representations of the pivot and neighboring input frames learned from the first stage. In the second stage of the proposed kinship verification framework:

• Input: $\mathcal{X}^o = [H^{piv}H^o]$ such that $o \in \{(piv - z), (piv + z)\} \land o \neq piv$ where $H^k = \phi(\mathbf{W}\mathbb{X}^k) \forall k \text{ in } \{1, \ldots, (2 * z + 1)\}$ and the regularization parameters α and γ .

- Output: Learned weight matrix \mathcal{W} and $\mathcal{H} = \phi(\mathcal{W}\mathcal{X})$
- Loss function (J_S) :

$$J_{S} = \underset{\mathcal{W},\mathcal{W}'}{\operatorname{arg\,min}} \| \mathcal{X} - \phi(\mathcal{W}'\mathcal{H}) \|_{F}^{2} + \alpha \sum_{c=1}^{C} \| \mathcal{W}\mathcal{X}_{c} \|_{2,p} + \gamma(\operatorname{Tr}(\mathcal{H}^{T}\mathcal{H}\mathbf{L}))$$
(5.10)

Third Stage - Learning Kinship-Specific Vidlet Representations

After learning the pairwise spatio-temporal representation of the frames in a vidlet in the second stage, the third stage of the proposed SMNAE based framework encodes the cumulative vidlet kinship representation. It aggregates the spatial and temporal representations to form the final compact representation for the vidlet. This learned representation for a subject is hierarchical in nature, including representation of a single frame, pair of frames, and group of frames.

For the final stage of the proposed kinship verification in unconstrained videos framework using SMNAE autoencoders:

- Input: $\mathcal{V} = [\mathcal{H}^1, \dots, \mathcal{H}^{(2*z)}]$ and the regularization parameters ζ and κ .
- Output: Learned weight matrix \mathcal{G} and $\Lambda = \phi(\mathcal{GV})$
- Loss function (J_T) :

$$J_{T} = \underset{\mathcal{G},\mathcal{G}'}{\operatorname{arg\,min}} \| \mathcal{V} - \phi(\mathcal{G}'\Lambda) \|_{F}^{2} + \zeta \sum_{c=1}^{C} \| \mathcal{G}\mathcal{V}_{c} \|_{2,p} + \kappa(\operatorname{Tr}(\Lambda^{T}\Lambda\mathbf{L}))$$
(5.11)

For a pair of test videos, similar steps are followed where the given pair of videos is first converted into vidlets. If two videos have different lengths, the frames of the shorter video are iteratively cycled to match the number of frames of the longer video. The vidlets are passed through the three stages of stacked SMNAEs of the proposed video-based kinship verification deep learning framework. The representation of all the vidlets is learned and provided to the SVM for classification.

Classification

The final spatio-temporal representation of streams of vidlets is obtained using the third stage learned weight matrix, \mathcal{G} . This learned representation Λ for each vidlet is

provided to a Support Vector Machine (SVM) [117] with radial basis function kernel for computing the decision boundary between the kin and non-kin class. The final classification score for the input pair of videos is a sum of the probability scores of each vidlet obtained as output from the SVM.

5.2.3 Implementation Details

Every video is pre-processed and contiguous frames are extracted. Face detection and alignment is performed on each frame of the video. This is followed by vidlet creation. For experimental purposes, three different values of z (pivot's neighboring frames) are analyzed with z = 1, z = 2, and z = 3.

In the first stage of the proposed kinship verification framework, three Supervised Mixed Norm autoencoders (SMNAEs) are stacked together, each containing 8192, 4096, and 2048 hidden nodes in their layers respectively. The spatial representation learned from the first stage of the proposed framework is used to extract the pairwise frame representation. This is used as input to the second stage of the proposed framework.

In the second stage of Figure 5.2, the third frame pair in the vidlet represents the pivot with z = 2. The spatial kinship representations learned from the first stage are concatenated in a pairwise fashion and sent to the second stage stacked SMNAEs. The second stage stacked SMNAE consists of two autoencoders, each containing 2304 and 1024 hidden nodes. Finally, all the pairwise spatio-temporal representations learned from the second stage of the proposed kinship verification framework are concatenated to form the input to the stacked autoencoder in the third stage.

The final stage learns the kinship-specific representation of the vidlets which aggregates the spatial and temporal information. The stacked SMNAE autoencoder used in this stage consists of two autoencoders, each containing 3072 and 2048 hidden nodes. A vector of size 2048 is extracted for each vidlet and is used as input to the SVM classifier. The score for the complete video is calculated as the sum of the probability scores of all the vidlets from the binary SVM classifier.

5.3 Kinship Video (KIVI) Database

Currently, only one kinship video database (UvA-NEMO Smile [63]) exists in the literature. This database was initially created for capturing smile and expression variations. However, it was captured in controlled settings with limited real-world



Figure 5.3: Sample kin-pair subjects from Kinship Video (KIVI) face database comprising of seven kin-relations. The number of respective kin pairs present in each kin-relations category in KIVI database is also indicated.

variations and is not suitable for kinship verification in uncontrolled videos. Therefore, to include realistic variations observed in the *wild* we collected the largest unconstrained kinship video (KIVI) face database.

The KIVI database consists of video sequences of 503 individuals forming a total of 355 positive kin pairs from 211 families. For creating the proposed KIVI database, popular celebrity families are selected and the ground truth is annotated manually by validating the kinship information from the Internet. For each subject in the database, there is one corresponding video. The average video duration is 18.78 seconds with the average frame rate of 26.79 frames per second (fps) and the total number of still frames in the database is over 250,000. The database includes seven kin-relations which are illustrated in Figure 5.3.

No. of Individuals	503	
	7 (Brother-Brother, Sister-Sister,	
No. of Kin Relations	Brother-Sister, Father-Son,	
	Father-Daughter, Mother-Son, and	
	Mother-Daughter)	
No. of Families	211	
Ethnicity of Familias	African-American (9) , Asian (74) , and	
Ethnicity of Fammes	Caucasian (128)	
Average Video Duration	18.78 seconds	
No. of Still Frames	252,804	
Average Frame Rate	26.79 fps	
No. of Video Kin Pairs	355 Positive and 355 Negative	
No. of Image Kin Pairs	212,018 Positive and 212,018 Negative	

Table 5.1: Characteristics of the unconstrained KIVI database.

The database contains videos with variations in illumination, pose, occlusion, ethnicity, and expression collected from the Internet. It has multi-ethnic kin pairs of Asian (74), Caucasian (128), and African-American (9) ethnicity. The characteristics of the KIVI database are summarized in Table 5.1. The KIVI database will be publicly available to the research community at http://iab-rubric.org/resources/ KIVI.html. For performance evaluation, unseen training and testing partitions of KIVI database are created with non-overlapping subjects. 214 kin subject pairs are chosen at random as the test partition while the remaining kin pairs form the train partition. An equal number of non-kin pairs are added in both the partitions.

5.4 Experimental Evaluation

Experimental evaluation of the proposed kinship verification framework in unconstrained videos is performed using the training and testing partitions of the KIVI database. For comparison, Deep+Shallow algorithm proposed by Boutellaa et al. [118] for kinship verification in videos is utilized. They proposed a fusion of VGG-Face and textural features (LBPTOP [119], LPQTOP, and BSIFTOP) and applied SVM for final classification. Additionally, comparative analysis is performed using existing image-based kinship verification algorithms. It is to be noted that these algorithms are developed to operate on still face images and not videos. Therefore, these existing algorithms are applied to each frame followed by score-level fusion of the frame-wise scores to produce the final classification of the video pair. The equal error rate (EER) is calculated for each experiment and the kinship verification accuracy is reported for all experiments as (100 - EER)%. The following image-based kinship verification algorithms are used for comparison purposes and the source codes have been obtained from the authors directly except [45].

- Neighborhood repulsed metric learning (NRML) [38]: A metric learning approach is employed where distance between kin pairs is minimized and non-kin pairs is maximized.
- Chromaticity-based approach [47]: The chrominance distance between a pair of kin images is used for classification.
- Color space approach [45]: Color information in the images is incorporated as feature and BSIF descriptor is computed in HSV color space for kinship verification.
- VGG-Face [120]: The face descriptors from the second last layer of VGG-16 convolutional neural network are extracted for kinship verification followed by cosine distance calculation.
- Stacked Denoising Autoencoder (SDAE): Traditional stacked denoising autoencoders are used to encode the kinship features in image. It is combined with Neural Network (NN) and Support Vector Machine (SVM) to learn the final classification of the input pair of images.
- Kinship verification via representation learning filtered contractive deep belief networks (KVRL-*fc*DBN) [3]: A two-stage deep belief network framework is used for kinship verification. It yields the state-of-the-art performance on five kinship image datasets.

5.4.1 Video-based Kinship Verification Results

Experimental Protocol

The objective of the video-based protocol is to compute the kinship verification performance of the proposed framework on a pair of input videos of the KIVI database. The experiments are performed with 40%-60% randomly created train-test partitions. There is no overlap between subjects or families in the training and testing folds. The test set consists of 214 positive kin pair videos while the train set contains 141 kin pair videos with an equal number of non-kin pair videos in each set. The entire video is processed and classified as kin or non-kin. The performance of existing image-based

	Kinship Verification Algorithm	Accuracy (%)
	Chromaticity based approach [47]	43.98
	Color space approach [45]	52.58
<u>1</u> 20	Deep+Shallow [118]	54.46
stiı	NRML [38]	55.61
E Xi	SDAE Autoencoder with NN	64.48
	SDAE Autoencoder with SVM	66.88
	VGG-Face [120]	70.09
	KVRL- fc DBN [3]	79.91
ed	$L_{2,p}$ Autoencoder with NN	78.51
SOC	$L_{2,p}$ Autoencoder with SVM	79.21
roț	SMNAE based framework with NN	81.78
Ч	SMNAE based framework with SVM	83.18

Table 5.2: Video-based kinship verification accuracy (%) of existing image-based and proposed kinship verification algorithms on KIVI face database.

techniques and the proposed video-based kinship framework using Supervised Mixed Norm autoencoders (SMNAEs) are evaluated on the test set of the KIVI database. As stated earlier, the existing image-based kinship verification algorithms are applied to each frame which is followed by score-level fusion of the frame scores to generate the final kin classification for the video.

Experimental Results

The results for video-based kinship verification using the proposed framework and existing image-based kinship verification algorithms are shown in Table 5.2 and Figure 5.4. The proposed framework with SMNAE and SVM yields the highest kinship verification accuracy of 83.18%. Sample video pairs which are correctly and incorrectly predicted by the proposed SMNAE framework are shown in Fig 5.5. The detailed analysis of the results is described below.

Comparison with existing approaches: From Table 5.2 and Figure 5.4, it is observed that the proposed multi-stage kinship verification in videos framework consisting of SMNAE with SVM as the classifier outperforms existing approaches with 83.18% video-based kinship verification accuracy. It surpasses the accuracy of existing video-based kinship verification algorithm [118] by over 30%.

It is to be noted that the proposed kinship verification framework with SMNAE achieves 3.27% higher accuracy as compared to the current state-of-the-art algorithm for image-based kinship verification algorithm KVRL-*fc*DBN [3] algorithm. This indicates that simple aggregation of frame-wise scores may not encode the spatio-



Figure 5.4: Video-based kinship verification performance of (a) existing image-based algorithms using frame level aggregation and (b) proposed framework using 3-Stage existing SDAE, 3-Stage $\ell_{2,p}$ AE with mixed norm regularization, and the proposed 3-Stage SMNAEs on the KIVI database.

Actual	Predicted Label		
Label	Kin	Non-Kin	
Kin			
Non- Kin	000		

Figure 5.5: Sample video pairs from kin and non-kin classes which are correctly and incorrectly classified by the proposed SMNAE framework.

temporal information present in videos. Hence, this highlights the need for developing video-specific kinship verification algorithms.

Effect of autoencoders in the proposed kinship verification framework: To examine the efficacy of different autoencoders as the representation learning unit in the video based kinship verification framework, the same experiments are performed using traditional stacked denoising autoencoders (SDAE) and $L_{2,p}$ autoencoders in the proposed framework. As observed in Table 5.2, $L_{2,p}$ autoencoder with SVM yields 12.33% higher classification accuracy as compared to traditional stacked denoising autoencoders. However, it is to be noted that SMNAE demonstrates around 4% higher performance than $L_{2,p}$ autoencoder for video-based kinship verification. The increase in performance of SMNAE as compared to $L_{2,p}$ autoencoder illustrates the efficacy of the pairwise class-based reconstructive penalty term in the loss function of SMNAE for verifying kin pair videos.

Effect of number of neighboring frames (z): The number of neighboring frames (z) to be considered with respect to a pivot frame in a vidlet is a crucial parameter. As shown in Figure 5.6, the number of neighboring frames selected in the temporal neighborhood affects the content in different frames thereby influencing the learned spatio-temporal representation. Therefore, we next analyze the effect of z in the proposed SMNAE based kinship verification framework.

The results are computed with z = 1 to z = 3. It is to be noted that z = 0 denotes



Figure 5.6: Variations across the pivot frame with respect to number of neighbors (z).

that there are no neighbors for the pivot frame and hence, is same as the traditional image-based kinship verification experiment. The proposed framework yields 83.18% kinship verification accuracy with z = 2 as compared to 82.24% with z = 1 and 80.14% with z = 3. As shown in Figure 5.6, it can be inferred that when the temporal neighborhood is very small, consisting of just one neighboring frame on both sides of the pivot frame (z = 1), the encoded spatio-temporal yields lower performance as compared to z = 2. With z = 3, the total number of contiguous frames including the pivot frame is 7. As the number of frame increases, the temporal representation is not accurately captured because the frames are not in close proximity of the pivot frame and this decreases the kinship classification performance. Therefore, this experiment illustrates that the best value for z is 2 for the KIVI database.

Effect of sparsity norm (p): For determining suitable value of p in $\ell_{2,p}$ norm term in Eqs. 5.9, 5.10, and 5.11, comparative analysis is performed with different values of p (0.2, 0.4, 0.6, 0.8, and 1). It is observed that the best performance of 83.18% is achieved when p = 0.8 as compared to 77.10% (p = 0.2), 78.50% (p = 0.4), 78.97% (p = 0.6), and 81.13% (p = 1). It is noted that smaller values of p produce lower kinship verification accuracy. This may suggest that with smaller values of p, the learned network becomes too sparse and hence, encoding inadequate information for the classifier. However, it is to be noted that different values of p can produce different performance on other databases.

Role of classifier for final kin vs non-kin decision: The effect of classifier after the third stage of the proposed kinship verification framework is analyzed. The proposed SMNAE based framework with a neural network (NN) classifier yields 1.4% lower accuracy as compared to the performance of SVM with the proposed framework. This result illustrates the effectiveness of SVM in learning the kin vs non-kin classification boundary as compared to the neural network.

Effect of kin-relation: Experimental evaluation is performed on all seven types of kin-relation in the input kin pair video. The classification accuracy of the proposed kinship verification framework for different kin-relations is shown in Table 5.3. It is observed that the mother-son pair is the most difficult to detect while mother-daughter video pairs are verified with the highest accuracy of 92.31%. It has been ascertained that humans are able to verify kin pairs easily if there is at least one female subject in the pair [18]. One reason postulated for this phenomenon is a higher degree of variation in facial images of males. A similar trend is noticed here as brother-brother and father-son kin pair videos have lower kinship verification accuracy as compared to other relationships due to the absence of a female subject. However, mother-son relationship kin pairs contradict this pattern and one reason for that can be the fewer number of samples of this relation as compared to other relationships in the KIVI database.

Role of fusion of vidlets: An input video pair of two subjects produces many vidlets. Therefore, in the proposed kinship verification framework, the final classification score for the input pair of videos is a sum of the probability scores of each vidlet obtained as output from the SVM. For comparative analysis, the performance of the proposed framework is computed by applying the max rule on the probability scores. In the max rule, the vidlet pair with the highest probability score is chosen for providing the final classification. Video-based kinship verification accuracy of 80.37% is obtained with the max rule as compared to 83.18% by using the sum rule. In sum rule fusion, all the vidlet pairs contribute equally and are combined to produce the final decision. The performance is higher as compared to max rule fusion where only one pair is



selected to produce the decision. This demonstrates that it is beneficial to integrate decisions from all the vidlet pairs to produce the final classification of kin or non-kin.

Figure 5.7: Image-based kinship verification performance of (a) existing image-based algorithms (b) first stage of SDAE, $L_{2,p}$ AE, and proposed SMNAE on still frames of KIVI database.

Kin Relationship	Accuracy (%)
Mother-Daughter	92.31
Sister-Sister	83.64
Brother-Sister	82.86
Father-Daughter	81.82
Brother-Brother	81.25
Father-Son	80.00
Mother-Son	77.78

Table 5.3: Video-based kinship verification performance on the seven kin-relations in the KIVI database using the proposed SMNAE based framework.

5.4.2 Image-based Kinship Verification Results

The proposed SMNAE based framework achieves state-of-the-art kinship verification accuracy in videos. However, in certain scenarios images may be the only source of information due to the unavailability of videos. In such scenarios, kinship can only be determined using image pair of the subjects. The performance of the proposed kinship verification framework on images is reported below.

Experimental Protocol

This protocol is created to evaluate the performance of the proposed kinship verification framework at frame-level (image-level) without taking into account the temporal information present in videos. Existing image-based kinship verification algorithms are applied to each frame. Corresponding frames of the input pair of video are extracted and kinship verification is performed on a per-frame basis. On average, the test set consists of 125,708 pairs of positive kin frames while the train partition contains 86,310 pairs of positive kin pair frames. An equal number of non-kin frame pairs are added in both test and training partitions.

Experimental Results

The kinship verification performance is evaluated on per-frame basis and the results are reported in Table 5.4 and Figure 5.7a and 5.7b.

Comparison with existing approaches: The performance of the proposed videobased kinship verification framework is evaluated on images using the first stage of the kinship verification framework as this stage of the framework is intended to encode the spatial features irrespective of the temporal information. It is observed that the first

	Kinship Verification Algorithm	Accuracy (%)
	Color space approach [45]	51.18
	Chromaticity based approach [47]	54.00
Existing	NRML [38]	54.47
	SDAE Autoencoder with NN	58.77
	SDAE Autoencoder with SVM	61.33
	VGG-Face [120]	69.32
	KVRL-fcDBN [3]	77.97
ed	First stage $L_{2,p}$ Autoencoder with NN	76.20
ropose	First stage $L_{2,p}$ Autoencoder with SVM	76.48
	First stage SMNAE with NN	77.41
L L	First stage SMNAE with SVM	79.09

Table 5.4: Image-based kinship verification accuracy of existing image-based and proposed kinship verification algorithms on still frames of the KIVI face database.

Table 5.5: Kinship verification accuracy (%) of existing and proposed kinship verification algorithms on existing kinship datasets. - denotes that the results are not reported in the original paper.

	Image				Video	
Mothod	Cornoll	UBKin	KEWI	KEWH	WVII Kin	UvA-
Method	Cornen	UDIXIII		171, 1411		Nemo
NRML	-	-	69.90	76.50	-	-
PDFL	71.90	67.30	70.10	77.00	-	-
DDMML	-	-	83.50	84.30	-	-
VTCL	-	-	80.50	82.30	-	93.65
KVRL	80.50	01.80	06.10	06.20	00.80	
fcDBN	09.00	91.00	90.10	90.20	90.00	_
Proposed	94.40	95.30	96.90	97.10	93.40	96.07

stage of stacked SMNAE with SVM classifier yields the highest classification accuracy of 79.09% on the KIVI database. Furthermore, it is observed that the state-of-theart algorithm for image-based kinship verification, KVRL-fcDBN [3] yields kinship verification accuracy of 77.97%. In [3], the authors have reported that KVRL-fcDBN achieves more than 90% classification accuracy on five publicly available kinship image databases. This decrease in the performance of KVRL-fcDBN [3] on KIVI database demonstrates the challenging nature of the database due to real-world variations in pose, illumination, occlusion, ethnicity, and expressions.

5.4.3 Performance on Existing Databases

To demonstrate the efficacy of the proposed SMNAE framework on existing publicly available kinship databases, five image-based and one video-based kinship databases are selected. The experimental evaluation on these databases is performed using the pre-defined protocols described in the respective papers. Comparative analysis on these databases is performed using existing kinship verification algorithms and the results are summarized in Table 5.5.

The first stage of the proposed SMNAE based framework, signifying the scenario when z = 0, is fine-tuned and evaluated on the following five publicly available imagebased kinship databases: (i) Cornell [1], (ii) UB KinFace [32], (iii) KinFaceW (KFW)-I [38], (iv) KFW-II [38], and (v) WVU Kin [3]. The performance of the proposed SMNAE is compared with the following existing kinship verification algorithms: (i) NRML [38], (ii) PDFL [42], (iii) DDMML [61], (iv) VTCL [64], and (v) KVRL-*fc*DBN [3]. The analysis of the results reported in Table 5.5 illustrates that the proposed SMNAE outperforms other existing algorithms on all five databases by 0.8% to 4.9% on the task of kinship verification.

The performance of the proposed three-stage SMNAE based framework is also evaluated on the kinship partition of UvA-Nemo Smile dataset [63]. This video dataset consists of 95 subjects having a kin relationship. Following the protocol described by Dibeklioglu et al. [63], the proposed SMNAE framework achieves 96.07% kinship verification accuracy. The three-stage SMNAE framework which encodes spatio-temporal kinship-specific representation outperforms the visual transformation aided technique [64] by 2.42%.

5.5 Conclusion

In this research, a deep learning based framework is proposed to detect for kinship in unconstrained videos. A novel supervised autoencoder (SMNAE) is proposed to learn kin-face representations with mixed norm regularization. This innovative autoencoder formulation introduces class-specific sparsity in the weight matrix for class-based discrimination. The proposed three-stage SMNAE encodes the spatio-temporal kinship cues present in video frames for video-based kinship verification. The comparative analysis of the proposed spatio-temporal kinship verification framework using SM-NAEs with existing state-of-the-art kinship verification algorithms demonstrates the efficacy of the proposed three-stage framework. The framework outperforms existing algorithms on publicly available video-based and five image-based kinship databases. Also, a new kinship video database termed as KIVI database is introduced which contains video sequences of 211 families, 355 true kin pairs, and seven kin-relations with more than 250,000 still frame images.

Chapter

Relationship between Kinship Verification and Face Verification

6.1 Introduction

Both kinship verification and face verification represent a binary classification problem where the input images are a pair of face images. However, kinship verification can be considered as a superset of face verification and a harder problem than face verification. This can be observed from the scenario of intra-class variation in the positive class. In face verification, the samples belonging to positive class belong to the same individual. However, in the problem of kinship verification, the samples belonging to the positive class are from different individuals (kin class). Thus, there is an inherent challenge in bringing the kin samples closer in the learned classifier boundary. Kinship verification can be delineated as two distinct but related problems: (a) kinship verification between different individuals and (b) self-kinship which characterizes age-invariant face verification associated with the same individual. In Chapters 3 and 4, we developed an algorithm to deal with the first situation where the input images belonged to different individuals. In this chapter, we will first consider the second situation, the problem of self-kinship. We will also look at how kinship verification can be used as a soft biometric in the problem of face verification.



Figure 6.1: An example of age-progressed pair of images of an individual. This is also an example of self-kinship as these age-separated face images can be verified using kinship verification algorithms.

6.2 Self-Kinship Problem (Age-Invariant Face Verification)

From an anthropological point of view, kinship information is based on the degree of genetic similarity between two individuals. Thus, it is logical to expect that an ideal kinship verification algorithm will give perfect kinship verification accuracy if the input pair of face images belong to the same individual. These pair of images belonging to the same individual can also be separated by age gap. Therefore, the classical age-invariant face verification research problem [121, 122] can be considered under the umbrella of kinship verification. This scenario is identified as the problem of *self-kinship* where the objective is to identify the age-progressed images of the same individual as kin. Figure 6.1 illustrates the problem of *self-kinship* where the age-progressed face images of an individual can be verified using kinship verification algorithms.

6.2.1 Self-Kinship Experimental Results

To validate the efficacy of KVRL framework in this situation, the experiment is performed on two databases: FG-NET database [123] and UB KinFace dataset[32] using the KVRL-SDAE and KVRL-DBN algorithms. The FG-NET database consists of 1,002 images from 82 subjects with an average of 12 images per subject. The database consists of 5,808 positive intra-class samples and 12,000 negative inter-class



Figure 6.2: ROC curves demonstrating the performance of Self-Kinship (age-invariant face verification) using the KVRL algorithm.

samples. Three-fold cross-validation is performed similar to Lu et al. [38] where each subject appears in the training or testing set exclusively. UB database also consists of images of parents when they were young and when they were old. The database consists of 200 images of individuals in each of the above-mentioned scenarios. A three-fold cross-validation is conducted with each pair present either in the training or testing dataset.

Figure 6.2 shows the results obtained from the self-kinship experiment. An equal error rate (EER) of 14.14% is observed on the UB dataset, while an EER of 16.45% is observed on the FG-Net Database using SDAE in the KVRL framework. A further reduction in EER of 10.95% on the UB dataset and 15.09% on the FG-Net Database is observed when using the DBN model in the KVRL framework. This result is considerably better than the 22.5% observed by the current state-of-the-art algorithm for kinship verification: multi-view neighborhood repulsed metric learning (MNRML) [38]. These algorithms outperform the previous best algorithm in the age-invariant face verification (self-kinship) experiment for the FG-Net Database [123] and perform equally good on the UB KinFace database [32].



Figure 6.3: Humans utilize kinship as a context to identify siblings of famous personalities.

6.3 Boosting Face Verification using Kinship

Soft biometrics modalities lack the individualization characteristics on their own but can be integrated within a verification system that uses the primary biometric trait such as face to boost the accuracy [124]. Soft biometric traits can often be based on association wherein the context of association can be used to increase the recognition performance in challenging image scenarios [125]. In this chapter, we propose kinship as a *context* that can be used as a soft biometric modality to improve the accuracy of face verification. Kinship cues are used by humans in daily life for recognition. For instance, we may recognize a person based on their familiarity with their kin even though we may not have met the person earlier. Such a scenario is depicted in Figure 6.3. To incorporate this context, we propose a formulation to incorporate kinship verification scores generated by the proposed framework to boost the performance of any face verification algorithm.

Figure 6.4 shows how the proposed KVRL-fcDBN framework is used to improve the performance of face verification algorithms using kin-verification scores. This formulation is generic in nature and independent of the kinship verification and face verification algorithms. As shown in Figure 6.5, given a probe face image, face verification score and kinship classification score are computed from the gallery data (claimed identity and associated kin image), which are then used in the proposed formulation. We demonstrate two methods for boosting the performance using Product of Likelihood Ratio (PLR) [126] and Support Vector Machine (SVM) [127].

PLR based Score Boosting Algorithm: Let s be the face matching score obtained by matching a probe image and a gallery image. k₁, k₂,..., k_n represent the kin scores obtained from the probe image and images of the gallery subject. The product of likelihood ratio [128] can be calculated as:



Figure 6.4: Illustrating the steps involved in the proposed context boosting algorithm where kinship verification scores generated from the KVRL framework are used to improve the face verification performance.



Figure 6.5: A probe image can have a match score (s) with an image in the gallery and a kin score (k) with the associated kin in the gallery to boost the face verification performance.

$$PLR = \frac{P(s \mid \omega_1)}{P(s \mid \omega_2)} \times \prod_{i=1}^{N} \frac{P(k_i \mid ks_1)}{P(k_i \mid ks_2)}$$
(6.1)

Here, ks_1 represents the true kin class, ks_2 represents the non-kin class, ω_1 represents the genuine class, ω_2 represents the impostor class. $P(s \mid \omega_1)$ and $P(s \mid \omega_2)$ represent the class conditional probability of the input vector. All four variables are modeled using mixture of Gaussian distributions.

SVM based Score Boosting Algorithm: Let p_i be the feature vector representing the concatenation of the face matching and kin verification scores i.e
 p_i = [s_i k_i]. A support vector machine can be trained on the combined score vector to boost the performance of face verification.

Since we are proposing a generic approach which is independent of the features used for face verification, we have used the commonly explored local binary patterns (LBP) [129] and histogram of oriented gradients (HOG) [130] for face verification.

6.3.1 Experimental Protocol

The WVU Kinship database is divided into training and testing sets. Similar to kinship verification experiments, the training partition consists of 60% of the dataset and the testing partition consists of the remaining 40% where the subjects are mutually independent and disjoint. In both the sets, two images of an individual are used as the probe, while the remaining are used as the gallery. Four images of the kin of the individual are kept in the gallery where the association between the kin in the gallery set is known. The proposed KVRL-fcDBN framework is used to generate the kinship scores between the probes and kin images using the fcDBN deep learning algorithm.

6.3.2 Results of Boosting Face Verification using Kinship as Context

The results from boosting the face verification performance using both PLR and SVM are shown in Figure 6.6. It is observed that HOG descriptor performs better than LBP for face verification on the WVU Kinship dataset. However, for both HOG and LBP, the face verification accuracy increases over 20% when kinship scores obtained using the proposed KVRL-fcDBN framework is used to boost the face verification scores. At 0.01% FAR, a performance of 59.4% is observed by using HOG descriptor. This improves to 79.3% when kinship scores are utilized using fcDBN as context and PLR algorithm is used. Similarly, the performance improves to 80.0% when SVM is used along with fcDBN. The improvement is more pronounced for true positive rate (TPR) at lower values of false positive rate (FPR). It is to be noted that the proposed experiment can be performed with any face verification algorithm or feature descriptor and these results suggest that incorporating kinship as soft biometric information improves the face verification performance.

6.4 Conclusion

In this chapter, we considered the relationship between kinship verification and face verification. The problem of kinship verification can be divided into parts, one where the pair of images belong to two different individuals and second where the pair of images belong to the same individual. We consider the second problem as selfkinship which is also known as age-invariant face verification. The proposed KVRL framework outperforms state-of-the-art self-kinship algorithm on multiple databases. We conclude by designing a system where a kinship verification system can be used to



b ROC using HOG descriptor.

Figure 6.6: ROC curves summarizing the results of Kinship aided Face Verification using PLR and SVM.

boost existing face verification systems. Our experiments demonstrate the viability of utilizing kinship as a context to improve face recognition.

Chapter

Conclusion and Future Work

7.1 Conclusion

Research in kinship verification has grown tremendously over the past five years. It has a broad application across different fields of security, social media, border control, and law. However, many of the initial databases proposed for this research suffer due to the lack of variation between faces. Additionally, limited samples have controlled the type of research that can be conducted in this field. In this dissertation, we have made an attempt to deal with the above-mentioned challenges. The major research contributions are listed below:

- Understanding human performance: A user study was conducted to understand how humans process kinship cues present in faces to make a decision of whether the presented pairs belong to the same family or not. The major results of this study were that females may be better than males in processing kin and binocular regions may not be the best region to provide significant kinship cues.
- KVRL Framework for Image-based kinship verification: A novel deep learning framework was designed to verify kin present in images. To alleviate the problem of limited samples present in existing datasets, representation of a face, suitable for kinship verification, was learned in an unsupervised manner. Utilizing the insights gained from the human study, a novel filter contractive RBM was proposed to learn a compact representation of a face. This representation was later used to finally classify kin using a neural network in an unsu-

pervised manner. The proposed framework outperforms existing algorithms on all existing image kinship datasets.

- SMNAE Framework for Video-based kinship verification: Supervised mixed norm autoencoders were proposed in this research for verifying kin present across two separate video frames. Specifically, a multi-stage deep learning based framework was proposed to utilize the spatio-temporal information present in the videos to verify kinship. Detailed analysis showcases the efficacy of the proposed framework in verifying kin as compared to existing image-based and video-based kinship verification algorithms.
- Database creation: To facilitate research in this field, two databases were created for kinship verification in images and videos in an unconstrained manner. WVUKin database was created which consist of 113 individuals with multiple images per person to account for variations in pose and illumination. KIVI database was also created consisting of 503 individuals and over 250,000 still frames. Both of these databases will be available for researchers to download and utilize in their own research

7.2 Future Research

Some challenges and solutions are discussed below to advance the research in the field of kinship verification:

- We assert that the researchers should now focus on extending the state-of-theart kinship verification performance for unconstrained environments with *wild* variations in occlusion, illumination, pose, ethnicity, and expressions. Majority of the current research focuses on (semi-)constrained images and needs to account for real-world variations in order to design a robust system.
- Additionally, adversarial training can be utilized to learn a discriminative kinship space for image-based kinship verification. Several papers have been proposed for paired training such as style training and cyclic unpaired translation which can also be applied here to learn the relationship between kin faces.
- The performance of video-based kinship verification can be improved by incorporating video frame selection techniques. It can also be improved by training recurrent neural networks to learn a representation of the individual present in

the video across multiple frames. Further, the video database can be extended for the deeper level of relationships, along with the lines of the FIW database. Thus, the database can include new kin-relations spanning across multiple levels of kinship hierarchy.

• There has been considerable interest in testing the bias of deep learning networks towards race and gender. Research can be carried out to test the deep learning networks trained for kinship verification for robustness against different ethnicities and kinship relations.

CHAPTER **C**

Appendix

Additional experiments are performed to evaluate the efficacy of the proposed SMNAE on two standard databases, MNIST and CIFAR-10. For object recognition, since the input is a single image, a minor modification is required. In Eq. 5.4, the parameter, M_{ij} is updated to denote class information and rest of the computation remains the same. The experiments are performed with the pre-defined experimental protocol of the MNIST and CIFAR-10 databases, and the performance of the proposed SMNAE is compared with the traditional KL-divergence based Autoencoder (KLD) and Group Sparse Autoencoder (GSAE) [48]. The results for both the databases are summarized in Table VII. On the MNIST database, SMNAE yields the lowest error of 0.98% as compared to 1.19% by GSAE, and 1.71% by KL-divergence based autoencoder. On CIFAR-10 image database, SMNAE achieves the highest classification accuracy of 79.50%. These experiments further show the usefulness of $\ell_{2,p}$ -norm based SMNAE approach on other object classification problems.

Table 8.1: Performance of the proposed SMNAE, KLD, and GSAE autoencoders on MNIST and CIFAR-10 databases.

Database	Metric	KLD	GSAE	SMNAE
MNIST	Error rate $(\%)$	1.71	1.19	0.98
CIFAR-10	Accuracy (%)	74.30	76.80	79.50

Bibliography

- R. Fang, K. D. Tang, N. Snavely, and T. Chen, "Towards computational models of kinship verification," in *IEEE International Conference on Image Processing*, 2010, pp. 1577–1580.
- [2] R. Fang, A. C. Gallagher, A. Loui, and T. Chen, "Kinship classification by modeling facial feature heredity," in *International Conference on Image Processing*, 2013, pp. 2983–2987.
- [3] N. Kohli, M. Vatsa, R. Singh, A. Noore, and A. Majumdar, "Hierarchical representation learning for kinship verification," *IEEE Transactions on Image Pro*cessing, vol. 26, pp. 289–302, 2017.
- [4] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceed-ings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.
- [5] L. Hogben, "The genetic analysis of familial traits," *Journal of Genetics*, vol. 25, no. 2, pp. 211–240, 1932.
- [6] J. Carsten, *Cultures of Relatedness: New Approaches to the Study of Kinship.* Cambridge University Press, 2000.
- [7] L. H. Morgan, "A conjectural solution of the origin of the classificatory system of relationship," *Proceedings of the American Academy of Arts and Sciences*, vol. 7, 1868.
- [8] C. M. Berman and B. Chapais, *Kinship and behavior in primates*. Oxford University Press, 2004.

- [9] M. Daly and M. I. Wilson, "Whom are newborn babies said to resemble?" *Ethology and Sociobiology*, vol. 3, no. 2, pp. 69–78, 1982.
- [10] N. J. Christenfeld and E. A. Hill, "Whose baby are you?" Nature, vol. 378, no. 6558, 1995.
- [11] S. Bredart and R. M. French, "Do babies resemble their fathers more than their mothers? A failure to replicate Christenfeld and Hill (1995)," *Evolution and Human Behavior*, vol. 2, pp. 129–135, 1999.
- [12] P. Bressan and M. F. Martello, "Talis pater, talis filius: perceived resemblance and the belief in genetic relatedness," *Psychological Science*, vol. 13, no. 3, pp. 213–218, 2002.
- [13] R. Burch and G. Gallup, "Perceptions of paternal resemblance predict family violence." *Evolutionary Human Behavior*, vol. 21, no. 6, pp. 429–435, 2000.
- [14] S. M. Platek, J. P. Keenan, G. G. Gallup, and F. B. Mohamed, "Where am I? The neurological correlates of self and other," *Brain research: Cognitive brain research*, vol. 19, no. 2, pp. 114–122, 2004.
- [15] L. T. Maloney and M. F. Dal Martello, "Kin recognition and the perceived facial similarity of children," *Journal of Vision*, vol. 6, no. 10, 2006.
- [16] M. F. Dal Martello and L. T. Maloney, "Where are kin recognition signals in the human face?" *Journal of Vision*, vol. 6, no. 12, 2006.
- [17] —, "Lateralization of kin recognition signals in the human face," *Journal of Vision*, vol. 10, no. 8, 2010.
- [18] G. Kaminski, S. Dridi, C. Graff, and E. Gentaz, "Human ability to detect kinship in strangers' faces: effects of the degree of relatedness," *The Royal Society Biological Sciences*, vol. 276, pp. 3193–3200, 2009.
- [19] D. McLain, D. Setters, M. P. Moulton, and A. E. Pratt, "Ascription of resemblance of newborns by parents and nonrelatives," *Evolution and Human Behavior*, vol. 21, no. 1, pp. 11 – 23, 2000.
- [20] R. Oda, A. Matsumoto-Oda, and O. Kurashima, "Effects of belief in genetic relatedness on resemblance judgments by Japanese raters," *Evolution and Human Behavior*, vol. 26, no. 5, pp. 441 – 450, 2005.
- [21] D. Lieberman, J. Tooby, and L. Cosmides, "The architecture of human kin detection." *Nature*, vol. 445, no. 7129, pp. 727–31, 2007.
- [22] G. Kaminski, F. Ravary, C. Graff, and E. Gentaz, "Firstborns' disadvantage in kinship detection," *Psychological Science*, vol. 21, no. 12, pp. 1746–1750, 2010.
- [23] S. M. Platek and S. M. Kemp, "Is family special to the brain? An event-related fMRI study of familiar, familial, and self-face recognition," *Neuropsychologia*, vol. 47, no. 3, pp. 849 – 858, 2009.
- [24] A. Kofman, "The Troubling Rise of Rapid DNA Testing," goo.gl/UuhZSN, 2016, [Online; accessed November 14, 2016].
- [25] K. Wagner, "Facebook says video is huge 100-million-hours-per-day huge," goo.gl/LvZb51, 2017, [Online; accessed March 14, 2017].
- [26] NIST, "Face Recognition Vendor Test (FRVT) Ongoing," https://www.nist. gov/programs-projects/face-recognition-vendor-test-frvt-ongoing/, 2011.
- [27] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1891–1898.
- [28] E. Zhou, Z. Cao, and Q. Yin, "Naive-deep face recognition: Touching the limit of LFW benchmark or not?" *arXiv preprint arXiv:1501.04690*, 2015.
- [29] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018.
- [30] K. Zhang, Y. Huang, C. Song, H. Wu, and L. Wang, "Kinship verification with deep convolutional neural networks," in *Proceedings of the British Machine Vision Conference*. BMVA Press, September 2015, pp. 148.1–148.12.
- [31] Y. Li, J. Zeng, J. Zhang, A. Dai, M. Kan, S. Shan, and X. Chen, "Kinnet: Fine-to-coarse deep metric learning for kinship verification," in *Proceedings of the Workshop on Recognizing Families In the Wild*, 2017, pp. 13–20.
- [32] S. Xia, M. Shao, and Y. Fu, "Kinship verification through transfer learning," in International Joint Conference on Artificial Intelligence, 2011, pp. 2539–2544.

- [33] M. Shao, S. Xia, and Y. Fu, "Genealogical face recognition based on UB KinFace database," in *IEEE Computer Vision and Pattern Recognition Workshops*, June 2011, pp. 60–65.
- [34] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding kin relationships in a photo," *IEEE Transactions on Multimedia*, vol. 14, no. 4, pp. 1046–1056, 2012.
- [35] X. Zhou, J. Hu, J. Lu, Y. Shang, and Y. Guan, "Kinship verification from facial images under uncontrolled conditions," in ACM Multimedia, 2011, pp. 953–956.
- [36] X. Zhou, J. Lu, J. Hu, and Y. Shang, "Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments," in ACM Multimedia, 2012, pp. 725–728.
- [37] N. Kohli, R. Singh, and M. Vatsa, "Self-similarity representation of weber faces for kinship classification," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2012, pp. 245–250.
- [38] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou, "Neighborhood repulsed metric learning for kinship verification," *IEEE Transactions on Pattern Analysis* and Machine Intelligence, vol. 36, no. 2, pp. 331–345, 2014.
- [39] J. Hu, J. Lu, J. Yuan, and Y.-P. Tan, "Large margin multi-metric learning for face and kinship verification in the wild," in Asian Conference on Computer Vision. Springer, 2014, pp. 252–267.
- [40] H. Yan, J. Lu, W. Deng, and X. Zhou, "Discriminative multimetric learning for kinship verification," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1169–1178, 2014.
- [41] Y. Guo, H. Dibeklioglu, and L. van der Maaten, "Graph-based kinship recognition," in *IEEE International Conference on Pattern Recognition*, 2014, pp. 4287–4292.
- [42] H. Yan, J. Lu, and X. Zhou, "Prototype-based discriminative feature learning for kinship verification," *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1–1, 2014.
- [43] Q. Liu, A. Puthenputhussery, and C. Liu, "Inheritable Fisher vector feature for kinship verification," in *IEEE International Conference on Biometrics Theory*, *Applications and Systems*, 2015, pp. 1–6.

- [44] P. Alirezazadeh, A. Fathi, and F. Abdali-Mohammadi, "A genetic algorithmbased feature selection for kinship verification," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2459–2463, 2015.
- [45] X. Wu, E. Boutellaa, M. B. López, X. Feng, and A. Hadid, "On the usefulness of color for kinship verification from face images," in *IEEE International Workshop* on Information Forensics and Security, 2016, pp. 1–6.
- [46] H. Yan, "Kinship verification using neighborhood repulsed correlation metric learning," *Image and Vision Computing*, vol. 60, pp. 91–97, 2017.
- [47] M. B. Lopez, E. Boutellaa, and A. Hadid, "Comments on the "Kinship Face in the Wild" data sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2342–2344, 2016.
- [48] M. Xu and Y. Shang, "Kinship measurement on face images by structured similarity fusion," *IEEE Access*, vol. 4, pp. 10280–10287, 2016.
- [49] A. Puthenputhussery, Q. Liu, and C. Liu, "SIFT flow based genetic fisher vector feature for kinship verification," in *International Conference on Image Process*ing, 2016, pp. 2921–2925.
- [50] X. Qin, X. Tan, and S. Chen, "Tri-subject kinship verification: Understanding the core of a family," *IEEE Transactions on Multimedia*, vol. 17, no. 10, pp. 1855–1867, 2015.
- [51] X. Zhou, Y. Shang, H. Yan, and G. Guo, "Ensemble similarity learning for kinship verification from facial images in the wild," *Information Fusion*, vol. 32, pp. 40–48, 2016.
- [52] H. Liu, J. Cheng, and F. Wang, "Kinship verification based on status-aware projection learning," in *IEEE International Conference on Image Processing*, 2017, pp. 1072–1076.
- [53] S. Mahpod and Y. Keller, "Kinship verification using multiview hybrid distance learning," Computer Vision and Image Understanding, vol. 167, pp. 28 – 36, 2018.
- [54] G. Guo and X. Wang, "Kinship measurement on salient facial features," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2322 –2325, 2012.

- [55] Q. Liu, A. Puthenputhussery, and C. Liu, "Inheritable fisher vector feature for kinship verification," in *IEEE International Conference on Biometrics: Theory*, *Applications and Systems*, 2015, pp. 1–6.
- [56] F. S. Khan, R. Muhammad Anwer, J. Weijer, A. D. Bagdanov, A. M. Lopez, and M. Felsberg, "Coloring action recognition in still images," *International Journal of Computer Vision*, vol. 105, no. 3, 2013.
- [57] A. Dehghan, E. G. Ortiz, R. Villegas, and M. Shah, "Who do I look like? determining parent-offspring resemblance via gated autoencoders," in *IEEE Computer Vision and Pattern Recognition*, 2014, pp. 1757–1764.
- [58] J. P. Robinson, M. Shao, Y. Wu, and Y. Fu, "Families in the wild (FIW): Largescale kinship image database and benchmarks," in ACM Multimedia, 2016, pp. 242–246.
- [59] L. Li, X. Feng, X. Wu, Z. Xia, and A. Hadid, "Kinship verification from faces via similarity metric based convolutional neural network," in *International Conference on Image Analysis and Recognition*, 2016, pp. 539–548.
- [60] S. Wang, J. P. Robinson, and Y. Fu, "Kinship verification on families in the wild with marginalized denoising metric learning," in *IEEE International Conference* on Automatic Face & Gesture Recognition, 2017, pp. 216–221.
- [61] J. Lu, J. Hu, and Y.-P. Tan, "Discriminative deep metric learning for face and kinship verification," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4269–4282, 2017.
- [62] Q. Duan, L. Zhang, and W. Zuo, "From face recognition to kinship verification: An adaptation approach," in *IEEE International Conference on Computer Vision Workshops*, Oct 2017.
- [63] H. Dibeklioglu, A. A. Salah, and T. Gevers, "Like father, like son: Facial expression dynamics for kinship verification," in *IEEE International Conference on Computer Vision*, 2013, pp. 1497–1504.
- [64] H. Dibeklioglu, "Visual transformation aided contrastive learning for videobased kinship verification," in *IEEE Conference on International Conference* on Computer Vision, 2017, pp. 2459–2468.

- [65] N. Kohli, D. Yadav, M. Vatsa, R. Singh, and A. Noore, "Supervised mixed norm autoencoder for kinship verification in unconstrained videos," *IEEE Transactions on Image Processing*, pp. 1–13, 2018.
- [66] H. Dibeklioglu, A. Salah, and T. Gevers, "Like father, like son: Facial expression dynamics for kinship verification," in *IEEE International Conference on Computer Vision*, Dec 2013, pp. 1497–1504.
- [67] M. Dawson, A. Zisserman, and C. Nellåker, "From same photo: Cheating on visual kinship challenges," in Asian Conference on Computer Vision. Springer, 2018, pp. 1–11.
- [68] L. Taylor and R. Sussman, "A preliminary study of kinship and social organization in a semi-free-ranging group of Lemur catta," *International Journal of Primatology*, vol. 6, pp. 601–614, 1985.
- [69] G. Murdock, *Social Structure*. Macmillan, 1949.
- [70] D. L. Pastor, "The quality of mother-infant attachment and its relationship to toddlers' initial sociability with peers." *Developmental Psychology*, vol. 17, no. 3, p. 326, 1981.
- [71] K. Kelly, A. Slade, and J. F. Grienenberger, "Maternal reflective functioning, mother-infant affective communication, and infant attachment: Exploring the link between mental states and observed caregiving behavior in the intergenerational transmission of attachment," *Attachment & human development*, vol. 7, no. 3, pp. 299–311, 2005.
- [72] M. Noriuchi, Y. Kikuchi, and A. Senoo, "The functional neuroanatomy of maternal love: mother's response to infant's attachment behaviors," *Biological psychiatry*, vol. 63, no. 4, pp. 415–423, 2008.
- [73] C. Lévi-Strauss, The Elementary Structures of Kinship:, ser. Beacon Paperback. no. BP 340. Beacon Press, 1969. [Online]. Available: https: //books.google.com/books?id=4QhYYQi6CoMC
- [74] K. Finkler, C. Skrzynia, and J. P. Evans, "The new genetics and its consequences for family, kinship, medicine and medical genetics," *Social Science & Medicine*, vol. 57, no. 3, pp. 403 – 412, 2003.

- [75] M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?" *Perspectives on Psychologi*cal Science, vol. 6, no. 1, pp. 3–5, 2011.
- [76] G. Somanath, M. V. Rohith, and C. Kambhamettu, "Vadana: A dense dataset for facial image analysis," in *International Conference on Computer Vision* Workshops, 2011, pp. 2175–2182.
- [77] G. Somanath and C. Kambhamettu, "Can faces verify blood-relations?" in Biometrics: Theory, Applications and Systems, 2012, pp. 105–112.
- [78] J. Rehnman and A. Herlitz, "Women remember more faces than men do." Acta Psychologica, vol. 124, no. 3, pp. 344–355, 2007.
- [79] —, "Higher face recognition ability in girls: Magnified by own-sex and ownethnicity bias," *Memory*, vol. 14, no. 3, pp. 289–296, 2006.
- [80] A. Herlitz and J. Loven, "Sex differences and the own-gender bias in face recognition: A meta-analytic review," Visual Cognition, vol. 21, no. 9-10, pp. 1306– 1336, 2013.
- [81] T. Susilo, L. Germine, and B. Duchaine, "Face recognition ability matures late: evidence from individual differences in young adults," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 39, no. 5, pp. 1212–1217, 2013.
- [82] J. Fleiss, B. Levin, and M. Paik, Statistical Methods for Rates and Proportions, ser. Wiley Series in Probability and Statistics. Wiley, 2004.
- [83] A. Herlitz, L.-G. Nilsson, and L. Backman, "Gender differences in episodic memory," *Memory and Cognition*, vol. 25, no. 6, pp. 801–811, 1997.
- [84] A. Alvergne, R. Oda, C. Faurie, A. Matsumoto, V. Durand, and M. Raymond, "Cross-cultural perceptions of facial resemblance between kin," *Journal of Vision*, vol. 9, no. 6, 2009.
- [85] C. Lewin and A. Herlitz, "Sex differences in face recognition-Women's faces make the difference," *Brain and Cognition*, vol. 50, no. 1, pp. 121–128, 2002.
- [86] D. B. Wright and B. Sladden, "An own gender bias and the importance of hair in face recognition," Acta Psychologica, vol. 114, no. 1, pp. 101–114, 2003.

- [87] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM Computational Survey, vol. 35, no. 4, pp. 399–458, 2003.
- [88] M. S. Keil, "I look in your eyes, honey : Internal face features induce spatial frequency preference for human face processing," *PLoS Computational Biology*, vol. 5, no. 3, 2009.
- [89] G. E. Hinton, "Learning distributed representations of concepts," in *Conference* of the Cognitive Science Society, vol. 1, 1986.
- [90] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.
- [91] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "Joint learning of words and meaning representations for open-text semantic parsing," in *International Conference on Artificial Intelligence and Statistics*, 2012, pp. 127–135.
- [92] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *International Conference on Machine Learning*, 2011, pp. 513–520.
- [93] G. Dahl, A. R. Mohamed, and G. E. Hinton, "Phone recognition with the meancovariance restricted boltzmann machine," in Advances in Neural Information Processing Systems, 2010, pp. 469–477.
- [94] G. Carneiro, J. Nascimento, and A. Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," in *IEEE International Conference on Image Processing*, vol. 21, no. 3, 2012, pp. 968–982.
- [95] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [96] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
- [97] Y. Bengio, "Learning deep architectures for AI," Foundations and trends in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009.

- [98] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [99] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations* in the Microstructure of Cognition, vol. 1, 1986, pp. 318–362.
- [100] A. Ng, "Sparse Autoencoders," http://web.stanford.edu/class/cs294a/ sparseAutoencoder.pdf/, 2011.
- [101] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [102] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive autoencoders: Explicit invariance during feature extraction," in *Proceedings of the* 28th International Conference on Machine Learning, 2011, pp. 833–840.
- [103] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th Annual International Conference on Machine Learning.* ACM, 2009, pp. 609–616.
- [104] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.
- [105] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *IEEE Computer Vision and Pattern Recognition*, 2011, pp. 529–534.
- [106] 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.
- [107] P. Viola and M. J. Jones, "Robust real-time face detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.
- [108] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Differentiating spontaneous from posed facial expressions within a generic facial expression recognition

framework," in *IEEE International Conference on Computer Vision Workshops*, 2011, pp. 868–875.

- [109] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 10, pp. 2108–2118, 2015.
- [110] A. Majumdar, R. Singh, and M. Vatsa, "Face verification via class sparsity based supervised encoding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1273–1280, 2017.
- [111] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.
- [112] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3869–3872.
- [113] Y. Yan, Z. Xu, G. Liu, Z. Ma, and N. Sebe, "Glocal structural feature selection with sparsity for multimedia data understanding," in ACM International Conference on Multimedia, 2013, pp. 537–540.
- [114] R. Jenatton, J. Mairal, F. R. Bach, and G. R. Obozinski, "Proximal methods for sparse hierarchical dictionary learning," in *International Conference on Machine Learning*, 2010, pp. 487–494.
- [115] Y. Nesterov, "Gradient methods for minimizing composite objective function," UCL, Tech. Rep., 2007.
- [116] M. Zhang, C. Ding, Y. Zhang, and F. Nie, "Feature selection at the discrete limit," in AAAI Conference on Artificial Intelligence, 2014, pp. 1355–1361.
- [117] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.
- [118] E. Boutellaa, M. B. López, S. Ait-Aoudia, X. Feng, and A. Hadid, "Kinship verification from videos using spatio-temporal texture features and deep learning," in *International Conference on Biometrics*, 2016, pp. 1–7.
- [119] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, 2007.

- [120] O. Parkhi, V. Andrea, and Z. Andrew, "Deep face recognition," in Proceedings of the British Machine Vision Conference, 2015, pp. 41.1–41.12.
- [121] D. Yadav, M. Vatsa, R. Singh, and M. Tistarelli, "Bacteria foraging fusion for face recognition across age progression," in *IEEE Computer Vision and Pattern Recognition Workshops*, 2013, pp. 173–179.
- [122] D. Yadav, R. Singh, M. Vatsa, and A. Noore, "Recognizing age-separated face images: Humans and machines," *PLOS ONE*, vol. 9, no. 12, pp. 1–22, 12 2014.
- [123] A. Lanitis, "Comparative evaluation of automatic age-progression methodologies," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 101:1–101:10, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1155/2008/239480
- [124] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Biometric Authentication*. Springer, 2004, pp. 731– 738.
- [125] S. Bharadwaj, M. Vatsa, and R. Singh, "Aiding face recognition with social context association rule based re-ranking," *International Conference on Biometrics*, 2014.
- [126] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification (2Nd Edition). Wiley-Interscience, 2000.
- [127] V. N. Vapnik and V. Vapnik, Statistical learning theory. Wiley New York, 1998, vol. 1.
- [128] 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.
- [129] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Anal*ysis and Machine Intelligence, vol. 28, no. 12, pp. 2037–2041, 2006.
- [130] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.