

**CLASSIFICATION OF WRINKLES
ON THE FOREHEAD AND AROUND EYES**

M.Sc. THESIS

Büşra ÇANAK

Department of Computer Engineering

Computer Engineering Programme

JUNE 2017

**CLASSIFICATION OF WRINKLES
ON THE FOREHEAD AND AROUND EYES**

M.Sc. THESIS

**Büşra ÇANAK
(504131510)**

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Assoc. Prof. Dr. Mustafa Ersel KAMAŞAK

JUNE 2017

**ALINDA VE GÖZ KENARLARINDA YER ALAN
KIRIŞIKLIKLARIN SINIFLANDIRILMASI**

YÜKSEK LİSANS TEZİ

**Büşra ÇANAK
(504131510)**

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Assoc. Prof. Dr. Mustafa Ersel KAMAŞAK

HAZİRAN 2017

Büşra ÇANAK, a M.Sc. student of ITU Graduate School of Science Engineering and Technology 504131510 successfully defended the thesis entitled “CLASSIFICATION OF WRINKLES ON THE FOREHEAD AND AROUND EYES”, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assoc. Prof. Dr. Mustafa Ersel KAMAŞAK**
Istanbul Technical University

Jury Members : **Assoc. Prof. Dr. Mustafa KAMAŞAK**
Istanbul Technical University

Asst. Prof. Dr. Yusuf YASLAN
Istanbul Technical University

Asst. Prof. Dr. Amaç GÜVENSAN
Yıldız Technical University

Date of Submission : **5 May 2017**

Date of Defense : **9 June 2017**

To my family,

FOREWORD

I want to thank to my advisor Assoc. Prof. Dr. Mustafa KAMAŞAK for his contribution, consideration and help during my research. In addition, I also want to thank Asst. Prof. Dr. Erkut ERDEM and Asst. Prof. Dr. Aykut ERDEM who were my advisors while i was studying in Hacettepe University. They were quite inspirational and thoughtful, besides they encouraged me to get my master's degree in the Image Processing area. Thanks to them, this research area helped me on my academic and working career. All in all, the special thanks belong to my beloved family for their endless patient, love and support during my childhood and education.

JUNE 2017

Büşra ÇANAK
(Computer Engineer)

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xiii
SYMBOLS	xv
LIST OF TABLES	xvii
LIST OF FIGURES	xix
SUMMARY	xxi
ÖZET	xxiii
1. INTRODUCTION	1
1.1 Purpose of Thesis	2
1.2 Literature Review	3
2. METHODS	7
2.1 Feature Extraction	7
2.1.1 Local Binary Pattern.....	7
2.1.2 Gabor Features and Connected Component Analysis	9
2.2 Classification	14
2.2.1 Support Vector Machines	14
2.2.2 K Nearest Neighbour.....	15
2.2.3 Evaluation.....	16
3. DATABASE COLLECTION	17
3.1 Database Collection.....	17
3.1.1 Forehead	17
3.1.2 Left Eye	19
3.1.3 Right Eye.....	19
3.2 Database Annotation	20
3.2.1 Raters.....	20
3.2.2 Categories	20
3.2.3 Raters' Reliability.....	21
3.3 Database Separation	23
4. EXPERIMENTS	25
4.1 Classification Using Local Binary Pattern	25
4.1.1 Forehead	25
4.1.2 Left Eye	26
4.1.3 Right Eye.....	26
4.2 Classification Using Gabor Filter Based On Connected Component Analysis	28
4.2.1 Forehead	28
4.2.2 Left Eye	29

4.2.3 Right Eye	30
5. DISCUSSIONS.....	31
6. CONCLUSION	37
REFERENCES.....	39

ABBREVIATIONS

Ac	: Actual label of the test samples
KNN	: K Nearest Neighbour
LBP	: Local Binary Pattern
Pre	: Prediction label of the test samples
RBF	: Radial Basis Function
SVM	: Support Vector Machine
RBF SVM	: Support Vector Machines based on Radial Basis Function
TP	: True Positive
TPR	: True Positive Rate

SYMBOLS

d	: Distance
u, v	: Displacement Vector Components
G(u,v)	: Gabor filter
G(x,y)	: Gaussian function
L	: Lagrangian function
LBP_(P,R)	: LBP with P points of R radius
K	: Mapping function
P	: Positive samples
q	: Train samples
p	: Test samples
x	: Train samples
y	: Train labels
α	: Weight factor
w	: Weight value

LIST OF TABLES

	<u>Page</u>
Table 3.1 : Number of images and person that contributed to the database for each region	18
Table 3.2 : Wrinkle severity scale	21
Table 3.3 : Measurement of κ coefficient	22
Table 3.4 : Results for each region.....	22
Table 3.5 : Number of labeled forehead images.....	24
Table 3.6 : Number of labeled left eye images.....	24
Table 3.7 : Number of labeled right eye images.	24
Table 4.1 : The classification results of the wrinkles on the forehead using LBP	26
Table 4.2 : The confusion for the forehead wrinkles. First column shows actual classes whereas first row shows predicted classes.....	26
Table 4.3 : The classification results of the wrinkles on the left eye using LBP...	27
Table 4.4 : The confusion matrix of the wrinkles on the left eye. First column shows actual classes whereas first row shows predicted classes.....	27
Table 4.5 : The classification results of the wrinkles on the right eye using LBP	27
Table 4.6 : The confusion matrix of wrinkles on the right eye. First column shows actual classes whereas first row shows predicted classes.....	28
Table 4.7 : The result of the classification of the wrinkles on the forehead using Gabor features	28
Table 4.8 : Confusion matrix of the forehead wrinkles. First column shows actual classes whereas first row shows predicted classes.....	29
Table 4.9 : The result of the classification of the wrinkles on the left eye using Gabor features.....	29
Table 4.10 : Confusion matrix of the wrinkles on the left eye. First column shows actual classes whereas first row shows predicted classes.....	29
Table 4.11 : The result of the classification of the wrinkles on the right eye using Gabor features	30
Table 4.12 : Confusion matrix of the wrinkles on the right eye. First column shows actual classes whereas first row shows predicted classes.....	30

LIST OF FIGURES

	<u>Page</u>
Figure 2.1 : How LBP operator works [1].	7
Figure 2.2 : Different LBP notations [2].	8
Figure 2.3 : Uniform LBP codes [3].	8
Figure 2.4 : LBP Histogram [3].	8
Figure 2.5 : Gabor Features based on Connected Component Analysis flow	9
Figure 2.6 : Original image on the left, the magnitude in the middle and the phase on the right.	11
Figure 2.7 : Sobel edge detection result	12
Figure 3.1 : The green points states the center of the two eyes.....	18
Figure 3.2 : The cropped and aligned forehead images with and without flash....	18
Figure 3.3 : The cropped and aligned left eye images with and without flash.....	19
Figure 3.4 : The cropped and aligned right eye images with and without flash. ..	20
Figure 3.5 : The image on the upper left side belongs to class 1. The image on the upper right side belongs to class 2. The third image belongs to class 3.....	23
Figure 3.6 : The image on the upper left side belongs to class 1. The image on the upper right side belongs to class 2. The third image belongs to class 3.....	23
Figure 3.7 : The image on the upper left side belongs to class 1. The image on the upper right side belongs to class 2. The third image belongs to class 3.....	24
Figure 5.1 : Box plot of the test samples for each fold of the wrinkles on the forehead. Horizontal line stands for the categories and vertical line stands for the wrinkle number.	32
Figure 5.2 : Box plot of the test samples for each fold of the wrinkles on the left eye. Horizontal line stands for the categories and vertical line stands for the wrinkle number.....	33
Figure 5.3 : Box plot of the test samples for each fold of the wrinkles on the right eye. Horizontal line stands for the categories and vertical line stands for the wrinkle number.....	34
Figure 5.4 : Wrinkle number and wrinkle length relation.....	34

CLASSIFICATION OF WRINKLES ON THE FOREHEAD AND AROUND EYES

SUMMARY

People communicate through their faces. They express themselves, feelings, give clues about their gender, age etc. via their faces. Face carries valuable information. Therefore it is important to be fulfilled about their appearance for people. It helps people to increase pleasure that they gain from life. Facial wrinkles appear in time due to loss of elasticity. On the other hand there are environmental issues that cause wrinkle on the face such as ultra-violet exposure or facial expressions (yawning, smiling etc.). The wrinkles can have different severities. While some of them are barely visible, some of them can be extremely deep. According to the clinicians (such as dermatologists, plastic surgeons), the treatment of the wrinkles is really based on their severity. The treatment for the mild wrinkles may not help to remove extreme wrinkles. Therefore, the classification of the wrinkle is a crucial point to find the optimal treatment. In addition, one cannot classify the wrinkles, if the wrinkles are not detected and localized.

For this purpose, cosmetic researchers have performed various studies. Their main concern is to provide standard systems to classify wrinkle severity. A group of dermatologists, plastic or aesthetic surgeons come together to grade wrinkle severities of the collected images. The grading process happens in two sessions mostly. The inter-rater reliability is calculated after the first session. The raters that have highest reliabilities are selected for the second session. Then they grade the wrinkles again. In the end the wrinkle severity systems are obtained. This process has two main drawbacks. First of all the grading systems are strongly related to the raters. If the raters are changed, the scores can be affected. In addition, the system is not automatic to reproduce the results in anywhere at any time.

When the aim is to develop reliable, repeatable and standard systems, there are significant studies in the computer society. However, the research is about the detection of the wrinkles mostly. Classification of the wrinkles is not a well studied area. The annotated wrinkle locations are being tried to find by the automatic systems. The studies are typically for the forehead wrinkles mostly. By annotating the collected images, Gabor filtering, Hessian filtering, Hybrid Hessian filtering etc. methods are being applied to detect wrinkles.

The aim of this study is to build a standard, repeatable, reliable, automatic system to detect and classify wrinkle severities. A database is required which has samples and labels for the wrinkles. Firstly, a database was collected for this purpose. The images were taken from three different parts of the face. These regions are forehead, left and right eye. The images were taken from the same view under the same light. The layout of the camera was used to provide the standardization. The images were taken from the volunteers. The volunteers were not wearing any make up and they also did not

have surgery in the last three years. The images were taken from 29 subjects with and without flash light.

The forehead images were labeled according to the eye brows. The eye images were labeled using eye and the ear points. These points were used to align and crop images for grading. The last version of the images only includes the wrinkle parts of the face to exclude any bias. The prepared images were graded by five female and five male volunteers. The volunteers were used their point of view while scoring. The categories that will be used were specified by using the previous research. Hence, the specified number for the categories was three. The absence of the wrinkle and the mild wrinkles belong to the first category while the moderate wrinkles belong to the second category. The extreme and deep wrinkles belong to the third category.

Before scoring, raters took a look to the whole images to have an initial information about the wrinkles. Then, they were asked to grade the wrinkle severities according to their observations. To calculate inter-rater reliability the Kappa (κ) coefficient were used. According to the calculated κ coefficients, the reliabilities were found 0.63, 0.79 and 0.69 for the forehead, left and right eye respectively. The results were found substantial to use. By using the scores, the images were divided into train and test sets. Three-fold were defined to use for the reproducibility. The train sets included the same number of the samples. The person independency was also taken into account.

In this study two different feature extraction and classification methods were used to detect and classify the wrinkles. The first method uses Local Binary Pattern (LBP) to extract features and Support Vector Machine (SVM) to train a model. 16×16 pixels and 8×8 pixels patches were used for Local Binary Pattern. The uniform LBP features were also used because of their informativeness. The features were trained using linear SVM and RBF SVM. The second feature extraction method was Gabor filter. However, the images were smoothed by Gaussian filter. Then the Gabor filter was applied to have its phase value. The erosion, dilation and Sobel edge detection were applied to the phase value in order. Finally, Connected Component Analysis were used to find the count of the wrinkles on the images to use as feature. The features were classified by using K Nearest Neighbour algorithm with majority voting.

When the experiments were examined, Gabor features outperformed LBP features in every category. While the highest accuracy of the classification of the wrinkles on the forehead were 64.5%, the accuracy of 79.3% were obtained using Gabor features. The accuracy was 64.4% for the classification of the wrinkles on the left eye when LBP was used. However the accuracy of 73% was obtained using Gabor features for the same region. The accuracies of the classification of the wrinkles on the right eye were 63% and 78.9% for the LBP and Gabor features respectively. While the LBP features mixed the second and third category with each other mostly, Gabor features were more successful to classify first and third categories. Gabor features mixed second category with the third category rarely.

The wrinkles around the mouth were also shot while collecting the database. However the classification results were not good. The images should have taken closer and with a proper angle. Then the wrinkle length was added to the wrinkle number for the classification. Nonetheless, these two features could not provide sufficient information for accurate classification.

ALINDA VE GÖZ KENARLARINDA YER ALAN KIRIŞIKLIKLARIN SINIFLANDIRILMASI

ÖZET

İnsanların günlük yaşamlarında en çok kullandıkları etkileşim yöntemi yüzyüze iletişimdir. Bir kişinin yüz ifadelerinden verdiği tepkiler, ruh durumu, cinsiyeti, yaşı gibi çeşitli bilgiler elde edilebilir. Çok yönlü kullandıkları bu iletişim aracından memnuniyetleri yaşam kalitelerini yükseltmeleri açısından önem taşımaktadır. Zamanla ya da tekrarlı mimik hareketlerinden ötürü insan yüzünde çeşitli bozulmalar meydana gelebilir. Bunlardan en önemlisi yüzde oluşan kırışıklıklardır. Yüzde yer alan kırışıklıkların oluşma sebepleri çeşitli etkenlere dayanmaktadır. Kırışıklıklar zamanla yüzde yer edinebileceği gibi morötesi (UV) ışıklar gibi dış etkenlerin sebebiyle de meydana gelebilir. Örneğin sürekli tekrarladığımız yüz hareketleri de bunlara sebep olabilir. Gülerken yaptığımız kas hareketlerinden dolayı göz çevresinde ve ağız kenarlarında bu tekrarlı hareketlerin sonucu olarak kırışıklıklar meydana gelebilir.

Yüzde yer alan bu kırışıklıkların giderilmesi araştırmacılar tarafından uzun sürelerdir başlıca araştırma konusu haline gelmiş alanlardan biridir. Kırışıklıkların giderilmesi için yapılan bu çalışmalarda gereken ilk kurallar kırışıklıkların tespiti, derecelendirilmesi ve giderilmesi şeklinde sıralanabilir. Kırışıklıkların giderilmesi başlıca amaç iken ancak ve ancak tespit edilmiş ve uygun yöntemler ile derecelendirilmiş kırışıklıklar için tedavi edilme yöntemlerinden bahsedilebilir. Duruma bu açıdan bakıldığında kırışıklık tespiti ve derecelendirilmesi belirli kırışıklık seviyelerine bulunacak tedavi yöntemi kadar önem taşımaya başlar. Kırışıklıkların tespiti uygulanacak bölgenin seçimine yardımcı olurken doğru tedavi önerilmesi açısından kırışıklıkların sınıflandırılması da büyük bir önem taşır. Kırışıklıkların dereceleri herkes için ve yine her bölge için farklı olacağından önerilecek tedavilerin başarılarından bahsedebilmek için doğru sınıflandırılma yapılmalıdır. Her seviye için ayrı bir tedavi yöntemi üretilmesi tedavilerin daha başarılı sonuçlara ulaşmasını sağlar. Örneğin hemen hemen hiç olmayan kırışıklıkların giderilmesi için önerilen tedavinin derecesi yüksek sınıflardan birine ait olan kırışıklığı gidermesi beklenemez.

Bu sebeple yapılan çalışmalar iki gruba ayrılarak incelenmiştir. Geçmiş çalışmalara bakıldığında kozmetik sektöründe dermatoloji uzmanları, estetik cerrahları ya da plastik cerrahlar tarafından yapılan çalışmalar göze çarpmaktadır. Dünya genelinde standart, güvenilir bir derecelendirme sistemi amacı güdülmüştür. Toplanan veri tabanları referans imgeler belirlenerek araştırmacılara derecelendirilmiştir. Bu derecelendirme seansları bir kaç kez tekrarlanarak araştırmacılar arasında güven oranları hesaplanılarak standart ve güvenilir bir sınıflandırma sistemi yaratılmak istenmiştir.

Söz konusu kozmetik çalışmalarını gerçekleştirilirken deneylere katılan derecelendiricilerin bakış açılarına göre bir sistem kurulmuştur. Bu durumda deneylerde kullanılan derecelendiriciler değiştirildiğinde farklı sonuçlar elde edilebilir. Yine

bu sistemler sonuçların tekrarlanabilirliği konusunda geride kalabilir. Sistemlerin otomatik olmaması kişileri anında sonuca götürmemesi de büyük bir eksiklik olarak kabul edilebilir.

Kırışıklıkların giderilmesine dair yapılan ikinci grup çalışmaların ise ürettiği otomatik sistemler ile tekrarlanabilirlik, genel geçerlilik ve güvenilirlik açısından daha başarılı örnekler sayılabilir. Bu konuda kırışıklıkların yerinin belirlenmesi adına çalışmalar öne çıkmaktadır. Çeşitli algoritmalar ile çıkarılan öznitelik vektörleri eğitilerek sınıflandırıcı olarak kullanılmıştır. Genel olarak sistemlerin kırışıklıkları sınıflandırmadan çok kırışıklıkların yerlerinin belirlenmesine yönelik çalışmalar mevcut olmuştur. Derecelendirilmiş örnek bir veri kümesinin yokluğu bunun temel sebeplerinden sayılabilir. Bu çalışmalarda genel olarak Gabor süzgeci, Hessian süzgeci gibi doku tabanlı yöntemler kullanılmıştır.

Yapılan çalışmadaki amaç standart, sonuçları tekrarlanabilir, güvenilir ve otomatik bir sistem tasarlanmasıdır. Bu nedenle çalışmanın gerçekleştirilebilmesi için gerekli olan ilk konu üzerinde çalışılabilecek bir veritabanının ihtiyacıdır. Kullanılmak istenen veritabanında farklı kırışıklık seviyeleri için çeşitli örnekler bulunması planlanmaktadır. Gönüllülerden oluşan veri kümesi kullanılarak kişilerin alın, sol ve sağ göz kenarlarında yer alan kırışıklıkların fotoğrafları çekilmiştir. Ardından alın imgeleri kaş orta noktalarına, göz imgeleri ise kulak ve göz bebeğine göre işaretlenip hizalanmış ve kesilmiştir. Hazırlanan imgeler beş kadın ve beş erkek olmak üzere toplam on adet gönüllüye derecelendirilmiştir. Derecelendirmede kullanılan imgelerin son halleri sadece ilgili bölgedeki kırışıklıkları içeren kısımları kapsamaktadır. Bu sayede derecelendiricilerin kişilerin yüzleriyle ifade ettiği diğer verilerden bağımsız olarak sadece kırışıklık bölgesine odaklanması sağlanmıştır. Gönüllülerin katılımıyla toplanan veri tabanında 29 kişinin alın ve göz kenarları flaş ışığı kullanılarak ve kullanılmayarak fotoğraflandırılmıştır. Derecelendirmede kullanılan sınıf sayısının belirlenmesinde daha önceden yapılan çalışmalar baz alınarak ve toplanan veritabanının boyutlarına göre üç farklı kategori kullanılmasına karar verilmiştir. Kullanılan üç kategori için tanımlar şu şekilde belirlenmiştir; kırışıklığın olmadığı ya da hafif kırışıklıkların olduğu imgeler bir numaralı kategoriye ait, aşırı kırışıklık gözlenen imgeler üç numaralı kategoriye ait, hafif ya da aşırı olarak belirlenemeyen iki durumun arasında kalan imgeler ise ikinci kategoriye ait olarak belirlenmiştir.

Derecelendiriciler hızlıca tüm veri setine göz gezdirmiş ve ardından kendi yargılarına ve bakış açılarına göre imgeleri derecelendirmiştir. Derecelendiriciler arasındaki güvenilirliği ölçmek için Kappa (κ) katsayısı hesaplanmıştır. Buna yönteme göre derecelendirme işlemleri 0.63, 0.79 ve 0.69 katsayı değerine alın, sol ve sağ göz kenarlarına denk gelecek şekilde sıralamaya sahip olmuştur. Elde edilen sonuçlar önemli derecede güvenilir bulunduğundan sınıflar çalışmaya uygun bulunmuştur. Derecelendirilen imgeler eğitim ve sınama kümesi olarak iki gruba ayrıştırılmıştır. Bu işlem üç kere tekrarlanmıştır. Buradaki amaç rastgele seçilecek olan veri kümesindeki imgelere bağlı kalmamak içindir. Sonuçlar üç kere tekrar edilerek tekrarlanabilirliği ölçülmüştür. Her eğitim kümesinde eşit sayıda örnek bulunmaktadır.

Bu çalışmada geçmiş çalışmalar göz önüne alınarak alın, sol ve sağ göz kenarlarında yer alan kırışıklıkların yerinin tespiti ve sınıflandırılması anlatılmıştır. Bu çalışma için iki farklı öznitelik çıkarma yöntemi iki farklı sınıflandırma yöntemi ile beraber kullanılmıştır. Yerel İkili Örüntü yöntemi kullanılan ilk yöntemdir. Bu yöntem gereği

imgeler 16×16 ve 8×8 piksellik yamalara parçalanarak çalışılmıştır. Bu yöntemin tekdüze olarak adlandırılan formu kullanılmıştır. Bu sayede elde edilen öznelik vektörleri Destek Vektör Makinelerinin doğrusal ve radyal tabanlı çekirdek fonksiyonu kullanılarak eğitilmiştir.

İkinci yöntem olarak ise Gabor süzgeci tercih edilmiştir. İmgelere Gabor süzgeci uygulanmadan önce imgelerde olan gürültülerden kurtulmak için Gaussian süzgeci ile bulanıklaştırılma uygulanmıştır. Gabor süzgecinden elde edilen faz değeri kullanılmaya uygun bulunup faz değerine aşındırma, genişletme, kenar bulma yöntemleri uygulanmıştır. Bu sayede elde edilen imgelerden öznelik vektörü olarak kırıxıklık sayısı elde edilmesini planlanmıştır. Bunun için işlem gören son imgeye Bağlı Bileşenler Yöntemi uygulanmıştır. Kırıxıklık sayısının öznelik vektörü olarak kullanıldığı bu yöntemde sınıflandırıcı olarak K En Yakın Komşuluk algoritması kullanılmıştır. 3 olarak belirlenen k değerine göre test örneklerinin ait oldukları sınıfların bulunması amaçlanmıştır.

Sonuçlar incelendiğinde Gabor süzgeci kullanılarak elde edilen başarının Yerel İkili Örüntü yönteminden her üç kırıxıklık bölgesi için daha yüksek olduğu görülmüştür. Yerel İkili Örüntünün kullanıldığı deneylerde 16×16 piksellik ve 8×8 piksellik yamalar kullanılmıştır. Alın bölgesindeki kırıxıklıkların sınıflandırıldığı deneylerde 16×16 piksellik yamalar 8×8 piksellik yamalardan daha yüksek sonuçlara ulaşmıştır. Kullanılan sınıflandırıcı yöntemleri değerlendirildiğinde ise doğrusal destek vektör makinesinin radyal fonksiyon tabanlı çekirdek kullanan destek vektör makinesinden daha başarılı olduğu gözlemlenmiştir. Elde edilen en yüksek başarı 64.5% ile 16×16 piksellik yamaların doğrusal destek vektör makinesi kullanılmasıyla elde edilmiştir.

Sol göz kenarındaki kırıxıklıkların derecelendirilmesi için yapılan deneylerde de 16×16 piksellik ve 8×8 yamalar uygulanmıştır. Deneylerde 16×16 piksellik yamalar 8×8 piksellik yamalardan daha yüksek sonuçlar üretmiştir. Yine sınıflandırıcılar dikkate alınarak yapılan karşılaştırmada ise doğrusal destek vektör makinesi radyal fonksiyon tabanlı çekirdek kullanan destek vektör makinesinden her seferinde daha yüksek başarılarla ulaştırmıştır. Elde edilen en yüksek başarı 64.4% oranındaki başarıyla 16×16 piksellik yamaların doğrusal destek vektör makinesi kullanılmasıyla elde edilmiştir.

Sağ göz kenarlarındaki kırıxıklıkların derecelendirilmesi için yapılan deneylerde ise en yüksek başarı yine 16×16 piksellik yamaların doğrusal destek vektör makinesi kullanılmasıyla elde edilen 63% oranındaki başarı olmuştur. Bu başarıyı 56.4% oranıyla doğrusal destek vektör makinesinin 8×8 piksellik yamalar kullanılmasıyla elde edilen sonuçlar takip etmiştir. Radyal fonksiyon tabanlı çekirdek kullanan destek vektör makinesinin 8×8 piksellik yamalar kullanılmasıyla yapılan deneylerde elde edilen başarı ise 53.6% oranında başarı yakalamıştır. Son olarak ise radyal fonksiyon tabanlı çekirdek kullanan destek vektör makinesinin 8×8 piksellik yamalar kullanılmasıyla yapılan deneylerde elde edilen 47% oranında başarı gelmektedir. Yapılan deneylerden elde edilen sonuçlara bakıldığında en çok karıştırılan sınıfların iki ve üç numaralı sınıflara ait olduğu gözlemlenmiştir. Kırıxıklığın yokluğunu iyi bir şekilde ayırt edebilirken var olan kırıxıklıkların seviyelendirilmesinde başarılı olamadığı gözlemlenmiştir. Eğer konu kırıxıklığın varlığı ve yokluğu olursa Yerel İkili Örüntü kullanılabilir olabileceği sonucuna varılabilir.

Gabor süzgeci ile elde edilen başarılar alın bölgesi için 79.3% olurken sol ve sağ göz kenarları için ise sırasıyla 73% ve 78.9% olarak elde toplanmıştır.

Karmaşıklık matrislerine göz atıldığında ise 1 ve 3 numaraları kategorilerin 2 numaralı kategoriye göre çok daha başarılı ayrıştırıldığı gözlemlenmiştir. Alın, sol ve sağ göz kenarlarındaki kırıksıklıklarının tespiti ve derecelendirilmesi işlemine ek olarak ağız çevresinde yer alan kırıksıklıkların tespit ve derecelendirilmesi için de imgeler fotoğraflandırılmıştır. Fakat öznitelik çıkarma yöntemlerinden başarılı sonuçlar elde edilememiştir. Ayrıca ağız bölgesinin çekilen fotoğraflarının daha yakın açıdan ve farklı bir işaretleme yöntemi kullanılarak yapılması gerektiği sonucuna ulaşılmıştır.

Diğer bir öneri olarak da kırıksıklık sayısına ek olarak kırıksıklık uzunluğu kullanılmasına karar verilmiştir. Bu yüzden kırıksıklıkları temsil etmek için hem kırıksıklık uzunluğu hem de kırıksıklık sayısı birlikte kullanılmıştır. Fakat elde edilen veriler incelendiğinde bu iki özelliğin birlikte kullanılmasının ayrıştırıcı bir özelliğe sahip olmadığı gözlemlenmiştir.

1. INTRODUCTION

People have expressed themselves via facial expressions, gestures. Face gestures is a component of interaction between people. Human face contains enormous information. The application of the facial techniques can be used by all kind of system services in daily life. Security, banking, vending machines, entertainment, biometrics, surveillance monitoring and cosmetology can be the prior examples of its usage. Therefore people pay attention to physical appearance especially facial appearance for self confidence. As a result, beautifying and enhancement of the face are crucial.

While face plays a key role for a person, massive and important information can be obtained from it. According to the face-based analysis, gender recognition, age estimation, facial expression detection, facial retouching can be applied. All of the facial techniques are mostly related to the facial wrinkles which plays an important role. It is an important clue for other face-based analysis techniques especially.

Wrinkles can be defined as fine or mild lines in the human face simply. They carry important clues for age, emotions, cosmetic research etc. They are caused by solar ultraviolet (UV) radiation, chronological ageing, repeated facial expressions or environmental factors. All of the effects caused the changes of elasticity of the skin, therefore the occurrence of the wrinkles cannot be inevitable. How they occurred does not matter for people because if the wrinkles do not occur as a consequence of environmental effects, it will occur in time eventually. The concern is how they can be removed or reduced. Cosmetic investigators pay lots of attention to this area for the patient's pleasure. Finding the perfect treatment for mild wrinkles does not help to handle with extreme wrinkles. The treatments must be related to the severity of the wrinkles to have successful solutions.

The severity of the wrinkles must be found to apply perfect treatment. On the other hand the solutions can be marked as invalid although it is a perfect way to specific severity. Therefore detection and classification of the wrinkles play a vital role for

whole treatment process. Recent studies about wrinkle assessment are split into two categories. First group includes cosmetic research which means that dermatologists, aesthetic surgeons define the wrinkle severities. These are not automatized systems. Their reproducibility, reliability and standardization are open to discuss. Other studies aim wrinkle localization mostly. The facial wrinkles which have different wrinkle severities are detected using this systems but classification or assessment are not well studied.

Automatized systems are expected to give same results in any other circumstances. They are reliable and accurate to reproduce the results. If an automatic system is developed, the standardization comes to specify the class of the wrinkles for all research.

1.1 Purpose of Thesis

A survey of studies about detecting and assessing wrinkles, show that there are two disciplines that work on this topic. First of all the cosmetic researchers have made a huge effort to assess wrinkle severities. However this studies strongly depend on the raters and it is a time consuming process. A group of dermatologists/surgeons collaborate to grade wrinkle severities. The inter and intra-rater reliabilities of the results are substantial. However, the effect of changing raters on system performance are not investigated. In addition to this, the process is not automated. Therefore, when the new input or rule is built, the whole process should be repeated.

Secondly, there are quiet significant wrinkle detection systems in computer vision area. They localize wrinkles using texture based methods or filtering processes. On the other hand, the studies do not work on classification of the wrinkles mostly.

In this research we came up with the idea that combines detection and assessment of wrinkle severities. We built an automatic, reliable, repeatable and standard system. By handling both detection and classification, the system gives a chance to find an affective treatment for the related severities. Better treatment requires better classification, while better classification needs better detection of wrinkles. The system will be helpful for

detection and classification parts quickly. For this reason, three regions were selected for investigation; forehead, left and right eye.

1.2 Literature Review

The cosmetic studies lean on the individual's observations. A wrinkle assessment system was proposed by taking into account dermatologists, aesthetic surgeons and plastic surgeons opinions [4]. The key point of the research is to build standard measurement system to assess the effectiveness of the treatments. The reference photos belong to the 11 different parts of a face image. Three plastic surgeons, three aesthetic surgeons and three dermatologists examined 76 deep wrinkles of a 38 different individuals and classified the photos into 6 categories (0-5). The agreement ratio is 92.7% to put a photo in a certain category . For the second part of the study, 130 photos are examined by eight observers who are four plastic surgeons and four lay people. The classification score is 89.4% in total. In addition to that, the accuracy of 87% is obtained after applying 40 silicon impressions to the individuals.

The aim of another study is to assess the efficacy of the treatments in cosmetic research. Therefore, post-treatment and pre-treatment processes are scored separately [5]. Perioral and periorbital wrinkles were examined by applying laser treatment in resurfacing photo-aged skin. Three main classes which were mild, moderate and severe are used to determine.

An another study examines nasolabial folds only [6]. Thirty wrinkle photos are classified into one of the five classes (1, 2, 3, 4 or 5), which are defined as absent, mild, moderate, severe and extreme by 5 investigators who are three dermatologists, one plastic surgeon and one ear-nose-throat specialist. Left and right side of the nasolabial fold's severity might be different for an image therefore these parts are displayed to the investigators separately. Photographic references and descriptions are used from a photo for scoring. Intra and inter observer reliability are calculated using kappa coefficient.

The nasolabial wrinkles were examined by using the same approach in [5]. The same three main classes are used exactly but one more class is added to define absence of a wrinkle [7]. In addition to this, three inter-classes are introduced. While main

classes have a reference photos which were specified by five committee members as a gold standard, inter-classes do not have a reference photos. Their assessment are left to the observers. Kappa (κ) coefficient is used to measure wrinkle grading system. Nine assessors graded volunteer's photos firstly. Then five of them who have highest reliability were selected to score clinical patient's photos for validation. Calculating the inter and intra-assessors agreement showed that proposed method can be used to grade wrinkle severities. The κ coefficient that belongs to previous study (which are 0.77 for the left and 0.81 for the right) are also similar in value that was obtained in this study. It indicates that the approach is reliable and reproducible.

Cosmetic researches take into account investigators' observation. One cannot say that it is most accurate way to measure efficacy of the system. κ coefficient is mostly used for calculation. If one observer changes his/her mind while scoring, the coefficients might change. The need for an automatic assessment system is undeniable. Instead of human based systems there must be computer-based technics to localize the wrinkles.

Due to lack of the databases which have real person wrinkles, some studies used replicas to obtain wrinkle folds [8,9]. Replica-based approaches were not successful because they could not get the form of actual skin morphological features. In addition, facial hair can not be provided while real life photos have these kind of difficulties. People also do not have perfect skin conditions, there always must be irregularities. Therefore, replicas are not reliable while the problem tries to find a solution for the actual human skin.

The classification of nasolabial wrinkles were examined by using nasolabial photo image of patients which were taken Visia camera [10]. While Visia cameras are able to take pictures with different degrees and illuminations, 0 degree and one illumination were chosen to use in the study. There are four severity classes but the number of the train sets are different for each class. K nearest neighbour with Euclidean distance are used to train the intensity profiles.

The marked point process is also employed to detect wrinkles [11]. FG-NET database and the images that were collected from the internet were used together. The examined region is forehead because it is claimed that the wrinkles on the forehead are found easier and obvious. However, if the wrinkles are not obvious enough, it affects

gradients. Therefore it fails. In addition the method is not suitable when the skin has so much irregularities and discontinuous lines. The approach works only well for deeper wrinkles which means extreme wrinkles.

Another study of wrinkles is about detecting forehead wrinkles using Gabor filter and geometric constraints [12]. The method also uses FG-NET ageing dataset. The proposed method combined Gabor and Hessian filter approaches in one way. It is stated that the proposed method is more successful when it is compared to Gaussian or Hessian filter separately.

The Cula method combines length of the lines and filter based detection in Wrinkle Index for severe wrinkles [13]. The method inspired from fingerprint detection. The wrinkles were treated as line segments like finger prints but not continuous with the same thickness. The difference between the wrinkles which are graded into same severity is extremely different from each other. Because each person has different skin features. The approach also takes care of this drawback with its robustness. The Frangi filter is a vessel enhancement filtering [14]. The desired idea is consisted of partial derivatives for detecting ridges. Hessian matrix is also used to extract directions.

Ng et al. examines forehead wrinkles [15]. The method which uses Hessian filter and directional gradient together is an expanded version of Frangi filter. The Bosphorus database were annotated by three users. It is stated that when the results are compared to the Cula method and Frangi filter, it achieved better results. While the proposed filter works well to detect medium and coarse wrinkles, it fails to detect fine wrinkles.

The same authors also examine detection of forehead wrinkles [16]. It is inspired from Hybrid Hessian Filter but it is an expanded version of it to overcome its weakness. Therefore the approaches also uses Hessian filter, however the results are better when the line tracking was added to the filtering. The database were annotated by three raters. It is stated that the proposed method achieved the accuracy of 84% which is better than the Cula method, Frangi filter and Hybrid Hessian Filter.

Moreover face imperfection is another notable study about wrinkles. Finding and removing the wrinkles were examined to imperfect facial beauty [17]. The proposed algorithm were includes Gabor features, Gaussian Mixture Model and Markov

Random Field Modelling as well. The effectiveness were presented by using database which was downloaded from the internet.

Another method defines five features which are length, number, width, area and depth to detect and assess wrinkles on the crow's feet area, forehead, upper lip and under the eye [18]. The approach uses Ranking Light and it is called Stephens Wrinkle Imaging using Ranking Light (SWIRL). It is found that the wrinkle area is the similar one to the clinical scoring. Because it includes length, number and width information also.

In this study instead of using medical dictionary such as periorbital lines, the regions are defined as forehead, left eye and right eye. The following chapter describes the feature extraction methods. Chapter three gives information about the database information and Chapter four explains the experiments. Chapter five discusses the results and the final chapter is the conclusion.

2. METHODS

This section describes the feature extraction and classification methods to detect and classify facial wrinkles. Feature extraction methods are mentioned in section 2.1 while the classification methods are described in section 2.2.

2.1 Feature Extraction

2.1.1 Local Binary Pattern

Local Binary Pattern (LBP) has been using in facial applications for ages [1, 2, 19, 20]. It is a texture-based method which plays a substantial role to represent a facial information. Texture information can be useful to detect mild wrinkle because of its successful applications for facial expression recognition [19].

3×3 pixels patches are built to calculate each pixel's value. According to the centred pixel's value, its neighbours' values are calculated. If neighbour's values are bigger than centred pixel, its new value becomes 1 otherwise 0. By turning clockwise, the binary code that is 8-bit length is calculated. Simple LBP operator can be seen in Figure 2.1. According to the binary codes, histogram values are calculated to represent the centred pixel. There are different forms of LBP that are indicated as: $LBP_{(P,R)}$ which is shown in Figure 2.2. While P stands for the points that will be used, R stands for the radius. In this study, 3×3 patches are used as $LBP_{(8,1)}$.

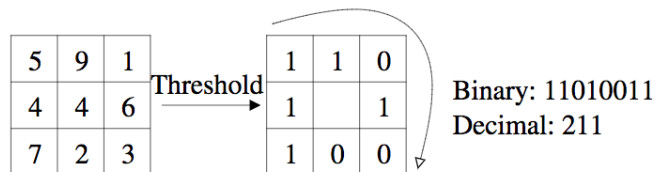


Figure 2.1 : How LBP operator works [1].

Another notable study about LBP includes uniform patterns [2]. It means that if there are only 0, 1 or 2 bitwise transition between 0 and 1 in 8-bit binary code, the code is called uniform. For instance, if the code is 00000011, there is one transition between 0

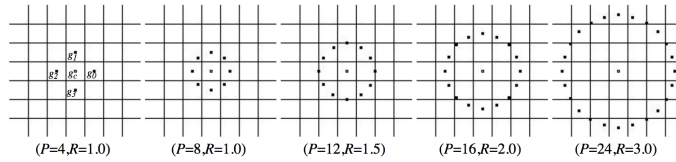


Figure 2.2 : Different LBP notations [2].

and 1. Therefore, it is an uniform code. The number of uniform patterns are less than %90 of all outcomes for $LBP_{(8,1)}$ notation [2]. The notation becomes $LBP^{u2}_{(8,1)}$ [1]. $u2$ implies that uniform patterns are assigned to its corresponding values while the rest of them are being assigned to the one different exact label. Another study also states that the uniform patterns carry more information about the photos which can be seen in Figure 2.3 [3]. There are 58 uniform codes for $LBP_{(8,1)}$ notation [21].

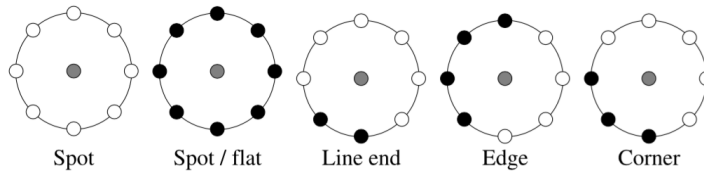


Figure 2.3 : Uniform LBP codes [3].

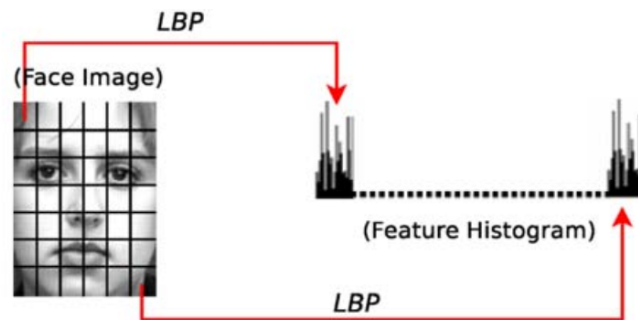


Figure 2.4 : LBP Histogram [3].

The images are divided into blocks. 3×3 pixels patch is applied to each pixel of the block to calculate uniform histogram vector. After obtaining the vectors, each vector are added one after another to build up a feature vector to represent the image as in Figure 2.4. 64×64 pixels and 8×8 pixels blocks can be constructed to divide images into blocks. The vector size becomes 1×59 pixels for each block. When the vectors are added in order, the feature vector size becomes $1 \times (59 \times 64)$ pixels. It is a quite small number when it is compared to non-uniform LBP method. All in all, uniform codes provide better description for the images while reducing the feature vector size.

2.1.2 Gabor Features and Connected Component Analysis

This feature extraction method uses Gabor Filter and Connected Component Analysis basically. While applying the methods, the enhancement algorithms are also applied to improve quality. Image filtering, erosion, dilation can be one of those algorithms. All of these methods are being mentioned in this section in detail. The method's order can be seen in Figure 2.5. Firstly the image is blurred to reduce noise effect. By applying Gabor filter to the filtered image, the phase value is obtained. Erosion, dilation and sobel filtering are applied to the phase image in order. Finally Connected Component Analysis is used to calculate the length of the wrinkles of the image.

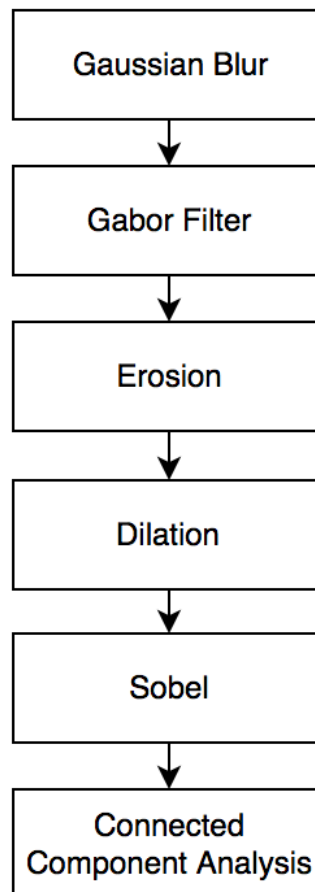


Figure 2.5 : Gabor Features based on Connected Component Analysis flow

Before applying feature extraction approaches, the image is filtered with Gaussian blur to overcome possible noise in the image [22]. The Gaussian function which is applied to each pixel of the image is represented as follows;

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.1)$$

where σ is assigned to 5 in this study. Gabor features are known that it is one of the closest methods to human visual system. Gabor filter includes different scales and orientations which is combined of gaussian kernel function and sinusoidal functions. The common usage for face recognition technics use 5 different scale and 8 different rotation [23].

Let $I(x,y)$ represent an $n \times m$ pixels image, ψ defines Gabor filter with frequency f and orientation θ . By using convolution the image can be filtered with Gabor filter ;

$$G_{u,v}(x,y) = I(x,y) * \psi^{u,v}(x,y) \quad (2.2)$$

The convolve output G can be separated into its imaginary and real parts;

$$E_{u,v}(x,y) = RE[G_{u,v}(x,y)] \text{ and } O_{u,v}(x,y) = IM[G_{u,v}(x,y)] \quad (2.3)$$

By using the separated parts phase and the magnitude of the image can be computed as follows:

$$A_{u,v}(x,y) = \sqrt{E_{u,v}^2(x,y) + O_{u,v}^2(x,y)} \quad (2.4)$$

$$\phi_{u,v}(x,y) = \arctan\left(\frac{O_{u,v}(x,y)}{E_{u,v}(x,y)}\right) \quad (2.5)$$

Five different scales and eight different orientations give forty filtered image samples. The faecture size that has to be carried for each image becomes 163840, if 64×64 pixels images are taken into account. To overcome this issue down-sampling or dimension selection methods can be applied [23]. Generally Gabor phase information is not taken into account in face applications because it varies simply. The magnitude and phase information of an image can be seen in Figure 2.6. In this study, phase information of an image is found proper to use. It is because the wrinkle number will be used as a feature and the bright part of the phase value represents the wrinkle lines in Figure 2.6. By applying other image processing technics these lines can be extracted to define wrinkle severity of the images.

In this study, 8 scales and 90 orientations are used to built up Gabor Filter Bank. The phase information is found informative to examine. At this point the erosion and

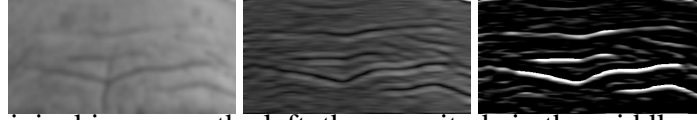


Figure 2.6 : Original image on the left, the magnitude in the middle and the phase on the right.

dilation processes are needed to be applied because of false detection. Connected component analysis tries to find the number of wrinkles in an image. The small artifacts can be false alarm. To prevent this, erosion process is applied firstly. Erosion which is applied to binary or grayscale images is an morphological operator to shrink foreground features of an image [24]. The patch is applied to the each pixel of the image. If its neighbours belong to the foreground, the pixel is assigned as foreground pixel otherwise not. Furthermore, dilation process is also similar to the erosion. The main difference is that while erosion tries to shrink the foreground, dilation tries to expand it [24]. If the center pixel of the 3×3 structuring element hits a foreground pixel, its neighbours are defined as foreground pixels. The structuring element can have different size or different neighbour options according to the problem.

Then, Sobel filter is applied to final image. Sobel filter can be used in edge detection algorithms. While wrinkle detection can be also defined as an edge detection, the 3×3 pixels Sobel patch is applied to the phase matrices. The Sobel matrices are defined as follows;

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * A \quad \text{and} \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \quad (2.6)$$

Let A is an input image, the matrix on both sides defines the sobel filters. While the outcome G_x represents horizontal derivative of the image, G_y represents vertical derivative of the same image. The total magnitude is defined as follows;

$$G = \sqrt{G_x^2 + G_y^2} \quad (2.7)$$

The image that is applied Sobel filter can be seen in Figure 2.7.

When the final images are obtained, the feature extraction method is applied to represent the image. The final image is also meaningful. It displays the possible wrinkle locations. When the final images with different severity scales are examined, one can simply understand that the number of the detected wrinkles increases in reverse order of severity scale. Therefore the count of the wrinkles are taken into account as

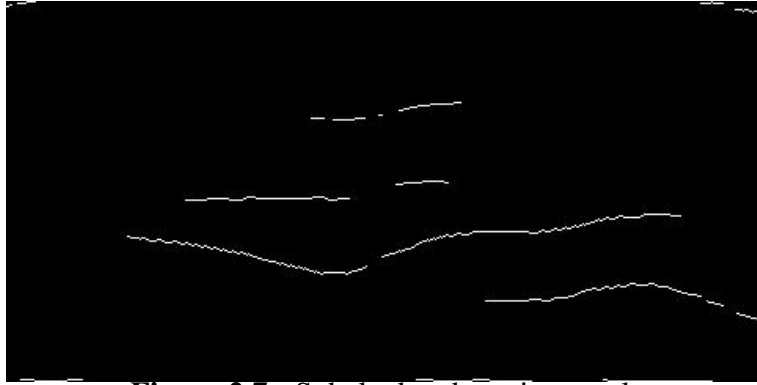


Figure 2.7 : Sobel edge detection result

feature to represent an image. Connected Component Labeling with using two-pass algorithm is applied to find the number of the wrinkles [25]. The labeling algorithm is shown in Algorithm1. The calculated label size defines as wrinkle count. Therefore, this number can be used as a feature for each image.

Data: Final image

Result: The number of detected wrinkles

linked = [];

labels = structure with dimensions of data, initialized with the value of

Background;

First pass;

for row in data: **do**

for column in row: **do**

if data[row][column] is not Background **then**

 neighbors = connected elements with the current element's value;

if neighbors is empty **then**

 linked[NextLabel] = set containing NextLabel;

 labels[row][column] = NextLabel;

 NextLabel += 1;

else

 Find the smallest label;

 L = neighbors labels;

 labels[row][column] = min(L);

for label in L **do**

 | linked[label] = union(linked[label], L);

end

end

else

end

end

end

Second pass;

for row in data: **do**

for column in row: **do**

if data[row][column] is not Background **then**

 | labels[row][column] = find(labels[row][column])

else

end

end

end

return labels;

Algorithm 1: Connected Component Labeling Algorithm [25]

2.2 Classification

Two different classification methods were used to train features. While Support Vector Machine (SVM) was used for LBP method, K Nearest Neighbour was used for Connected Component Analysis method.

2.2.1 Support Vector Machines

SVM is a supervised learning method to classify given input values according to its actual labels. The data can be linearly separable or non-linearly separable [26]. Support vectors are the closest ones to the decision boundary. While x_i defines the features of the class, $y_i \in \{+1, -1\}$ states the label of the binary classification. $\mathbf{w}^T \mathbf{x} + b = 0$ is used for optimal hyperplane equation which maximize the $\frac{1}{\|\mathbf{w}\|}$ margin. The feature points are stayed in these equations as in equation 2.8 and 2.9;

$$\mathbf{w}^T \mathbf{x} + b \geq 1 \quad \text{if } y_i = +1 \quad (2.8)$$

$$\mathbf{w}^T \mathbf{x} + b \leq -1 \quad \text{if } y_i = -1 \quad (2.9)$$

The main purpose of it is to try to find the best hyperplane which maximizes the margin to separate data correctly. The farthest margin is found, the less error is obtained while classifying the new values. f, g constraints functions are added to produce the new Lagrangian function with slack variable of α .

$$L(x, \alpha) = f(x) - \alpha g(x) \quad (2.10)$$

In general;

$$L(x, \alpha) = f(x) - \sum_i \alpha_i g_i(x) \quad (2.11)$$

In this case;

$$L_P = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_i \alpha_i y_i (x_i \cdot \mathbf{w} + b) + \sum_i \alpha_i \quad (2.12)$$

$$\frac{\partial L_P}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i x_i = 0 \quad (2.13)$$

$$\frac{\partial L_P}{\partial b} = \sum_i \alpha_i y_i = 0 \quad (2.14)$$

At this point, Lagrangian Dual problem is applied to the derived form of the equation. The aim is to try to maximize the α that is dual variable instead of trying to minimize w and b . Furthermore the equation can be finalized as in Equation 2.15;

$$L_D(a_i) = \sum_i^l \alpha_i - \frac{1}{2} \sum_i^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2.15)$$

This equation is to separate the classes which are linear. When the data is not linearly separable, it is mapped into a higher space by using kernel function K to obtain linear separation. The equation becomes;

$$L_D(a_i) = \sum_i^l \alpha_i - \frac{1}{2} \sum_i^l \alpha_i \alpha_j y_i y_j \mathbf{K}(x_i \cdot x_j) \quad (2.16)$$

where K stands for mapping function which can be polynomial, radial basis function (RBF) or sigmoid. The RBF is defined as follows;

$$K(x, y) = \exp \left\{ -\frac{\|x-y\|^2}{2\sigma^2} \right\} \quad (2.17)$$

2.2.2 K Nearest Neighbour

KNN is a supervised learning algorithm that can be used for both classification and regression. In this study it is used to classify features that are extracted by using Gabor features based on Connected Component Analysis. This method produces wrinkle number as an output which is specified as feature for each image. KNN algorithm is defined as follows;

Data: Final image

Result: The number of detected wrinkles

1. Calculate the distance from test feature to all train features.
2. Select the closest k of them.
3. By using majority voting find the test sample label.

Algorithm 2: KNN algorithm

The distance is calculated using Euclidean formula for multidimensional vectors;

$$d(p, q) = d(q, p) = \sqrt{\sum_1^n (q_i - p_i)^2} \quad (2.18)$$

where d stands for distance function. q and p define test and train feature vectors. n is the number of test samples. After calculating the distance for a test sample, the k closest values are selected for majority voting. In this study k is selected as 3. This

means that the chosen values are the closest values to the test sample. If two or three out of the three samples belong to the same class, the test sample is labeled with that class. If all outputs belong to the different class, the calculated distances are taken into account. It is assumed that the closest distance defines the test sample more.

2.2.3 Evaluation

Each train set includes same number of images. Therefore the test images for each category can be different. For that reason, the accuracies must be calculated for each category separately. Then the main accuracy to represent the result can be calculated by dividing the sum of the ratios to category numbers. The accuracy is calculated using true positive rate for each category as follows;

$$TPR = \frac{TP}{P} \quad (2.19)$$

where as TPR defines true positive rate, TP stands for true positive samples and P is the number of the positive cases. The correctly classified samples divided to total positive numbers for the category. Then the accuracy for the category is obtained. Since the study specified three main category, there must be three accuracies in total. The ratio of the accuracies gives the exact accuracy of the proposed method.

3. DATABASE COLLECTION

In this research an annotated wrinkle database is required from different parts of faces. Three different regions of the face are selected to investigate. These are forehead, right eye and left eye which are called forehead lines, periorbital lines (left and right) respectively. The selected regions have to be annotated according their wrinkle severities. Owing to the fact that there is no database that can be used for this purpose. New database was collected and annotated. All images were taken from volunteers. In addition, every volunteer signed the informed consent form.

Database must be collected different people that are at various ages. Besides, wrinkles happen when people get older, the other environmental effects can also be a reason. Therefore one can not be say that old people have wrinkles. The rules were specified before collecting database. In computer vision applications light and angle are the crucial topics that affect system reliability. Therefore the images that are taken must be under same light. The angle of the desired region and camera must be the same for all images.

3.1 Database Collection

CANON professional photo machine was used. Images were taken without and with flash light consecutively. Therefore every subject has two images for every region. When the test and train sets are separated, one subject has to be in one set with its all images. Person independent database are provided. Every region has different number of subjects. The number of images that were taken from volunteers are listed in Table 3.1.

3.1.1 Forehead

The middle point of the eyebrows was marked manually to align images. Middle point can be seen in Figure 3.1. By using the annotated point images were cropped and scaled to 482×242 pixels. The middle point is at fixed point of images which

Table 3.1 : Number of images and person that contributed to the database for each region

	Number of Person	Number of Images
Forehead	29	58
Left Eye	29	58
Right Eye	29	58

is (242, 242). The main drawback of the forehead is hair of the volunteers. While taking photos, this situation was handled effectively. Therefore, this drawback effect was reduced. The samples of the cropped and aligned images with and without flash are shown in Figure 3.2.



Figure 3.1 : The green points states the center of the two eyes.



Figure 3.2 : The cropped and aligned forehead images with and without flash.

3.1.2 Left Eye

For this purpose, the layout of the camera was used to take photos. Same angles were used for each volunteers while shooting. The eye and the ear were aligned using camera coordinates. After taking photos, the left eye and the left ear were marked to align images. By using the marked coordinates the photos were cropped by size of 242×482 pixels. The left eye images were also taken with and without flash light which can be shown in Figure 3.3.



Figure 3.3 : The cropped and aligned left eye images with and without flash.

3.1.3 Right Eye

The left and right eye images were taken separately because the previous researches show that the wrinkle severity of the left and right eye can be different for people. For this purpose, the layout of the camera was used to take photos. Same angles were used for each volunteers while shooting. The eye and the ear were aligned using camera coordinates. After taking photos, the right eye and the right ear were marked to align images. By using these marked coordinates the photos were cropped by size of 242×482 pixels. The right eye images were also taken with and without flash light which are shown in Figure 3.4



Figure 3.4 : The cropped and aligned right eye images with and without flash.

3.2 Database Annotation

3.2.1 Raters

The database consists of 58 forehead images of 29 people, 58 left eye images of 29 people and 58 right eye images of 29 people. The participants were volunteers not patients. The participants were not wearing make up and did not have any facial surgery in three years. The database was annotated by 10 educated volunteers. Five of the raters were females and five of them were males. To exclude any bias, the cropped parts of the face were only shown to the raters. In addition, the left and right eye of the images were also rated separately. All of the images were scored in one session. There were any time gap between scoring processes.

Firstly, all images were shown to the raters quickly. The decision were left to their observation. Then the images were shown in order. The total process took half an hour approximately.

3.2.2 Categories

According to previous researches which are mentioned in Section 2.2 the wrinkles can be graded such as 0 to 11, 0 to 5. When the number of images in the database are taken into the account, the class number was defined as three. The classes are specified with the number of 1, 2 and 3. The absence of the wrinkle did not counted for this research.

Instead of this, absence of the wrinkles and mild wrinkles were categorized into the same class that is assumed as 1. The class 2 consists of moderate wrinkles. The final class 3 includes the severe and extreme wrinkles. The wrinkle severity scale can be seen in Table 3.2. The reference images were not provided to the raters. The decision was left to the their observation and perspective.

Table 3.2 : Wrinkle severity scale

Categories	Wrinkle Severities
1	absence - mild
2	moderate
3	severe - extreme

3.2.3 Raters' Reliability

To use the wrinkle severity as label, there must be agreement between raters. There are different reliability measurement systems for this purpose. The Fleiss's Kappa coefficient were used to measure inter-rater reliability [27]. The method is a statistical approach which was derived from Scott's pi statics to measure reliability [28]. The measurement system is also available for the systems that have multiple raters. The agreement is defined as follows;

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \quad (3.1)$$

$\bar{P} - \bar{P}_e$ defines the actual agreement while $1 - \bar{P}_e$ states the obtainable agreement by chance. If the κ equals 1, it means that all raters agreed to one option. If the κ coefficient is less than or equals to 0, it states that there is not any agreement between the raters. For the rest of the κ values, the mentioned strength of agreement chart is used [29]. It is listed in Table 3.3.

Let n be the number of raters, N be number of samples and k be the categories. The \bar{P}_e and \bar{P} can be defined as follows [30];

$$p_j = \frac{1}{Nn} \sum_i^N n_{i,j} \quad (3.2)$$

$$\bar{P}_e = \sum_j^N p_j^2 \quad (3.3)$$

$$P_i = \frac{1}{n(n-1)} \sum_j^k n_{i,j}(n_{i,j} - 1) \quad (3.4)$$

$$\bar{P} = \frac{1}{N} \sum_i^N P_i \quad (3.5)$$

The calculated κ coefficients for each region can be seen in Table 3.4

Table 3.3 : Measurement of κ coefficient

κ	Agreement
0	poor
0.0 - 0.20	slight
0.21 - 0.40	fair
0.41 - 0.60	moderate
0.61 - 0.80	substantial
0.81 - 1	almost perfect

Table 3.4 : Results for each region

Regions	κ coefficients
forehead	0.623
left eye	0,790
right eye	0,696

In this study, $n = 10$, $N = 58$ and $k = 3$ are used. The coefficients are 0.623, 0.790, 0.696 which are related to forehead, left and right eye respectively. The all κ coefficients are in substantial interval. The inter-rater reliability was found sufficient for this study. Therefore the images were graded with the reliability of 0.623, 0.790, 0.696 by the raters. The graded images can be seen in Figure 3.5, 3.6, 3.7 for each region.



Figure 3.5 : The image on the upper left side belongs to class 1. The image on the upper right side belongs to class 2. The third image belongs to class 3.



Figure 3.6 : The image on the upper left side belongs to class 1. The image on the upper right side belongs to class 2. The third image belongs to class 3.

3.3 Database Separation

The scored images information can be shown in Table 3.5, 3.6 and 3.7. While dividing test and train sets, person independency was provided as there are two images of a person in the database. One of them is with flash light and the other one was taken without flash light. Therefore;

- If a person was selected for train set, all images of the same person are put in the train set or vice versa.

The other issue is to keep train set size the same for all categories;

- For each category 10 images were selected for training.

Rest of the images were selected for testing.



Figure 3.7 : The image on the upper left side belongs to class 1. The image on the upper right side belongs to class 2. The third image belongs to class 3.

Table 3.5 : Number of labeled forehead images.

	1	2	3
train	10	10	10
test	10	8	10

Table 3.6 : Number of labeled left eye images.

	1	2	3
train	10	10	10
test	8	8	12

Table 3.7 : Number of labeled right eye images.

	1	2	3
train	10	10	10
test	10	10	8

4. EXPERIMENTS

Two feature extraction methods were applied to train and test sets. The results will be shown in the following sections. The 3-fold cross validation is applied to provide reliability and reproducibility. The results of each fold will also be listed on the tables of the related sections.

4.1 Classification Using Local Binary Pattern

LBP features were extracted using 8×8 pixels blocks and 16×16 pixels blocks. 1×59 pixels feature vector is obtained from one block. Since the images are 242×482 pixels, the block number becomes 1800 for each image. The feature vector is built up the size of 1×106.200 pixels to represent an image. When the 16×16 pixels blocks are used, 450 blocks are obtained. Therefore, the feature vector is 1×26.550 pixels. Features are trained using multi-class SVM that uses two different versions which are linear SVM and RBF SVM. The experiments were repeated using linear and RBF SVM for each fold. The results for forehead, left eye and right eye can be seen in Table 4.1, 4.3 and 4.5 respectively.

4.1.1 Forehead

The classification results of wrinkles on the forehead are listed in Table 4.1. The most successful accuracy is 64.5%, when the 16×16 pixels blocks are trained with linear SVM. 8×8 pixels blocks with linear SVM follows the best result with the accuracy of 57.2%. When all results of linear SVM are examined, 16×16 pixels blocks outperform 8×8 pixels blocks for each fold. The 16×16 pixels blocks using RBF SVM took the third place with the accuracy of 48.9%, while 8×8 pixels blocks using RBF SVM took fourth place with the accuracy of 44.3%. The linear SVM also outperforms RBF SVM for each folds.

The confusion matrix was constructed for the most successful fold of the result that has highest accuracy (Linear SVM with 16×16 pixels blocks). The results can be

seen in Table 4.2. Confusion matrix helps to understand which classes cannot be separated correctly. According to the Table 4.2, mild wrinkles were distinguished easily from extreme and moderate wrinkles.

Table 4.1 : The classification results of the wrinkles on the forehead using LBP

Forehead	8 × 8		16 × 16	
	Linear	RBF	Linear	RBF
1	58.3	42.5	61.7	46.7
2	52.5	41.2	67.5	45
3	60.8	49.2	64.2	55
Total	57.2	44.3	64.5	48.9

Table 4.2 : The confusion for the forehead wrinkles. First column shows actual classes whereas first row shows predicted classes.

Act. - Pred.	1	2	3
1	9	1	0
2	1	5	2
3	2	4	4

4.1.2 Left Eye

For the classification of the wrinkles on the left eye, features are extracted with 16 × 16 pixels and 8 × 8 pixels blocks using linear and RBF SVM. The results are listed in Table 4.3. According to the results, the linear SVM with 16 × 16 pixels blocks achieved the highest accuracy of 64.4%. Linear SVM with 8 × pixels blocks, RBF SVM with 16 × 16 pixels blocks and RBF SVM with 8 × 8 pixels blocks follows the highest result with the accuracies of 61.6%, 51.9% and 42.2% respectively.

The confusion matrix was constructed by using the most successful method which is linear SVM with 16 × 16 pixel blocks. It is listed in Table 4.4. The moderate and extreme wrinkles are the most confused ones when they are compared to mild wrinkles.

4.1.3 Right Eye

The features are extracted 8 × 8 pixels and 16 × 16 pixels blocks to classify the wrinkles on the left eye. Linear and RBF SVM were used to train features. The results are listed

Table 4.3 : The classification results of the wrinkles on the left eye using LBP

Right Eye	8 × 8		16 × 16	
	Linear	RBF	Linear	RBF
1	73.6	55.6	77.8	65.3
2	59.7	32	67.5	43.1
3	51.4	36	64.2	47.2
Total	61.6	42.2	64.4	51.9

Table 4.4 : The confusion matrix of the wrinkles on the left eye. First column shows actual classes whereas first row shows predicted classes.

Act. - Pred.	1	2	3
1	7	1	0
2	0	7	1
3	0	5	7

in Table 4.5. Linear SVM with 16 × 16 pixels blocks achieved the highest accuracy of 63% as same as the forehead classification. Linear SVM with 8 × pixels blocks, RBF SVM with 8 × 8 pixels blocks and RBF SVM with 16 × 16 pixels blocks follows the highest result with the accuracies of 56.4%, 53.6% and 47% respectively.

The confusion matrix was constructed with linear SVM with 16 × 16 pixels blocks. One can see that the mild wrinkles. The moderate and extreme wrinkles were mixed mostly.

Table 4.5 : The classification results of the wrinkles on the right eye using LBP

Left Eye	8 × 8		16 × 16	
	Linear	RBF	Linear	RBF
1	65	63.3	70	47.5
2	58.3	55.8	70.8	51.7
3	45.8	41.7	48.3	41.7
Total	56.4	53.6	63	47

Table 4.6 : The confusion matrix of wrinkles on the right eye. First column shows actual classes whereas first row shows predicted classes.

Act. - Pred.	1	2	3
1	10	0	0
2	2	7	1
3	2	2	3

4.2 Classification Using Gabor Filter Based On Connected Component Analysis

The features were extracted by using Gabor Filter based on Connected Component Analysis. The feature is the number of the detected wrinkles on the images. By using this feature K Nearest Neighbour algorithm was applied with Majority Voting. K is 3 in the experiments. When the wrinkle number for the test samples were examined, it was understood that the images that belong to the same person (with and without flash light) have the same number of the wrinkle. Therefore instead of calculating the KNN for both images only one image was used per person.

4.2.1 Forehead

The classification result of wrinkles on the forehead was listed in Table 4.7. The result was the accuracy of 79.3%. The classification of the mild and extreme wrinkles were the accuracies of 93.3% and 86.7% respectively, while the moderate wrinkle classification was the accuracy of 58.3%. The confusion matrix can be seen in Table 4.8 which belongs to most successful test set. The moderate wrinkles were confused with the extreme wrinkles. The mild and extreme wrinkles were classified much better.

Table 4.7 : The result of the classification of the wrinkles on the forehead using Gabor features

	Gabor Filter - Connected Component Analysis			Mean
	1	2	3	
1	100	50	80	76.7
2	80	75	80	78
3	100	50	100	83.3
Mean	93.3	58.3	86.7	79.3

Table 4.8 : Confusion matrix of the forehead wrinkles. First column shows actual classes whereas first row shows predicted classes.

Act. - Pred.	1	2	3
1	10	0	0
2	0	4	4
3	0	0	10

4.2.2 Left Eye

The classification of the wrinkles on the left eye was the accuracy of 73% which can be seen in Table 4.9. The mild wrinkles were classified the accuracy of 91.6% while moderate and extreme wrinkles were classified the accuracies of 66.7% and 60.7% respectively. The confusion matrix were listed in Table 4.10. The result was the same as the classification results. The moderate and extreme wrinkles were confused mostly. The classification rate of the forehead wrinkles was perfect.

Table 4.9 : The result of the classification of the wrinkles on the left eye using Gabor features

	Gabor Filter - Connected Component Analysis			Mean
	1	2	3	
1	75	75	66	72.2
2	100	50	66	72
3	100	75	50	75
Mean	91.6	66.7	60.7	73

Table 4.10 : Confusion matrix of the wrinkles on the left eye. First column shows actual classes whereas first row shows predicted classes.

Act. - Pred.	1	2	3
1	8	0	0
2	0	6	2
3	0	6	6

4.2.3 Right Eye

The classification of the wrinkles on the right eye achieved the average accuracy of 78.9%. The all test samples of mild wrinkles were classified correctly. Moderate and extreme wrinkles were classified the accuracies of 46.7% and 83.3% respectively. The confusion matrix were listed in Table 4.12. The mild and extreme wrinkles were classified correctly while moderate wrinkles were confused with the extreme wrinkles mostly.

Table 4.11 : The result of the classification of the wrinkles on the right eye using Gabor features

	Gabor Filter - Connected Component Analysis			
	1	2	3	Mean
1	100	40	75	78.3
2	100	40	100	80
3	100	60	75	78.3
Mean	100	46.7	83.3	78.9

Table 4.12 : Confusion matrix of the wrinkles on the right eye. First column shows actual classes whereas first row shows predicted classes.

Act. - Pred.	1	2	3
1	10	0	0
2	0	4	4
3	0	0	10

5. DISCUSSIONS

This section gives a brief summary of the methods and the results of the experiments. The dataset includes three face regions which are forehead, left eye and right eye. Another crucial part of the face image is mouth for the wrinkle detection. These wrinkles are commonly occurred getting age in time and daily activities such as speaking, laughing, yawning. The wrinkles on the mouth were also shot to examine in the scope of the study. The points of the camera were aligned using nose and lip points therefore the wrinkles were shot from front side. Moreover when the results are examined, the wrinkles on the mouth were not found distinguishable enough. For that reason, the methods were not qualified to detect and classify the wrinkles on the mouth. LBP method uses patches to create feature vector of the image. Wrinkles are close to the nose and the lip. The patches also includes the nose and lip parts in addition to the region of interest. It leads to the wrong classification results. When the images were cropped to include only wrinkle part but not the lip and nose lines, the cropped images became meaningless. All of the severities started to looked alike. Meaningful results were not obtained from those experiments. The shooting angle and the length between camera and the face were not qualified enough for wrinkles on the forehead. The images should have only included the wrinkles. With the appropriate angle and distance this can be achieved in the future works.

Two different feature extraction methods were used for each part of the face. The test and train sets were divided into three mixed groups (3-fold cross validation). The results were consistent between groups. When the LBP were used for the classification of the wrinkles on the forehead, 16×16 pixels patches using linear SVM achieved most successful results. If the patch size is taken into the account, the 16×16 pixels patches outperformed 8×8 pixels patches for linear and RBF SVM methods. If the training methods are considered, linear SVM achieve higher results than RBF SVM. The obtained results for training methods and patch size are valid for each fold. The confusion matrix of the LBP method also states that the mild wrinkles

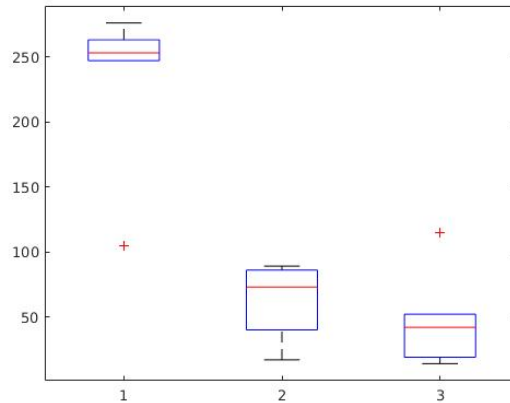


Figure 5.1 : Box plot of the test samples for each fold of the wrinkles on the forehead. Horizontal line stands for the categories and vertical line stands for the wrinkle number.

were distinguished much better than moderate and extreme wrinkles. The wrinkles that belong to second and third categories were mixed mostly. The Gabor features based on Connect Component Analysis method outperformed LBP method for each fold. In addition according to the confusion matrix the mild and extreme wrinkles were classified successfully while moderate wrinkles were misclassified mostly as in Figure 5.1. All in all the count of the wrinkles on the forehead is a much better way than LBP features to distinguish wrinkle severity.

16×16 pixels patches using linear SVM outperforms the other methods when the LBP features were used to classify the wrinkles on the left eye. If the patch size is considered for linear and RBF SVM, 16×16 pixels patches achieve higher results than 8×8 pixels patches for each fold. If the training methods are taken into accounts, linear SVM also outperforms RBF SVM for each fold. According to the confusion matrix, mild and moderate wrinkles were classified more accurate while extreme wrinkles misclassified with the moderate wrinkles mostly. The Gabor features based on Connected Component Analysis also achieved higher results than LBP method. This also proves that using wrinkle number is more distinguishable than LBP features for the classification of the wrinkle on the left eye. The confusion matrix of the Gabor features also states that the mild and moderate wrinkles were classified much better than extreme wrinkles as in Figure 5.2. In the end the Gabor features based on Connected Component Analysis achieved higher results than LBP features.

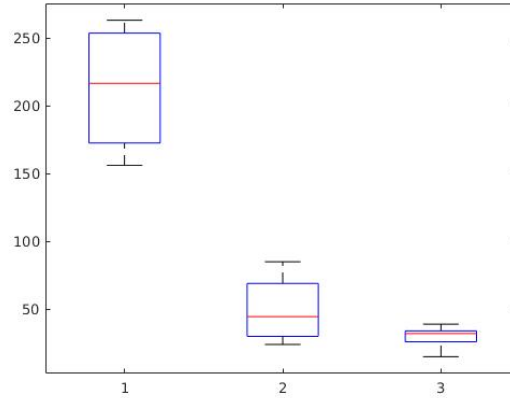


Figure 5.2 : Box plot of the test samples for each fold of the wrinkles on the left eye. Horizontal line stands for the categories and vertical line stands for the wrinkle number.

16×16 pixels patches using linear SVM outperformed other methods while classifying wrinkle on the right eye using LBP features. While 16×16 pixels patches using linear SVM achieves higher accuracy than 8×8 pixels patches using linear SVM, 8×8 pixels patches using RBF SVM were more successful than 16×16 pixels patches using RBF SVM. This results repeat themselves for each fold. The confusion matrix also states that mild wrinkles were classified successfully while moderate and extreme wrinkles were mixed each other mostly. The Gabor features based on the Connected Component Analysis outperforms LBP features. The method proves that the count of the wrinkles are more successful than LBP features to classify wrinkles on the right eye. In addition, its confusion matrix shows that mild and extreme wrinkles can be classified accurately while moderate wrinkles were being mixed with the other categories as in Figure 5.3. All in all Gabor features based on Connected Component Analysis is more useful than LBP features to classify wrinkles on the right eye.

In the scope of the study using number of the wrinkles was found more distinguishable than LBP features. According to the previous researches about detecting wrinkles, LBP text features were applied. The results were not excellent. Then another text descriptor method were applied which was Gabor features. Gabor features gives an opportunity to visualize the wrinkles on the images. After applying additional methods to improve the efficiency, the number of the wrinkles on an image was found as a feature. The trick is that one can easily understand that the wrinkle number of the mild category is less than the others. On the other hand the wrinkle number of the first category (mild wrinkles) was extremely high. Besides, extreme and moderate

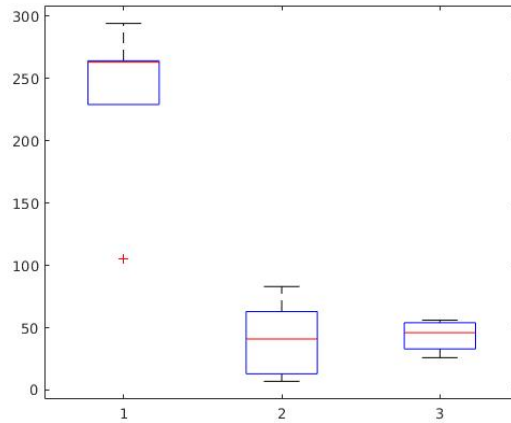


Figure 5.3 : Box plot of the test samples for each fold of the wrinkles on the right eye. Horizontal line stands for the categories and vertical line stands for the wrinkle number.

wrinkles had less wrinkle count. The extreme wrinkle count was lesser than moderate wrinkles. Therefore the moderate wrinkles were the one that was mixed with the extreme wrinkles. To improve this method, the length of the wrinkles was also taken into account. The length and the number of the wrinkles were accepted as a feature of the images. Unfortunately, the results were not satisfactory. As in Figure 5.4, when the categories were tried to be displayed on the 2D coordinate system, the data were complex to separate linearly. For this reason the length and the number of the wrinkle cannot be used as a feature to represent an image for the Gabor features based on the Connected Component Analysis method.

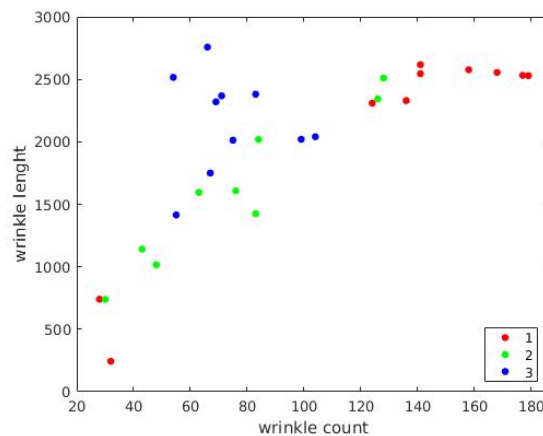


Figure 5.4 : Wrinkle number and wrinkle length relation

The main drawback of the study was the database. There are few images for each category. If the database were expanded, the results may be better.

The second consideration was the feature for the Gabor features using Connected Component Analysis. The wrinkle length was the main and only feature. For the classification of the wrinkles on the forehead, wrinkle length was also calculated. The wrinkle length and the wrinkle count were used together. The features were plotted in Figure 5.4. X coordinate represents wrinkle count while Y coordinate represents wrinkle length. The features could not be separated linearly or grouped. Therefore using the length and the number of the wrinkles could not be used to classify the wrinkles.

The images were shot with and without flash light. While Gabor features were being used, the calculated wrinkle number were the same for both images. Therefore, it can be said that the method is not affected by the flash light or any light.

6. CONCLUSION

We communicate through our faces. Therefore, the treatments and improvements for the facial beauty is a must to improve the pleasure that people gain from life. The researchers try to find a way to detect and classify wrinkles. Detection and classification of the facial wrinkles were examined together in this research. Previous researches are split into two categories which are cosmetic and computer systems researches. The first part includes human observation to detect and classify wrinkles. The dermatologists, surgeons are the observers of the topics. The latter category are responsible to build automatic systems.

In this research, a database was collected and graded according to its wrinkle severities. Three facial parts were specified for detection and classification. These are forehead, left eye and the right eye. The images were scored one of the three categories by the observers. The categories are mild, moderate and extreme. The Kappa (κ) coefficient was used to find the inter-rater reliability.

For the feature extraction two different methods were used. The first method is LBP. 16×16 and 8×8 pixels patches were used for extraction. After that, LBP features were trained using linear and RBF SVM. Other feature extraction method is Gabor filter. On the other hand in this approach the Gabor features were improved using erosion, dilation, edge detection and connected component analysis. The aim is to obtain wrinkle number of an image as a feature. In addition, the feature were classified using KNN and majority voting.

When the results are observed, the Gabor features outperformed LBP features for each category. The highest accuracy was 64.5% by using LBP features for the wrinkle on the forehead. On the other hand the Gabor features achieved the accuracy of 79.3%. Moreover, the LBP method achieved the accuracy of 64.4% for the classification of the wrinkles on the left eye. However, Using Gabor features gives the higher accuracy which is 73%. Furthermore, while the LBP features achieved the accuracy of 63% for the classification of the wrinkles on the right eye, the Gabor features achieved the

accuracy of 78.9%. All in all the Gabor features based on Connected Component Analysis is more successful than the LBP features to classify facial wrinkles on the forehead, left eye and right eye.

Flash light did not affect the Gabor filtering method. In addition, using wrinkle length and wrinkle number together did not achieved better results than using wrinkle number alone. When mouth images were examined, the methods failed for classification because the images were shot using wrong angle and distance.

To conclude, the study examined the detection and the classification of the wrinkles on the forehead, left eye and the right eye. The Gabor filters based on Connected Component Analysis achieved higher scores.

REFERENCES

- [1] **Ahonen, T., Hadid, A. and Pietikainen, M.** (2006). Face Description with Local Binary Patterns: Application to Face Recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12), 2037–2041, <http://dx.doi.org/10.1109/TPAMI.2006.244>.
- [2] **Ojala, T., Pietikäinen, M. and Mäenpää, T.** (2002). Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7), 971–987, <http://dx.doi.org/10.1109/TPAMI.2002.1017623>.
- [3] **Pietikäinen, M., Hadid, A., Zhao, G. and Ahonen, T.** (2011). *Local Binary Patterns for Still Images*, Springer Nature.
- [4] **Lemperle, G., Holmes, R.E., Cohen, S.R. and Lemperle, S.M.** (2001). A Classification of Facial Wrinkles, *Plastic and Reconstructive Surgery*, 108(6), 1735–1750, <https://doi.org/10.1097%2F00006534-200111000-00048>.
- [5] **Fitzpatrick, R.E.** (1996). Pulsed Carbon Dioxide Laser Resurfacing of Photoaged Facial Skin, *Archives of Dermatology*, 132(4), 395, <https://doi.org/10.1001%2Farchderm.1996.03890280047007>.
- [6] **Day, D.J., Littler, C.M., Swift, R.W. and Gottlieb, S.** (2004). The Wrinkle Severity Rating Scale, *American Journal of Clinical Dermatology*, 5(1), 49–52, <https://doi.org/10.2165%2F00128071-200405010-00007>.
- [7] **Shoshani, D., Markovitz, E., Monstrey, S. and Narins, D.** (2008). The modified fitzpatrick wrinkle scale: A clinical validated measurement tool for nasolabial wrinkle severity assessment, *DERMATOLOGIC SURGERY*, 34(1), S85–S91, <http://dx.doi.org/1854/12788>.
- [8] **John, H.** (2004). The wrinkle and its measurement: A skin surface Profilometric method, *Micron*, 35(3), S201–S219.
- [9] **Lodén, M., Buraczewska, I. and Halvarsson, K.** (2007). Facial anti-wrinkle cream: influence of product presentation on effectiveness: a randomized and controlled study, *Skin Research and Technology*, 13(2), 189–194, <https://doi.org/10.1111%2Fj.1600-0846.2007.00220.x>.
- [10] **Lluncor, D.** Nasolabial Wrinkle Classification, *University of California*.

- [11] **Batool, N. and Chellappa, R.** (2012). Modeling and Detection of Wrinkles in Aging Human Faces Using Marked Point Processes, *Proceedings of the 12th International Conference on Computer Vision - Volume 2, ECCV'12*, Springer-Verlag, Berlin, Heidelberg, pp.178–188, http://dx.doi.org/10.1007/978-3-642-33868-7_18.
- [12] **Batool, N. and Chellappa, R.** (2015). Fast detection of facial wrinkles based on Gabor features using image morphology and geometric constraints, *Pattern Recognition*, 48, 642 – 658, <https://hal.inria.fr/hal-01096629>.
- [13] **Cula, G.O., Bargo, P.R., Nkengne, A. and Kollias, N.** (2012). Assessing facial wrinkles: automatic detection and quantification, *Skin Research and Technology*, 19(1), e243–e251, <https://doi.org/10.1111%2Fj.1600-0846.2012.00635.x>.
- [14] **Frangi, A.** (2001). Three-dimensional model-based analysis of vascular and cardiac images, *Ph.D. thesis*, University Medical Center Utrecht.
- [15] **Ng, C.C., Yap, M.H., Costen, N. and Li, B.** (2014). Automatic Wrinkle Detection Using Hybrid Hessian Filter., *D. Cremers, I.D. Reid, H. Saito and M.H. Yang, editors, ACCV (3)*, volume9005 of *Lecture Notes in Computer Science*, Springer, pp.609–622, <http://dblp.uni-trier.de/db/conf/accv/accv2014-3.html#NgYCL14>.
- [16] **Ng, C., Yap, M.H., Costen, N. and Li, B.** (2015). Wrinkle Detection Using Hessian Line Tracking, *IEEE Access*, 3, 1079–1088.
- [17] **Batool, N. and Chellappa, R.** (2014). Detection and Inpainting of Facial Wrinkles Using Texture Orientation Fields and Markov Random Field Modeling, *IEEE Trans. Image Processing*, 23(9), 3773–3788.
- [18] **Jiang, L.I., Stephens, T.J. and Goodman, R.** (2013). SWIRL, a clinically validated, objective, and quantitative method for facial wrinkle assessment, *Skin Research and Technology*, n/a–n/a, <https://doi.org/10.1111%2Fsrt.12073>.
- [19] **Shan, C., Gong, S. and McOwan, P.W.** (2009). Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study, *Image Vision Comput.*, 27(6), 803–816, <http://dx.doi.org/10.1016/j.imavis.2008.08.005>.
- [20] **Pietikainen, M., Pietikainen, M., Ojala, T., Pietikäinen, M. and Harwood, D.** (1996). A comparative study of texture measures with classification based on feature distributions, *Pattern Recognition*, 29, 51–59.
- [21] **Huang, D., Shan, C., Ardabilian, M., Wang, Y. and Chen, L.** (2011). Local Binary Patterns and Its Application to Facial Image Analysis: A Survey, *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 41(4), 1–17, <http://liris.cnrs.fr/publis/?id=5004>.

- [22] **Nixon, M. and Aguado, A.S.** (2008). *Feature Extraction & Image Processing, Second Edition*, Academic Press, 2nd edition.
- [23] **Štruc, V. and Pavešić, N.**, (2010). From Gabor Magnitude to Gabor Phase Features: Tackling the Problem of Face Recognition under Severe Illumination Changes, In-Tech, Vienna, pp.215–238.
- [24] **Soille, P.**, (2004). *Erosion and Dilation*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp.63–103, http://dx.doi.org/10.1007/978-3-662-05088-0_3.
- [25] **Shapiro, L.; Stockman, G.** (2002). *Computer Vision*, Prentice Hall.
- [26] **Flach, P.** (2012). *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*, Cambridge University Press, New York, NY, USA.
- [27] **Fleiss, J. et al.** (1971). Measuring nominal scale agreement among many raters, *Psychological Bulletin*, 76(5), 378–382.
- [28] **Scott, W.A.** (1955). Reliability of content analysis: The case of nominal scale coding, *Public Opinion Quarterly*, 19(3), 321–325.
- [29] **Landis, J.R. and Koch, G.G.** (1977). The Measurement of Observer Agreement for Categorical Data, *Biometrics*, 33(1).
- [30] **Fleiss, J.L.** (1971). Measuring nominal scale agreement among many raters., *Psychological Bulletin*, 76(5), 378–382, <https://doi.org/10.1037%2Fh0031619>.

CURRICULUM VITAE

Name Surname: Büşra Çanak

Place and Date of Birth: Lüleburgaz 16.10.1990

E-Mail: canakbusra@gmail.com

EDUCATION:

- **B.Sc.:** 2013, Hacettepe University, Engineering Faculty, Computer Engineering Department

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- Çanak B., Kamaşak M.E., 2017. Classification of the Wrinkles on the Forehead. *SIGNAL PROCESSING AND COMMUNICATIONS APPLICATIONS CONFERENCE*, May 15-18, 2017 Antalya, Turkey.