Hareket Geçmişi Görüntüsü Yöntemi ile Türkçe İşaret Dilini Tanıma Uygulaması
Turkish Sign Language Recognition Application Using Motion History Image

Özge Yalçınkaya[1], Anıl Atvar[2], Pınar Duygulu[3]
[1]Bilgisayar Mühendisliği Bölümü, İhsan Doğramacı Bilkent Üniversitesi, Ankara, Türkiye
ozge.yalcinkaya@cs.bilkent.edu.tr
[2]HAVELSAN A.Ş., Ankara, Türkiye
aatvar@havelsan.com.tr
[3]Bilgisayar Mühendisliği Bölümü, Hacettepe Üniversitesi, Ankara, Türkiye
pinar@cs.hacettepe.edu.tr

Özetçe **—İşitme ve konuşma engelli bireylerin toplum içerisinde diğer bireylerle sağlıklı şekilde iletişim kurabilmeleri açısından işaret dili çok önemli bir role sahiptir. Ne yazık ki işaret dilinin toplumda sadece duyarlı insanlar tarafından bilindiği ve bu sayının da azlığı dikkat çekmektedir. Yaptığımız çalışma kapsamındaki amaç, geliştirdiğimiz sistem sayesinde işitme veya konuşma engeli mevcut olan bireylerin diğer bireylerle olan iletişiminde iyileşme sağlamaktır. Bu amaç doğrultusunda kameradan alınan işaret diline ait hareket bilgisi tanınabilmekte ve o hareketin ne anlama geldiği daha önceden eğitilen işaret diline ait hareket bilgileri ile karşılaştırılarak bulunabilmektedir.**

Hareket bilgilerinin kameradan alınan görüntülerden çıkarılması aşamasında "Hareket Geçmişi Görüntüsü" yöntemi kullanılmıştır. Bu bağlamdaki sınıflandırma işlemi için de "En Yakın Komşuluk" algoritması kullanılmıştır. Sonuç olarak geliştirilen sistem, eğitim kümesini kullanarak işaret dili hareketi için bir metin tahmin etmektedir. Toplamdaki sınıflandırma başarısı %95 olarak hesaplanmıştır.

Anahtar Kelimeler—işaret dili, hareket tanıma, Hareket Geçmişi Görüntüsü, en yakın komşuluk.

Abstract**—Recognizing sign language is an important interest area since there are many speech and hearing impaired people in the world. They need to be understood by other people and understand them as well. Unfortunately, the number of people who have the knowledge of sign language is not many. In order to communicate with handicapped people, existence of some automatized systems may be helpful. Therefore, in this work, we aimed to implement a system that recognizes the sign language and converts it to text to help people while communicating with each other where the input scene is taken from camera.**

We produced a training data which includes eight different sign language videos. After that, we used "Motion History Images"(MHI) to extract the motion information from them. A classification is done by using nearest neighbor approach after extracting the features from MHI of videos. As a result, by using training data, our system predicts the text for given sign language. The overall classification accuracy is computed as 95%.

Keywords—sign Language, motion recognition, Motion History Image(MHI), nearest neighbor.

978-1-5090-1679-2/16/$31.00 ©2016 IEEE

I. GİRİŞ

İnsanların harekete bağlı gerçekleştirdiği davranışların, takip ve tanıma işlemlerinin bilgisayarlı görü teknikleriyle gerçekleştirilmesi, son dönemlerde hem akademik hem de endüstriyel alanda çok önemli bir konu haline gelmiştir [1, 2, 3].

İnsan ile bilgisayar arasındaki iletişimi daha sağlıklı ve etkileşimli hale getirme misyonu, bu alandaki çalışmaların daha çok popülerlik kazanmasına olanak sağlamıştır. Geliştirilen sistemler genel olarak gövde, kafa, yüz ifadeleri ve el hareketlerini tanımlama üzerine yoğunlaşmıştır. İnsan ile bilgisayar arasındaki iletişim için özellikle el/kol/baş hareketlerinin tanımlanabiliyor olması çok önemlidir [4].

İnsanların gerçekleştirdiği hareket ve harekete bağlı davranışların bilgisayarlar tarafından tanınabiliyor olması, insanlık adına faydalı, birçok olumlu yeniliği de beraberinde getirmektedir. İşitme veya konuşma engeli olan bireylerin kullandıkları işaret diline ait hareketlerin bilgisayarlar tarafından tanınıp metin formatına çevrilmesi, insanlar arasındaki iletişimi iyileştirmekte ve kolaylaştırmaktadır.

Temel olarak işaret dili el pozisyonu ve el/kol/baş hareketlerini içeren işitme veya konuşma engeli olan bireylerin kullandığı görsel bir dildir. İşaret dilinde bulunan el pozisyonu ve el/kol/baş hareket davranışları, bir harf, bir kelime veya kelime grubuna denk gelmektedir. Bu nedenle yapılan hareketlerin doğru bir biçimde algılanması işaret dilindeki karşılığının bulunması ve bunun metin olarak ifade edilmesi açısından önem arz etmektedir.

İşaret dilini tanımlamaya yönelik geliştirilen sistemlerin başlangıcı 1990'lı yıllara kadar gitmektedir. Bu sistemlerde tanıma problemini çözmek için iki temel yaklaşım uygulanmıştır. Bunlardan ilki cihaz tabanlı geliştirilen bir sistem olup, hareket takibini yapacak aygıtlar aracılığıyla hareket tanımı gerçekleştiren sistemlerdir. Diğer uygulanan yaklaşım ise bir kayıt cihazından elde edilen görüntüler üzerinden hareket tanıma işleminin gerçekleştirilmesi yöntemidir.

Yaptığımız bu çalışmada kameradan alınan dinamik el/kol/baş hareketlerini hareket tanıma algoritmaları kullanarak hareketin karşılık geldiği metne çevirme işlemi gerçekleşmektedir. Hareket tanımaya yönelik literatürde bulunan algoritmalardan Hareket Geçmişi Görüntüsü (Motion History Image) [5] algoritması kullanılmıştır.

Hareket Geçmişi Görüntüsü yönteminde videoda bulunan her bir çerçevenin, algoritmaya verilen formülizasyon ile hareket geçmişini temsil eden bir görüntü elde edilmesi amaçlanmaktadır. Bu sayede videodaki hareketi temsil eden bir resim elde edilmiş olunur. Ardından da bu görüntüden özellik çıkarma işlemi gerçekleştirilir. Bizim çalışmamızda da bu yöntem kullanılarak, yine bizim oluşturduğumuz eğitim kümesinden ilgili hareketin en iyi temsiliyetinin bulunması sağlanmıştır. Daha sonra yeni gelen test videolarındaki hareketlerin karşılığını bulmak amacıyla en yakın komuşuluk sınıflandırma yöntemi tercih edilmiştir. Yönteme ilişkin detay "III. YÖNTEM" başlığı içerisinde açıklanmıştır.

Geliştirdiğimiz bu uygulama sayesinde işitme veya konuşma engeli olan bireylerin iletişim kalitesinin artırılması hedeflenmektedir. Örnek olarak bir iş görüşmesi esnasında işitme veya konuşma engeli olan bireyin işveren ile herhangi bir çevirmen bireye ihtiyaç duymadan iletişim kurabilmesi, bu geliştirilen sistem aracılığı ile imkanlı hale gelebilmektedir. Bizim uygulamamız işaretleri anlık olarak kameradan algılayıp metne dönüştürüp karşı tarafa görsel olarak gösterebilmektedir.

II. BAĞLANTILI ÇALIŞMALAR

İşaret dilini tanımaya yönelik farklı ülkelerden birçok çalışma bulunmaktadır. Bunlardan bir tanesi Starner ve çalışma arkadaşlarının [6] yaptığı çalışmadır. Bu çalışmada Amerika'ya özgü işaret dilini anlamak için iki tane gerçek zamanlı saklı Markov model tabanlı sistem sunulmuştur.

İkinci olarak Mekala ve çalışma arkadaşlarının [7] yaptığı çalışma da ise işaret diline ait harflerin hızlı ve gerçek zamanlı tanınması gerçekleştirilmiştir. Bu çalışma kapsamında yapay sinir ağları kullanılarak el işaretlerinin takibinin yapılması ve bu işaretlerin tanınarak karşılık geldiği metine veya sese dönüştürülmesi sağlanmıştır.

Literatürde Türkçe'ye özgü işaret diline ait de birçok başarılı çalışma bulunmaktadır. Bunlardan bir tanesi Haberdar ve çalışma arkadaşlarının [8] Saklı Markov Model(Hidden Markov Models) yöntemini kullanarak geliştirdikleri, Türkçe'ye özgü işaret dilindeki hareketleri evrensel özellikler üzerinden tanıyan sistemdir. Bu sistemde kameralar yardımıyla el hareketlerinin takibi yapılabilmekte ve kameradan elde edilen çerçevelerde ten-tonu belirleme algoritması kullanılarak yüz ve el tespit edilebilmektedir. Bu aşamadan sonra ise Saklı Markov Model kullanılarak el/kol/baş hareketlerinin tanınması gerçekleştirilebilmektedir.

Diğer bir çalışmada Ari ve çalışma arkadaşlarının [9] çok çözünürlüklü aktif şekil model takipçisi yöntemi(MR-ASM) kullanarak geliştirdikleri ve işaret dili için çok önemli olan yüz ifadelerinin tanınmasını sağlayan sistemdir. Bu çalışmada yüzde bulunan işaret noktaları kullanılarak harekete dair özellikler çıkarılmaktadır. Bu özelliklerin sınıflandırılmasında da Destekçi Vektör Makinesi(SVM) kullanılmıştır. Sonuç olarak 7 adet işaret dili hareketi kullanarak %90 düzeyinde bir sınıflandırma başarısı elde etmişlerdir.

Memis ve çalışma arkadaşlarının sunduğu tanımlama sisteminde de Kinect sensörleri üzerinden elde edilen uzaysal-zamansal özellikler kullanılmıştır. Geliştirilen sistemde hareket farkları temel alınarak çıkartılan birikimli hareket görüntüleri kullanılmıştır. Türkçe işaret dili kategorisindeki 1002 hareket üzerinde %90 civarında bir sınıflandırma başarıları mevcuttur.

Ek olarak, Kim ve çalışma arkadaşlarının [11] yaptığı çalışmada ise işaret dilinin tanınmasına yönelik olarak hibrit yapay sinir ağları yöntemi kullanılmıştır. Bu sistemde iki tür yapay sinir ağı kullanılmıştır. Bunlar CNN(Convolutional Neural Network) model ve WFMM modeldir. Bu sistemde özellik çıkarımı için hareket geçmişi ağırlığı yöntemi kullanılmıştır.

III. YÖNTEM

Öncelikle her bir video için Hareket Geçmişi Görüntüsü oluşturulmuştur. Burada kullanılan temel mantık, ön planda bulunan nesnelerin ikili görüntülerini kullanarak Hareket Geçmişi Görüntüsü yöntemi ile birlikte hareketleri tanımlayan şablonları oluşturmaktadır.

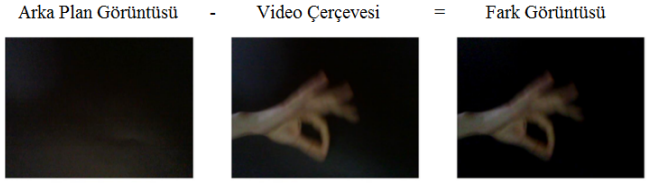
Bu çalışma kapsamında videolardan Hareket Geçmişi Görüntüsü bilgisini çıkarmak adına, Türkçe'ye özgü olan işaret dili için 8 farklı hareketi barındıran bir eğitim kümesi oluşturulmuştur. Bu aşamadan sonra da 800 boyutlu özellik vektörleri, her bir video için tanımlayıcı olarak elde edilmiştir.

A. Hareket Geçmişi Görüntüsü'nü Elde Etme

İşaret diline ait hareketleri içeren videolardan Hareket Geçmişi Görüntüsü bilgisini elde etmek ve ön plan bilgisini çıkarmak amacıyla daha önce elde edilen arka plan bilgisi kullanılmıştır. Temel olacak, her bir çerçeve arka plan görüntüsünden çıkartılır. Sonuç olarak ön plandaki nesnenin piksel değerleri elde edilmiş olur. Şekil 1' de Türkçe'ye özgü işaret diline ait 'fayda' hareketi ve bu harekete ait ön plan bilgilerinin nasıl çıkarıldığı gösterilmektedir. Her bir video içerisinde sabit siyah bir arka plan kullanılmıştır.

Ardından ikili ön plan bilgisini elde edebilmek için, eşik değeri 40 seçilmiştir. Bu şekilde eşik değerinden büyük olan piksellerin değeri 1 değerine eşitlenirken, eşik değerinden küçük olan piksellerin değeri de 0 değerine eşitlenmiştir. Şekil 2' de eşit değerinin fark görüntüsüne uygulandıktan sonraki hali gösterilmiştir.

İkili fark görüntülerinin elde edilmesinin ardından videodaki her bir çerçeve için verilen formule [5] göre Hareket Geçmişi Görüntüsü bilgisi hesaplanır. Bu hesaplamada bütün ikili fark görüntüleri τ-çerçeve sırasına göre videonun Hareket Geçmişi Görüntüsü bilgisini çıkarır. Hτ: t değeri 1 ile τ arası değerler alır (1).



Şekil 1: 'fayda' hareketi için arka plan çıkarma işlemi

Şekil 2: 'fayda' hareketi için eşik değerine göre çıkartılmış olan ikili fark görüntüsü

$$H\tau(x,y,t) = \begin{cases} \tau & \text{eğer } D(x,y,t) = 1 \\ Max\ (0, H\tau(x,y,t-1) - 1) & \\ & \text{aksi halde} \end{cases} \quad (1)$$

Bu formülizasyon [5] kullanılarak bir hareketin parçası olan tüm çerçeveler bir sonuç resminin içine aktarılmıştır. Burada basitçe, çerçevedeki piksel değeri eğer 1 ise, bulunulan çerçevenin zamansal numarası o piksele aktarılmaktadır. Eğer o piksel 0 ise, bir önceki çerçevedeki değerine bakılarak herhangi bir hareket olup olmadığı bilgisi öğrenilir. Bir hareket var ise, yani o piksel 0'dan farklı ise, bir önceki çerçevenin zamansal numarası bulunulan piksele aktarılır.

Böylece, şu an işlem yapılan çerçeveye ilişkin hareket bilgisi eklenirken bir önceki çerçevelerde bulunan hareket bilgileri de eklenmiş olmaktadır. Hareket Geçmişi Görüntüsü yönteminin bir hareket için ürettiği sonuç görüntü Şekil 3'te gösterilmiştir.

Şekilde de görüldüğü gibi Hareket Geçmişi Görüntüsü yöntemi hareket geçmişine ait bilgileri tutmaktadır. Ek olarak, bu görüntüler için tanımlayıcılar belirlenmeden önce görüntüdeki maksimum değere göre normalizasyon işlemi gerçekleştirilmiştir..

B. Tanımlayıcı

Hareket Geçmişi Görüntüsü yöntemi ile her video için videoyu temsil edecek olan tanımlayıcılar oluşturulmuştur.

Şekil 3: 'fayda' hareket sınıfına ait Hareket Geçmişi Görüntüsü sonucu

Çalışmamızda her bir Hareket Geçmişi Görüntüsü sonucu 20x20 boyutundaki hücrelere bölünmüş ve her bir hücre için ortalama ile varyans değerleri hesaplanmıştır. Sonuç olarak her bir video için 1x800 boyutunda tanımlayıcılar elde edilmiştir.

Hücre boyutlarının farklı boyutlardaki seçimi, hesaplama aşamasında farklı sonuçların ortaya çıkmasına sebep olabilmektedir. Oluşturduğumuz veri kümesi için en iyi olacak hücre boyutunu belirlemek ve en uygun olanı uygulamak için yaptığımız çalışmanın detayı Bölüm V. içerisinde açıklanmıştır.

C. Değerlendirme

Her bir videoyu temsil eden tanımlayıcılar elde edildikten sonra çapraz doğrulama yapılarak en yakın komşuluk sınıflandırılması gerçekleştirilmiştir. Burada her bir video test örneği olarak kullanılmış ve her birinin hangi sınıfa ait olduğu geriye kalan videolar ile uzaklığı ölçülerek gerçekleştirilmiştir. Karşılaştırmanın sonunda en yakın videonun sınıf numarası karşılaştırmada kullanılan test örneğine atanmıştır.

Sınıflandırmanın doğruluğunu bulmak için yapılan değerlendirmede kendi hazırladığımız eğitim kümesi kullanılmıştır. Buna ek olarak gerçek-zamanlı ve kullanıcı ile etkileşimli olan bir uygulama geliştirilerek, gerçek kullanıcılar tarafından uygulamanın kullanımı sırasında, algoritmanın testini yapma şansı da elde edilmiş oldu.

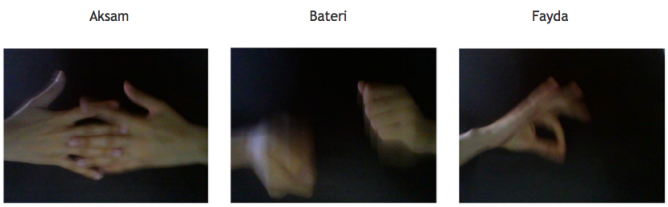
IV. VERİ KÜMESİ

Literatürde Türkçe'ye özgü işaret diline yönelik data eksikliği olduğundan eğitim ve test aşamaları için kendi veri kümemizi oluşturduk. Eğitim kümesini oluşturmak için ilk aşamada Türkçe'ye özgü işaret dilinden 8 farklı hareket seçildi. Bu hareketlerin ifade ettiği anlamlar şu şekildedir: "acıkmak, akşam, arkadaş, bateri, direksiyon, fayda, makas, sevmek".

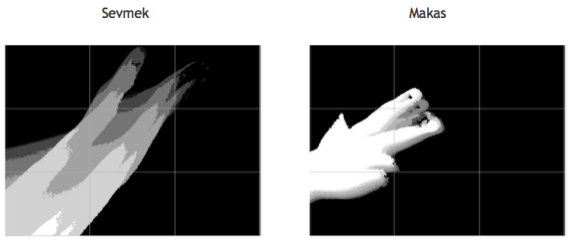
Ardından arka planı siyah olacak şekilde web-cam aracılığıyla video çekimleri gerçekleştirildi. Her bir hareket sınıfı için 5 tane video çekildi. Sonuç olarak 8 ayrı sınıf için toplamda 40 video çekilmiş olundu. 3 farklı hareketin örnek çerçeveleri Şekil 4'te gösterilmiştir.

V. DENEYSEL SONUÇLAR

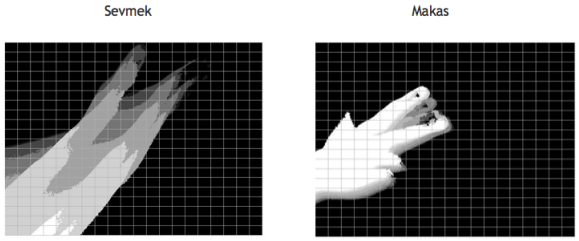
Tanımlıyıcıların hücre boyutunun doğruluk oranını etkilediğinden daha önceki başlıklarda bahsedilmişti. Kendi eğitim kümemiz üzerinden Hareket Geçmişi Görüntüsü yöntemi uygulanırken farklı hücre boyutları kullanıldı.



Şekil 4: "akşam", "bateri", "fayda" ifadelerine karşılık gelen 3 farklı hareket



Şekil 5: 3x3 boyutundaki hücreler



Şekil 6: 20x20 boyutundaki hücreler

Videolar için tanımlayıcı oluşturması aşamasında hücrelerin ortalama ve varyans değerleri hesaplanırken büyük boyutlu hücre seçimi yapılması hareket tanımını yeterli düzeyde temsil edemediği için daha küçük doğruluk oranının çıkmasına sebep olmuştur. Şekil 5'te gösterilen "sevmek" ve "makas" hareketlerine ait tanımlayıcı oluşturulması aşamasında hücrelerin boyu 3x3 seçildiği için tanımlayıcılar hareketi iyi temsil edememiştir.

Hücre boylarının bu kadar büyük olması neticesinde bu iki hareketin birbirine benzerliği ortaya çıkmıştır. Bunun aksine Şekil 6'da gösterildiği üzere yine aynı iki hareketin tanımlayıcılarını oluşturma aşamasında hücrelerin Hareket Geçmişi Görüntüsü düzlemini 20x20 olacak şekilde bölmesi durumunda ise, bu iki hareketin temsiliyetinin birbirinden rahatlıkla ayrılabildiği gözlemlenmiştir.

Tüm sınıflandırma doğrulukları konfüzyon matrisinden hesaplanmıştır. Farklı hücre boyutları için doğruluk sonuçları Tablo I'de verilmiştir. Bu tablodan anlaşılabileceği üzere, hücre sayısı arttıkça, hareketin detayları daha iyi belirlenebildiği için, sınıflandırma doğruluk oranları da aynı doğrultuda artmaktadır.

Örneğin, Hareket Geçmişi Görüntüsü düzlemi 3x3 olacak şekilde hücrelere bölündüğünde, hücre boyutlarının çok büyük olması sebebiyle gerekli detaylı bilgi elde edilememektedir. Bu durumun doğruluk sonucunun daha düşük çıkmasına, yani yanlış sınıflandırılmalara neden olduğu gözlenmektedir.

Fakat hücre boyutları küçültülüp, sayısı arttırıldığında, örneğin 20x20'lik bir bölümlendirme yapıldığında, doğruluk sonucu artmakta yani hareketler daha doğru sınıflandırılmaktadır.

Hücre Sayısı	Doğruluk
3x3	82%
8x8	87.5%
15x15	92.5%
20x20	**95%**

Tablo I: Hücre sayısına göre elde edilen doğruluk sonuçları

VI. ÖZET VE TARTIŞMA

Bu çalışma işitme veya konuşma engelli bireylerin toplumdaki diğer bireylerle daha kolay ve sağlıklı iletişim kurmalarına yardımcı olacak bir çalışmadır. Bu çalışma kapsamında, 8 farklı Türkçe'ye özgü işaret dilinde bulunan hareket ile eğitim kümesi oluşturuldu. Ardından Hareket Geçmişi Görüntüsü yöntemi video ve hareketleri temsil etmek amacıyla kullanılırken, sınıflandırma aşaması için de en yakın komşuluk yaklaşımı uygulandı. Sonuç olarak %95 başarı oranıyla bu eğitim kümesinde bulunan bu 8 hareketin tanınması işlemi başarıyla gerçekleştirilebilmiştir.

Sunulan yöntemin doğru sonuçlar verebilmesi için sabit bir arka plan kullanımı zorunluluğu, sistemin kullanımını kısıtlamaktadır. Fakat bu çalışma bir ilk basamak olarak düşünülebilir. Basit bir yöntem ve basit bir veri kümesi kullanılarak, ileriki aşamalar için taban oluşturacak bir çalışma yapılmıştır. Gelecekte arka plandan bağımsız bir yöntem ile bu çalışmanın geliştirilmesi mümkündür. Böylece kullanımı daha kolay bir sistem ortaya çıkacaktır.

KAYNAKÇA

[1] C. Cedras, M. Shah, Motion-based recognition a survey, Image and Vision Computing 13 (1995) 129–155.

[2] T. B. Moeslund, Computer vision-based human motion capture– a survey, University of Aalborg Technical Report LIA 99 (1999).

[3] J. K. Aggarwal, Q. Cai, Human motion analysis: A review, Computer vision and image understanding 73 (1999) 428–440.

[4] K. Imagawa, H. Matsuo, R.-i. Taniguchi, D. Arita, S. Lu, S. Igi, Recognition of local features for camera-based sign language recognition system, in: Pattern Recognition, 2000. Proceedings. 15th International Conference on, volume 4, IEEE, 2000, pp. 849–853.

[5] J. W. Davis, A. E. Bobick, The representation and recognition of human movement using temporal templates, in: Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, IEEE, 1997, pp. 928–934.

[6] T. Starner, J. Weaver, A. Pentland, Real-time american sign language recognition using desk and wearable computer based video, Pattern Analysis and Machine Intelligence, IEEE Transactions on 20 (1998) 1371–1375.

[7] P. Mekala, Y. Gao, J. Fan, A. Davari, Real-time sign language recognition based on neural network architecture, in: System Theory (SSST), 2011 IEEE 43rd Southeastern Symposium on, IEEE, 2011, pp. 195–199.

[8] H. Haberdar, S. Albayrak, Real time isolated turkish sign language recognition from video using hidden markov models with global features, in: Computer and Information SciencesISCIS 2005, Springer, 2005, pp. 677–687.

[9] I. Ari, A. Uyar, L. Akarun, Facial feature tracking and expression recognition for sign language, in: Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on, IEEE, 2008, pp. 1–6.

[10] A. Memis, S. Albayrak, Turkish sign language recognition using spatio-temporal features on kinect rgb video sequences and depth maps, in: Signal Processing and Communications Applications Conference (SIU), 2013 21st, IEEE, 2013, pp. 1– 4.

[11] H.-J. Kim, S.-J. Park, S.-K. Lee, Sign language recognition using motion history volume and hybrid neural networks, International Journal of Machine Learning and Computing 2 (2012) 750–753.