REPUBLIC OF TURKEY

AKDENIZ UNIVERSITY

MULTIMODAL BIOMETRIC VERIFICATION SYSTEM USING FUSION OF FACE AND SIGNATURE INFORMATION

Kağan ÖZTÜRK

INSTITUTE OF NATURAL AND APPLIED SCIENCES

DEPARTMENT OF COMPUTER ENGINEERING

MASTER THESIS

JANUARY 2020

ANTALYA

REPUBLIC OF TURKEY

AKDENIZ UNIVERSITY

MULTIMODAL BIOMETRIC VERIFICATION SYSTEM USING FUSION OF FACE AND SIGNATURE INFORMATION

Kağan ÖZTÜRK

INSTITUTE OF NATURAL AND APPLIED SCIENCES

DEPARTMENT OF COMPUTER ENGINEERING

MASTER THESIS

JANUARY 2020

ANTALYA

REPUBLIC OF TURKEY
AKDENIZ UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MULTIMODAL BIOMETRIC VERIFICATION SYSTEM USING FUSION OF FACE AND SIGNATURE INFORMATION

Kağan ÖZTÜRK

DEPARTMENT OF COMPUTER ENGINEERING

MASTER THESIS

This thesis was unanimously accepted by the jury on 03/01/2020

Asst. Prof. Dr. Mustafa Berkay YILMAZ (Supervisor)
Asst. Prof. Dr. Hüseyin Gökhan AKÇAY
Asst. Prof. Dr. Umut TOSUN

ÖZET

YÜZ VE İMZA BİLGİLERİ KULLANARAK ÇOK KİPLİ KULLANICI DOĞRULAMA

Kağan ÖZTÜRK

Yüksek Lisans Tezi, Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Mustafa Berkay YILMAZ

Ocak 2020; 39 sayfa

Biyometrik doğrulama sistemleri, bir kişinin kimliğini doğrulamak için yaygın olarak kullanılır. Yüz, imza, parmak izi ve iris popüler biyometrikler arasındadır. Biyometrik sistemler güvenlik, alışveriş ve finans gibi çeşitli uygulamalarda kullanılırlar. Özellikle adli uygulamalarda çok düşük hata oranlarına sahip olmaları gerekmektedir. Kabul edilebilir sonuçlar elde etmek için çeşitli problemlerle başa çıkmak zorundadırlar.

Tek kipli biyometrik sistemlerin limitlerinin üstesinden gelmek için, çok kipli bir doğrulama sistemi sunulmuştur. İmza ve yüz özellikleri tek kipli biyometrik sistemler oluşturmak için kullanılmıştır. Daha sonra, bu sistemlerin skor seviyeleri kombinasyonu kullanılarak hata oranları azaltılmak istenmiştir. Çok kipli doğrulama sisteminin performansını değerlendirmek için çeşitli saldırı ve gürültü prosedürleri uygulanmıştır.

Bilgisayarlı görü problemlerinde büyük başarı elde eden tekrarlayan ve evrişimsel yapay sinir ağlarının kullanımı çevrimdışı imza doğrulaması için incelenmiştir. İki farklı ağ mimarisinin kombinasyonlarının, düşük hata oranı elde etmek için kullanılabileceği gösterilmiştir.

Kimlik doğrulama için kullanıcı-bağımsız ve kullanıcı-bağımlı yaklaşımlar araştırılmıştır. İki evrişimsel sinir ağı mimarisi, kullanıcı-bağımsız imza öznitelikleri öğrenmek için kullanılmıştır. Daha sonra, kullanıcı-bağımlı sınıflandırıcılar bir kimlik talebini kabul etmek veya reddetmek için eğitilmiştir.

Yüz doğrulama sistemi geliştirmek için transfer öğrenme yaklaşımı kullanılmıştır. Kullanıcı-bağımsız yüz öznitelikleri, eğitilmiş bir evrişimsel sinir ağı kullanılarak çıkarılmıştır. Ardından, kullanıcı-bağımlı sınıflandırıcılar bu öznitelikler kullanılarak doğrulama işlemi yapmak üzere eğitilmişlerdir.

Bu tezde, yüz ve imza doğrulama sistemleri birbirinden bağımsız olarak geliştirilmiş ve performansları değerlendirilmiştir. Daha sonra, bu iki sistemden gelen bilgiler çok kipli bir biyometrik sistem tarafından kaynaştırılmıştır. Sonuçlar, çok kipli yaklaşımın, sistemi saldırılara ve gürültüye karşı daha güçlü hale getirerek, tek kipli sistemlerden daha yüksek doğruluk oranına ulaşmak için kullanılabileceğini göstermektedir.

ANAHTAR KELİMELER: Çok Kipli Biyometrik Sistem, İmza Doğrulama, Skor Kaynaştırma, Yüz Doğrulama

JÜRİ: Dr. Öğr. Üyesi Mustafa Berkay YILMAZ
Dr. Öğr. Üyesi Hüseyin Gökhan AKÇAY
Dr. Öğr. Üyesi Umut TOSUN

ABSTRACT

MULTIMODAL BIOMETRIC VERIFICATION SYSTEM USING FUSION OF FACE AND SIGNATURE INFORMATION

Kağan ÖZTÜRK

MSc Thesis in Computer Engineering
Supervisor: Asst. Prof. Dr. Mustafa Berkay YILMAZ
January 2020; 39 pages

Biometric verification systems are widely used to verify the identity of a person. Face, signature, fingerprint and iris are among popular biometrics. They are used in a wide range of applications such as security, shopping and finance. It is required them to have very low error rates especially in forensic applications. They have to deal with several problems to obtain acceptable results.

In order to overcome limitations of unimodal biometric systems, a multimodal verification system is presented. Signature and face traits are used to build unimodal biometric systems. Then, score level combination of these system is utilized to reach lower error rates. Several attack and noise procedures are applied to evaluate performance of the multimodal verification system.

The usage of recurrent and convolutional neural network architectures, that have achieved great success in a broad range of computer vision tasks, are investigated for offline signature verification. It is shown that, combinations of these two different approaches can be used to achieve state of the art results.

User-independent and user-dependent approaches are investigated to perform authentication. Two convolutional neural network architectures are deployed to learn user-independent signature features. Then, user-dependent classifiers are trained to accept or reject an identity claim.

A transfer learning approach is utilized to develop a face verification system. User-independent face features are extracted from a pre-trained convolutional neural network. Then, these features are fed into user-dependent classifiers to perform verification.

In this thesis, face and signature verification systems are developed and their performances are evaluated separately. Then, a multimodal biometric system, which fuses information coming from two biometrics, is proposed. Results show that, multimodal approach can be used to obtain higher accuracy than unimodal systems and make the system robust against spoof attacks and noise.

KEYWORDS: Face Verification, Multimodal Biometric System, Score Level Fusion, Signature Verification

COMMITTEE: Asst. Prof. Dr. Mustafa Berkay YILMAZ
Asst. Prof. Dr. Hüseyin Gökhan AKÇAY
Asst. Prof. Dr. Umut TOSUN

ACKNOWLEDGEMENTS

I would like to express my very great appreciation to my supervisor Asst. Prof. Dr. Mustafa Berkay YILMAZ for his support, patience and guidance. He has always believed and supported me. I also would like to thank our chairman Prof. Dr. Melih GÜNAY for his encouragement. Finally, I would like to thank my family and friends for their support.

LIST OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols:

μ : Mean

$\mathcal{N}$: Normal Distribution

σ : Standart Deviation

$\mathcal{U}$: Uniform Distribution

Abbreviations:

CNN	: Convolutional Neural Networks
DER	: Distinguishing Error Rate
DMML	: Deep Multitask Metric Learning
EER	: Equal Error Rate
FAR	: False Acceptance Rate
FRR	: False Rejection Rate
HOG	: Histogram of Oriented Gradient
LBP	: Local Binary Patterns
LDA	: Linear Discriminant Analysis
UD	: User-Dependant
UI	: User-Independent
RBP	: Recurrent Binary Pattern
RNN	: Recurrent Neural Network

TEXT OF OATH

I declare that this study "MULTIMODAL BIOMETRIC VERIFICATION SYSTEM USING FUSION OF FACE AND SIGNATURE INFORMATION", which I present as master thesis, is in accordance with the academic rules and ethical conduct. I also declare that I cited and referenced all material and results that are not original to this work.

03/01/2020

Kağan ÖZTÜRK

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Biometrics refers to recognition of a person based on their physiological or behavioral traits. Physiological traits include, face, iris, DNA, palm veins, fingerprint. Behavioral traits are related to patterns of human activities, such as signature and voice.

Biometric systems can be used for identification or verification. Identification systems try to find an identity, based on query sample, among users in database. On the other hand, the aim of verification systems is to accept or reject the claimed identity of a person.

The first step in a biometric verification system is enrollment. Users register to the system by providing several biometric samples. In verification phase, a person, who may either be enrolled to the system or not, claim an identity in the database by giving a query sample. The aim of verification systems is to accept or reject the request by comparing reference samples and query sample of the claimed person.

Biometric verification systems are used for authentication in a wide range of security applications. Unimodal biometric systems use only one of these biometrics to verify the identity. There are several challenges that unimodal systems have to deal with, such as spoofing attacks, intra-class variance, non-universality, noisy data. These challenges can be overcome by multimodal biometric systems making use of multiple biometric traits to make a decision. In this work face and signature traits are examined.

Signature verification is commonly applied technique in legal and financial areas to verify the identity of a person. Depending on the acquisition method, signature verification systems are divided into two categories: online and offline. In online systems, signatures are obtained during the writing phase via an electronic device. In addition to time information of pixel locations in a signature image, angle and pressure of pen can be captured by some devices. In offline case, signatures are obtained after the writing process is finished. A digital image represents a signature sample.

Offline signature verification can be said to be more challenging than online systems. Availability of dynamic information in online systems make the imitation of a signature more difficult. While some simple signatures can be imitated easily by impostor in offline case, additional information in online systems make it more robust to forgeries.

In offline signature literature there are two types of forgeries: random forgery and skilled forgery. In the first case, forger does not have the any information about the shape of

the signature and writer of it. Signatures of other users than being imitated is often considered as random forgeries. In skilled forgeries, impostors have access to some signatures of the user and practice some time to imitate them. While it is easy to detect random forgeries, skilled forgeries make verification task difficult. Genuine signatures of three users can be seen in Figure 1.1. Skilled forgeries of these users is shown in Figure 1.2.

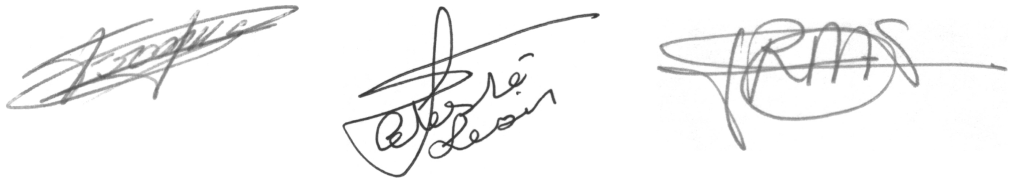

Figure 1.1. Genuine signatures

Figure 1.2. Skilled forgeries

Face recognition is one of the most studied topics in computer vision. Although some biometrics are more reliable than face recognition technology, such as fingerprint and iris, it has become the most popular biometric in recent years. While usage of face identification systems is crucial for governments to maintain public security, face verification systems are mainly used in user authentication applications.

There are three main steps in face recognition systems: face detection, feature extraction and classification. First, the location of a face in an image is found by a detection system. Then, a feature extraction method is applied to reduce the dimension of an image while preserving important features. This step is vital, since the performance of classification step heavily depends on the quality of features. Finally, a classification can either be used for identification or verification.

Face recognition systems have to deal with several challenges: illumination (different level of lighting condition), pose variation (frontal, non-frontal), aging, occlusion (glasses, hat, beard, mustache etc.), expression. While humans are remarkably good at face

recognition without being effected some of these problems, it is troublesome to extract robust features that can overcome these variations in a face verification system.

Classification approaches in verification systems can be divided in two categories: user-independent (UI) and user-dependent (UD) classification. In the first case, a single classifier is developed to make decision for all users. In the latter case, a classifier is trained for each user. It is expected that UD classifiers are capable of catching using specific information and perform better than UI approach. However, complexity of UD systems grows as the number of users increase.

The performance of a verification system can be measured by following evaluation metrics:

- False Acceptance Rate (FAR): FAR shows that a verification system how often falsely accepts an identity.
- False Rejection Rate (FRR): FRR is a measure of the likelihood that a verification system will reject a query by a genuine user.
- Equal Error Rate (EER): EER is found by looking at the point where FAR and FRR are equal.
- Distinguishing Error Rate (DER): DER is the average of FAR and FRR.

A biometric system consists of 4 parts: sensor, feature extraction, matching and decision. Data is acquired via an appropriate device. For example, a camera is used to obtain a face image. Then, features of data are extracted for ease the work of next steps. Proper methods are determined to extract robust features based on data and task. Next, extracted features are compared in matching module and a score is produced. Finally, decision is made by looking the score. For instance, a face verification system gives a similarity score for a query and either accept or reject the request.

Biometric systems can be combined to be more powerful at different levels: feature, matching and decision. An example fusion of face and fingerprint data at different levels can be seen in Figure 1.3.

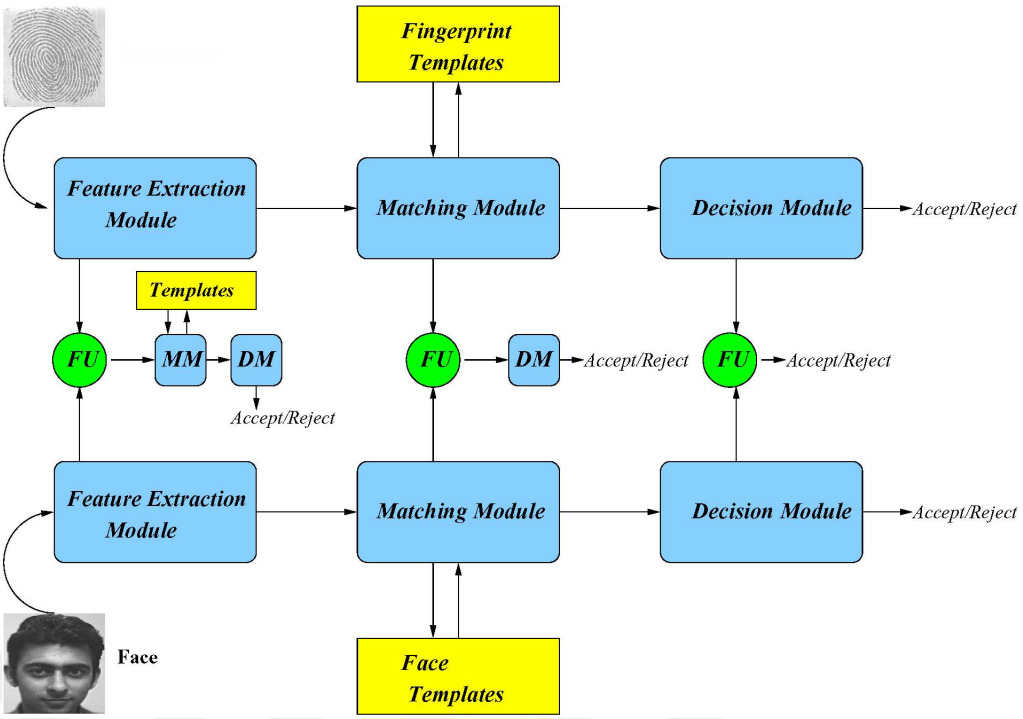

Figure 1.3. Multimodal fusion at different levels (Ross and Jain 2004)

2. LITERATURE REVIEW

While the purpose of recognition systems is to find the identity of a query sample, if it exists in database, the aim of verification systems is to check if a query sample belongs to the claimed identity. Although these systems have different goals, there are many overlapped techniques used by both systems. Hence, literature of recognition systems can be utilized by verification systems.

Face verification and signature verification systems are well-studied topics. These systems have high intra-class variation that makes verification task difficult. Also, they are susceptible to spoof attacks and noise. Multimodal biometric systems can overcome these problems because, they have multiple and independent biometrics. As they have more information about users, they can achieve better accuracy than unimodal systems.

2.1. Convolutional Neural Networks

Convolutional neural networks (CNN) have achieved great success on various computer vision tasks recently. It is first introduced in (LeCun et al. 1998) with the backpropagation learning algorithm. They proposed LeNet-5 for handwritten digit recognition. The architecture of LeNet-5 is shown in Figure 2.1. It is trained using 60000 images of size 32×32.

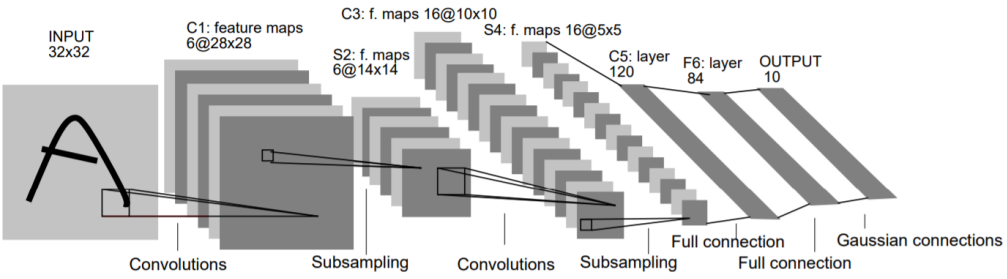

Figure 2.1. LeNet-5 architecture (LeCun et al. 1998)

In (Krizhevsky et al. 2012), a CNN architecture proposed to recognize 1000 objects. 1.2 million training images are used. It has 60 million trainable parameters in 5 convolutional and 3 fully-connected layers. Input size of the network is 224×224. The network is implemented on two graphical processing units. They showed that CNNs can be very powerful in large scale computer vision tasks.

Siamese neural network (Bromley et al. 1994) is a CNN architecture that takes two input images and gives a similarity score. It is composed of two shared layers and at the end of the network they are concatenated to measure the distance. Zagoruyko and Komodakis (2015) propose 2-channel network to compare two images. Images are joined in the third dimension. For image size of $M \times N$, input size of the network is $M \times N \times 2$. They showed that this architecture can achieve better results than the siamese network architecture.

2.2. Face Verification

Recent works, that use convolutional neural networks, improved the performance of face verification systems. Taigman et al. (2014) align faces to a 3D model before fed into the convolutional neural network. CNN is trained on the SFC dataset that includes 4.4 million face images from 4000 people. Then, they train a siamese network, using the trained CNN without the top layer, to classify whether two images belong to the same person or not. The performance of the proposed work on the LFW dataset (Huang et al. 2008) is 97.35% and 91.4% on the YTF dataset (Wolf et al. 2011).

In (Sun et al. 2015), the network is trained with both identification and verification signals to learn features. They add these supervisory signals to early convolutional layers as well as fully connected layers. An ensemble of 25 networks are used for 25 different face patches. They achieved 99.47% and 93.2% verification accuracies on LFW and YTF datasets respectively.

Schroff et al. (2015) propose FaceNet that learns mapping from face images to a Euclidean space where distances show the measure of face similarity. The network is trained to give small distance to face images belonging to the same person and large distance to images of different people. They used triplet loss where three images (anchor, positive and negative) are compared. The goal is to make anchor closer to positive than negative sample. An online triplet selection method is presented for fast convergence. They reported 99.63% accuracy on LFW and 95.1% accuracy on YTF datasets.

Parkhi et al. (2015) propose a data collection method from Internet. After collection procedure, they trained the network to classify 2622 identities each has up to 1000 images. Once the training is done, the top layer is removed and a triplet loss training procedure is employed similar to in (Schroff et al. 2015). It obtains 98.95% accuracy on LFW and

97.3% accuracy on YTF datasets.

Wen et al. (2016) propose a new supervision signal called center loss to classify face images. It learns a center for features of each class and penalizes distances between features and their corresponding class centers. The CNN is trained under the joint supervision of the softmax loss and center loss. The proposed method achieves 99.28% and 94.9% accuracies on LFW and YTF datasets respectively.

2.3. Signature Verification

Signature verification systems are divided into two categories as online and offline. In online systems, signatures are usually obtained by an electronic tablet and pen. Signatures are represented as a sequence over time, containing the position of the pen and depending on acquisition device some additional information can be included such as angle of a pen and pressure. In contrast, offline verification systems obtain signatures after writing process is done via scanners. Thus, they only deal with static digital images of signatures.

Offline Signature Verification has been of great interest to researchers. In (Yılmaz and Yanıkoğlu 2016), scores of user-dependent and user-independent classifiers are fused to make decision on query signatures. Scale invariant feature transform, histogram of oriented gradients (HOG) and local binary patterns (LBP) features are extracted to represent signatures. They achieve 7% EER on GPDS-160 (Ferrer et al. 2005) using 12 reference signatures of subjects.

Hu et al. (2017) proposed a user-independent verification system utilizing fusion of LBP, HoG and statistical gray-level co-occurrence matrix. Random Forest is utilized for classification. They achieve EER of 7.42% on 140 subjects from GPDS-160 database using 12 reference signatures per subject.

Ribeiro et al. (2011) propose restricted boltzmann machines to learn features from signatures. They did not test the performance of features and only report visual representation of weights. In (Khalajzadeh et al. 2012) CNN is used for verification of Persian signatures. They only consider classifying different subjects and did not consider skilled forgeries on test.

Soleimani et al. (2016) propose deep multitask metric learning (DMML) for offline signature verification. DMML consists of shared layer for all subjects and user-specific layers at the end. Given a pair of signatures, it learns a distance metric to determine

whether they belong to the same user or not. It achieves EER of 20.94% on subset of GPDS-960 database (Ferrer et al. 2012) including 300 users.

Hafemann et al. (2017) propose a novel loss function including skilled forgery information to learn writer-independent features. The CNN is trained to classify different users and distinguish genuine and forgery signatures simultaneously. They achieve 1.72% EER on GPDS-160 using 12 references per user.

2.4. Multimodal Biometric

Multimodal biometric systems can be used to achieve better performance than unimodal systems. Fusion of different biometrics can be combined at different levels such as sensor level, feature level, score level and decision level. Awang et al. (2013) propose a feature level fusion of face features and online signature features for recognition. Linear Discriminant Analysis (LDA) is utilized to overcome high dimensionality problem of combined features. They use ORL face database (Cambridge 2016) and SuSIG signature database (Kholmatov and Yanikoglu 2009) for their experiment. They achieve 97.5% recognition accuracy.

In (Kazi et al. 2012), face and signatures are combined at score level. They analyze the performance of a normalized cross-correlation matcher and simple sum of scores fusion techniques. They show that fusion of face and signature scores improved the accuracy rate about 10% than unimodal system.

In (Lumini and Nanni 2017), overview of fusion approaches is presented for multimodal verification systems. Score level fusion methods are explored. They report that better results can be achieved using combination of different techniques.

Multiple traits are used to build a identification system in (Soleymani et al. 2018). Feature level combination of iris, fingerprint and face samples are utilized. They use CNN to extract features from each trait. Then, features are joint together for optimization. They report that optimization of joint representation concurrently superior to independent optimization of different models.

Kartik et al. (2008) present a multimodal system that uses face, speech and signatures of subjects. A score level fusion method is applied and the performance of the system against noise is reported. In (Rattani et al. 2007), feature level fusion of face and fingerprint is presented. They propose a method to combine features of each traits while reducing the

dimension. They report that the proposed fusion method at feature level perform better than score level fusion approach.

3. MATERIAL AND METHOD

3.1. Signature Verification

3.1.1. Preprocessing

In order to help training neural networks a simple preprocessing technique is applied to signature images. First, pixel values are inverted by subtracting them from 255 so that background pixels become 0. Then, connected components consisting of less than 40 pixels are removed with the assumption that they are not characteristic of signatures of writers and can be considered as noise. Minimum and maximum x, y coordinates of bounding box are found by removing rows and columns with zero values before and after them. Lastly, the image is resized to a fixed size to feed a convolutional neural network. A signature image and its preprocessed form is shown in Figure 3.1.

Figure 3.1. Preprocessing

3.1.2. Two-channel network

Siamese and two-channel neural networks are used to measure distance between two images. They take two images as inputs and they are trained for making decision whether two input images are similar or not.

In the proposed model, a two-channel network is utilized to perform UI signature verification. First channel of the input is used for genuine signatures and second channel is used for query (genuine or skilled forgery) signatures. During training, the network is forced to learn features for reference and query signatures differently, since first channel is only used for genuine signatures. Once the training is finished, the network can also be used to extract features and train user-dependent models. As it allows user-dependent and user-independent verification concurrently in a single forward propagation, it is efficient to combine these two approaches and obtain a robust verification system. Structure of the two-channel convolutional neural network can be seen in Figure 3.2.

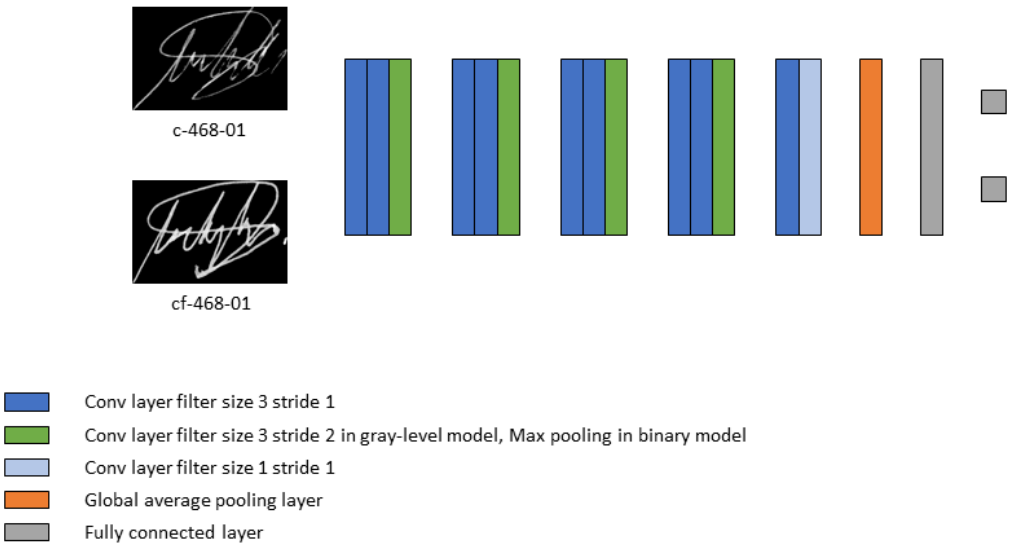

Figure 3.2. The proposed 2-channel CNN

Two similar CNN architectures are built for verification of binary and gray-level signature images separately. While gray-level model reduces dimension by using convolutional layer with stride of 2, binary model replaces these layers with max pooling layers to avoid overfitting.

Each convolutional layer is followed by batch normalization (Ioffe and Szegedy 2015) and then ReLU activation function (Nair and Hinton 2010) is applied. The network optimizes its parameters with the objective of minimizing binary cross-entropy loss. Adam optimizer (Kingma and Ba 2014) is used to train the network. Dropout (Srivastava et al. 2014) layers are applied to regularize the network. Hidden units in layers are given in Table 3.1.

In Table 3.2, convolutional layers denoted with C_3, C_6, C_9 and C_{12} are only applied to gray-level model. In binary model, max-pooling layers are deployed in place of these convolutional layers. Lower hidden unit values are given for binary model in C_4, C_5, C_7, C_8, C_{10}, C_{11} and higher values are given for gray-level model.

Two images are concatenated in the third dimension to feed the network with input size of $100 \times 150 \times 2$. Genuine pairs and genuine-skilled forgery pairs of each user are used to train the network. Once the training finished, it is employed to give a score between 0 and 1 indicating the probability of query signature being skilled forgery. Additionally, features extracted from GAP layer are used to train UD classifiers per user. It is expected that combination of these UI and UD approaches can be utilized to obtain better results.

Table 3.1. Number of hidden units in 2-channel CNNs

Layers	Hidden units
Convolution C_1 & C_2	30
C_3 or Max-Pooling	30 (C_3)
Dropout (0.5)	
Convolution C_4 & C_5	60 or 30
C_6 or Max-Pooling	60 (C_6)
Dropout (0.5)	
Convolution C_7 & C_8	100 or 60
C_9 or Max-Pooling	100 (C_9)
Dropout (0.5)	
Convolution C_{10} & C_{11}	150 or 100
C_{12} or Max-Pooling	150 (C_{12})
Dropout (0.5)	
Convolution C_{13} & C_{14}	200
GAP	
Fully-Connected	200
Dropout (0.5)	
Fully-Connected (softmax)	2

3.1.3. Recurrent binary patterns

Recurrent binary pattern (RBP) network is a user-dependent recurrent neural network (RNN) that learns sequential relations between LBP histograms over image windows. LBP features extracted from image windows are applied to a BiLSTM layer so that the network can capture sequential information, which is expected to be useful for making decision about a query signature.

LBP finds local features in an image by considering binary relationship between a center pixel and its neighbors. Calculation of LBP for a pixel is shown in Equation 3.1.

$$f_{LBP}(x_c, y_c) = \sum_{n=0}^{L-1} 2^n \xi(i_n, i_c) \tag{3.1}$$

In Equation 3.1, i_n is the intensity of n^{th} neighbor pixels and i_c is the intensity of the center pixel. L is the number of neighbor pixels considered for calculation. Function $\xi(\cdot)$ gives a binary result. If $i_n \geq i_c$ the result is 1 otherwise it is 0. In this work, $L = 4$ pixels are chosen to form neighbor groups resulting in 16 different $f_{LBP}(\cdot)$ codes between 0 and 15.

After the codes are calculated, histogram for an $M \times N$ image is built according to Equation 3.2.

$$f_H(k) = \sum_{m=1}^{M} \sum_{n=1}^{N} \gamma(f_{LBP}(x_m, y_n), k) \tag{3.2}$$

In Equation 3.2, $K = 2^L - 1$ is the maximum LBP value. Histogram element k is between 0 and K. Function $\gamma(x, y)$ is equal to 1 if $x = y$.

All neighborhoods up to Chebyshev distance 4 from the center pixel are considered. Neighborhoods for each distance are calculated separately since considering all 80 neighbors would result in 2^{80} codes. 4 equidistant pixels are chosen to form neighborhoods per distance, similar to (Yılmaz 2015). One example group for Chebyshev distance 2 is shown in Figure 3.3. In this case 4 different groups can be formed in which pixels are equidistant.

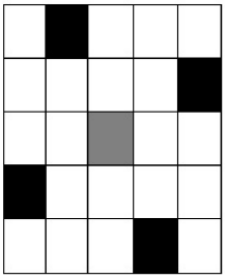

Figure 3.3. An example neighbor group for Chebyshev distance 2

There are 2 different neighbor groups can be chosen for Chebyshev distance 1, 4 neighbor groups for Chebyshev distance 2, 6 neighbor groups for Chebyshev distance 3 and 8 neighbor groups for Chebyshev distance 4. In total, there are 20 histograms and $2^4 - 1$ different LBP codes are considered. The case when all neighbors are zero is omitted. Therefore, $20 \times 15 = 300$ dimensional vector is used to represent signatures.

In order to reduce the dimensionality and obtain better results, most and least frequent LBP codes are removed from signature representations. A subset of training set is used to

detect these codes. Most frequent 40 and least frequent 40 codes are discarded resulting in 220 dimensional feature vector.

An RNN layer learns sequential information from 5 horizontal and 5 vertical image windows. 10% overlapping windows are considered. It treats windows as time steps so, each feature consists of 10 time steps. Features are divided by the maximum value for normalization. The proposed network is shown in Figure 3.4.

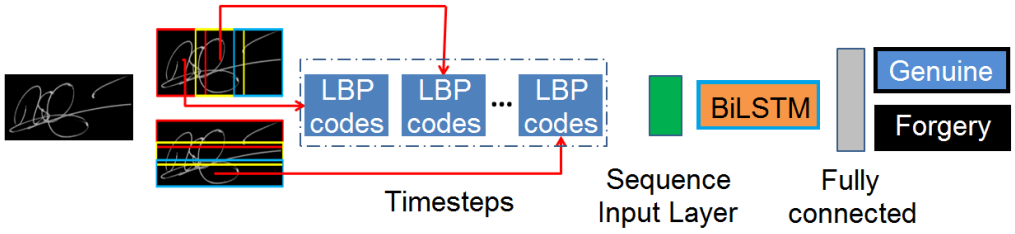

Figure 3.4. UD-RBP network architecture

A BiLSTM layer is a type of RNN architecture that learns bidirectional information in time steps. The proposed BiLSTM layer is used to learn sequence in LBP-coded image windows. It has 300 hidden units. Hyperbolic tangent and sigmoid functions are used as state and gate activation functions respectively.

3.1.4. Forgery identification CNN

In Forgery identification CNN (FI CNN), there are $|\tau|$ subjects used to train the network. Genuine signatures and skilled forgeries of users are considered as separate classes resulting in $2|\tau|$ outputs. The proposed CNN is shown in Figure 3.5.

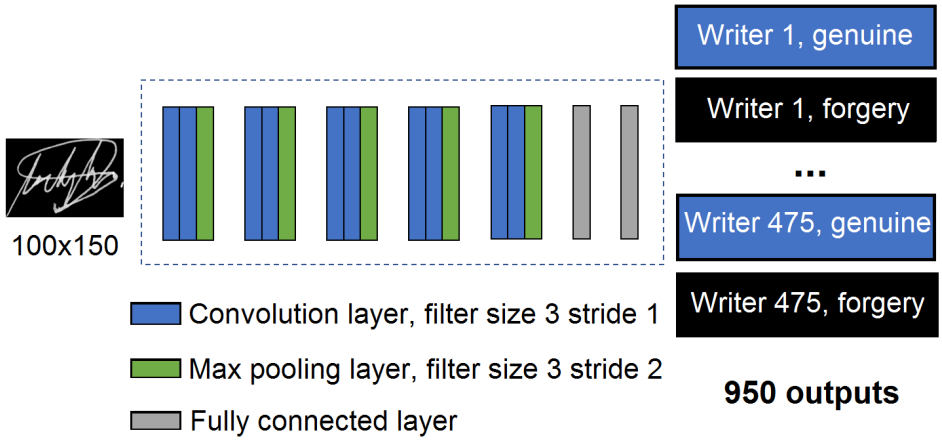

Figure 3.5. Architecture of the proposed CNN with forgery outputs

Table 3.2. Number of hidden units in the forgery identification CNN

Layers	Hidden units
Convolution	15
Convolution	15
Max Pooling	
Convolution	30
Convolution	30
Max Pooling	
Convolution	60
Convolution	60
Max Pooling	
Convolution	120
Convolution	120
Max Pooling	
Convolution	240
Convolution	240
Max Pooling	
Fully-Connected	1000
Dropout (0.5)	
Fully-Connected (softmax)	950

Signatures images are resized to 100×150 to feed the CNN. There are 10 convolutional layers, 5 max pooling layers and 2 fully-connected layers in the network. A dropout layer is used for regularization. Convolutional layers and the first fully-connected layer are followed by batch normalization layer and then ReLU activation function. At the end of the network, softmax function is used to normalize outputs. In training phase, the network minimizes categorical cross-entropy cost function using adam optimizer.

Once the training is done, features extracted from output of the first fully-connected layer are used to train UD classifiers per user. Number of hidden units in corresponding layers of the CNN is shown in Table 3.2.

3.1.5. Other models

SigNet-F (Hafemann et al. 2017) and SigNet-SPP-300dpi (Hafemann et al. 2018) are used to compare results with the proposed methods. In SigNet-F, skilled forgeries are used during training. The network has two cost function: categorical cross-entropy and binary cross-entropy. While categorical cross-entropy forces the network to recognize different users, binary cross-entropy term forces it to learn differences between genuine signatures and skilled forgeries. Weighted sum of these two terms is used to combine them. They also noted that, the model achieves better results if they are not penalized the network for misclassification of user of a skilled forgery. In this case, categorical cross-entropy term is ignored.

SigNet-SPP-300dpi is a CNN architecture that accept input with various sizes. They show that, the model achieves comparable results with SigNet-F. They also investigate the effect of different resolutions. It is shown that, higher resolutions can improve the results if skilled forgeries are used to train the network.

3.2. Face Verification

Transfer learning approach is considered to deploy a face verification system. Pre-trained VGG-Face (Parkhi et al. 2015) is utilized to extract features. The network is trained to classify 2622 identities. Up to 1000 face images per subject, in total 982803 images, is used for training. The architecture of VGG-Face can be seen in Figure 3.6.

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
type	input	conv	relu	conv	relu	mpool	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv
name	–	conv1_1	relu1_1	conv1_2	relu1_2	pool1	conv2_1	relu2_1	conv2_2	relu2_2	pool2	conv3_1	relu3_1	conv3_2	relu3_2	conv3_3	relu3_3	pool3	conv4_1
support	–	3	1	3	1	2	3	1	3	1	2	3	1	3	1	3	1	2	3
filt dim	–	3	–	64	–	–	64	–	128	–	–	128	–	256	–	256	–	–	256
num filts	–	64	–	64	–	–	128	–	128	–	–	256	–	256	–	256	–	–	512
stride	–	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	1
pad	–	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1

layer	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
type	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	softmx
name	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3	relu5_3	pool5	fc6	relu6	fc7	relu7	fc8	prob
support	1	3	1	3	1	2	3	1	3	1	3	1	2	7	1	1	1	1	1
filt dim	–	512	–	512	–	–	512	–	512	–	512	–	–	512	–	4096	–	4096	–
num filts	–	512	–	512	–	–	512	–	512	–	512	–	–	4096	–	4096	–	2622	–
stride	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1
pad	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

Figure 3.6. The architecture of VGG-Face (Parkhi et al. 2015)

While the pre-trained network can only recognize people in its training set, features extracted from an intermediate layer can be useful to recognize different people. It is assumed that, training set is large enough to learn distinguishing features not only for person in training set but also for different set of people. The last fully-connected layer before

the softmax layer is used for extracting features. Three different classification approaches are considered: one-vs-all, one-vs-one and verification.

3.2.1. One-vs-all model

One-vs-all models are trained for each registered subject in a system. N number of reference face images are used as positive samples and images of other subjects are used as negative samples to train a 1-vs-all verification model.

For a query face image Q claiming an identity c, the trained model of the user M^c gives a verification score $M^c(Q)$ between 0 and 1. If $M^c(Q) > \theta$ model verify the claimed identity otherwise query is rejected. A global or user-based threshold value θ can be used to make decision. For a system with k enrolled users, recognition can be done as in Equation 3.3.

$$\begin{aligned} max\{M^i_{1-vs-all}(Q), i = 1, ..., k\} > \theta \Rightarrow \\ \text{assign Q to subject i,} \\ max\{M^i_{1-vs-all}(Q), i = 1, ..., k\} \leq \theta \Rightarrow \\ \text{reject Q.} \end{aligned} \tag{3.3}$$

3.2.2. Verification model

Verification model is similar to 1-vs-all model. The only difference is the selection of negative samples to train each classifier. For training set T, another set T' is considered such that $T \cap T' = \emptyset$. Additional to the face images of other subjects in T, T' is also used as negative samples. It is assumed that, these images in T' can be collected readily from publicly available databases and can lead to performance gain, especially if T consists of few subjects.

Similar to 1-vs-all model, evaluation of k models is needed to recognize a query image according to Equation 3.3.

3.2.3. One-vs-one model

For k users, there are $k \times (k-1)/2$ 1-vs-1 models trained to compare users with each other. A perfect binary tree is utilized to make a decision. One subject is eliminated at

each level and therefore $k-1$ evaluation is needed to identify a query. An example of a tree for $k=5$ users is given in Figure 3.7.

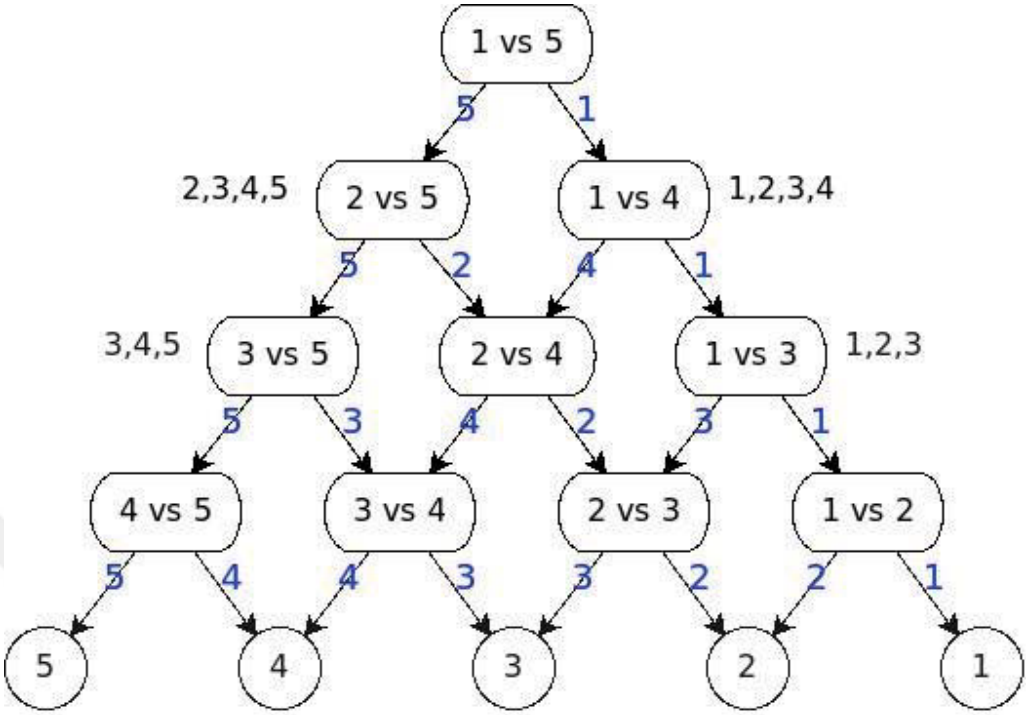

Figure 3.7. A perfect binary tree for 5 users (Anonymous 2 2019)

According to Equation 3.4, evaluation starts at the root node. Query is assigned to a subject at each level and decision is made at the leaves of the tree.

$$\begin{aligned} M_{i-vs-j}(Q) > \theta \Rightarrow \\ \text{assign Q to i or j whoever gets higher score,} \\ M_{i-vs-j}(Q) \leq \theta \Rightarrow \text{reject Q.} \end{aligned} \tag{3.4}$$

3.3. Multimodal Verification

A multimodal verification system which handles signature and face images is introduced. Each modality is verified independently and a score level combination is applied to make the final decision to either accept or reject. For signature verification, two-channel CNN with writer-independent and writer-dependent combination described in Section 3.1.2. is utilized.

Score obtained from signature model is denoted as S_{sign}. For face verification, user-dependent models trained with VGG-Face descriptor described in Section 3.2. is utilized. Score obtained from the face model is denoted as S_{face}. Proposed system applies a basic score level combination rule to generate the final score $S_{final} = \alpha S_{sign} + (1 - \alpha) S_{face}$ where the weight is experimentally found over a validation set in a user-independent manner. Multimodal verifier is illustrated in Figure 3.8.

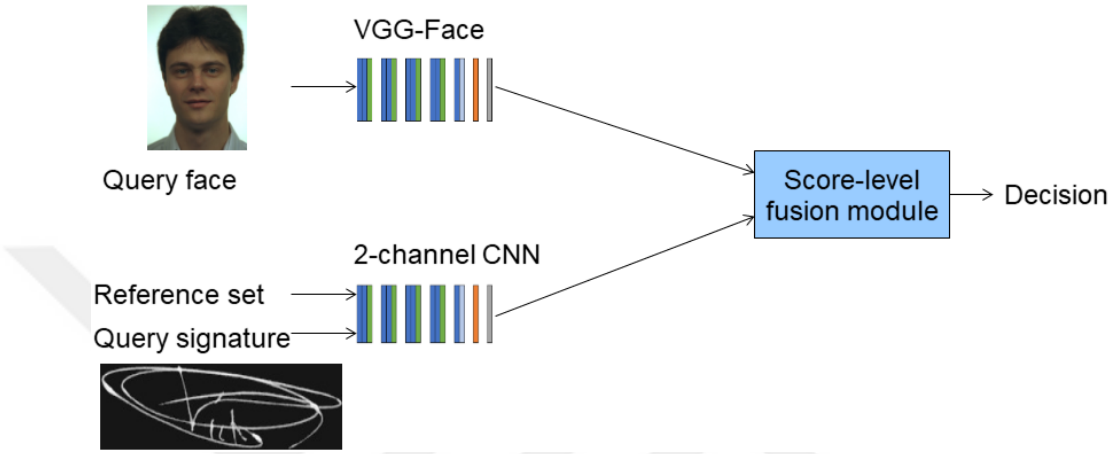

Figure 3.8. Proposed multimodal verifier

Two types of attacks are investigated. First one is the usual verification task in a controlled environment and is denoted as $Attack_1$. A forger can try to imitate a signature and can only show his own face to the camera. A genuine input thus consists of a genuine signature and face pair, a forgery input consists of a skilled forgery signature and the face image of the forger.

In the second kind of attack, it is assumed that in addition to presenting signature forgery and forger's face, a forger can clone a genuine template but for only one of the modalities. For example; the forger can either show the image of the claimed identity to the camera or carbon-copy the signature of the claimed identity, but not both at the same time. This kind of attack is called $Attack_2$. A genuine input thus consists of a genuine signature and face pair, while a forgery input may consist of skilled forgery signature and the face image of forger, a skilled forgery signature and face mask image of the claimed identity, a carbon-copied signature of the claimed identity and the face image of the forger.

$Attack_2$ is a superset of $Attack_1$ and it is impossible to overcome for unimodal face or signature verifiers. As a result, it is not necessary to experiment on a unimodal basis

as the error rate can be theoretically found. Assume a unimodal system has an equal error rate ϵ. Also assume the probability of template-copy attack is p over forgery inputs. Probability of usual $Attack_1$ is thus $1-p$. FRR is in any case ϵ. However, FAR will be $p(1-\epsilon)+\epsilon(1-p)$ where $(1-\epsilon)$ is true accept rate of the system which becomes false accept in case of template-copy attack. Numerically put; if a face verification system has EER of 2% and probability p that an attacker successfully shows the mask of the claimed identity to the system is 0.3, FAR will be $(0.3 \times 0.98) + (0.7 \times 0.02) = 30.8\%$. Average error rate of the system will thus be $(FAR + FRR)/2 = 16.4\%$

Varying levels of noise has been added to face images as described in Section 4.3., to measure the robustness of the proposed multimodal verification system. Note that face image noise removal or reduction is not of interest in this thesis, hence the aim is to investigate the contribution of signature when face image becomes noisy.

4. RESULTS AND DISCUSSION

4.1. Unimodal Signature Verification Models

GPDS-960 (Ferrer et al. 2012) and GPDS-Synthetic (Ferrer et al. 2017) databases are equipped to conduct experiments. GPDS-960 database consists of gray-level offline signature images of 881 users. There are 24 genuine signatures and 30 skilled forgeries provided for each user. The database is both used to train and test proposed signature verification models. Gray-level values are converted to binary in the experiments with binary model.

GPDS-Synthetic database consists of 10000 subjects. Each subject has 24 reference signatures and 30 skilled forgeries generated synthetically. Offline signature images are provided with gray-level values. First 300 subjects are used on test to evaluate the performance of the models trained on GPDS-960 database.

4.1.1. Binary and gray-level comparison of 2-channel CNN

GPDS-960 database is divided into different subsets as shown in Table 4.1. Test set T is divided into different subsets as T_1 and T_2 to separate reference and query signatures. V_1 and V_2 are used to determine hyperparameters of CNN and UD SVMs. Gray-level values are converted to binary for conducting experiments for binary model.

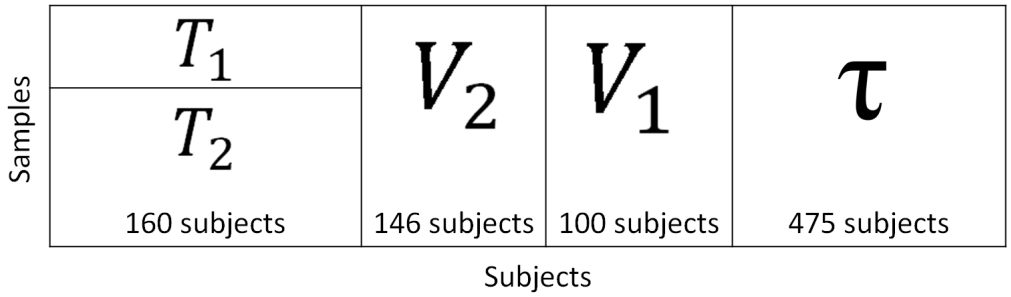

Figure 4.1. Database partition for 2-channel CNN

Last 475 subjects τ are used to train UI CNN. While all possible 24×23 pair of genuine signatures are used in training for each user, randomly selected subset of all possible permutations 24×30 of genuine-forgery pairs are used to balance samples of two classes. In total, $475 \times 24 \times 23 \times 2$ pairs are used in training. For validation, 50 positive and 50 negative pairs are randomly selected from each user in V_1.

Once the model is trained, it does not require any further training process. It can make decision for signatures of users that are not included in the development set τ. While a single reference is sufficient to give a similarity score, multiple references can be used to increase stability. Average score is calculated in this case.

It is needed to have a threshold value to make a binary decision using CNN scores. Threshold value can be independent from users or user-based value can be chosen. In order to report EER on test set, threshold values are determined by looking at test scores.

UD SVMs are trained with 200 dimensional features extracted from the GAP layer of the UI CNN. They are trained using 5 and 12 references per user. Hyperparameters of classifiers are decided based on the performance on validation set V_2. Skilled forgeries are not included in training set. Reference signatures of other users are used to create genuine-forgery pairs. For a query signature, signature representations are extracted from CNN as many as reference numbers. Then, an average score is calculated from SVM scores. UD and UI threshold values are used to report EER. UD SVM and UI CNN are fused to obtain robust results. Weighted sum of scores are combined. V_2 is used to learn combination weight.

Table 4.1 shows the results of gray-level training. Combination of UI CNN and UD SVM are given in Table 4.2. Results of training with binary signatures can be seen in Table 4.3 and combination of UI and UD approaches for binary models is shown in Table 4.4. Binary and gray-level models are both tested on gray-level and binary test set T. Results for reference numbers $N = 1$, $N = 5$ and $N = 12$ are reported.

Table 4.1. Results with gray-level training

	N	UI threshold EER		UD threshold EER	
		UI	UD	UI	UD
Gray T	1	8.74%	-	6.81%	-
	5	7.39%	6.52%	5.75%	4.72%
	12	7.20%	4.29%	5.78%	2.88%
Binary T	1	32.74%	-	29.74%	-
	5	31.92%	23.49%	27.26%	19.65%
	12	31.22%	17.95%	26.80%	15.03%

Table 4.2. Combination results with gray-level training

	N	UI threshold EER	UD threshold EER
Gray T	5	5.38%	3.92%
	12	4.13%	2.94%
Binary T	5	21.57%	18.21%
	12	18.08%	14.73%

Table 4.3. Results with binary training

	N	UI threshold EER		UD threshold EER	
		UI	UD	UI	UD
Gray T	1	32.15%	-	28.69%	-
	5	30.38%	14.03%	25.90%	11.01%
	12	30.18%	11.15%	25.75%	8.30%
Binary T	1	24.97%	-	21.22%	-
	5	22.32%	15.46%	18.95%	11.41%
	12	21.64%	12.14%	18.47%	9.31%

Table 4.4. Combination results with binary training

	N	UI threshold EER	UD threshold EER
Gray T	5	14.10%	10.85%
	12	11.12%	8.26%
Binary T	5	15.40%	11.31%
	12	11.86%	9.22%

The best result is obtained when the model is trained and tested with gray-level signatures since they have more information than binary signatures. While UD SVMs are always better than UI CNN, UI CNN gives compatible results, even using only 1 reference signature.

4.1.2. Experimental protocol for RBP

UD RBPs are trained for each subject. Reference signatures as positive examples and random forgeries as negatives are used to train the network for making a binary decision. Skilled forgeries are not included in training set. Global and user-based thresholds are used to report verification EER.

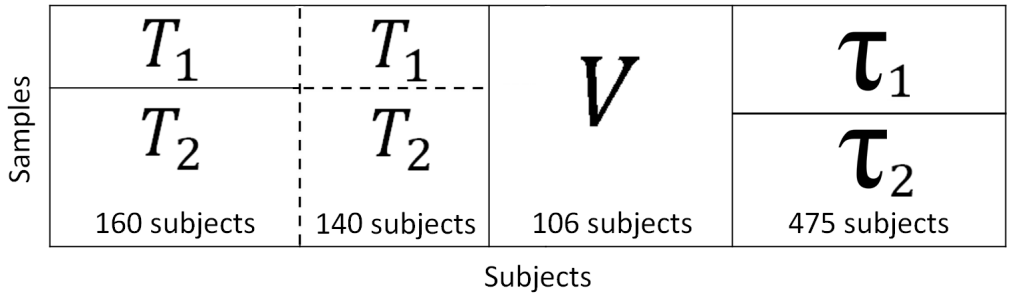

Figure 4.2. Database partition for RBP and forgery identification CNN

The subsets of GPDS-960 are given in Figure 4.2. GPDS-160 (first 160 subjects) and GPDS-300 (first 300 subjects) are used to test the performance of the RBP network. The hyperparameters of the network are set depending on the performance on validation set V. Test set is divided into two subsets as T_1 and T_2. T_1 represents the reference signatures used to train the network. T_2 represents the query samples composed of genuine signatures and skilled forgeries. $N = 5$ and $N = 12$ number of references are considered to train the model.

4.1.3. Experimental protocol for forgery identification CNN

The subset τ composed of 475 subjects are used for training the CNN. τ_1, consisting of 20 genuine signatures and 20 skilled forgeries of each user in τ, is used to train the network. Remaining samples τ_2 are utilized to set hyperparameters of the network.

Once the training is done, UD classifiers are deployed using features extracted from the first fully-connected layer of the CNN. Validation set V is utilized to determine hyperparameters of the UD SVMs. $N = 5$ and $N = 12$ number of references are considered to train the classifiers. The same test subsets in Section 4.1.2. is utilized for comparison.

4.1.4. Experimental results

GPDS-160, GPDS-300 and GPDS-Synthetic databases are used to measure the performance of the proposed methods with using different number of references. Reference and query samples are determined randomly and repeated 3 times. Average EERs are reported for all models. User-independent and user-dependent thresholds are used to report ERRs. Results are shown in Tables 4.5, 4.6, 4.7, 4.8, 4.9 and 4.10.

SigNet-F and SigNet-SPP-300dpi is utilized for comparison with the proposed models. The same test protocol is applied to all UD classifiers. Score combinations of different models are provided. Combination weights are learned from the validation set V. Since it is learned from another set than test set, some combination results do not lead to improvement and therefore are not included in tables.

Table 4.5. EER results for $N = 5$ on GPDS-160

Method	UI threshold	UD threshold
RBP size 300	6.75%	5.16%
(1) RBP size 220	5.77%	4.29%
(2) FI CNN	7.34%	4.34%
(3) SigNet-F	4.28%	2.83%
(4) SigNet-SPP	4.86%	3.44%
1 & 2	3.38%	1.82%
1 & 3	2.41%	1.34%
2 & 3	2.97%	1.87%
3 & 4	3.92%	2.53%
1 & 2 & 3	2.05%	1.11%
1 to 4	1.93%	1.08%

Table 4.6. EER results for $N = 12$ on GPDS-160

Method	UI threshold	UD threshold
RBP size 300	6.32%	4.72%
(1) RBP size 220	5.56%	4.24%
(2) FI CNN	5.08%	3.26%
(3) SigNet-F	3.51%	2.11%
(4) SigNet-SPP	4.39%	2.68%
(5) 2-channel CNN (UI & UD)	4.13%	2.94%
1 & 2	2.75%	1.63%
1 & 3	2.11%	0.98%
2 & 3	2.19%	1.27%
3 & 4	3.42%	1.99%
3 & 5	1.76%	0.88%
1 & 2 & 3	1.66%	0.81%
1 to 4	1.66%	0.81%
1 to 5	1.11%	0.57%

Table 4.7. EER results for $N = 5$ on GPDS-300

Method	UI threshold	UD threshold
RBP size 300	5.55%	3.99%
(1) RBP size 220	4.58%	3.29%
(2) FI CNN	6.26%	3.45%
(3) SigNet-F	4.38%	2.83%
(4) SigNet-SPP	5.83%	4.22%
1 & 2	2.79%	1.19%
1 & 3	2.22%	1.15%
2 & 3	3.06%	1.78%
3 & 4	4.21%	2.83%
1 & 2 & 3	1.86%	0.84%
1 to 4	1.90%	0.84%

Table 4.8. EER results for $N = 12$ on GPDS-300

Method	UI threshold	UD threshold
RBP size 300	5.50%	3.79%
(1) RBP size 220	4.59%	3.18%
(2) FI CNN	4.40%	2.46%
(3) SigNet-F	3.64%	2.23%
(4) SigNet-SPP	5.12%	3.48%
1 & 2	2.10%	1.00%
1 & 3	1.90%	0.88%
2 & 3	2.30%	1.20%
3 & 4	3.56%	2.24%
1 & 2 & 3	1.54%	0.61%
1 to 4	1.53%	0.61%

Table 4.9. EER results for $N = 5$ on GPDS-Synthetic

Method	UI threshold	UD threshold
(1) RBP size 220	31.73%	31.16%
(2) FI CNN	31.55%	28.42%
(3) SigNet-F	27.47%	24.91%
(4) SigNet-SPP	35.00%	33.21%
1 & 2	28.82%	25.90%
1 & 3	26.50%	24.04%
2 & 3	25.90%	22.48%
3 & 4	27.07%	24.27%
1 & 2 & 3	25.08%	22.13%
1 to 4	25.08%	22.13%

Table 4.10. EER results for $N = 12$ on GPDS-Synthetic

Method	UI threshold	UD threshold
(1) RBP size 220	26.32%	24.22%
(2) FI CNN	25.68%	23.52%
(3) SigNet-F	18.93%	16.98%
(4) SigNet-SPP	28.71%	26.83%
1 & 2	22.00%	20.01%
1 & 3	17.95%	16.16%
2 & 3	18.18%	15.62%
3 & 4	18.92%	16.85%
1 & 2 & 3	17.50%	14.80%
1 to 4	17.65%	14.93%

The performance of RBP network using histogram selection method is evaluated. Experimental results show that proposed technique to reduce dimension improves the accuracy.

GPDS-Synthetic EERs are much higher than results on GPDS-300 and GPDS-160. It can be noted that, since GPDS-Synthetic database is composed of generated signatures using modeled pens, the characteristics of signatures do not similar to each other and there is no fine-tuning procedure applied to this test set.

The results show that, lower EERs can be achieved by using combinations of different models. While the best single model is SigNet-F, EER can be reduced by half using combination of RBP network and Signet-F.

4.2. Unimodal Face Verification

Samples of 15 subjects from Yale (Anonymous 4 2019), 40 subjects from Essex (Anonymous 1 2019) and 40 subjects from ORL (Anonymous 3 2019) databases are used for the experiments. Each subject has 10 frontal face images. While images in Yale and ORL are gray-scale, Essex database consists of color images.

10 subjects from Yale, 20 subjects from Essex and 20 subjects from ORL are considered as users enrolled to the system defined as T in Section 3.2.2.. The remaining 45

Table 4.11. EER results for Yale

	Global threshold			User-based threshold		
N	1	3	5	1	3	5
1-vs-all	7.6e−3%	2.5e−4%	6.3e−5%	1.5e−3%	0	0
1-vs-1	2.4e−2%	4.7e−3%	1.4e−3%	2.1e−2%	1.9e−4%	0
V_1	1.6e−3%	0	0	7.4e−4%	0	0
V_2	8.2e−3%	6.3e−5%	0	6.7e−3%	0	0

Table 4.12. EER results for ORL

	Global threshold			User-based threshold		
N	1	3	5	1	3	5
1-vs-all	1.7e−3%	5.0e−4%	0	0	0	0
1-vs-1	6.8e−3%	7.2e−4%	2.2e−4%	1.1e−3%	2.9e−5%	0
V_1	1.0e−3%	5.3e−4%	0	8.8e−5%	0	0
V_2	3.7e−3%	8.2e−4%	0	1.8e−4%	0	0

subjects are used as random negatives T'.

Random negatives are divided into two groups as T'_1 (22 subjects) and T'_2 (23 subjects). T'_1 is used to train the verification model V. T'_2 is used to test all models.

Two different training protocols are applied for the model V. In the first one V_1, samples of other users in T are considered as negative examples with samples from T'_1. In the other case V_2, negative examples only consist of samples from T'_1. While we have to train all models again when a new user register for V_1, we do not have to update model for V_2 since T'_1 is fixed.

Test results are reported for $N = 1$, $N = 3$ and $N = 5$ reference samples. References samples are selected randomly 3 times. Average EERs are reported using global and user-based threshold. Results for each database are given in Tables 4.11, 4.12 and 4.13.

Although the results are close to each other, verification models achieve better results than $1 - vs - 1$ and $1 - vs - all$ approaches. It is worth noting that, when a new user enrolls, we have to train all models except V_2.

Table 4.13. EER results for Essex

	Global threshold			User-based threshold		
N	1	3	5	1	3	5
1-vs-all	0	0	0	0	0	0
1-vs-1	3.2e−4%	0	0	0	0	0
V_1	0	0	0	0	0	0
V_2	0	0	0	0	0	0

4.3. Multimodal Verification

GPDS-960 signature and color FERET (Phillips et al. 2000) face databases are used for experiments of multimodal verifier. A face detection method (Viola and Jones 2001) is applied to subjects of color FERET database. An example of original image and face detected image is shown in Figure 4.3.

Figure 4.3. Original image and detected face

After the faces are detected, first 300 subjects with the highest number of samples are considered for experiments. Each subject are matched with the first 300 subjects from GPDS-960 so that there are 300 artificial subjects having both face and signature samples.

Two different levels of noise and transformation are applied to face images to evaluate robustness of multimodal verifier. For the first level, each of the following is independently applied with a probability of 0.33: Rotation (angle $\sim N(\mu = 15, \sigma = 1)$), scaling (factor $\sim |N(\mu = 1, \sigma = 0.1)|$), shear ($\lambda \sim \mathcal{U}[0, 0.5]$) and white noise for each color

channel independently $\sim N(\mu = 0, \sigma = 0.005)$.

For the second level, following processes are independently applied with a probability of 0.5: Rotation (angle $\sim N(\mu = 20, \sigma = 1)$), scaling (factor $\sim |N(\mu = 1, \sigma = 0.2)|$), shear ($\lambda \sim \mathcal{U}[0, 0.8]$) and white noise for each color channel independently $\sim N(\mu = 0, \sigma = 0.005)$. Shear transformation is applied parallel to the y axis according to Equation 4.1.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{4.1}$$

While rotating the face images, the following procedure is applied to get rid of the empty spaces occurring as an artifact of rotation. First, face image is padded with the value of the last matrix element on each dimension as the padding value (i.e. the first or last row / column / corner for that direction, accordingly) for half the size of the image in each direction. Then, padded image is rotated and the same number of rows and columns used in padding are removed. This procedure is illustrated in Figure 4.4.

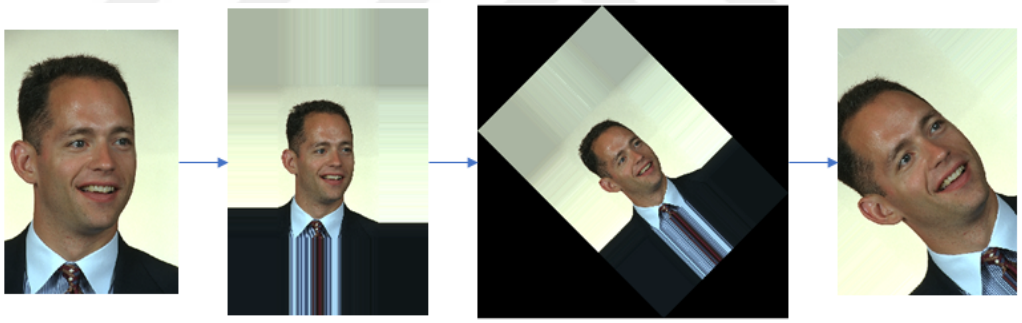

Figure 4.4. Rotation procedure

For the signatures; first 300 subjects of GPDS-960 are utilized with two-channel CNN of writer-independent and writer-dependent combination. Number of reference signatures per subject is only taken as 12 and cross-validations with random partitions and reference set is performed.

For the faces; first 300 subjects having a necessary amount of image samples, obtained from the processed color FERET database are utilized. Number of reference faces per subject is 3 and cross-validations are performed similar to signatures. While combining the scores of two modalities, all cross-validation reference-set selections are considered to generate all possible multimodal pairs.

Combination weight α is learned on last 100 subjects of the dataset used for multimodal tests. All multimodal test results are thus obtained on the first 200 signature and face subjects. Note that those signature and face subjects are matched as single multimodal subjects according to the same order among signature and face subjects.

Unimodal EER results for the mentioned test protocol are shown in Table 4.14 for $Attack_1$ and $Attack_2$ scenarios, while multimodal EER results and weights are shown in Tables 4.15 and Table 4.16 for $Attack_1$ and $Attack_2$ scenarios respectively.

Table 4.14. Unimodal EER results

Attack type	Signature	Face	Noisy Face	Noisy Face (2)
$Attack_1$ EER	2.89%	1.03%	2.15%	6.86%
$Attack_2$ EER	21.61%	7.53%	7.67%	9.24%

Table 4.15. Multimodal EER results for $Attack_1$ scenario

Signature and $\rightarrow$	Clean Face	Noisy Face	Noisy Face (2)
EER	0.34%	0.72%	1.37%
α (combination weight)	0.2%	0.2%	0.4%

Table 4.16. Multimodal EER results for $Attack_2$ scenario

Signature and $\rightarrow$	Clean Face	Noisy Face	Noisy Face (2)
EER	2.87%	3.60%	7.56%
α (combination weight)	0.5%	0.5%	0.4%

5. CONCLUSION

In this thesis, unimodal signature and face verification systems are investigated separately. Several models are proposed for offline signature verification tasks. Usage of different deep learning architectures, that have achieved great success on various computer vision tasks, is analyzed. Score level combinations of models are used to lower error rates. A pre-trained network is deployed to further improve verification results by combining unimodal biometric systems.

Two CNN architectures are utilized to perform signature verification. In the first one, a CNN architecture is used to distinguish different users and skilled forgeries concurrently. Skilled forgeries are considered as separate classes and the network is forced to identify them. In the second case, a two-channel architecture is presented to compare two signatures. It is important to note that, after the training is done, this network does not require any further training procedure to verify people who are not included in training set. However, for the first one, it is necessary to train a classifier for recognizing test set.

A novel RBP network is proposed. It is assumed that, it can help to capture sequential information in signatures. It is shown that, CNN and RBP architectures are very complementary and combination of them is an effective way to obtain better results.

User-independent and user-dependent approaches are analyzed to train verifiers. It is reported that, the models trained in user-independent manner can be used to train user-dependent classifiers. Furthermore, combination of two approaches is used to improve results.

A score level combination approach is presented for unimodal biometrics. Several noise and attack procedures are applied to evaluate robustness of the multimodal biometric system. It is shown that, the usage of multimodal system is critical to overcome limitations of unimodal systems.

Signature databases have small number of samples for each subject compared to computer vision task in which deep learning techniques have been very successful. It is worth noting that, although the most important factor of success for deep learning is the amount of data, it can be very powerful even when the training set is small.

As a future work, in order to overcome the data limitation of signature databases, usage of synthetic signatures for training offline signature verification system will be in-

vestigated. Moreover, generator methods will be investigated to produce synthetic signatures to improve performance of the classifiers.

6. REFERENCES

Anonymous 1: Essex Face Recognition Data. http://cswww.essex.ac.uk/mv/allfaces [Last access date: 01.12.2019].

Anonymous 2: Handwritten Digit Recognition Based on SVM. http://liuhongjiang.github.io/tech/blog/2012/12/29/svm-ocr [Last access date: 01.12.2019].

Anonymous 3: The ORL AT&T Database of Faces. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html [Last access date: 01.12.2019].

Anonymous 4: The Yale Face Database. http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html [Last access date: 01.12.2019].

Awang, S., Yusof, R., Zamzuri, M. F. and Arfa, R. 2013. Feature level fusion of face and signature using a modified feature selection technique. International Conference on Signal-Image Technology and Internet-Based Systems, pp. 706-713, 2-5 December, Kyoto, Japan.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E. and Shah, R. 1993. Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7 (4): 669-688.

Ferrer, M. A., Alonso, J. B. and Travieso, C. M. 2005. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (6): 993-997.

Ferrer, M. A., Diaz, M., Carmona-Duarte, C. and Morales, A. 2017. A behavioral handwriting model for static and dynamic signature synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (6): 1041-1053.

Ferrer, M. A., Vargas, J. F., Morales, A. and Ordonez, A. 2012. Robustness of offline signature verification based on gray level features. *IEEE Transactions on Information Forensics and Security*, 7 (3): 966-977.

Hafemann, L. G., Oliveira, L. S. and Sabourin, R. 2018. Fixed-sized representation learning from offline handwritten signatures of different sizes. *IJDAR*, 21 (3): 219-232.

Hafemann, L. G., Sabourin, R. and Oliveira, L. S. 2017. Learning features for offline handwritten signature verification using deep convolutional neural networks. *Pattern Recognition*, 70 (3): 163-176.

Hu, J., Guo, Z., Fan, Z. and Chen, Y. 2017. Offline signature verification using local features and decision trees. *International Journal of Pattern Recognition and Artificial Intelligence*, 31 (3): 175-190.

Huang, G. B., Mattar, M., Berg, T. and Learned-Miller, E. 2008. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. University of Massachusetts, Technical Report 07-49, Amherst, Massachusetts.

Ioffe, S. and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning, pp. 448-456, 6-11 July, Lille, France.

Kartik, P., Prasad, R. V. and Prasanna, S. M. 2008. Noise robust multimodal biometric person authentication system using face, speech and signature features. Annual IEEE India Conference, pp. 23-27, 11-13 December, Kanpur, India.

Kazi, M., Rode, Y., Dabhade, S., Al-Dawla, N., Mane, A., Manza, R. and Kale, K. 2012. Multimodal biometric system using face and signature: a score level fusion approach. *Advances in Computational Research*, 4 (1): 99-103.

Khalajzadeh, H., Mansouri, M. and Teshnehlab, M. 2012. Persian signature verification using convolutional neural networks. *International Journal of Engineering Research and Technology*, 1 (2): 7-12.

Kholmatov, A. and Yanikoglu, B. 2009. Susig: an online signature database, associated protocols and benchmark results. *Pattern Analysis and Applications*, 12 (3): 227-236.

Kingma, D. P. and Ba, J. 2014. Adam: A method for stochastic optimization. International Conference on Learning Representations, 7-9 May, San Diego, California.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems, pp. 1097-1105, 3-6 December, Lake Tahoe, Nevada.

LeCun, Y., Bottou, L., Bengio, Y.,and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86 (11): 2278-2324.

Lumini, A. and Nanni, L. 2017. Overview of the combination of biometric matchers. *Information Fusion*, 33 (1): 71-85.

Nair, V. and Hinton, G. E. 2010. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on International Conference on Machine Learning, pp. 807-814, Haifa, Israel.

Parkhi, O. M., Vedaldi, A. and Zisserman, A. 2015. Deep face recognition. *BMVC*, 1 (3): 6-15.

Phillips, P. J., Moon, H., Rizvi, S. A. and Rauss, P. J. 2000. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (10): 1090-1104.

Rattani, A., Kisku, D. R., Bicego, M. and Tistarelli, M. 2007. Feature level fusion of face and fingerprint biometrics. 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–6, 27-29 September, Crystal City, Virginia.

Ribeiro, B., Goncalves, I., Santos, S. and Kovacec, A. 2011. Deep learning networks for off-line handwritten signature recognition. Proceedings of the 16th Iberoamerican Congress conference on Progress in Pattern Recognition. pp. 523–532, 15-18 November, Pucon, Chile.

Ross, A. and Jain, A. K. 2004. Multimodal biometrics: an overview. 2004 12th European Signal Processing Conference, pp. 1221-1224, 6-10 September, Vienna, Austria.

Schroff, F., Kalenichenko, D. and Philbin, J. 2015. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815-823, 7-12 June, Boston, Massachusetts.

Soleimani, A., Araabi, B. N. and Fouladi, K. 2016. Deep multitask metric learning for offline signature verification. *Pattern Recognition Letters*, 80 (3): 84-90.

Soleymani, S., Dabouei, A., Kazemi, H., Dawson, J. and Nasrabadi, N. M. 2018. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification. International Conference on Pattern Recognition, pp. 3469–3476, 20-24 August, Beijing, China.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of Machine Learning Research*, 15 (1): 1929–1958.

Sun, Y., Wang, X. and Tang, X. 2015. Deeply learned face representations are sparse, selective, and robust. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2892–2900, 7-12 June, Boston, Massachusetts.

Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. 2014. Deepface: Closing the gap to human-level performance in face verification. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701– 1708, 23-28 June, Columbus, Ohio.

Viola, P. and Jones, M. 2001. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511-518, 8-14 December, Kauai, Hawaii.

Wen, Y., Zhang, K., Li, Z. and Qiao, Y. 2016. A discriminative feature learning approach for deep face recognition. European Conference on Computer Vision, pp. 499-515, 11-14 October, Amsterdam, The Netherlands.

Wolf, L., Hassner, T. and Maoz, I. 2011. Face recognition in unconstrained videos with matched background similarity. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 529-534, Colorado Springs, Colorado.

Yılmaz, M. B. 2015. Offline signature verification with user-based and global classifiers of local features. PhD dissertation, Sabancı University, İstanbul, 95 p.

Yılmaz, M. B. and Yanıkoğlu, B. 2016. Score level fusion of classifiers in off-line signature verification. *Information Fusion*, 32 (1): 109–119.

Zagoruyko, S. and Komodakis, N. 2015. Learning to compare image patches via convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4353–4361, 7-12 June, Boston, Massachusetts.

CURRICULUM VITAE

KAĞAN ÖZTÜRK

kaganozturk1992@gmail.com

EDUCATION

Master of Science 2017-2020	Akdeniz University Institute of Natural Sciences, Department of Computer Engineering, Antalya
Bachelor of Science 2010-2015	Yıldız Technical University Faculty of Electrical and Electronics, Electronics and Communication Engineering, İstanbul

WORK EXPERIENCE

Research Assistant 2018-Present	Alanya Alaaddin Keykubat University Faculty of Engineering, Department of Computer Engineering, Antalya

PUBLICATIONS

Papers delivered in International Conferences and Printed as a Proceedings

1- Ozturk, K. and Yilmaz M. B. 2017. A Comparison of Classification Approaches for Deep Face Recognition. International Conference on Computer Science and Engineering (UBMK), pp. 227-232, 5-8 October, Antalya.

2- Yilmaz M. B and Ozturk, K. 2018. Hybrid User-Independent and User-Dependent Offline Signature Verification with a Two-Channel CNN. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 526-534, 18-22 June, Salt Lake City, Utah.

3- Yilmaz M. B and Ozturk, K. 2019. Recurrent Binary Patterns and CNNs for Offline Signature Verification. Proceedings of the Future Technologies Conference, pp. 417-434, 24-25 October, San Francisco, California.