



**GebzeYüksek Teknoloji Enstitüsü
Bilgisayar Mühendisliği Bölümü**

Teknik Rapor
Seminer Raporu

TM # :	
Başlık :	Videolarda İnsan Eylemlerinin Tanınması
Anahtar Kelimeler :	Bilgisayarla Görme, İnsan Eylemlerinin Tanınması, İ Dönüşümü, Hareket Örüntüleri.
Yazarlar :	İlktan Ar
Tarih :	16.01.2009
Yayın Dili :	Türkçe

Özet:

Bu çalışmada Bilgisayarla Görme alanında günümüzdeki en dikkat çekici konulardan birisi olan Videolarda İnsan Eylemlerinin Tanınması konusu araştırılmıştır. Konu üzerinde yapılan çalışmalara değinildikten sonra güncel yaklaşımlardan öne çıkan hareket örüntüleri temelli yaklaşım ve \mathfrak{R} dönüşüm temelli yaklaşım hakkında detaylı bilgi verilmiştir. Genel kullanıma açık bir eylem videosu veritabanından seçilen videolar üzerinde bu iki yaklaşım uygulanarak eylemleri sınıflandırmak için çeşitli öznitelikler elde edilmiştir. Sistemimiz, bu öznitelikleri ile Adaboost ve türevi algoritmalar aracılığıyla eğitilmiş ve farklı test videolar üzerinde denenerek sistemin sınıflandırma başarısı elde edilmiştir.

Anahtar Kelimeler : Bilgisayarla Görme, İnsan Eylemlerinin Tanınması, \mathfrak{R} Dönüşümü, Hareket Örüntüleri.

Abstract:

In this work, we researched “Human Activity Recognition in Videos” which is one of the most popular Computer Vision’s topic in nowadays. After review of literature, two actual approach which are motion pattern based approach and \mathfrak{R} transform based approach were examined in details. Videos taken from a public database were used to implement this approaches so attributes to classify these videos were captured. Our system, was trained with these attributes by Adaboost and similar algorithms. At the end we obtained classification performance of our system with test-videos.

Keywords : Computer Vision, Human Activity Recognition, \mathfrak{R} Transform, Motion Patterns.

Videolarda İnsan Eylemlerinin Tanınması

İlktan Ar

GYTE Bilgisayar Mühendisliği Bölümü,
ilktana@khas.edu.tr

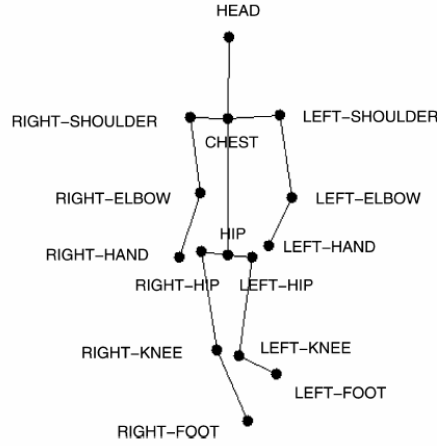
1. Sunuş

Videolardaki insan eylemlerinin tanınması günümüzdeki en dikkat çeken bilgisayarla görme araştırmalarından biridir. Çok çeşitli kullanım alanları olan bu konuda çok değişik çalışmalar yapılmaktadır. Örneğin spor müsabakalarında yapılan insan eylemlerinin neler olduğunun tanımlanması (teniste vuruş stili v.s.), dans ve bale kareograflerinin tanımlanması, ortopedik hastaların klinik incelemeleri, insanın fiziksel durumunun incelenmesi, yaşlı insanların bakımı gibi konular bu tip çalışmaların uygulama alanını oluşturmaktadır. Ayrıca konu, yayılma alanları olarak ise dudak okuma, el ifadeleri anlama, yüz hareketleri anlama v.s. gibi alt dallara da bölünebilmektedir.

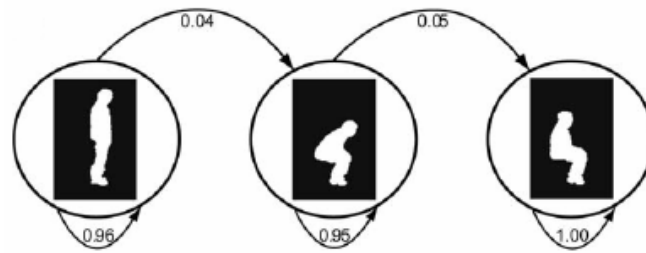
Yaptığımız çalışmada ana problem bir insan tarafından yapılan eylemin bir sabit kamera aracılığıyla çekilmiş videosundan bu eylemin ne olduğunun tespit edilmesidir. Verilen ana problem tanımımızla bazı kısıtlamalar ile sistemimizi sınırlandırmaya çalıştık. Çok sayıda hareketli veya sabit kamera kullanan yapılarla farklılığımız bu tanımda bellidir. Konunun karmaşıklığı ise videolardaki farklı özelliklerde insanlar, tek bir kamera açısının kullanımı, eylemleri tanımlayan özniteliklerin neler olduğu, bu özniteliklerin seçici olup olmadığı, seçilen özniteliklerin nasıl kullanılacağı gibi bir çok alt problemden kaynaklanmaktadır. Problemimiz için varolan çözümler incelendiği çok farklı yaklaşımların olduğu ve bu yaklaşımlardan hiçbirinin her zaman iyi bir şekilde çalışmadığı ortaya çıkmaktadır. Ayrıca varolan çözümlerin çok farklı kısıtlamalar ile çalıştığından dolayı birbirleriyle tam olarak karşılaştırılamaması da ayrı bir problemdir. İlerleyen paragraflarda aktarılan literatür araştırmasından da anlaşılacağı üzere gerçek zamanda çalışan başarılı bir şekilde videolardaki insan eylemlerini tanımlayacak bir sistem ihtiyacı mevcuttur.

Konumuzdaki önceki çalışmalar çeşitli “survey” tipi araştırmalardan derlenilmiştir. Bu yaklaşımın benimsenmesinin esas amacı çok çeşitli ve çok fazla çalışma olmasıdır. Bu şekilde önemli olan yaklaşımları daha rahat inceleyebileceğimiz düşünülmüştür ayrıca her çalışma hakkında detaylı bilgi verilmesi yerine öne çıkan çalışmaların ana yaklaşımları hakkında bilgi verilmiştir. D.M.Gavrila (1998) yaptığı çalışmasında insan eylemlerinin tanınmasının gelecekte Bilgisayarla Görme alanında ön plana geleceğini belirtmiş ve bu ana kadar yapılan çalışmaları büyük ölçüde özetlemiştir [1]. Bu çalışmadan incelendiği üzere bu yıla kadar yapılan çalışmalar yoğun olarak insanın uzuvların tespit edilmesi konusunda yoğunlaşmıştır. Uzuvlar bulunduktan sonra bunların konumlarına göre çeşitli metrikler tanımlanmakta ve yapılan hareket tahmin edilmeye çalışılmaktadır. Özetle vücut parçalarının 2D veya 3D ortamda bulunup vücudun modelinin oluşturulması ve eylemin bu noktalara göre modellenmesi ilk zamanlarda en

öne çıkan yaklaşımdır. Şekil 1’de bu çalışmalardan alınan örnek bir şekil gösterilmiştir. Liang Wang ve diğerleri (2002), insan hareket analizi bazındaki 1997-2001 yılları arasındaki çalışmaları özetlemiştir [2]. Bu özetleme ile insan hareket analizinde 3 ana alt başlık tanımlanmıştır. Bunlar yapılan sistemlerin izlediği basamakları da sırasıyla göstermektedir: insan tespiti, insan takibi ve yapılan eylemin anlaşılması. Bizim açımızda önemli olan yapılan eylemin anlaşılması konusunda verilen çalışmalarda sıkça değinilen yapılar ise şablon eşleştirmesi, optik akış ve “hidden markov models”dir. Şablon eşleştirmesi basitliğiyle, optik akış ise karmaşıklık ve zaman almasıyla, HMM ise komplike yapısı ve hesaplama isteğinin fazlalığıyla eleştirilmesiyle beraber sunulan tüm çalışmalarda tam ve başarılı çözümler bulunmamaktadır. Şekil 2’de Liang Wang ve diğerlerinin, HMM kullanılan çalışmaları tanımlamak için verdikleri örnek gösterilmektedir. Bu örnekte silüet anahtar framer arasındaki geçiş olasılıkları verilmiştir.



Şekil 1. 2D Vücut Kısımları Örneği



Şekil 2. Oturma Eyleminin Geçiş Olasılıklarıyla HMM Örneği

Thomas B. Moeslund ve diğerleri (2006) yaptıkları çalışma da, 2000-2006 yılları arasındaki insan hareket analizi çalışmalarına değinmiştir [3]. Bu zamana kadar yapılan çalışmalarda insan takibi ve poz tahmini alanlarında başarılı sonuçlar elde edildiği belirtilmiştir. Fakat tanıma konusunda çok farklı ve çeşitli yaklaşımlar olmasına rağmen bu konunun daha olgunlaşmadığı vurgulanmıştır. Sundukları araştırma kapsamında silüet tabanlı özellikler ve bunları kullanan çalışmalar ön plandadır. Hareket enerji resimleri tabanlı (zaman) çalışmalar ise bunları takip etmektedir.

Yakın zamanda yapılan arařtırmalar incelendiğinde hareket özellikleri çıkartılarak bu özelliklerin yorumlayacak sistemler ve çok yeni sayılabilecek az sayıda yayın içeren Ɣ dönüşüm yüzeyi tabanlı sistemler karşımıza çıkmaktadır. P.Viola ve diğeri 2003 yılında hareket ve görünüş dokularını kullanarak yayaları tespit etmişlerdir [4]. Bu çalışmanın dikkat çeken özelliği günümüzde insan eylemlerinin tanınması konusunda yapılan hareket tabanlı sistemler tarafından sıkça referans verilen bir kaynak olmasıdır. Sunulan sistemler bu çalışma kapsamında ilerleyen kısımlarda detaylı olarak anlatılmakta beraber çeşitli sahnelerde çekilmiş yayaların resimlerinden görünüş tabanlı filtreler aracılığıyla özellikler çıkartmaktadır. Bu çıkartılan özelliklerin yanında yürüyen yayaların hareket özellikleri yakalamak amacıyla takip eden iki frame arasındaki hareket örüntüleri bulunmaktadır. Çıkartılan bu hareket ve görünüş temelli özellikler boosting çatısı altında birleştirilerek sistem en verimli özelliklerle eğitilmiştir. Boosting olarak AdaBoost algoritması kullanılarak ikili sınıflandırma yapılmıştır (yaya ve yaya değil şeklinde). Hareket örüntüleri görünüş temelli özelliklere baskın çıkmıştır. Sistemin başarı düzeyi verilen test seti üzerinde %95 seviyesindedir. Ying Wang ve diğeri tarafından 2007 yılında insan eylemlerinin tanınması için Ɣ dönüşümü ve HMM tabanlı bir sistem sunmuşlardır [5]. Bu sistemde siyah-beyaz silüet resimlerde Ɣ dönüşümü ile düşük seviyede özellikler elde edilmiştir. Elde edilen özelliklere PCA (Principal Component Analysis) uygulanarak anlamlı özellikler seçilmiş ve bu özellikler zaman bilgisi ile pekiştirilmesi açısından eğitim ve test aşamasında HMMs (Hidden Markov Model) kullanılmıştır. Sistem eğilme, yürüme, taşıma, zıplama ve hızlıca hareket etme eylemlerinde test edilmiş ve verilen veri seti bazında %90 civarında başarı elde edilmiştir. R. Souvenir ve J. Babbs tarafından 2008 yılında 13 farklı eylemi tanıyan Ɣ dönüşüm yüzeyi tabanlı bir sistem sunmuşlardır [6]. Sunulan sistem Ɣ dönüşümü ile frame bazında düşük seviye özellikler bulunmuş daha sonra bu özellikler Ɣ dönüşüm yüzeylerine dağıtılmıştır. Sistemde tek bir kamera açısı kullanılmakla beraber bu bakış noktasındaki yüzeyler farklı bakış noktalarındaki yüzeylerin tahmin edilmesinde kullanılmıştır. Elde edilen özelliklerin boyut indirgemesinde isomap algoritması kullanılmıştır. Boyut indirgesini takiben veri miktarı da azaltıldıktan sonra (downsampling) eğitim örnekleri, test örnekleri ile karşılaştırılmıştır. Bazı eylemler test aşamasında yanlış sınıflara atansa da genel olarak başarı düzeyi %80 düzeyindedir.

Sunuş kısmında konunun ana ve alt problemleri tanımlanmış, geçmişten günümüze gelen konu üzerindeki çalışmalarda nelere yoğunlaştığı gözlemlenmiştir. Sonuç olarak bu seminer-2 çalışmasında yapmayı hedeflediğimiz insan eylemlerinin tanınmasında öne çıkan hareket örüntüleri temelli yaklaşımı ve Ɣ dönüşüm yüzeyi temelli yaklaşımı deneyerek bunlardan gelecekteki çalışmamız için bir yol haritasının çıkartılmasıdır. Çalışmamızda açık bir veritabanı (IXMAS) kullanılarak insan eylemleri veritabanı oluşturulmuş ve oluşturulan bu veritabanında Ɣ dönüşüm yüzeyi özellikleri ve hareket örüntüleri çıkartılmıştır. Her iki yaklaşım sonucunda elde edilen özellikler boosting yapısı kullanılarak sistemimiz eğitilmiş ve test sonuçları elde edilmiştir. İzlenen yöntemlere ufak katkılarda bulunularak başarı sonuçlarımızın artırılması hedeflenmiştir.

2. Veri Seti

Çalışmamız kapsamında kullanılan veri seti IXMAS'dır. IXMAS "INRIA Xmas Motion Acquisition Sequences" baş harfleriyle tanımlanan 2006 yılında tamamlanıp genel kullanıma açılan ücretsiz bir video veritabanıdır. IXMAS veritabanında 12 aktörün 3 farklı diziliş ve konum bakımından icra ettiği 14 farklı eylemin videoları bulunmaktadır. Bu eylemler aktör bazında tek bir videoda ardı ardına icra edilmiştir. 5 kamera farklı noktalara yerleştirilmiştir. Yapılan çekimler .png resim dosyası olarak, .pbm formatında silüet resimler ve frame bazında eylemleri gösteren doğruluk tablosu bulunmaktadır.

IXMAS veritabanından rastgele 8 aktör seçildikten sonra, bu aktörlerin yaptığı 2 eylemin videolarının frameleri alınmıştır. Bu iki eylem hem tanınması hem de kolayca fark edilmesi açısından dairesel yürüme ve el sallama eylemleri olarak seçilmiştir. Seçilen 8 aktörün 2 eylemini barındıran frameler aktör bazında test ve eğitim setleri olarak bölümlendi. Bu bölümlenme de farklı 6 aktörün eylemleri eğitim setini geriye kalan 2 farklı aktörün eylemleri de test seti olarak tanımlandı. Bu işlemler için sadece beş kameradan alınan video yerine tek bir kameradan alınan video tercih edildi. Yapılan seçimde aktörlerin seçilen kameraya karşı duruşları farklıydı. Şekil 3'de IXMAS veritabanından alınan bir örnek frame verilmiştir.

(<https://charibdis.inrialpes.fr/html/sequences.php> adresiyle IXMAS'a internetten erişilebilmektedir.)



Şekil 3. IXMAS Veritabanından Alınan Bir Örnek Frame

3. İncelenen Yöntemler

3.1. Radon Dönüşümü

Radon dönüşümü S.R.Deans tarafından 1983'de tanımlanmıştır [7]. Tanım olarak bir resmin Radon dönüşümü bu resmin değişik açılarda alınan doğrular boyunca oluşturulan bir projeksiyon kümesiyle belirlenmektedir. Daha basitçe resmin belirtilen yöndeki projeksiyonlarını hesaplar da denilebilir. Dijital resim işlemede, Hough dönüşümü gibi resimdeki doğruların bulunmasında kullanılır.

Radon dönüşümü, $I(x,y)$ resim ve δ Dirac delta fonksiyonu (giriş 0 ise çıkış 1, diğer durumda tam tersi) olmak üzere, $g(p,\theta)$ olarak denklem (1) ile ifade edilmiştir. Bu formülden $g(p,\theta)$ nun I resminin verilen polar koordinat ile tanımlanan doğru üzerindeki integrali olduğu da görülmektedir.

$$G(\rho,\theta) = \sum_x \sum_y I(x,y)\delta(x\cos\theta + y\sin\theta - \rho). \quad (1)$$

Radon dönüşünün bir çok özelliği bulunmaktadır. Bunlardan bazıları şeklin sunumuyla ilgilidir. Bunlardan bizim için önemli olanları kısaca sıralarsak resim bazında ölçeklemeye (scaling), yer değiştirmeye (translation) ve dönmeye (rotation) duyarlı olmasıdır. Bu duyarlılık faz bazında düşünülmelidir. Aşağıda verilen Şekil 4'te siyah beyaz bir 2 boyutlu şekil ve onun Radon dönüşüm sonucu verilmiştir.



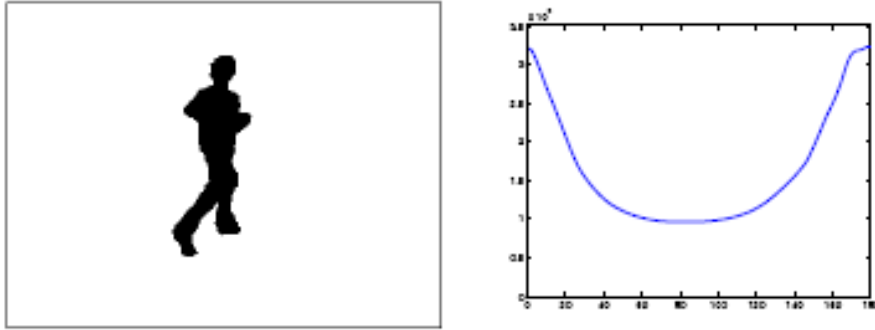
Şekil 4. 2D Bir Resim ve Bu Resmin Radon Dönüşümü

3.2. \mathfrak{R} Dönüşümü

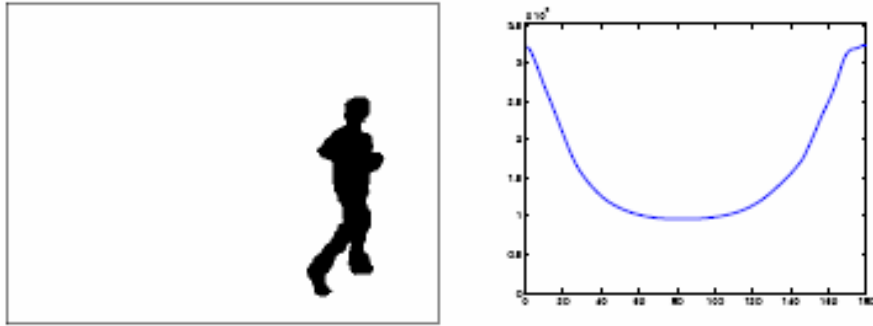
Radon dönüşümünün ölçeklemeye, yer değiştirmeye ve dönmeye duyarlılığını kaldırmak için bazı çalışmalar yapılmıştır. Bu çalışmalar sonucunda S.Tabbone ve diğerleri tarafından 2006 yılında Radon dönüşümü geliştirilerek \mathfrak{R} dönüşümü tanımlanmıştır [8]. $g(p,\theta)$ polar koordinatlarla tanımlanan Radon dönüşümü olmak üzere $\mathfrak{R}(\theta)$ ile \mathfrak{R} dönüşümü denklem (2) ile verilmiştir:

$$R(\theta) = \sum_p g^2(\rho,\theta). \quad (2)$$

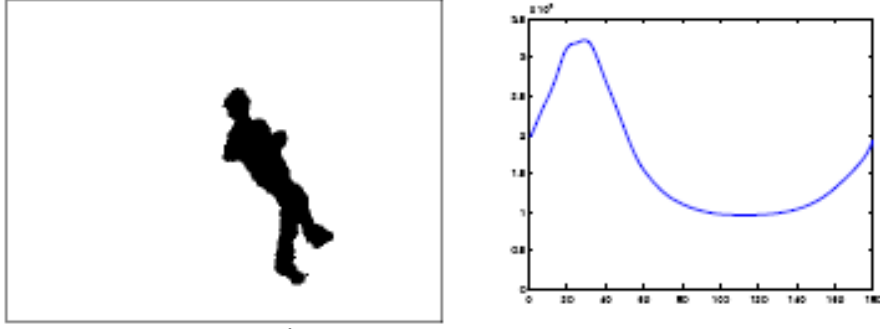
\mathfrak{R} dönüşümü bu haliyle Radon dönüşümdeki bazı eksiklileri gidermiştir. Şekil 5'te verilen yürüyen adam şekli ve bunun 0-180 derece aralığında \mathfrak{R} dönüşümünün sonucu bulunmaktadır. Şekil 6'da verilen aynı yürüyen adam şeklinin düzlemde yer değiştirmesi sonucunda elde edilen \mathfrak{R} dönüşümünün sonucunun Şekil 5'tekinden farkının olmadığı gözlemlenmektedir. Böylece \mathfrak{R} dönüşümünün yer değiştirme durumunda aynı sonucu verdiği anlaşılmaktadır. Şekil 7'de ise aynı şekil bu sefer -30 derece döndürülmüş ve \mathfrak{R} dönüşümü incelenmiştir. \mathfrak{R} dönüşümünde -30 derece faz değişiminin olduğu görülmektedir. Bu insan hareketlerinde sık görülmeyen bir durumdur. Şekil 8'de küçültülmüş olarak verilen yürüyen insan şekli ve bunun \mathfrak{R} dönüşümü görülmektedir. Ölçekleme \mathfrak{R} dönüşümünün şeklini değiştirmemiş fakat genlik boyutla orantılı bir şekilde değişmiştir.



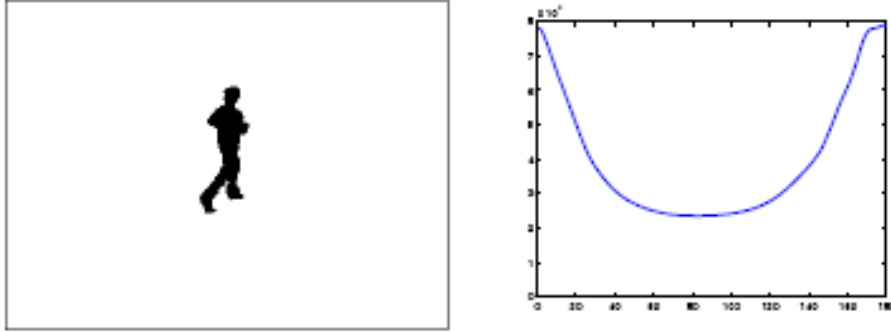
Şekil 5. Orijinal Resim ve \mathfrak{R} Dönüşümü



Şekil 6. Yer Değiştirilmiş Resim ve \mathfrak{R} Dönüşümü



Şekil 7. -30^0 Döndürülmüş Resim ve \mathfrak{R} Dönüşümü



Şekil 8. Küçültülmüş Resim ve \mathfrak{R} Dönüşümü

\mathfrak{R} dönüşümü yer değiştirmeye karşı değişmez bir halde bulunurken ölçeklemeye karşı değişmezliği denklem (3) ile normalize edilerek sağlanmaktadır.

$$R'(\theta) = \frac{R(\theta)}{\max_{\theta'} (R(\theta'))} \quad (3)$$

Özetle \mathfrak{R} dönüşümünün Radon dönüşümüne oranla ölçekleme ve yer değiştirme gibi geometrik dönüşümlerde daha dayanıklı bir yapı sunduğu ortadadır.

3.3. \mathfrak{R} Dönüşüm Yüzeyi

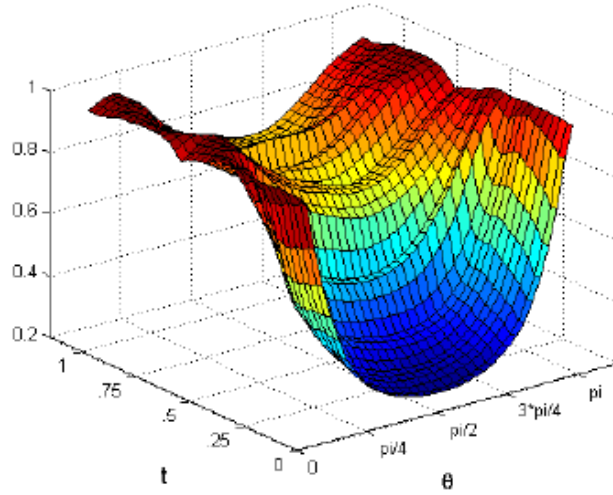
Bir eylemi tasvir etmek için o eylemi temsil eden videodaki her frame'e tek tek \mathfrak{R} dönüşümü uygulanır. Çıkan sonuçlar ise frame bazında normalize edilir. Bu elde edilen sonuçlar zaman ekseninde ardı ardına eklenilerek bir yüzey elde edilmektedir. Bu yüzeye \mathfrak{R} dönüşüm yüzeyi denilir. t frame'in zamanını göstermek üzere $\mathfrak{R}'(\theta)$ 0-180 derece aralığında normalize \mathfrak{R} dönüşümü ise $I(x,y,t)$ ile verilen silüet resimlerinden oluşan videonun \mathfrak{R} dönüşüm yüzeyi aşağıdaki denklem (4) ile tanımlanır:

$$S(\theta, t) = R'_t(\theta). \quad (4)$$

Şekil 9’da bir aktörün oturma eylemini icra ederken seçilen bazı silüet frame’leri (anahtar frameler) verilmiştir. Şekil 9’da bazı silüet frameleri verilen tüm videodaki silüet frame’lere normalize \mathcal{R} dönüşümü uygulanıp sonuçlar ardı ardına eklenip elde edilen \mathcal{R} dönüşüm yüzeyi ise Şekil 10’da verilmiştir.



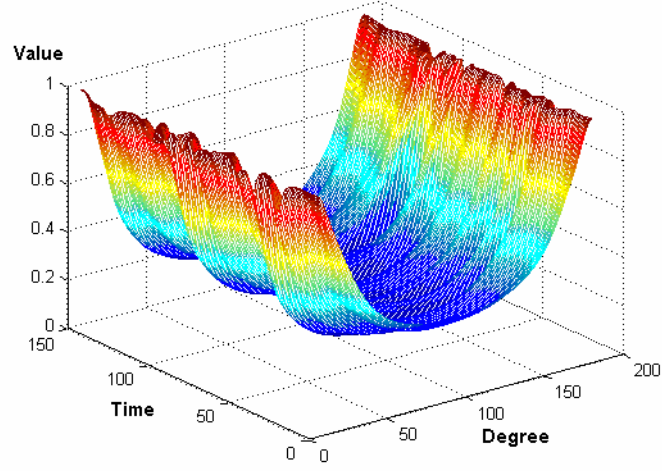
Şekil 9. Silüet Olarak Oturma Eyleminin Ana Frameleri



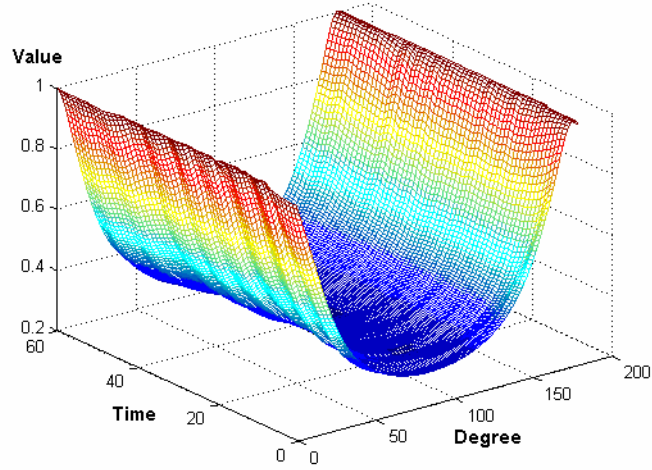
Şekil 10. Oturma Eyleminin \mathcal{R} Dönüşüm Yüzeyi

3.3.1. Çalışma Kapsamında \mathcal{R} Dönüşüm Yüzeyinin Kullanım Şekli

Deneysel sonuçlar elde edilmesi ve \mathcal{R} dönüşüm yüzeylerinin başarısının test edilmesi amaçlı olarak çalışma kapsamında eğitim ve test setindeki tüm el sallama ve yürüme frame’lerinden zaman sıralamaları da dikkate alınarak \mathcal{R} dönüşüm yüzeyleri oluşturulmuştur. Bu şekilde ki kullanımda elde edilen yüzeylere görüş bazlı \mathcal{R} dönüşüm yüzeyleri denilmiştir. Şekil 11’de seçilen bir aktörün yürüme eylemi için görüş bazlı \mathcal{R} dönüşüm yüzeyi, Şekil 12’de ise seçilen bir aktörün el sallama eylemi için görüş bazlı \mathcal{R} dönüşüm yüzeyi örnek olarak verilmiştir.

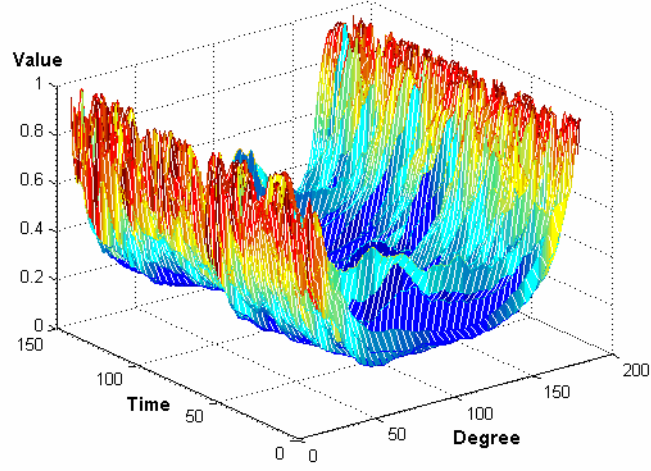


Şekil 11. Yürüme Eylemi için Görünüş Bazlı \mathfrak{R} Dönüşüm Yüzeyi

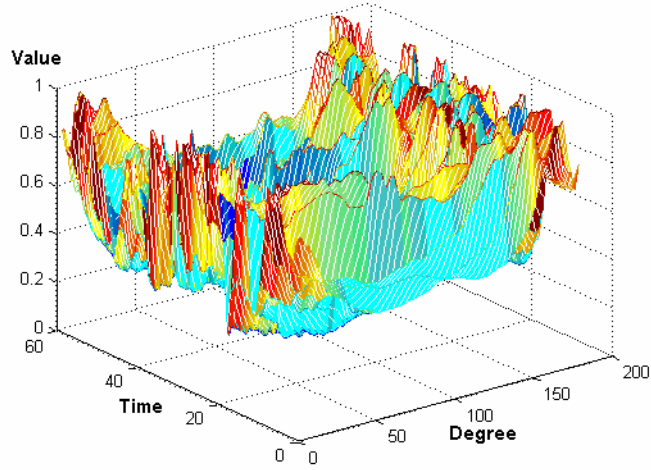


Şekil 12. Oturma eyleminin \mathfrak{R} dönüşüm yüzeyi

Görünüş bazlı \mathfrak{R} dönüşüm yüzeyleri tüm aktörler için ayrı ayrı oluşturulduktan sonra literatür kapsamında yapılmayan bir deneme gerçekleştirilmiştir. Bu deneme kapsamında aynı eylemi oluşturan ardışık silüet resimler arasındaki fark alınmış ve oluşan yeni silüet resimlerinden (fark) \mathfrak{R} dönüşüm yüzeyleri oluşturulmuştur. Şekil 13’de seçilen bir aktörün yürüme eylemi için fark bazlı \mathfrak{R} dönüşüm yüzeyi, Şekil 14’de ise seçilen bir aktörün el sallama eylemi için fark bazlı \mathfrak{R} dönüşüm yüzeyi örnek olarak verilmiştir.



Şekil 13. Yürüme Eylemi için Fark Bazlı \mathfrak{R} Dönüşüm Yüzeyi



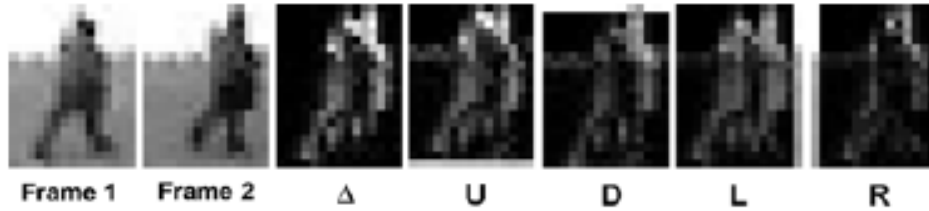
Şekil 14. El Sallama Eylemi için Fark Bazlı \mathfrak{R} Dönüşüm Yüzeyi

3.4. Hareket Örüntüleri

Sunuş kısmında anlatıldığı üzere hareket bazlı yaklaşımlar günümüzde insan eylemlerinin tanınması konusunda oldukça sık olarak karşımıza çıkmaktadır. Araştırmamız kapsamında hareket örüntüleri örneklerini de deneyerek sonuçlarını incelememizin faydalı olacağı düşünülmüştür. Bu kapsamda yoğun olarak referans verilen çalışmalardan önde gelen biri olan Paul Viola ve diğerlerinin (2003) çalışmasındaki hareket örüntüleri incelenmiştir.

Paul Viola ve diğerleri (2003), yayaları görünüş ve hareket bilgilerinden (yürüyüş) tespit etmek için bir sistem geliştirmiştir [4]. Hareket örüntülerini bulmak için Paul Viola ve Michael Jones (2001)'un statik yüz bulma amaçlı oluşturdukları sistemde kullandıkları filtreleri harekete uyarlamışlardır [9]. Bu sistemde hareket bilgilerini elde etmek için bir dizi işlem uygulanmaktadır. Bu işlemler aşağıda sırasıyla anlatılmıştır:

- Hareket örüntülerini yakalamak amacıyla çok çeşitli filtreler oluşturulmuştur. Bu filtreler hareket bilgisi üzerinde çalışmak üzere tanımlandığı için verilen çalışmada öncelikle filtrelerin çalışacağı 5 farklı resim yaratılmıştır. Bu resimler gri seviyedeki videolardan alınan ardışık frame'lerin farklarından oluşturulmuştur. Δ, U, R, L, D ile tanımlanan bu resimlerde: Δ ardışık iki frame arasındaki (aynı boyutta resimler) mutlak farkı hesaplanarak oluşturulmuştur, U ardışık iki frame'den ilk frame aynı kalmak kaydıyla ikinci frame bir piksel yukarı kaydırılmış ve bu haldeyken iki frame'nin mutlak farkı hesaplanarak oluşturulmuştur, R ardışık iki frame'den ilk frame aynı kalmak kaydıyla ikinci frame bir piksel sağa kaydırılmış ve bu haldeyken iki frame'nin mutlak farkı hesaplanarak oluşturulmuştur, L ardışık iki frame'den ilk frame aynı kalmak kaydıyla ikinci frame bir piksel sola kaydırılmış ve bu haldeyken iki frame'nin mutlak farkı hesaplanarak oluşturulmuştur, D ardışık iki frame'den ilk frame aynı kalmak kaydıyla ikinci frame bir piksel aşağı kaydırılmış ve bu haldeyken iki frame'nin mutlak farkı hesaplanarak oluşturulmuştur. Şekil 15'te ardışık iki frame ve bunlardan elde edilen Δ, U, R, L, D mutlak fark resimleri örnek olarak verilmiştir.



Şekil 15. Ardışık 2 Frame ve Bunlardan Elde Edilen Δ, U, R, L, D Mutlak Fark Resimleri

- Hareketin yönünü tespit etmek için $S=\{U, R, L, D\}$ olmak üzere r_i penceresinde mutlak farkların toplamı hesaplanılır. Aşağıdaki denklem ile ifade edilmiştir:

$$f_i = r_i(\Delta) - r_i(S). \quad (5)$$

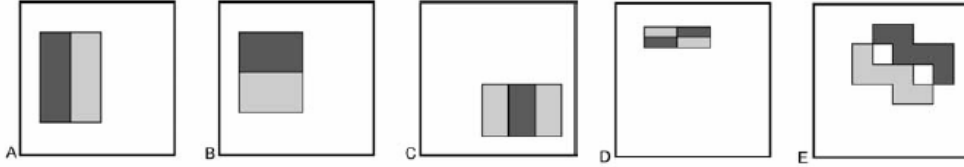
- Hareketin büyüklüğünü tespit etmek için $S=\{U, R, L, D\}$ olmak üzere r_k penceresinde toplam değer hesaplanılır. Aşağıdaki denklem ile ifade edilmiştir:

$$f_k = r_k(S). \quad (6)$$

- Çalışma kapsamında *motion shear* değerleri hesaplamak üzere çeşitli filtreler tanımlanmıştır. Şekil 16'da verilen filtreler $S=\{\Delta, U, R, L, D\}$ olmak üzere tüm fark

resimlerine uygulanmıştır. Filtreleme sonucunda oluşan resimdeki her pikselin renk değeri toplanmıştır. Bu işlem aşağıdaki denklem ile ifade edilmiştir:

$$f_i = \phi_j(S). \quad (7)$$



Şekil 16. Paul Viola ve Diğerleri Tarafından Tanımlanan Filtreler [4]

3.4.1. Çalışma Kapsamında Hareket Örüntülerinin Kullanım Şekli

Çalışmamızda Paul Viola ve diğerleri tarafından yürüyen yayaların hareket örüntüleri çıkartmak amaçlı kullanılan filtreleri kullandık. Kullanım şekli itibariyle farklı bir yaklaşım izledik. Paul Viola ve diğerleri kendi sistemlerinde hareketi tanımlamak amacıyla değişik videolardan yürüme eylemini içeren sadece 2 frame almışlardır. Biz kendi çalışmamızda bunu bir hareket dizisi kabul edip bu hareketi içeren tüm frameleri alıp inceledik. Paul Viola ve diğerleri kendi sistemlerinde çeşitli pencere boyutlarında arama yapmışlar ve milyonlarca özellik oluşturmuşlardır. Yaptığımız çalışmada ise bu pencere yapısı kullanılmayarak silüet resimler ve gri seviyedeki aynı resimler kullanılarak VE işlemi uygulanmıştır. Böylece çeşitli pencere boyutlarında aktör arama işlemine gerek kalmamıştır. Aşağıdaki Şekil 17’de bu VE işlemi sonucu oluşturulan örnek resimler verilmiştir.



Şekil 17. Silüet ve Orijinal resimlerin birleşimi

Eğitim ve test setindeki tüm framerde silüet ve orijinal resimler anlatıldığı üzere birleştirilmiştir. Her ardışık frame çifti için Paul Viola ve diğerleri (2003) tarafından tanımlanan Δ, U, R, L, D mutlak fark resimleri oluşturulmuştur. Daha sonra f_i, f_k ve f_j özellikleri hesaplanarak her frame çifti için 33 ayrı öznelik değeri oluşturulmuştur. Ayrıca görüntü piramidi yapısı kullanılarak aynı resimler 0,8 oranında 4 defa küçültülerek bu öznelik çıkartımı işlemi tekrar edilmiştir. Sonuçta her frame çifti için 165 öznelik elde edilmiştir. Bu işlem doğal olarak tüm eylemi içeren ardışık framer

için tekrar edilmiştir. Böylece aktörün yaptığı eylemi temsil etmek üzere bir öznitelik dizisi oluşturulmuştur.

Hareket örüntülerinden oluşan öznitelik dizileri oluşturulduktan sonra farklı bir yaklaşımda denenmiştir. Dikkat edildiği üzere yapılan Paul Viola ve diğerleri (2003) tarafından çalışmada herhangi bir normalizasyon değeri kullanılmamıştır. Ufak bir unsur olarak gözlenen bu açık yapılan bu çalışma içerisinde fark edilerek işleme katılmıştır. Normalizasyon değişkeni olarak seçilen değer her frame çifti için ayrı ayrı olarak hesaplanan Δ mutlak fark resminin büyüklüğüdür. Bir önceki paragrafta tanımlanan öznitelik dizileri bir de elde edilen değerlerin normalize hali alınarak tekrardan oluşturulmuştur.

4. Boosting

Boosting basitçe orta düzeyde başarılı zayıf hipotezleri birleştirerek (weak learner) yüksek başarılı bir hipotez oluşturma metodu olarak tanımlanır [10]. Günümüzde çok değişik alanlarda kullanılmakta olan çeşitli boosting algoritmaları mevcuttur. *Adaboost*, Freund ve Schapire'nin 1995 yılında tanımladığı bir boosting algoritmasıdır. Adaboost algoritması $(x_1, y_1), \dots, (x_m, y_m)$ şeklinde bir eğitim setini giriş parametresi olarak kabul eder. Her x_i değeri bir eğitim örneğinin öznitelikleri olarak düşünülebilir y_i değerleri ise etiket değerleridir. Bu etiket değerleri pozitif ve negatif sınıflar olarak tanımlanır (+1,-1). Adaboost verilen zayıf öğrenici tekrar tekrar $t=1, \dots, T$ çağırır. Algoritmanın ana amaçlarından birisi eğitim seti üzerinde bir dağılım veya ağırlık kümesini güncellemektir. İlk olarak her ağırlık eşittir ama her tekrarda (round) yanlış sınıflandırılan özelliklerin ağırlığı artırılarak zayıf öğrencinin bu eğitim setindeki zor örnekler üzerinde yoğunlaşması zorlanır. Zayıf öğrencinin görevi dağılıma uygun olan bir zayıf hipotez bulmaktır. Bu zayıf hipotezin iyiliği hatasıyla ölçülür. Tabii ki hata ölçümünde bu dağılım yerine ağırlıklarda kullanılabilir. Zayıf hipotez oluşturulduğunda Adaboost bu zayıf hipoteze bir önem değeri atar. Bu önem değeri hata oranı ile hesaplanmaktadır. Sonuçta test setindeki örnekler için bir oylama yapısı kullanarak her bir zayıf hipoteze verdiği sonucun ağırlıkları toplanarak bu örneğin sınıfı bulunur.

Adaboost algoritmasının avantajları olarak gerçekleştirilmesinin basitliği, çeşitli öğrenme görevlerinde rahatlıkla uygulanabilmesi, büyük öznitelik kümelerinde başarılı bir şekilde öznitelik seçimini yapabilmesi, pratikte *overfit* problemi yaşamaması olarak sıralanabilir. Adaboost algoritması greedy yaklaşımından ötürü tam optimal sonucu vermeyebilmektedir. Adaboost algoritmasından değişik yaklaşımlarla yeni algoritmalar çıkartılmıştır. Çalışma kapsamında kullanılanlar ise Real Adaboost ve Gentle Adaboost olarak tanımlanan iki sınıf tabanında çalışan ve ağaç bazlı zayıf öğrenici kullanan iki örneğidir. Bunlardan ilki *Real Adaboost*, Adaboost algoritmasının genelleştirilmesi olarak bilinmekte ve temel boosting algoritması olarak sayılmaktadır [11]. *Gentle Adaboost*, Real Adaboost algoritmasının daha stabil ve güvenilir halidir. Gürültü içeren verilerde daha başarılı sonuçlar vermektedir [12].

5. Deneysel Sonular

Yapılan alıřma kapsamında birbirinden farklı iki yaklařım izlenmiřtir. Bunlar \mathcal{R} dnüşüm yzeyleri ve hareket rüntüleridir. Bu iki yaklařımın nasıl uygulandıđı ilgili bařlıklar altında verilmiřtir. Silüet bazlı resimlerden elde edilen görünüş ve fark bazlı \mathcal{R} dnüşüm yzeylerinde eylemi temsil eden her frame için 180 adet farklı znitelik bulunmaktadır. Kullanılan veri seti bařlıđında da anlatıldıđı üzere rastgele seilen 6 aktrn yrme frameleri pozitif rnekler, el sallama frameleri negatif rnekler olarak seilerek eđitim seti olarak boosting algoritmalarına verilmiřtir. Hareket rüntleri için de eđitim ve test setindeki yrme ve el sallama rnekleri aynı mantık erevesinde ele alınmıřtır. Fakat bunlarda ardıřık iki framenin farkı kullanıldıđından veri seti 16 frame daha azdır. Durma parametresi olarak 200-100 iterasyon veya eđitim setindeki hata oranının %0 olması seilmiřtir. Eđitim kısmı tamamlandıktan sonra farklı iki aktrn eylemlerini ieren test setindeki framelerin etiketlenme bařarısı incelenmiřtir. Boosting bařlıđı altında anlatılan iki ayrı boosting algoritması kullanılmıřtır.

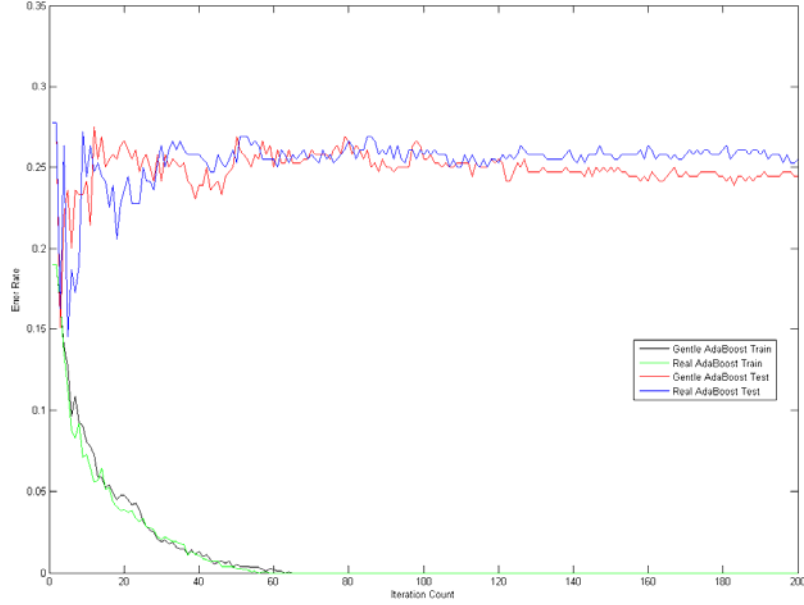
řekil 18’de görünüş bazlı normalize \mathcal{R} dnüşümüyle her frame için oluřturulan 180 adet znitelik boosting algoritmalarına belirtilen eđitim setiyle verilmiřtir. Yaklařık 66. iterasyonda eđitim setindeki hata oranı ok dřük bir dzeye gelmesine karřın bu durumda oluřan zayıf hipotezler kullanılarak test seti için sınıflandırılma yapıldıđında bařarı oranı %74-%76 aralıđındadır. Bu deneyde kullanılan zayıf đrenici 3 katmanlı bir ađatır.

řekil 19’da görünüş bazlı normalize \mathcal{R} dnüşümüyle her frame için oluřturulan 180 adet znitelik boosting algoritmalarına belirtilen eđitim setiyle verilmiřtir. Son iterasyonda eđitim setindeki hata oranı ok dřük bir dzeye gelmesine karřın bu durumda oluřan zayıf hipotezler kullanılarak test seti için sınıflandırılma yapıldıđında bařarı oranı %76-%77 aralıđındadır. Bu deneyde kullanılan zayıf đrenici 1 katmanlı bir ađatır.

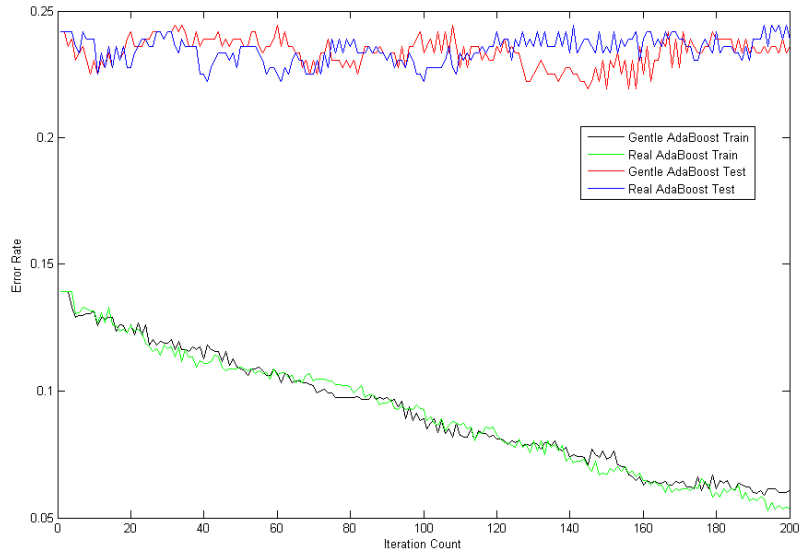
řekil 20’de görünüş ve fark bazlı normalize \mathcal{R} dnüşümleriyle her frame için oluřturulan znitelikler birleřtirilmiřtir (180+180). Buradaki amacımız boosting algoritmamızın hem fark hem de görünüş bazlı zelliklerden baskın olanları semesidir. Son iterasyonda eđitim setindeki hata oranı ok dřük bir dzeye gelmesine karřın bu durumda oluřan zayıf hipotezler kullanılarak test seti için sınıflandırılma yapıldıđında bařarı oranı %72-%76 aralıđındadır. Bu deneyde kullanılan zayıf đrenici 1 katmanlı bir ađatır.

řekil 21’de hareket rüntleri bařlıđında detaylı olarak anlatılan znitelikler kullanılmıřtır. Eylemleri temsil eden tm frameler için 135 adet znitelik seilmiř ve bu tm eđitim seti ve test seti için hesaplandıktan sonra bunlar boosting algoritmalarına verilmiřtir. Yaklařık 70. iterasyonda eđitim setindeki hata oranı ok dřük bir dzeye gelmesine karřın bu durumda oluřan zayıf hipotezler kullanılarak test seti için sınıflandırılma yapıldıđında bařarı oranı %95-%93 aralıđındadır. Eđitim setindeki hata oranı azaltılmaya alıřıldıđında test setindeki bařarı oranı dalgalanma gstermekle beraber %94-%92 aralıđındadır. Bu deneyde kullanılan zayıf đrenici 3 katmanlı bir ađatır.

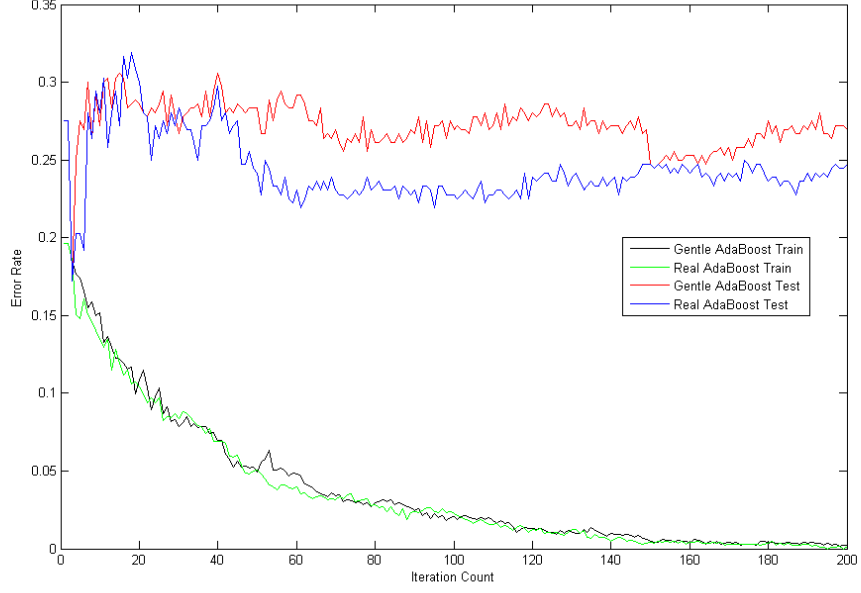
Şekil 22’de hareket örüntüleri başlığında detaylı olarak anlatılan normalize edilmiş öznelikler kullanılmıştır. Bu durumda son iterasyonda test seti üzerindeki başarı oranı %94-%95 aralığındadır. Bu deneyde kullanılan zayıf öğrenici 3 katmanlı bir ağaçtır.



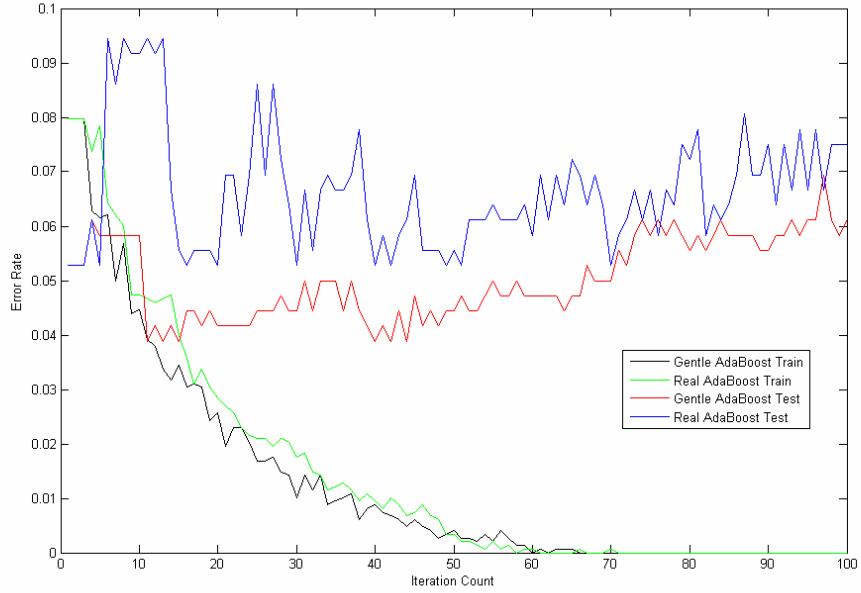
Şekil 18. Görünüş Bazlı Normalize \mathcal{R} Dönüşümünün Boosting Sonucu Başarı Grafiği



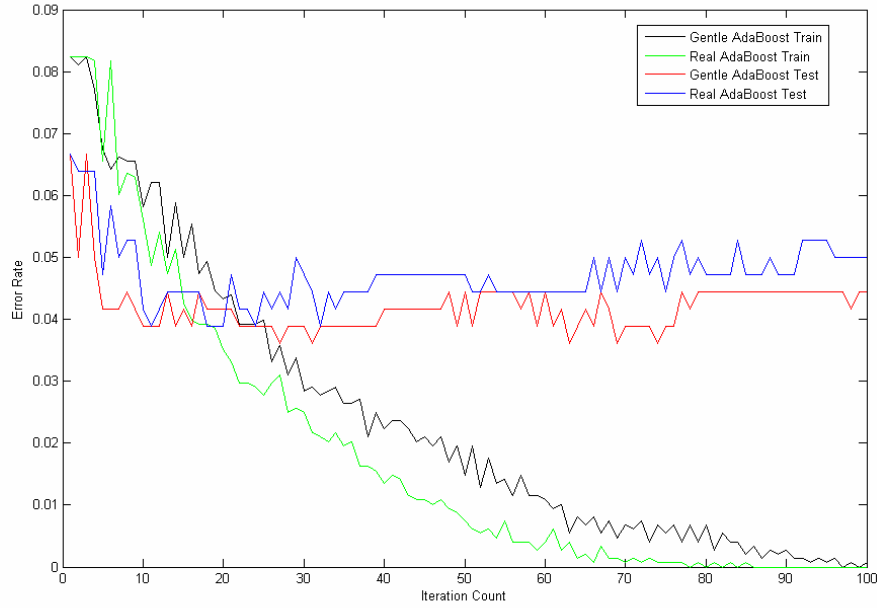
Şekil 19. Fark Bazlı Normalize \mathcal{R} Dönüşümünün Boosting Sonucu Başarı Grafiği



Şekil 20. Fark ve Görünüş Bazlı Normalize \mathfrak{R} Dönüşümü Birleşiminin Boosting Sonucu Başarı Grafiği



Şekil 21. Hareket Örüntülerinden Elde Edilen Öznitelikler Kullanıldığında Boosting Sonucu Başarı Grafiği



Şekil 22. Hareket Örüntülerinden Elde Edilen Normalize Öznitelikler Kullanıldığında Boosting Sonucu Başarı Grafiği

Şekil 22'deki grafikte verilen sonuçlar incelendiğinde sistemin bu eğitim ve test seti bazında iyi sonuçlar ürettiği söylenebilir. Bu durumda son iterasyonda en başarılı sonucu veren Gentle AdaBoost algoritmasının seçtiği ilk beş zayıf öğrenicinin kullandığı özniteliklerin neler olduğu anlamak için resim piramidi bazında kaçınıcı seviyedeki resmi kullandığı, hangi karşılaştırma resmini seçtiği ve hangi filtreyi kullandığını aşağıdaki tabloda görmekteyiz.

Filtre	Hedef	Ölçek
C	U	3.seviye
D	U	1.seviye
C	U	5.seviye
Fk	D	1.seviye
B	D	1.seviye

Tablo 1. Gentle AdaBoost Algoritmasının Seçtiği İlk Beş Zayıf Öğrenici

6. Sonuç ve Gelecekte Yapılacak Çalışmalar

Bu çalışma kapsamında günümüzde Bilgisayarla Görme alanında en güncel konulardan biri olan videolarda insan eylemlerinin incelenmesi konusuna değinilmiştir. Bu konu kapsamında günümüzde karşımıza çıkan iki önemli yaklaşım ele alınmıştır. 3 dönüşüm yüzeyleri silüet resimlerde çalışan ve bu konuda çok fazla yayın bulunmayan yeni bir yaklaşımdır. Deneysel sonuçlar incelendiğinde hem görünüş hem de hareket temelli 3 dönüşüm yüzeylerinin çok başarılı sonuçlar vermediği ortadadır. Gelişmekte olan bu yaklaşımın en avantajlı noktaları gürültü, kesik/eksik bilgi içeren silüet resimlerde daha az performans ve başarı kaybı yaşanmasıdır. Ayrıca 3 dönüşüm yüzeylerinin silüet resimlerde çalışmasından dolayı kaynaklanan renk ve dolayısıyla bilgi kaybı bu sistemlerin başarı oranlarını etkilemektedir. Her ne kadar yer değiştirme ve ölçeklemeye karşı duyarsız bir yaklaşım olmasına rağmen aktörün kameraya karşı duruş açısına çok duyarlı olduğu yapılan deneyler sonucunda tekrar ortaya çıkmıştır. Sonuçta elde edilen 3 dönüşüm yüzeylerinin tam olarak yorumlanamaması ise açık bir noktadır. Hareket örüntüleri ise orijinal yaklaşımda sadece 2 ardışık frameyi incelemesinden ötürü eylemi tam olarak tasvir edememektedir. Yaptığımız çalışmada bu yaklaşımı tüm eylem üzerine uyguladığımızda elde ettiğimiz sonuç ümit vericidir. Hareket örüntüleri kayıplı ve/veya gürültü içeren veride büyük bir performans kaybı yaşamakta ve yüksek düzeyde eğitim verisine ihtiyaç duymaktadır. Boosting konusu günümüz çalışmalarda çok değişik alanlarda karşımıza sıkça çıkmaktadır, temel amacı boyut indirgemesi merkezli olarak anlamlı özelliklerin seçimidir. Fakat girişte verilen özelliklerin belirleyici olması sistemlerin genel performanslarını doğrudan etkilemektedir.

Gelecekte yapılması planlanan çalışmalar ise varolan hareket örüntülerinin daha da geliştirilerek birbirlerine benzer eylemlerin (yürüme, koşma, hızlı yürüme v.b.) ayırt edilmesinde çoklu sınıflandırmayı destekleyecek kuvvetli özelliklerin tespit edilmesidir. 3 dönüşüm yüzeylerinin bize sağladığı yüksek boyutlu verinin daha az boyutlu ve anlamlı bir şekilde ifade edilmesinin sağlanması ise bir diğer hedefimizdir. Nihai amacımız ise bu iki yöntemi daha da kuvvetlendirerek birbirlerinin zayıf yönlerini kapatacak bir şekilde birleştirmektir şeklinde özetlenebilir. Bu amaçlar tamamlandıktan sonra ise önerilen yapının gerçek zamanda veya buna en yakın performansta çalıştırılabilir hale getirilmesi ise bu çalışmanın son noktası olarak görülmektedir.

Kaynakça

- [1] D.M. Gavrilu, “*The Visual Analysis of Human Movement: A Survey*”, Computer Vision and Image Understanding: CVIU, vol.73, no.1, pp.82-98, 1996.
- [2] Liang Wang, Weiming Hu ve Tieniu Tan, “*Recent Developments in Human Motion Analysis*”, Pattern Recognition Letters, vol.26, no.3, pp.355-368, 2002.
- [3] Thomas B.Moeslund, Adrian Hilton ve Volker Krüger, “*A Survey of Advances in Vision-based Human Motion Capture and Analysis*”, Computer Vision and Image Understanding, vol.104, no.2, pp.90-126, 2006.
- [4] P. Viola, M.J. Jones ve D. Snow, “*Detecting Pedestrians Using Patterns of Motion and Appearance*”, IEEE International Conference on Computer Vision (ICCV), vol.2, pp.734-741, 2003.
- [5] Ying Wang, Kaiqi Huang ve Tieniu Tan, “*Human Activity Recognition Based on R Transform*”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007), pp.1-8, 2007.
- [6] R. Souvenir ve J. Babbs, “*Learning The Viewpoint Manifold for Action Recognition*”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), pp.1-7, 2008.
- [7] S.R. Deans, “*Applications of the Radon Transform*”, Wiley Interscience Publications, New York, 1983.
- [8] S. Tabbone, L. Wendling ve J.-P. Salmon, “*A New Shape Descriptor Defined on The Radon Transform*”, Computer Vision and Image Understanding, vol.102, no.1, pp.42-51, 2006.
- [9] Paul Viola ve Michael Jones, “*Robust Real-time Object Detection*”, International Journal of Computer Vision, 2001.
- [10] Yoav Freund ve Robert E. Schapire, “*Game Theory, On-Line Prediction and Boosting*”, Annual Workshop on Computational Learning Theory, pp.325-332, 1996.
- [11] Robert E. Schapire ve Yoram Singer, “*Improved Boosting Algorithms Using Confidence-Rated Predictions*”, Annual Workshop on Computational Learning Theory, pp.80-91, 1998.
- [12] Jerome Friedman, Trevor Hastie ve Robert Tibshirani, “*Additive Logistic Regression: A Statistical View of Boosting*”, Annals of Statistics, vol.28, no.2, pp.337-407, 2000.