

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DAVRANIŞSAL VE FİZİKSEL MULTİ-BİYOMETRİ İLE KİŞİ
KİMLİKLENDİRME VE HAREKET TANIMA ÜZERİNE
YAKLAŞIMLAR

Onur Can KURBAN

DOKTORA TEZİ
Elektronik ve Haberleşme Mühendisliği Anabilim Dalı
Elektronik Programı

Danışman
Prof. Dr. Tülay YILDIRIM

Haziran, 2022

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**DAVRANIŞSAL VE FİZİKSEL MULTI-BİYOMETRİ İLE KİŞİ
KİMLİKLENDİRME VE HAREKET TANIMA ÜZERİNE YAKLAŞIMLAR**

Onur Can KURBAN tarafından hazırlanan tez çalışması 29.06.2022 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektronik ve Haberleşme Mühendisliği Anabilim Dalı Elektronik Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Tülay YILDIRIM
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Tülay YILDIRIM, Danışman
Yıldız Teknik Üniversitesi

Doç. Dr. Nihan KAHRAMAN, Üye
Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Berk GÖKBERK, Üye
Boğaziçi Üniversitesi

Doç. Dr. Bülent BOLAT, Üye
Yıldız Teknik Üniversitesi

Prof. Dr. Çiğdem EROĞLU ERDEM, Üye
Marmara Üniversitesi

Danışmanım Prof. Dr. Tülay YILDIRIM sorumluluğunda tarafımda hazırlanan Davranışsal ve Fiziksel Multi-biyometri ile Kişi Kimliklendirme ve Hareket Tanıma Üzerine Yaklaşımlar başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Onur Can KURBAN

İmza

*Bu tez, eđitim hayatım boyunca üzerinde emeđi bulunan tüm öđretmenlerime ve
hocalarıma ithaf olunmuştur.*



TEŞEKKÜR

Tüm yüksek öğrenimim boyunca akademik bilgi ve tecrübeleri ile bana her zaman destek olan ve beni yönlendiren, üzerimde çok büyük emekleri olan, öğrencisi olmayı büyük bir onur ve şans addettiğim, tez danışmanım ve hocam Prof. Dr. Tülay YILDIRIM'a çok teşekkür ederim.

Tez izleme jürilerim Doç. Dr. Nihan KAHRAMAN ve Dr. Öğr. Üyesi Berk GÖKBERK hocalarıma, değerli katkıları ve desteklerinden dolayı teşekkürü bir borç bilirim. Çalışmakta olduğum THY Teknoloji ve Bilişim A.Ş. ile THY Teknik A.Ş.'deki yöneticilerim ve çalışma arkadaşlarıma tez dönemindeki desteklerinden dolayı teşekkür ederim.

Tez sürecinde birlikte çalışma fırsatı bulduğum ve yardımlarını hiçbir zaman esirgemeyen meslektaşım Dr. Nurullah ÇALIK'a ve Ahmet BİLGİÇ'e çok teşekkür ederim. Tüm bu süreçler boyunca desteğini her zaman hissettiğim, beraber çalışmayı büyük bir şans olarak gördüğüm, değerleri arkadaşım Ali Rıza YILMAZ'a da çok teşekkür ederim.

Tüm eğitim-öğrenim hayatım boyunca en büyük destekçilerim olan babam Mustafa KURBAN ve annem Seriha KURBAN ile beni akademik alanda her daim destekleyen ve cesaretlendiren ablam Doç.Dr. Banu ÖZDEMİR'e çok teşekkür ederim. Varlıkları ile hayatımdaki herşeyi çok daha anlamlı ve değerli kılan, gönlümün süruru eşim Derya KURBAN ve biricik kızım Fatma Erva KURBAN'a eşsiz katkılarından dolayı teşekkür ederim. Bu tez benden çok onların başarısıdır.

Onur Can KURBAN

İÇİNDEKİLER

SİMGE LİSTESİ	viii
KISALTMA LİSTESİ	ix
ŞEKİL LİSTESİ	x
TABLO LİSTESİ	xiv
ÖZET	xvii
ABSTRACT	xix
1 GİRİŞ	1
1.1 Literatür Özeti	1
1.2 Tezin Amacı	5
1.3 Hipotez	6
2 BİYOMETRİK VERİLER	7
2.1 Silüet Verileri ve Enerji Görüntüsü	7
2.2 Termal Görüntüler	9
2.3 Eklemler arası uzaklık (İskelet) Verileri	10
3 ENERJİ GÖRÜNTÜSÜ VE YÜZ GÖRÜNTÜLERİ İLE SANAL VERİ SETİ OLUŞTURULMASI	12
3.1 Eurocom Face Veri Seti	12
3.2 Davranışsal Biyometrik Veriler ve Yüz görüntüleri ile sanal Veri Seti Eşleme	13
3.3 Davranışsal Biyometrik Hareketlerin Başarımını Arttırmaya Yönelik Çalışmalar	16
3.3.1 Oluşturulan Yeni Enerji Görüntüleri	19
4 HAREKET ENERJİ GÖRÜNTÜSÜ VE HAREKET GEÇMİŞİ GÖRÜNTÜLERİNİN ELDE EDİLMESİNDE YENİ BİR YÖNTEM – ADAPTİF ENERJİ GÖRÜNTÜSÜ	22
4.1 Adaptif Enerji Görüntüsü Yöntemi	22

4.2	Kullanılan Veri Tabanları ve Yöntemler	28
4.2.1	Karşılaştırma	28
4.2.2	Sınıflandırıcılar	34
4.3	DeneySEL Sonuçlar	36
4.3.1	Mevcut Sonuçlarla Karşılaştırma	37
4.3.2	TT'lerin CNN Tabanlı Karşılaştırması	39
4.3.3	Hand-Crafted Tabanlı Karşılaştırma	44
4.4	Adaptif HE Tartışması	49
5	SİLÜET, TERMAL VE İSKELET (STI) VERİ SETİNİN OLUŞTURULMASI	50
5.1	Kinect ve Termal Kameralar ile Toplanan Görüntüler	50
5.1.1	Kinect Kamera ile Toplanan Silüet Görüntüleri	50
5.1.2	Termal Kamera ile Kaydedilen Görüntüler	51
5.2	Bazı Katılımcılardan 2.Oturum Verilerinin Alınması	53
5.3	Kinect Kamera ile Toplanan Verilerdeki Hatalar	53
5.3.1	Hareket Verisinde Kayıp	53
5.3.2	Siluet Tanınmasında Gecikme	54
5.3.3	Tespit Edilen Kayıt Sorunları İçin Uygulanan Çözümler	55
5.3.4	Görüntülerde Derinlik Bilgisinin Tutulması	55
5.4	Kayıtlarda Yapılan Ön İşlemler ve Veri Setlerinin Hazırlanması	56
5.4.1	Enerji Görüntülerinin Hazırlanması için Uygulanan Ön İşlemler	56
5.4.2	Termal Kayıtlar, Uygulanan Ön İşlemler ve Eğitime Uygun Veri Setlerinin Oluşturulması	58
5.4.3	İskelet Görseli Kayıtları ve Eğitime Uygun Veri Setinin Hazırlanması	59
5.4.4	Eğitime Hazır Veri Setleri	60
6	EĞİTİM AĞLARI VE PERFORMANS SONUÇLARI	61
6.1	Ağ Seçim Kriterleri	61
6.1.1	GoogleNet Ağ Mimarisi	61
6.1.2	ResNet-18 Ağ Mimarisi	63
6.1.3	MobileNetV2 Ağ Mimarisi	63
6.2	Eğitim Sonuçları	65
6.2.1	GoogleNet Eğitim Sonuçları	65
6.2.2	ResNet-18 Sonuçları	67
6.2.3	MobileNetV2 Eğitim Sonuçları	69
6.3	Termal-3D Veri Setinde Eksik Veri Testi	70
6.3.1	Aynalanmış Görüntü (Mirrored-Face) Eklenmiş Eğitimler	71
6.3.2	Tüm Yüz Açılarının Bir Arada Eğitilmesi	72

6.4	Eklemler Arası Uzaklık Verilerinin Görüntü Yerine XY Koordinatları Olarak 1D Formatta Konvolüsyonel Bir Ağa Uygulanması (1D-SkeletonNet)	73
6.5	Çoklu Biyometrik Eğitimler	74
6.5.1	Kişi Tanıma Füzyon Eğitimleri	74
6.5.2	Hareket Tanıma Füzyon Sonuçları	77
7	SONUÇ VE ÖNERİLER	79
A	EK TABLOLAR	81
B	EK FORMLAR	84
	KAYNAKÇA	88
	TEZDEN ÜRETİLMİŞ YAYINLAR	93

SİMGE LİSTESİ

D_m	Gözler arası mesafe bilgisi
W_i	Her bir karenin ağırlığı
$B_t(x, y)$	İki boyutlu değişken kümesi
R	Korelasyon değeri
γ	Maksimum ölçekleme faktörü
m_i	M inci kare
Z	Normalize değerler
μ	Ortalama değer
S_k	Scale faktörü
$\sigma_{a,b}$	Standart sapma
N	Videoda ki toplam kare sayısı
$\mu_{1,2,3,4,R}$	Zamansal fonksiyonlar

KISALTMA LİSTESİ

CNN	Convolutional Neural Network
DNN	Depth Neural Network
EER	Equal Error Rate
FAR	False Accepted Rate
FC	Fully Connected
FPS	Frame Per Second
FRR	False Rejected Rate
GesEI	Gesture Energy Image
HE	Hareket Enerjisi
HoG	Histogram of Oriented Gradients
ID	Identification
k-NN	K-Nearest Neighbors Algoritması
LBP	Local Binary Pattern
LDA	Linear discriminant Analysis
MLP	Multi Layer Perceptron
N/A	Not Applicable
PCA	Principal Component Analysis
ReLU	Rectified Linear Units
RGB	Red-Green-Blue
RGB-D	Red-Green-Blue-Depth
ROC	Receiver Operating Characteristic
STI	Silüet-Termal-İskelet
SVM	Support Vector Machine
TT	Temporal Template

ŞEKİL LİSTESİ

Şekil 2.1	Bir yürüyüş çevrimine ait siluet görüntüleri ve bu siluetler ile üretilen yürüyüş enerji görüntüsü [28]	8
Şekil 2.2	Termal yüz görüntü örnekleri	10
Şekil 2.3	Kinect eklem noktalarının model tabanlı çıkarımla belirlenmesi. a) Ham derinlik verisi, b) çıkarımı yapılan vücut parçaları, c) çıkarımı yapılan eklem noktaları [31]	11
Şekil 2.4	Kinect kamera ile elde edilen iskelet yapısı	11
Şekil 3.1	KinectFaceDB varyasyon görünümü: a)önden ışık b)natürel c)ağız açık d)gülümseme	13
Şekil 3.2	a) orijinal resim b) ölçek ve poz normalizasyonu c) ilgi alanı d) aydınlatma normalizasyonu	13
Şekil 3.3	S hareketinin yapılışı	14
Şekil 3.4	S hareketine ait ROC eğrisi ve FAR-FRR grafikleri	15
Şekil 3.5	VGG Face ROC eğrisi ve FAR-FRR grafikleri	15
Şekil 3.6	Çoklu Biyometrik ROC eğrisi ve FAR-FRR grafikleri	16
Şekil 3.7	S hareketine ait a) başlangıç anı b) orta nokta anı c) bitiş noktası görselleri	17
Şekil 3.8	a)GesEI ile elde edilen görüntü b)fGesEI ile elde edilen görüntü c)mGesEI ile elde edilen görüntü	18
Şekil 3.9	GesEI, fGesEI ve mGesEI ile elde edilen görüntülere ait ROC eğrileri kıyaslaması	18
Şekil 3.10	a)1. Hareket b)2. Hareket c)3. Hareket d)4. Hareket için elde edilen sonuçlar	19
Şekil 3.11	a) b) c) ve d) yukarıda tanımlandığı sırayla konulmuştur. e) fGesEI f)GesEI görüntüleridir	20
Şekil 3.12	GesEI yöntemi üzerinde farklı yöntemlerin kıyaslanması grafiği . . .	21
Şekil 4.1	Hareket enerjisi hesaplanmasında kullanılan katsayılar ve grafikleri [24]	23

Şekil 4.2	NATOPS veri setinden 24 farklı hareket üzerinde $(1 - R_N)$ işlemi ile TT grafikleri elde edilmiştir. Hareket çerçeveleri sabit olmasına rağmen bazı hareketler daha erken bazı hareketler daha geç bitmiştir. Bazı hareketlerde başlangıç kısımları vurgulanırken, bazılarında ise hem başlangıç hem de orta kısımlar vurgulanmıştır. Önerilen yöntem hareketlere uyarlanabilir olduğu için elde edilecek TT'ler veri setine ve hareketlerin içeriğine göre değişiklik gösterecektir	26
Şekil 4.3	[38]'te gösterildiği gibi NATOPS veri tabanındaki 24 hareketin kalıpları. Oklarla gösterilen yönde hareketler tamamlanır	29
Şekil 4.4	İlk satır RGB görüntüler yer almaktadır. İkinci satırdaki görüntüler, Kinect kamera ile toplanan ham derinlik görüntüleridir. Son satır ise RGB görüntüleri ve ham derinlik görüntüleri ile oluşturulan maskelenmiş görüntüleri göstermektedir	30
Şekil 4.5	NATOPS veri tabanı kullanılarak iki farklı enerji görüntüsü veri seti oluşturulmuştur. Bu görüntü, maskelenmiş derinlikli video akışlarında μ_R kullanılarak oluşturulan "NATOPS <i>FullBody</i> " (sol taraf) ve "NATOPS <i>MovingLimbs</i> " (sağ taraf) veri kümelerinin örneklerini göstermektedir	30
Şekil 4.6	[39]'da gösterildiği gibi SBU kinect veri tabanındaki 8 etkileşim kalıbı. a) yaklaşma, b) uzaklaşma, c) tekmeleme, d) yumruk atma, e) itme, f) sarılma, g) el sıkışma h) nesne değiş tokuşu. Yumruklama-itme, sarılma-yaklaşma ve tokalaşma-nesne değiş tokuşu gibi bazı hareketler birbirine yapısal olarak benzerlik gösterebilmektedir. Toplam hareket çeşitliliği daha az olsa bile, bu benzerlikler veri setini daha zorlu hale getirmektedir	31
Şekil 4.7	SBU Kinect veri tabanı kullanılarak iki farklı HE veri seti oluşturulmuştur. Bu görüntülerde, derinlikli video akışlarında μ_R fonksiyonu kullanılarak oluşturulan "SBU <i>FullBody</i> " (sol taraf) ve "SBU <i>Moving-Limbs</i> " (sağ taraf) veri kümelerinin örneklerini göstermektedir	32
Şekil 4.8	[35]'te gösterildiği gibi BodyLogin Silhouettes and Skeletons veri kümesindeki S-hareketi ve kullanıcı tanımlı hareketlerin kalıpları . .	32
Şekil 4.9	[37]'te gösterildiği gibi BodyLogin Posture, Build ve Dynamics veri kümesindeki sol-sağ el hareketi, çift elle kavis hareketi ve denge hareketlerinin yapısal kalıpları sırayla sunulmuştur	33

Şekil 4.10	Adaptif TT ve mevcut TT'ler kullanılarak üretilen HE görüntülerinin örnekleri sırasıyla NATOPS, SBU Kinect ve BodyLogin veri kümeleri için sunulmaktadır. Örnekler, soldan sağa μ_2 , μ_3 , μ_4 ve μ_R fonksiyonları kullanılarak elde edilen enerji görüntüleridir. Fonksiyonların çıktıları bu örneklerle daha iyi bir şekilde görülebilmektedir	34
Şekil 4.11	TT tarafından üretilen enerji görüntülerinin CNN ağlarında sınıflandırması için genel çerçeve. VGG-F ve ResNet-18 ağları probleme göre revize edilmiş ve yeniden eğitilmiştir. CNN'lerin FC katmanları, sınıf sayısına göre değiştirilmiştir	35
Şekil 4.12	ResNet-18 ağının Validation-Loss grafikleri. Sırasıyla μ_2 , μ_3 , μ_4 ve μ_R fonksiyonlarının Validation-Loss değişiklikleri karşılaştırması. Son iterasyonlarda eğriler alt şekil içinde büyütülmüştür. Alt şekil kutusunda farklılıklar daha net bir şekilde gösterilir. a) Hareket tanıma için NATOPS <i>Full-Body</i> . b) Kişi kimliklendirme için NATOPS <i>Full-Body</i> . c) Hareket tanıma için NATOPS <i>Moving-Limbs</i> . d) Kişi kimliklendirme için NATOPS <i>Moving-Limbs</i>	43
Şekil 4.13	ResNet-18 ağına ait Validation-Loss grafikleri. Sırasıyla μ_2 , μ_3 , μ_4 ve μ_R fonksiyonlarının Validation-Loss değişiklikleri karşılaştırması. Son iterasyonlarda eğriler alt şekil içinde büyütülmüştür. Alt şekil kutusunda farklılıklar daha net bir şekilde gösterilir. a) Hareket Tanıma için SBU <i>Full-Body</i> . b) Hareket Tanıma için SBU <i>Moving-Limbs</i> . c) Hareket tanıma için BLD-SS <i>Kullanıcı-Tanımlı</i> . d) Kişi kimliklendirme için BodyLogin <i>Full-Body</i>	44
Şekil 5.1	İskelet görüntüsü	51
Şekil 5.2	Kinect kamera ile kaydedilen hareketlerin formatı	52
Şekil 5.3	Termal kameraya bakarken başını önce sağa tam sonra sola tam çevirme – 10 tekrar	53
Şekil 5.4	Kinect kamera ile kaydedilen hareketlerde ilk karelerde oluşan kayıp örnekleri	54
Şekil 5.5	Geç tanınan silüet kareleri örnekleri (Hareket kaybı olmadan)	54
Şekil 5.6	30 FPS hızına ulaşmadan başlanan daire hareketinde silüet tanınmanın gecikmesi	55
Şekil 5.7	Segmente edilen görüntülerdeki uzuvların ve gövdenin kameraya göre olan derinlik bilgisi	56
Şekil 5.8	Kinect Kamera ile kaydedilen videolardan elde edilen kareler. 30 fps altında silüet tanınmaması ve örnekler arası eşikleme tonlarında ki farklılıklar gösterilmektedir	57

Şekil 5.9	Derinlik bilgileri korunmuş olarak arka plandan ayrıştırılmış hareket frame örnekleri	57
Şekil 5.10	Mevcut fonksiyonlar ve korelasyon fonksiyonu ile üretilen hareket HE görüntülerinin görsel kıyaslaması	58
Şekil 5.11	Termal kamera kayıtlarından frame'lerin elde edilmesi ve arka plan görsellerinden ayrıştırılması	58
Şekil 5.12	Arka plandan temizlenmiş frontal termal yüz görselleri. Frontal termal yüz görselleri, yüzün karşıdan yaklaşık 30 derece sağa ve sola bakış açılarını içerecek şekilde oluşturulmuştur	59
Şekil 5.13	Termal left, frontal ve right side görselleri ve bu görsellerin gri seviyeye çevrildikten sonra birleştirilmesi ile oluşturulan Termal-3D yüz veri setine ait görsel örnekleri	60
Şekil 5.14	İskelet yapısı görselleri. Her bir görsel farklı kişiye aittir. Durağan haldeki iskelet yapısı görselleri alınarak oluşturulmuştur	60
Şekil 6.1	State-of-the-art derin ağlar için harcanacak eğitim zamanı ve doğruluk grafiği	62
Şekil 6.2	GoogleNet mimari yapısı	62
Şekil 6.3	ResNet-18 mimari yapısı	64
Şekil 6.4	MobileNet-V2 mimari yapısı	64

TABLO LİSTESİ

Tablo 3.1	Kullanılan veri setlerine ait EER sonuçları	16
Tablo 4.1	Korelasyon Değerleri için İlişki Durumları	25
Tablo 4.2	NATOPS veri seti kullanılarak hareket tanıma konusunda önerilen ve mevcut yaklaşımların performans sonuçları (% olarak sınıflandırma doğruluğu açısından). Bu tablo, daha önceki çalışmalarda kullanılan yöntem ve özelliklerle elde edilen doğruluk oranlarını göstermektedir. Mevcut yöntemler ve önerilen yaklaşım için farklı özellikler kullanılmasına rağmen tüm veriler NATOPS veri setinden elde edilmiştir	38
Tablo 4.3	SBU veri seti kullanılarak hareket tanıma konusunda önerilen ve mevcut yaklaşımların performans sonuçları (% olarak sınıflandırma doğruluğu açısından). Bu tabloda, önceki çalışmalarda kullanılan yaklaşımların ve özelliklerin performansı önerilen yöntem ile karşılaştırılmıştır. Mevcut yöntemler ve önerilen yaklaşım için farklı özellikler kullanılmasına rağmen tüm veriler SBU Kinect veri setinden elde edilmiştir	39
Tablo 4.4	CNN ağlarını kullanan NATOPS <i>Full-Body</i> veri setinde önerilen adaptif TT ve mevcut TT fonksiyonlarının başarı oranlarının karşılaştırılması. Her bir CNN ağı için hareket tanıma ve kişi kimliklendirme sonuçları tabloda ayrı ayrı sunulmaktadır. Kalın işaretli sonuçlar en iyi ortalama performansı gösterir	40
Tablo 4.5	CNN ağlarını kullanan NATOPS <i>Moving-Limbs</i> veri kümesinde önerilen adaptif TT ve mevcut TT fonksiyonlarının başarı oranlarının karşılaştırılması. Her bir CNN ağı için hareket tanıma ve kişi kimliklendirme sonuçları tabloda ayrı ayrı sunulmaktadır. Kalın işaretli sonuçlar en iyi ortalama performansı gösterir	40
Tablo 4.6	SBU Kinect veri kümelerinde mevcut TT fonksiyonlarının ve önerilen adaptif TT'nin, aynı ağlar üzerinden bireysel başarı oranlarının karşılaştırılması. Tablonun ilk satırı, referans çalışmadaki sonuçları içerir. Diğer satırlarda bu çalışma kapsamında yapılan karşılaştırma sonuçları yer almaktadır	41

Tablo 4.7	CNN ağlarını kullanan SBU Full-Body ve SBU <i>Moving-Limbs</i> veri kümelerinde adaptif TT ve mevcut TT fonksiyonlarının bireysel başarı oranlarının karşılaştırması. Her bir veri seti ve CNN ağı için hareket tanıma sonuçları sırasıyla tabloda sunulmaktadır. Koyu vurgulanmış sonuçlar en iyi ortalama başarıyı göstermektedir . . .	42
Tablo 4.8	CNN ağlarını kullanan BodyLogin veri kümelerinde adaptif TT ve mevcut TT fonksiyonlarının bireysel başarı oranlarının karşılaştırması. Hareket tanıma ve kişi kimliklendirme sonuçları, sırasıyla her bir CNN ağı için tabloda sunulmaktadır. Koyu vurgulanmış sonuçlar en iyi ortalama başarıyı göstermektedir . . .	42
Tablo 6.1	Veri seti eğitimlerinde kullanılan GoogleNet parametreleri	63
Tablo 6.2	ResNet-18 eğitimlerinde kullanılan parametreler	64
Tablo 6.3	GoogleNet ile yapılan eğitimlerin ortalama sonuçları. Farklı eğitim/validasyon oranları ve tüm veri setlerine ait ortalama sonuçlar gösterilmektedir	66
Tablo 6.4	Kişi tanımda hareketlerin ayrıntılı başarımları ve ortalama sonuçları	66
Tablo 6.5	ResNet-18 ile yapılan eğitimlerin ortalama sonuçları. Farklı eğitim/validasyon oranları ve tüm veri setlerine dair ortalama sonuçlar gösterilmektedir	68
Tablo 6.6	MobileNetV2 ile yapılan eğitimlerin ortalama sonuçları. Farklı eğitim/validasyon oranları ve tüm veri setlerine ait ortalama sonuçlar gösterilmektedir	69
Tablo 6.7	3 farklı yüz açısından sadece birinin olması durumu	70
Tablo 6.8	3 farklı yüz açısından herhangi 2 tanesinin olması durumu	70
Tablo 6.9	1. katmana ayna görüntüsü eklenmiş kayıplı durum kıyaslaması . .	72
Tablo 6.10	3. katmana ayna görüntüsü eklenmiş kayıplı durum kıyaslaması . .	72
Tablo 6.11	Termal veri setinin farklı içeriklere göre eğitilmesi ile elde edilen sonuçların kıyaslandığı tablo. Termal(RGB): sadece ön yüz açısı ile yapılan eğitim sonuçlarını, 3-Side Termal RGB: yüzün 3 farklı açısını içeren veri seti ile yapılan eğitim sonuçlarını, Termal-3D: yüzün 3 farklı açısının birer katmana konularak tek bir görüntü elde edilmesi ile yapılan eğitimlerin sonucunu göstermektedir	73
Tablo 6.12	İskelet verilerinin XY koordinatları olarak kullanılması sonucu elde edilen 1D-SkeletonNet eğitim sonuçları	73
Tablo 6.13	İskelet verilerinin koordinat olarak eğitiminin yapıldığı ağ yapısı . .	74
Tablo 6.14	Kişi tanıma için Termal-3D ve enerji görüntüsü veri setlerinin için skor level füzyon sonuçları	75

Tablo 6.15 Kişi Tanımada Termal-3D ve enerji görüntüsü füzyon sonuçları her bir hareket için ayrıntılı olarak verilmiştir (Tüm ağların sonuçları sırayla eklenmiştir)	76
Tablo 6.16 Kişi tanıma için 1D iskelet ve enerji görüntüsü veri setlerinin için skor level füzyon sonuçları	76
Tablo 6.17 Kişi Tanımada her bir hareket için 1D iskelet ve enerji görüntüsü füzyon sonuçları (Tüm ağlar sırayla verilmiştir)	77
Tablo 6.18 1D-iskelet veri seti ve Termal-3D veri setlerinin skor level füzyon başarımları sonuçları	77
Tablo 6.19 Eğitim ağları arasında hareket tanıma için füzyon başarımları sonuçları	78
Tablo A.1 NATOPS <i>Full – Body</i> veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları	81
Tablo A.2 NATOPS <i>Moving – Limbs</i> veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları	81
Tablo A.3 NATOPS <i>Full – Body</i> veri setinde kişi kimliklendirme için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları	82
Tablo A.4 NATOPS <i>Moving – Limbs</i> veri setinde kişi kimliklendirme için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları	82
Tablo A.5 SBU KINECT <i>Full – Body</i> veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları	82
Tablo A.6 SBU KINECT <i>Moving – Limbs</i> veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları	82
Tablo A.7 BodyLogin BLD-SS <i>User – Defined</i> veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları	83
Tablo A.8 BodyLogin <i>Full – Body</i> veri setinde kişi kimliklendirme için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları	83

Davranışsal ve Fiziksel Multi-biyometri ile Kişi Kimliklendirme ve Hareket Tanıma Üzerine Yaklaşımlar

Onur Can KURBAN

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Doktora Tezi

Danışman: Prof. Dr. Tülay YILDIRIM

Son yıllarda insan davranışının analizi en popüler araştırma alanlarından biri haline gelmiştir. Bu analiz, ölçülebilir davranışsal ve fiziksel hareket bilgilerini kullanarak insanların belirli bir zaman aralığındaki davranışlarını anlamayı amaçlar. Davranış analizi, gözlemin zamanına bağlı olarak hareketler, jestler, etkileşimler, etkinlik veya kişi tanıma ve doğrulama olarak sınıflandırılır. Bu analiz sadece bir kişinin davranışını değil, aynı zamanda kişilerarası ve kişiler-nesne arasındaki etkileşimi de araştırır. İncelenen davranışsal hareketler ayrıca üst veya alt beden, el veya kol hareketleri ve yüz ifadeleri gibi alt kategorilere ayrılabilir. Karmaşık insan aktivitelerinin tanınabilmesi, bazı önemli uygulamaların gerçekleştirilmesini de mümkün kılar. Anormal faaliyetleri tespit etmek için havaalanları ve metro istasyonları gibi halka açık yerlerde otomatik gözetim sistemleri; hastaların, yaşlıların ve çocukların gerçek zamanlı izlenmesi; jest tabanlı insan-bilgisayar arayüzleri bu uygulamalara örnek olarak verilebilir.

Bu çalışmanın ilk aşamasında, hareket boyunca meydana gelen değişiklikleri hesaba katan ve hareketteki değişimleri daha çok ortaya koyabilecek yeni bir enerji görüntüsü yöntemi üzerinde araştırma yapılmıştır. Buna ek olarak davranışsal biyometrinin çoklu biyometriye sağladığı performansı etkileri incelenmiştir. Bu araştırmalar sonucunda RGB-D verilerinden elde edilen maskelenmiş derinlikli video akışlarındaki hareket dizisi bilgisine dayalı olarak eylem tanıma ve kişi tanımlama için yeni bir enerji görüntüsü oluşturma yaklaşımı önerilmiştir. Önerilen yöntemde, bir hareket dizisi sırasında meydana gelen değişiklikleri belirlemek için ilk kare referans alınmakta ve tüm kareler arasındaki korelasyon katsayıları hesaplanmaktadır. Elde edilen

katsayılar, harekete uyumlu bir zamansal şablon oluşturmak için kullanılır. Adaptif zamansal şablon olarak adlandırılan bu fonksiyon tarafından oluşturulan enerji görüntüleri, literatürde yer alan ve sabit olarak belirli aralıkları vurgulayan mevcut zamansal şablonlara kıyasla hareketleri daha iyi bir temsil edebilmektedir. Bu şekilde, eylemleri daha iyi temsil edebilen yeni bir enerji görüntüsü ile birden fazla TT'nin füzyon ihtiyacı ortadan kaldırabilen ve işlem maliyetini azaltan pratik bir yöntem önerilmiştir. RGB-D tabanlı çalışmalar literatürde daha çok hareket tanıma için kullanılmaktadır. Bu çalışmada yer alan bir diğer katkı da, mevcut ve önerilen yöntemlerin kişi kimliklendirme içinde incelenmiş olmasıdır.

Tez kapsamında önerilen fonksiyonla oluşturulan enerji görüntüleri, hareket süresince değişim olan aralıkları vurgularken, değişim olmadığı veya çok az olduğu aralıkları bastırır. Önerilen fonksiyon kullanılarak elde edilen enerji görüntüleri, evrişimli sinir ağlarına ve farklı hand-crafted sınıflandırıcılara girdi olarak verilmiştir. Önerilen yöntem BodyLogin, NATOPS ve SBU Kinect veri setleri üzerinde gözlemlenmiş ve mevcut yöntemlerle karşılaştırılmıştır. Bunun dışında davranışsal biyometrinin çoklu-biyometrik sistemlere katkısını incelemek için davranışsal, termal ve fiziksel verilerden oluşan yeni bir veri seti oluşturulmuş ve önerilen yöntem bu yeni veri seti üzerinde test edilmiştir.

Elde edilen sonuçlar, önerilen adaptif zamansal şablon yönteminin, literatürde mevcut olan şablonlara kıyasla daha yüksek performans ve daha kısa işlem süresi sağladığını göstermektedir. Kişi tanımlama sonuçları, RGB-D sensörleri ile elde edilen görüntülerin kişi tanımlaması için etkin bir şekilde kullanılabileceğini göstermiştir.

Anahtar Kelimeler: Davranışsal biyometri, enerji görüntüsü, termal yüz, derin öğrenme

Approaches to Person Identification and Motion Recognition with Behavioral and Physical Multi-Biometrics

Onur Can KURBAN

Department of Electronics and Communication Engineering

Doctor of Philosophy Thesis

Supervisor: Prof. Dr. Tülay YILDIRIM

In recent years, the analysis of human behaviour has become one of the most popular research areas. This analysis aims to understand people's behaviour over a period of time using measurable behavioural and physical movement information. Behaviour analysis is classified as motions, gestures, interactions, activity or person recognition and verification, depending on the time of observation. This analysis explores not only a person's behaviour but also interpersonal and person-object interactions. The behavioural movements studied can be subdivided into subcategories such as upper or lower body, hand or arm movements, and facial expressions. The ability to recognize complex human activities also makes it possible to perform some important applications. Automatic surveillance systems in public places such as airports and subway stations to detect abnormal activity; real-time monitoring of patients, the elderly and children; gesture-based human-computer interfaces are considered examples of these applications.

In the first phase of this study, a new energy imaging method that takes into account the changes occurring during the motion and can emphasize the changes in the motion has been researched. In addition, the effects of behavioral biometrics on multi-biometrics performance were examined. As a result of these studies, a new energy image generation approach has been proposed for action recognition and person identification based on the motion sequence information in masked depth video streams obtained from RGB-D data. In the proposed method, the first frame is referenced to determine the changes that occur during the motion sequence and

the correlation coefficients between frames are calculated. The obtained coefficients are used to create a temporal template compatible with the motion. Energy images generated by this function, called adaptive temporal template, can better represent movements compared to existing temporal templates in the literature that consistently emphasize certain intervals. In this way, a new energy image that can better represent the actions and a practical method that can eliminate the need for fusion of multiple TTs and reduce the cost of operation is proposed. RGB-D-based studies are mostly used for motion recognition in the literature. Another contribution in this study is the examination of existing and proposed methods for person identification.

The energy images generated with the function proposed in the thesis emphasize the intervals that change during the motion while suppressing the intervals where there is no or little change. The energy images obtained using the proposed function are given as input to convolutional neural networks and different state-of-the-art classifiers. The proposed method was observed on BodyLogin, NATOPS and SBU Kinect datasets and compared with existing methods. Apart from this, a new dataset consisting of behavioural, thermal and physical data was created to examine the contribution of behavioural biometrics to multi-biometric systems and the proposed method was tested on this new dataset.

The results show that the proposed adaptive temporal template method provides higher performance and shorter processing time compared to the templates available in the literature. Person identification results showed that images obtained with RGB-D sensors can be used effectively for person identification.

Keywords: Behavioral biometrics, energy images, thermal face, deep learning

1.1 Literatür Özeti

Biyometri son yıllarda önem verilen ve geniş çapta çalışmaların yapıldığı bir konudur. Biyometri, kişilere ait fiziksel ve davranışsal farklılıkların, onları tanımak ve ayırt edebilmek için kullanılması olarak tanımlanabilir. Anahtar, şifre, akıllı kart sistemlerindeki güvenlik açığı, kaybolma ve unutma risklerine karşılık biyometrik veriler daha rahat ve güvenilir bir kullanım alanı sağlamaktadır. Günümüzde, bilgisayarla görme sistemlerinde, giriş-çıkış sistemlerinde, güvenlik takibi gibi birçok alanda biyometrik bilgiler kullanılmaktadır.

Biyometrik alanda yapılan çalışmaların genişlemesi ve ilgi görmesi, çeşitli biyometrik verilerin kullanılmasını arttırmıştır. Biyometri denildiğinde, akla yüz, parmak izi, iris ve el damar verileri gelmekteydi. Ancak bu alandaki çalışmaların artması ve çoklu biyometrik verilerin daha güvenilir ve doğru sonuçlar vermesi son yıllarda yürüyüş analizi, boy uzunluğu, eklem uzunlukları, kıyafet, ten rengi, ırk ve cinsiyet gibi bilgilerinde eklenmesini sağlamıştır. Biyometrik verilerdeki bu çeşitliliğin artması verilerin alt başlıklar altında toplanmasını gerekli kılmıştır.[1, 2]

Biyometrik veri türlerini alt başlıklar altında inceleyecek olursak;

I.Hard Biyometri: İris, parmak izi, parmak damarı, avuç içi damarı, yüz gibi biyometrik veriler hard biyometri olarak sınıflandırılmaktadır. Hard biyometri verileri her insanda kendine özgü, güvenilirliği ve ayırt edilebilirliği yüksek, değiştirilebilirliği zor olan veri grubunu oluşturmaktadır. Hard-biyometri verileri için karmaşık sensör yapılarına ve yüksek çözünürlüklü görüntüleme sistemlerine ihtiyaç duyulmaktadır. Buna ek olarak hard-biyometrik verilerin alınması için işbirlikçi çalışma sistemi oluşturulmalıdır. Yani biyometrik veri alınacak kişi ile işbirliği içinde olunmalı ve çok yakın bir mesafede temasa geçilmelidir [3].

II.Behavioral (Davranışsal) Biyometri: Kişilerde kendilerine has yapılar göstermekle birlikte, istenildiğinde değiştirilme veya taklit edilebilme özelliği olan biyometrik veri

grubudur. Davranışsal biyometriye örnek olarak yürüyüş, klavye ve mouse kullanımı, konuşma, imza, bir şey yiyip içmek gibi hareketler gösterilebilir. Davranışsal biyometrik veriler, hard biyometrik verilere oranla daha az ayırt edilebilen ve daha değişken olabilen verilerdir. Ancak elde etmek için kişi ile işbirliğinin zorunlu olmaması ve karmaşık sensörlere ihtiyaç duymaması, bu verileri kullanım alanlarına göre avantajlı kılmaktadır [4].

III.Soft Biyometri: Elbise, boy uzunluğu, eklemler arası uzunluklar, cinsiyet, saç rengi ve uzunluğu, üzerinde bulunan eşya, gibi tanımlama ve kimliklendirme işlemlerinde kullanılabilen özelliklerdir. Soft biyometri verileri, kişisel özgünlüğü ve ayırt etme güvenilirliği az olan verilerdir. Bu yüzden tek başına kullanılmaları güvenilir sonuçlar vermemektedir. Soft biyometri verileri daha çok diğer biyometri verilerinden elde edilen sonucun başarısını arttırmaya yönelik yardımcı bilgiler olarak değerlendirilebilir. Soft biyometrinin en belirgin özelliği işbirlikçi çalışma sistemine ve karmaşık sensörlere ihtiyaç duymamasıdır [5].

Bu veri türleri içerisinde hard-biyometri alanında çoklu biyometri ile oldukça yüksek başarımlı sonuçlarına ulaşılmıştır. Özellikle parmak izi-parmak damarı, yüz-iris, parmak izi-avuç içi damarı, parmak izi-yüz gibi değişik verileri birleştirilerek yüksek başarımlı sonuçlar elde edilmiştir. Ancak bu sistemlerin yüksek kimliklendirme başarımlarının yanı sıra kullanım alanlarının belirli bir sınır içerisinde olması ve kullanıcı ile temas ihtiyacı duyması bu sistemlerin dezavantajı olmaktadır [6].

İnsanlar ise doğal hayatları içerisinde, karşısındaki kişileri tanımak ve ayırt edebilmek için davranışsal ve fiziksel farklılıkları kullanabilmektedir. Bu işlem genel olarak kişinin zihninde daha önceden görmüş olduğu bir davranış veya fiziksel özelliği tanımlaması ile mümkün olmaktadır. İnsanlar görmüş oldukları davranışsal ve fiziksel biyometri verilerini, zihinlerinde hedef kişi için kaydetmiş olurlar. Daha sonra farklı bir ortamda o kişiyi gördüğünde, zihninde kayıtlı olan biyometrik veriler ile kişileri tanıyabilirler. Tanınan kişinin davranışında veya fiziksel yapısındaki değişiklikler hafızada güncellenerek tutulabilir. Üstelik bunun için karşısındaki kişinin yüzünü görmesi de gerekmez. Bu yöntem ile kimliklendirme ve tanıma, özellikle kişi ile yakın temasın veya işbirliğinin sağlanamadığı havaalanı gibi güvenlik düzeyi yüksek ortamlarda, hard biyometrik verilerin elde edilemediği adli vakalarda ve şüpheli tanıma işlemlerinde, bir bölgeden geçen veya belirli bir bölge içerisinde aranan bir kişinin takip edilebilmesi gibi işlemlerde kullanılabilmektedir [7]. Özellikle son 2 yıldır Covid-19 ile temassız sistemlere ilgi çok daha artmıştır [8, 9]. Covid-19 kısıtlamaları göz önüne alındığında parmak izi, parmak damar izi, avuç içi damar tarama, yüz tarama, iris tanıma gibi yakın temas içeren veya maske engeli olabilecek biyometri türlerinin kullanımı, sağlık sebebiyle durdurulmuş veya yerini temassız ve

mesafe kurallarına uygun olan davranışsal ve fiziksel biyometri çeşitlerine bırakmıştır [10–12].

Bilgisayar destekli davranış analizi, ölçülebilir davranışsal ve fiziksel hareket bilgilerini kullanarak insanların belirli bir zaman aralığındaki davranışlarını anlamayı amaçlar. Davranış analizi, gözlemin zamanına bağlı olarak hareketler, jestler, eylemler, etkileşimler, etkinlik veya kişi tanıma ve doğrulama olarak sınıflandırılır. Bu analiz sadece bir kişinin davranışını değil, aynı zamanda kişilerarası etkileşimleri ve kişi-nesne arasındaki etkileşimi de araştırır. Analiz edilen davranışsal hareketler, üst-alt vücut, el-kol hareketleri ve yüz ifadeleri gibi alt kategorilere ayrılabilir. Karmaşık insan faaliyetlerinin tanınması, birkaç önemli uygulamanın inşasını da mümkün kılar. Anormal faaliyetleri tespit etmek için havaalanları ve metro istasyonları gibi halka açık yerlerde otomatik gözetim sistemleri; hastaların, yaşlıların ve çocukların gerçek zamanlı izlenmesi; jest tabanlı insan-bilgisayar ara yüzleri bu uygulamalara örnek olarak verilebilir [13]. Bu alanda yapılan çalışmalar incelendiğinde, [14]'de Bobick ve Davis, bir görüntü dizisinde hareketin nerede meydana geldiğini gösteren binary bir hareket-enerji görüntüsü (MEI) yapısını sunmuştur. Bu sayede, hareketin etkileşim alanı görülebilmektedir. Bu yöntem, hareketi tanımak için yaygın olarak kullanılmaktadır. [15]'te, yürüme enerjisi görüntüsü (GEI) olarak adlandırılan uzaysal-zamansal yürüyüş temsili için MEI yöntemi kullanılmıştır. İnsan yürüme özellikleri, yürüyüş salınımlarına göre kişi tanıma için karakterize edilmiştir. Ancak, bu gösterim ayrıntılı bilgileri kaybeder ve zamansal değişiklikleri içermez. Daha sonra, [16]'te Davis, hareket yoğunluğuna bağlı bir fonksiyon ile, hareket geçmişi görüntüsü (MHI) adı verilen skaler değerli bir görüntü üretti. Buna ek olarak yürüyüş geçmişi görüntüsü (GHI), önceki çalışmalarda yer alan GEI ve MHI'ye dayalı olarak geliştirilmiştir [17]. Bu yöntem, GEI'deki zamansal değişiklik eksikliğini ortadan kaldırır. Ancak bu yöntemde yetersiz eğitim döngüleri sorunlara yol açmaktadır. [18]'de GEI ile bir yürüyüş anı görüntüsü (GMI) oluşturulmuştur. GMI, tüm yürüyüş döngülerinin her önemli anındaki yürüme olasılığı görüntüsüdür. Tüm yürüyüş döngülerinin önemli anlarının ortalaması, GEI'de değerlendirildiği gibi alınır. Ancak bu önemli anları farklı periyotlara sahip döngülerden seçmek GMI için kolay değildir. [19]'de Chen ve ark. hareketi kümelerle ayırmış, baskın enerji görüntüleri (DEI) elde etmiştir. GEI ve DEI tarafından elde edilen görüntüler arasındaki farkı alarak, çerçeve farkı-enerji görüntüsü (FDEI) yöntemini önermiştir. Yakın tarihli bir çalışmada, [20]'de eylem tanıma için Poonkodi ve Vadivu tarafından Zamansal Fark Çerçevesinin Korelasyonu (CTDF) yöntemi önerilmiştir. Bu yöntem, birbirini izleyen üç silüet çerçevesi arasındaki korelasyona göre yerel olarak ilgi noktaları üretir. Mekânsal-zamansal komşuluklar, Harris matematiksel operatörü kullanılarak elde edilir. [21]'de, korelasyon katsayısına dayalı çerçeve farkı tekniği

kullanılarak arka plan çıkarma işlemi yapılmış ve siluet görüntüleri elde edilmiştir. Bu görüntülerden öznitelikler çıkarmak için Öklid mesafe dönüşümü ve Shannon entropisi kullanılır. Bu özelliklerin kombinasyonları, çeşitli insan eylemlerini tanımak için ileri beslemeli bir sinir ağına girdi olarak verilir. [22]'de ise insan hareketini kısa videolar aracılığıyla tanımak için 3B CNN modeli kullanılmıştır. Önerilen mimari, faaliyetlerin mekânsal konumu ile zaman-hareket özelliklerini dikkate alarak eylem sınıflandırması sağlar. CNN ile uzamsal özellikler çıkarılırken, konvolüsyonel Long-Short Time Memory katmanı sayesinde çerçeveler arasında zamansal bir ilişki kurulur. Ve son olarak, [23]'de Chen ve ark. insan iskelet eklemlerinin mekânsal ilişkilerini ve zaman içinde uzun vadeli zamansal bilgileri sağlayan iskelet tabanlı bir insan eylemi tanıma yöntemi sunmuştur. Bu yaklaşımda, iskelet dizileri renkli hareketli görüntülere dönüştürülür ve eylem sınıflandırması için sık bir CNN modeline girdi olarak uygulanır. Özetle, enerji görüntüleri son yıllarda popüler bir çalışma konusu haline gelmiştir ve bu alanda yeni yöntemlerin geliştirilmesi devam etmektedir.

Hareket tanıma ile ilgili daha önce yapılan çalışmalar, özellikle yukarıda bahsedilen son makaleler, yeni araştırmacılar için fikir oluşturmakta ve araştırmacıların bu fikirleri geliştirip daha verimli bir yapıya taşıdığını bizlere göstermektedir. Bu gelişmelere ek olarak, Ijjina ve ark. geleneksel MHI ve MEI yöntemlerinden farklı olarak, RGB-D videodaki hareket dizisi bilgisine dayalı eylemleri tanımak için hareketin çeşitli zamansal bölgelerini (başlangıç, orta ve bitiş) vurgulayan üç özel zamansal şablon (TT) önermiştir [24].

Enerji görüntülerini kullanan başka bir çalışmada, Abdelbaky ve Aly, temel bileşen analiz açısından (PCANet) öğrenilen uzamsal ve zamansal özellikleri, bag of features (BoF) ve yerel olarak kümelenmiş tanımlayıcıların vektörü (VLAD) kodlama şemaları ile birleştirerek bir insan eylemi tanıma yöntemi önermektedir [25]. Bu çalışmada, whitening transformation (WT) kullanılarak her bir video boyutu azaltılmış ve PCA-Net aracılığıyla hem uzamsal hem de zamansal özellikler öğrenilmiştir. Kısa zamanlı hareket enerjisi görüntüleri (ST-MEI), zaman akışını hesaplamak için kullanılmıştır. Mishra et al. [26]'deki çalışmalarında üyelik olarak statik 5 parçalı bir fonksiyon kullanmışlardır. Burada hareket akışını iyi modellemek için hareket vurgusunun azaldığı yerleri düşük katsayılarla çarparken farklı kareleri yüksek değerlerle çarpmayı amaçlamışlardır. Ancak Ijjina'nın çalışmasında olduğu gibi geçici kalıpların sayısındaki artış ve Mishra'nın çalışmasında olduğu gibi özel aralıklar için sabit kalıpların kullanılması bazı sorunları da beraberinde getirmektedir. Her şablon, farklı hareket türleri için farklı sonuçlar üretecek olan farklı eylem alanlarına odaklanır. Sonuç olarak, bu yöntemler vurgulanan alanlarla örtüşen eylemler için iyi sonuçlar verebilirken, vurgulanan alan dışındaki eylemler için temsil kabiliyetini azaltacaktır. Eylem alanı herhangi bir TT tarafından vurgulanan alandan

daha genişse, eylemi daha iyi temsil etmek için birden fazla TT'nin birleştirilmesi gerekir. Bu zorlukların üstesinden gelmek için, hareket boyunca meydana gelen değişikliklere göre oluşturulan, harekete uyum sağlayan yeni bir zamansal şablon temsili öneriyoruz. Önerilen yöntemde, bir hareket dizisi sırasında meydana gelen değişiklikleri belirlemek için ilk kare referans alınmakta ve tüm kareler arasındaki korelasyon katsayıları hesaplanmaktadır. Elde edilen katsayılar, harekete uyumlu adaptif bir TT oluşturmak için kullanılır. Adaptif TT fonksiyonu tarafından oluşturulan enerji görüntüleri, belirli dönemleri vurgulayan mevcut TT'ler tarafından oluşturulan enerji görüntülerinden daha iyi hareket temsiline sahip olacaktır. Bu şekilde, enerji görüntülerinde eylemlerin daha iyi temsil edilmesini sağlayan ve birden fazla TT'nin füzyon ihtiyacını ortadan kaldırdığı için işlem maliyetini de azaltan pratik bir yöntem önerdik.

Ayrıca RGB-D tabanlı çalışmalar literatürde daha çok hareket tanıma için kullanılmaktadır [27]. Bu çalışmadaki diğer bir katkı ise, mevcut ve önerilen yöntemlerin kişi tespiti için de kullanılmış olmasıdır.

Buna ek olarak davranışsal ve fiziksel biyometri çalışmalarında kullanılması amacıyla silüet, termal yüz ve eklemler arası uzaklığa bağlı iskelet verileri içeren yeni bir biyometrik veri tabanı oluşturulmuş ve bu çalışma içerisinde test edilmiştir.

1.2 Tezin Amacı

Biyometri konusu son yıllarda oldukça önem verilen ve geniş çapta çalışmaların yapıldığı bir konudur. Biyometri, kişilere özel olan fiziksel ve davranışsal farklılıkları onları tanımak ve kimliklendirmek için kullanabilmektedir. Anahtar, şifre, akıllı kart sistemlerindeki güvenlik açığı, kaybolma ve unutma risklerine karşılık biyometrik veriler daha rahat ve güvenilir bir kullanım alanı sağlamaktadır. Günümüzde, bilgisayarla görme sistemlerinde, giriş çıkış sistemlerinde, güvenlik takibi ve tanıma gibi birçok alanda biyometrik bilgiler kullanılmaktadır.

Biyometrik alanda yapılan çalışmaların genişlemesi ve ilgi görmesi, ayrıt edilebilir biyometrik verilerin kullanılmasını arttırırken, verilerin türlerini de kendi içerisinde ayırma gereksinimi duymaktadır. Daha önceden biyometri denildiğinde geleneksel olarak akla gelen yüz, parmak izi, iris ve el damar verileri gelmekteydi. Ancak bu alandaki çalışmaların artması ve çoklu biyometrik verilerin daha güvenilir ve doğru sonuçlar vermesi günümüzde yürüyüş analizi, boy uzunluğu eklem uzunlukları, kıyafet, ten rengi, ırk ve cinsiyet gibi bilgilerinde eklenmesini sağlamıştır.

Bu çalışmada yapılması amaçlanan işlem, soft biyometri ve davranışsal biyometri ile

kiři tanıma ve kimliklendirmeye yönelik daha az kamera kaydı ve kinect kamera ile daha yüksek başarımlı sağlayacak bir yöntemin geliştirilmesi olacaktır. Yapılan çalışmalar incelendiğinde çok sayıda kamera kullanılmış olması hem işlem yoğunluğu hem de malzeme maliyetini arttırmaktadır. Bu çalışmada optimum kamera ile soft ve behavioral biyometriye ek olarak hard biyometri verisi olan yüz tanıma sistemi de eklenerek hem kimliklendirme performansının artışı hem de malzeme maliyetlerinin düşürülmesi amaçlanmıştır.

1.3 Hipotez

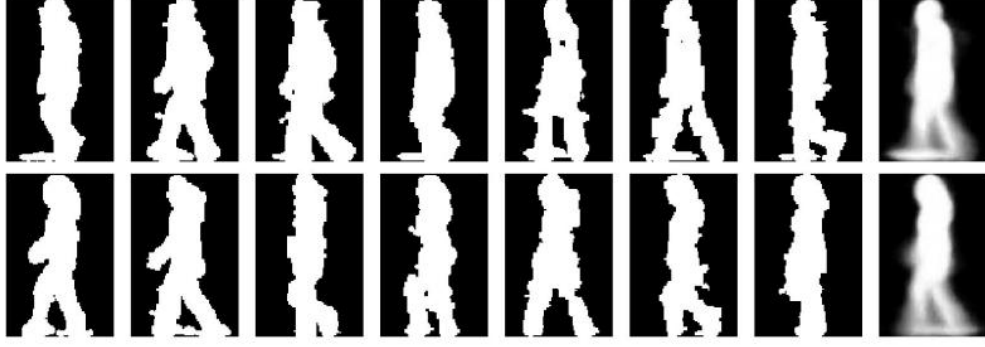
Davranışsal biyometrik verilerden olan enerji görüntüsünden hareket ve kiři tanıma üzerinde başarımlı ve hareketi temsil kabiliyetini arttıran yeni yöntemlere ihtiyaç vardır. Bu doğrultuda hem hareket değişimlerini daha iyi temsil edebilecek hemde çoklu biyometride başarımlı arttıracak bir adaptif fonksiyon oluşturulmuştur. Bu fonksiyon, farklı veri setlerine uyum sağlarken, farklı eğitim oranlarında daha yüksek başarımlar elde edebilmektedir. Buna ek olarak işlem maliyetinide azaltmaktadır. Bu yönüyle, mevcut yöntemlerden kıyasla daha kullanışlıdır.

2 BİYOMETRİK VERİLER

2.1 Silüet Verileri ve Enerji Görüntüsü

Silüet verileri, davranışsal bir biyometrik özellik olarak kullanılabilir. Ancak bu biyometrik verileri parmak izi, göz irisi gibi ömür boyu korunan veya değişmez veri kaynakları değildir. Kişinin günlük giyim tarzına, yaşına ve duruş şekline göre değişkenlik gösterebilir. Bu sebepten dolayı, silüet verilerinin tek başına bir biyometrik veri olarak kullanılmasından ziyade, belirli bir zaman aralığındaki silüetlerden elde edilen enerji görüntüleri, biyometrik veri olarak tercih edilmektedir. Çünkü kişilerin silüetleri anlık olarak değişkenlik gösterse bile uzun süreli davranışları daha kişiseldir ve daha uzun süre korunabilmektedir. Daha iyi bir ifade ile yürüyüş salınımı (gait) bir kişiyi tanımak için kullanılan bir biyometri çeşididir. Yürüyüş biyometrisinin avantajı parmak izi, yüz, iris gibi biyometrik verilerinin aksine, kişi ile işbirliği kurulmaksızın alınabilir olmasıdır. Yürüyüş biyometrisi kişiye has özellik ve bilgiler sağlamakta ancak bir davranışsal biyometri olduğu içinde koşullara göre değişim gösterebilmektedir. Örneğin hastalık, yaralanma, kıyafet, ayakkabı, yer ve çevresel koşullara göre değişimler göstererek kişinin yürüyüş silüetinde farklılıklar olabilmektedir. Yürüyüş biyometrisi bu zorluklar sebebiyle her ne kadar bir parmak izi veya iris biyometrisi gibi özel olmasa da, görsel gözlemde kişilere ait yeri doldurulamayacak karakteristik bilgiler taşımaktadır.

Yukarıda bahsedildiği gibi davranışsal biyometriklerden bilgi çıkartmada kullanılan bir yöntemde silüet verilerinden elde edilen enerji görüntüsüdür. Örneğin, düzenli bir insan yürüyüşünde, insan hareketinin kararlı bir frekansta tekrarlandığı ve çevrimsel hareket oluşturduğu bilinmektedir. Genel bir yürüme çevriminde poz sırası aynı şekilde ilerlemektedir. Yani kollar ve bacaklar benzer bir şekilde ileriye ve geriye gitmektedir. Ancak yürüyüş çevriminin fazı, gövde şekli, uzuvların uzunlukları ve salınım fazı farklılık göstermektedir. Bu varsayımlar altında, uzay-zamansal bilgiyi, sıralı bir görüntü sırası yerine tek bir 2B yürüyüş görüntüsünde temsil etmek Şekil-2.1 görüldüğü gibi mümkündür.



Şekil 2.1 Bir yürüyüş çevrimine ait siluet görüntüleri ve bu siluetler ile üretilen yürüyüş enerji görüntüsü [28]

Yürüyüş enerji görüntüsü için ilk olarak insanların normal yürüyüş görüntüsü üzerinden siluetler çıkartılmaktadır. Toplanan siluetler üzerinde ilk olarak bazı ön işlemler yapılır. Bunlar; boyut normalleştirme (her siluet resmini orantılı olarak yeniden boyutlandırarak tüm siluetlerin aynı yüksekliğe sahip olması sağlanır) ve yatay hizalama (üst yarım siluet parçasını yatay ağırlık merkezine göre merkezileştirme) işlemleridir. Bu sayede eşit boyutlarda anlık yürüyüş görüntüleri elde edilmektedir. Bütün görüntülerin toplanıp toplam sayıya bölünmesi ile yürüme frekansı ve fazı belirlenirken, zaman serisi sinyalinden ise maksimum entropi spektrum kestirimi yapılabilir. Binary görüntülerin her biri t zamanına bağlı iki boyutlu bir değişken $B_t(x, y)$ olarak tanımlanırsa, yürüyüş enerji görüntüsü aşağıdaki gibi tanımlanabilir.

$$G(x, y) = \left(\frac{1}{N}\right) \sum_{t=1}^n B_t(x, y) \quad (2.1)$$

Burada N yürüyüş çevrimine ait görüntü sayısını, t o andaki görüntü sırasını, x ve y ise iki boyutlu görüntüye ait koordinat bilgilerini belirtmektedir. Daha öncede belirtildiği gibi yürüyüş enerji görüntüsü, kişiye ait majör siluet görüntüsünü ve bir yürüyüş çevrimi içerisindeki değişimleri içermektedir. Buna yürüyüş enerji görüntüsü denmesinin sebebi ise;

1. Her siluet görüntüsü o andaki insanın yürüme alanıyla normalize edilmiş enerji görüntüsüdür.
2. Son görüntü kişinin bir yürüyüş çevrimine ait zamana göre normalize edilmiş toplam enerjisini ifade eder.
3. Son görüntüde yüksek yoğunluklu her bir piksel, bu noktada insan hareketinin daha

sık meydana geldiği anlamına gelir (yani daha yüksek enerji) [28].

Bu özellikler çerçevesinde siluetlerden elde edilen enerji görüntüsünün hem hareket tanıma hem de kişi tanıma için kullanılabilecek bir biyometrik şablon olduğu görülmektedir.

2.2 Termal Görüntüler

Yüz tanıma özellikle son 20 yıldır, görüntü işleme ve örüntü tanıma uygulamalarında en çok çalışılan alanlardan biri olmuştur. Görülebilir spektrum içerisinde yapılan bu çalışmalar sonucunda oldukça yüksek başarıyla uygulamalara ulaşılmıştır. Ancak, gerçek dünya koşullarında görülebilir spektrum içerisinde yapılan uygulamalarda bazı zorluklarla da karşılaşmıştır. Özellikle ışığın geliş açısına göre oluşan gölgelenme ve parlamalar, yüz ifadeleri ve yüz maskesi gibi aldatıcı zorluklar görülebilir spektrumdaki uygulamalarda olumsuz etkiler oluşturmaktadır. Bu tür durumların üstesinden gelmek için daha karmaşık çözüm yöntemlerinin ve yeni metodların geliştirilmesi üzerine çalışılmaktadır. Hatta 3D tarayıcılar ile yüz modelleri çıkartılarak yüz tanıma uygulamaları yapılmaktadır. Ancak bu gibi uygulamalar sistem maliyeti ve işlem yükünü de arttırmaktadır.

Bu alanda bir diğer seçenek ise termal görüntülerdir. Kızılötesi spektrumdan elde edilen termal görüntüler, görünür spektrumdaki aydınlatma ile meydana gelen değişimlere büyük ölçüde tepkisizdir. Bu özelliğiyle dikkat çeken bir yöntem olmuştur. Yüz tanıma bağlamında, kızılötesi kameralar kullanılarak elde edilen veriler, görünür spektrumda çalışan daha yaygın kameralardan elde edilen verilere kıyasla belirgin avantajlara sahiptir. Örneğin, kızıl ötesi spektrum ile elde edilen termal yüz görüntüleri her hangi bir aydınlatma olmaksızın veya değişik aydınlatmalı ortamlarda dahi değişimden etkilenmemektedir. Ve yine termal kızılötesi spektrumunda alınan görüntüler, duman ya da toz dağılımından ve yansıyan görünür ışıktan daha az etkilenmektedir. Görülebilir spektrumun aksine, kızılötesi spektrum görüntüleme sadece dış görünüşü değil, aynı zamanda cilt altına ait anatomik bir biyometri bilgisini de (örneğin, bir damar ağı) elde etmek için kullanılabilir. Bunlara ek olarak, termal görüntüler, görünür spektruma kıyasla yüz maskelerini tespit etmek için de kullanılabilir. Şekil 2.2’de termal yüz görüntüleri örnekleri sunulmuştur.

Termal görüntülerin avantajları gibi bazı dezavantajları da bulunmaktadır. Bunlar genel olarak kişinin metabolizmasına, hastalığına, egzersiz yapmasına, gözlük kullanması, hava sıcaklığına göre bazı değişiklikler göstermektedir. Ayrıca kıyafet, kızarıklık, baş ağrısı veya diş enfeksiyonu gibi sıkıntılarda lokal olarak görüntüde farklılıklar oluşturabilmektedir. Ancak çoklu biyometri kullanımı, bu tür



Şekil 2.2 Termal yüz görüntü örnekleri

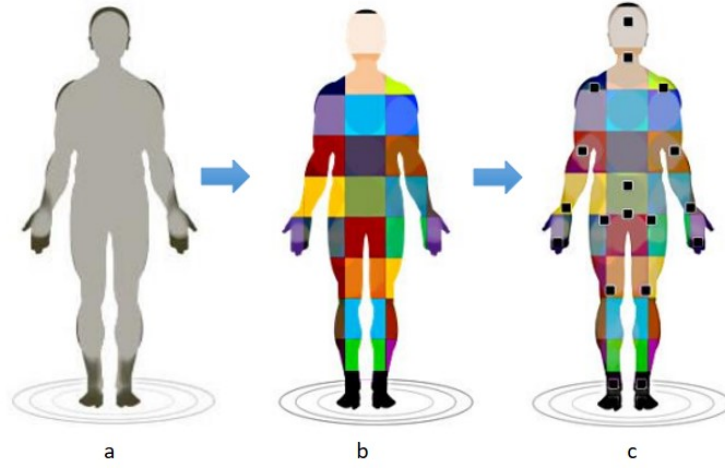
olumsuzlukların giderilmesinde veya etkisinin azaltılmasında başarılı olmaktadır [29].

2.3 Eklem arası uzaklık (İskelet) Verileri

Kinect kameralar, görüş alanındaki ortamın ham derinlik görüntülerini verebilmektedir. Elde edilen derinlik verilerinden, pozun şekline göre kişinin iskeleti çıkartılabilir ve ilerleyen derinlik kareleri boyunca bu iskeletin takibi sağlanabilir. Eklem bütünü vücut hareketi takip edilerek, oluşan yeni vücut pozuna göre konumlandırılır. Her bir eklem vücutun belli bir bölgesini temsil eder. Bu sayede kişi uzuvlarının Kinect uzayında nerede olduğu göreceli olarak anlaşılabilir. İnsan vücudu boy ve cüsselerindeki çeşitlilik, giyilen kıyafetlerdeki değişikliğe göre esnek olup farklı şekildeki duruşlara müsaittir. Ham derinlik görüntülerinden insan pozisyonları çıkarımı model tabanlı bir çıkarım kullanılmaktadır. Shotton ve ekibi [30] tarafından geliştirilen bu yöntem Kinect SDK ile kullanıcılara 20 eklemli bir iskelet yapısını hazır olarak sunmaktadır. Farklı boylar, cüsseler, kıyafetler, saç tipleri vb. etmenler hesaba katılarak çeşitli görünümüne göre derinlik verisindeki silüet parçalarının vücutun hangi bölgesine denk geldiği tayin edilmektedir.

Kinect, birbirinden farklı karakter modelleri olan karar ağaçlarını kullanarak, vücut kısımlarını birbirinden ayıracak şekilde etiketlenmiş, belirli vücut parçasına ait olabilecek alanları belirler. Bu şekilde ham derinlik verilerindeki her pikselin karar ağacındaki bir düğümlerle eşleşmesi sağlanır ve vücut parçaları oluşturularak vücut uzuvları belirlenir. Belirlenen bu uzuvlara bir eklem noktası atanır. Takip eden derinlik kareleri boyunca bu eklem noktaları güncellenerek eklem ve vücut takibi gerçekleştirilir. Bu durum Şekil 2.3'teki gibidir.

Kinect, kişiye ait 20 eklem noktasından hiyerarşik olarak iskelet görüntüsü elde eder. Her bir eklem ön tanımlı olarak bir vücut bölgesine aittir. Bu iskelet modelleri, baş, boyun, omurga, merkez kalçanın yanı sıra sol ve sağ yan eklem: el, bilek, dirsek, omuz, kalça, diz, ayak bileği ve ayak gibi aşağıdaki merkez-vücut bağlantılarını izler. Bu bağlantılar Şekil 2.4'te gösterilmektedir. Kullanıcı görünür alanından çıkmadığı veya eğilme, dönme gibi hareketlerle duruş bozulmadığı sürece, eklem noktaları vücutta aynı konumu göstermeye devam eder. Elde edilen bu iskelet yapısı kişiye ait



Şekil 2.3 Kinect eklem noktalarının model tabanlı çıkarımla belirlenmesi. a) Ham derinlik verisi, b) çıkarımı yapılan vücut parçaları, c) çıkarımı yapılan eklem noktaları [31]

fiziksel bir biyometri verisi olarak kullanılabilir. İskelet biyometrisi hem kişiye ait bir yapı içermekte hem de dış görünüşten bağımsız olarak, temassız bir biyometri olarak kullanılabilir.



Şekil 2.4 Kinect kamera ile elde edilen iskelet yapısı

3

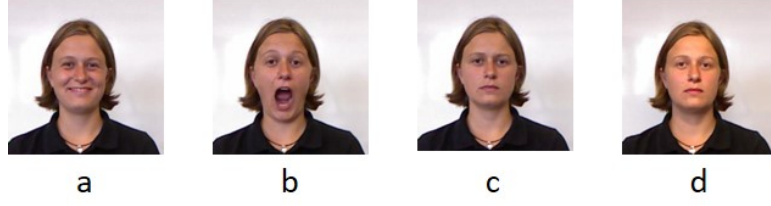
ENERJİ GÖRÜNTÜSÜ VE YÜZ GÖRÜNTÜLERİ İLE SANAL VERİ SETİ OLUŞTURULMASI

Bu bölümde enerji görüntülerinin kişi tanıma üzerinde olumlu etkilerinin incelenmesi amacıyla multi biyometrik füzyon eğitimleri yapılmıştır. Hali hazırda yüz görüntüleri ve enerji görüntüleri içeren bir veri seti bulunmadığı için bu özellikleri içeren 2 farklı veri seti birbirleri ile eşleştirilmiş ve sanal bir veri seti oluşturulmuştur. Bu sayede enerji görüntüsünün kişi tanıma üzerindeki etkileri incelenmiştir. İlk olarak yüz tanıma performansı Eurocom Face veri seti kullanılarak incelenmiştir. Ardından BodyLogin hareket veri seti kullanılarak enerji görüntüleri oluşturulmuş ve bu veri setinin performansı incelenmiştir. En son olarak ise veri setleri eşleştirilerek oluşan sanal veri setinin performansı incelenmiştir.

3.1 Eurocom Face Veri Seti

Yüz tanıma uygulaması için Eurocom face veri kümesi [32] kullanılmıştır. Bu veri kümesi, 14 kadın ve 38 erkek olmak üzere 52 kişinin, farklı ifade ve farklı aydınlatma koşullarındaki yüz görsellerini içermektedir. Kinect kamera kullanılarak toplanan veri seti, aralarında 15 gün bulunan 2 farklı çekim seansını içermektedir. Bu çalışmada, EURECOM veri kümesindeki ilk 40 kişiden iki seansta çekilmiş olan ve Şekil 3.1'de gösterilen, önden ışık (light on), doğal (naturel), gülümseme (smile) ve ağız açık (open mouth) görüntüleri kullanılmıştır. Bu sayede her bir kişi için farklı ışık ve ifadelerde 8 resim içeren 40 kişilik bir veri kümesi oluşturulmuştur.

Bu veri kümesi 2015 yılında Parkhi ve diğerleri tarafından önerilen VGG Face modeline uygulanmıştır [33]. VGG Face, 16 katmandan oluşan bir derin öğrenme modelidir ve 2600 kişinin 2,6 milyon resmi ile oluşturulmuştur. Sınıflandırma eğitimi öncesinde bazı ön işlemler uygulanmıştır. Bunlardan ilki, yüz üzerindeki karakteristik (landmark) noktalarından yararlanmak için *DlibC++* algoritması [34] ile yüz üzerindeki 68 noktaya ait pozisyon bilgileri elde edilmesidir. Şekil 3.2.a ve Şekil 3.2.b'de, elde edilen yüze ait bu karakteristik noktalar işaretlenmiştir. Bir diğer



Şekil 3.1 KinectFaceDB varyasyon görünümü: a)önden ışık b)natürel c)ağız açık d)gülümseme

ön işlem ise, gözlerin yatay düzlemde aynı pozisyona oturtulmasıdır. Bunun için poz düzeltme ve ölçekleme işlemleri uygulanır (Şekil 3.2.b). Poz düzeltme için göz koordinatları kullanılarak çevrilme açısı hesaplanır. Ölçekleme faktörü γ gözler arası maksimum mesafe bilgisi (1) kullanılarak elde edilmektedir.

$$\gamma = maks(D_1, D_2, \dots, D_m) \quad (3.1)$$

$$S_k = \left(\frac{\gamma}{D_k} \right) \quad (3.2)$$



Şekil 3.2 a) orijinal resim b) ölçek ve poz normalizasyonu c) İlgi alanı d) aydınlatma normalizasyonu

3.2 Davranışsal Biyometrik Veriler ve Yüz görüntüleri ile sanal Veri Seti Eşleme

Davranışsal biyometri için BodyLogin veri seti kullanılmıştır [35]. Bu veri seti, yaşları 18 ile 33 arasında değişen, 13 kadın ve 27 erkek olmak üzere toplam 40 kişiden alınan farklı hareketlerin silüet görüntülerini içermektedir. Her bir kişi ortalama 3 saniye süren bir takım hareketleri 20 kez tekrarlamıştır. Bu veri seti içerisinde özellikle kimlik doğrulamada daha zor olduğu belirtilen “iki el ile havada S çizme” hareketi çalışma için tercih edilmiştir. Hareketin yapılışı şeması Şekil 3.3’te gösterilmektedir.

S hareketi her ne kadar kolay görünse de hareketin yapılışındaki benzerlik ayırt ediciliği zorlaştırmaktadır. Veri seti içerisinde yer alan bir diğer hareket ise kişilerin



Şekil 3.3 S hareketinin yapıışı

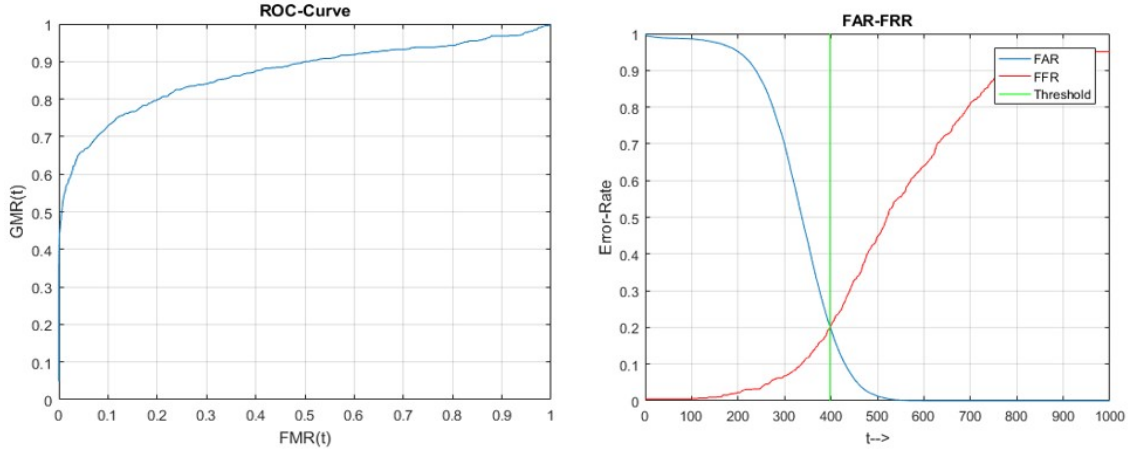
belirlediği hareketlerdir. Bu durum kimlik doğrulama işlemlerinde daha kolay olarak belirtilmektedir. Çünkü her kullanıcı kendine ait bir hareket kullanmaktadır. Bu başarıyı arttırsa bile taklit edilme durumunda sistemin kandırılma ihtimalini arttırmaktadır. Çünkü her ne kadar karmaşık olsa da farklı hareketler kullanıldığında, bu biyometrik bir tanımlamadan öte hareket tanıma işlemine girecektir ve hareket taklit edildiğinde bile doğru hareket olarak değerlendirilebilir. Bu durum da False Accepted Rate (FAR) oranının artması anlamına gelmektedir. Farklı kişilerin aynı hareketi yaptığı değerlendirme yönteminde ise sınıflar arası özellik farklılıkları, kişinin hareketi yaparken kendine has olan davranışlarından oluşacaktır. Belirlenen hareket, başkası tarafından taklit edilse bile kişiye ait farklılıklar sebebiyle ayırt edilebilecektir. Bu durum ise FAR oranının azalmasını sağlayacaktır. Bu sebeplerden dolayı S hareketi davranışsal biyometrik veri olarak tercih edilmiştir.

S hareketi 30 frame per second (fps) ile ortalama 3 sn'lik kayıtlardan oluşmaktadır. Bu da katılımcıların tekrar ettiği her bir S hareketinin ortalama 90 görüntüden oluşmasını sağlamaktadır. Bu görüntülerden yürüyüş enerji görüntüsü yönteminde olduğu gibi tüm görüntülerin toplanması ve ortalamasının alınması ile bir enerji görüntüsü üretilmiştir. Bu görüntüye hareket enerji görüntüsü (Gesture Energy Image) olarak adlandırılmış ve 3.3 nolu denklemdaki gibi formüle edilmiştir.

$$GesEI(x, y) = \left(\frac{1}{N}\right) \sum_{t=1}^N B_t(x, y) \quad (3.3)$$

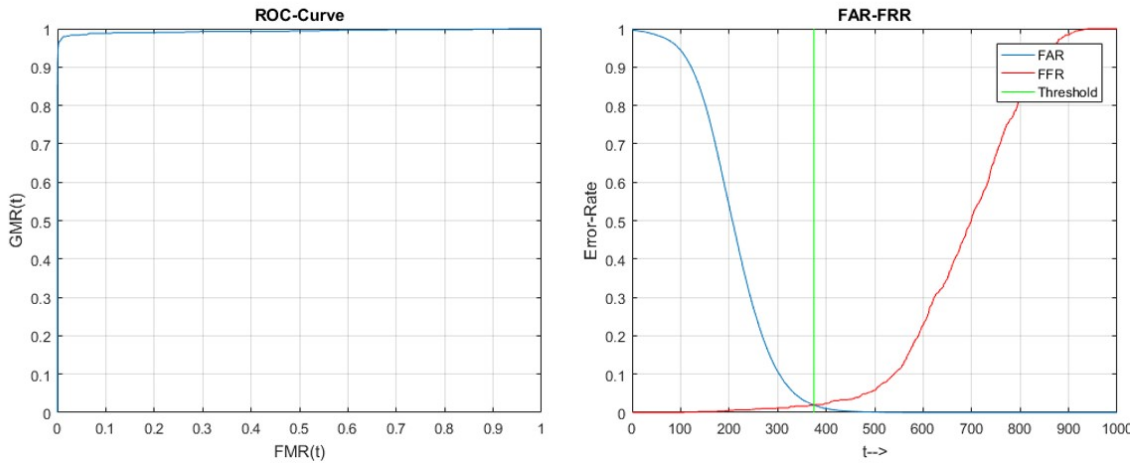
Sonrasında bu görüntüye ait özellikler PCA ile indirgenmiş ve ROC analizi yapılmıştır. Şekil 3.4'te S hareketine ait ROC eğrisi ve False Accepted Rate – False Rejected Rate (FAR-FRR) grafikleri gösterilmiştir. ROC curve eğrisini oluşturan Genuine Match Rate (GMR) – False Match Rate oranları ve hata oranına bağlı optimum eşik değeri incelendiğinde S hareketinin tek başına 0.2 oranında hata oranına sahip olduğu görülmektedir. Kişi tanıma için uygun bir biyometrik veri olmakla birlikte daha önce belirtildiği gibi kişi tanıma ve ayırt etmede zor bir veri olduğu da görülmektedir.

İkinci olarak toplam 40 kişiye ait farklı ışık seviyelerini ve farklı ifadeleri içeren yüz



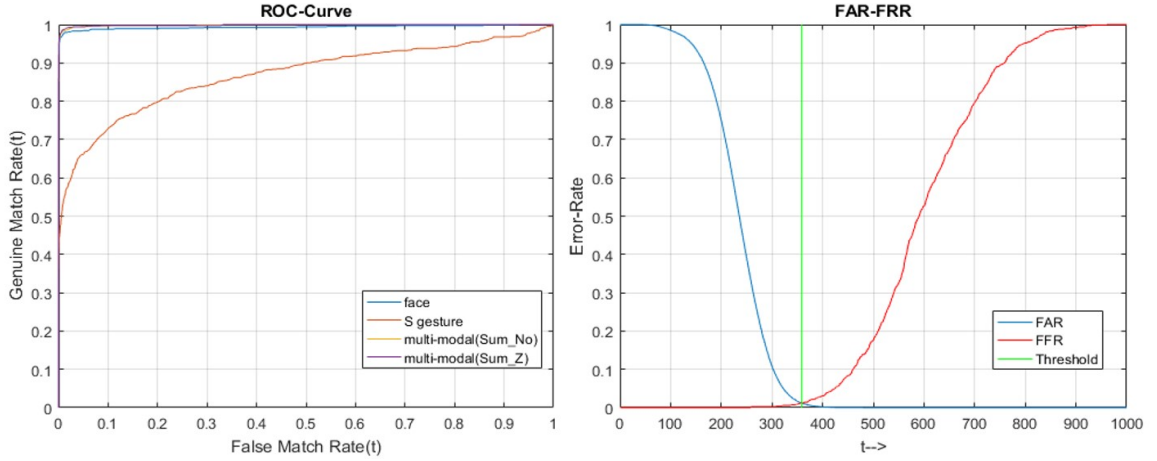
Şekil 3.4 S hareketine ait ROC eğrisi ve FAR-FRR grafikleri

verisine ait VGG Face verileri PCA ile indirgenmiş ve elde edilen öz niteliklere ait ROC analizi yapılmıştır. Elde edilen sonuçlar Şekil 3.5'te sunulmuştur. Eğitilmiş VGG Face ağından elde edilen özellikler kişi tanımda %97 gibi bir başarıma ve aynı şekilde oldukça bir düşük hata oranlarına ulaşmıştır.



Şekil 3.5 VGG Face ROC eğrisi ve FAR-FRR grafikleri

Bir sonraki aşamada ise VGG Face ağından elde edilen yüz verileri ile S hareketine ait veriler multi-biyometrik bir inceleme için eşleştirilmiş ve sanal bir veri seti oluşturulmuştur. Sanal veri seti olarak adlandırılmasının sebebi, yüz veri setlerindeki kişiler ile hareketi yapan kişiler aslında aynı değildir. Ancak silüet görüntülerinde belirgin bir yüz görüntüsünün olmaması bu iki veri setini eşleştirmeye olanak sağlamaktadır. Bu sayede hareket ve yüz görüntüsü ile yapılacak bir füzyonda başarımın ne şekilde etkileneceği daha doğru ve güvenilir olarak incelenebilecektir. Score seviyesinde birleştirilmiş olan çoklu-biyometri performansı incelenmiştir. Oluşturulan bu veri setine ait başarımlar analizleri ve grafikleri Şekil 3.6'de gösterilmektedir.



Şekil 3.6 Çoklu Biyometrik ROC eğrisi ve FAR-FRR grafikleri

Şekil 3.6’de ki grafiklerden de görüldüğü gibi oluşturulan sanal veri setine ait biyometrik sınıflandırma başarımı bu biyometrik verilerin tek başına kullanıldığı durumlara göre daha yüksek, hata oranları ise daha düşük çıkmıştır. Soldaki grafikte yüz, S hareketi, multi-modal ve multi-modal Z normalizasyonu için Roc eğrileri sunulmuştur. Score seviyesi birleştirme işlemi için Sum Rule ve Z normalizasyonu uygulanmıştır. En iyi sonuçlar Sum Rule Z-normalizasyon ile elde edilmiştir. Tablo 3.1’te yapılan kıyaslamalara ait EER sonuçları gösterilmektedir.

Tablo 3.1 Kullanılan veri setlerine ait EER sonuçları

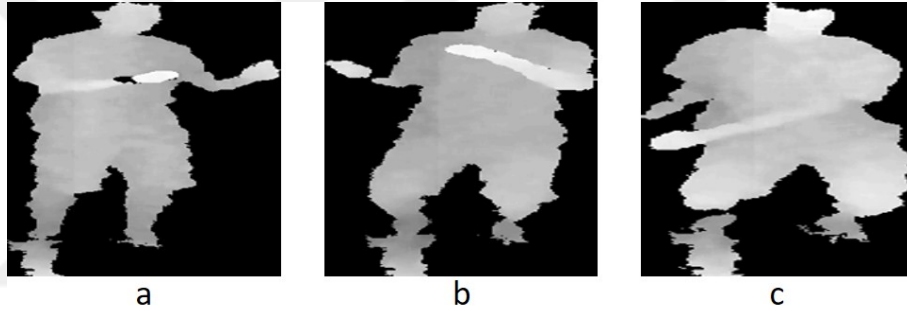
Veri setleri	FACE	S GESTURES	Multi Modal Sum Rule	Multi Modal Sum Rule-Z normalization
EER	0.0194	0.2157	0.0144	0.0128

Bu sonuçlar yüz görüntüsü gibi belirgin bir biyometrik verinin daha davranışsal biyometrik veri olan hareket silüeti ile birleştirilmesinin olumlu avantajlar sağladığını göstermektedir. Hatanın azaltılması, güvenilirliğin artırılması ve kişi tanımda harekete bağlı davranışsal biyometrinin olumlu etkilerini göz önüne sermektedir. Üstelik davranışsal biyometrinin aslında aynı kişiye ait olmamasına rağmen eşleştirilmiş olduğu biyometrik verinin ayırt edici özelliklerin arttığı gözlemlenmiştir [36]. Bu çalışma Polonya’da düzenlenen INISTA-2017 uluslararası konferansına “A Multi-Biometric Recognition System Based On Deep Features of Face and Gesture Energy Image” başlığı ile sunulmuş ve kabul edilmiştir.

3.3 Davranışsal Biyometrik Hareketlerin Başarımını Arttırmaya Yönelik Çalışmalar

Bu bölümde yukarıda anlatılan davranışsal hareketlerin başarımlarını arttırmaya yönelik çalışmalardan bahsedilecektir. GesEI olarak tanımladığımız hareketlere ait

enerji görüntüleri, o harekete ait görüntülerin toplanıp ortalamasının alınması ile elde edilmektedir. Bu yöntem ilk olarak yürüyüş analizi için uygulanmıştır. Yürüyüş hareketi incelendiğinde bu hareketin periyodik olarak tekrar ettiği görülmektedir. Yürüyüş analizi için ayırt edici bir enerji görüntüsü oluşturmada, silüetlerin genel ortalamasının yeterli olabileceği düşünülebilir. Ancak bu çalışmada kullanılan hareketler tek seferlik ve periyodikliği bulunmayan hareketleri de içermektedir. Örneğin: yürüme hareketinde tekrarlanan bir süreç vardır ve hareketin bitiş ve yeniden başlangıcı arasında enerji farkı oldukça azdır. Ancak S hareketi başlangıç ve sonu itibari ile herhangi bir periyodiklik ve tekrar içermemektedir. Hareketin başlangıç noktası, orta nokta ve bitiş noktaları arasındaki bilgi ve enerji yoğunluğunun ve değişiminin çok daha yüksek olduğunu Şekil 3.7’de görselde daha net görülmektedir. Bu yüzden yürüyüşte olduğu gibi hareketleri art arda toplamak yerine, enerji farklılığının daha yüksek olduğu kısımların daha çok vurgulanabileceği enerji görüntüsü oluşturma yöntemleri üzerinde yoğunlaşmıştır.

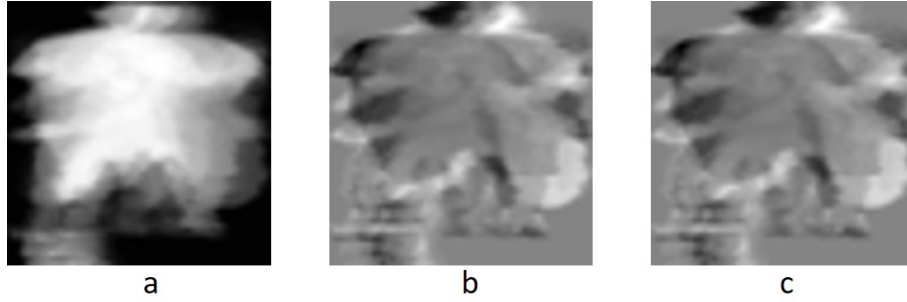


Şekil 3.7 S hareketine ait a) başlangıç anı b) orta nokta anı c) bitiş noktası görselleri

İlk olarak enerji farklılığının daha yüksek olduğu ilk ve son resimlerden başlayarak ortaya doğru farklarının alınması düşünülmüştür. Bu işlem ifade (3.4)’te gösterildiği gibi uygulanmıştır. Bu denklem GesEI’den farklı olarak “fGesEI” olarak adlandırılmıştır. fGesEI, en uzak resimlerin farklarının alınması ve sonraki fark resimleri ile toplanması ile elde edilen bir enerji görüntüsünü ifade etmektedir. N harekete ait toplam resim sayısını, I ise harekete ait resimleri ifade etmektedir. Bu şekilde elde edilen görüntüler Şekil 3.8’de gösterilmektedir. Ayrıca ifade (3.5)’te fGesEI’den farklı olarak başlangıç ve orta nokta arasındaki resimlerin farkını alarak sona doğru kaydırma yöntemi uygulanmış ve bu denklemde mGesEI olarak ifade edilmiştir.

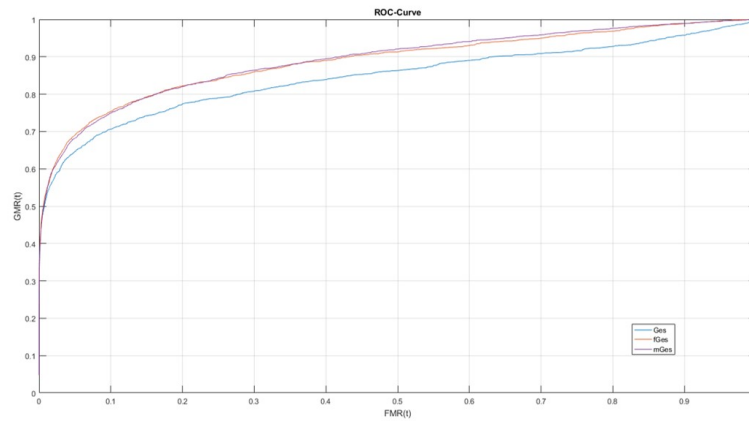
$$fGesEI(x, y) = \left(\frac{1}{N}\right) \sum_{t=1}^N (I_{N-i} - I_i) \quad (3.4)$$

$$mGesEI(x, y) = \left(\frac{1}{N}\right) \sum_{t=1}^N (I_i - I_{\frac{N}{2}+i}) \quad (3.5)$$



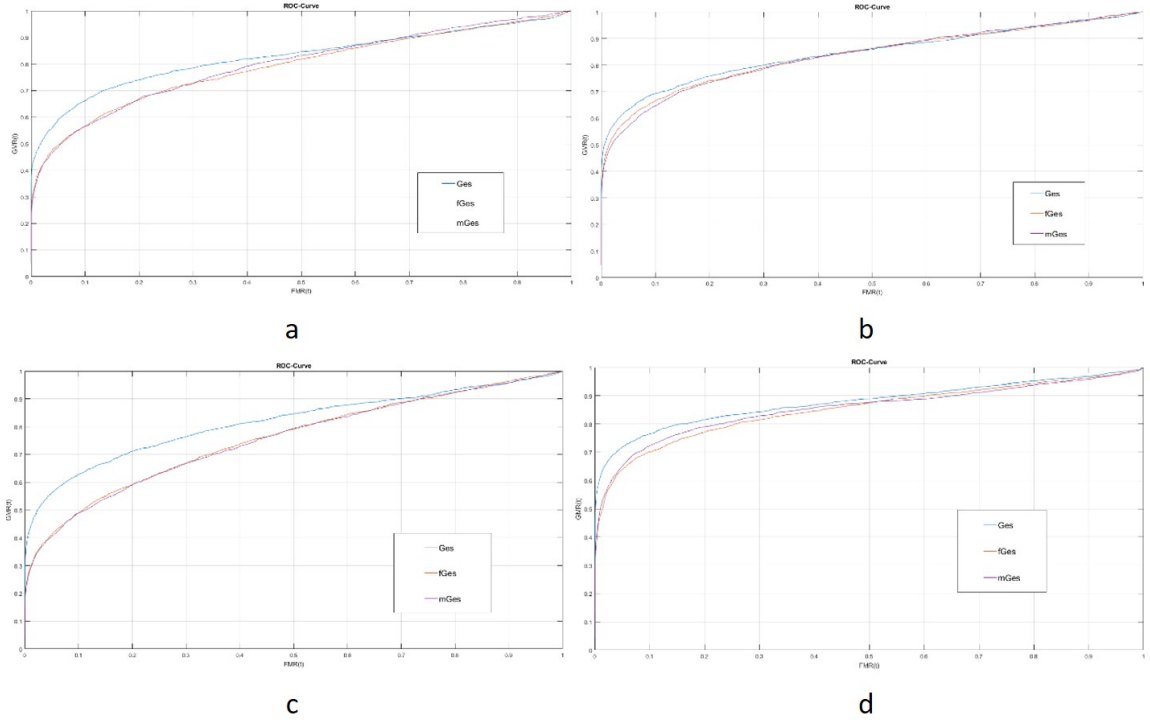
Şekil 3.8 a)GesEI ile elde edilen görüntü b)fGesEI ile elde edilen görüntü c)mGesEI ile elde edilen görüntü

Ancak elde edilen enerji görüntüsü görsellerinde mGesEI ve fGesEI arasında belirgin bir farklılık görülmemiştir. Bu işlemler uygulanmadan önce 224x224 boyutundaki hareket resimleri 4'te 1 oranında boyutlandırılarak 56x56 boyutuna düşürülmüştür. Bu sayede yapılacak işlemlerde zaman maliyeti azaltılmaya çalışılmıştır. Şekil 3.8'da görülen üç farklı enerji görüntüsüne ait başarımlar incelendiğinde, S hareketi için fGesEI yöntemi ile üretilen görüntülerin daha iyi başarımlar gösterdiği Şekil 3.9'da ki sonuçlardan anlaşılmaktadır.



Şekil 3.9 GesEI, fGesEI ve mGesEI ile elde edilen görüntülere ait ROC eğrileri kıyaslaması

fGesEI ve mGesEI yöntemlerinin başka hareketlerde de başarımlar sağlayıp sağlayamadığını anlamak için Body Login database içerisinde yer alan diğer 4 hareket içinde kullanılmıştır. Alınan sonuçlar Şekil 3.10'de sırası ile gösterilmektedir.



Şekil 3.10 a)1. Hareket b)2. Hareket c)3. Hareket d)4. Hareket için elde edilen sonuçlar

Sonuçlar incelendiğinde fGesEI metodunun S hareketi dışında 2. hareket için benzer sonuçlar verdiği diğer 3 hareket için GesEI'nin daha iyi sonuçlar elde ettiği görülmüştür. Bu sonuçlar, hareketlerin özelliğine göre tek bir yöntemden ziyade farklı enerji görüntülerinin üretilebileceğini desteklemektedir. Bunun sonucunda yeni hesaplamalar yapılarak her hareket için yeni görüntüler oluşturulmuştur. Ancak elde edilen sonuçlar hangi yöntemin hangi harekette daha iyi sonuca ulaştıracığı konusunda net bir sonuç üretememiştir.

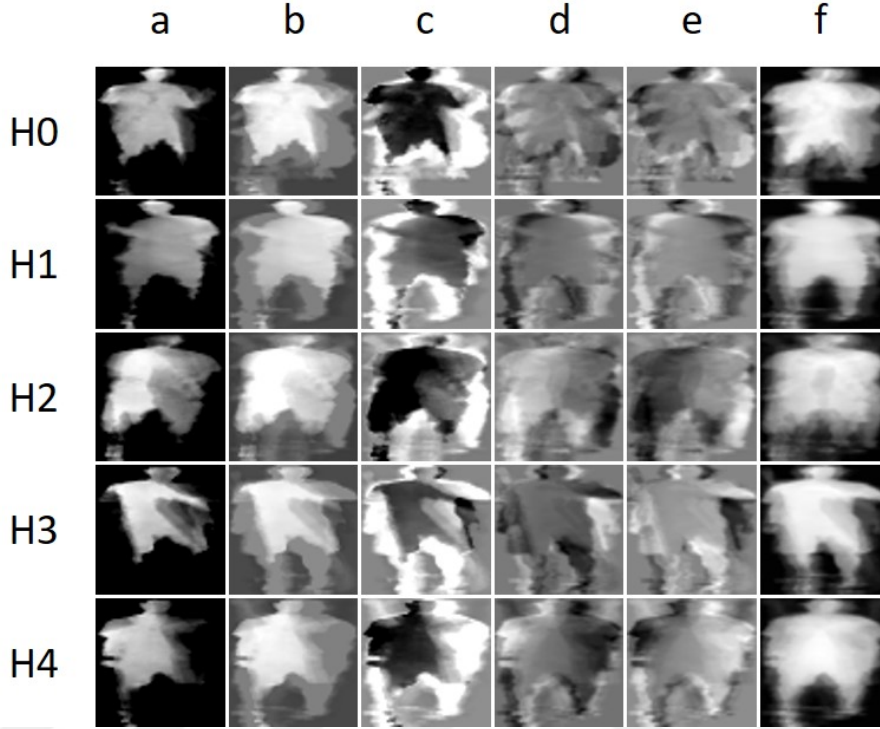
3.3.1 Oluşturulan Yeni Enerji Görüntüleri

fGesEI S hareketi için daha iyi bir başarımlı sağlarken, farklı hareketlerde farklı hesaplamaların daha yüksek başarımlara ulaştırabileceği fikrini oluşturmuştur. Bu bağlamda 5 farklı hareket için 6 farklı enerji görüntüsü oluşturulmuştur.

- (a) GesEI ve fGesEI görüntülerinin farkı alındı.
- (b) GesEI ve fGesEI görüntülerinin toplamalarının ortalaması alındı.
- (c) GesEI ve fGesEI görüntülerinin farkı ve toplamalarının ortalamasının farkı alındı.
- (d) Hareketlerin ilk yarısının ortalaması ile ikinci yarısının ortalamasının farkı alındı.

(e) Orijinal fGesEI görüntüsü.

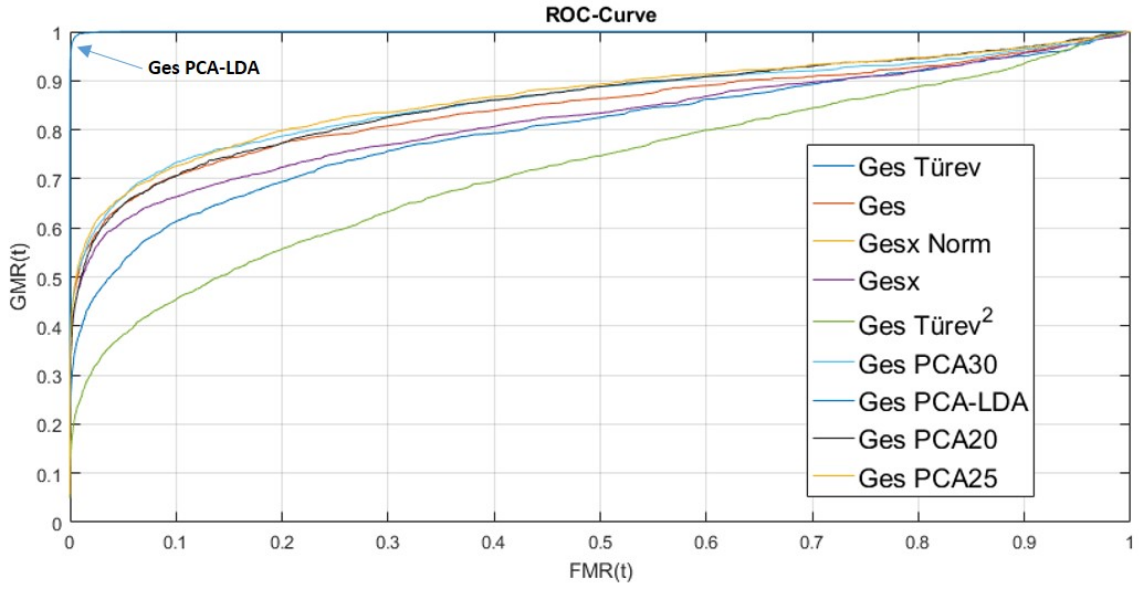
(f) Orijinal GesEI görüntüsü.



Şekil 3.11 a) b) c) ve d) yukarıda tanımlandığı sırayla konulmuştur. e) fGesEI
f)GesEI görüntüleridir

Şekil 3.11’de gösterilen bu görüntülerden E seçeneğindeki fGesEI yöntemi H0 için en iyi sonucu vermiştir. B seçeneğinde açıklanan yöntem ile elde edilen enerji görüntüsü ise H1 hareketi için en iyi sonucu vermiştir. Bunun dışındaki diğer hareketlerde GesEI yöntemi ile üretilen enerji görüntüsü en iyi sonuca ulaşmıştır. Bu sebeple GesEI yöntemi ile elde edilecek görüntüler üzerinde değişik parametreler ve metotlar denenmesi düşünülmüştür. Elde edilen kıyaslamalar Şekil 3.12’de gösterilmiştir.

Şekil 3.12’de GesEI yöntemi ile üretilen enerji görüntüleri öncesinde yapılan bazı ön işlemlerin kıyaslaması yapılmıştır. Bunlar, normal GesEI, orijinal boyutta görüntüden enerji görüntüsünün üretildiği Gesx ve Gesx Norm(Normalizasyonlu görüntü), Görüntü üzerinde farklı değerlerde PCA ve LDA yöntemi ile feature reduction yapılması ve görüntünün birinci ve ikinci türevlerinden elde edilen enerji görüntüleridir. Sonuçlar kıyaslandığında PCA ile seçilen öz niteliklerin azalması başarıyı arttırmaktadır. Enerji görüntüsü öncesi oluşturulan dizilerde türev alınması ise başarıyı olumsuz etkilemiştir. Resimlerin orijinal boyutlarında (Gesx ve Gesx Norm) küçültülmeden değerlendirilmesi yine başarıyı olumsuz etkilemiştir. Bu işlemler içinde en iyi sonuç PCA-LDA modelinde elde edilmiştir. PCA ile



Şekil 3.12 GesEI yöntemi üzerinde farklı yöntemlerin kıyaslanması grafiği

çıkartılan öznitelikler LDA ile düzenlendikten sonra elde edilen başarımların sonucu diğer denemelere kıyasla çok daha yüksek doğruluğa ulaşmıştır. S hareketi için %98 üzerinde başarımların sonucu üretmiştir. Bu yöntem diğer hareketler için üretilen görüntüler üzerinde de denenmiştir. Diğer hareketlerde ise %90-%98 aralığında sonuçlara ulaşılmıştır.

HAREKET ENERJİ GÖRÜNTÜSÜ VE HAREKET GEÇMİŞİ GÖRÜNTÜLERİNİN ELDE EDİLMESİNDE YENİ BİR YÖNTEM – ADAPTİF ENERJİ GÖRÜNTÜSÜ

Bu bölümde önceki bölümlerde yapılan enerji görüntüsünden hareket hakkında daha ayırt edici bilgiler elde edilmesi ve kişi tanıma içinde kullanılabilir seviyede başarıma ulaşması konusunda yeni bir yaklaşım anlatılacaktır. Literatür çalışmaları incelendikten sonra çalışmalarda görülen bir eksiklik üzerine yoğunlaşmış ve önceki yöntemlerden daha etkili olabileceği görülen yeni bir enerji görüntüsü yöntemi bu bölümde anlatılmıştır.

4.1 Adaptif Enerji Görüntüsü Yöntemi

Hareket enerji görüntüsü, ilk olarak yürüyüş hareketi için kullanılmış olup hareket süresince kişilerin siluet görüntülerinin çıkartılıp, bütün görüntülerin toplanması ve toplam görüntü sayısına bölünmesi ile elde edilmektedir. Binary görüntülerin her biri t zamanına bağlı iki boyutlu bir değişken $B_t(x, y)$ olarak tanımlanırsa, hareket enerji görüntüsü aşağıdaki gibi tanımlanabilir.

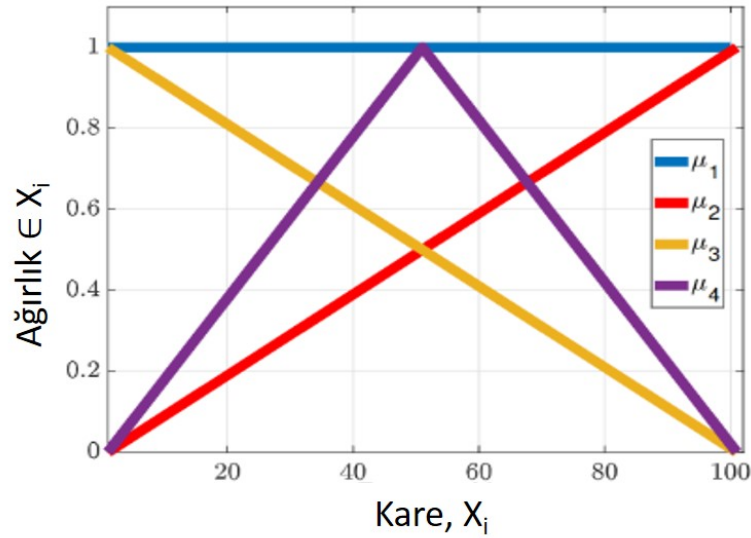
$$G(x, y) = \left(\frac{1}{N}\right) \sum_{t=1}^N B_t(x, y) \quad (4.1)$$

Yürüyüş gibi tekrarlanan hareketler yerine tek seferlik ve periyodik olmayan hareketler için ise hareket geçmişi enerjisi (Motion History Image - MHI) yöntemi tercih edilmektedir. Bu yöntemde, hareketi oluşturan her bir görüntü, hareketin yoğunlaştığı anları ve hareket bilgisini belirgin hale getirecek şekilde katsayılar ile çarpılır. Hareketin türüne göre frameler üzerinde önceden belirlenmiş katsayılar uygulanabilir. Yapılan çalışmalarda hareketin türüne göre belirlenen katsayıların, sınıflandırma

başarımını olumlu yönde etkilediği görülmüştür. Ancak hangi hareket için hangi katsayı yönteminin seçileceği deneysel olarak tespit edilebilmektedir. Geleneksel hareket enerjisi (HE) hesaplama yöntemi (4.2)'nolu denklemde ifade edildiği gibidir.

$$HE = \left(\frac{1}{255}\right) \sum_{t=2}^N W_i \cdot m(i) = \left(\frac{W_i}{255}\right) \sum_{t=2}^N m(i) \quad (4.2)$$

Burada n hareketini oluşturan toplam görüntü sayısını, m i. andaki görüntüyü, W_i ise i anındaki görüntünün piksel değerleri için bir katsayıyı ifade etmektedir. Yapılan güncel çalışmalarda hareket enerji görüntüsü ve hareket geçmişi görüntüsü elde etmede yeni yöntemler kullanıldığı görülmektedir. Yapılan çalışmalarda HE hesaplanmasında 3 farklı yöntem önerilmiştir [24]. Bu yöntemlerde kat sayı ifadesi $\mu(i) = \frac{W_i}{255}$ olarak ifade edilecek şekilde düzenlenmiştir. Geleneksel yöntem Şekil 4.1 üzerinde $\mu_1 = 1$ olarak gösterilmektedir. Literatürdeki diğer güncel yöntemlere ait grafikler ve açıklamaları Şekil 4.1'de sunulmuştur.



Şekil 4.1 Hareket enerjisi hesaplanmasında kullanılan katsayılar ve grafikleri [24]

$$\mu_1(i) = 1, \forall i \in [0 n] \quad (4.3)$$

$$\mu_2(i) = \frac{i}{n}, \forall i \in [0 n] \quad (4.4)$$

$$\mu_3(i) = 1 - \frac{i}{n}, \forall i \in [0 n] \quad (4.5)$$

$$\mu_4(i) = \begin{cases} \frac{2i}{n}, & 0 < i \leq \frac{n}{2} \\ 2 - \frac{2i}{n}, & \frac{n}{2} < i \leq n \end{cases} \quad \forall i \in [0 n] \quad (4.6)$$

Şekil 4.1’de gösterilen ve yukarıdaki denklemlerle ifade edilen her bir fonksiyon değeri, hareket boyunca elde edilen n adet görüntüyle çarpılmakta ve ortalaması alınarak hareket enerji görüntüsü elde edilmektedir. Bir önceki bölümde denemelerini yapmış olduğumuz enerji görüntüsü oluşturma yöntemlerinde olduğu gibi, bu fonksiyonlarda da yöntemlerin başarımı uygulandığı hareketlere göre değişiklik göstermektedir. Yani yapılan hareketin şekline, zamanlamasına veya hareketi yapan kişinin hız veya tekrar ediliş şekline göre seçilmesi gereken fonksiyon değişmektedir. Şekil 4.1’de yer alan μ_2, μ_3 ve μ_4 nolu fonsiyonlar incelendiğinde bu fonksiyonların hareketin başlangıç, orta veya son kısmını yüksek katsayı ile vurguladığı ve diğer alanları ise düşük katsayılar ile baskıladığı görülmektedir. Bu durum ise en başarılı sonucu verecek HE fonksiyonunun belirlenmesinde tüm yöntemlerin denemesini veya her bir hareketin hangi aralıkta daha yoğun bilgi içerdiğinin incelenmesini gerektirmektedir.

Tez çalışmasının bu bölümünde, hem deneme yanılma yönteminin giderilmesi hem de hareket içerisindeki değişim bilgisini en iyi şekilde temsil edebilecek yeni bir HE hesaplama yöntemi önerilmektedir. Bu yöntemde, yapılan harekete adapte olarak katsayıların üretilmesini ve bu katsayılar ile enerji görüntüsünün elde edilmesini amaçlanmıştır. Bu sayede, hareketin kendi doğasında oluşan değişimler tespit edilmiş ve hareket ile ilişkili bir katsayılandırma yöntemi geliştirilmiştir. Hareketin içerisinde bilgi yoğunluğu taşıyan kareler daha yüksek katsayılara sahip olurken durağan olan kareler ise daha düşük katsayılar ile eşleştirilecektir. Bu katsayılandırma yöntemi görüntüler arası korelasyonun hesaplanmasına dayanmaktadır. Bilindiği üzere korelasyon, iki yada daha fazla değişken arasındaki doğrusal değişim ilişkisini göstermektedir. Bu sayede, bir hareketin farklı zamanlardaki görüntüleri arasındaki değişiklikler korelasyon ile incelenebilir. Yani iki görüntü arasında korelasyon düşükse bireyin hareketinde bir değişim olduğu, korelasyon yüksek ise bireyin hareketinde değişim olmadığı sonucuna ulaşılabilir. Bunu daha iyi ifade etmek için Tablo 4.1’de açıklanan “Pearson Momentler Çarpımı Korelasyon Katsayıları” incelenebilir.

Tablo 4.1 Korelasyon Değerleri için İlişki Durumları

R (Korelasyon Değeri)	İlişki Durumu
0.00	İlişki yok
0.01 – 0.29	Düşük Düzeyde İlişkili
0.30 – 0.69	Orta Düzeyde İlişkili
0.70 – 0.99	Yüksek Düzeyde İlişkili
1	Mükemmel İlişki

$$P(A, B) = \frac{1}{N-1} \sum_{i_1}^N \left(\frac{A_i - \mu_A}{\sigma_A} \right) \cdot \left(\frac{B_i - \mu_B}{\sigma_B} \right) \quad (4.7)$$

$$P(A, B) = \frac{cov(A, B)}{\sigma_B \cdot \sigma_A} \rightarrow R = \begin{pmatrix} P(A, A) & P(A, B) \\ P(B, A) & P(B, B) \end{pmatrix} \quad (4.8)$$

Tablo 4.1, (4.7) ve (4.8) nolu denklemler birlikte incelendiğinde, (4.8)'de $P(A, A)$ ve $P(B, B)$ değişkenin kendisi ile olan ilişki değerini belirttiği için "1" değerini almaktadır. $P(A, B)$ ve $P(B, A)$ değerleri birbirine eşit olup A ve B değişkenleri arasındaki korelasyon değerini (R) vermektedir.

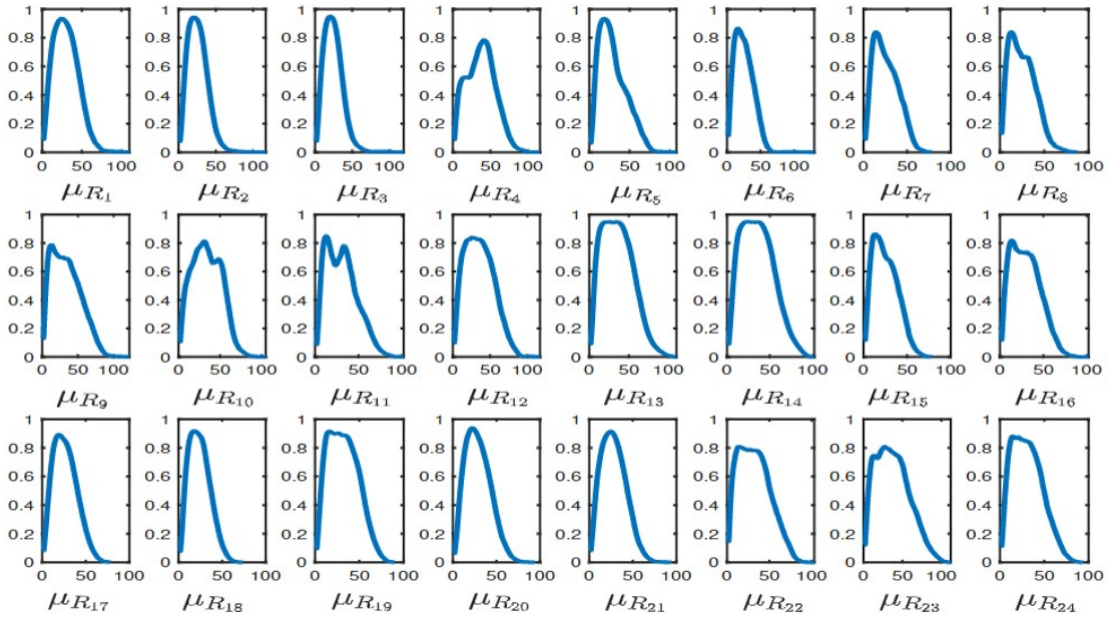
Bir hareket boyunca elde edilen tüm görüntülerin, referans olarak seçilecek bir görüntü (A) ile korelasyonları incelenecek olursa, ve seçilecek referans görüntü hareket başlamadan önceki durağan poz olarak seçilirse, bireydeki herhangi bir değişim sonucunda korelasyon değeri düşecek, bireyin tekrar durağan pozisyona geçmesiyle korelasyon değeri tekrar yükselecektir. Bu aralıkta kalan kısım bireyin hangi aralıkta ve hangi oranda bir değişim yaptığı hakkında bilgi verecektir. Bir hareket boyunca tüm görüntüler aynı kişiye ait olacağından, (4.8)'de yer alan $P(A, B)$ veya $P(B, A)$ değerleri Tablo 4.1'de belirtilen aralıklardan *düşük*, *orta* veya *yüksek düzeyde ilişki* ifade edecek şekilde bir değer alacaktır. Bu değerlerin daha iyi temsil edilebilmesi için hareket boyunca üretilen korelasyon değerleri 0 – 1 aralığına normalize edilir ve R_N olarak ifade edilir.

Elde edilen korelasyon katsayıları, değişimin çok olduğu görüntülerde düşmekte, değişimin azaldığı görüntülerde ise artmaktadır. Ancak enerji görüntüsünde, değişimin çok olduğu görüntülerde daha belirgin, az olduğu görüntülerde ise daha bastırılmış olması istenmektedir. Bu amaçla kat sayılar $1 - R_N$ olarak hesaplanır. Bu yöntemde, bir hareket dizisi sırasında meydana gelen değişiklikleri belirlemek için ilk kare referans alınmakta ve tüm kareler arasındaki korelasyon katsayıları hesaplanmaktadır. Elde edilen katsayılar, harekete uyumlu bir zamansal şablon (Temporal Template-TT) fonksiyonu oluşturmak için kullanılır.

Adaptif TT fonksiyonu tarafından oluşturulan enerji görüntüleri, şekil 4.1'deki gibi belirli aralıkları vurgulayan mevcut TT'ler (statik-TT) tarafından oluşturulan enerji görüntülerine kıyasla hareketlerin daha iyi bir temsiline sahip olacaktır. Bu şekilde birden fazla TT'nin hesaplanması ve füzyon işlemi ihtiyacı ortadan kalkarak, işlem maliyeti azaltılacaktır. Bu amaçla önerilen yöntem ise denklem (4.9) ve (4.10)'da ifade edilmektedir.

$$\mu_{corr}(i) = (1 - R_N(i)), \forall i \in [0 n] \quad (4.9)$$

$$EI_{corr} = \sum_{i=1}^n \mu_{corr}(i).m(i) \quad (4.10)$$



Şekil 4.2 NATOPS veri setinden 24 farklı hareket üzerinde $(1 - R_N)$ işlemi ile TT grafikleri elde edilmiştir. Hareket çerçeveleri sabit olmasına rağmen bazı hareketler daha erken bazı hareketler daha geç bitmiştir. Bazı hareketlerde başlangıç kısımları vurgulanırken, bazılarında ise hem başlangıç hem de orta kısımlar vurgulanmıştır. Önerilen yöntem hareketlere uyarlanabilir olduğu için elde edilecek TT'ler veri setine ve hareketlerin içeriğine göre değişiklik gösterecektir

Önerilen bu yöntem sayesinde enerji görüntüsü elde edilirken uygulanan katsayılar Şekil 4.2'de gösterildiği gibi hareketin kendi doğasına göre otomatik olarak oluşacağından deneysel yöntemlerle kat sayı belirlenmesine gerek kalmayacaktır. Bu ağırlıklar NATOPS veri setindeki 24 farklı eylemden elde edilmiştir. Her hareket için oluşturulan uyarlanabilir TT'ler burada görülebilir. Ayrıca tüm şablonlar incelendiğinde hareketlerin tamamlanma sürelerinin, etki alanlarının ve değişim

yoğunluğunun farklı olduğu görülmektedir. Grafikler 0–100 kare ölçeğinde gösterilse de bazı hareketler 70–80 kare aralığında bitmekte, ancak hareket bittikten sonra video kayıtları devam ettiği için toplam kare sayısı hareket aralığından fazladır. Grafiklerde sona doğru 0 seviyesinde devam eden çerçeve katsayıları bunun sonucudur.

Örneğin Şekil 4.2’de ilk satırın soldan ikinci ve üçüncü şablonları yaklaşık olarak 30. karede en üst katsayıya ulaşmış, yaklaşık olarak 60. karede sona ermiş ve 100. kareye kadar devam etmiştir. Bu şablonda hareketin orta kısmı vurgulanmış, başlangıç ve bitiş kısımları bastırılmıştır. Ayrıca ikinci satırın soldan 6. şablonu ve son satırın 7. şablonu incelendiğinde hareketin 100. kareye kadar devam ettiği görülmektedir. Burada hareketlerin başlangıç ve orta kısımları vurgulanmış, hareketlerin bitiş kısımları bastırılmıştır.

Önerilen bu adaptif TT’ler ile enerji görüntüleri oluşturulduğunda, önceki yöntemlerde olduğu gibi uygun zamansal şablonu seçmek için deneysel yöntemlere veya hareket değişim alanlarını incelemeye gerek kalmayacaktır.

Bu tür RGB-D tabanlı çalışmalar, literatürde daha çok hareket tanıma için kullanılmaktadır [27]. Bu çalışmadaki diğer bir katkı da, kişi tespiti için mevcut ve önerilen yöntemlerin de incelenmiş olmasıdır. Performans analizi için VGG-F, ResNet-18 ve el yapımı öznitelik çıkarıcılar ve sınıflandırıcılar kullanılmıştır.

Bir CNN türü olan VGG-F ve ResNet-18 ağı, yöntemimizi değerlendirmek ve hareket tanıma ve kişi tanımlama için [24]’de sunulan mevcut geçici şablon temsilleriyle karşılaştırmak için kullanılır. Bu kıyaslamalarda İlk olarak, eğitimin sonuçları literatürde daha önce yapılmış çalışmalarla karşılaştırılmıştır. Ardından önerdiğimiz adaptif-TT’nin avantajı ve performansı önceki TT’ler ile aynı koşullar altında test edildi. Son olarak, adaptif-TT ile elde edilen performansın, CNN ağlarının gücünden gelen bir avantaj mı olduğu yoksa önerilen uyarlamalı yöntemden mi kaynaklandığını daha iyi anlamak için mevcut TT’ler ve yöntemimiz arasındaki performans, hand-crafted öznitelik çıkarıcılar ve sınıflandırıcılar ile incelenmiştir. Bu sayede çok sayıda ve oldukça farklı eğitim oranlarında kıyaslama testleri yapılmıştır.

Deneyler için “BodyLogin” [35, 37], “Naval Air Training and Operating Procedures Standardization” (NATOPS) [38] ve “SBU Kinect” [39] veritabanları kullanılmıştır. Sonuçlar, önerdiğimiz uyarlamalı yöntemin, mevcut üç şablon ve onların füzyonundan daha yüksek performans ve daha kısa işlem süresi sağladığını göstermektedir. Kişi tanımlama amaçlı yapılan eğitim sonuçları ise RGB-D sensörler ile elde edilen görüntülerin, kişi tanımlama için etkin bir şekilde kullanılabileceğini göstermiştir. Özetle bu yöntemin avantajları şu şekilde sıralanabilir:

- 1) Eylemin tamamlanma süresine uyarlanabilir.
- 2) Hareket sırasında değişikliklerin daha fazla olduğu aralıklar belirlenir ve vurgulanır. Tersine, herhangi bir değişiklik içermeyen aralıklar bastırılır.
- 3) Deneysel yöntemlerle bir eylem için hangi statik TT'nin daha uygun olduğunu belirlemeye gerek yoktur.

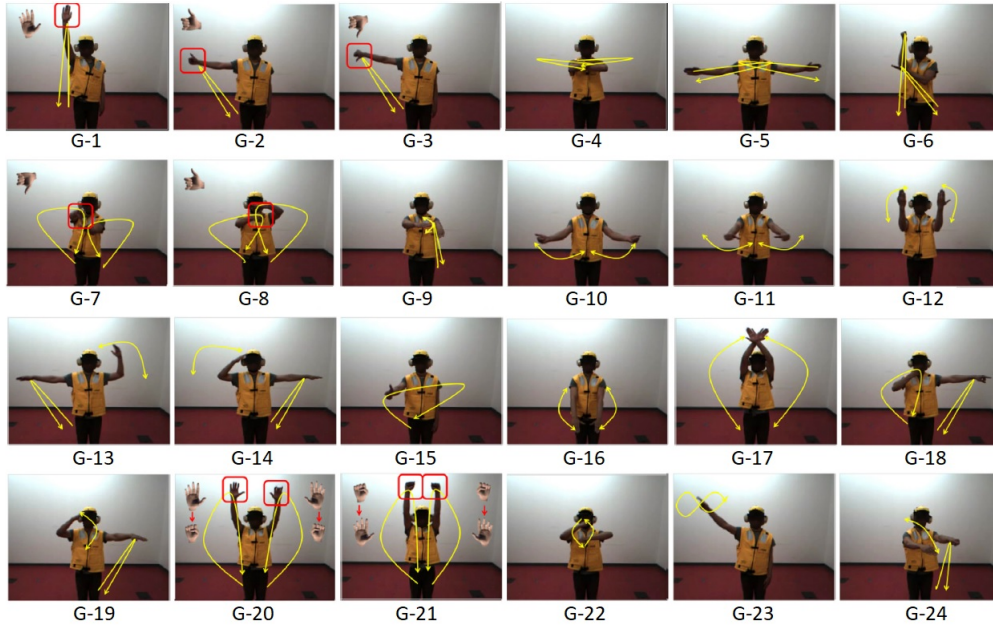
4.2 Kullanılan Veri Tabanları ve Yöntemler

4.2.1 Karşılaştırma

Bu çalışmada, önerilen yöntemin katkısını gözlemlemek için üç veri tabanı kullanılmıştır. NATOPS ve SBU Kinect veri tabanlarındaki görseller, kişilerin hem tüm vücut hareketine göre hem de sadece hareket eden uzuvlarına göre değerlendirildi. Bu yöntem ile her iki veri tabanından ikişer farklı enerji görüntüsü veri seti oluşturulmuştur. Ayrıca BodyLogin veri tabanı kullanılarak bir veri seti oluşturulmuştur. Bu şekilde beş farklı veri seti elde edilmiştir. Mevcut yöntemler ile önerilen yöntemimiz arasındaki doğruluk performansı bu veri setleri kullanılarak incelenmiştir. Veri setlerinin detayları ve özellikleri alt bölümlerde açıklanmıştır. Karşılaştırmalar için kullanılan eğitim ağları ve parametreleri de bu bölümün devamında açıklanmıştır.

4.2.1.1 NATOPS Veri Tabanı

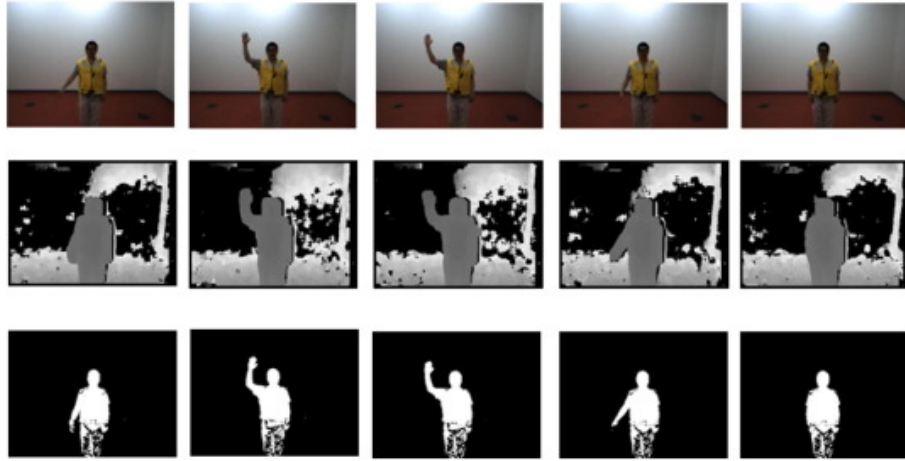
NATOPS veri tabanı, hava taşıtlarının yönlendirilmesi için kullanılan 24 hareketten oluşur. Bu hareketler NATOPS kılavuzuna göre düzenlenmiştir. Bu veri setindeki hareketler, vücudun üst kısmında yer alan kol hareketlerini ve *başparmak yukarı/aşağı*, *eller açık/kapalı* gibi el hareketlerini içerir. Veri tabanında bulunan 24 hareket 20 farklı kişi tarafından 20 tekrar yapılarak oluşturulmuştur. Tüm örneklerin ortalama uzunluğu 2.34 saniyedir ($std.sap. = 0.62$). Toplanan veri seti iki farklı formatta kaydedilmiştir. Birincisi, özellik veri kümesidir (CSV formatı). Buna vücut ve el özellikleri ile segmentasyon etiketleri dâhildir. İkincisi ise video veri setidir. Bu video veri seti, stereo kamera ile kaydedilmiş görüntüleri, derinlik haritalarını ve jestlerin maske görüntülerini içerir [38]. Bu hareketler, Şekil 4.3'te gösterildiği gibi 320×240 çözünürlükte 20 FPS'de bir Microsoft Kinect kamerası kullanılarak kaydedilmiştir. Çalışmamızda motivasyonumuz görüntülerden HE elde etmeye dayandığı için video veri seti tercih edilmiştir. Bu nedenle, hareket tanıma için kişi-hareket çiftleri oluşturulmuştur. Değerlendirme için farklı oranlarda eğitim ve test grupları oluşturulmuştur.



Şekil 4.3 [38]'te gösterildiği gibi NATOPS veri tabanındaki 24 hareketin kalıpları. Oklarla gösterilen yönde hareketler tamamlanır

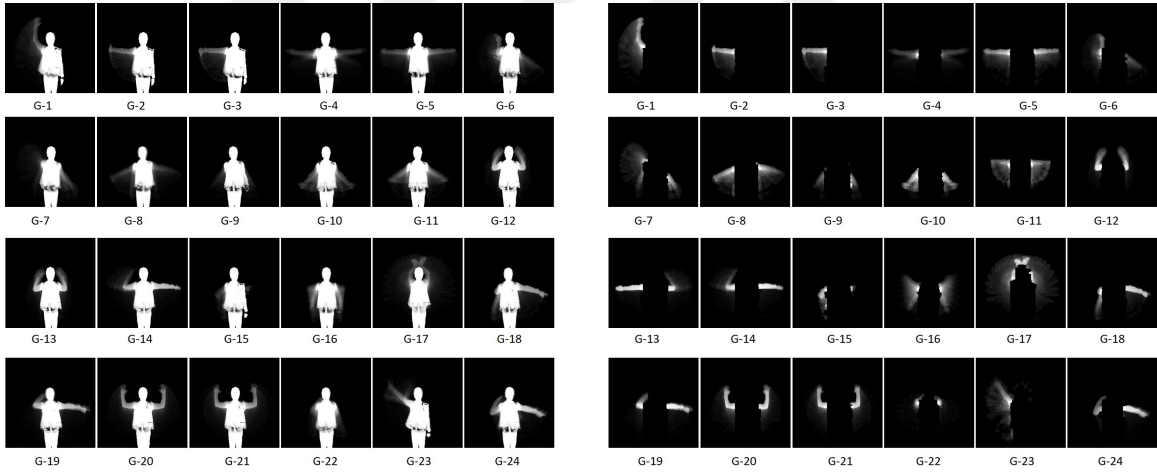
Bu veri seti üzerinde önceki bir çalışma olan [24]'de RGB özelliklerinin derinlik özelliklerinden daha başarılı olduğu sunulmuştur. Bunun nedeni, kollar ve vücut arasındaki yüksek kontrast farkına neden olan deneklerin giydiği sarı renk yeleştiği olabilir. Ek olarak, derinlik şablonunda deneklerin kolları vücutlarıyla örtüşmektedir. Bu örtüşme nedeniyle, kol hareketleri RGB şablonlarında derinlik şablonlarından daha iyi bir şekilde gözlemlenebilir. Ancak bu veri tabanı oluşturulurken denekler özel kıyafetler giymiştir. Bu avantaj her koşulda sağlanamayacak olup, denekler aynı kol ve vücut rengine sahip giysiler giydiğinde kaybolacaktır. Bu durumda RGB şablonlarından sağlıklı bir değerlendirme yapılamayabilir. Bu nedenle bu çalışmada deneklerin giydiği giysinin rengi ne olursa olsun her durumda aynı sağlamlığı gösterebilen maskeli derinlik şablonlarının kullanılması tercih edilmiştir. Video veri setindeki maskelenmiş derinlik görüntülerinin elde edilmesi Şekil 4.4'de gösterilmiştir. Burada, ilk satır RGB görüntüleri içerir. İkinci satırdaki görüntüler, Kinect kamera ile toplanan ham derinlik görüntüleridir. Bu ham görüntüler büyük miktarda arka plan gürültüsü içerir. Bu gereksiz bilgiyi kaldırmak için RGB görüntüleri ve ham derinlik görüntüleri maskelenir. Bu şekilde oluşturulan maskelenmiş derinlik veri seti son satırda gösterilmiştir.

Veri setleri oluşturulurken iki farklı enerji görüntüsü oluşturulmuştur. İlk görüntü türünde hareketli uzuvlar ve vücut gövdesi birlikte kullanılmıştır. Bu enerji görüntüleri Şekil 4.5'in sol tarafında gösterilmiştir. Bu veri kümesi bundan sonra NATOPS *Full-Body* olarak anılacaktır. İkinci tip görüntüde, sadece hareketli uzuv (el-kol) bilgisi kullanılır ve [24]'de kullanıldığı gibi gövde bilgisi görüntüden çıkarılır. Bu görüntüler



Şekil 4.4 İlk satır RGB görüntüler yer almaktadır. İkinci satırdaki görüntüler, Kinect kamera ile toplanan ham derinlik görüntüleridir. Son satır ise RGB görüntüleri ve ham derinlik görüntüleri ile oluşturulan maskelenmiş görüntüleri göstermektedir

Şekil 4.5'in sağ tarafında gösterilmektedir. Bu veri seti ise bundan sonra NATOPS *Moving-Limbs* olarak anılacaktır. Bu şekilde NATOPS veri tabanı kullanılarak iki farklı enerji görüntüsü veri seti üretilmiştir.



Şekil 4.5 NATOPS veri tabanı kullanılarak iki farklı enerji görüntüsü veri seti oluşturulmuştur. Bu görüntü, maskelenmiş derinlikli video akışlarında μ_R kullanılarak oluşturulan "NATOPS *FullBody*" (sol taraf) ve "NATOPS *MovingLimbs*" (sağ taraf) veri kümelerinin örneklerini göstermektedir

4.2.1.2 SBU Kinect Veri Tabanı

Bu veri seti, iki kişinin karşılıklı etkileşimini içeren sekiz hareketten oluşmaktadır. Bu hareketler [39]'da açıklandığı gibi sırasıyla *yaklaşma*, *uzaklaşma*, *tekmeleme*, *yumruk atma*, *itme*, *sarılma*, *el sıkışma* ve *nesne değiş tokuşu* şeklindedir ve Şekil 4.6'da gösterilmiştir. Bir nesneyi değiş tokuş etmek ve el sıkışmak gibi bazı eylemler yapısal olarak birbirine benzerlik gösterebilmektedir. Bu benzerlikler bu veri tabanın

zorluklarını artırmaktadır.



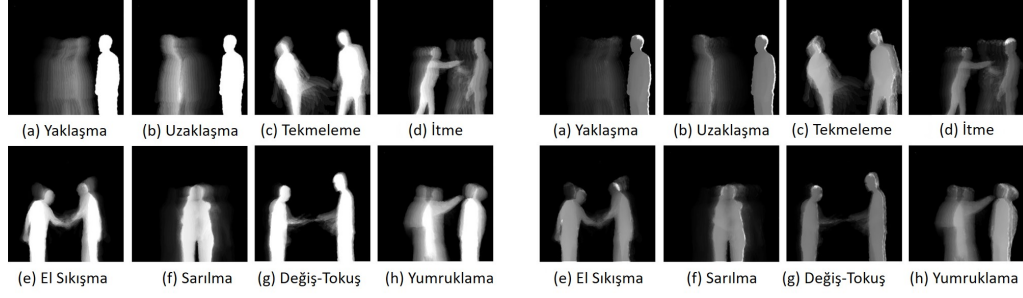
Şekil 4.6 [39]'da gösterildiği gibi SBU kinect veri tabanındaki 8 etkileşim kalıbı. a) yaklaşma, b) uzaklaşma, c) tekmeleme, d) yumruk atma, e) itme, f) sarılma, g) el sıkışma h) nesne değiş tokuşu. Yumruklama-itme, sarılma-yaklaşma ve tokalaşma-nesne değiş tokuşu gibi bazı hareketler birbirine yapısal olarak benzerlik gösterebilmektedir. Toplam hareket çeşitliliği daha az olsa bile, bu benzerlikler veri setini daha zorlu hale getirmektedir

Bu eylemlerin bazılarında (itme, tekmeleme, yaklaşma, uzaklaşma ve yumruk atma gibi), kişilerden biri hareketi başlatırken diğer kişi buna tepki verir. Veri seti kapsamında her etkileşim için, saniyede 15 kare (FPS) ile 640×480 piksel çözünürlükte RGB-D özellikte video görüntüleri elde edilmiştir. Hareketin tipine bağlı olarak her hareket 20 ile 40 kare aralığında tamamlanır. Bu veri tabanındaki her bir etkileşim 21 farklı çift ile elde edilmiştir. NATOPS veri setinde olduğu gibi, SBU Kinect veri tabanı kullanılarak iki farklı HE veri seti oluşturulmuştur. İlk veri setinde vücudun tamamına ait enerji görüntüsü bilgisi kullanılmıştır. Bu enerji görüntüsü örnekleri Şekil 4.7'nin sol tarafında gösterilmektedir.

Bu veri seti bundan sonra "SBU *Full-Body*" olarak anılacaktır. İkinci veri setinde ise sadece "vücudun hareketli uzuvları" bilgisi kullanılmış ve görüntülerden 'sabit vücut' bilgisi kaldırılmıştır. Bu görüntülerin bazıları Şekil 4.7'nin sağ tarafında gösterilmektedir. Bu veri kümesi bundan sonra "SBU *Moving-Limb*" olarak anılacaktır. Şekil 4.7'de gösterildiği gibi, SBU *Moving-Limbs*, NATOPS *Moving-Limb* veri setinin aksine tüm vücut görüntüsü içerir. Bunun sebebi ise, SBU Kinect veri setindeki etkileşim sırasında vücudun tamamında aktif hareketler oluşmaktadır.

4.2.1.3 BodyLogin Veri Tabanı

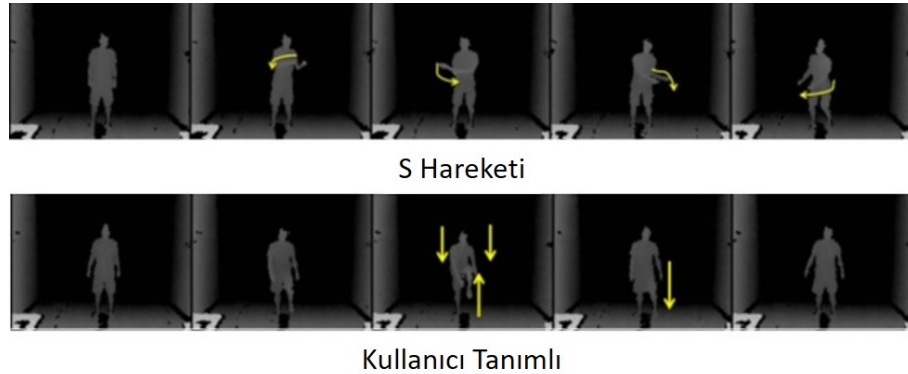
BodyLogin veritabanı, farklı zamanlarda kaydedilen farklı veri kümesi gruplarından oluşturulmuştur. Bu çalışmada, bu veri setlerinden ikisi birleştirilmiş ve kullanılmıştır. Bunlardan ilki "*Silhouettes and Skeletons (BLD-SS)*" veri setidir ve 40 farklı bireyin (18-33 yaş arası 27 erkek ve 13 kadın) tarafından gerçekleştirilen iki tür hareket içerir.



Şekil 4.7 SBU Kinect veri tabanı kullanılarak iki farklı HE veri seti oluşturulmuştur. Bu görüntülerde, derinlikli video akışlarında μ_R fonksiyonu kullanılarak oluşturulan "SBU *FullBody*" (sol taraf) ve "SBU *Moving-Limbs*" (sağ taraf) veri kümelerinin örneklerini göstermektedir

Veri seti oluşturulurken her bireyden yaklaşık 3 saniye uzunluğunda 20 örnekle iki farklı kısa hareket yapması istenmiştir. Her iki hareket de vücudun hem alt hem de üst kısımlarını kapsayacak şekilde yapılmıştır. Hareketlerin yapısal değişimleri Şekil 4.8'de gösterilmektedir.

1. *BLD-SS S-Hareketi*: Denek iki eliyle havada bir S şekli çizer. Bu jest basit olmasına rağmen tüm kullanıcılar tarafından benzer şekilde yapılır. Hareket tanıma için kolay olmasına rağmen, kişi kimliklendirme için daha zor bir harekettir.
2. *BLD-SS Kullanıcı Tanımlı Hareket*: Önceki hareketten farklı olarak, bireyler herhangi bir talimat olmadan kendi hareketlerini seçer. Potansiyel olarak karmaşık olmasına rağmen, bu işlem çoğu kullanıcı için benzersizdir. Bu, veri setini kişi kimliklendirme için kolaylaştırır [35].



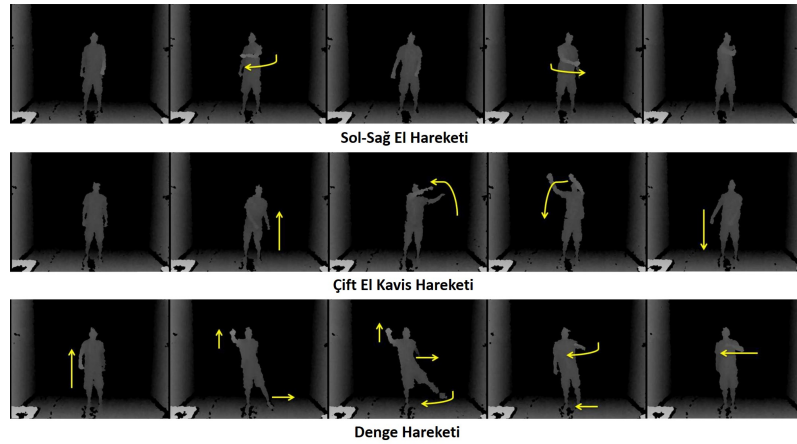
Şekil 4.8 [35]'te gösterildiği gibi BodyLogin Silhouettes and Skeletons veri kümesindeki S-hareketi ve kullanıcı tanımlı hareketlerin kalıpları

İkinci veri seti, 36 farklı birey (18-33 yaş arası 25 erkek ve 11 kadın) tarafından gerçekleştirilen, üç tür hareketi içeren '*Posture, Build and Dynamics (BLD-PBD)*'

veri setidir. Veri seti oluşturulurken, her bireyden yaklaşık 3 saniye uzunluğunda 20 örnekle üç benzersiz kısa hareket yapması istenmiştir. Bu hareketler (sol-sağ hareketi, çift el hareketi ve dengede durma) yapısal olarak hem üst hem de alt vücut hareketlerini içermektedir.

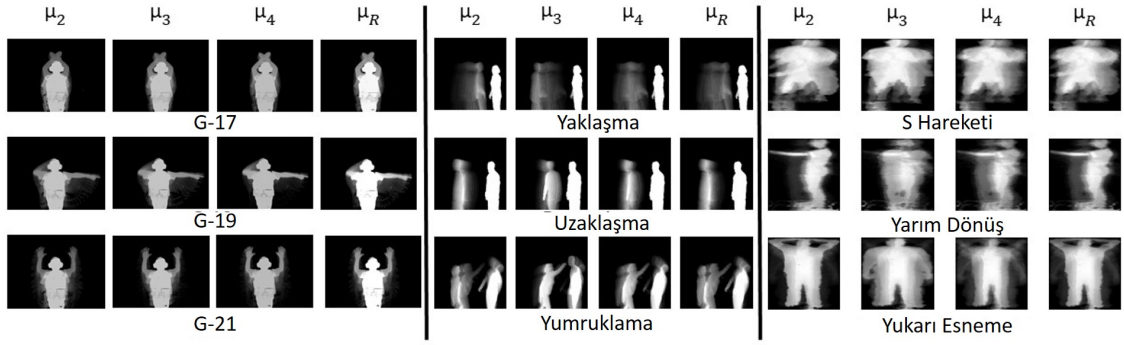
BLD-PBD Sol-Sağ el hareketi: Bireyler sol eli ile sağ omzuna, ardından sağ eli ile sol omzuna dokunurlar. *BLD-PBD Çift elle kavis hareketi:* Bireyler iki eliyle soldan sağa bir yay şekli çizer. *BLD-PBD Denge hareketi:* Bireyler dengede durmaya çalışır ve sağ kolunu öne doğru kaldırırken sol bacağına ileri doğru dairesel çevirir ve sol kolunu geri çeker.

Bu hareketler, yaklaşık 2 metre mesafeden bir Windows Kinect kamerası (Versiyon-1) kullanılarak elde edilmiştir. Kinect kamera SDK'sı kullanılarak, 640 X 480 çözünürlüklü derinlik görüntüleri ve iskelet eklem koordinatları 30 FPS'de kaydedilmiştir [37]. Hareketlerin yapısal değişimi Şekil 4.9'da gösterilmektedir. Bu iki veri kümesindeki bu hareketler, dört hareket içeren yeni bir veri kümesi oluşturmak için birleştirilmiştir. Birleştirilen bu veri setleri kişi kimliklendirme için kullanılmıştır. 40 farklı hareket içerdiğinden dolayı hareket tanıma için yalnızca *BLD-SS Kullanıcı Tanımlı hareket* veri kümesi kullanılmıştır. Önceki veri kümelerinden farklı olarak, veri kümesinin yapısı nedeniyle 'BodyLogin Full-Body' adlı yalnızca bir tür HE veri kümesi oluşturuldu.



Şekil 4.9 [37]'te gösterildiği gibi BodyLogin Posture, Build ve Dynamics veri kümesindeki sol-sağ el hareketi, çift elle kavis hareketi ve denge hareketlerinin yapısal kalıpları sırayla sunulmuştur

Adaptif TT ve önceki çalışmalarda yer alan TT'ler tarafından üretilen enerji görüntülerine ait örnekler Şekil 4.10'de gösterilmiştir. Kullanılan her veri kümesi için üç adet örnek eklenmiştir. Bu örnekler μ_2 , μ_3 , μ_4 ve μ_R fonksiyonları ile üretilmiştir. Bu fonksiyonlardaki farklılıkların HE görüntüleri üzerindeki etkileri bu şekil ile daha iyi incelenebilir ve yöntemler arasındaki farklılıklar daha ayrıntılı bir şekilde anlaşılabilir.



Şekil 4.10 Adaptif TT ve mevcut TT'ler kullanılarak üretilen HE görüntülerinin örnekleri sırasıyla NATOPS, SBU Kinect ve BodyLogin veri kümeleri için sunulmaktadır. Örnekler, soldan sağa μ_2 , μ_3 , μ_4 ve μ_R fonksiyonları kullanılarak elde edilen enerji görüntüleridir. Fonksiyonların çıktıları bu örneklerle daha iyi bir şekilde görülebilmektedir

4.2.2 Sınıflandırıcılar

Önerilen yöntemi ve diğer TT algoritmalarını test etmek için üç sınıflandırıcı seçilmiştir. Birinci ve ikinci sınıflandırıcılar, derin öğrenme tabanlı olan ve öznetelik çıkarımı için literatürde sıklıkla kullanılan VGG-F ve ResNet-18'dir. Üçüncü sınıflandırıcı olarak, farklı kernel tiplerine (RBF, Poly ve Linear) sahip bir SVM ve k değeri sırayla 3, 5 ve 7 olan bir k-NN'den oluşmaktadır.

4.2.2.1 VGG-F Ağı

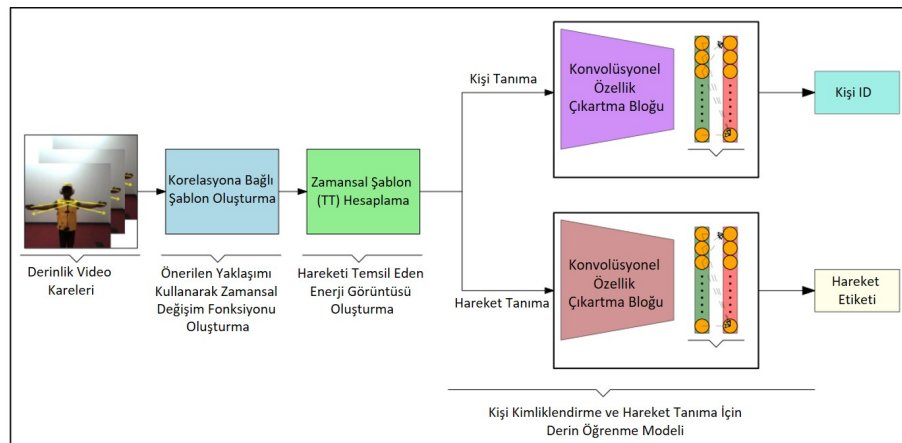
VGG-F ağı, Visual Geometry Group [40] tarafından önerilen üç ağdan biridir. 2014 ImageNet yarışmasında bu ağ, yerelleştirme ve sınıflandırma alanlarında sırasıyla birinci ve ikinci sırayı aldı. VGG-F ağ mimarisi, hızlı ve pratik deneyler gerçekleştirmek için geliştirilmiştir. Birçok çalışma, algoritmalarını geliştirmek için VGG-F ağını bir uçtan uca ağ olarak kullanmıştır [41, 42]. CNN'ler, bu çalışmalarda önerilen yöntemleri test etmek için etkili bir sınıflandırıcı olarak kullanılmaktadır. Önceden eğitilmiş ağların kullanımı özellikle bu tür çalışmalar için tercih edilmektedir. Ancak çalışmamız içerisinde, önceden eğitilmiş bir ağ yerine, [43]'da önerilen Glorot (veya Xavier) rastgele ağırlık ile yeniden eğitime başlatmayı tercih ettik. VGG-F ağı, batch modunda bir geri yayılım algoritması kullanılarak eğitilmiştir. Batch size ve eğitim epoch sayısı, tüm eğitim ağları için sırasıyla 16 ve 50 olarak seçilmiştir. Bu sayılar, önceki yayınlarda uygulanan değerler göz önünde bulundurularak tercih edilmiştir.

4.2.2.2 ResNet-18 Ağı

VGG tipi ağlara kıyasla, ResNet tipi ağlarda bilgi akışı için alternatif yollar bulunmaktadır. Örneğin, bir katmandaki ağırlıklar öğrenmeyi olumsuz etkiliyorsa,

kısayol bağlantıları ile bu katmanlar atlanabilir. Bu şekilde model, kaybolan gradyan problemini en aza indirerek daha derin ağlarda veri öğrenebileceği uygun bir çözüm bulmaya çalışır. ResNet modeli yayınlandığında 5 farklı model sunulmuştu [44]. Bunlar arasında ResNet-18 ve 50, backbone ağları olarak en çok dikkat çeken ve birçok çalışmada tercih edilen ağlar olarak karşımıza çıkmıştır. Bu çalışmada, öznetelikli siluet görüntülerinin sınıflandırmak ve hızlı bir karşılaştırma yapmak için ResNet-18 ağı tercih edilmiştir. ResNet-18 modeli, Glorot ağırlık başlatma yöntemine göre [43] rastgele ağırlıklarla MATLAB ortamında oluşturulmuştur. Önceden eğitilmiş bir model kullanılmamasının nedeni, TT yöntemlerinin yeteneklerini daha adil bir şekilde ortaya koymaktır. ResNet modelinde görüntü girdisi olarak $224 \times 224 \times 3$ boyutu kullanılmaktadır. TT siluet görüntüleri tek kanal olduğu için bu görüntüler, R, G, B kanallarına gri seviyeli olarak ayrı ayrı yerleştirilmiştir. Bu sayede 3 derinlikli bir görüntüye dönüştürülmüştür. Ağın son nöron katmanı olan Logit katmanı, problemimizdeki sınıf sayısına karşılık gelecek şekilde yeni bir Fully-Connected katmanı ile değiştirilmiştir. Daha sonra model, her bir problem için %80, %50 ve %20 eğitim oranları için rastgele sıralanmış 5 farklı küme üzerinde eğitilmiş ve test edilmiştir. Bu 5 sonucun ortalaması alınarak TT yöntemleri arasında başarımların kıyaslaması yapılmıştır. Hiperparametre analizi için 48 minibatch'den oluşan otuz epoch seçilmiştir. Loss fonction olarak ise cross-entropi kullanılmıştır.

CNN sınıflandırıcı kullanımının bir özeti Şekil 4.11'de verilmiştir. Sınıflandırmalar bu sistematik yaklaşımla yapılmaktadır. Ağlar, kişi kimliklendirme (Person Identificaiton-PI) ve hareket tanıma (Motion Recognition-MR) görevleri için ayrı ayrı eğitilmiştir.



Şekil 4.11 TT tarafından üretilen enerji görüntülerinin CNN ağlarında sınıflandırılması için genel çerçeve. VGG-F ve ResNet-18 ağları probleme göre revize edilmiş ve yeniden eğitilmiştir. CNN'lerin FC katmanları, sınıf sayısına göre değiştirilmiştir

4.3 Deneyisel Sonuçlar

Bu bölümde, önceki TT fonksiyonları ve önermiş olduğumuz adaptif TT fonksiyonu arasında hem hareket tanıma hem de kişi tanımlama için deneyisel sonuç karşılaştırması yapılacaktır. Araştırmacılar için erişime açık olan NATOPS Gesture, SBU Kinect Interaction ve BodyLogin veri kümelerini bu kıyaslama için kullanılmıştır. Kullanılan veri kümeleri, bir Microsoft Kinect derinlik sensörü kullanılarak yakalanan RGB-D videoları ve görüntüleri içermektedir. Kinect sensörü tarafından yakalanan derinlik bilgisi düşük doğrulukta olduğundan bu kıyaslama kapsamında göre göz ardı edilerek, derinlik video çerçeveleri maskelenerek binary hale getirdi. Bu şekilde hesaplanan HE görüntüleri, bir silüete benzer şekilde bireylerin zamansal konumlarına da sahiptir.

Şekil 4.1’de verilen üyelik fonksiyonlarından ilki olan μ_1 , Bobick ve Davis tarafından [14]’de bir enerji görüntüsü oluşturmak için önerilmiş ve sonraki çalışmalarda farklı yöntemlerin de temelini oluşturmuştur. [24]’de, Ijjina ve ark. bir HE görüntüsü oluşturmak için belirli zamansal bölgelere vurgu yaparak μ_1 yöntemine alternatif sağlayan μ_2 , μ_3 ve μ_4 olmak üzere üç yeni üyelik fonksiyonu önermiştir. Bu fonksiyonlarla elde edilen HE görüntüleri, μ_1 fonksiyonu ile elde edilen HE görüntülerine kıyasla daha etkili başarımlar göstermiştir. Ayrıca μ_2 , μ_3 ve μ_4 ’ün μ_1 ’e göre avantajları detaylı olarak gösterilmiştir. Ancak bu çalışmada bizim dikkatimizi çeken kısım, bu üç fonksiyonun başarımları, hareket türü ve veri kümesine göre tamamen deneyisel olarak elde edilmiştir. Yani başka bir ifade ile bu fonksiyonlardan herhangi biri bir hareket veya veri kümesinde iyi sonuç verirken, diğerleri μ_1 yönteminden daha düşük başarımlar elde edebilmektedir. Bu sebeple tüm yöntemlerin füzyonu uygulanmaktadır. Çalışmamızda ise, μ_R fonksiyonunu kullanarak, bu üç fonksiyonun deneyisel seçim zorluklarını ortadan kaldıracak ve HE görüntüsünün oluşturulmasında hareketlerdeki değişiklikleri daha iyi temsil edebilecek bir yöntem önerdik. Bu bağlamda kıyaslama tablosunda μ_1 ile μ_R arasında bir karşılaştırma yapmanın gerekli olmadığını düşünmekteyiz.

Tüm veri setlerinin değerlendirilmesi için 2 ve 5 katlı çapraz doğrulama yöntemleri kullanıldı. 2 katlı çapraz doğrulamada, veriler rastgele iki eşit parçaya bölündü: bir grup eğitim için, diğer grup test için kullanıldı (%50 test). 5 kat çapraz doğrulamada veriler rastgele beş farklı gruba ayrıldı. İlk değerlendirmede dört grup eğitim için, bir grup test için kullanıldı (%20 test). İkinci değerlendirmede bir grup eğitim için, dört grup test için kullanıldı (%80 test). Bu işlemler beş kez tekrarlanmış ve tüm performansların ortalaması alınmıştır. Üretilen tüm HE görüntüleri, kişi kimliklendirme ve hareket tanınması için hem kolay hem de zorlu durumlarda değerlendirilmiştir. Bu bölüm, üç değerlendirme prosedürü olarak

düzenlenmiştir. Birinci bölümde, önerilen yöntemle elde edilen sonuçlar önceki çalışmalarla karşılaştırılmıştır. Bu sayede önerilen yöntem ile literatürdeki çalışmalar arasındaki performans ortaya konmuştur. İkinci bölümde, mevcut ve önerilen TT fonksiyonları aynı CNN ağlarına uygulanmıştır. Bu şekilde her bir TT fonksiyonunun aynı koşullar altında performansı karşılaştırılmış ve önerilen yöntemin katkısı gözlemlenmiştir. Üçüncü bölümde, birinci ve ikinci bölümlerdeki başarının önerilen yöntemin sağlamlığından mı yoksa CNN sınıflandırıcılarının gücünden mi geldiğini anlamak için farklı ve hazırlanmış yöntemler kullanılmıştır. Böylece farklı öznetelik çıkarıcılar ve sınıflandırıcılar ile TT fonksiyonlarının performansları karşılaştırılmıştır. Bu değerlendirmeler sayesinde daha geniş bir eğitim-test ortamı oluşturulmuş ve önerilen yöntemin katkısı incelenmiştir.

4.3.1 Mevcut Sonuçlarla Karşılaştırma

Bu bölümde CNN ağlarından elde edilen eğitim sonuçları literatürde bildirilen diğer son teknoloji yöntemlerle karşılaştırılmıştır. Daha önceki çalışmalarda sunulan sonuçlar genellikle 5 kat çapraz doğrulama (%20 test) sonuçları olduğundan karşılaştırmalar bu sonuçlara göre yapılmıştır. Tablo 2, NATOPS veri kümesi kullanılarak yapılan hareket tanıma sonuçları ile önerilen yöntemin performans karşılaştırmalarını göstermektedir. Önerilen yaklaşım ile hareket tanıma için, VGG-F ağı kullanılarak %99,58 ve ResNet-18 ağı kullanılarak %96,49 doğruluk elde edilmiştir. Mevcut çalışmalarla karşılaştırıldığında, aynı veri kümesi üzerinden farklı veriler ve yöntemler kullanılarak elde edilen hareket tanıma performansı önerdiğimiz yöntemle çok yüksek başarı elde etmiştir. Yüksek performansın temel nedenleri şu şekilde sıralanabilir: kullanılan CNN ağının özellik çıkarma kabiliyetinin yüksek olması, kullanılan veri seti özelliklerinin daha basit olması ve karmaşık RGB-D verilerinden ziyade silüetteki değişikliklere odaklanılması ve adaptif TT yöntemi sebebiyle hareket değişimlerinin daha fazla ön plana alınmış olması, olarak değerlendirilebilir. Tablo 4.2'deki karşılaştırmalar farklı veri türlerini içermektedir. Bazı çalışmalar iskelet ve el özelliklerini kullanırken, diğerleri RGB-D özelliklerini kullanır. Biz de çalışmamızda maskeli derinlik özelliklerini kullandık. Mevcut yöntemler ve önerilen yaklaşım için farklı özellikler kullanılsa da, tüm veriler NATOPS veri setinden elde edildiğinden çalışmalar arasında karşılaştırmalar yapılabilir.

SBU Kinect veri setinden Adaptif TT yöntemiyle elde edilen HE görüntüleri ile önceki çalışmalarda elde edilen başarımların sonuçları Tablo 4.3'te karşılaştırılmalı olarak gösterilmiştir. Sonuçlar, önerilen yaklaşımın %96,56 doğruluk elde ettiğini ve önceki çalışmalara göre daha iyi sonuçlar verdiğini göstermektedir. NATOPS veri kümesinde olduğu gibi SBU Kinect veri kümesinde de farklı türde veriler kullanılmış olsa da,

Tablo 4.2 NATOPS veri seti kullanılarak hareket tanıma konusunda önerilen ve mevcut yaklaşımların performans sonuçları (% olarak sınıflandırma doğruluğu açısından). Bu tablo, daha önceki çalışmalarda kullanılan yöntem ve özelliklerle elde edilen doğruluk oranlarını göstermektedir. Mevcut yöntemler ve önerilen yaklaşım için farklı özellikler kullanılmasına rağmen tüm veriler NATOPS veri setinden elde edilmiştir

Yaklaşım-Yöntem	Kullanılan Özellikler	Başarım %
Song et al.[45]	Skeletal & hand	75.37
CRF, Song et al. [46]	Skeletal & hand	53.30
HMM, Song et al. [46]	Skeletal & hand	77.67
HCRF, Song et al. [46]	Skeletal & hand	78.00
Couples HCRF, Song et al. [46]	Skeletal & hand	86.00
Linked HCRF, Song et al. [46]	Skeletal & hand	87.00
TTs Fusion Top-1, Ijjina et al. [24]	RGB-D raw video streams	72.58
TTs Fusion Top-2, Ijjina et al. [24]	RGB-D raw video streams	86.58
2D-MRCN, Elbouchaki et al. [47]	RGB-D video streams	89.83
3D-CDCN, Elbouchaki et al. [47]	RGB-D video streams	94.54
MultiD-CNN, Elbouchaki et al. [47]	RGB-D video streams	95.87
Önerilen Yaklaşım (ResNet-18 ağı)	Masked depth video streams	96.49
Önerilen Yaklaşım (VGG-F ağı)	Masked depth video streams	99.58

hepsi aynı kaynaktan alındığı için karşılaştırılabilir niteliktedir. Tablo 4.3'te, Khairi ve ark. [48], VGG-16 ağında biraz daha iyi sonuçlar elde etmiş olduğu görülmektedir. Ancak [48]'de önerilen CNN akışlarını birleştirme yöntemi, [24]'de olduğu gibi birden fazla ağı eğitimi ve bu ağların kaynaşmasını gerektirir. Bu kaçınılmaz olarak süreç maliyetlerini artırır. Yine RGB+DMI max fusion yöntemi önerdiğimiz yöntemle kıyasla daha yüksek başarımlar elde etmiştir. Ancak bu iki yöntemin ortaya atılma zamanları düşünüldüğünde önerdiğimiz yöntem Singh'in yönteminden çok daha önce oluşturulduğu görülmektedir.

Önerdiğimiz yöntem, daha düşük işlem maliyetleri içermekte ve SBU Kinect verileriyle yapılan karşılaştırmalarda önemli ölçüde daha başarılı sonuçlar elde etmiştir. BodyLogin veri seti ile hareket tanıma ve kişi kimliklendirme konusunda özel bir çalışma olmadığı için önceki çalışmalarla karşılaştırma yapılamamıştır. BodyLogin sonuçları 2. ve 3. karşılaştırma başlıklarında mevcut TT'ler ile önerdiğimiz adaptif TT'nin karşılaştırılması tablolarında sunulmuştur. Önerilen yöntem her ne kadar önceki çalışmalardan daha iyi sonuçlar verse de, aşağıdaki soruları gündeme getirmiştir. Adaptif TT'nin mevcut TT'lere kıyasla hareketi vurgulama yeteneği ne kadar daha iyiydi? Mevcut TT'ler aynı ağlara uygulansaydı sonuç ne olurdu? Bu soruların cevabı incelenebilirse, adaptif TT'nin diğer TT'lere göre başarısı daha adil bir şekilde karşılaştırılmış olacaktır. Bu amaçla, tüm fonksiyonlar (önceden kullanılan μ_2 - μ_3 - μ_4 ve önerilen μ_R HE görüntüleri üretmek için kullanılmış ve aynı CNN ağlarında eğitilmiştir. Sonuçlar bir sonraki başlıkta gösterilmiş ve değerlendirilmiştir.

Tablo 4.3 SBU veri seti kullanılarak hareket tanıma konusunda önerilen ve mevcut yaklaşımların performans sonuçları (% olarak sınıflandırma doğruluğu açısından). Bu tabloda, önceki çalışmalarda kullanılan yaklaşımların ve özelliklerin performansı önerilen yöntem ile karşılaştırılmıştır. Mevcut yöntemler ve önerilen yaklaşım için farklı özellikler kullanılmasına rağmen tüm veriler SBU Kinect veri setinden elde edilmiştir

Yaklaşım-Yöntem	Kullanılan Özellikler	Başarım %
Raw skeleton, Yun et al.[39]	Skeletal Data	49.70
Raw skeleton, Ji et al. [49]	Skeletal Data	79.40
Joint features, Yun et al. [39]	Skeletal Data	80.30
Joint features, Ji et al. [49]	Skeletal Data	86.90
Hierarchical RNN, Du et al. [50]	Skeletal Data	80.35
Cluster analysis of pose, Edwards et al. [51]	Skeletal Data	83.90
Deep LSTM, Zhu et al. [52]	Skeletal Data	86.03
Generative topic model, Huynh-The et al. [53]	Skeletal Data	90.30
STA-LSTM, Song et al. [54]	Skeletal Data	91.51
ST-LSTM + Trust Gate, Liu et al. [55]	Skeletal Data	93.30
Radius-margin bound, Lin et al. [56]	Skeletal Data	93.40
TTs Fusion of {begin and middle}, Iijina et al. [24]	RGB-D Raw Video Stream	90.98
Multi-Task Learning Network, Ke et al. [57]	Concatenated Skeleton Sequences	93.57
Non-Linear Mapping (ConvNet + ELM classifier), Gharahdaghi et al.[58]	Depth Features	78.00
Non-Linear Mapping (ConvNet + ELM classifier), Gharahdaghi et al.[58]	RGB Features	81.50
Deeply Coupled ConvNet (Inception-v3 + LSTM), Singh et al. [59]	RGB Features	76.60
Deeply Coupled ConvNet (Inception-v3 + LSTM), Singh et al. [59]	Dynamic Motion Image Features	91.40
Deeply Coupled ConvNet (Inception-v3 + LSTM), Singh et al. [59]	RGB + DMI Max Fusion	98.70
Combining CNN streams (VGG-F), Khaire et al.[48]	RGB + Depth + Skeletal Data	96.26
Combining CNN streams (VGG-16), Khaire et al.[48]	RGB + Depth + Skeletal Data	96.67
Önerilen Yaklaşım (ResNet-18 ağı)	Depth video stream	81.25
Önerilen Yaklaşım (VGG-F ağı)	Depth video stream	96.56

4.3.2 TT'lerin CNN Tabanlı Karşılaştırması

Bu bölümde, tüm TT fonksiyonlarından elde edilen HE görüntüleri VGG-F ve Resnet-18 ağı ile eğitilmiş ve sadece hareket tanıma için değil aynı zamanda kişi tanıma için de kıyaslamaya alınmıştır. Çalışmamızda veri setleri ile kişi kimliklendirme üzerine çalışılmış olması bu veri setleri için literatürde bir ilk olmuştur. Hareket tanıma olduğu gibi kişi tanıma içinde geniş eğitim-test oranları hazırlanmıştır. Hareket tanıma için tüm bireylerin hareketleri aynı şekilde etiketlenirken, her bireyin hareketleri kişi tanımlaması için ayrı ayrı etiketlenmiştir. Yani, tüm bireylerin bir hareketine ait örnekler hareket tanıma için aynı etikete sahiptir; Öte yandan, doğru kimliklendirme için eylemi gerçekleştiren kişiye göre farklı etiketlere sahiptir. NATOPS veri tabanı kullanılarak oluşturulan iki veri kümesi için performans karşılaştırması Tablo 4.4 ve Tablo 4.5'te verilmiştir. Mevcut fonksiyonlar hareket tanıma için VGG-F ağına daha iyi sınıflandırma başarımına ulaşmıştır. ResNet-18 ağına ise, ortalama performans genellikle VGG-F'den daha düşük olmasına rağmen, önerilen adaptif TT diğer yöntemlerden daha iyi performans göstermiştir. Ek olarak, kişi tanıma elde edilen sonuçlarda ise, her iki CNN ağındaki NATOPS *Full-Body* veri setinde mevcut fonksiyonlar daha iyi performans göstermiştir. Daha önceki çalışmalar dikkate alındığında, NATOPS veri tabanı ile kişi sınıflandırması konusunda herhangi bir çalışma yapılmadığı görülmektedir. Bu nedenle önerilen yöntem yalnızca mevcut fonksiyonlar ile karşılaştırılabilmiştir.

Tablo 4.4 CNN ağlarını kullanan NATOPS *Full-Body* veri setinde önerilen adaptif TT ve mevcut TT fonksiyonlarının başarı oranlarının karşılaştırılması. Her bir CNN ağı için hareket tanıma ve kişi kimliklendirme sonuçları tabloda ayrı ayrı sunulmaktadır. Kalın işaretli sonuçlar en iyi ortalama performansı gösterir

Zamansal Fonk.	VGG-F						ResNet-18					
	Hareket Tanıma			Kişi Kimliklendirme			Hareket Tanıma			Kişi Kimliklendirme		
Test Oranı(%)	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%
μ_2	99.81	99.43	97.97	99.99	100	99.96	96.26	93.95	89.46	100	99.96	99.92
μ_3	99.75	99.40	98.02	100	100	99.99	96.17	92.97	87.31	99.95	99.98	99.96
μ_4	99.59	99.30	97.56	100	99.99	99.88	95.23	91.16	86.01	99.95	99.90	99.71
μ_R	99.58	99.11	97.94	100	100	100	96.49	94.10	88.30	100	100	99.95

Tablo 4.5 CNN ağlarını kullanan NATOPS *Moving-Limbs* veri kümesinde önerilen adaptif TT ve mevcut TT fonksiyonlarının başarı oranlarının karşılaştırması. Her bir CNN ağı için hareket tanıma ve kişi kimliklendirme sonuçları tabloda ayrı ayrı sunulmaktadır. Kalın işaretli sonuçlar en iyi ortalama performansı gösterir

Zamansal Fonk.	VGG-F						ResNet-18					
	Hareket Tanıma			Kişi Kimliklendirme			Hareket Tanıma			Kişi Kimliklendirme		
Test Oranı(%)	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%
μ_2	99.19	98.87	96.51	98.80	98.06	94.84	91.76	88.66	88.70	97.92	95.98	91.33
μ_3	99.16	98.58	96.34	98.96	98.13	95.40	92.24	88.73	89.11	97.60	96.77	91.88
μ_4	97.99	97.06	93.79	97.72	96.16	90.68	83.57	81.15	81.97	93.13	89.83	78.54
μ_R	98.73	97.90	95.14	98.76	97.55	94.02	87.60	83.76	85.38	95.63	93.10	84.65

Tablo 4.5 incelendiğinde, NATOPS *Moving-Limbs* veri seti üzerinde VGG-F ağındaki μ_2 ve ResNet-18 ağındaki μ_3 fonksiyonlarının hareket tanımada daha başarılı olduğu görülmektedir. Önerdiğimiz μ_R fonksiyonu, her iki ağıdaki tüm TT'ler arasında ortalama bir başarıya sahipken, μ_4 hareket tanımada en kötü sonucu vermiştir. Ayrıca kişi kimliklendirme için sonuçların benzer olduğu görülmektedir. NATOPS veri tabanından oluşturulan ve CNN ağlarına uygulanan veri setleri arasında *Full-Body* veri seti üzerinde en iyi sonuç μ_R fonksiyonundan elde edilmiştir. *Moving-Limbs* veri setinde μ_R ortalama sonuçlara ulaşırken en iyi sonuçlar μ_3 fonksiyonundan elde edilmiştir.

NATOPS *Moving-Limbs* veri setinde, insanlar sabitken kol ve bacaklarını hareket ettirdikleri için vücut silueti bilgisi kaldırılmıştır. Mevcut TT'lerde, özellikle μ_2 ve μ_3 fonksiyonlarında, herhangi bir örnekteki vücut siluetinin vurgusu ve ortalaması eşittir. Yani, 100 karelik bir video akışındaki sabit bir gövde silueti için fonksiyonlar aynı ortalama alacaktır. Bu iki fonksiyonun hareket değişikliğinden bağımsız olarak sabit bir uzunluğu vardır. Bu, vücut silueti bilgisinin fonksiyonlar arasında ayrım özelliğini azaltacaktır. Ancak, μ_R fonksiyonunda, gövde silueti hareketli karelerin sayısı ile orantılı olarak vurgulanacaktır. Bu durumda siluet harekete göre değişen değerlere sahip olacaktır. Tablo 4.5'teki sonuçlar incelendiğinde en iyi sonuçların μ_2 ve μ_3 olduğu ve bu değerlerin birbirine çok yakın olduğu görülmektedir. Ancak, μ_4 ve μ_R fonksiyonlarının performansı daha düşüktür. Bu durum, μ_R fonksiyonunun hareket değişiklikleri nedeniyle siluet bilgisinden daha fazla yararlandığını ortaya koymaktadır. μ_R performansı, vücut silueti bilgisinin kaldırılmasıyla azalmış olabilir.

Dolayısıyla μ_R 'nin siluet gövdesindeki değişiklikleri de vurguladığı ve μ_R 'nin *Full-Body* veri setindekiler gibi gövdenin yer aldığı derinlik görüntülerinde daha iyi sonuçlar verdiği söylenebilir.

SBU Kinect veri tabanındaki yer alan yaklaşma ve uzaklaşma hareketleri gibi bazı örneklerde, bireylerden biri hareketsiz veya çok az hareketli bir durumda iken, diğer birey çok daha aktif hareket etmektedir. Bireylerin birbirlerine karşı pozisyonları değişebilir, ancak etkileşimler aynı kaldığından, HE değerlendirmek için yatay olarak aynalanmış etkileşimler de kullanılabilir. Böylece hareket tanıma için örnek sayısı iki katına çıkarılabilir. Ancak bu durum veri setinin kişi kimliklendirme için kullanımını kısıtlamaktadır. Bu nedenle, SBU Kinect veri seti yalnızca hareket tanıma için kullanılmıştır.

Tablo 4.6'da, önerdiğimiz yöntem ve mevcut TT fonksiyonları ile elde edilen sonuçların karşılaştırması gösterilmektedir. Daha önce Tablo 4.3'te, SBU Kinect veri tabanını kullanarak mevcut TT'lerin ve adaptif TT'nin performansının bir karşılaştırmasını sunmuştuk. Ancak Tablo 4.3'teki sonuçlar mevcut TT'lerin füzyonu ile elde edilen sonuçları içermektedir. Tablo 4.6'da ise her bir TT'nin bireysel performansı sunulmaktadır. [24]'de sunulan mevcut TT sonuçları ilk satırda verilmiştir. Sonraki satırlarda tüm TT'lerden elde edilen HE görüntülerine ait performans sonuçları sırayla eklenmiştir. Tüm TT'ler aynı parametreler ile başlatılarak CNN ağlarına uygulanmıştır. Böylece adaptif TT ile mevcut TT fonksiyonları arasında eşit koşullar altında bir karşılaştırma yapılmıştır. Tüm TT'lerin performansı CNN ağlarına bağlı olarak değişirken, adaptif TT'nin performansı tüm karşılaştırmalarda en iyi sonucu vermiştir. Sınıflandırıcılar arasındaki en iyi sonucun ise VGG-F ağından elde edildiği görülmektedir.

Tablo 4.6 SBU Kinect veri kümelerinde mevcut TT fonksiyonlarının ve önerilen adaptif TT'nin, aynı ağlar üzerinden bireysel başarı oranlarının karşılaştırması. Tablonun ilk satırı, referans çalışmadaki sonuçları içerir. Diğer satırlarda bu çalışma kapsamında yapılan karşılaştırma sonuçları yer almaktadır

Veri Seti	Kullanılan Özellikler	Sınıflayıcı modeli	Test Oranı	5-Katlı Çapraz Val. Ort. Skoru			
				μ_2	μ_3	μ_4	μ_R
SBU in [24]	RGB-D data	ConvNet with ELM Classifier	20%	75.58	85.06	79.85	-
SBU <i>Full Body</i>	Maskelenmiş Derinlik Verisi	VGG-F	20%	96.25	96.41	95.94	96.56
		ResNet-18	20%	79.38	73.75	70.63	81.25
SBU <i>Moving Limbs</i>	Maskelenmiş Derinlik Verisi	VGG-F	20%	95.31	94.22	94.06	95.94
		ResNet-18	20%	71.88	68.75	67.50	79.38

Tablo 4.7'de, SBU Kinect veri tabanından elde edilen 2 farklı veri seti kullanılarak tüm TT'lerin karşılaştırması sunulmaktadır. Her test grubu, her fonksiyon için aynı eğitim ve test sırası ile rastgele beş kez eğitildi ve ortalama performansları karşılaştırıldı. Önerdiğimiz yöntem genellikle daha iyi sonuçlar vermiştir, ancak verilerin %80'i ile test edilen zorlu grupta en iyi sonuç μ_2 ile elde edilmiştir. Bu durum, mevcut

fonksiyonların füzyonunun daha başarılı olduğu gösterse de, hareket tanıma için birden fazla fonksiyonun kullanılması uzun süreli olacak ve işlem yükü artacaktır. Önerdiğimiz yaklaşım bu sorunlara daha optimum bir çözüm sunmaktadır.

Tablo 4.7'nin son sütununda değer *N/A* (Not Applicable) olarak görülmektedir. Bu, elde edilen sonuçların değerlendirme için geçerli olmadığını göstermektedir. Bu nedenle son sütuna ait değerler tabloya dahil edilmemiştir. NATOPS *Moving-Limbs* veri kümesi sonuçlarının aksine, adaptif TT, SBU *Moving-Limbs* veri kümesinde en iyi sonuçlar vermiştir. Bunun nedeni, daha önce bahsedilen vücut silüetlerinin SBU *Moving-Limbs* veri setinde tamamen kaybolmamasıdır. SBU Kinect veritabanı, NATOPS'dan farklı olarak, iki kişi arasındaki etkileşim hareketlerini içerir, dolayısıyla kişilerin tüm bedenleri hareket halindedir. Tamamen kaybolmayan vücut silüetleri ise adaptif TT'nin doğruluk performansını etkilemiş olabilir.

Tablo 4.7 CNN ağlarını kullanan SBU Full-Body ve SBU *Moving-Limbs* veri kümelerinde adaptif TT ve mevcut TT fonksiyonlarının bireysel başarı oranlarının karşılaştırması. Her bir veri seti ve CNN ağı için hareket tanıma sonuçları sırasıyla tabloda sunulmaktadır. Koyu vurgulanmış sonuçlar en iyi ortalama başarıyı göstermektedir

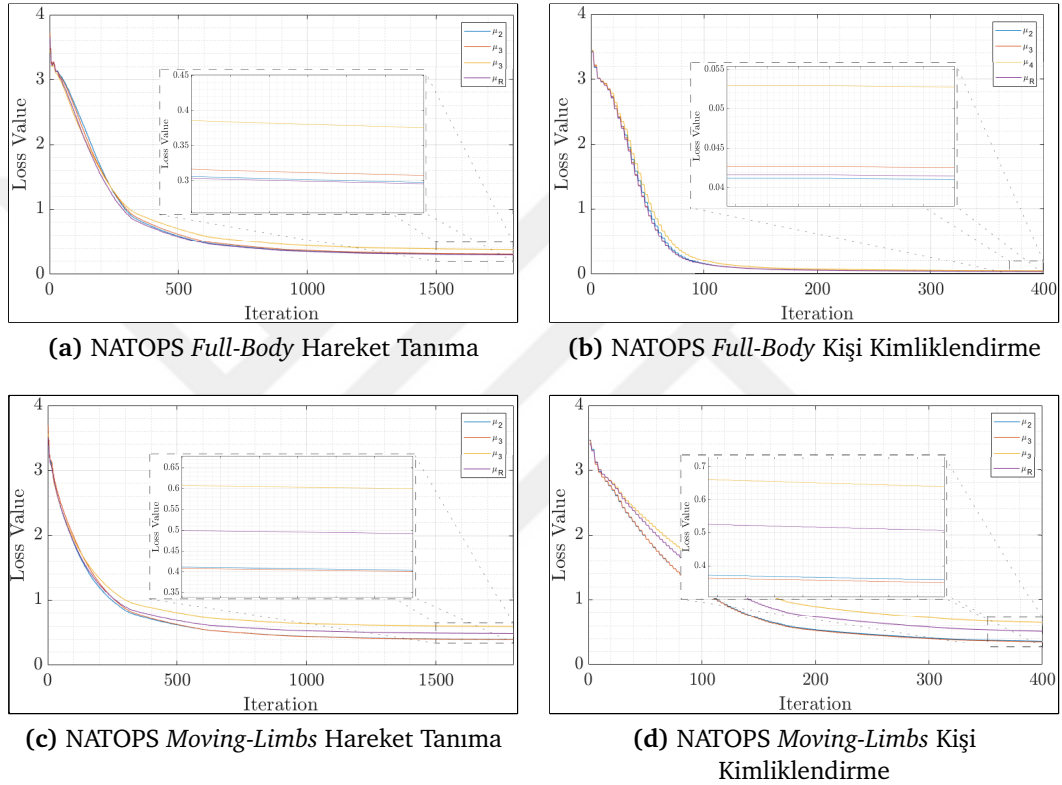
Zamansal Fonk.	VGG-F						ResNet-18					
	SBU Full Body			SBU Moving Limbs			SBU Full Body			SBU Moving Limbs		
Test Oranı(%)	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%
μ_2	96.25	85.25	65.31	95.31	79.25	58.28	79.38	67.75	49.06	71.88	66.25	N/A
μ_3	96.41	85.50	65.16	94.22	70.50	53.44	73.75	65.25	45.78	68.75	55.75	N/A
μ_4	95.94	75.50	60.78	94.06	74.25	55.31	70.63	63.25	49.53	67.50	58.75	N/A
μ_R	96.56	86.00	64.69	95.94	79.25	57.66	81.25	69.00	54.69	79.38	66.25	N/A

VGG-F ve ResNet-18 sınıflandırıcıları, önceki veri kümelerinde olduğu gibi BodyLogin veri kümesi için de eğitilmiştir. Üretilen enerji görüntüleri kullanılarak hem kolay hem de zorlu durumları değerlendirmek için hareket tanıma ve kimliklendirme için %20 test, %50 test ve %80 test işlemleri uygulanmıştır. Eğitim beş kez tekrarlanmış ve ortalama puanlar Tablo 4.8'de sunulmuştur. Kişi tanımlaması için BodyLogin Full-Body veri seti kullanılmıştır. Tablo 4.8'de her iki CNN ağında adaptif TT'nin en iyi performansı %20 ve %50 testte elde ettiği görülmektedir. %80 test işleminde VGG-F ağındaki μ_2 işlevi ve ResNet-18 ağındaki μ_4 işlevi en iyi sonuçları elde etmiştir.

Tablo 4.8 CNN ağlarını kullanan BodyLogin veri kümelerinde adaptif TT ve mevcut TT fonksiyonlarının bireysel başarı oranlarının karşılaştırması. Hareket tanıma ve kişi kimliklendirme sonuçları, sırasıyla her bir CNN ağı için tabloda sunulmaktadır. Koyu vurgulanmış sonuçlar en iyi ortalama başarıyı göstermektedir

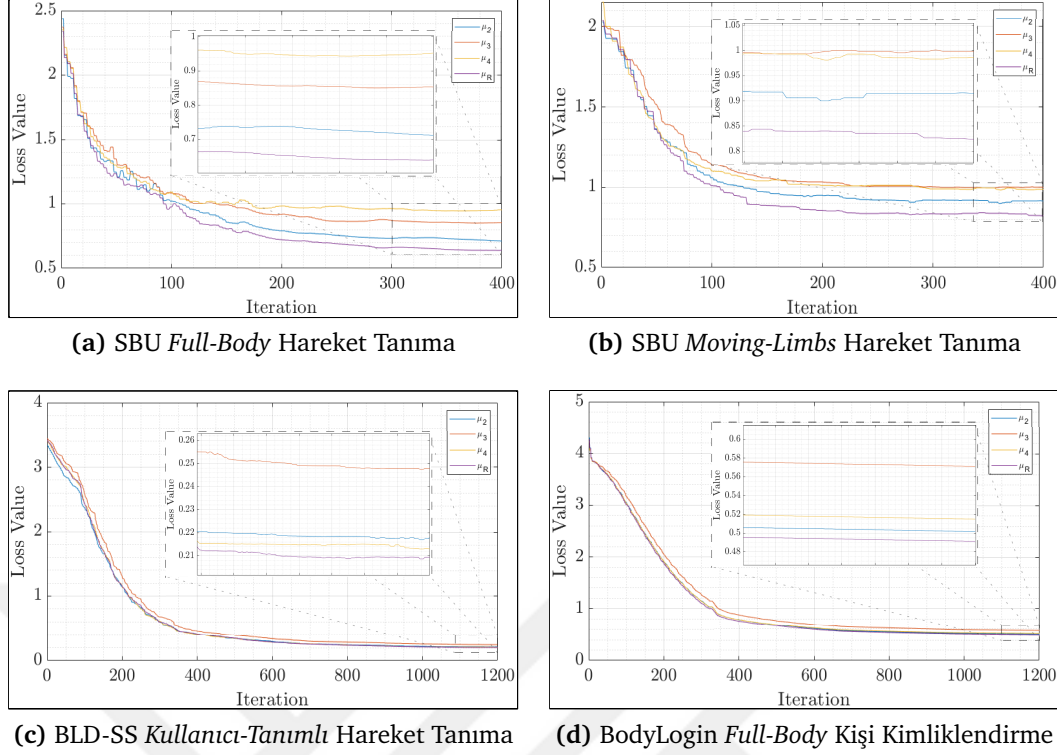
Zamansal Fonk.	VGG-F						ResNet-18					
	Kişi Kimliklendirme BodyLogin Full body			Hareket Tanıma BLD-SS User-Defined			Kişi Kimliklendirme BodyLogin Full body			Hareket Tanıma BLD-SS User-Defined		
Test Oranı(%)	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%
μ_2	99.63	98.75	94.25	98.87	97.29	85.74	97.38	95.90	85.97	94.25	87.63	68.22
μ_3	99.00	98.05	91.63	98.41	96.02	83.52	96.88	94.40	81.19	91.63	85.45	64.67
μ_4	99.25	98.55	94.00	98.72	96.79	85.20	98.13	95.66	86.22	93.63	86.16	66.27
μ_R	99.63	98.80	93.81	99.12	97.28	86.13	98.13	96.05	85.72	94.88	88.28	67.99

Hareket tanıma için kullanılan BLD-SS Kullanıcı Tanımlı veri setinin sonuçları incelendiğinde önerilen yöntemin diğer fonksiyonlara göre daha başarılı olduğu görülmektedir. μ_R işlevi, VGG-F ağında %20 ve %80 test işlemlerinde ve ResNet-18 ağında %20 ve %50 test işlemlerinde en iyi performansı elde etmiştir. μ_2 işlevi, BLD-SS Kullanıcı Tanımlı veri kümesinde VGG-F ağı kullanılarak %50 test işlemi ve ResNet-18 ağı kullanılarak %80 test işleminde en iyi sonuçları elde etmiştir. Bu tablolara ek olarak ResNet-18 ağındaki eğitimin validasyon kayıp eğrileri Şekil 4.11 ve Şekil 4.12’de sunulmuştur.



Şekil 4.12 ResNet-18 ağının Validation-Loss grafikleri. Sırasıyla μ_2 , μ_3 , μ_4 ve μ_R fonksiyonlarının Validation-Loss değışiklikleri karşılaştırması. Son iterasyonlarda eğriler alt şekil içinde büyütülmüştür. Alt şekil kutusunda farklılıklar daha net bir şekilde gösterilir. a) Hareket tanıma için NATOPS Full-Body. b) Kişi kimliklendirme için NATOPS Full-Body. c) Hareket tanıma için NATOPS Moving-Limbs. d) Kişi kimliklendirme için NATOPS Moving-Limbs

Eğriler arasındaki farkları daha iyi görebilmek için son kısımlar büyütülmüş ve alt şekillerde gösterilmiştir. Grafiklerdeki bazı eğitimlerde iterasyonların farklı olduğu görülmektedir. Bunun sebebi, bazı eğitimlerde sınıf başına düşen görüntü sayısının batch size oranıdır. Bu oran daha yüksek olan eğitimlerde, yinleme sayısı daha yüksek seçilmiştir. Bir veri kümesi için sınıf sayısı daha fazla olmasına rağmen, sınıf başına örnek sayısının az olması ve aynı batch size’da eğitim yapılması nedeniyle farklı yinleme değerleri ortaya çıkmıştır. BodyLogin veri setlerinden elde edilen



Şekil 4.13 ResNet-18 ağına ait Validation-Loss grafikleri. Sırasıyla μ_2 , μ_3 , μ_4 ve μ_R fonksiyonlarının Validation-Loss değışiklikleri karşılaştırması. Son iterasyonlarda eğriler alt şekil içinde büyütölmüştür. Alt şekil kutusunda farklılıklar daha net bir şekilde gösterilir. a) Hareket Tanıma için SBU Full-Body. b) Hareket Tanıma için SBU Moving-Limbs. c) Hareket tanıma için BLD-SS Kullanıcı-Tanımlı. d) Kişı kimliklendirme için BodyLogin Full-Body

sonuçlara göre adaptif TT, iki farklı CNN ağında eğitilen tüm TT'ler arasında daha yüksek performansı göstermiştir. Bu sonuçlar, μ_R fonksiyonunun hareketleri temsil etmede daha iyi olduğunu kanıtlamıştır. Sonuçlar μ_R 'ın başarısını gösterse de, bir başka soru da, CNN ağlarının μ_R işlevinin performansı üzerindeki etkisinin ne olduğudur? Yani CNN ağları dışında bir yöntem foknsiyonun başarımı üzerinde değışik bir sonuca neden olur mu? Bu sonuçlara ek olarak Hand-Crafted öznitelik çıkarıcılar ve sınıflandırıcılar kullanılarak elde edilen başarının CNN ağının genelleme kabiliyetine mi yoksa önerilen yönteme mi bağılı olduğunu bir sonraki başlık altında incelenmiştir.

4.3.3 Hand-Crafted Tabanlı Karşılaştırma

Bu bölümde önerilen yöntem ve mevcut yöntemler kullanılarak üretilen HE görüntülerinden temel bileşen analizi (PCA), yerel ikili modeller (LBP) [60] ve gradyan histogramı (HoG) [61] yöntemleri ile özellik çıkarımı yapılmıştır. Bu özellikler daha sonra derin sinir ağı tabanlı çok katmanlı bir algılayıcı, k-en yakın

komşu (k-NN) ve karar destek makinası (SVM) gibi sınıflandırıcılara girdi olarak verilmiştir.

CNN ağlarında yapılan eğitimlere benzer şekilde, bu sınıflandırıcılarda da hem kolay hem de zorlu durumları değerlendirmek için %20, %50 ve %80 gibi farklı oranlarla eğitimler beş kez tekrarlanarak ortalama performans puanları elde edilmiştir.

Kullanılan İlk yöntem, HoG özniteliklerini çıkarmaktır. Bu yöntem, hand-crafted yöntemler arasında yaygın olarak kullanılan özelliklerden biridir. İlk olarak [58]'te önerilen HoG, hem sınıflandırmada [60, 61] hem de nesne algılamada çalışmalarında [62, 63] başarılı sonuçlar vermiştir. HOG, bir görüntüyü $N \times N$ boyutlarında küçük kare bloklara böler ve her blok içindeki yönlendirilmiş gradyanların histogramını hesaplar. Daha sonra bloklar arasında özellik normalizasyonu gerçekleştirilir. LBP ilk olarak Ojala ve ark. [64] tarafından sunulmuştur. Algoritma rotasyona karşı bağımsız olması ve aydınlanma problemlerine karşı dayanıklı olduğu için doku analizi problemlerinde sıklıkla tercih edilmiştir [65, 66]. Algoritma, x_{ij} komşuluğundaki pikselleri, x_{ij} 'den büyükse 1, değilse 0 olarak değerlendirir. Belirli bir başlangıç noktasından, bu ikili değerlerden üretilen ondalık değer, x_{ij} 'nin dönüştürülmüş değerini temsil eder. LBP'de, HoG algoritmasında olduğu gibi, işlenen görüntü $N \times N$ kare bloklara bölünür. Her bloğun histogramı daha sonra füzyon edilmiş görüntü özelliklerini elde etmek için kullanılır. Hem HoG hem de LBP özniteliklerini çıkarmak için, 224×224 görüntünün sekizde birine karşılık gelen N değeri 28 olarak seçilmiştir.

Sınıflandırıcı olarak, modifiye edilmiş bir MLP tarzı Derin Sinir Ağı (DNN) [62] modeli de kullanılmıştır. Standart MLP yapısı içinde batch normalizasyon (BNM) [63] ve rektifiye lineer birim (ReLU) gibi yenilikçi katmanlar kullanılarak geliştirilmiştir. LBP ve HoG vektörlerini sınıflandıran ağın mimarisi $v_i \rightarrow FC_1 \rightarrow bnorm \rightarrow ReLU \rightarrow FC_2 \rightarrow softmax$. şeklinde verilmiştir. $v_i \in \mathbb{R}^{1 \times 1 \times N_f}$ çıkarılan öznitelik vektörüdür, burada N_f vektörün uzunluğudur. FC1 ve FC2 tam bağlantılı katmanlardır ve $2 \times N_f$ ve N_c nöronları içerirler. Ağın eğitimi için 100 iterasyon seçilmiştir ve minibatch değeri kullanılan veri setinin 1/5'i olarak tanımlanmıştır. Eğitim aşamasında Adam optimizier [64] yöntemi kullanılmıştır.

Bu üç yöntemle elde edilen öznitelik vektörlerinin uzunlukları oldukça yüksektir. Bu nedenle vektörler, PCA ve lineer diskriminant analizi (LDA) kullanılarak Sınıf-1'e (Class-1) indirgenmiştir. Elde edilen öznitelikler DNN, kernel SVM ve k-NN algoritmalarının varyasyonları ile sınıflandırılmış ve adaptif TT'nin etkileri önceki fonksiyonlarla karşılaştırılmıştır. CNN ağlarında olduğu gibi, %20, %50 ve %80 oranlarında eğitimler yapıldı ve bu eğitimlerin beş kez tekrarlanmış ortalamaları hesaplandı. Hand-crafted yöntemlerde bu şekilde elde edilen tüm deneysel sonuçların

tek bir tabloda gösterilmesi, anlaşılmasını zorlaştırmaktadır. Bu nedenle, sonuçlara ilişkin yorumlar burada belirtilmiştir. Ancak test sonuçları makalenin sonundaki ek bölümünde sunulmuştur. Detaylı sonuçlar Ek-A'daki tablolarda görülebilir.

Tabloların daha kolay gözlemlenebilir olması için sonuçlar renklendirilmiştir. Bu renklendirmede **MAVİ** renk karşılaştırma grubundaki en iyi sonuçları, **YEŞİL** renk en iyi 2. sonuçları, **BEYAZ** renk 3. en iyi sonucu ve **KIRMIZI** renk ise karşılaştırma tablosundaki en kötü sonucu ifade etmektedir.

Tablo A.1'de NATOPS *Full-Body* veri seti üzerinde hareket tanıma sonuçları verilmiştir. Sonuçlar incelendiğinde PCA ile elde edilen özelliklerin %50 ve %80 oranlarında test edildiği daha zorlu durumlarda önermiş olduğumuz adaptif yöntemin çok daha iyi olduğu görülmektedir. Ancak HOG özellikleri için önerdiğimiz yöntem en kötü performansı göstermiştir. Buna rağmen HOG özelliklerinde en iyi sonuca ulaşan μ_2 fonksiyonunun performansı ile birçok durumda %1'in altında fark oluşmuştur. LBP özellikleri için en iyi performans μ_3 fonksiyonu ile üretilen HE görüntülerinden elde edilmiştir. Bu kıyaslamada da Adaptif fonksiyonumuz bir çok kıyaslamada %0.5'in altında farklılık göstermiştir. Tablo A.1'in genelinde ise μ_4 fonksiyonunun en kötü sonuçları verdiği görülmektedir. Hatta μ_4 'e ait birçok sonuçların çoğunda bu fark en iyi sonuca kıyasla %2'nin de üzerindedir.

Tablo A.1 ve Tablo A.2 sonuçları incelendiğinde, elde edilen hareket tanıma sonuçları, PCA özellikleri için ResNet-18 sonuçlarına benzerken, LBP ve HOG özellikleri ile elde edilen sonuçlar da VGG-F ağ sonuçlarına benzerlik göstermektedir. NATOPS veri seti üzerinde öznitelik seçme yöntemlerinin oldukça etkili olduğu görülmektedir. Tablo A.2'deki tüm test sonuçları incelendiğinde en iyi sonucun μ_2 ve μ_3 fonksiyonlarının, en kötü sonucun ise μ_4 'ün olduğu görülmektedir. Daha önce açıklandığı gibi *MovingLimbs* veri setinde vücut silüeti olmadığı için ve μ_3 çok yakın performans elde etmiş ve μ_R bu eksiklikten olumsuz etkilenmiştir. En iyi ve en kötü sonuçlar arasında %12'ye kadar ulaşan değişimler görülmüştür. Önerdiğimiz μ_R fonksiyonu ise Tablo A.2 sonuçlarında ortalama bir değerde kalmıştır. Bir veri seti üzerinde μ_2 , μ_3 , μ_4 yöntemlerini tek tek denemek ve süreci uzatmak yerine μ_R fonksiyonu ile zaman maliyeti azaltılabilir. Öznitelik çıkarma yöntemleri *Moving-Limbs* veri setinin performansı üzerinde bir etkiye sahipken, TT fonksiyonlarının sıralanması üzerinde hiçbir etkisi olmamıştır. Bu veri setindeki sonuçlar ve sıralamalar, CNN ağlarından elde edilenlere benzerdir. Bu sonuca göre fonksiyonların sınıflandırıcılardan çok fazla etkilenmediği söylenebilir.

Tablo A.3'teki NATOPS *Full-Body* veri seti kullanılarak yapılan kişisel kimliklendirme sonuçları incelendiğinde, tüm fonksiyonlar PCA özelliği ile farklı sınıflandırıcılarda neredeyse mükemmel benzerlikte performans elde etmiştir. Adaptif TT fonksiyonu,

LBP özelliklerini kullanan tüm hand-crafted sınıflandırıcılarda ortalamanın altında kalmıştır ancak HoG özelliklerinin kullanıldığı varyasyonlarda ise en iyi performans gösteren fonksiyon olmuştur. Tablo A.3, özellikle zorlu durumlarda HoG özelliklerinde önerdiğimiz fonksiyonun daha gürbüz bir şekilde hareketi temsil ettiğini göstermektedir. LBP özelliklerinde en iyi ve gürbüz performans μ_2 fonksiyonu ile elde edilmiştir. HoG ve PCA özellikleri ile elde edilen sonuçlar CNN ağlarında elde edilen sonuçlara benzerdir. Bu sonuçlara göre, hand-crafted sınıflandırıcılardaki tüm TT fonksiyonları, NATOPS *Full-Body* veri setinde CNN ağlarında ulaşılan kişi kimliklendirme ile benzer performans göstermiştir.

Natops *Moving-Limbs* veri seri sonuçlarının yer aldığı Tablo A.4 incelendiğinde, HoG özellikleri ile yapılan sınıflandırmalar varyasyonlarında önerdiğimiz yöntem oldukça iyi performans göstermiştir. Özellikle DNN ve k-NN ile yapılan varyasyonlarında diğer yöntemlere kıyasla neredeyse tüm oranlarda iyi performansı adaptif TT yöntemimiz göstermiştir diyebiliriz. PCA ve LBP özelliklerine ait kıyaslamalarda μ_2 , μ_3 ve μ_R arasında daha dağılımlı bir sonuç tablosu görülmektedir. Ancak μ_2 özellikle SVM varyasyonların PCA ve LBP özelliklerinde daha yüksek performans göstermiştir. μ_4 fonksiyonu LBP özellikleriyle yapılan sınıflamalarda önceki tablolara göre daha iyi performans gösterse de, genel sıralamada yine en düşük başarıyı gösterdiği görülmektedir. Bu durum göstermektedir ki, Şekil 4.1’de ifade edilen μ_4 grafiği her ne kadar vurgu alanı olarak Şekil 4.2’de belirtilen μ_R grafiklerine benzese de hareketin vurgulanma bölgelerindeki ufak değişiklikler bile performansı ciddi anlamda etkilemektedir.

Sınıflandırıcı ve öznitelik çıkarma yöntemlerine bağlı olarak tüm fonksiyonların performanslarında değişme olabildiği gözlemlenmiştir. Bununla birlikte, genel performansta, adaptif TT fonksiyonu bu varyasyondan daha az etkilenmiştir. Bu durum, önerilen μ_R fonksiyonunun HE görüntülerini mevcut TT fonksiyonlarından daha iyi temsil edebileceğini göstermektedir.

Tablo A.5’deki SBU Full-Body veri setinin sonuçları incelendiğinde, sonuçlar, CNN ağlarının bu veri setindeki performansında olduğu gibi, hareket tanıma için mevcut fonksiyonlara kıyasla çok daha iyi olduğunu göstermektedir. Performans skorları CNN ağları kadar yüksek olmasa da μ_R fonksiyonu hand-crafted varyasyonların neredeyse tamamında en iyi veya ortalama üstü iyi sonuçlar elde etmiş ve farklı sınıflandırıcı ve öznitelik çıkarma yöntemlerinde gürbüz bir sonuç göstermiştir. Diğer yöntemlerin performansları, önceki tablolarda olduğu gibi sınıflandırıcı ve özellik çıkarma varyasyonlarına göre daha değişken sonuçlara sahiptir. Veri setinin %20’sini eğitim %80’ini test olarak kullandığımız zorlu durumlarda ise Adaptif fonksiyonumuz çok daha kararlı olmuştur. Tablo A.6’daki SBU *Moving-Limbs* sonuçları incelendiğinde,

μ_R fonksiyonunun enerji görüntülerinden hareket tanıma konusunda özellikle HoG özelliklerinin kullanıldığı eğitim varyasyonlarında çok başarılı sonuçlar üretti görülmektedir. LBP özelliklerinin kullanıldığı varyasyonlarda, mevcut fonksiyonlar eğitim oranı, sınıflandırıcı türü ve parametre varyasyonlarından çok daha etkilenmiş ve bazen iyi bazen en kötü sonuçlar arasında değişkenlik göstermiştir. Adaptif fonksiyonumuz ise LBP verileriyle yapılan eğitim varyasyonlarından çok daha az etkilenmiştir. Bu durum önerdiğimiz yöntemin HE görüntülerinde hareketi daha iyi temsil edebildiğini göstermektedir. Tablo A.5 ve Tablo A.6'da bazı sınıflandırıcı ve özellik çıkarıcı varyasyonları için karşılaştırılabilir anlamlı sonuçlara ulaşılamaması sebebiyle tablolarda bu alanlar *N/A* (Not applicable) olarak işaretlenmiştir.

NATOPS ve SBU veri seti sonuçları incelendiğinde, mevcut fonksiyonlar arasında seçim yapmak için ampirik bir yaklaşım ile gözlem yapmak gerekmektedir. Yani değişen varyasyonlardan hangisinde hangi fonksiyon daha iyi olduğunu tüm yöntemleri deneyerek görmek gerekmektedir. Bu durum işlem maliyetini önemli ölçüde arttırmakta ve uygulanabilir olmaktan çıkmaktadır. Öte yandan, SBU *Moving-Limbs* veri seti için μ_R fonksiyonunun sonuçları incelendiğinde, ilk iki en iyi sonuç uyarlamalı TT fonksiyonu ile elde edilmiştir. Bu da μ_R fonksiyonunun hem hareketi daha iyi temsil ettiğini hem de daha düşük bir zaman maliyeti sağladığını göstermektedir. BodyLogin veri seti ile hareket tanıma için sınıflandırma sonuçları Tablo A.7'de gösterilmektedir. μ_R fonksiyonu, hand-crafted varyasyonlarının büyük çoğunluğunda en iyi sonucu çıkartmış genel ortalamada en iyi 2 sonuç ağırlığını korumuştur. Ayrıca hiçbir varyasyonda en kötü sonuca düşmemiştir. Diğer fonksiyonların sonuçlar sınıflayıcı türü, parametresi ve eğitim oranı gibi varyasyonlardan çok daha fazla etkilenmiş ve değişken sonuçlar üretmiştir. Diğer veri setlerinde en kötü performans gösteren μ_4 fonksiyonu, BodyLogin *Full-Body* hareket tanıma testlerinde μ_R fonksiyonundan sonra en iyi ikinci performans gösteren fonksiyon olmuştur. Bu durum mevcut yöntemlerin veri setinden ne derece etkilendiğini de göstermektedir.

Tablo A.8 de ise BodyLogin *Full-Body* kişi kimliklendirme sonuçlarına yer verilmiştir. Kişi tanımlama sonuçlarına göre, adaptif TT fonksiyonu, μ_R , %20 test prosedüründe LBP ve HOG özelliklerinde en iyi başarıyı ve PCA özellikleri ile ikinci en iyi başarıyı elde etmiştir. Tüm sınıflandırıcı ve özellik çıkarım varyasyonlarında, BodyLogin veri seti ile kişi kimliklendirme genellikle, birinci ve ikinci en iyi sonuçlar μ_R fonksiyonundan elde edilmiştir. BodyLogin veri setinde kişi tanımlama için mevcut TT fonksiyonları incelendiğinde, adaptif TT'ye en yakın performans ise μ_4 fonksiyonu ile elde edildi. μ_2 fonksiyonu genellikle ortalama sonuçlar üretilirken ve tüm fonksiyonlar arasında en kötü sonuç μ_3 ile elde edildi.

Bölüm-4'te yer alan bu çalışma Pattern Recognition dergisinde "Human and action recog-

nition using adaptive energy images” başlığı ile kabul edilmiş ve yayınlanmıştır [67].

4.4 Adaptif HE Tartışması

Önerilen yöntem NATOPS, SBU Kinect ve BodyLogin veri kümeleri üzerinde test edilmiştir. Bu veri kümelerini kullanarak üst vücut ve tüm vücut hareketi tanıma, kişi tanımlama ve iki kişi etkileşimi üzerinde eğitimler ve testler yapılmıştır. Tüm sonuçlar incelendiğinde, kişi kimliklendirme ve hareket tanıma için HE görüntülerinin korelasyon katsayısındaki değişime dayalı bir üyelik fonksiyonu olan μ_R yönteminin daha iyi sonuçlar elde ettiği görülmüştür. Önerilen yöntemle elde edilen sonuçların, CNN ağlarının sınıflandırma gücünden mi yoksa önerilen yöntemin sağlamlığından mı kaynaklandığı sorgulanmıştır. Bu nedenle önerilen fonksiyon ve mevcut üyelik fonksiyonları, farklı öznitelik çıkarma yöntemleri, farklı ağlar ve bunların farklı parametrik varyasyonları ile elde edilen makine öğrenme yapıları üzerinde test edilmiştir. Farklı makine öğrenimi ve özellik çıkarma varyasyonları CNN'lere kıyasla daha düşük sonuçlar verse de, CNN'ler ve diğer ağlar aynı veri kümeleri için benzer sıralama sonuçları verir. CNN'lerin öznitelik çıkarma ve sınıflandırma yeteneklerinin diğerlerinden daha iyi olduğu açıktır. Ancak tüm sınıflandırıcılarda aynı veri setleri kullanılarak benzer sıralama sonuçları elde edilmesi, önerilen adaptif TT fonksiyonunun sınıflandırma ağlarından bağımsız olarak diğer fonksiyonlardan daha sağlam olduğunu göstermektedir. NATOPS veri seti incelendiğinde, hareketler başlangıç ve orta bölgelerinde daha belirgindir. Başlangıç ve orta alanları vurgulayan şablonların, önerdiğimiz şablondan daha iyi sonuç vermesinin nedeni bu olabilir. Ancak, her veri kümesi NATOPS gibi belirgin bir şekilde belirli aralıkları içeren eylemleri içermemektedir. SBU Kinect ve BodyLogin veri setlerinde, NATOPS veri setlerinden farklı olarak, aksiyonlar hareketin başından sonuna kadar devam eder. Böyle bir durumda belirli aralıkları vurgulayan şablonlar yetersiz kalacaktır. Elde edilen sonuçlarda görüldüğü gibi, önerilen yöntem SBU Kinect ve BodyLogin veri setlerinde hareket tanıma ve kişi tanımlama için önceki çalışmalara ve mevcut şablonlara göre daha etkilidir. Sonuç olarak yaklaşımımızın hareketleri daha etkin temsil etme gücüne sahip olduğu söylenebilir.

5

SİLÜET, TERMAL VE İSKELET (STI) VERİ SETİNİN OLUŞTURULMASI

Tez kapsamında hem daha önce geliştirdiğimiz yöntemleri test edebileceğimiz hem de halka açık bir veri seti oluşturarak bu alandaki çalışmalara katkı sağlayacak bir veri seti oluşturmayı amaçladık. Bu konuda 2 farklı motivasyon noktamız bulunmaktadır. İlk olarak, 2. bölümde oluşturulan sanal veri seti tarzında hem hareketi hem de yüz görüntüsü içeren kombine bir veri setinin olmamasıdır. İkinci motivasyon noktamız ise 3. bölümde kullanılan veri setlerinde görüldüğü gibi bir çok veri setinde kişi sayısının özellikler 20-30-40 aralığında sınırlı kalmış olmasıdır. Bu iki konu bu alanda yeni bir veri seti oluşturma konusunda bize motivasyon oluşturmuştur.

Veri seti oluşturulmadan önce atılacak ilk adım ise o veri setini insanların onayları ve etik kurallar çerçevesinde toplamaktır. Bu amaçla üniversite etik kuruluna başvuruda bulunulmuş ve veri seti oluşturulması konusunda etik onay alınmıştır. Veri seti oluşturulması sürecine gönüllü olarak katılan katılımcılar bu konuda bilgilendirilmiş ve Aydınlatılmış Onam Formu doldurtularak kayıtları toplanmıştır. Bu onam formu dışında katılımcılarım herhangi bir kişisel verisi kayıt altına alınmamış, kayıtlar sıralı olarak ID numarasına göre yapılmıştır. Silüet, termal görüntü ve iskelet verilerinden herhangi bir şekilde katılımcılara yönelik kişisel veri içeren bir bilgi yer almamaktadır. Katılımcılara verilen ve imzalı onayları alınan belgeler EK-B kısmında sunulmuştur.

Bu bölümde Kinect V1 kamera ve Flir Termal kamera ile elde edilen veri setlerinin ne şekilde toplandığı hakkında ayrıntılı bilgiler verilecektir.

5.1 Kinect ve Termal Kameralar ile Toplanan Görüntüler

5.1.1 Kinect Kamera ile Toplanan Silüet Görüntüleri

Bu bölümde Microsoft Kinect V1 kamera ve bu kameraya ait arayüz yazılımı ile silüet ve iskelet veri setlerinin elde edilme süreci, bu süreçte yaşanan aksaklıklar, veri kayıpları ve bu durumlar için üretilen çözümler anlatılacaktır.

Derinlik görüntülerinin toplandığı bu süreçte aynı zamanda eklem noktaları arasındaki bağlantıların oluşturulması ile elde edilen iskelet verileri de toplanmıştır. İskelet verisi Şekil 5.1’de gösterilmiştir. Silüet veri seti 7 farklı hareketten oluşmaktadır. Bunlar sırası ile "yürüme, iki el ile havaya Z harfi çizme, iki el ile havada Daire çizme, tek elini kaldırıp bileğini havada çevirme, tek ayağı ile bileğini çevirme, sağa ve sola sırayla eğilip yerden bir cisim alma, yürürken önündeki engelin etrafından geçme" şeklindedir.

Bu hareketler tercih edilirken vücudun tamamının, sadece üst ekstremlerinin ve sadece alt ekstremlerinin hareketleri olacak şekilde oluşturulmuştur. Bu sayede vücudun hem tamamı hem de belirli bölgelerinin hareketini içeren bir veri seti oluşturulmuştur. Kinect kamera ile kayıt edilen tüm görüntüler Şekil 5.2’de gösterilen formattaki gibi toplanmıştır.



Şekil 5.1 İskelet görüntüsü

5.1.2 Termal Kamera ile Kaydedilen Görüntüler

Flir E8 modeli termal kamera kullanılarak yüze ait termal görüntü veri seti oluşturulmuştur. Kamera 30 cm’den yaklaşık 20 metreye kadar 320×240 çözünürlükte termal görüntü sağlamaktadır. Görüntüler anlık frame’ler yerine video modunda kesintisiz olarak kayıt edilmiştir. Bu veri seti kişilerin kameranın yaklaşık 1 metre uzağına geçerek başlarını sağa ortaya ve sola döndürmeleri ve 10 kez tekrar etmesi ile toplanmıştır. Bu sayede yüzün hem karşıdan görünümü hem de sağa ve sola 90° açıya kadar tüm aralıkları elde edilmiş olmaktadır. Görüntülerin video formatında kayıt edilmesi, hareket anında tüm açıların alınabilmesini sağlamakla birlikte katılımcının etkileşim süresini de azaltmıştır. Bu sayede sadece 10 kere başını oynatması yeterli olmuştur. Kayıt formatı Şekil 5.3’te gösterilmektedir.

Bu kayıtlar sayesinde 78 kişilik bir veri seti oluşturulmuştur. Veri seti her bir kişi için 7 farklı hareketi (ayak bileği çevirme, el bileği çevirme, iki el ile dışa doğru daire çizme, önce sağa sonra sola eğilme, yürüme, çift el ile Z şekli çizme ve engel etrafından yürüme), 180° termal yüz görüntülerini ve kişinin durağan halde iskelet görüntülerini



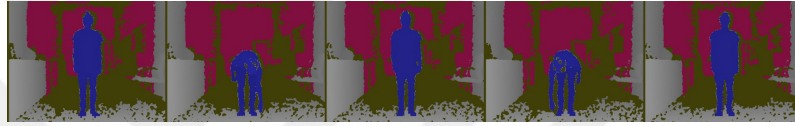
(a) 2 çift adım(sağ-sol-sağ-sol) atarak kameraya doğru yürüme – 10 tekrar



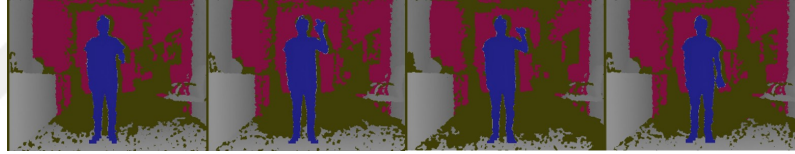
(b) Havada iki el ile Z harfi çizme - 10 tekrar



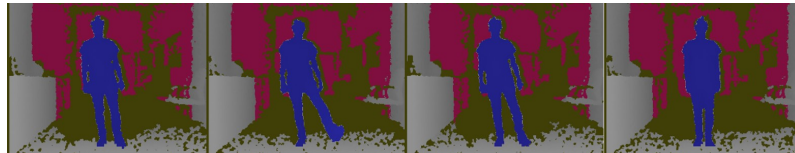
(c) İki elini kaldırıp dışa doğru tam daire çizme – 10 tekrar



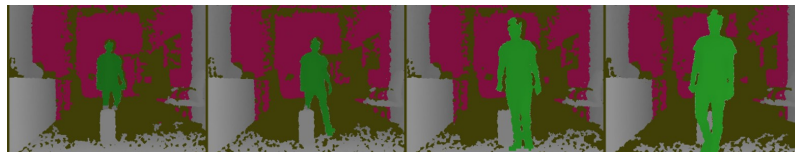
(d) Sağa eğilip yerden bir cisim alıp doğrulmak, sola eğilip yerden bir cisim alıp doğrulmak – 10 tekrar



(e) Tek elini kaldırıp sallamak – 10 tekrar



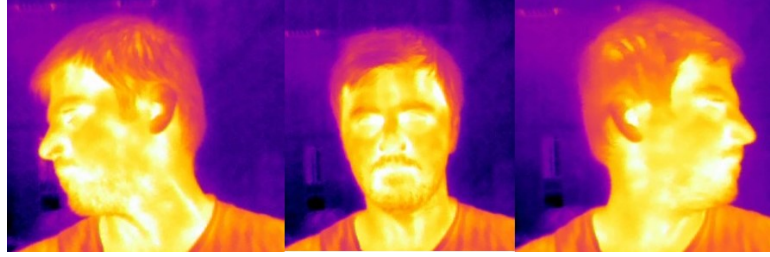
(f) Tek ayağını yerden az bir miktar kaldırıp daire çizmek – 10 tekrar



(g) Yürürken önünde duran engelin etrafından geçmek – 10 tekrar

Şekil 5.2 Kinect kamera ile kaydedilen hareketlerin formatı

içermektedir. Her görüntüden ve hareket setinden 10 tekrar yaptırılmıştır. Her kişiye bir ID verilmiştir. Veriler arasında bağlantı olması için hareket verilerine atanan ID numarası termal yüz ve iskelet verisinde de aynı ID ile kaydedilmiştir.



Şekil 5.3 Termal kameraya bakarken başını önce sağa tam sonra sola tam çevirme – 10 tekrar

5.2 Bazı Katılımcılardan 2.Oturum Verilerinin Alınması

Katılımcılardan tekrar ulaşılabilen belirli bir kısımdan 2. oturum verileri alınmıştır. Bir önceki verilerin alınmasında yaşanan hatalar göz önünde bulundurularak hareketler 30FPS değerine ulaşıldıktan sonra kayıt edilmiştir. Bu sayede 2. kayıtlarda benzer hataların yaşanmamasına dikkat edilmiştir. 2. oturumda elde edilecek veri sayısı 20 kişi sınırlı kalmıştır.

5.3 Kinect Kamera ile Toplanan Verilerdeki Hatalar

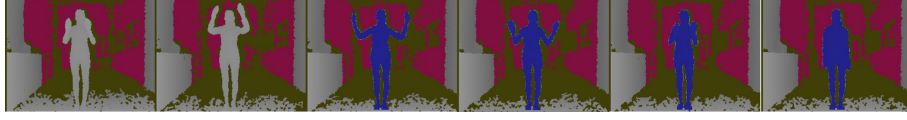
Kinect kamera ile 7 farklı hareket için 10'ar adet örnek “.xed” formatında videolar şeklinde Kinect SDK'sı ile ayrı ayrı toplanmıştır. Bu videolar yine SDK üzerinden framelelere ayrılarak dosyalanmıştır. Videoların framelelere çevrilme aşamasında bazı zorluklarla karşılaşmıştır. Bunları genel olarak; - Hareket verisinde kayıp, - Siluetin tanınmasında gecikme, olarak iki başlık altında toplayabiliriz.

5.3.1 Hareket Verisinde Kayıp

Hareket verilerindeki kayıplar ele alındığında buna sebep olan nedenler araştırılmıştır. Kayıt yapılan bilgisayarın donanımsal özelliğine veya sabit disk doluluğuna bağlı olarak kayıt tuşuna basıldığında programın eşzamanlı olarak kayda başlamadığı durumlar olduğu, bu sebeple kaydın yaklaşık 0.5-1 sn aralığında gecikmeli başladığı tespit edilmiştir. Hareketlerin ortalama 3-4 sn sürdüğü göz önüne alındığında, bazı kayıtlarda bu gecikme sebebiyle başlangıç karesinin eller yanlarda serbest olarak görülmesi gerekirken, ilk karenin hareketin içerisinde başladığı görülmüştür. Bu sebeple bazı hareket videolarının başlangıç kısımlarında kayıplara oluşmuştur. Örnek hareket verileri Şekil 5.4'te görselleştirilmiştir.



(a) Z-hareketi videosundaki ilk frame hareket ortasından başlamıştır



(b) Daire hareketi videosundaki ilk frame hareket ortasından başlamıştır

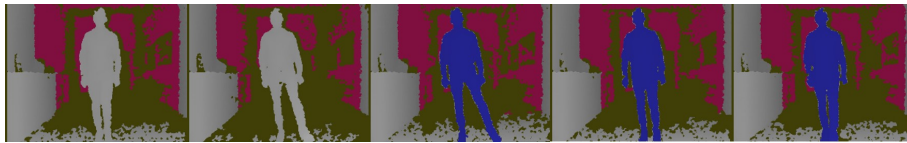
Şekil 5.4 Kinect kamera ile kaydedilen hareketlerde ilk karelerde oluşan kayıp örnekleri

5.3.2 Siluet Tanınmasında Gecikme

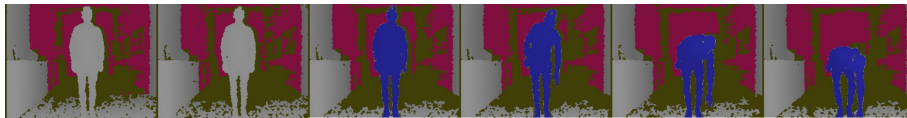
Videolar kayıt edilirken SDK ekranındaki görüntüde öncelikle siluetin tanınması beklenmiş ve katılımcılara ondan sonra harekete başlama komutu verilmiştir. Ancak bu videoların tekrar SDK'da açılıp framelere dönüştürülmesi sürecinde, siluetlerin tanınması gecikmekte hatta bazı kişilerde hareketin sonuna kadar siluet tanınması gerçekleşmemektedir. Bunun sebepleri incelendiğinde birkaç farklı nedeni olduğu tespit edilmiştir.

1.Sebep; siluet SDK üzerinde tanınmış olmasına rağmen video kayıt moduna geçildiğinde bu özellikler kaydedilmemekte ve kayıt edilen videolar tekrar açıldığında bu siluet tanıma algoritması sıfırdan çalıştırılmaktadır. Yani SDK ekranında siluet tanınmış olsa dahi video kaydına aktarılmamaktadır. Bunun aksine her kaydedilen videoda bu özelliğin arka planda yeniden çalıştırıldığı tespit edilmiştir.

2.Sebep; video kaydında, 640×480 piksel çözünürlük ve 30 FPS kayıt hızı seçilmiştir. Ancak kayıt başlatıldıktan sonra ilk 1,5 sn boyunca kayıt hızı 30 FPS altında kalmakta ve kaydedilen görüntülerde siluet tanınması 30 FPS hızına çıktıktan sonraki süreçte başlamaktadır. Aşağıdaki şekillerde hareket kaydı olmadan silüetin geç tanınmasına örnekler verilmiştir.



(a) Ayak ile daire hareketinde silüetin geç tanınması



(b) Yaklaşık 1 sn sonra silüetin tanınması

Şekil 5.5 Geç tanınan silüet kareleri örnekleri (Hareket kaybı olmadan)

3. Sebep; Video 30 FPS hızına çıktıktan sonra silüet tanınmasına geçmektedir. Ancak hareketin bu süreçten önce başladığı durumlarda (örneğin 18 FPS’de iken kolların yukarıya kaldırılması) katılımcıların hareketlerindeki duruma göre silüet ve iskelet görüntülerinin tanınması gecikebilmektedir.



Şekil 5.6 30 FPS hızına ulaşmadan başlanan daire hareketinde silüet tanınmanın gecikmesi

5.3.3 Tespit Edilen Kayıt Sorunları İçin Uygulanan Çözümler

Kayıt sırasından yaşanan veri kaybı ve silüet tanınması problemleri konusunda bir takım çözüm yöntemleri oluşturulmuş ve başarıyla uygulanmıştır. Bunlar aşağıdaki sıralanmıştır.

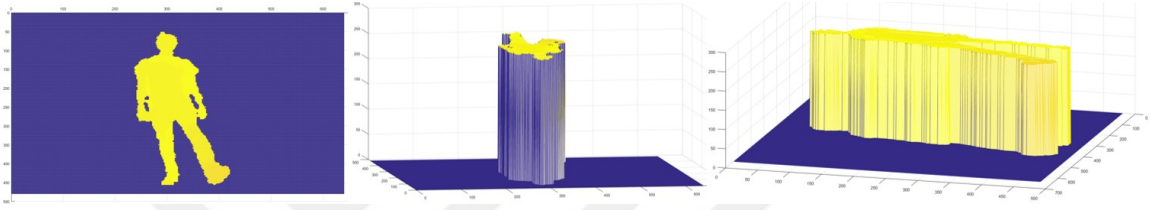
a) Hareket kayıplarında oluşan eksiklikleri gidermek için, özellikle düşük sayılı verilerin çoğaltılmasında kullanılan Veri Artırma (Data Augmentation) yöntemi tercih edilmiştir. Eksik hareket bilgisi içeren örnekler için tam kaydedilen örneklerden kayıplı kısma uygun olacak kareler alınmış ve morfolojik genişletme-aşındırma uygulanarak ufak değişiklikler oluşturulmuştur. Bu kareler eksik kısımların tamamlanmasında kullanılmıştır.

b) Silüet tanınmasında oluşan gecikmeler için görüntüler üzerinde görüntü segmentasyonu (image segmentation – c-means algoritması), görüntü eşikleme ve HSV renk tonları üzerinde eşikleme yöntemleri uygulayarak arka plan görüntüleri ile silüet görüntülerinin ayrılması sağlanmıştır. Eşikleme ve görüntü segmentasyonu işlemi sonrasında görüntülerdeki ton farklılıklarını gidermek için tüm silüet görüntüleri binarize edilmiştir. Orijinal görüntü ile birleştirilen binarize silüet görüntüleri birleştirilmiş ve arka plan tamamen ortadan kaldırılarak gri tonda silüet görüntüleri elde edilmiştir. Bu görüntüler silüet tanınmış görüntülerdeki renk tonlarına göre yeniden renklendirilmiştir. Bu sayede silüet tanınmasının gecikmesi sebebiyle oluşan farklılık giderilmiştir. Sonrasında tüm görüntüler gri renk seviyesine alınarak HE görüntüleri oluşturulması için kullanılmıştır. Bu konun ayrıntıları ön işleme başlığı altında daha ayrıntılı olarak anlatılmıştır.

5.3.4 Görüntülerde Derinlik Bilgisinin Tutulması

Kinect kamera ile elde edilen derinlik görüntülerindeki silüetler üzerinde derinlik bilgilerini tutacak şekilde segmentasyon edilmektedir. Böylece katılımcıların hareket

esnasındaki görüntüleri, sadece binary bir görüntü olarak değil aynı zamanda uzuvların kameraya uzaklık bilgisini içeren formatta saklanmaktadır. Oluşturulan veri seti üzerinde yapılacak eğitimlerde sadece binary görüntülerden ziyade derinlik bilgisi içerecek şekilde HE görüntüleri oluşturulmuştur. Bunun özellikle hareket tanıma performansını daha iyi yönde etkileyeceği varsayılmıştır. Aşağıdaki şekilde ayak hareketine ait bir karede görülen derinlik farkı gösterilmektedir. Vücut yüzeyindeki derinlik farklılıkları çok belirgin görünmese de hareket eden ayağın vücuda göre derinlik farklılığı çok net görülmektedir. Bu farklılık lokalde ton farklılıkları yani piksel değerlerinde farklılık göstermemiştir. Bu farklılıklar HE görüntüsü oluşturulurken korunmaktadır. Böylece hem hareketin yoğun olduğu kısımlar hem de derinliğin farklı olduğu uzuvlar HE görüntüsünde daha belirgin olarak temsil edilecektir.



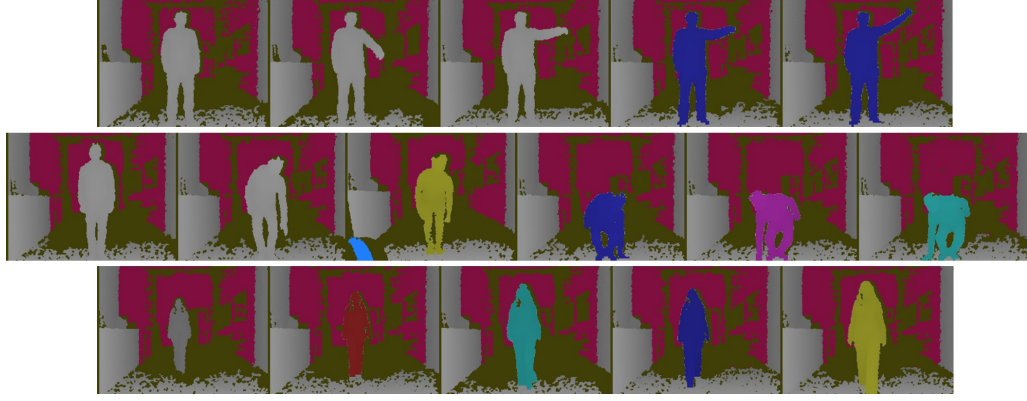
Şekil 5.7 Segmente edilen görüntülerdeki uzuvların ve gövdenin kameraya göre olan derinlik bilgisi

5.4 Kayıtlarda Yapılan Ön İşlemler ve Veri Setlerinin Hazırlanması

5.4.1 Enerji Görüntülerinin Hazırlanması için Uygulanan Ön İşlemler

Kinect kamera ile kaydedilen hareket görüntülerinde oluşan silüet tanınmasındaki gecikmeler ve farklı kişi ve tekrarlar da silüet eşiklerinin farklı tonlara sahip olması gibi zorluklarla karşılaşmıştır. Silüet tanınması kısmı, silüetin arka plandan ayrılması ve HE görüntüsünü oluşturulması için kullanılmaktadır. Ancak Kinect SDK'sı bu işlemi belirli bir standartta yapmak yerine değişken renk tonları ile yapmaktadır. Bu durum arka planın ayrıştırılma işlemini zorlaştırmaktadır. Bu zorluklar, görüntüler üzerinde komplement alınması, HSV düzlemi, morfolojik işlemler ve histogram dağılımı gibi farklı yöntemler ile giderilmiştir. Bu sayede veri kaybı en aza indirilmiştir. Özellikle Kinect kayıtlarında harekete başlanmış olmasına rağmen çekim hızının 30 fps'nin altında kaldığı durumlarda oluşan silüetin tanınmaması durumları bu sayede çözülmüş ve veri kaybı oluşması engellenmiştir. Aynı bir augmentation işlemine ihtiyaç duyulmamıştır. Bir diğer problem hareketlerin farklı tonlarda kaydedilmiş olmasıdır. Bu süreçte farklı eşikleme değerleri kullanılarak bu problem de giderilmiştir.

Kayıt edilen kareler, farklı tonlardan ve arka planları ayrıştırılmış ve vücudun derinlik bilgisi kaybedilmeden silüet görüntüleri elde edilmiştir. Mesela, yürüme hareketine ait bir karede ayak ile vücut arasındaki derinlik farkı veya eğilme hareketinde baş ve vücut



Şekil 5.8 Kinect Kamera ile kaydedilen videolardan elde edilen kareler. 30 fps altında silüet tanınmaması ve örnekler arası eşikleme tonlarında ki farklılıklar gösterilmektedir

arasındaki derinlik farkı korunmuştur. Eğer bu kareler tamamen “1” ve “0” değerleri olarak binarize edilmiş olsaydı hareketlerdeki uzuv ve vücutlar arası derinlik bilgisi kaybedilmiş olacaktı. Bu amaçla bu derinlik bilgisi kaybedilmeden silüet görselleri çıkartılmıştır.



(a) El bilek hareketi

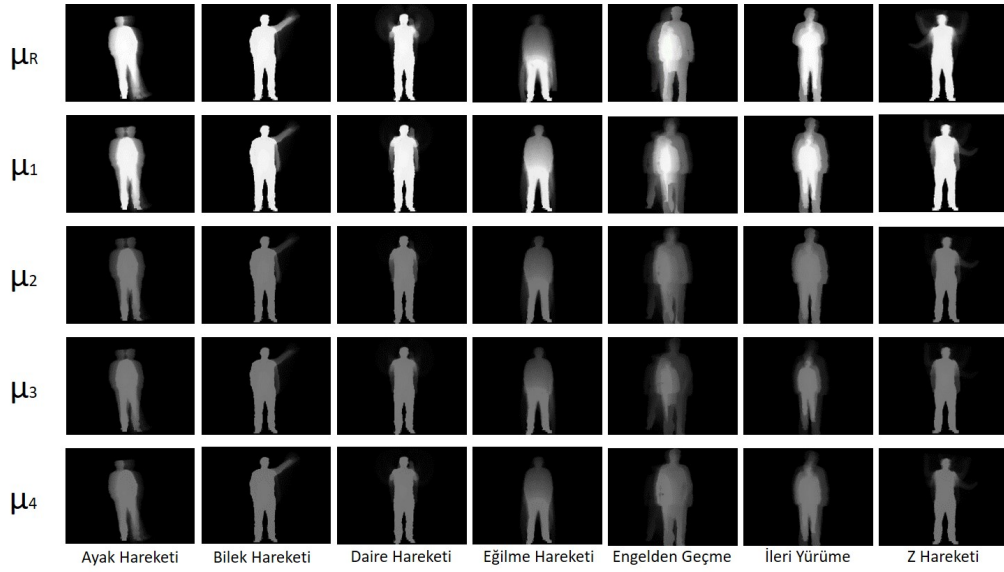


(b) Eğilme hareketi

Şekil 5.9 Derinlik bilgileri korunmuş olarak arka plandan ayrıştırılmış hareket frame örnekleri

Elde edilen bu karelerden, bir önceki bölümde anlatılan ve mevcut HE fonksiyonları olarak bahsedilen yöntemler ve korelasyon katsayılarına dayalı önermiş olduğumuz adaptif HE fonksiyonu kullanılarak enerji görüntüleri elde edilmiştir. Elde edilen hareket enerji görselleri Şekil 5.10’da gösterilmektedir.

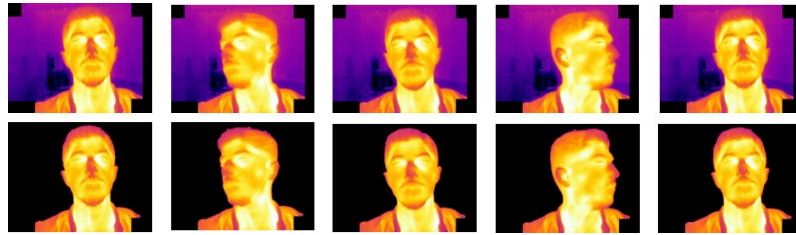
Bu yöntemler sayesinde hareket enerjisine ait veri setleri oluşturulmuştur. İlk 32 kişi için daire çizme, eğilme, ileri yürüme ve z çizme hareketleri bulunmaktadır. 33. Kişiden 77. Kişiye kadar olan 45 kişilik grupta ise tüm hareket örnekleri yer almaktadır. 2. Session verileri için ise 18 kişiden oluşan bir veri seti oluşturulmuştur. Toplanan verilerde 1. Session’dan 1 kişi 2. Session’dan ise 2 kişinin video kayıtlarında eksiklik olduğu için veri setine dâhil edilememiştir.



Şekil 5.10 Mevcut fonksiyonlar ve korelasyon fonksiyonu ile üretilen hareket HE görüntülerinin görsel kıyaslaması

5.4.2 Termal Kayıtlar, Uygulanan Ön İşlemler ve Eğitime Uygun Veri Setlerinin Oluşturulması

Termal kamera ile toplanan 360 derecelik görüntüler hareket videolarında olduğu gibi karelere ayrılmıştır. Video kayıtları sırasında ekran kenarlarında yer alan çekim bilgileri karelere ayrılırken kapatılmıştır. Kareler üzerinde RGB eşikleme, histogram dağılımı ve morfolojik yöntemler kullanılarak yüz bölgesi arka plan görsellerinden ayrıştırılmıştır. Şekil 5.11’de örnek görseller sunulmuştur.

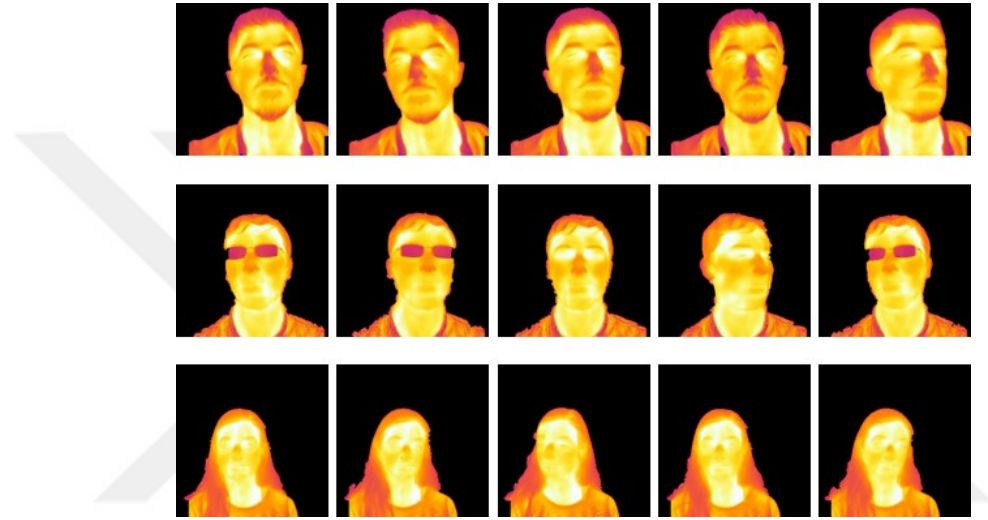


Şekil 5.11 Termal kamera kayıtlarından frame’lerin elde edilmesi ve arka plan görsellerinden ayrıştırılması

Elde edilen termal yüz görselleri 180 derece genişliğe sahip olduğu için yüzün Önü, Sağı ve Solu olacak şekilde 20’şer kare içeren 3 kümeye ayrılmıştır. Bu kümeler oluşturulurken her bir küme yaklaşık 30 derecelik açısal hareket içerecek şekilde düzenlenmiştir. Yani ön yüze ait görseller direk kameraya bakış haricinde sağa veya sola 30’ar derecelik çevrilmeleri de içermektedir. Aynı şekilde hem sağ hem de sol yan görseller yaklaşık 30’ar derecelik öne dönüş görsellerini içerecek şekilde 20’şer örneklik kümelere ayrılmıştır. Bu sayede oluşturulan 3 veri seti gerçek yaşamda karşılaşılabılır açısal durumlara daha yakın olacak şekilde düzenlenmiştir. Bu 3 veri

setine ait örnek görseller daha önce Şekil 5.3'te sunulmuştu.

Bu ayrıştırma, Ön (frontal), Sol (left) ve Sağ (right) side görsellerden örnekler alınmış ve %90 üstü korelasyon sağlayan görseller 3 ayrı kümeye ayrılmıştır. Yani tam frontal bakışa sahip görselin %90 üstü korelasyon uyumu olan görseller frontal veri kümesine alınmıştır. Aynı şekilde left ve right side bir görüntüye %90 üstü korelasyon sağlayan görseller de sırayla left ve right side veri kümelerine alınmıştır. Bu oran yaklaşık olarak 30 dereceye tekabül etmektedir. Şekil 5.12'de yaklaşık 30 derecelik ön yüz görüntüleri verilmiştir. Bu sınıflandırma ile doğal çekime en yakın şekilde olabilecek bir veri seti oluşturulmaya çalışılmıştır.

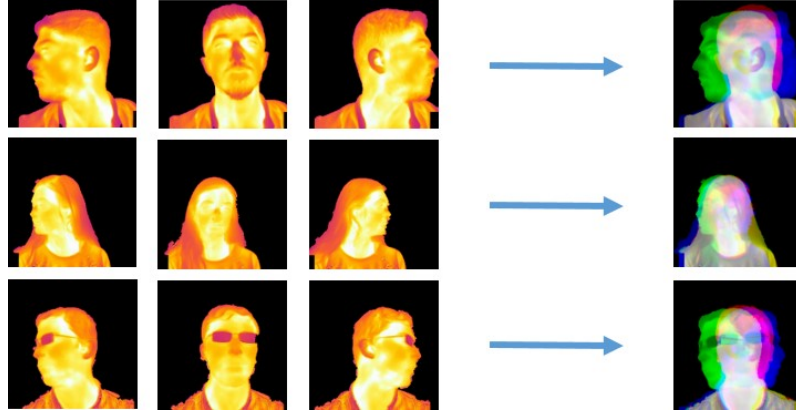


Şekil 5.12 Arka plandan temizlenmiş frontal termal yüz görselleri. Frontal termal yüz görselleri, yüzün karşıdan yaklaşık 30 derece sağa ve sola bakış açılarını içerecek şekilde oluşturulmuştur

Elde edilen bu görüntülerin her biri bir veri seti olarak kaydedilmiştir. Bunun haricinde bu 3 veri setinden ise yeni bir veri seti daha oluşturulmuştur. Oluşturulan yeni veri setindeki görseller RGB görsellerdeki gibi 3 katman içermektedir. 1.katmana left side bakışın gri görseli, 2.katmana frontal side bakışın gri görseli ve 3.katmana ise right side bakışın gri görseli yerleştirilmiştir. Bu sayede 3 farklı yöne ait termal yüz görüntüleri tek bir kare içerisinde toplanmıştır. Bu yeni veri setine ait görseller ise Şekil 5.13'te sunulmaktadır. Bu yeni veri setine ise Termal-3D veri seti adı verilmiştir.

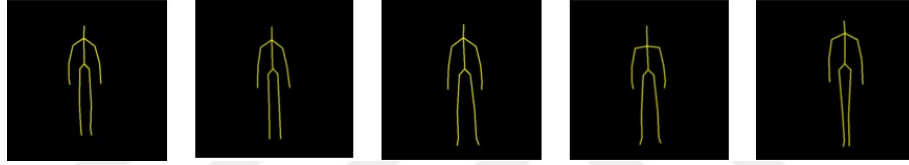
5.4.3 İskelet Görseli Kayıtları ve Eğitime Uygun Veri Setinin Hazırlanması

Kinect kamera ile hareketlerin kaydedilmesi sırasında eklemler arası uzaklık bilgileri de kaydedilmiştir. Kinect SDK'sı ile kayıt edilen görüntülerdeki tahmini 21 eklem noktası tespit edilir. Tespit edilen bu eklem noktaları birleştirilerek bir iskelet görüntüsü elde edilir. Herhangi bir arka plan görseli içermedikleri için herhangi bir ön işlem uygulanmadan kayıt edilmiştir. İskelet veri seti oluşturulurken kişilerin durağan



Şekil 5.13 Termal left, frontal ve right side görselleri ve bu görsellerin gri seviyeye çevrildikten sonra birleştirilmesi ile oluşturulan Termal-3D yüz veri setine ait görsel örnekleri

hallerdeki iskelet yapıları tercih edilmiştir. Bu veri setine ait görseller Şekil 5.13'te sunulmaktadır.



Şekil 5.14 İskelet yapısı görselleri. Her bir görsel farklı kişiye aittir. Durağan haldeki iskelet yapısı görselleri alınarak oluşturulmuştur

5.4.4 Eğitime Hazır Veri Setleri

Yapılan tüm kayıtlar, kayıtların ayrıştırılması, hatalı ve problemli görüntülerin düzenlenmesi, görüntülerin arka planlarının temizlenmesi ve uygulanan ön işlemler sonucunda tüm veriler eğitime hazır olacak şekilde veri setlerine ayrılmıştır. Yapılan bu işlemlerin sonucunda;

- Hareket enerjisine ait 5 farklı fonksiyon kullanılarak elde edilmiş enerji görüntülerinden; kişi tanıma yönelik 1 veri seti, hareket tanıma yönelik 1 veri seti olmak üzere **2 hareket veri seti**,
- Termal yüz görüntülerinden elde edilen Termal-3D veri seti ve left-frontal-right side yüz görsellerini içeren görüntülerden oluşan 2. bir veri seti olmak üzere **2 termal veri seti**,
- Kişilerin ayakta durağan haldeki eklemler arası uzaklığı içeren **1 iskelet veri seti**, olmak üzere **5 farklı veri seti** oluşturulmuştur.

Bu bölümde, 4. bölümde anlatılan veri setleri üzerinde yapılan eğitimler, bu eğitimler için seçilen ağlar ve performans değerlendirmeleri anlatılacaktır. Eğitimler, kişi kimliklendirme, hareket tanıma, termal yüz tanıma, iskelet verisi ile kişi kimliklendirme, eksik veri durumundaki performans ve bu verilerin füzyon performanslarını içermektedir.

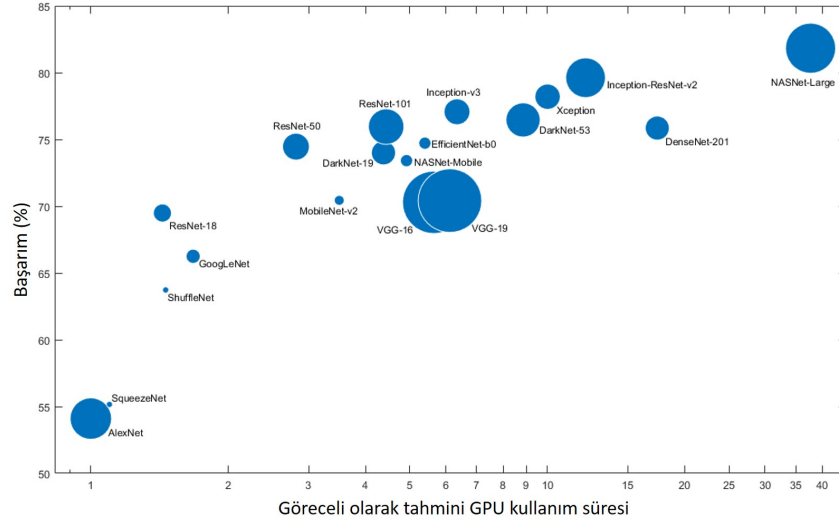
6.1 Ağ Seçim Kriterleri

Oluşturulan veri setleri için öncelikli olarak güncel çalışmalarda kullanılan ve başarılı sonuçlar verebilecek state-of-art ağlar incelenmiştir. Eğitim ağları tercihinde öncelikli olarak eğitim süresi, ağın parametre boyutu ve performans değişkenlerine bakılmıştır. Şekil 6.1’de state-of-art derin ağların performans ve hızlarına ait bilgiler veren bir grafik sunulmuştur. Görselde yer alan noktaların genişliği o ağın disk üzerinde kapladığı boyutsal alanı ifade etmektedir. Grafik bu bağlamda incelendiğinde hem ağın disk üzerinde kapladığı alan tasarrufu hem de ortalama zaman-performans için en iyi ağların GoogleNet, Resnet-18 ve MobileNet-V2 ağları olduğu görülmektedir. Bu nedenle eğitimler seçilen bu ağlar ile gerçekleştirilmiştir.

Elde edilecek sonuçlara göre, veri setlerinden en iyi performansı alabilmek ve HE görüntüsü fonksiyonları arasında kıyaslama yapabilmek için ne tür bir ağ kullanılması veya oluşturulması gerektiğine dair bilgi ve tahminler de bu sayede belirlenmiş olacaktır.

6.1.1 GoogleNet Ağ Mimarisi

GoogleNet ağı 144 katmandan bir CNN ağıdır. Bu ağ 1000 farklı nesnenin sınıflandırılması için eğitilmiştir. Veri setlerinin eğitiminde hızlı sonuç almak için GoogleNet ağı fine-tuned edilmiştir. Fully-Connected öncesi ağ üzerindeki Layer’larda herhangi bir değişiklik yapılmamıştır. Fully-Connected sonrası kısımlar olan Softmax,



Şekil 6.1 State-of-the-art derin ağlar için harcanacak eğitim zamanı ve doğruluk grafiği

Classification Layer ve Output layer'lar, kullanılacak veri setindeki sınıf sayısına göre yeniden düzenlenmiş ve eğitim bu şekilde gerçekleştirilmiştir.

Şekil 6.2'de GoogleNet'e ait olarak tüm modelin katmanlarıyla ilgili bilgiler verilmektedir. Sütunlardaki bilgiler sırasıyla: Katmanda yapılan işlem tipi (type), filtre boyutu/adım kaydırma büyüklüğü (patch size/stride), çıkış boyutu (output size), derinlik filtre boyutu (depth), boyut azaltma için filtre boyutu (reduce), hesaplanan parametre sayısı (params), işlem yükü (ops) olarak gösterilmiştir.

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Şekil 6.2 GoogleNet mimari yapısı

Eğitimlerde, eğitim katsayısı (learning rate) olarak $1e^{-4}$ ile başlandı ve her 10 epoc'da bir yarıya bölünerek azaltıldı. Eğitimlerde minimum oturma eğilimine göre 20 veya

30 epoc ilerlendi. Tüm eğitimler 5 kere tekrar edildi ve ortalama sonuçlar alındı. Kullanılan parametrik değerler Tablo 6.1’de açıklanmıştır. Eğitimlerde örnek seçimi rastgele yapılmıştır ancak fonksiyonlar arası eşit bir inceleme olması için tüm enerji fonksiyonlarına aynı sıralama ile train ve validasyon olarak ayrılmıştır. Yani birinci fonksiyon için yapılan rastgele sıralama diğer fonksiyonlarda da aynı sıralama ile uygulanmıştır.

Tablo 6.1 Veri seti eğitimlerinde kullanılan GoogleNet parametreleri

GoogleNet	Eğitim Başlangıç Oranı	Çözüm	Mini-Batch Boyutu	Max-Epochs	Eğitim Oranları ve Tekrar Sayıları
Enerji Görüntüsü (Kişi tanıma)	$1e^{-4} / 2, 5e^{-5}$ aralığında değişken	Sgdm	20	20-30 aralığında	70% x5, 50% x5, 30% x5
Enerji Görüntüsü (Hareket tanıma)	$2e^{-4} / 5e^{-5}$ aralığında değişken	Sgdm	20-30 aralığında	20-30 aralığında	70% x5, 50% x5, 30% x5
İskelet Görüntüsü	$1e^{-4} / 2, 5e^{-5}$ aralığında değişken	Sgdm	10	30	70% x5, 50% x5, 30% x5
Termal(RGB)	$1e^{-4} / 2, 5e^{-5}$ aralığında değişken	Sgdm	10	30	70% x5, 50% x5, 30% x5
Termal-3D	$1e^{-4} / 2, 5e^{-5}$ aralığında değişken	Sgdm	10	20	70% x5, 50% x5, 30% x5

Elde edilen sonuçlar deneysel sonuçlar başlığı altında sunulmuştur. Hareket enerjisi veri setleri önermiş olduğumuz μ_R fonksiyonu ve μ_{1-4} mevcut fonksiyonları ile [“%70 train / %30 validasyon”, “%50 train / %50 validasyon”, “%30 train / %70 validasyon”] olmak üzere 3 farklı oranda eğitime tabi tutulmuştur. Ayrıca termal yüz veri setleri ve iskelet veri seti de aynı oranlarda eğitime tabi tutulmuştur.

6.1.2 ResNet-18 Ağ Mimarisi

ResNet-18 ağı 78 katmandan oluşan GoogleNet’e kıyasla daha küçük bir ağ mimarisine sahiptir. Güncel birçok derin ağlara kıyasla hem zaman/performans hem de disk üzerinde kapladığı alan itibarıyla etkili bir ağ olduğu görülmektedir. ResNet-18 ağı GoogleNet’e kıyasla daha küçük bir ağ olduğu için eğitimler fine-tuned yerine tamamen eğitilecek şekilde organize edilmiştir. Kullanılan ResNet-18 mimari yapısı Şekil 6.3’te gösterilmiştir. Bu eğitimlerde sadece ResNet-18 mimarisi kullanılmış olup tamamen kullanılan veri setleri göre baştan eğitilmiştir. Pre-trained parametreler kullanılmamıştır. Kullanılan parametrik değerler Tablo 6.2’de verilmiştir.

Bu mimaride fully-connected ve softmax kısımları eğitilecek veri setinin sınıf sayısına göre yeniden düzenlenmiştir. Örnek olarak, enerji görüntüleri kişi tanıma için çıkışı etkileyen sınıf sayısı 77 olurken, aynı verilerden üretilen hareket tanıma veri setinde ise 7 sınıf olacak şekilde ağı çıkışı düzenlenmiştir.

6.1.3 MobileNetV2 Ağ Mimarisi

MobileNet-v2, 53 katman derinliğine sahip evrişimli bir sinir ağıdır. 32 filtreli FC katmanı ve 19 residual bottleneck katmanı içerir. MobileNetV2, mobil cihazlarda

Layer Name	Output Size	ResNet-18
conv1	$112 \times 112 \times 64$	$7 \times 7, 64, \text{stride } 2$
conv2_x	$56 \times 56 \times 64$	$3 \times 3 \text{ max pool, stride } 2$ $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
average pool	$1 \times 1 \times 512$	$7 \times 7 \text{ average pool}$
fully connected	1000	$512 \times 1000 \text{ fully connections}$
softmax	1000	

Şekil 6.3 ResNet-18 mimari yapısı

Tablo 6.2 ResNet-18 eğitimlerinde kullanılan parametreler

ResNet-18	Eğitim Başlangıç Oranı	Çözüm	Mini-Batch Boyutu	Max-Epochs	Eğitim Oranları ve Tekrar Sayıları
Enerji Görüntüsü (Kişi tanıma)	$1e^{-4} / 2, 5e^{-5}$ aralığında değişken	Sgdm	20	20-30 aralığında	70% x5, 50% x5, 30% x5
Enerji Görüntüsü (Hareket tanıma)	$2e^{-4} / 5e^{-5}$ aralığında değişken	Sgdm	20-40 aralığında	15	70% x5, 50% x5, 30% x5
İskelet Görüntüsü	$1e^{-4} / 2, 5e^{-5}$ aralığında değişken	Sgdm	10	30	70% x5, 50% x5, 30% x5
Termal(RGB)	$1e^{-4} / 2, 5e^{-5}$ aralığında değişken	Sgdm	10	30	70% x5, 50% x5, 30% x5
Termal-3D	$1e^{-4} / 2, 5e^{-5}$ aralığında değişken	Sgdm	10	20	70% x5, 50% x5, 30% x5

iyi performans göstermeyi amaçlayan bir ağı mimarisidir. Bu çalışmada kullanılan ağlar arasında en küçük boyutlu ağ MobileNetV2 ağıdır. ImageNet veritabanından bir milyondan fazla görüntü ile eğitilmiş 224×224 girişe sahiptir. Ağın mimari yapısı Şekil 6.4'de gösterilmiştir. MobileNetV2 ağı ResNet-18 ağına olduğu gibi bu tez içerisindeki veri setleri ile tekrar eğitilerek kullanılmıştır.

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	-

Şekil 6.4 MobileNet-V2 mimari yapısı

6.2 Eğitim Sonuçları

Yukarıda anlatılan 3 CNN ağı ile gerçekleştirilen eğitimler tablolarda 3 gruba ayrılmıştır. Birinci grupta iskelet veri seti ve termal veri setlerine ait sonuçlara yer verilmiştir. İkinci grupta mevcut fonksiyonlar ve Adaptif fonksiyon ile elde edilmiş HE görüntülerinden kişi kimliklendirme sonuçları karşılaştırılmıştır. Üçüncü grupta ise aynı HE görüntülerinden hareket tanıma sonuçları karşılaştırılmalı olarak sunulmuştur.

Bunlara ek olarak enerji görüntüsü fonksiyonlarının her bir hareket özelinde kişi tanıma için performansları karşılaştırılmalı olarak verilmiştir. Bu sayede hangi fonksiyonun hangi hareket için daha başarılı olduğu ayrıntılı olarak incelenmiştir.

6.2.1 GoogleNet Eğitim Sonuçları

GoogleNet ağının sınıflama performansının kullandığımız bu veri setleri üzerinde çok etkili olduğu görülmüştür. Sonuçlar incelendiğinde önermiş olduğumuz yöntemin mevcut fonksiyonlara kıyasla daha iyi sonuçları verdiği diğer fonksiyonların ise veri setlerine göre değişkenlik gösteren başarımlar ürettiği görülmüştür. Önermiş olduğumuz yöntemin hareket değişimine adaptif olması hareketteki değişimleri daha belirgin yaparken durağanlıkları daha fazla baskılamaktadır. En zor hareketlerden olan engel hareketinde en iyi performans adaptif TT ile elde edilmiştir. Adaptif TT kişi tanımada, %70 oranında eğitimde en iyi sonuçları alırken, %50 oranında eğitimde 2. en iyi sonucu elde etmiştir. Zorlu durum olan %30 oranında eğitim grubunda ise kişi tanıma başarısının μ_2 ve μ_1 'den sonra geldiği görülmektedir. Sonuçların ayrıntıları Tablo 6.3'te ve Tablo 6.4'te sunulmuştur.

GoogleNet ağında önerdiğimiz yöntemin başarılı sonuçlar elde etmesinde, bu ağın filtrelerinin özellik çıkartma kapasitesinin yüksek olması ve fine-tuned olarak eğitilmesinin etkili olduğu söylenebilir. Ancak birçok makale de bu tür kıyaslamalar daha sık ve az katmanlı, daha genel özellikleri çıkartabilecek ve yapılacak kıyaslamaların daha net görülmesini sağlayacak ağlar ile gerçekleştirilmiştir. Bu sebepten enerji görüntülerinin sınıflandırılmasında ve kullanılan yöntemler arasındaki farklılıkların incelenmesi için daha küçük çapta bir ağın tamamen eğitilip, fonksiyonların hareketi temsil etme üzerindeki performansı bu şekilde de incelenebilir.

Tablo 6.3'te üretilen 5 farklı veri setinin GoogleNet ağında yapılan eğitim sonuçları verilmiştir. Özellikle hareket tanıma sonuçlarında adaptif TT'nin hareketi temsil etme yeteneğinin tüm eğitim oranlarında diğer fonksiyonlardan açık ara önde olduğu görülmektedir. Aynı zamanda iskelet ve termal veri setlerinde de yüksek başarımlar oranlarında sonuçlar elde edilmiştir. Çoklu biyometri eğitimlerinde bu veri setlerinin

başarımı arttıracak etki oluşturacağı söylenebilir.

Tablo 6.3 GoogleNet ile yapılan eğitimlerin ortalama sonuçları. Farklı eğitim/validasyon oranları ve tüm veri setilerine ait ortalama sonuçlar gösterilmektedir

Eğitim Ağı	Veri Seti	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
GoogleNet-FineTuned	İskelet	93,23%	83,62%	-
“	Termal(RGB)	99,44%	99,05%	98,52%
“	Termal-3D	100%	99,95%	99,81%
Eğitim Ağı	Veri Seti (Kişi Tanıma)	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
GoogleNet-FineTuned	Enerji Gör. (μ_R)	95,55%	92,50%	88,60%
“	Enerji Gör. (μ_1)	95,17%	92,48%	88,94%
“	Enerji Gör. (μ_2)	95,43%	92,86%	89,25%
“	Enerji Gör. (μ_3)	90,39%	86,44%	81,51%
“	Enerji Gör. (μ_4)	93,54%	89,79%	86,11%
Eğitim Ağı	Veri Seti (Hareket Tanıma)	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
GoogleNet-FineTuned	Enerji Gör. (μ_R)	96,84%	98,50%	97,75%
“	Enerji Gör. (μ_1)	94,36%	97,41%	96,34%
“	Enerji Gör. (μ_2)	90,74%	95,39%	93,65%
“	Enerji Gör. (μ_3)	91,65%	95,55%	93,98%
“	Enerji Gör. (μ_4)	95,71%	97,52%	96,37%

Tablo 6.4 Kişi tanımda hareketlerin ayrıntılı başarımları ve ortalama sonuçları

	TT Yöntemleri / 5 farklı eğitim ortalaması (Tüm yöntemlerde aynı sıralama kullanılmıştır)														
	%70					%50					%30				
	μ_R (%)	μ_1 (%)	μ_2 (%)	μ_3 (%)	μ_4 (%)	μ_R (%)	μ_1 (%)	μ_2 (%)	μ_3 (%)	μ_4 (%)	μ_R (%)	μ_1 (%)	μ_2 (%)	μ_3 (%)	μ_4 (%)
Hareket	μ_R (%)	μ_1 (%)	μ_2 (%)	μ_3 (%)	μ_4 (%)	μ_R (%)	μ_1 (%)	μ_2 (%)	μ_3 (%)	μ_4 (%)	μ_R (%)	μ_1 (%)	μ_2 (%)	μ_3 (%)	μ_4 (%)
Ayak	97,04	96,30	95,70	94,96	96,59	93,87	93,42	94,58	93,07	92,89	90,60	90,92	88,51	86,48	88,32
Bilek	98,82	98,52	99,26	97,04	99,26	97,51	98,31	98,84	95,82	96,98	93,46	95,62	96,00	91,75	93,33
Daire	98,27	98,44	98,87	95,24	98,27	97,35	98,03	97,97	95,06	97,50	95,66	96,85	96,55	90,32	95,25
Eğilme	94,63	96,45	95,24	88,40	93,51	92,31	93,30	92,31	84,05	88,78	88,05	90,43	89,24	79,81	86,05
Engel	89,19	88,59	87,85	80,30	80,59	84,98	82,49	84,09	74,31	74,22	76,70	74,86	76,70	67,24	68,25
Yürüme	92,81	89,78	93,42	80,87	88,31	85,30	85,09	86,81	70,65	81,66	80,74	78,63	84,30	67,24	76,73
Z	98,10	98,10	97,66	95,93	98,27	96,16	96,73	95,43	92,11	96,52	94,99	95,25	93,47	87,76	94,84
Ort(Tüm)	95,55	95,17	95,43	90,39	93,54	92,50	92,48	92,86	86,44	89,79	88,60	88,94	89,25	81,51	86,11

İskelet veri seti Tablo 6.1'deki parametreler ile GoogleNet ağı üzerinde eğitilmiştir. Bu veri setinde sınıflar arası örnekler, HE görüntülerine kıyasla birbirlerine çok daha fazla benzemektedir. Bu nedenle, ağı eğitim ilerlemesi çok daha yavaş olmuş ve ancak 30 epoctan sonra %90'ın üzerine çıkabilmiştir. Bu veri seti farklı mini-batch değerleri ile tekrar ele alınabilir. Aynı zamanda sınıflar arası benzerliklerin çok yüksek olması bu tür verilerin sınıflandırılmasında, HE görüntülerinin aksine daha etkili özellik çıkartabilecek ağlara ihtiyaç olduğunu göstermektedir. Bunu ispatlayabilmek için farklı katmanlı ağlarda da denemeler yapılabilir.

Termal veri setinde kişilerin termal görüntülerinden 2 farklı veri seti oluşturulmuştur. Bu veri setleride Tablo 6.1'deki parametreler ile GoogleNet ağında eğitilmiştir. Bu görsellerde, yani kişi yüzlerinde net farklılıklar olduğu düşünüldüğünde, HE görüntüsü veri setinde olduğu gibi ağın sınıflama başarımı yüksek olmuştur. Nihai sonuçlara yaklaşık 10 epoc içinde yaklaşılmıştır. Termal-3D olarak ifade ettiğimiz, yüzün 3 farklı bölgesinin birer katman olarak birleştirildiği veri seti, sadece ön yüzün yer aldığı termal (RGB) veri setine göre kişi tanımada daha yüksek performans göstermiştir.

Termal-3D veri seti diğer veri setlerine kıyasla multi-biyometrik bir özelliğe sahiptir. 3 farklı açıdan yüz biyometrisi içermekle multi-instance yapıya uymaktadır. Ayrıca yan profil görselleri kulak biyometrisini de içerdiği için multi-modal biyometric sınıfına da dahil edilebilir. Veri seti oluşumu olarak sample-level ve features-level aralığında füzyon işlemi de uygulanmıştır. Eğer özellik çıkarımı sadece sınıflayıcı seviyesinde düşünülürse, veri seti sample-level füzyon'a uymaktadır. Bir diğer bakış açısında ise, termal görsellerde arka planın çıkartılması ve termal yüze ait gri seviye özelliklerinin elde edilmesi bir özellik çıkartımı olarak düşünülebilir. Böyle bir durumda ise 3 farklı yöne ait yüz görüntülerinin birleştirilmesi ile oluşturulan veri seti features level-füzyon'a dâhil olmaktadır. Bu yapısı itibariyle başarı sonuçları diğer veri setlerine göre çok daha yüksektir.

6.2.2 ResNet-18 Sonuçları

ResNet-18 ağı, ağ mimarisi bölümünde anlatıldığı gibi ve Tablo 6.2'deki parametreler ile eğitilmiştir. ResNet-18 ağında eğitimler tamamen veri setine göre yeniden yapıldığı için GoogleNet eğitimlerine kıyasla daha uzun sürmüştür. 5 veri seti için elde edilen sonuçlar Tablo 6.5'te sunulmuştur. GoogleNet ağına kıyasla daha düşük başarımlar elde edilmiştir. Bunun sebebi GoogleNet ağında fine-Tuned yapılırken bir çok parametre önceden eğitim değerlerini korumuştur. Başka bir ifade ile GoogleNet çok sayıda veri ile eğitilmiş parametrelere sahipken, ResNet-18 ağı elimizdeki sınırlı veri seti ile eğitilmiştir. Veri setimizdeki kısmi veri miktarındaki azlıkta da bu sonucun bir sebebi olabilir. Bunun haricinde hareket tanıma eğitimlerinde, %50 eğitim oranında MiniBatchSize değeri 30 olarak, %30 eğitim oranında ise MiniBatchSize 20 olarak kullanılmıştır. Initial learning rate %50 eğitim için $1e^{-4}$ iken, %30 eğitim için $2e^{-04}$ olarak başlatıldı. %50 ile %30 eğitimler arasındaki fark bu parametrik değişikliklerden kaynaklı olabilir.

Tablo 6.5'teki Resnet-18 sonuçları incelendiğinde GoogleNet sonuçlarına benzer şekilde iskelet veri setinin eğitiminde ağın zorlandığı gözlemlenmiştir. Termal veri setleri içinde Termal-3D'nin daha iyi sonuçlara ulaştığı gözlemlenmiştir. Bu sonuçlar,

Tablo 6.5 ResNet-18 ile yapılan eğitimlerin ortalama sonuçları. Farklı eğitim/validasyon oranları ve tüm veri setlerine dair ortalama sonuçlar gösterilmektedir

Eğitim Ağı	Veri Seti	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
ResNet-18 (re-trained)	İskelet	93,30%	92,18%	-
“	Termal(RGB)	97,32%	94,59%	86,68%
“	Termal-3D	99,91%	99,77%	97,86%
Eğitim Ağı	Veri Seti (Kişi Tanıma)	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
ResNet-18 (re-trained)	Enerji Gör. (μ_R)	93,86%	92,78%	85,28%
“	Enerji Gör. (μ_1)	94,74%	92,53%	86,55%
“	Enerji Gör. (μ_2)	93,37%	91,02%	84,17%
“	Enerji Gör. (μ_3)	93,98%	91,22%	84,64%
“	Enerji Gör. (μ_4)	93,22%	90,46%	83,84%
Eğitim Ağı	Veri Seti (Hareket Tanıma)	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
ResNet-18 (re-trained)	Enerji Gör. (μ_R)	99,05%	97,35%	98,27%
“	Enerji Gör. (μ_1)	98,34%	95,50%	97,02%
“	Enerji Gör. (μ_2)	98,15%	94,26%	96,54%
“	Enerji Gör. (μ_3)	98,31%	95,76%	96,60%
“	Enerji Gör. (μ_4)	98,74%	96,01%	97,05%

ResNet-18 gibi bir ağı bu tür bir veri setinde yüksek başarımlar sağladığı ve GoogleNet gibi daha büyük boyutlu bir ağa ihtiyaç olmadığı da göstermiştir. Ayrıca Termal-3D veri seti, ön yüz içeren Termal(RGB) veri setine kıyasla, zorlu eğitim durumlarında (%30) çok daha iyi sonuçlara ulaşmıştır. Termal yüz görüntülerinde multi-biyometrik özellik taşıyan bir veri setinin kişi özelliklerini daha iyi temsil ettiği söylenebilir.

HE görüntüleriyle kişi tanıma sonuçların incelendiğinde korelasyon katsayıları tabanlı μ_R ve μ_1 yöntemlerinin en iyi sonuçlara ulaştığı ve yine diğer yöntemlerinde yakın sonuçlara ulaştığı görülmektedir. %50 eğitim oranında ResNet-18 ağı daha iyi başarımlar göstermiştir. %30 eğitim oranında ise GoogleNet'den daha düşük bir başarımlar göstermiştir. Hareket tanıma sonuçlarında önermiş olduğumuz Adaptif TT diğer yöntemlere göre yine tüm eğitim oranlarında en iyi sonuca ulaşmıştır. Sonuçlar genel olarak GoogleNet ağı sonuçlarından daha yüksektir. Yani ResNet-18 ağına yapılan hareket tanıma performansı, GoogleNet ağına kıyasla daha yüksek bir çok daha hızlı bir şekilde sonuçlanmıştır. Bu durumda, hareket tanıma için enerji görüntüsü veri setinde ResNet-18 ağına daha etkili olduğu söylenebilir.

6.2.3 MobileNetV2 Eğitim Sonuçları

Üçüncü CNN modeli olan MpbileNetV2 eğitimlerinde HE görüntüleri ile kişi tanıma veri setinde, eğitimler *overfitting*'e doğru gitmiştir. Bu sebeple mobileNetV2 ağına 0.5 parametrelili *dropout layer* eklenmiştir. Ayrıca *L2Regularization* parametresi $1e^{-4}$ den $1,5e^{-4}$ 'e alınmış ve *shuffle* parametresi *once* yerine *every epoch* olacak şekilde değiştirilmiştir. Bu duruma göre yeniden yapılan eğitimlerde *overfitting* oluşumunun engellendiği ve buna bağlı olarak sınıflandırma başarımında artış görülmüştür. Sonuçlar Tablo 6.6'da sunulmuştur.

Tablo 6.6 MobileNetV2 ile yapılan eğitimlerin ortalama sonuçları. Farklı eğitim/validasyon oranları ve tüm veri setlerine ait ortalama sonuçlar gösterilmektedir

Eğitim Ağı	Veri Seti	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
MobileNetV2 (re-trained)	İskelet	95,010%	92,14%	-
"	Termal(RGB)	99,44%	98,70%	97,46%
"	Termal-3D	100%	99,92%	99,67%
Eğitim Ağı	Veri Seti (Kişi Tanıma)	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
MobileNetV2 (re-trained)	Enerji Gör. (μ_R)	98,15%	96,58%	90,21%
"	Enerji Gör. (μ_1)	98,48%	96,86%	90,17%
"	Enerji Gör. (μ_2)	98,50%	96,30%	90,20%
"	Enerji Gör. (μ_3)	97,93%	95,50%	88,02%
"	Enerji Gör. (μ_4)	97,65%	96,18%	89,22%
Eğitim Ağı	Veri Seti (Hareket Tanıma)	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
MobileNetV2 (re-trained)	Enerji Gör. (μ_R)	99,43%	99,52%	99,34%
"	Enerji Gör. (μ_1)	99,37%	99,32%	98,83%
"	Enerji Gör. (μ_2)	99,14%	99,04%	98,44%
"	Enerji Gör. (μ_3)	99,14%	99,33%	98,84%
"	Enerji Gör. (μ_4)	99,32%	99,25%	98,96%

Termal görüntülere ait sonuçlar incelendiğinde, diğer ağlarda olduğu gibi Termal-3D veri seti sadece ön yüzü içeren Termal(RGB) veri setinden daha yüksek başarımlar elde etmiştir. Özellikle düşük eğitim oranı olan %30 eğitimde de bu başarımlarını sürdürmüştür. Kişi tanıma eğitimlerinde ise %70 eğitimde μ_2 fonksiyonu, %50 eğitimde ise μ_1 fonksiyonu en iyi sonuçları vermiştir. Korelasyon fonksiyonu bu 2 eğitimde de ortalamanın üzerinde başarımlar elde etmiştir. Zorlu eğitim oranı olan %30 eğitimde ise μ_R fonksiyonu ile elde edilen enerji görüntüleri en iyi temsiliyeti göstermiştir. Bu sonuçlar, MobileNetV2 ağına önerdiğimiz yöntemin yüksek oranlarda ortalama üstünde ve zorlu eğitim koşullarında ise kişi bilgisini daha iyi temsil edebildiğini göstermektedir. Diğer yöntemler eğitim verisindeki azalmadan daha fazla etkilenirken, korelasyon katsayısı tabanlı μ_R fonksiyonunun temsiliyet gücü

eğitim oranındaki değişimlerden daha az etkilenmiştir.

Hareket tanıma sonuçlarında ise önerdiğimiz yöntem MobileNetV2 ağı üzerinde tüm eğitim oranlarından en iyi sonucu göstermiş ve hareket temsiliyeti olarak diğer fonksiyonlardan çok daha iyi sonuç elde etmiştir. Bunun yanında eğitim oranlarındaki azalmadan da en az etkilenen fonksiyon olmuştur. Hareket değişimine adaptif olabilen μ_R fonksiyonunda, %70 eğitimden %30 eğitime geçildiğinde görülen değişim %0.09 olmuştur. Bu oran diğer fonksiyonlar için sırasıyla μ_1 için %0.54, μ_2 için %0.70, μ_3 için %0.30 ve μ_4 için %0.38 şeklindedir.

6.3 Termal-3D Veri Setinde Eksik Veri Testi

Termal-3D bir önceki bölümde Şekil 5.13'te gösterildiği gibi 3 farklı açıdan yüz görüntülerinin birleştirilmesi ile oluşturulmuş bir veri setidir.

Bu veri setinde enrollment süreci, başın sağdan-sola 180° çevrilmesi ve video olarak kayıt edilmesi ile tamamlanmıştır. Ancak bu tür bir sistem enrollment sürecinden farklı olarak iki farklı şekilde de çalışabilir. Birincisi, bu çalışmada anlatıldığı gibi denekten veri setinin kaynağındaki gibi başını sağdan sola tam çevrim yapması istenir ve bu hareket video olarak kayıt edilebilir. Bu video karelerinden sağ-sol-ileri yüz görüntüleri yakalanıp Termal-3D görüntüsü oluşturulabilir. İkinci yöntemde ise denek sağ-sol-ileri olarak kamera karşısında tek karelik resim örneği verebilir. Bu 3 görüntü ile Termal-3D görüntüsü oluşturulabilir. Ancak bu iki yöntemde de 3 farklı açığa ait yüz görüntülerinde veri kaybı olabileceği durumlar göz önüne alınarak eksik veriler ile test denemeleri yapılmış ve ağların performansları incelenmiştir. Sonuçları sırası ile Tablo 6.7 ve Tablo 6.8'de verilmektedir.

Tablo 6.7 3 farklı yüz açılarından sadece birinin olması durumu

	Only Left			Only Front			Only Right		
									
Eğitim Oranları	70%	50%	30%	70%	50%	30%	70%	50%	30%
GoogleNet	35,41%	35,64%	32,45%	14,03%	13,56%	12,08%	28,40%	28,00%	31,09%
ResNet-18	4,46%	2,65%	4,99%	2,29%	1,90%	2,26%	3,46%	4,03%	2,84%
MobileNetV2	34,63%	31,71%	24,02%	13,07%	11,95%	9,00%	7,79%	5,35%	5,79%

Tablo 6.8 3 farklı yüz açılarından herhangi 2 tanesinin olması durumu

	Only Front-Left			Only Front-Right			Only Left-Right		
									
Eğitim Oranları	70%	50%	30%	70%	50%	30%	70%	50%	30%
GoogleNet	82,12%	80,65%	76,29%	73,91%	49,51%	48,13%	76,80%	76,52%	73,91%
ResNet-18	10,30%	5,92%	3,86%	9,78%	6,88%	6,73%	23,29%	21,84%	12,15%
MobileNetV2	95,37%	83,35%	82,69%	47,06%	43,64%	39,37%	82,90%	81,66%	62,67%

3 farklı yüz açısının birleşimi ile oluşturulan veri seti ile eğitilmiş ağlar üzerinde tablolarda da görüldüğü gibi, farklı açılardan bir veya bir kaçının kayıplı olması durumunda ne gibi sonuçlar elde edileceği incelenmiştir.

GoogleNet ağında bir adet yüz açısının yer aldığı durumlarda, başarımlar çok ciddi oranda düşmüştür. 2 farklı açının yer aldığı durumlarda ise daha kıyaslanabilir sonuçlar elde edilmiştir. Özellikle %70 eğitim grubunda sonuçlar daha değerlendirilebilir seviyededir. 2 farklı açının olduğu durumlar içinde ise en etkili sonuç, yüzün ön kısmı ve sol kısmının yer aldığı Front-Left grubundan elde edilmiştir.

ResNet-18 ağında ise sonuçlar kıyaslama yapılamayacak kadar düşük çıkmıştır. Bu durum ResNet-18 ağının veri kaybında kullanım açısından uygun olmadığını göstermektedir.

Mobinet-V2 ağı için sonuçlar incelendiğinde, özellikle left yüz görselinin yer aldığı 1. katman görüntüsü ağı başarımında oldukça önemli bir konumdadır. 2. katman ise başarımında 1. katmana kıyasla az etki etmekte, 3. katman ise en az katkıyı sağlamaktadır. Bu sebeple en iyi sonuçlar front-left grubunda elde edilmiş, hatta %70 eğitim grubunda %95'in üzerinde bir başarımla GoogleNet'den daha iyi sonuçlara ulaşılmıştır. Burada GoogleNet görüntü katmanları arasında 1 ve 2. katmanlardan daha iyi sonuçlar elde etsede genel olarak daha dengeli bir şekilde özellik çıkarttığı söylenebilir. ResNet-18'de ise 1. ve 3. katmanlardan daha fazla özellik çıkartıldığı ve başarımın bu katmanlardaki bilgilere göre şekillendiği belirtilebilir.

Tablo 6.7'nin sonuçları iki farklı yüz açısı kaybında sistemin artık çalışamaz derecede etkilendiğini göstermektedir. Ancak Tablo 6.8'in sonuçları, verinin eksik geldiği katmana veya eğitim yapılan ağı mimarisine göre sistemin hala kullanılabilir sonuçlar ürettiğini göstermektedir. Özellikle GoogleNet ve MobileNetV2 ağlarında 1.ve 2. katman görüntüleri çok daha belirgin başarımlar sağlamıştır. ResNet-18 ise tüm durumlarda veri kaybına karşı kabul edilebilir bir sonuç üretememiştir.

6.3.1 Aynalanmış Görüntü (Mirrored-Face) Eklenmiş Eğitimler

Sağ veya sol görüntüsü eksik olan görüntülerde, sağ veya sol yüz görüntülerinin aynalanmış (mirrored-face) görüntüleri eklenerek veri kaybı bir nebze olsa giderilmek istenmiştir. Bu yöntem ile kayıpsız görüntüler ile eğitilmiş ağlarda testler yapılmış ve katkı sağlayıp sağlamadığı konusunda sonuçlar kıyaslanmıştır. Elde edilen sonuçlar Tablo 6.9 ve Tablo 6.10'da sunulmuştur.

Tablo 6.9'un sol tarafına bakıldığında, 1. katmanda oluşan (sol yüz) eksikliğini gidermek için sağ yüzün ayna görüntüsü 1. katmana eklenmiştir. Sonuçları

incelendiğinde, ayna görüntüsü eklemenin genel itibari ile daha iyi sonuçlar verdiği görülmektedir. Bu denemeler sadece sağ yüzün yer alıp sol yüzün yer almadığı “Only Front-Right” ve “Only Right” veri grupları üzerinde test edilmiştir.

Tablo 6.9 1. katmana ayna görüntüsü eklenmiş kayıplı durum kıyaslaması

Eğitim Oranları	Only Front-Right						Only Right					
	70%	/Mirror	50%	/Mirror	30%	/Mirror	70%	/Mirror	50%	/Mirror	30%	/Mirror
GoogleNet	73,91%	69,22%	49,51%	69,09%	48,13%	64,71%	28,40%	42,20%	28,00%	40,68%	31,09%	39,17%
ResNet-18	9,78%	11,77%	6,88%	10,10%	6,73%	8,26%	3,46%	4,63%	4,03%	3,17%	2,84%	3,25%
MobileNetV2	47,06%	74,55%	43,64%	71,56%	39,37%	64,08%	7,79%	26,32%	5,35%	25,63%	5,79%	18,44%

Tablo 6.10 3. katmana ayna görüntüsü eklenmiş kayıplı durum kıyaslaması

Eğitim Oranları	Only Front-Left						Only Left					
	70%	/Mirror	50%	/Mirror	30%	/Mirror	70%	/Mirror	50%	/Mirror	30%	/Mirror
GoogleNet	82,12%	68,31%	80,65%	68,68%	76,29%	66,03%	35,41%	44,71%	35,64%	42,08%	32,45%	41,48%
ResNet-18	10,30%	16,23%	5,92%	11,44%	3,86%	8,26%	4,46%	4,46%	2,65%	3,79%	4,99%	3,56%
MobileNetV2	95,37%	76,71%	83,35%	72,75%	82,69%	65,07%	34,63%	25,50%	31,71%	26,21%	24,02%	18,55%

Benzer işlemler 1. katmanda yer alan sol yüzün yer aldığı ancak 3. katmanda yer alan sağ yüzün yer almadığı “Only Left” ve “Only Front-Left” durumları için de test edilmiştir. 3. katmana ayna görüntüleri eklenmiş verilerin test edildiği bu veri setlerinde ise çoğunlukla başarımın düştüğü görülmüştür. Elde edilen sonuçlara göre özellikle GoogleNet ve MobileNetV2 ağlarında başarımın azaldığı görülmektedir. ResNet-18 ağında ise sonuçlar olumlu yönde değişmiştir. Genel olarak başarımın yüksek olduğu durumlar orijinal olmayan görüntülerin eklenmesinden olumsuz etkilenmiştir. Bu durum için ağların oldukça ayırt edici özellikler çıkartabildiği ve ufak farklılıklara karşı hassas olabileceği söylenebilir. Bunun yanında düşük başarımlı sonuçlarda artış olması ise veri yetersizliği durumlarında kullanılabilecek bir yöntem olabileceğini göstermektedir.

6.3.2 Tüm Yüz Açılarının Bir Arada Eğitilmesi

Eksik veri durumlarının ayna görüntüsü ile giderilmeye çalışıldığı eğitimlerin haricinde, 3. bir eğitim ise tüm yüz açıların (sağ, sol ve ön yüz) bir arada kullanılması ile yapılmıştır. Bu eğitimde kullanılan veri seti "3-Side Termal RGB" olarak isimlendirilmiştir. Bu veri seti eğitiminde tüm yüz açıları kullanılmış ancak testleri tek yüz açısı ile yapılmıştır. Yani bir kişiye ait eğitim sınıfında 3 farklı yüz açısı yer alırken, aynı sınıfın test kümesinde 3 açıdan herhangi biri yer almıştır. Tüm açıların ortalama sonucu alınarak başarım sonucu elde edilmiştir. Eğitim sonuçları Tablo 6.11’de kıyaslamalı olarak sunulmuştur. Sadece ön yüz görsellerinin kullanıldığı “Termal RGB” veri setine kıyasla başarımının arttığı ve zorlu eğitim oranlarında daha iyi sonuçlar verdiği de görülmektedir. Ancak sağ sol ve ön yüzün birleştirilmesi ile

oluşturmuş olduğumuz “Termal-3D” veri setinin başarımı hem farklı ağlarda hem de farklı eğitim oranlarında hala en iyi sonucu vermiştir.

Tablo 6.11 Termal veri setinin farklı içeriklere göre eğitilmesi ile elde edilen sonuçların kıyaslandığı tablo. Termal(RGB): sadece ön yüz açısı ile yapılan eğitim sonuçlarını, 3-Side Termal RGB: yüzün 3 farklı açısını içeren veri seti ile yapılan eğitim sonuçlarını, Termal-3D: yüzün 3 farklı açısının birer katmana konularak tek bir görüntü elde edilmesi ile yapılan eğitimlerin sonucunu göstermektedir

Eğitim Ağı	Veri Seti	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
GoogleNet-FineTuned	Termal(RGB)	99,44%	99,05%	98,52%
/Yeni GoogleNet-FineTuned	3-Side Termal RGB	99,93%	99,92%	99,79%
GoogleNet-FineTuned	Termal-3D	100%	99,95%	99,81%
ResNet-18 (re-trained)	Termal(RGB)	97,32%	94,59%	86,68%
/Yeni ResNet-18 (re-trained)	3-Side Termal RGB	99,38%	98,62%	94,73%
ResNet-18 (re-trained)	Termal-3D	99,91%	99,77%	97,66%
MobileNetV2 (re-trained)	Termal(RGB)	99,44%	98,70%	97,46%
/Yeni MobileNetV2 (re-trained)	3-Side Termal RGB	99,97%	99,90%	99,51%
MobileNetV2 (re-trained)	Termal-3D	100%	99,92%	99,67%

6.4 Eklemler Arası Uzaklık Verilerinin Görüntü Yerine XY Koordinatları Olarak 1D Formatta Konvolüsyonel Bir Ağa Uygulanması (1D-SkeletonNet)

GoogleNet, ResNet-18 ve MobileNet ağları üzerinde iskelet görselleri ile eğitim yapılmıştır. Diğer veri türlerine oranla düşük başarımlar elde edilirken, eğitim sürecinde ise çok daha uzun zaman harcanmıştır (her bir yüzdelik için yaklaşık 22 saat). İskelet görüntülerinde, veri yoğunluğunun az olması ve görsellerin benzerlik göstermesi bu durumun bir nedeni olarak düşünülmüştür. Bu sebeple, sadece eklemlerin koordinat verileri kullanılarak 1D verilerin eğitilmesinin etkisi üzerinde durulmuştur. Bu kapsamda 20 farklı eklem noktalarına ait XY koordinat verileri alınmış ve sınıflama için 1D bir konvolüsyonel ağ üzerinde eğitilmiştir. Elde edilen sonuçlar Tablo 6.12’de sunulmuştur.

Tablo 6.12 İskelet verilerinin XY koordinatları olarak kullanılması sonucu elde edilen 1D-SkeletonNet eğitim sonuçları

Eğitim Oranları	1	2	3	4	5	Ortalama / std
%70	98.70	96.97	96.97	96.97	99.57	97.84 / 1.22
%50	97.40	98.96	98.44	97.92	98.18	98.18 / 0.58
%30	95.92	95.73	95.92	95.18	95.18	95.58 / 0.38

Tablo 6.12’deki sonuçlar incelendiğinde, daha önce 3 farklı ağda görsel verilerin eğitilmesine oranla çok daha iyi sonuçlara ulaşıldığı görülmektedir. Önceki eğitimlerdeki gibi 5 kez eğitim rastgele seçim ile yapılmış ve ortalaması alınmıştır. Sonuçlar sadece daha iyi çıkmakla kalmayıp her bir yüzdelik için toplam eğitim süresi

22 saatten 5 dk gibi bir süreye düşürülmüştür. Bunun bir sebebi kullanılan verinin azalmış olması, bir diğer sebebi ise kullanılan 1D konvolüsyonel ağın daha basit tasarlanmış olmasıdır. Bu eğitimler için aşağıdaki gibi bir ağ oluşturulmuştur.

Tablo 6.13 İskelet verilerinin koordinat olarak eğitiminin yapıldığı ağ yapısı

Layers	Size	Output
Input Layer	[40 1 1]	
Conv	Filter – 3x1, 16	40x1x16
Relu	40x1x16	
Fully Conn-1	1x1x1024	1024x640
Fully Conn-2	1x1x1024	1024x1024
Fully Conn-3	1x1x77	77x1024
Softmax	1x1x77	
ClassOutput	Classification Output	

Tablo 6.13'te sunulan ağda girişe uygulanan 1D eklem koordinatları 1 konvolüsyon ve 1 relu katmanından geçtikten sonra 1024 nöron içeren 2 gizli katman ve 77 sınıfa göre ayarlanan bir çıkış katmanı içeren fully connected katmanlarına bağlanmıştır. Bu sonuçlar doğrultusunda çoklu biyometrik eğitimlerde iskelet verilerinin görüntü yerine koordinat olarak kullanılması tercih edilmiştir.

6.5 Çoklu Biyometrik Eğitimler

6.5.1 Kişi Tanıma Füzyon Eğitimleri

Elde edilen veri setleri ile başarıyı arttırmak ve sonuçlarını incelemek amacıyla çoklu biyometrik eğitimler de yapılmıştır. Çoklu biyometrik eğitimlere ilk olarak termal görüntüler ve HE kişi tanıma veri setleri ile başlanmıştır. Bu sayede kişi tanıma performansı ve doğruluğunun artırılması hedeflenmiştir.

Termal görüntülerden türetilen ve en yüksek başarıyı elde eden Termal-3D veri seti ile kişi tanıma için kullanılan enerji görüntüsü veri setleri arasında skor seviyesinde füzyon işlemi gerçekleştirilmiştir. Skor seviyesi füzyonu için Z-Score normalizasyonu kullanılmıştır.

$$Z = \left(\frac{x - \mu}{\sigma} \right) \quad (6.1)$$

Eşitlikte Z normalize skor değerini, x normalizasyon öncesi skor değerlerini, μ ortalama değerleri, σ standart sapma değerini ifade etmektedir. Bu formül ile elde

edilen ortalamanın üzerindeki skorlar pozitif, altındaki skorlar ise negatif olarak ele alınır. Her bir veri setinden elde edilen skor matrisi 0.5 katsayısı çarpılmış ve sonuçlar toplanmıştır. Elde edilen skor matrisi denklem 6.1'e göre normalize edilmiştir.

Tablo 6.14 Kişi tanıma için Termal-3D ve enerji görüntüsü veri setlerinin için skor level füzyon sonuçları

Eğitim Ağı Kişi Tanıma	Termal-3D ve Enerji Görüntüsü Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
GoogleNet-FineTuned	Enerji Gör. (μ_R)	100	100	99,96
"	Enerji Gör. (μ_1)	100	100	99,97
"	Enerji Gör. (μ_2)	100	100	99,99
"	Enerji Gör. (μ_3)	100	99,98	99,96
"	Enerji Gör. (μ_4)	100	99,99	99,97
Eğitim Ağı Kişi Tanıma	Termal-3D ve Enerji Görüntüsü Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
ResNet-18 (re-trained)	Enerji Gör. (μ_R)	100	99,99	99,62
"	Enerji Gör. (μ_1)	99,98	99,98	99,68
"	Enerji Gör. (μ_2)	100	99,99	99,51
"	Enerji Gör. (μ_3)	100	99,99	99,72
"	Enerji Gör. (μ_4)	100	99,98	99,50
Eğitim Ağı Kişi Tanıma	Termal-3D ve Enerji Görüntüsü Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
MobileNetV2 (re-trained)	Enerji Gör. (μ_R)	100	100	99,98
"	Enerji Gör. (μ_1)	100	100	99,95
"	Enerji Gör. (μ_2)	100	100	99,99
"	Enerji Gör. (μ_3)	100	100	99,97
"	Enerji Gör. (μ_4)	100	100	99,97

Tablo 6.14'te kişi tanıma için Termal-3D ve HE görüntülerinin skor seviye füzyon eğitim sonuçları verilmiştir. Sonuçlar beklendiği gibi her iki veri setinin bireysel başarımlarından daha yüksek çıkmıştır. GoogleNet ve MobileNetV2 için, ilk iki ölçekte %100 çıkarken %30 eğitim - %70 test sonuçlarında da neredeyse aynı sonuçlara ulaşmıştır. Enerji görüntüsü yöntemlerinin neredeyse tamamında benzer sonuçları elde edilmiştir. Bu kadar yakın sonuçlar için hangi yöntemin daha iyi olduğunu söylemek pek mümkün değildir. Tablo 6.15'te ise kişi tanımada kullanılan her bir hareket sınıfının başarımları ayrı ayrı verilmiştir.

İkinci olarak Tablo 6.16 ve Tablo 6.17'de sunulan enerji görüntüsü ve 1D iskelet verileri ile yapılan çoklu biyometrik eğitim sonuçları yer almaktadır. Füzyon işleminin başarımı arttırdığı ancak 5 yöntem arasında ayırt edilebilir ciddi bir fark olmadığı görülmektedir. Bir önceki tabloda yer alan Termal-EIs füzyonunun daha iyi olduğu görülmüştür. Özellikle zorlu şartlar olarak değerlendirilen %30 veri ile eğitimlerde Termal görüntü kullanılarak yapılan füzyonda başarımın çok daha yüksek çıktığı görülmüştür. Termal görüntüler ile yapılan eğitimlere kıyasla EIs ve 1D iskelet verileri ile kişi kimliklendirme zorlu şartlarda daha düşük başarım elde etmiştir. Bu sonuçlar, termal yüz verilerinin kişi tanıma üzerindeki etkisinin çok daha fazla olduğunu

Tablo 6.15 Kişi Tanımda Termal-3D ve enerji görüntüsü füzyon sonuçları her bir hareket için ayrıntılı olarak verilmiştir (Tüm ağların sonuçları sırayla eklenmiştir)

ResNet-18	TT Yöntemleri / 5 farklı eğitim ortalaması (Tüm yöntemlerde aynı sıralama kullanılmıştır)														
	%70					%50					%30				
Hareket	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)
Ayak	100	100	100	100	100	100	100	100	100	100	99,75	99,75	99,68	99,81	99,75
Bilek	100	100	100	100	100	100	100	100	100	100	100	99,94	100	100	99,81
Daire	100	100	100	100	100	100	100	100	100	100	99,93	99,89	99,85	99,93	99,74
Eğilme	100	100	100	100	100	100	100	100	100	100	99,82	99,93	99,89	99,89	99,85
Engel	100	99,85	100	100	100	99,91	99,91	100	100	99,91	99,17	99,56	98,73	99,56	98,67
Yürüme	100	100	100	100	100	100	99,95	99,95	99,95	99,95	98,81	99,78	98,44	98,99	98,78
Z	100	100	100	100	100	100	100	100	100	100	99,89	99,96	99,96	99,85	99,93
Ort(Tüm)	100	99,98	100	100	100	99,99	99,98	99,99	99,99	99,98	99,62	99,68	99,51	99,72	99,50
GoogleNet	TT Yöntemleri / 5 farklı eğitim ortalaması (Tüm yöntemlerde aynı sıralama kullanılmıştır)														
	%70					%50					%30				
Hareket	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)
Ayak	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Bilek	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Daire	100	100	100	100	100	100	100	100	100	100	99,96	99,96	99,96	99,93	99,96
Eğilme	100	100	100	100	100	100	100	100	100	100	100	99,96	100	99,96	100
Engel	100	100	100	100	100	100	100	100	99,91	100	99,87	100	100	100	100
Yürüme	100	100	100	100	100	100	100	100	99,95	99,95	99,85	99,89	99,96	99,85	99,85
Z	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99,96
Ort(Tüm)	100	100	100	100	100	100	100	100	99,98	99,99	99,96	99,97	99,99	99,96	99,97
MobileNetV2	TT Yöntemleri / 5 farklı eğitim ortalaması (Tüm yöntemlerde aynı sıralama kullanılmıştır)														
	%70					%50					%30				
Hareket	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)
Ayak	100	100	100	100	100	100	100	100	100	100	100	99,94	99,94	100	100
Bilek	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Daire	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Eğilme	100	100	100	100	100	100	100	100	100	100	100	99,96	100	100	100
Engel	100	100	100	100	100	100	100	100	100	100	99,87	99,81	100	99,87	99,87
Yürüme	100	100	100	100	100	100	100	100	100	100	100	100	99,96	99,93	99,93
Z	100	100	100	100	100	100	100	100	100	100	100	99,96	100	100	100
Ort(Tüm)	100	100	100	100	100	100	100	100	100	100	99,98	99,95	99,99	99,97	99,97

göstermektedir.

Tablo 6.16 Kişi tanıma için 1D iskelet ve enerji görüntüsü veri setlerinin için skor level füzyon sonuçları

Eğitim Ağları Kişi Tanıma	Enerji Görüntüsü 1D İskelet Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
GoogleNet-Fine Tuned	Enerji Gör. (μ_R)	99.68	99.70	97.80
	Enerji Gör. (μ_1)	99.75	99.76	97.76
	Enerji Gör. (μ_2)	99.74	99.77	97.71
	Enerji Gör. (μ_3)	99.60	99.66	97.49
	Enerji Gör. (μ_4)	99.73	99.67	97.56
Eğitim Ağları Kişi Tanıma	Enerji Görüntüsü 1D İskelet Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
ResNet-18 (re-trained)	Enerji Gör. (μ_R)	98.92	98.74	96.50
	Enerji Gör. (μ_1)	98.92	98.74	96.49
	Enerji Gör. (μ_2)	98.96	98.70	96.45
	Enerji Gör. (μ_3)	98.87	98.74	96.42
	Enerji Gör. (μ_4)	98.94	98.73	96.50
Eğitim Ağları Kişi Tanıma	Enerji Görüntüsü 1D İskelet Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
MobileNetV2 (re-trained)	Enerji Gör. (μ_R)	99.61	99.43	97.23
	Enerji Gör. (μ_1)	99.58	99.48	97.24
	Enerji Gör. (μ_2)	99.60	99.52	97.21
	Enerji Gör. (μ_3)	99.57	99.39	97.19
	Enerji Gör. (μ_4)	99.58	99.49	97.20

Üçüncü olarak, Termal-3D veri setiyle eğitilmiş ağların çıktıları ile iskelet veri setine

Tablo 6.17 Kişi Tanımda her bir hareket için 1D iskelet ve enerji görüntüsü füzyon sonuçları (Tüm ağlar sırayla verilmiştir)

ResNet-18	TT Yöntemleri / 5 farklı eğitim ortalaması (Tüm yöntemlerde aynı sıralama kullanılmıştır)														
	%70					%50					%30				
Hareket	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)
Ayak	99.56	99.41	99.56	99.56	99.41	99.29	99.29	99.29	99.29	99.29	97.02	97.27	97.02	97.27	96.89
Bilek	99.70	99.85	99.85	99.70	99.85	99.38	99.47	99.38	99.47	99.47	97.59	97.46	97.46	97.02	97.46
Daire	98.79	98.70	98.70	98.53	98.70	98.39	98.39	98.44	98.34	98.39	96.18	96.07	96.10	96.10	96.14
Eğilme	98.27	98.18	98.27	98.27	98.35	98.34	98.29	98.23	98.29	98.29	96.03	96.03	96.10	96.03	96.10
Engel	99.56	99.56	99.41	99.56	99.56	99.29	99.11	99.02	99.20	99.02	96.76	96.83	96.63	96.76	97.14
Yürüme	98.10	98.10	98.10	97.92	98.10	98.18	98.23	98.18	98.23	98.18	95.70	95.62	95.70	95.62	95.66
Z	98.44	98.61	98.87	98.53	98.61	98.34	98.39	98.34	98.34	98.44	96.25	96.18	96.14	96.14	96.10
Mean(Tüm)	98.92	98.92	98.96	98.87	98.94	98.74	98.74	98.70	98.74	98.73	96.50	96.49	96.45	96.42	96.50
GoogleNet	TT Yöntemleri / 5 farklı eğitim ortalaması (Tüm yöntemlerde aynı sıralama kullanılmıştır)														
	%70					%50					%30				
Hareket	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)
Ayak	100	100	100	99.70	100	99.91	99.73	99.91	99.64	99.64	98.41	97.78	97.97	97.84	98.22
Bilek	100	100	100	99.85	100	99.91	100	100	99.91	99.91	99.05	99.37	98.92	98.86	98.92
Daire	99.74	99.83	99.57	99.65	99.83	99.74	99.84	99.84	99.84	99.90	97.74	98.00	98.14	97.44	97.77
Eğilme	99.39	99.57	99.65	99.31	99.31	99.48	99.58	99.48	99.69	99.69	97.22	96.99	96.85	96.73	97.03
Engel	99.85	100	100	99.70	100	99.64	99.64	99.82	99.47	99.29	97.71	97.40	97.40	97.14	96.95
Yürüme	99.22	99.39	99.39	99.48	99.48	99.53	99.64	99.48	99.38	99.53	96.85	96.99	97.22	96.92	96.66
Z	99.57	99.48	99.57	99.48	99.48	99.69	99.84	99.84	99.69	99.74	97.63	97.77	97.48	97.48	97.37
Mean(Tüm)	99.68	99.75	99.74	99.60	99.73	99.70	99.76	99.77	99.66	99.67	97.80	97.76	97.71	97.49	97.56
MobilenetV2	TT Yöntemleri / 5 farklı eğitim ortalaması (Tüm yöntemlerde aynı sıralama kullanılmıştır)														
	%70					%50					%30				
Hareket	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)	UR(%)	U1(%)	U2(%)	U3(%)	U4(%)
Ayak	100	100	100	99.85	100	99.47	99.56	99.64	99.47	99.47	97.78	97.65	97.52	98.03	97.90
Bilek	99.85	100	99.85	100	100	99.82	99.91	99.91	99.91	100	98.41	98.48	98.41	98.22	98.67
Daire	99.57	99.57	99.57	99.48	99.57	99.64	99.74	99.74	99.58	99.58	97.66	97.40	97.40	97.51	97.37
Eğilme	99.05	99.22	99.22	99.22	99.22	99.06	99.17	99.06	99.22	99.12	96.40	96.59	96.51	96.40	96.55
Engel	100	99.85	100	100	100	99.64	99.56	99.73	99.47	99.73	96.95	97.14	97.02	97.14	96.89
Yürüme	99.22	98.96	99.13	99.05	98.96	98.96	98.96	99.06	98.91	99.01	96.36	96.18	96.47	96.10	96.25
Z	99.57	99.48	99.39	99.39	99.31	99.43	99.48	99.48	99.17	99.53	97.07	97.25	97.14	96.88	96.77
Mean(Tüm)	99.61	99.58	99.60	99.57	99.58	99.43	99.48	99.52	99.39	99.49	97.23	97.24	97.21	97.19	97.20

ait 20 eklem x-y koordinatlarını içeren 1D-iskelet verilerinin eğitildiği ağı çıktılar, skor seviye füzyon olarak ele alınmış ve sonuçları Tablo 6.18’de sunulmuştur. Kişi tanımanın son füzyon tablosunda ilk olarak GoogleNet ağı ve 1D skeleton ağı sonuçlarının füzyon sonuçları ardından Resnet-18 ağı ve 1D skeleton ağı füzyon sonuçları ve en son olarak MobilenetV2 ağı ve 1D skeleton ağı sonuçlarının füzyon sonuçları verilmiştir. Füzyon sonuçları iskelet veri seti bazında olumlu bir artış göstermiştir.

Tablo 6.18 1D-iskelet veri seti ve Termal-3D veri setlerinin skor level füzyon başarımları sonuçları

Füzyon: 1D-İskelet Termal-3D Kişi tanıma			
Eğitim Ağı	5 farklı eğitim ortalaması		
GoogleNet(Fine-Tuned)	%70 Eğitim	%50 Eğitim	%30 Eğitim
1D-SkeletonNet	99,83	99,95	99,60
Resnet-18 (re-trained)	%70 Eğitim	%50 Eğitim	%30 Eğitim
1D-SkeletonNet	99,31	99,19	96,62
MobileNetV2 (re-trained)	%70 Eğitim	%50 Eğitim	%30 Eğitim
1D-SkeletonNet	99,74	99,64	98,52

6.5.2 Hareket Tanıma Füzyon Sonuçları

Hareket tanıma için ağlar arası füzyon başarımları incelenmiştir. Bu eğitimlerde GoogleNet, Resnet-18 ve MobileNetV2 ağlarında elde edilen hareket tanıma skorları Z-score yöntemine göre füzyon edilmiştir. Bu sayede multi-network bir füzyon

işleminin başarımı incelenmiştir. Tablo 6.19’da bu ağların füzyon başarımları verilmiştir.

Bireysel ağ başarımları ve çaprazlama füzyon sonuçları incelendiğinde, füzyon sonrası başarımda önemli bir artış olduğu, özellikle GoogleNet ve Resnet ağlarının sonuçlarında önemli oranda iyileşme olduğu görülmektedir. Füzyon sonuçlarının tamamında adaptif TT fonksiyonu ile üretilen enerji görüntüleri tüm foksasyonlar içinde en iyi sonucu elde etmiştir. Tüm bu sonuçlar, önermiş olduğumuz bu yöntem hareket tanımada en zorlu eğitim durumlarında ve farklı füzyon koşullarında hareket tanıma için önceki yöntemlere göre çok daha başarılı olduğunu göstermiştir.

Tablo 6.19 Eğitim ağları arasında hareket tanıma için füzyon başarımları sonuçları

Eğitim Ağları	Hareket Tanıma Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
GoogleNet-MobileNet Füzyon	Enerji Gör. (μ_R)	99.80	99.63	99.48
	Enerji Gör. (μ_1)	99.53	99.49	98.99
	Enerji Gör. (μ_2)	99.19	99.06	98.40
	Enerji Gör. (μ_3)	99.40	99.22	98.64
	Enerji Gör. (μ_4)	99.70	99.55	99.23
Eğitim Ağları	Hareket Tanıma Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
ResNet-18 - GoogleNet Füzyon	Enerji Gör. (μ_R)	99.61	99.10	99.18
	Enerji Gör. (μ_1)	99.31	98.32	98.31
	Enerji Gör. (μ_2)	98.87	97.26	97.42
	Enerji Gör. (μ_3)	98.50	97.31	97.41
	Enerji Gör. (μ_4)	99.44	98.49	98.47
Eğitim Ağları	Hareket Tanıma Füzyon	5 farklı eğitim ortalaması		
		%70 Eğitim	%50 Eğitim	%30 Eğitim
MobileNetV2-ResNet-18 Füzyon	Enerji Gör. (μ_R)	99.73	99.65	99.59
	Enerji Gör. (μ_1)	99.50	99.32	99.18
	Enerji Gör. (μ_2)	99.43	98.97	98.88
	Enerji Gör. (μ_3)	99.25	99.42	99.18
	Enerji Gör. (μ_4)	99.52	99.41	99.25

7 SONUÇ VE ÖNERİLER

Makine öğrenimi modelleri, kişi tanıma ve hareket tanıma gibi yüksek boyutlu problemler için yüksek performans göstermeleri nedeniyle bilim ve endüstride standart araçlar haline gelmiştir. Bu nedenle modeller ve ağlar birçok alanda uygulamaya hazır olarak sunulmaktadır. Özellikle hareket tanıma ve biyometrik kişi kimliklendirme alanında bir çok ağ modeli sunulmaktadır. Biyometrik sistemlerde ise son yıllarda temassız ve kişinin belirli bir şarta bağlı kalma eğilimi azaltılmaktadır. Özellikle Covid-19 döneminde temas etme problemleri göz önüne alındığında parmak izi, avuç içi, yüzün yaklaştırıldığı yüz ve iris tanıma gibi sistemler kullanım dışı kalmıştır. Bunların yerine uzaktan görüntüleme, hareket analizine ve termal görüntülere yönelik yeni sistemler kurulmuştur. Bu tür durumlara hızlı adaptasyon için yeni ağlar yerine var olan ağlarda bu verilerin girdi olarak kullanılması eğitim ve test süreçlerini daha pratik hale getirmiştir. Bu tür uygulamalarda, girdinin elde edilen performans üzerindeki etkisi giderek daha önemli hale gelmektedir. Bunu göstermek için, enerji görüntüsü elde etmede kullanılan zamansal şablonlara yeni bir yaklaşım önerdik ve bunu hareket dizilerinin korelasyon katsayılarına göre oluşturduk. Önerilen μ_R fonksiyonu ile hareket dizilerinin korelasyon katsayıları hesaplandı ve hareketteki varyasyon aralıkları hızlı ve basit bir şekilde belirlendi. Bu sayede önceki fonksiyonlardan farklı olarak, hareketin belirli aralıkları değil, tüm değişimleri dikkate alabilen bir enerji görüntüsü oluşturulmuştur. Bu yaklaşım, enerji görüntüsünün baz alındığı hareketteki değişiklikleri daha iyi temsil etme yeteneğini artırdı. Yöntemimiz, herkese açık veri kümeleri kullanılarak farklı makine öğrenimi ve CNN modelleri altında test edilmiş ve sağlam sonuçlar vermiştir.

Eğitim ağlarının farklı varyasyonlarından elde ettiğimiz sonuçlar, yöntemimizin sınıflandırma ağlarından bağımsız olarak önceki fonksiyonlardan daha iyi performans sağladığını ve hareketleri daha iyi temsil edebildiğini göstermiştir. Eğitim sürecinde, ayarlanmış bir makine öğrenimi modelini veya önceden eğitilmiş mevcut CNN modellerini kullanmakta özgür davrandık. Bu sayede ileride bu yöntemi kullanmak isteyen araştırmacılar için bu fonksiyonu farklı ağlarda kullanmak kolaylaşmaktadır.

Belirli aralıkları vurgulayan mevcut yaklaşımlar dikkate alındığında, sağlamlıkları ve genelleme kapasiteleri sınırlı görünmektedir. Belirli aralıkların kullanılması, verideki fark oluşturan bilgi kaybına neden olacaktır. Ancak önerdiğimiz yöntem, grafik işlem birimi (GPU) üzerinde paralel olarak çalışabilen bir CNN modeli ile gerçek zamanlı eylem tanıma ve kişi tanımlama için daha sağlam bir şekilde kullanılabilir. Özetle, çalışmamızın katkıları (1) hareketlerin tamamlanma zamanına uyum sağlayarak hareketsiz bitiş kısımlarını bastırmak, (2) harekette değişiklik alanlarını belirleyip vurgulayabilen işlemi sağlamak ve (3) önceki çalışmalardan farklı olarak, Hangi mevcut TT'nin hangi eylem için daha uygun olduğunu belirlemek için ampirik yöntemlerden kaçınarak zaman kaybını önlemek olarak sıralanabilir.

Son iki bölümde ise toplanan davranışsal, termal ve fiziksel verileri ile literatüre yeni bir veri tabanı kazandırılmıştır. Bu veriler ile oluşturulan veri setleri üzerinde önerdiğimiz yöntem hem hareket tanıma hemde kişi kimliklendirme için test edilmiş ve yüksek performans elde edilmiştir. Buna ek olarak tüm veri setleri arasında çoklu-biyometri testleri yapılmış ve başarımları etkisi incelenmiştir. Bu konuda multi-model, multi-instance, multi-feature olarak isimlendirilen çoklu-biyometrik yöntemler yer almaktadır. Bu veri tabanı, elde edilen 5 farklı veri seti ve ham veriler olacak şekilde tüm araştırmacılara ve genel kullanıma açık olarak sunulacaktır.

A

EK TABLOLAR

Renkli kareler sırasıyla **En iyi** **2inci** 3üncü **En kötü** skorları ifade etmektedir.

Tablo A.1 NATOPS *Full – Body* veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları

NATOPS Hareket Tanıma Full Body	DNN			RBF			SVM LIN			POL			3			k-NN 5			7			
	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	
	PCA																					
Features	U2	94.59	92.71	88.02	99.24	98.81	94.42	98.54	97.48	90.7	99.13	98.29	93.04	97.84	96.85	85.02	97.61	95.98	75.63	97.04	94.90	64.6
	U3	94.10	91.31	86.84	99.39	98.98	94.89	98.34	97.36	90.67	99.26	98.52	93.56	97.97	96.82	85.31	97.29	95.92	75.44	96.82	94.80	64.67
	U4	92.36	91.24	86.05	99.00	98.43	93.92	97.92	96.59	89.69	98.79	97.99	92.08	97.33	96.01	84.13	96.68	94.86	75.14	96.31	93.58	65.11
	UR	95.55	93.07	88.59	99.06	98.58	94.51	98.39	97.50	92.26	98.79	98.11	92.85	97.94	97.09	86.32	97.52	96.23	76.27	97.02	95.17	65.30
Features	LBP																					
	U2	94.28	93.36	88.22	95.53	94.58	89.98	93.73	93.03	88.52	93.18	92.50	88.25	96.36	95.52	90.84	96.27	95.28	90.52	96.15	95.19	90.02
	U3	94.35	93.85	88.66	95.59	94.89	90.35	94.43	93.70	89.03	92.78	92.67	88.48	96.50	95.59	91.46	96.25	95.47	91.01	95.94	95.30	90.64
	U4	94.19	92.94	87.34	95.35	94.36	89.24	93.75	92.82	87.81	93.49	92.28	87.34	96.09	94.94	89.92	95.89	94.95	89.49	95.72	94.60	89.17
	UR	94.03	93.20	87.98	95.53	94.41	89.78	94.27	93.25	88.49	92.02	92.55	88.19	96.17	95.35	90.64	96.10	95.14	90.38	95.98	94.95	89.84
Features	HOG																					
	U2	98.72	98.29	95.25	98.68	98.44	96.96	98.28	98.05	96.15	98.34	98.06	95.99	98.81	98.68	97.22	98.79	98.63	97.13	98.73	98.54	97.09
	U3	98.67	98.13	94.74	98.48	98.12	96.49	98.21	97.92	95.82	98.50	97.86	95.51	98.65	98.42	96.86	98.73	98.29	96.67	98.63	98.23	96.52
	U4	98.09	97.84	94.31	98.44	97.85	96.02	98.27	97.49	95.09	98.07	97.45	94.94	98.73	98.23	96.37	98.65	98.10	96.26	98.61	98.05	96.20
	UR	98.13	97.43	94.09	98.14	97.78	96.15	97.75	97.48	95.30	98.07	97.50	95.40	98.23	98.05	96.34	98.17	97.94	96.28	98.09	97.87	96.29

Tablo A.2 NATOPS *Moving – Limbs* veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları

NATOPS Hareket Tanıma. Moving Limbs	DNN			RBF			SVM LIN			POL			3			k-NN 5			7		
	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test
	PCA																				
Features																					
U2	93.95	91.75	83.86	94.90	93.15	86.16	94.60	92.23	83.88	94.64	92.10	83.51	95.27	93.58	86.70	95.21	93.39	86.42	94.98	93.23	86.32
U3	93.73	91.48	83.84	94.83	93.09	86.33	94.60	92.20	84.05	94.51	92.27	83.29	95.10	93.46	86.88	95.18	93.43	86.75	94.96	93.32	86.63
U4	83.52	80.09	71.07	85.32	82.32	75.13	84.11	80.94	72.40	85.20	81.35	72.57	86.31	82.95	75.41	86.29	83.01	75.35	86.09	82.82	75.11
UR	87.54	84.11	75.18	88.95	86.4	78.99	88.28	85.46	76.78	88.78	85.64	76.08	90.77	87.34	79.7	90.33	87.28	79.62	89.94	86.93	79.32
LBP																					
U2	92.31	91.49	86.69	94.14	92.74	88.57	91.64	90.73	86.92	92.31	91.39	86.7	94.81	93.83	89.13	94.70	93.55	88.51	94.44	93.12	87.75
U3	93.00	91.79	86.92	94.47	92.93	88.7	91.84	91.01	86.95	93.23	91.73	86.97	95.22	94.17	89.33	95.07	93.81	88.7	94.89	93.61	88.03
U4	90.50	89.08	83.28	91.69	90.47	85.36	88.92	87.86	83.17	90.26	88.75	82.34	93.21	91.62	85.82	92.96	91.53	85.32	92.49	90.91	84.78
UR	90.54	89.11	83.84	92.21	90.84	86	89.14	88.68	84.1	90.47	89.29	83.89	93.47	91.96	86.71	93.49	91.78	86.08	92.94	91.34	85.51
HOG																					
U2	97.56	96.77	93.05	97.82	97.41	93.86	97.88	97.16	93.09	97.78	96.89	92.72	98.01	97.62	94.22	97.99	97.57	94.07	97.91	97.44	93.97
U3	97.13	96.22	92.07	97.40	96.75	93.26	97.28	96.45	92.27	97.39	96.18	91.97	97.36	97.03	93.59	97.51	96.95	93.45	97.35	96.91	93.31
U4	93.89	92.25	86.65	94.57	93.67	88.85	94.35	93.01	87.34	94.36	92.75	87	95.01	93.57	88.83	94.89	93.86	88.99	94.86	93.73	88.92
UR	94.88	93.82	89.04	95.77	94.96	90.63	95.65	94.59	89.48	95.48	94.43	88.85	95.68	95.1	90.78	95.66	95.11	90.85	95.46	95.06	90.73

Tablo A.3 NATOPS *Full – Body* veri setinde kişi kimliklendirme için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları

NATOPS Kişi Kimlik. Full Body	DNN			RBF			SVM LIN			POL			3			k-NN 5			7		
	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%
	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test
Features	PCA																				
U2	100	100	100	100	100	99.97	100	100	99.59	100	100	99.53	100	100	100	100	100	100	100	100	100
U3	100	100	100	100	100	100	100	100	99.78	100	100	99.66	100	100	100	100	100	100	100	100	100
U4	100	100	100	100	100	100	100	100	99.91	100	100	99.81	100	100	100	100	100	100	100	100	100
UR	100	100	100	100	100	100	100	100	99.69	100	100	99.72	100	100	100	100	100	100	100	100	100
LBP	LBP																				
U2	100	99.90	99.44	100	100	98.31	99.38	99.10	92.94	99.84	99.70	92.31	100	100	98.84	100	100	98.84	100	100	98.97
U3	99.84	99.80	98.81	99.69	99.65	97.19	99.69	99.25	88.84	99.69	99.30	87.13	99.69	99.60	98.22	99.69	99.60	98.28	99.69	99.60	98.34
U4	99.84	99.60	99.34	99.84	99.70	97.38	99.38	98.60	89.28	99.53	98.50	86.88	99.84	99.75	98.13	99.84	99.70	98.03	99.84	99.65	98.03
UR	99.53	99.70	97.22	99.69	99.55	91.00	98.59	96.75	78.34	98.75	97.50	75.84	99.69	99.75	92.75	99.69	99.69	98.80	92.72	99.69	99.75
HOG	HOG																				
U2	100	100	99.72	100	100	91.88	100	99.80	86.37	100	99.85	82.84	100	100	94.5	100	100	94.59	100	100	94.59
U3	100	100	99.47	100	99.90	89	99.69	99.60	86.88	100	99.55	85.81	100	99.95	90.5	100	99.95	90.5	100	99.95	90.50
U4	99.84	99.85	97.91	99.84	99.80	78.88	99.84	99.65	76.25	99.84	99.75	75.84	100	99.85	80.38	100	99.85	80.38	100	99.85	80.38
UR	100	100	99.91	100	100	99.84	100	99.60	99.75	100	99.50	99.69	100	100	99.84	100	100	99.84	100	100	99.84

Tablo A.4 NATOPSMoving – *Limbs* veri setinde kişi kimliklendirme için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları

NATOPS Kişi Kimlik. Moving Limbs	DNN			RBF			SVM LIN			POL			3			k-NN 5			7		
	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%
	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test
Features	PCA																				
U2	99.69	99.70	97.09	99.53	99.50	90.72	97.34	97.95	83.47	96.72	96.90	79.91	99.53	99.60	93.13	99.53	99.60	93.31	99.53	99.60	93.28
U3	99.53	99.25	97.84	99.69	99.20	86.84	98.91	98.00	78.00	98.75	97.20	73.97	99.84	99.25	89.13	99.84	99.25	89.22	99.84	99.25	89.28
U4	99.69	99.25	95.44	98.44	98.30	64.91	97.97	95.70	54.75	96.72	94.00	50.84	98.59	98.70	69.31	98.54	98.75	69.28	98.59	98.70	69.28
UR	100	99.80	98.87	99.38	99.35	80.81	99.38	97.40	73.13	98.91	95.85	68.56	99.69	99.50	83.44	99.53	99.50	83.41	99.69	99.50	83.59
LBP	LBP																				
U2	99.22	98.95	95.69	99.69	99.20	92.41	98.75	97.00	85.00	98.91	97.75	83.94	100	99.40	94.34	99.84	99.35	94.63	99.69	99.30	94.38
U3	99.06	99.00	96.94	99.22	99.30	92.66	98.28	97.10	83.69	98.59	97.05	81.50	99.22	99.25	94.31	99.22	99.25	94.28	99.22	99.15	94.28
U4	99.69	99.40	96.81	99.53	99.25	94.28	98.28	97.45	83.91	98.28	97.35	83.75	99.69	99.55	95.75	99.53	99.60	95.78	99.69	99.55	95.91
UR	99.84	99.25	96.81	99.69	99.65	89.69	97.34	97.10	77.75	98.13	96.90	74.16	100	99.65	91.81	99.69	99.65	91.78	99.84	99.70	91.88
HOG	HOG																				
U2	99.69	99.30	94.09	99.69	99.00	85.50	99.53	98.45	85.00	99.53	98.25	83.84	99.69	99.25	86.06	99.69	99.30	86.03	99.69	99.30	86.13
U3	99.22	98.95	97.81	99.06	98.85	97.88	98.13	96.85	97.09	98.28	97.10	96.91	99.22	99.05	98.03	99.22	99.00	98.00	99.22	99.05	98.09
U4	99.22	98.95	95.03	99.06	98.75	99.05	95.66	98.13	94.44	97.50	96.30	93.66	99.22	99.20	95.87	99.22	99.20	95.91	99.22	99.20	95.97
UR	99.84	99.75	98.06	99.84	99.15	98.47	99.38	97.60	98.06	99.53	97.30	97.47	99.84	99.35	98.75	99.84	99.40	98.78	99.84	99.35	98.81

Tablo A.5 SBU KINECT *Full – Body* veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları

SBU Hareket Tanıma Full Body	DNN			RBF			SVM LIN			POL			3			k-NN 5			7		
	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%
	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test
Features	PCA																				
U2	73.75	65.25	N/A	72.50	61.00	N/A	72.50	64.00	N/A	66.88	55.75	N/A	73.75	65.25	N/A	75.00	64.75	N/A	74.38	64.75	N/A
U3	65.63	59.25	N/A	66.88	58.50	N/A	63.13	55.88	N/A	65.00	52.50	N/A	66.88	60.25	N/A	68.13	59.50	N/A	67.50	59.75	N/A
U4	71.25	64.25	N/A	73.13	57.50	N/A	71.88	61.00	N/A	67.50	52.50	N/A	71.88	59.25	N/A	76.25	62.25	N/A	75.63	62.75	N/A
UR	77.50	66.75	N/A	70.63	63.00	N/A	72.50	64.38	N/A	67.50	55.50	N/A	78.13	64.75	N/A	78.13	65.50	N/A	77.50	65.00	N/A
LBP	LBP																				
U2	88.13	77.25	61.88	88.75	79.50	70.78	85.63	75.00	61.56	84.38	75.75	70.00	88.75	81.25	73.75	88.75	81.00	74.06	89.38	81.75	74.84
U3	86.25	76.25	61.25	84.38	71.00	61.56	78.13	67.25	59.53	77.50	69.00	56.09	87.50	72.50	61.88	87.50	81.00	62.66	86.25	71.75	61.72
U4	86.88	76.75	62.81	85.63	78.00	71.56	82.50	73.50	62.66	83.13	74.00	70.00	87.50	81.25	74.06	87.50	80.75	76.25	88.75	75.94	75.94
UR	85.63	72.25	61.25	88.75	81.00	72.66	85.00	76.25	63.28	85.63	78.50	71.25	90.00	83.75	75.16	90.63	85.00	75.94	90.00	85.00	77.03
HOG	HOG																				
U2	87.50	81.00	72.50	78.75	62.00	72.97	71.88	61.75	70.94	71.88	61.75	67.34	79.38	62.50	73.59	79.38	63.75	73.44	80.00	63.75	72.97
U3	80.63	73.75	68.44	81.88	48.25	66.72	78.13	47.25	64.69	78.13	45.75	63.28	84.38	48.50	68.44	83.75	48.50	69.53	84.38	48.25	69.53
U4	86.25	80.75	71.88	80.00	70.00	71.56	77.50	64.50	69.53	74.38	65.50	66.09	81.88	70.50	72.34	81.88	70.50	72.50	81.88	70.50	72.19
UR	87.50	77.25	72.34	81.88	63.00	73.59	74.38	61.75	73.59	76.25	60.50	72.66	84.38	64.50	74.69	84.38	63.50	74.53	84.38	63.25	74.38

Tablo A.6 SBU KINECT *Moving – Limbs* veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları

SBU Hareket Tanıma Moving Limbs	DNN			RBF			SVM LIN			POL			3			k-NN 5			7		
	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test
	PCA																				
Features	PCA																				
U2	73.75	N/A	N/A	69.38	N/A	N/A	68.13	N/A	N/A	64.38	N/A	N/A	68.75	N/A	N/A	69.38	N/A	N/A	68.75	N/A	N/A
U3	70.00	N/A	N/A	71.25	N/A	N/A	69.38	N/A	N/A	66.25	N/A	N/A	68.75	N/A	N/A	69.38	N/A	N/A	70.63	N/A	N/A
U4	69.38	N/A	N/A	68.13	N/A	N/A	66.88	N/A	N/A	71.88	N/A	N/A	70.63	N/A	N/A	71.25	N/A	N/A	72.50	N/A	N/A
UR	74.38	N/A	N/A	70.63	N/A	N/A	71.88	N/A	N/A	68.75	N/A	N/A	73.13	N/A	N/A	74.38	N/A	N/A	76.25	N/A	N/A
LBP	LBP																				
U2	89.38	72.00	N/A	86.88	81.00	69.38	85.63	73.25	N/A	81.88	77.50	67.34	89.38	81.75	70.63	90.00	72.50	N/A	90.63	82.50	70.47
U3	85.00	75.25	N/A	85.00	82.00	71.09	81.25	74.50	N/A	80.00	75.75	68.59	86.88	84.25	71.72	86.25	72.00	N/A	86.88	84.25	72.34
U4	88.13	73.50	N/A	87.50	80.00	70.47	83.13	73.75	N/A	84.38	73.75	67.5	89.38	83.75	76.75	73.50	N/A	N/A	89.38	84.50	70.78
UR	82.50	73.25	N/A	88.13	82.50	69.53	83.75	74.25	N/A	84.38	76.75	67.66	88.75	83.25	70.63	89.38	72.25	N/A	90.00	84.00	70.47
HOG	HOG																				
U2	85.00	77.00	70.31	79.38	N/A	71.41	73.13	N/A	70.47	76.25	N/A	68.75	83.13	N/A	72.5	82.50	N/A	72.81	82.50	N/A	72.5
U3	85.00	73.75	69.53	80.00	67.75	67.81	73.75	64.25	63.91	71.25	63.50	62.81	82.50	69.00	69.06	82.50	69.00	68.59	83.13	69.00	68.13
U4	85.63	75.75	70.78	78.13	70.00	70.00	70.63	64.50	69.22	73.75	66.00	66.09	80.00	71.75	72.03	80.63	71.00	71.88	80.63	71.50	71.88
UR	88.13	79.50	72.97	84.38	N/A	75.31	78.13	N/A	73.91	80.00	N/A	80.00	86.25	N/A	75.47	86.25	N/A	75.78	86.25	N/A	75.47

Tablo A.7 BodyLogin BLD-SS *User – Defined* veri setinde hareket tanıma için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları

BodyLogin Hareket Tanıma Full Body	DNN			RBF			SVM LIN			POL			3			KNN 5			7		
	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test
	PCA																				
Features																					
U2	99.63	99.30	94.75	99.75	98.75	84.69	99.00	98.10	89.47	96.38	91.85	61.53	99.88	99.15	87.09	99.88	99.25	87.19	99.88	99.10	87.44
U3	99.13	97.45	91.00	98.75	97.20	71.03	98.50	96.93	82.06	98.25	93.35	49.00	98.87	97.80	75.63	99.00	97.70	75.84	98.87	97.60	75.72
U4	99.62	98.75	94.69	99.50	98.60	88.59	98.75	98.10	87.47	98.25	95.45	65.72	99.63	98.95	90.94	99.75	98.95	90.97	99.75	98.90	91.12
UR	99.63	99.40	95.59	99.63	99.25	85.72	99.13	98.18	89.72	97.50	93.70	61.69	100	99.20	88.38	99.88	99.25	88.5	99.75	99.30	88.28
LBP	LBP																				
U2	96.25	94.05	81.75	100	99.75	97.94	100	99.20	97.72	100	99.40	97.81	100	99.85	98.06	100	99.85	98.06	100	99.85	98.09
U3	97.38	93.20	81.13	100	99.75	97.44	99.88	99.15	96.47	99.88	99.40	97.31	100	99.75	97.63	100	99.75	97.59	100	99.75	97.66
U4	97.38	93.35	80.91	100	99.85	97.5	100	99.15	96.97	100	99.40	97.34	100	99.90	97.56	100	99.90	97.66	100	99.90	97.44
UR	96.88	92.95	81.75	100	99.70	97.81	100	99.25	97.44	100	99.50	97.78	100	99.75	97.88	100	99.75	97.91	100	99.75	97.84
HOG	HOG																				
U2	100	99.70	98.31	99.88	99.40	98.66	99.75	97.90	98.47	99.63	97.85	98.25	99.88	99.60	98.75	99.88	99.60	98.84	99.88	99.60	98.94
U3	99.62	99.20	96.72	99.88	98.85	98.06	99.63	97.95	97.63	99.38	97.35	96.78	100	99.10	98.06	99.88	99.05	98.13	99.88	99.05	98.09
U4	99.75	99.85	97.94	99.88	99.40	98.84	99.63	98.40	98.75	99.63	98.10	98.66	99.88	99.50	98.94	99.88	99.50	98.97	99.88	99.50	99.06
UR	100	99.65	98.72	99.88	99.50	98.75	99.63	98.50	98.59	99.63	98.25	98.5	99.88	99.70	98.78	99.88	99.60	98.78	99.88	99.60	98.75



Tablo A.8 BodyLogin *Full – Body* veri setinde kişi kimliklendirme için sınıflandırıcı ve özellik çıkartma varyasyon sonuçları

BodyLogin Kişi Kimlik. Full Body	DNN			RBF			SVM LIN			POL			3			KNN 5			7		
	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test	20% Test	50% Test	80% Test
	PCA																				
Features																					
U2	98.60	97.36	83.09	98.50	97.35	67.13	97.03	93.54	58.6	96.22	92.13	56.26	98.66	97.84	69.62	98.78	97.85	69.71	98.75	97.93	69.83
U3	98.22	96.75	82.56	98.10	96.29	68.38	96.47	92.76	63.37	95.63	91.38	61.14	98.38	96.90	70.11	98.41	96.96	70.30	98.41	97.00	70.24
U4	99.01	97.53	87.18	99.16	97.69	70.83	97.88	94.95	60.90	97.19	93.75	58.64	99.19	98.16	73.48	99.25	98.23	73.48	99.25	98.25	73.67
UR	98.63	97.70	87.22	98.82	97.81	73.16	97.66	94.65	64.72	96.88	93.59	62.11	98.97	98.24	75.58	98.97	98.40	75.54	99.10	98.38	75.63
LBP	LBP																				
U2	86.04	77.26	N\A	99.22	97.89	88.27	98.63	94.46	86.02	98.56	94.91	87.25	99.31	98.18	89.19	99.41	98.23	89.25	99.31	98.23	89.24
U3	85.63	78.40	N\A	99.13	97.43	87.94	98.78	94.40	85.46	98.75	95.19	87.08	99.16	97.79	88.31	99.19	97.81	88.47	99.22	97.83	88.37
U4	84.47	77.46	N\A	99.28	98.01	88.35	98.75	95.20	85.84	98.88	95.84	87.53	99.31	98.31	88.82	99.34	98.35	88.91	99.38	98.33	88.87
UR	84.82	76.56	N\A	99.28	97.84	88.57	98.53	95.10	86.76	98.88	95.39	87.85	99.41	98.26	89.09	99.38	98.25	89.07	99.38	98.24	89.07
HOG	HOG																				
U2	98.88	98.00	87.83	99.07	97.56	89.25	98.16	94.58	88.51	97.91	93.93	87.71	99.66	99.08	90.52	99.66	99.05	90.63	99.69	99.11	90.75
U3	98.75	97.69	87.47	98.88	97.60	88.4	98.13	95.65	87.6	98.28	94.96	86.37	99.56	98.53	90.26	99.59	98.53	90.36	99.59	98.54	90.38
U4	99.16	98.19	89.9	99.22	97.89	90.94	98.63	95.48	89.76	98.34	94.96	88.94	99.59	98.95	90.7	99.56	98.96	90.69	99.56	98.95	90.73
UR	98.94	97.96	90.21	99.32	97.83	90.68	98.44	95.65	89.57	98.35	94.99	88.82	99.72	98.99	90.75	99.75	98.96	90.79	99.75	98.94	90.89

B

EK FORMLAR



Ek-2. Deneklerin Gönüllülüğü ve Aydınlatılmış Onam Formu

.././20..

DENEKLERİN GÖNÜLLÜLÜĞÜ VE AYDINLATILMIŞ ONAM FORMU

Prof. Dr. Tülay Yıldırım tarafından yürütülen Onur Can Kurban'ın araştırmacılığını yaptığı yeni bir biyometrik sistem tasarımı amaçlayan, 'Davranışsal ve Fiziksel Multibiyometri ile Kişi Kimliklendirme/Tanıma Üzerine Yaklaşımlar' konulu, doktora tezinde, deneklerden alınacak biyometrik veriler isimsiz olarak kaydedilecek ve deneklerin özel hayatını, kişilik haklarını etkilemeyecektir. Araştırma süresince deneklerden gönüllü olarak yürüyüş görüntüsü, farklı açılardan termal yüz görüntüsü ve eklemler arası uzaklık gibi biyometrik veriler istenecektir. Gönüllü olmayan, sözlü rızası olmayan hiçbir denekten çalışmaya katılması istenmeyecektir. Deneklerden sözlü olarak çalışmada biyometrik verilerinin verilmesinin rızası, yazılı olarak Bilgilendirilmiş Olur Formu, Deneklerin Gönüllülüğü ve Aydınlatılmış Onam Formu ve Gönüllü Katılım Formu talep edilecek ve deneklerin imzalı onayları istenecektir.

Yukarıda bulunan bilgileri okudum ve onaylıyorum.

Katılımcının Adı ve Soyadı, Adresi: Tarih, İmza:

Araştırmacının Adı ve Soyadı, Adresi: Tarih, İmza:

Arş.Gör. Onur Can Kurban, YTÜ Davutpaşa Kampüsü Elektrik-Elektronik Fakültesi Elektronik ve Haberleşme Mühendisliği Bölümü B014 Esenler/İstanbul

Ek-3. Bilgilendirilmiş Olur Formu

.././20..

BİLGİLENDİRİLMİŞ OLUR FORMU

1. Aşağıda imzası olan ben " Davranışsal ve Fiziksel Multibiyometri ile Kişi Kimliklendirme/Tanıma Üzerine Yaklaşımlar" başlıklı çalışmaya katılmayı kabul ediyorum.
2. Bu çalışmayı yürüten Prof. Dr. Tülay Yıldırım çalışmanın yapısı, amacı ve muhtemel süresi, hakkında ayrıntılı sözlü ve yazılı bilgi verdi.
3. Araştırmacı Onur Can Kurban'a çalışmasıyla ilgili her soruyu sorma fırsatını buldum. Cevapları ve bana verilen bilgiyi anladım.
4. Araştırmacı Onur Can Kurban'a bilgilerin ayrıntılarını açıklamama ve benimle ilgili sırları koruması şartıyla benimle bu çalışmayı yapmasına izin veriyorum.
5. Çalışma boyunca tüm kurallara uyacağıma, araştırmacı Onur Can Kurban ile tam bir uyum içinde çalışacağıma ve konuyla ilgili herhangi bir sorun çıktığında hemen onu arayacağımı kabul ediyorum.
6. Bu çalışma sonuçlarının kullanılmasını kısıtlamayacağımı ve yayın, rapor ve benzeri bilimsel dokümanlarda kullanabileceğimi kabul ediyorum.
7. Bu çalışmadan istediğim zaman çıkabileceğimi anladım.

OKUDUM VE ONAYLADIM.

Katılımcının Adı ve Soyadı, Adresi: Tarih, İmza:

Araştırmacının Adı ve Soyadı, Adresi: Tarih, İmza:

Arş. Gör. Onur Can Kurban, YTÜ Davutpaşa Kampüsü Elektrik Elektronik Fakültesi
Elektronik ve Haberleşme Mühendisliği Bölümü B014 Esenler/İstanbul

GÖNÜLLÜ KATILIM FORMU

Davranışsal ve Fiziksel Multibiyometri ile Kişi Kimliklendirme/Tanıma Üzerine Yaklaşımlar çalışmasına kullanılmak üzere;

Farklı açılarda yürüyüş görüntülerinizi gönüllü olarak vermeyi kabul ediyor musunuz?

Evet ☐ Hayır ☐

Farklı açılarda Termal Yüz görüntülerinizi gönüllü olarak vermeyi kabul ediyor musunuz?

Evet ☐ Hayır ☐

Eklemler arası uzaklık görüntülerinizi gönüllü olarak vermeyi kabul ediyor musunuz?

Evet ☐ Hayır ☐

Yukarıda verdiğiniz biyometrik verilerin akademik yayınlarda basılı olarak kullanılmasına ve isimsiz olarak bir veri tabanında yer almasına gönüllü olarak onay veriyor musunuz?

Evet ☐ Hayır ☐

Gönüllünün Adı ve Soyadı, Adresi: Tarih, İmza:

Araştırmacının Adı ve Soyadı, Adresi: Tarih, İmza:

Arş. Gör. Onur Can Kurban, YTÜ Davutpaşa Kampüsü Elektrik Elektronik Fakültesi
Elektronik ve Haberleşme Mühendisliği Bölümü B014 Esenler/İstanbul

- [1] S. Dargan, M. Kumar, "A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities," *Expert Systems with Applications*, vol. 143, p. 113 114, 2020.
- [2] Y. Zhang, Y. Huang, L. Wang, S. Yu, "A comprehensive study on gait biometrics using a joint cnn-based method," *Pattern Recognition*, vol. 93, pp. 228–236, 2019.
- [3] D. A. Reid, M. S. Nixon, S. V. Stevenage, "Identifying humans using comparative descriptions," in *4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011)*, IET, 2011, pp. 1–6.
- [4] K. Moustakas, D. Tzovaras, G. Stavropoulos, "Gait recognition using geometric features and soft biometrics," *IEEE Signal processing letters*, vol. 17, no. 4, pp. 367–370, 2010.
- [5] D. A. Reid, M. S. Nixon, S. V. Stevenage, "Soft biometrics; human identification using comparative descriptions," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 36, no. 6, pp. 1216–1228, 2013.
- [6] K. Niinuma, U. Park, A. K. Jain, "Soft biometric traits for continuous user authentication," *IEEE Transactions on information forensics and security*, vol. 5, no. 4, pp. 771–780, 2010.
- [7] A. Iosifidis, A. Tefas, I. Pitas, "Activity-based person identification using fuzzy representation and discriminant learning," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 530–542, 2011.
- [8] A. M. S. Rahma, R. F. Hassan, S. Hamandi, "Design a multi biometric system for safe access to buildings," *Al-Qadisiyah Journal Of Pure Science*, vol. 26, no. 4, pp. 275–287, 2021.
- [9] Y. Zhou, "Evaluation of biometric recognition in the covid-19 period," in *2021 2nd International Conference on Computing and Data Science (CDS)*, IEEE, 2021, pp. 243–248.
- [10] S. Carlaw, "Impact on biometrics of covid-19," *Biometric Technology Today*, vol. 2020, no. 4, pp. 8–9, 2020.
- [11] A. Goode, "Biometrics in payments: The challenge of covid and customer choice," *Biometric Technology Today*, vol. 2021, no. 7-8, pp. 7–10, 2021.
- [12] M. Y. Sikkandar, "Design a contactless authentication system using hand gestures technique in covid-19 panic situation," *Annals of the Romanian Society for Cell Biology*, pp. 2149–2159, 2021.
- [13] J. K. Aggarwal, M. S. Ryoo, "Human activity analysis: A review," *ACM Computing Surveys (CSUR)*, vol. 43, no. 3, p. 16, 2011.

- [14] A. F. Bobick, J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
- [15] J. Han, B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 28, no. 2, pp. 316–322, 2006.
- [16] J. W. Davis, "Hierarchical motion history images for recognizing human motion," in *Detection and Recognition of Events in Video, 2001. Proceedings. IEEE Workshop on*, IEEE, 2001, pp. 39–46.
- [17] J. Liu, N. Zheng, "Gait history image: A novel temporal template for gait recognition," in *2007 IEEE International Conference on Multimedia and Expo*, Jul. 2007, pp. 663–666.
- [18] Q. Ma, S. Wang, D. Nie, J. Qiu, "Recognizing humans based on gait moment image," in *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007.*, IEEE, vol. 2, 2007, pp. 606–610.
- [19] C. Chen, J. Liang, H. Zhao, H. Hu, J. Tian, "Frame difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognition Letters*, vol. 30, no. 11, pp. 977–984, 2009.
- [20] M. Poonkodi, G. Vadivu, "Action recognition using correlation of temporal difference frame (ctdf)—an algorithmic approach," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 7, pp. 7107–7120, 2021.
- [21] P. Ramya, R. Rajeswari, "Human action recognition using distance transform and entropy based features," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 8147–8173, 2021.
- [22] Y. A. Andrade-Ambriz, S. Ledesma, M.-A. Ibarra-Manzano, M. I. Oros-Flores, D.-L. Almanza-Ojeda, "Human activity recognition using temporal convolutional neural network architecture," *Expert Systems with Applications*, vol. 191, p. 116 287, 2022.
- [23] J. Chen, W. Yang, C. Liu, L. Yao, "A data augmentation method for skeleton-based action recognition with relative features," *Applied Sciences*, vol. 11, no. 23, p. 11 481, 2021.
- [24] E. P. Ijjina, K. M. Chalavadi, "Human action recognition in rgb-d videos using motion sequence information and deep learning," *Pattern Recognition*, vol. 72, pp. 504–516, 2017.
- [25] A. Abdelbaky, S. Aly, "Two-stream spatiotemporal feature fusion for human action recognition," *The Visual Computer*, vol. 37, no. 7, pp. 1821–1835, 2021.
- [26] S. R. Mishra, T. K. Mishra, G. Sanyal, A. Sarkar, S. C. Satapathy, "Real time human action recognition using triggered frame extraction and a typical cnn heuristic," *Pattern Recognition Letters*, vol. 135, pp. 329–336, 2020.
- [27] B. Liu, H. Cai, Z. Ju, H. Liu, "Rgb-d sensing based human action and interaction analysis: A survey," *Pattern Recognition*, vol. 94, pp. 1–12, 2019.

- [28] J. Han, B. Bhanu, "Individual recognition using gait energy image," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 2, pp. 316–322, 2005.
- [29] R. S. Ghiass, O. Arandjelović, A. Bendada, X. Maldague, "Infrared face recognition: A comprehensive review of methodologies and databases," *Pattern Recognition*, vol. 47, no. 9, pp. 2807–2824, 2014.
- [30] J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.
- [31] M. O. Şen, "Çoklu kinect kullanımıyla elde edilen iskelet hareket verilerinin birleştirilmesi ve karşılaştırılması," M.S. thesis, İnönü Üniversitesi Fen Bilimleri Enstitüsü, 2016.
- [32] R. Min, N. Kose, J.-L. Dugelay, "Kinectfacedb: A kinect database for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 11, pp. 1534–1548, 2014.
- [33] O. M. Parkhi, A. Vedaldi, A. Zisserman, "Deep face recognition," 2015.
- [34] V. Kazemi, J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1867–1874.
- [35] J. Wu, P. Ishwar, J. Konrad, "Silhouettes versus skeletons in gesture-based authentication with kinect," in *2014 International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, 2014, pp. 99–106.
- [36] O. C. Kurban, T. Yildirim, A. Bilgiç, "A multi-biometric recognition system based on deep features of face and gesture energy image," in *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, IEEE, 2017, pp. 361–364.
- [37] J. Wu, P. Ishwar, J. Konrad, "The value of posture, build and dynamics in gesture-based user authentication," in *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, IEEE, 2014, pp. 1–8.
- [38] Y. Song, D. Demirdjian, R. Davis, "Tracking body and hands for gesture recognition: Natops aircraft handling signals database," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, IEEE, 2011, pp. 500–506.
- [39] K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, D. Samaras, "Two-person interaction detection using body-pose features and multiple instance learning," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, IEEE, 2012, pp. 28–35.
- [40] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *British Machine Vision Conference*, 2014.
- [41] S. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, P. Frossard, "Universal adversarial perturbations," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 86–94.

- [42] M. Xie, N. Jean, M. Burke, D. Lobell, S. Ermon, "Transfer learning from deep features for remote sensing and poverty mapping," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, 2016, pp. 3929–3935.
- [43] X. Glorot, Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. Society for Artificial Intelligence and Statistics, 2010, pp. 249–256.
- [44] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [45] Y. Song, D. Demirdjian, R. Davis, "Continuous body and hand gesture recognition for natural human-computer interaction," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 1, p. 5, 2012.
- [46] Y. Song, L.-P. Morency, R. Davis, "Multi-view latent variable discriminative models for action recognition," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, IEEE, 2012, pp. 2120–2127.
- [47] A. Elboushaki, R. Hannane, K. Afdel, L. Koutti, "Multid-cnn: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in rgb-d image sequences," *Expert Systems with Applications*, vol. 139, p. 112 829, 2020.
- [48] P. Khaire, P. Kumar, J. Imran, "Combining cnn streams of rgb-d and skeletal data for human activity recognition," *Pattern Recognition Letters*, vol. 115, pp. 107–116, 2018.
- [49] Y. Ji, G. Ye, H. Cheng, "Interactive body part contrast mining for human interaction recognition," in *Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on*, IEEE, 2014, pp. 1–6.
- [50] Y. Du, W. Wang, L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1110–1118.
- [51] M. Edwards, X. Xie, "Generating local temporal poses from gestures with aligned cluster analysis for human action recognition," in *UK Computer Vision Student Workshop (BMVW)*, BMVA Press, 2015, pp. 1–1.
- [52] W. Zhu *et al.*, "Co-occurrence feature learning for skeleton based action recognition using regularized deep lstm networks," in *AAAI*, vol. 2, 2016, p. 6.
- [53] T. Huynh-The *et al.*, "Pam-based flexible generative topic model for 3d interactive activity recognition," in *Advanced Technologies for Communications (ATC), 2015 International Conference on*, IEEE, 2015, pp. 117–122.
- [54] S. Song, C. Lan, J. Xing, W. Zeng, J. Liu, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," in *AAAI*, vol. 1, 2017, pp. 4263–4270.
- [55] J. Liu, A. Shahroudy, D. Xu, G. Wang, "Spatio-temporal lstm with trust gates for 3d human action recognition," in *European Conference on Computer Vision*, Springer, 2016, pp. 816–833.

- [56] L. Lin, K. Wang, W. Zuo, M. Wang, J. Luo, L. Zhang, "A deep structured model with radius-margin bound for 3d human activity recognition," *International Journal of Computer Vision*, vol. 118, no. 2, pp. 256–273, 2016.
- [57] Q. Ke, M. Bennamoun, S. An, F. Sohel, F. Boussaid, "A new representation of skeleton sequences for 3d action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3288–3297.
- [58] A. Gharahdaghi, F. Razzazi, A. Amini, "A non-linear mapping representing human action recognition under missing modality problem in video data," *Measurement*, vol. 186, p. 110 123, 2021.
- [59] T. Singh, D. K. Vishwakarma, "A deeply coupled convnet for human activity recognition using dynamic and rgb images," *Neural Computing and Applications*, vol. 33, no. 1, pp. 469–485, 2021.
- [60] T. Ojala, M. Pietikäinen, T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [61] R. K. McConnell, *Method of and apparatus for pattern recognition*, US Patent 4,567,610, Jan. 1986.
- [62] N. Calik, M. A. Belen, P. Mahouti, "Deep learning base modified mlp model for precise scattering parameter prediction of capacitive feed antenna," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, e2682, 2019.
- [63] S. Ioffe, C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [64] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [65] Z. Guo, L. Zhang, D. Zhang, "Rotation invariant texture classification using lbp variance (lbpv) with global matching," *Pattern recognition*, vol. 43, no. 3, pp. 706–719, 2010.
- [66] L. Nanni, A. Lumini, S. Brahnam, "Survey on lbp based texture descriptors for image classification," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3634–3641, 2012.
- [67] O. C. Kurban, N. Calik, T. Yildirim, "Human and action recognition using adaptive energy images," *Pattern Recognition*, vol. 127, p. 108 621, 2022.

TEZDEN ÜRETİLMİŞ YAYINLAR

Makale

1. Kurban, O.C., Calik, N., Yildirim, T. (2022). Human and action recognition using adaptive energy images. Pattern Recognition, 127, 108621.

Konferans Bildirisi

1. Kurban, O.C., Yildirim, T., Bilgic, A. (2017, July). A multi-biometric recognition system based on deep features of face and gesture energy image. In 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA) (pp. 361-364). IEEE.