

VIDEO DİZİLERİNDE HAREKET TANIMA

Celal Murat KANDEMİR¹, Nihat ADAR²

ÖZET: Bu çalışmada, video dizilerinden insan hareketlerinin tanınması için yapay sinir ağı (YSA) ve saklı markov modelini (SMM) temel alan insan hareket tanıma sistemleri modellenmiştir. Hareket tespit ve tanıma sistemi üç aşama içermektedir. Birinci aşamada, video çerçevelerinde insan pozunu tespit edilmektedir. İkinci aşamada, hareketlere ait poz dizileri elde edilmektedir. Üçüncü aşamada, YSA ve SMM tabanlı tanıma modelleri hareket tanıma için kullanılır. Önerilen YSA modeli yüksek tanıma oranlarına sahiptir. Bununla birlikte, SMM modeli için yeni eylemler önceki eğitilmiş SMM'lerde herhangi bir değişiklik yapmadan kolayca eklenebilir.

ANAHTAR KELİMELEER: Hareket Tanıma, Poz Sembolü, Poz Dizisi, YSA, SMM.

ACTION RECOGNITION IN VIDEO SEQUENCE

ABSTRACT: In this study, human activity recognition systems based on Artificial Neural Networks (ANN) and Hidden Markov Model (HMM) are modeled for recognition of the human activities from video sequences. The action recognition system models consist of three stages. At the first stage, human pose is detected in video frames. At the second stage, the pose sequences are obtained. At the third stage, ANN, and HMM based recognition models are used for action recognition. Suggested ANN model has higher recognition rate. However, for the HMM models new actions can be easily added without making any changes in the previous trained HMM.

KEYWORDS: Action Recognition, Pose Symbols, Pose Sequence, ANN, HMM.

¹ Eskişehir Osmangazi Üniversitesi, Eğitim Fakültesi, Bilgisayar Öğretim Teknolojileri Bölümü, Meşelik Kamp., ESKİŞEHİR

² Eskişehir Osmangazi Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilg.Müh.Böl., 26480 Meşelik Kamp., ESKİŞEHİR

I. GİRİŞ

Günümüzde video görüntülerinden oluşan veri miktarı güvenlik ve izleme kamera kayıtları gibi yaygın kullanım nedeniyle çok artmıştır. Bu kadar çok verinin gerçek zamanlı ve doğru olarak anlamlandırılabilmesi ancak bilgisayar tabanlı aktivite tanıma sistemleri ile mümkün olabilmektedir. Aktivite tanıma yöntemlerinin zamansal ve yersel küçük değişikliklere karşı gürbüz bir davranış sergilemesi için ihtiyaç duyduğu iki önemli gereksinimi öğrenme ve tanıma algoritmalarıdır.

Yapay sinir ağları, zamanla değişen verilerin analizinde sıklıkla kullanılan bir yöntemdir. Bununla birlikte ağların eğitilmesi için genellikle çok büyük veri kümelerine ihtiyaç duyulmaktadır. Zaman-gecikme işlevselliğinin dahil edilebilmesi amacıyla çalışmalar yapılmıştır. YangveAhuja zaman-gecikmeli yapay sinir ağlarını el hareketi tanımda uygulanmış ve yüksek tanıma oranları elde etmiştir [1].

Video dizisindeki imgeler ile ilgilenildiğinde test ve referans çerçeveler arasındaki birbiriyle uyuşmayan zaman ölçeğinde çalışılabilmesi bir avantaj sağlamaktadır. Myersvd. Q uzunluğundaki test örüntüsü ve P şablonun referans örüntüleri verildiğinde test ve referans örüntüleri arasında en iyi eşleşmenin dinamik programlama ile bulunulabileceğini ileri sürmüştür [2]. Raovd. çalışmasında, video dizisindeki hareketli nesnelere ait yörüngelerin yer-zaman eğrilerini hesaplamış ve dinamik zaman eğritmesi yöntemi kullanarak bunları eşleştirmiştir [3]. Genel anlamda dinamik zaman eğritmesi, iki farklı işaret arasındaki örüntü benzerliğinin, kesin kısıtlamalar altında dinamik eşleşmesi yoluyla hesaplanmasında kullanılmaktadır. Bu yöntem ses tanımadan sonra insan aktivitelerinin tanınmasında uygulanmıştır [4].

Fourier dönüşümü, temel bileşenler analizindeki gibi veri ayrıştırmak için kullanılabilir. Cunadovd. yürümeyi gösteren imge dizisindeki insan bacak hareketine ait eğrileri Hough dönüşümü yardımıyla çıkarmıştır. Bu eğrilerin eğimi basit bir harmonik hareketi takip ettiği görülmüş ve bu harmonik hareketler yürüyüş parametreleri olarak kullanılmıştır. Eğrilerdeki kayıp noktalar en küçük kareler yöntemiyle birleştirilerek bacak eğrilerinin eğim değişikliklerini göstermek hareketleri tanımak için Fourier dönüşümden yararlanılmıştır [5]. Diğer bir çalışmada ise insan silüetinin uç noktalarından faydalanılarak çıkartılan yıldız gösterime fourier dönüşüm uygulanmasıyla yürüyüş çevrimlerinin sezimi ve tanınması sağlanmıştır [6].

Şablon eşleme tekniğinde, bir imge dizisi durağan şekil örüntülerine dönüştürülür ve bu örüntüler önceden tanıma için hazırlanmış örüntü şablonları ile karşılaştırılarak uygun eşleşmenin bulunması sağlanır. Şablon eşleme yöntemi, harekete ait dinamik karakteristiklerin kaybolmasına sebep olabilmekte ve hareketin süresindeki bir değişime çok hassas olmaktadır. Bobick vd. insan hareketini tanımak için hareket geçmişi imgeleri ve hareket enerji imgeleri olmak üzere iki zamansal şablon yaratmıştır. Hareket

enerji imgeleri (HEİ), ikilik imgeler olup imge dizisi içinde hareketin olduğu yeri göstermektedir. Hareket geçmişi imgeleri (HGİ) ise yerel hareket geçmişinin bir fonksiyonu olarak her bir bölgedeki ışık şiddetinin gösterildiği skalar imgelerdir [7]. Bazı durumlarda hareket enerji imgeleri ile ayrılması kolay olmayan hareketler için her iki şablonun birlikte değerlendirilmesi gerekebilmektedir

Collinsvd. test dizilerinden anahtar çerçeveleri çıkararak döngüsel yürüyüş analizini gerçekleştirmiştir. Bu çerçeveler, eğitilmiş çerçevelerle normalleştirilmiş bağlantı yöntemi kullanılarak karşılaştırılmış ve nesne sınıflandırması bu karşılaştırma sonucunda elde edilen değerler üzerinden en yakın komşuluk yöntemiyle gerçekleştirilmiştir [8].

Hareketi gösteren imge dizileri gürültülü olabilmektedir. Aktiviteler kısmen veya tamamen örtülmüş veya arkaplan iyi bir bölütleme imkân tanımayacak kadar parazitli olabilir. Benzer durumlar için tanımda kullanılmak için olasılıksal yapılar geliştirilmiştir. Bu yapılarda, tanıma kararı en yüksek gerçekleşme olasılığına sahip aktivitenin tercih edilmesi temeline dayanmaktadır. Bu yapılar içerisinde en yaygın kullanılanı ve bilineni saklı markov modelleridir. Saklı markov modeli, eğitim tabanlı tanıma tekniği olarak ses tanımada başarı ile kullanılmaktadır. Saklı markov modeli yardımıyla hareket tanıma problemi bir örüntü tanıma problemine dönüştürülmektedir. Yamatovd. saklı markov modeli kullanılarak insan hareketinin tanınması konusundaki ilk çalışmalardan birini gerçekleştirmiştir [9]. Bregler, insan hareketlerini değişik soyutlama seviyelerinde olasılıksal parçalarına ayırmış ve olasılık dağılımının uzay, zaman ve soyutlama seviyeleri üzerindeki yayılımlarının nasıl olduğunu göstermiştir [10]. Yürüyüş stili ve insan aktivitesi tanıma için farklı saklı markov modelleri araştırmalarda kullanılmıştır [11].

Bregler'in çalışmasında, beklenti enbüyükleme algoritması ile elde edilen birbiriyle uyumlu hareket alanlarını, dinamik sistemlerle ifade etmiş ve karmaşık insan hareketleri saklı markov modeli ile gösterilmiştir. Saklı markov modelinde art arda gelen her bir faz basit hareketleri göstermektedir ve bunlar "movemes" olarak adlandırılmıştır. IvanovveBobick, tanıma problemini iki seviyeye ayırdıkları stokastik bir ayırma yöntemi geliştirmişlerdir. Birinci seviyede, saklı markov modelleri gibi bağımsız olasılıksal olay sezicileri yardımıyla mümkün olan düşük-seviye olay öznitelikleri tahmin edilerek stokastik bağlama-duyarsız ayırma mekanizmasına girdi sağlanıyordu. Bu mekanizma ile tanıma sürecine önsel bilgi ve zamansal kısıtların aktarılabilmesine olanak sağlamıştır [12]. Arkaplan çıkarma, vücut pozunu bulma ve hareket tanıma olmak üzere üç adımdan meydana gelen çalışmada yer-zaman boyutunda tanımlanmış olan insan pozları "movelet" olarak adlandırılmıştır [13]. Bir movelet, vücudun ana bölümlerine bağlı imge parçalarının, birbirini izleyen çerçevelerdeki şekli, hareketi ve örtüşmesinden elde edilmiş bir koleksiyondur. Movelet'lerin bulunması için veri kümesi içindeki kişilere her bir vücut parçası farklı renkte olacak şekilde özel kıyafetler giydirilerek farklı hareketler yaptırılmıştır. Tanıma işlemi, video dizisi içinde yaralan en uygun aktivite ve kod kelimeleri dizisinin saklı markov modeli kullanılarak tahmin edilmesiyle gerçekleştirilmektedir. Chenvd. insan pozları için tanımlayıcı olarak

yıldız iskeleti yapısını kullandıkları saklı markov modeli tabanlı bir aktivite tanıma sistemi önermişlerdir [14].

Hatun ve Duygulu'nun gerçekleştirdiği çalışmada her bir video çerçevesindeki pozlar yönlü gradyanların histogramı (histogram of oriented gradients) olarak tanımlanmaktadır. Hareketlerin sınıflandırılıp tanınması için çok büyük boyutlu poz uzayında poz kelimelerinin kümelenerek hareketleri oluşturan pozların daha az boyutlu uzaylarda temsil edilmesine yarayan kelime çantası (Bag of Words) yöntemi kullanılmaktadır. Bu yöntem pozların hareket içerisindeki dizilimleri yerine hareket içerisinde bulunup bulunmadığına göre tanıma gerçekleştirmektedir [15]. Ancak pozların dizilimi hareketi tanımlamanın önemli bir verisidir. Bu çalışmada önerilen sıralı poz dizilimi ile çalışılan SMM yöntemi, Güçlü ve Adar'ın yaptıkları çalışmada kelime çantası yöntemi ile kıyaslanmıştır. Kelime çantası yöntemi ile yapılan çalışmada %60, sıralı poz dizilimi ile yapılan SMM yönteminde %90 başarılı tanıma gerçekleştirildiği gösterilmiştir [16].

İnsan hareketleri, bir dizi pozun zaman düzleminde arka arkaya gelmesiyle gösterilebilmektedir. Konuşma dilinde seslerin heceleri, hecelerin de birleşerek kelimeleri oluşturmalarına benzer yapıda insan pozları da bir araya gelerek insanın gerçekleştirdiği hareketleri oluşturmaktadır [17].

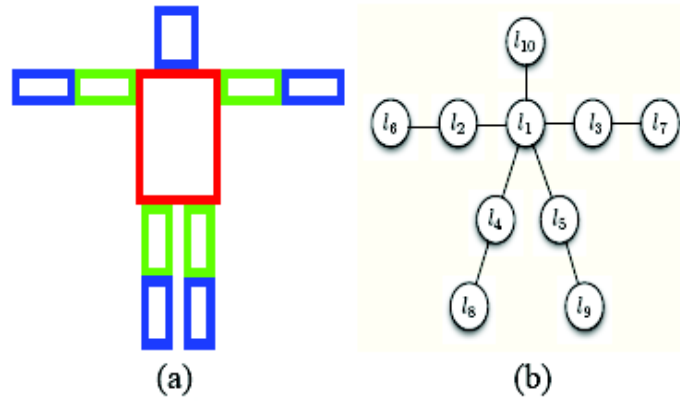
Bu çalışmada, hareketi oluşturan pozların tanımlanmasında resimsel yapı modeli kullanılmıştır. Kenar-tabanlı biçim değiştirir model yaklaşımıyla algoritma yardımıyla insan pozlarına ait öznitelik vektörleri bulunur. Tüm çerçevelere ait öznitelik vektörlerinden oluşan öznitelik matrisi K-ortalama yöntemi ile sınıflandırılarak poz kod kitabı oluşturulur. Video dizisindeki çerçevelerde elde edilen öznitelik vektörleri kod kitabındaki kendisine en çok benzeyen (en düşük mesafeye sahip) kod kelimesini ifade eden sembollere eşlenmesi sonucunda harekete ait poz sembol dizisi elde edilir. Her bir çerçevede bulunan pozlar ile elde edilmiş olan poz dizileri ile hareketi tanımak için iki yöntem önerilmiştir. Bu yöntemlerden saklı markov modelinde her bir hareket için ayrı model kullanılmakta ve bu modellerin sonucuna göre tanıma kararı üretilmektedir. Bu yaklaşımda her bir hareket farklı alt model ile tanındığından tanıma işlemi paralel ve gerçek zamanlı gerçekleştirilebilir. Çalışmada, tanıma için önerilen diğer yöntem olan Yapay sinir ağı modelinde ise poz dizileri ile öncelikle eğitim uygulanmakta, daha sonra test poz dizileri ile hareket tanıma başarımı sorgulanmaktadır. Bu yöntemde, test aşaması çok hızlı çalışmakla beraber yeni hareketlerin sisteme tanıtılması sürecinde önceki eğitim setinde yapılmış tüm işlemler baştan tekrar edilmek zorundadır.

II. HAREKETE AİT POZ SEMBOL DİZİLERİNİN BULUNMASI

Kenar-tabanlı biçim-değiştirir modelde, insan vücudu her bir parçanın sabit büyüklükte yönlü dörtgenler ile gösterildiği K parçadan oluşmaktadır. Temel olasılıksal model, ağaç-yapılı koşullu rastsal alandır. Her bir parça, imgedeki yeri ve yönü cinsinden $l_i=[x_i, y_i, \theta_i]$ olarak gösterilmektedir. Parçaların sabit büyüklükte en üst noktası (x_i, y_i) verilmiş olan yönlü dörtgenlerden oluştuğu varsayılmaktadır. Buna göre K parçalı model $L=(l_1, l_2, \dots, l_k)$ ile gösterilebilir. $K=10$ adet parçadan oluşmuş insan modeli Şekil 1 (a)'da görüldüğü gibi ifade edilebilmektedir [18].

$l_k=[x_i, y_i, u_i, v_i]$ şeklindeki ifadede (x_i, y_i) parçaların üst son noktasını gösterirken (u_i, v_i) vücut içinde yukarıdan aşağı yönü gösteren birim vektördür.

Video dizilerindeki çerçevelerde insan resimsel yapı modelini elde etmek için bir araç geliştirilmiştir. Video dizilerinde resimsel yapıyı oluşturan l_k yönlü dörtgenlerinin yerleri ilk çerçevede tespit edilmektedir. Resimsel yapıyı oluşturan parçalar, bir sonraki çerçevede daha önce bulunduğu bölgeden piksel bazında çok uzak mesafede bulunamayacağı için, hareket vektörleri elde etme yaklaşımlarında kullanılan üç adım arama metodu yardımıyla [19] bu parçalar belli yakınlıktaki piksel mesafelerinde aranarak her bir çerçevedeki insan resimsel modeli bulunmaktadır. Bu araç yardımıyla resimsel yapıyı oluşturan parçalar otomatik olarak bulunmaktadır.



Şekil 1. İnsan vücut modeli (a) insan vücudunun 10 parça ile ifade edilmesi, (b) parçaların ağaç yapısındaki yönsüz çizgesi.

II.1. Öznitelik Vektörleri

Video dizisini oluşturan çerçevelerde insan pozlarının resimsel yapı modelleri, öznitelik vektörleri olarak ifade edilmektedir. Farklı hareketlere ait video dizilerindeki tüm öznitelik vektörleri $F = f_1, f_2, f_3, \dots, f_T$ k-ortalamları kullanılarak sınıf merkezleri hesap edilir. Bu sınıf merkezlerine

$P = P_1, P_2, P_3, \dots, P_N$ poz sembolleri atanarak poz sembollerinden (kod kelimeleri) oluşmuş bir kod kitabı elde edilir.

Resimsel yapı modelinde l_k , k . vücut parçasını göstermek üzere her bir parça eşitlik (1) ile ifade edilmektedir.

$$l_k = [x_k, y_k, u_k, v_k, tip_k]; \quad k = 1..10 \quad (1)$$

Buna bağlı olarak i . çerçevedeki poza ait f_i , öznitelik vektörü de eşitlik (2)'de belirtildiği şekilde gösterilebilir.

$$f_i = [l_1 \quad l_2 \quad l_3 \quad l_4 \quad l_5 \quad l_6 \quad l_7 \quad l_8 \quad l_9 \quad l_{10}]^T \quad (2)$$

İnsan figürüne ait resimsel yapıyı oluşturan parçaların (x, y) değerleri çerçeveden çerçeveye değişebildiği için resimsel yapıda bulunan kafaya ait parçanın (x, y) değerleri merkez alınarak diğer parçaların (x, y) değerleri kafaya göre tekrar hesaplanır. l_{10} , kafayı ifade eden resimsel yapı modeli parçası olmakla birlikte her bir parça için $\Delta x, \Delta y$ değerleri eşitlik (3) ve (4) ile hesaplanır.

$$\Delta x_k = x_{10} - x_k \quad (3)$$

$$\Delta y_k = y_{10} - y_k \quad (4)$$

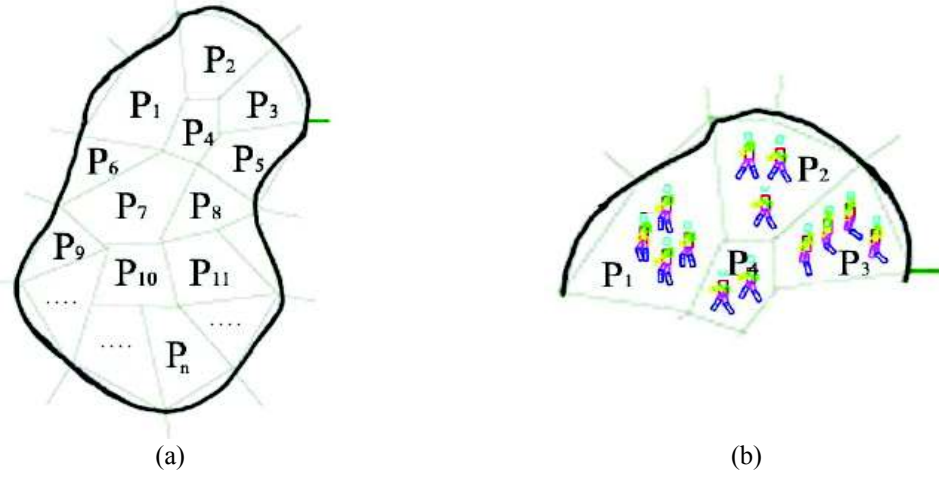
Eşitlik (3) ve (4) kullanılarak gerçekleştirilen normalizasyondan sonra resimsel yapı modeline ait vücut parçalarına ait bileşenler eşitlik (5) ile verilen vektör ile ifade edilebilir.

$$l_k = [\Delta x_k, \Delta y_k, u_k, v_k, tip_k] \quad (5)$$

II.2. Öznitelik Vektörleri İle Sembol Eşleşmesi

Hareket tanıma için günlük hayattaki beş temel harekete ait videolardan oluşturulmuş bir veritabanından yararlanılmıştır. Video veritabanındaki farklı hareketlere ait 5000 çerçevede insan pozları tespit edilerek pozlara ait öznitelik vektörleri çıkarılmıştır. Bu öznitelik vektörleri K-ortalama yöntemi kullanılarak n adet kod kelimesine sahip kod kitapları oluşturacak şekilde sınıflandırılmışlardır. Farklı kod kelimesi sayıları belirlenirken kol ve bacak gibi insan vücudunun eklemli bölgelerinin 180° lik bir yay boyunca 6° lik, $4,5^\circ$ lik ve yaklaşık olarak $3,6^\circ$ lik bir açıyla çerçeveden çerçeveye değiştiği varsayımına göre K-ortalama sınıflandırma yöntemi kullanılarak 30, 40 ve 50 kod kelimesine sahip kod kitapları elde edilmiştir. Video çerçevelerinde bulunan öz nitelik vektörleri, K-ortalama ile oluşturulan kod kitabında

yer alan kod kelimelerinden en çok benzeşenine vektör niceme tekniği kullanılarak atanarak, karşılık gelen poz sembolleri ile eşleştirilmektedir.



Şekil 2. $g_j \in R^n$ merkezli n adet kod kelimesine sahip kod kitabı.

Vektör niceme için, R^n öznitelik uzayında sınıfların merkezini temsil eden $g_j \in R^n$ kod kelimeleri Şekil 2' (a)'da görülmektedir. Şekil 2 (b)'de de görüleceği gibi bir video dizisini oluşturan çerçevelere ait farklı öznitelik vektörleri, merkeze olan uzaklıklarının birbirlerine yakın olmasından dolayı kod kitabında aynı poz sembolü ile eşleşebilirler.

Video dizilerini oluşturan çerçevelerden elde edilen her bir f_i öznitelik vektörü, öznitelik uzayında bu vektöre en yakın kod kelimesine atanmış sembole dönüştürülür. Bunun anlamı, f_i öznitelik vektörü eğer $j = \arg \min_j d(f_i, g_j)$ ise P_j poz sembolüne dönüştürülmesidir. $d(f_i, g_j)$ öznitelik vektörü ile sınıf merkezinde bulunan kod kelimesi arasındaki mesafedir ve eşitlik (6) ile hesaplanmaktadır.

$$d(f_i^k, g_j^k) = \arg \min_{1 \leq j \leq J} \sum_{k=1}^{10} (f_i^k - g_j^k) \quad (6)$$

Eşitlikte g_j^k kod kelimesi vektörü ve f_i öznitelik vektörleri, insan vücudunun resimsel yapısını oluşturan k adet ($k=10$) alt vektörden oluşmaktadır. J : k -ortalama ile elde edilen sınıf sayısını gösterir.

Vektör nicemleme sonrası her bir harekete ait video dizisi hareket gözlem vektörüne dönüştürülmektedir. Eşitlik (7)'deki $O_{hareket}$ vektörü bir kişiye ait gerçekleştirdiği hareket için elde dilmiş olan hareket gözlem vektörüdür.

$$O_{hareket} = P_1 P_1 P_1 P_1 P_2 P_2 P_2 P_3 P_3 P_3 P_3 P_4 P_4 P_4 P_5 P_5 P_6 P_6 P_6 P_5 P_5 P_5 P_3 P_3 P_3 P_3 P_2 P_2 P_2 \quad (7)$$

Elde edilen bu vektörün gerçekleştirdiği hangi tip harekete karşılık geldiğinin (yürüme, tempolu yavaş koşma TYK, koşma, alkışlama, el sallama) bulunması ise bir sonraki poz dizilerinden anlam çıkarma bölümünde ele alınacaktır.

III. POZ DİZİLERİNDEN ANLAM ÇIKARMA

Önerdiğimiz tanıma modelinde, hareketlerin tamamında elde edilen öznelik vektörlerinin birlikte sınıflandırılması ile tek bir hareket kod kitabı oluşturularak farklı hareket sınıflarında aynı cins öznelik vektörlerinin kod kelimesi olarak bulunması engellenmektedir. Bu nedenle testlerde tanıma amaçlı olarak birbirinden farklı kod kelimeleri içeren hareketlere ait video dizileri kullanılmıştır.

Hareketlere ait gözlem vektörleri boyutlarını normalize etmek amacıyla en büyük boyuta sahip olan gözlem vektörünün boyutuna eşitlenmektedir. Bu amaçla, gözlem vektörlerine ortak poz ve tekrarlı poz gözlem vektörü ekleme üzere iki farklı yöntem kullanılmıştır.

Ortak poz gözlem vektörü ekleme yönteminde P_0 poz sembolü gözlem vektörlerinin sonuna eklenmektedir. Tekrarlı poz gözlem vektörü ekleme yönteminde ise hareketin devam ettiği varsayımından yola çıkılarak harekete ait gözlem vektöründeki ilk elemandan itibaren semboller tekrar girilmektedir.

Her bir hareket için normalize edilmiş poz dizi vektörleri, eğitim ve test hareket gözlem vektörleri olarak gruplandırılmışlardır. Eğitim hareket gözlem vektörleri önerilen SMM ve YSA'na uygulanarak her bir hareket için modeller eğitilerek, modeller için hareket tanımda gerekli olan parametre ve ağırlıklar hesaplanmıştır. Eğitimi tamamlanarak oluşturulmuş olan modele test hareket gözlem vektörleri uygulanarak eğitimlerin başarımı ölçülmüştür.

III.1. SMM Hareket Tanıma

Gerçekleştirilen saklı markov modeli tabanlı hareket tanıma sisteminde yürüme, koşma, TYK, el sallama ve alkışlama için ayrı birer saklı markov modeli eğitilmekte ve hareket tanınacak video dizisindeki insan pozları hareket gözlem vektörüne dönüştürülerek eğitilmiş olan saklı markov modellerine uygulanmaktadır.

N durum ($Q = \{q_1, q_2, \dots, q_N\}$) ve M çıkış poz sembollerine ($O = \{x_1, x_2, \dots, x_M\}$) sahip bir SMM, üç parametre ($\lambda_k = (A, B, \pi)$) ile tamamen ifade edilebilmektedir. Önerilen SMM modelinde durum sayısı $N = 5$ ve M gözlem sembol sayısı ise sırasıyla 30, 40 ve 50 alınmaktadır.

$$A = \{a_{mn} \mid a_{mn} = P_k(s_{t+1} = q_n \mid s_t = q_m)\} \quad (8)$$

Eşitlik (8) a_{mn} , q_m durumundan q_n durumuna geçiş olasılığını göstermektedir. $b_n(m)$, x_m sembolünün q_n durumundaki olasılığını t de video dizisindeki çerçeve anını göstermek B matrisi aşağıdaki eşitlik (9)'da verilmektedir.

$$B = \{b_n(m) \mid b_n(m) = P_k(x_m \mid s_t = q_n)\} \quad (9)$$

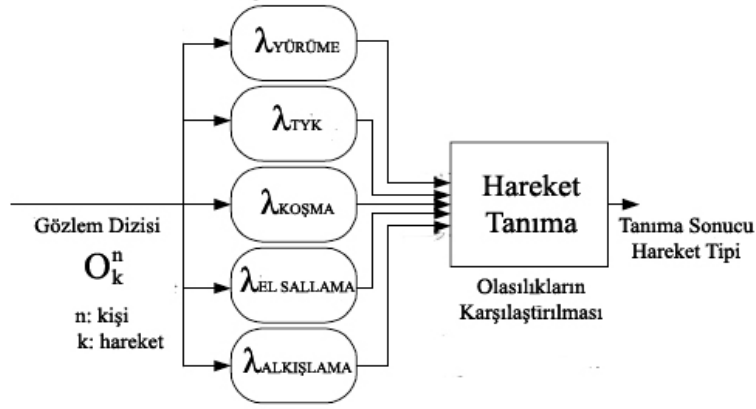
π ise ilk kullanım için başlangıç durum dağılımı vektörüdür ve $\pi = \{\pi_m \mid \pi_m = P_k(s_1 = q_m)\}$ ifadesiyle tanımlanmaktadır.

SMM tabanlı tanıma işlemi $P_k(\lambda \mid O)$ değerinin hesaplandığı eğitim işlemi ve eğitim sonucu elde edilen λ_k modelinin ürettiği O gözlem dizisine ait olasılığın hesap edildiği tanıma işleminden meydana gelmektedir.

Tanınması istenen K adet hareket için elde edilmiş hareket gözlem vektörlerinin kullanılmasıyla (A, B, π) model parametrelerinin bulunması ileri-geri arama ve Baum-Welch Tahmin algoritmaları ile gerçekleştirilir.

O gözlem dizisinin hangi hareket olduğunun tanınması ve hareketi en iyi temsil eden kod kelimesi dizisinin ne olduğunun bulunması için izleyen SMM çözümü uygulanmıştır. Her bir hareket $k \in (1 \dots K)$ için en uygun durum (kod kelimesi) dizisini aramak amacıyla eşitlik (10)'daki olasılık eşitliğini en büyükleyecek viterbi algoritması uygulanarak sonuç elde edilmektedir.

$$q_{k1}^*, q_{k2}^*, \dots, q_{kT}^* = \max_{q_{k1}, q_{k2}, \dots, q_{kT}} P(q_{k1}, q_{k2}, \dots, q_{kT} \mid O_k, A_k), q_{kt} \in (P_1, \dots, P_i) \quad (10)$$



Şekil 3. Hareket tanıma için kullanılan SMM yapısı.

Her bir hareket için yukarıda anlatıldığı şekilde eğitilen SMM modeli Şekil 3'te görüldüğü gibi paralel kullanılarak, her bir SMM modelden gelen tanıma olasılıkları hesaplanmakta ve hesaplanan bu değer eşitlik (11)'de verilen karar ölçütü karşılaştırılarak hareketin tanınması işlemi gerçekleştirilmektedir.

$$\arg \max \text{Log}(P(O | \lambda_i)) \quad (11)$$

SMM ile hareket tanıma sisteminin ilk aşaması olan eğitim aşamasında, her bir hareketi gerçekleştiren 125 kişi-hareketten 75 kişi-hareket (yaklaşık 3000 çerçeve) ile eğitim yapılarak hareketlere ait (A, B, π) model parametreleri hesap edilmektedir. Model parametreleri hesap edilirken saklı katmanlardaki durum sayısı beş, gözlem sembolleri ise değişik sayıda kod kelimesi içeren kod kitaplarına bağlı olarak sırasıyla 30, 40 ve 50 olarak alınmaktadır. SMM ile hareket tanıma sisteminin ikinci aşaması olan test aşamasında ise her bir hareketi gerçekleştiren 125 kişi-hareketten eğitimde kullanılmayan 50 kişi-hareket (yaklaşık 2000 çerçeve) kullanılarak sistem başarımı incelenmektedir.

III.2. İleri Beslemeli Yapay Sinir Ağı Modeli İle Hareket Tanıma

Elde ettiğimiz poz dizilerinden, bu dizilerin hangi harekete ait olduğunun bulunmasında kullandığımız diğer yöntem de yapay sinir ağlarıdır. Bu seçimi yapmamızın nedenleri arasında, eğitim aşamasının uzun sürmesine rağmen test aşamasının oldukça hızlı gerçekleşmesi ve hata yüzeyi üzerinde minimum hatayı bulmaya çalışırken bölgesel minimumlara takılma oranının çok düşük olması sayılabilir.

Poz dizilerinden hareket tanıma için kullanılan ileri besleme YSA modeli bir adet gizli katmandan oluşmaktadır. Hareket veritabanında bulunan beş temel harekete ait video dizilerinde hareketlerin

başlangıç ve bitişleri arasında en fazla 38 çerçeve (tempolu yavaş koşu) bulunmaktadır. Her çerçevede bulunan insan resimsel yapı modeline ait öznitelik vektörü boyutu 50 olmaktadır. Bu yüzden giriş ve gizli katmandaki yapay nöron sayısı $38 \times 50 = 1900$ adet olarak alınmaktadır. Transfer fonksiyonu olarak Tan-sigmoit transfer fonksiyonu kullanılmıştır. Veri setinin büyüklüğü ve yavaş bir öğrenme hızı olmasına rağmen daha az bellek isteyen ölçekli-eşlenik gradyan yöntemi (Scaled Conjugate Gradient) ile eğitim gerçekleştirilmiştir [20]. Çıkış katmanı ise 5 adet nörondan oluşmakta olup her biri eğitim ve testlerde kullanılacak olan hareketlere karşılık gelmektedir. Eğitim esnasında eğitimi yapılan harekete ait çıkış vektör satırı 1 diğer satırlar 0 olarak eğitilmiştir. Test aşamasında uygulanan poz dizisine bağlı olarak elde edilen çıkış vektörü satırında 1'e en yakın değeri veren satır, poz dizisinin hangi harekete ait olduğunu göstermektedir.

IV. TEST SONUÇLARI

Yapılan test sonuçları Çizelge 1 – 4 'te verilmiştir. Çizelge 1-3 'te farklı kod kelime sayısı için; 5 hareketin Ortak poz (O), Tekrarlı poz (T), SMM ve YSA başarımları yüzdeleri verilmiştir. Örneğin Çizelge 1 satır 1 de test edilen yürüme poz dizimleri hem Ortak pozlu SMM yöntemi ile hem de Tekrarlı poz SMM yöntemi ile %80 doğrulukla yürüme olarak tanınmıştır. Test için kullanılan yürüme poz dizimleri diğer hareketlerden sadece TYK için Ortak poz SMM ve Tekrarlı poz SMM yöntemlerinde %20 oranında hatalı olarak tanınmıştır. “-” işareti ile verilen yüzdeler ise testlerde ilgili seçenekte tanıma gerçekleşmediğini gösterir.

Çizelgelerden de görüldüğü gibi, normalizasyon yöntemi olarak ortak veya tekrarlı poz ekleme yöntemlerinden hangisi kullanılırsa kullanılsın, YSA modelleri yüzde olarak en iyi tanıma oranlarını vermişlerdir. Eğitim ve testlerin yapıldığı model bazında sonuçlar incelendiğinde ise ortak poz kullanılan modellerdeki hareket tanıma işlemlerinin başarımları, tekrarlı poz kullanılan modellerdekine göre daha başarılı olduğu görülmektedir. Kod kelimeleri kullanma yönteminde (yeterli sayıda poz kelimesi oluşturulması halinde, tüm hareketlerdeki tüm pozlar sınırlı sayıda olacaktır) yeni hareketlerin her iki yöntemle göre tanıma sistemine eklenmesinde sadece ilave eğitim süreçlerinin çalışılması yeterli olacaktır. Yeni hareketi YSA tanıma yöntemine eklemek için tüm eğitim setini tekrar uygulamak gerekecektir. SMM tanıma yönteminde ise diğer eğitilmiş hareketlerde herhangi yenileme yapmaksızın sadece yeni hareketin SMM parametrelerini oluşturarak sisteme eklemek yeterli olacaktır. Bu özelliği sayesinde genel kod kelimeleri tanımlanmış SMM yönteminin gömülü sistem olarak, gerçek zamanlı ve genişletilebilir yapısı ile tanıma aracı olarak geliştirilmesinin daha uygun olacağı değerlendirilmektedir.

Çizelge 1. Kod kelime sayısı n=30 için model başarımları.

		Yürüme (%)		Koşma (%)		TYK (%)		El Sallama (%)		Alkış (%)	
		O	T	O	T	O	T	O	T	O	T
Yürüme	SMM	80	80	-	-	20	20	-	-	-	-
	YSA	90	90	10	10	-	-	-	-	-	-
Koşma	SMM	10	-	80	70	10	30	-	-	-	-
	YSA	-	-	100	90	-	10	-	-	-	-
TYK	SMM	20	10	40	50	40	40	-	-	-	-
	YSA	-	-	-	10	100	90	-	-	-	-
El Sallama	SMM	-	-	-	-	-	-	100	100	-	-
	YSA	-	-	-	-	-	-	90	80	10	20
Alkış	SMM	-	-	-	-	-	-	-	10	100	90
	YSA	-	-	-	-	-	-	-	-	100	100

Çizelge 2. Kod kelime sayısı n=40 için model başarımları.

		Yürüme (%)		Koşma (%)		TYK (%)		El Sallama (%)		Alkış (%)	
		O	T	O	T	O	T	O	T	O	T
Yürüme	SMM	90	90	-	-	10	10	-	-	-	-
	YSA	90	90	10	10	-	-	-	-	-	-
Koşma	SMM	10	10	80	80	10	10	-	-	-	-
	YSA	-	-	100	90	-	10	-	-	-	-
TYK	SMM	-	-	30	40	70	60	-	-	-	-
	YSA	-	-	-	20	100	80	-	-	-	-
El Sallama	SMM	-	-	-	-	-	-	100	100	-	-
	YSA	-	-	-	-	-	-	90	90	10	10
Alkış	SMM	-	-	-	-	-	-	-	10	100	90
	YSA	-	-	-	-	-	-	-	-	100	100

Çizelge 3. Kod kelime sayısı n=50 için model başarımları.

		Yürüme (%)		Koşma (%)		TYK (%)		El Sallama (%)		Alkış (%)	
		O	T	O	T	O	T	O	T	O	T
Yürüme	SMM	90	90	-	-	10	10	-	-	-	-
	YSA	100	90	-	10	-	-	-	-	-	-
Koşma	SMM	-	-	100	90	-	10	-	-	-	-
	YSA	-	-	100	80	-	20	-	-	-	-
TYK	SMM	-	-	20	20	80	80	-	-	-	-
	YSA	-	-	-	30	100	70	-	-	-	-
El Sallama	SMM	-	-	-	-	-	-	100	100	-	-
	YSA	-	-	-	-	-	-	90	90	10	10
Alkış	SMM	-	-	-	-	-	-	-	-	100	100
	YSA	-	-	-	-	-	-	-	-	100	100

Birbirine uzak olan hareketlerde tekrarlı poz kullanımı ortak poz kullanımına göre bir olumsuzluk yaratmazken; birbirine yakın hareketlerde özellikle tekrar eklenen pozlar içinde yakın olan hareketteki pozlarla benzer olan pozların çokluğu tanımayı olumsuz yönde etkilemektedir. Kod kitabının büyüklüğü ise tanıma oranlarıyla doğru orantılıdır ve kod kelime sayısı arttıkça poz dizilerindeki benzer poz sayısı azaldığından hatalı hareket tanıma oranları, doğru tanıma yönünde iyileşmeye sebep olmaktadır.

Çizelgelerde izlendiği gibi; test amaçlı kullanılan video dizilerindeki 5 temel hareketin tanıma oranlarına göre kendi içinde de gruplandığı görülmüştür. Birbirine yakın olan, yürüme, koşma ve TKY bir grupken, el sallama ve alkışlama diğer bir grubu oluşturmaktadır. Bu yüzden de hatalı tespit edilen hareket dizileri sadece ilgili buldukları grup içinde hatalı tanınmaktadır. Örneğin videoda gerçekleşen hareket yürüme ise bu hareket eğer doğru tanınmadı ise koşma veya TKY hareketlerinden biri olarak hatalı tanınmıştır. Benzer durum el sallama ve alkışlama hareketleri için de geçerlidir.

Çalışmada önerilen SMM yönteminin bir diğer uygulama avantajı da modele uygulanan hareket vektörlerinin YSA modellerde olduğu gibi aynı uzunlukta olmaları gerekmektedir. Bu yüzden test hareket videolarından elde edilen poz dizileri herhangi bir normalizasyona gerek kalmadan da SMM modellere uygulanabilmektedir. Çizelge 4'te test poz dizilerinin normalizasyon yapılmadan SMM modele uygulanmasının sonuçları verilmiştir. Bu testlerde hareketlerin tanınması için kullanılan SMM modellerinden en düşük tanıma oranları, normalize edilmemiş poz dizileri ile eğitimi ve testleri yapılmış olan SMM modeli ile elde edilmektedir. El sallama ve alkışlama hareketlerinde ortak ve tekrarlı poz eklenerek eğitim ve testlerin yapıldığı SMM modelleri ile yaklaşık aynı doğru tanıma sayılarına sahipken, yürüme, koşma ve TYK hareketlerinde tanıma oranları ortak ve tekrarlı poz eklenerek eğitim ve testlerin yapıldığı SMM modellerine göre daha kötüdür.

Kod kelime sayısının artması, ortak poz ve tekrarlı poz eklenerek eğitim ve testlerin yapıldığı SMM modellerinde tanıma oranlarını iyileştirici yönde katkı sağlarken, normalizasyon yapılmadan poz dizilerinin uygulanarak eğitim ve testlerinin yapıldığı SMM modelinde tanıma oranları arasındaki fark, özellikle birbirine yakın hareketler arasında azaldığı için hareketlerin hatalı algılanması yönünde olumsuz katkıda bulunmaktadır.

Çizelge 4. Poz dizilerinin normalizasyon yapılmadan SMM modele uygulanması.

Kod Kitabı Büyüklüğü	Yürüme (%)			Koşma (%)			TYK (%)			El Sallama (%)			Alkışlama (%)		
	30	40	50	30	40	50	30	40	50	30	40	50	30	40	50
Yürüme	100	100	80	-	-	-	-	-	20	-	-	-	-	-	-
Koşma	-	-	-	90	80	70	10	20	30	-	-	-	-	-	-
TYK	10	-	-	40	40	40	50	60	60	-	-	-	-	-	-
El Sallama	-	-	-	-	-	-	-	-	-	100	100	100	-	-	-
Alkışlama	-	-	-	-	-	-	-	-	-	-	-	-	100	100	100

V. SONUÇ VE ÖNERİLER

Önerilen YSA ve SMM modellerinde öznitelik vektörü olarak insan resimsel yapı ifadesindeki vücut parçalarının birbirlerine olan uzaklıkları ve vücut parçalarının yönleri kullanılmaktadır. Her bir hareket için ayrı kod kitabı oluşturularak bu kod kitaplarının bir arada kullanılmasıyla model eğitimi ve testlerinin yapıldığı çalışmalarda yeni bir hareketin model tarafından tanınabilmesi için o hareketin kod kelimeleri elde edilerek model kod kitabına eklenmesi gerekmektedir. Kod kelimelerinin hareket temelli üretilmesi farklı hareketlere ait kod kitaplarında aynı kod kelimelerinin bulunmasına sebep olabilmektedir. Önerdiğimiz tanıma modelinde, hareketlerin tamamına ait pozlar sınıflandırılmak suretiyle kod kitabı oluşturulduğundan, yeni bir hareket modele uygulandığında mevcut kod kitabına bağlı olarak yeni bir poz dizisi elde edilip, modelin eğitimi bu yeni poz dizisiyle yapılarak yeni hareketlerin de tanınmasına olanak sağlanmaktadır. Yeni hareketlere ait pozlar elde edildiğinde eğer pozlara ait öznitelik vektörleri kod kitabında bulunan kod kelimelerinden tolerans dışında bir mesafede ise yeni kod kelimesi model kod kitabına eklenmektedir. Böylelikle kod kitabının dinamik bir şekilde güncellenmesi sağlanmaktadır.

Hareketlerin video içerisindeki gerçekleşme süreleri farklılık göstermektedir. Bu yüzden bu hareketlere ait poz dizi vektörlerinin tüm hareketlerde aynı uzunlukta olabilmesi için girdi vektör boyutlarının normalize edilmesi gerekmektedir. Çalışmalarımızda ortak poz ekleme ve tekrarlı poz ekleme olmak üzere iki farklı yöntem önerilmiş ve kullanılmıştır.

YSA eğitimlerinde ortak poz ekleme yöntemiyle eğitim verileri oluşturulmuş olan model, tekrarlı poz ekleme yöntemiyle eğitim verileri oluşturulmuş olan modele göre daha hızlı bir şekilde eğitilebilmektedir.

Eğitim ve testlerin yapıldığı model bazında sonuçlar incelendiğinde görüldüğü gibi, normalizasyon yöntemi olarak ortak veya tekrarlı poz ekleme yöntemlerinden hangisi kullanılırsa kullanılsın, YSA modelleri yüzde olarak en iyi tanıma oranlarını vermişlerdir. Ayrıca, ortak poz kullanılan modellerdeki hareket tanıma işlemlerinin başarımları, tekrarlı poz kullanılan modellerdekine göre daha başarılı olduğu görülmektedir. Her iki yöntem kullanılarak yapılan tanıma işlemlerinde hatalı olarak tanınan hareketlerin birbirine yakın hareket grubu içinde hatalı tanındığını göstermiştir.

Önerilen SMM tanıma yöntemi gerçek zamanlı tanıma yapabilen, kolayca genişletilebilir gömülü sistem olarak gerçekleştirmek mümkündür. Yeni bir hareket tanıtmak için önceki eğitilmiş hareketlerde değişiklik yapmaya gerek kalmadan sadece yeni hareket için yeni bir SMM eklenebilmesi kolayca genişletilebilir olmasını sağlamaktadır. Gerçek zamanlı hareket tanıma için ise her bir hareket farklı bir gömülü sistem ile gerçekleyerek tanıma kararı tüm hareketler için paralel olarak üretilmiş sonuçlara göre verilebilir. Bu avantajları nedeniyle gerçek zamanlı tanıma yapabilen genişleyebilir, paralel sistem mimarisinde gömülü mimarili SMM tanıma makinesi geliştirilmesi gelecek çalışma olarak planlanmaktadır.

VI. KAYNAKLAR

- [1] M. H. Yangand, N. Ahuja, "Recognizing hand gesture using motion trajectories", Proceedings of the IEEE Conference on Computer Vision and Image Understanding, Fort Collins, Colorado, June 1999, pp. 468-472.
- [2] C. Myers, L. Rabinierand A. Rosenberg, 'Performance tradeoffs in dynamic time warping algorithms for isolated word recognition', *IEEE Trans. on Audio Speech Lang. Process.*, pp.623-635, 1980.
- [3] C. Rao, A. Yilmazand M. Shah, "View-invariant representation and recognition of actions", *Int. J. of Computer Vision*, Vol. 50, No.2, pp. 203-226, 2002.
- [4] N. Cuntoor, A. Kale and R. Chellappa, 'Combining multiple evidences for gait recognition', Proceedings of International Conference on Acoustics, Speech and Signal Processing, 2003, Vol. 3, pp. 113-116.
- [5] D. Cunado, M. S. Nixonand J. N. Carter, 'Using gait as a biometric, via phase-weighted magnitude spectra,' Proceedings of 1st International Conference on Audio and Video Based Biometric Person Authentication, 1997, pp. 95-102.
- [6] H. Fujiyoshi, A. J. Liptonand T. Kanade, "Real time human motion analysis by image skeletonization", *IEICE Trans Inf Sys*, pp. 113-120, Vol. E87-D, No.1, 2004.

- [7] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates", *IEEE Trans. on Pattern Anal. Mach. Intell.*, Vol. 23 (3), pp. 257-267, 2001.
- [8] R. T. Collins, R. Gross and J. Shi, 'Silhouette based human identification from body shape and gait', 5th International Conference on Automatic Face and Gesture Recognition, May 2002.
- [9] J. Yamato, J. Ohya and K. Ishii, "Recognizing human action in time sequential images using hidden markov models", *Proceedings Computer Vision and Pattern Recognition*, 1992, pp. 379-385.
- [10] C. Bregler, "Learning and Recognizing Human Dynamics in Video Sequences", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 1997, pp. 568-574.
- [11] A. Kale, A.N. Rajagopalan, N. Cuntoor and V. Krueger, 'Gait based Recognition of Humans Using Continuous HMMs', *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, May 2002, pp. 336-341.
- [12] Y. A. Ivanov and A. F. Bobick, "Recognition of visual activities and interactions by stochastic parsing", *IEEE Trans. on Pattern Anal. Mach. Intell.*, Vol.22, No.8, pp.852-872, 2000.
- [13] X. Feng and P. Perona, "Human action recognition by sequence of motion code words", *Proceedings of 1th International Symposium on 3D Data Processing Visualization and Transmission*, 2002, pp. 717-721.
- [14] H. S. Chen, H. T. Chen, Y. W. Chen, and S. Y. Lee, "Human action recognition using star skeleton", *Proc. of 4th ACM International Workshop on Video Surveillance and Sensor Networks 2006 (VSSN-2006)*, in conjunction with ACM Multimedia 2006, Santa Barbara, CA, USA, October 27, 2006.
- [15] K. Hatun, P. Duygulu, "A new representation for action recognition using sequence of pose words", 19th International Conference on Pattern Recognition, 2008.
- [16] Güçlü T., Adar N., "Recognizing human actions by using visual pose words", *Int. Symposium on Innovations in Intelligent Systems and Applications (INISTA 2010)*, Haziran 2010 Kayseri.
- [17] Kandemir C.M., "Yüksek Başarılı, Bilgisayarla Görü Uygulamaları Programlaması", *Doktora Tezi, Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü*, 125s., 2009.
- [18] D. Ramanan, C. Sminchisescu, "Training deformable models for localization", *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, Vol. 1, 206-213.
- [19] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion compensated interframe coding for video conferencing," *Proc. Nat. Telecommun. Conf.*, New Orleans, LA, Nov. 29-Dec. 3 1981, pp. G5.3.1-G5.3.5.
- [20] H. Demuth, M. Beale, *Neural Network Toolbox for use with MATLAB*, 2002.