

# GELİŞTİRİLMİŞ YOĞUN GEZİNGELER KULLANARAK YALITILMIŞ İŞARET DİLİ TANIMA

## ISOLATED SIGN LANGUAGE RECOGNITION USING IMPROVED DENSE TRAJECTORIES

Oğulcan Özdemir, Necati Cihan Camgöz, Lale Akarun  
Bilgisayar Mühendisliği Bölümü, Boğaziçi Üniversitesi  
{ogulcan.ozdemir, cihan.camgoz, akarun}@boun.edu.tr

**Özetçe** —İşaret dili tanıma, işitme engellilerin ana iletişim dili olan işaret dilini, bilgisayar uygulamalarında kullanmalarını mümkün hale getirdiği için, son yıllarda araştırmanın ilgi odağı olmuştur. Bu çalışmada, işaret dili tanıma için büyük çaplı eylem tanımada kullanılan Geliştirilmiş Yoğun Gezinge (Improved Dense Trajectories - IDT) öznitelikleri kullanan bir yöntem önerilmiştir. Önerilen bu yöntemde her bir görüntüyü tek bir betimleyici ile gösterebilmek için Fisher Vektör (FV) gösterimi kullanılmıştır. Bu betimleyicilerin yedi farklı bileşen kombinasyonu ile deney düzenekleri kurulmuş ve doğrusal Destek Vektör Makineleri (Support Vector Machines - SVM) kullanılarak 200 farklı işaret için eğitim ve sınav yapılmıştır. Deneylerin sonucunda önerilen yöntem, IDT'nin Optik Akışların Histogramı (Histogram of Optical Flow - HOF) ve Hareket Sınırlarının Histogramı (Motion Boundary Histogram - MBH) bileşenleri beraber kullanıldığında %80,43 başarı sağlamıştır.

**Anahtar Kelimeler**—geliştirilmiş yoğun gezinmeler, işaret dili tanıma, işmar tanıma

**Abstract**—Sign language recognition has been the focus of research in recent years because it has enabled the use of sign languages, which are the main medium of communication for the hearing impaired, for human-computer interaction. In this work, we propose a method to recognize signs using Improved Dense Trajectory (IDT) features which were previously used in large-scale action recognition. Fisher Vectors (FV) are used to represent sign samples in the proposed method. Seven different combinations of features were compared using a test set of 200 signs, using a Support Vector Machine (SVM) classifier. The best combination yielded 80.43% recognition performance when Histogram of Optical Flow (HOF) and Motion Boundary Histogram (MBH) components were used together.

**Keywords**—improved dense trajectories, sign language recognition, gesture recognition

### I. GİRİŞ

İşaret dili, duyma engellilerin ana iletişim aracıdır. Toplumun geri kalanının işaret dili bilmemesi nedeniyle duyma engelliler yazılı materyaller veya işaret dili çevirmenleri aracılığıyla duyma engeli olmayan insanlar ile iletişim sağlamaktadırlar. Duyma engellilerin okuma yazma oranının düşük olması ve işaret dili çevirmeni sayısının yeterli olmaması nedeniyle duyma engelliler günlük yaşantılarında zorluklarla karşılaşmaktadırlar. Bu problemi ortadan kaldırmak için son yıllarda işaret dili tanıyan ve konuşma diline çeviren insan bilgisayar etkileşim platformları geliştirilmiştir [1].

Pek çok işaret dili tanıma uygulaması sınırlı dağarcıkta yalıtılmış işaretleri tanıma üzerine çalışmaktadır [2], [3]. Bu uygulamalardan bazıları tüm işaret dağarcığı içinden tanıma yapmayı amaçlayarak kullanıcının işaretlerini konuşma diline çevirmektedir [2], [4]. İşaret dili uygulamalarındaki başka bir yaklaşım ise kullanıcılarla soru-cevap biçiminde etkileşim sağlamak ve kullanıcıları yönlendirerek hem etkileşimi hem de tanıma işlemini kolaylaştırmayı hedeflemektedir. [3].

Mevcut işaret dili tanıma uygulamaları giyilebilir sensörler ve görme tabanlı sensörleri kullanmasına bağlı olarak iki gruba ayrılmaktadır [5]. Giyilebilir sensörlerin pahalı oluşu ve her kullanıcı için kalibrasyona ihtiyaç duyması nedeniyle sistemlerde görsel algılayıcıların kullanılması tercih edilmektedir.

Görme tabanlı işaret dili tanıma sistemleri beş temel bileşenden oluşmaktadır. Bunlar işaretçinin üst vücut pozunun kestirimi, öznitelik çıkarımı, özniteliklerin zamansal modellenmesi, özniteliklerin normalizasyonu ve işaretlerin sınıflandırılmasıdır. Renkli tabanlı kameralar kullanan sistemlerde, işaretçinin bulunması ve üst vücut pozunun kestirimi renk belirsizliğini nedeniyle hala zorlu bir iştir. Buna ek olarak renkli görsel algılayıcılar ortamın aydınlanmasından çokça etkilenmektedir. Bu nedenle araştırmacılar, günümüzde yaygınlaşan Microsoft Kinect v.b. kullanıcı derinlik kameralarını kullanmaya yönelmişlerdir. Derinlik imgeleri sayesinde kullanıcıları arka plandan ayırmak zor bir işlem olmaktan çıkmıştır. Ayrıca Shotton v.d. [6] tarafından geliştirilmiş rassal karar ağaçları tabanlı üst vücut poz kestirimi algoritması sayesinde, derinlik imgelerinden çerçeve tabanlı vücut poz kestirimi gerçek zamanlı olarak yapılabilmektedir.

İşaret dili kullanıcıları, el şekilleri, el hareketleri, el pozisyonu, üst vücudun pozunu ve yüzsel işmarları kullanarak birbirleriyle iletişim kurmaktadırlar. Benzer bir şekilde işaret dili tanıma sistemleri işaretlerden ellerin şeklini, hareketini, pozisyonunu, üst vücudun pozunu ve yüzsel işmarları betimleyen öznitelikler çıkarmakta ve bunları işaretleri tanımak için kullanmaktadırlar.

İşaret dilleri, her diğer konuşma dili gibi, zaman serileri olduğundan çıkartılan özniteliklerin zamanda değişiminin modellenmesi, işareti tanımak için büyük önem taşımaktadır. Zaman modelleme yöntemlerinden Saklı Markov Modelleri (Hidden Markov Models - HMMs) [7], [8] ve Dinamik Zaman Bükmesi (Dynamic Time Warping - DTW) [1], [9] işaret dili tanımada sıkça kullanılmaktadır. Uzamsal ve zamansal olarak modellenen işaretler sistemlerin başarımını arttırmak ve kullanıcıdan bağımsız olarak çalışabilmesini sağlamak için normalize edilmekte [10] ve ardından sınıflandırılmaktadır.

Bu çalışmada, büyük çaplı eylem tanıma için Wang v.d. [11] tarafından önerilmiş Geliştirilmiş Yoğun Gezingerler (Improved Dense Trajectories - IDT) yöntemi işaret dili tanıma için kullanılmıştır. Geliştirilmiş Yoğun Gezingerlerin kullanımındaki motivasyon, bu betimleyicinin işaret dili örneklerini tanımakta sıkça kullanılan üst vücut hareketleri ve el şeklini yoğun bir şekilde tanımlıyor olmasıdır. IDT daha önce işaret dili tanımak için kullanılmamış olsa da, benzer bir alan olan işmar tanımada, 20 sınıflık bir problemde işmarları ayırma ve tanıma için, kayan pencere yöntemi ile birleştirilerek kullanılmıştır [12].

Yalıtılmış işaret dili örneklerinin tanınması üzerine bu çalışmamızda, IDT'nini, işaret dili tanımak için kullanılabilirliği ve IDT'nin bileşenleri olan, HOG (Histogram of Oriented Gradients), HOF (Histogram of Optical Flow) ve MBH'in (Motion Boundary Histogram) tanıma başarımına etkisi incelenmiştir. Çin İşaret Dili veritabanı DEVISIGN'in 200 sınıflık bir alt kümesinde yapılan deneylerde kullanıcı bağımsız olarak tanıma deneyleri yapılmış ve IDT'nin MBH ve HOF bileşenlerin kullanıldığı kurulumda %80,43 oranında başarımlar elde edilmiştir.

Bölüm II'de Geliştirilmiş Yoğun Gezingerler anlatılmaktadır. Bölüm III'de IDT kullanılarak işaret dili tanımak için önerilen yöntem anlatılmıştır. Bölüm IV'de ise önerilen yöntemin değerlendirilmesi amacıyla yapılan deneylerin kurulumu, deneylerin yapıldığı veri kümesi ve deney sonuçlarından bahsedilmiştir. Son olarak Bölüm V'de yapılan deneyler sonucunda çıkartılan vargılar aktarılmaktadır.

## II. GELİŞTİRİLMİŞ YOĞUN GEZİNGELER

Yoğun Gezingerler (Dense Trajectories), Wang v.d. [11] tarafından, eylem tanıma probleminde görüntüleri tanımlamak için geliştirilmiş bir betimleyicidir. Önerilen yöntem öncelikle görüntülerden yoğun bir biçimde uzamsal öznitelikleri örneklemektedir. Ardından örneklenen öznitelikler yoğun optik akış alanları kullanılarak zamansal olarak takip edilmekte ve eylemi betimleyici gezingerler elde edilmektedir. Son olarak elde edilen gezingerler arasından statik ve ani yer değişikliği özellikleri gösteren gezingerler gürültü olarak kabul edilerek elenmektedir.

Gezingerlerin uzamsal öznitelikleri Yönlü Gradyan Histogramu (Histogram of Oriented Gradients - HOG) [13] kullanılarak betimlenmektedir. Gezingerlerin hareketleri ise Optik Akış Histogramu (Histogram of Optical Flow - HOF) [14] ve Hareket Sınırlarının Histogramu (Motion Boundary Histogram - MBH) [15] kullanılarak betimlenmektedir.

Yoğun Gezingerler kullanılarak eylem algılama işleminde karşılaşılan en büyük problemlerden biri kameranın hareketli olmasından dolayı ortaya çıkan kamera hareketini betimleyen gezingerlerdir. Bu problemi ortadan kaldırmak adına Wang v.d. gezingerleri çıkartırken kameranın hareketini de hesaba katan Geliştirilmiş Yoğun Gezingerleri (Improved Dense Trajectories) önermiştir [16]. Kamera hareketinden kaynaklanan gezingerleri bulabilmek için, ardışıl çerçeveler arasında Hızlandırılmış Gürbüz Öznitelikler (Speeded Up Robust Features - SURF) [17] ve Rasgele Örneklemlerin Mutabakatı (Random Sample Consensus - RANSAC) [18] kullanılarak homografi hesaplanmaktadır. Hesaplanan homografiye uyan gezingerler kameranın hareketinden kaynaklandığı varsayılarak elenmektedir.

Ancak, yapılan bu işlemde sonra görüntüdeki insanların çerçeveyi baskın olarak doldurduğu durumlarda hesaplanan homografinin kamera hareketi kestiriminde etkin olmadığı görülmüştür. Bundan dolayı homografi hesabı yapılmadan önce insan sezici kullanıldıktan sonra, sonuçlar dahilinde insanların bulunduğu sınırlayıcı kutular (bounding box) çerçeveden kesilmektedir. Elde edilen çerçeveler üzerinden tekrar bir homografi hesaplandıktan sonra, yeni bir optik akış hesaplanmaktadır ve önceden belirlenen yer değişim eşik değerinin altında kalan gezingerler kamera hareketi olarak varsayılmakta ve elenmektedir.

## III. YÖNTEM

Yalıtılmış işaret dili örneklerinin tanındığı bu çalışmada önerilen tanıma sisteminin akış diyagramı Şekil 1'de verilmiştir. İlk olarak işaret dili örneklerine ait görüntülerden Geliştirilmiş Yoğun Gezinger öznitelikleri [16] elde edilmektedir.

Çıkartılan gezingerler boyut olarak büyük olduğu için, gezingerlerin tümünün aynı anda işlenmesi hem işlem gücü hem de bellek alanı kısıtlarından dolayı mümkün olmamaktadır. Bu problemleri aşmak için eğitim aşamasında, eğitim kümesindeki görüntülere ait gezingerlerden  $N * k$  adet örnek gezinger seçilmektedir.  $N$  örnekleme oranını temsil ederken,  $k$  gezingerlere uygulanacak Gauss Karışım Modeli (Gaussian Mixture Model - GMM) için küme sayısını temsil etmektedir.

Yapılan rastgele örneklemeden sonra elde edilen 396 boyutlu gezinger vektörlerinin (96 boyutlu HOG, 108 boyutlu HOF ve 192 boyutlu MBH) boyutlarını azaltmak için %99 enerji korunumlu Temel Bileşen Analizi (Principal Component Analysis - PCA) uygulanarak  $D$  boyutlu öznitelik vektörleri elde edilmiştir.

Önişlemeye tabi tutulan öznitelikleri kullanarak görüntüler, Öznitelikler Torbası Modeline (Bag of Features Model - BOF) benzer bir öznitelik havuzlama yöntemi olan ve eylem tanıma probleminde BOF modelinden daha yüksek başarıma sahip olan Fisher Vektör (Fisher Vector - FV) [19] kullanılarak betimlenmiştir. Fisher Vektörlerin hesaplanmasında kullanılan GMM, eğitim aşamasında rastgele olarak seçilen gezingerler kullanılarak oluşturulmuştur.

Veri kümesindeki her bir örneği betimleyecek Fisher Vektörler, örneklere ait görüntülerden elde edilen gezingerlere rastgele seçilmiş gezingerler ile eğitilen PCA ve GMM'ler kullanılarak elde edilmektedir. Son olarak her bir işaret örneği  $2Dk$  boyutlu Fisher vektörler ile betimlenmektedir.

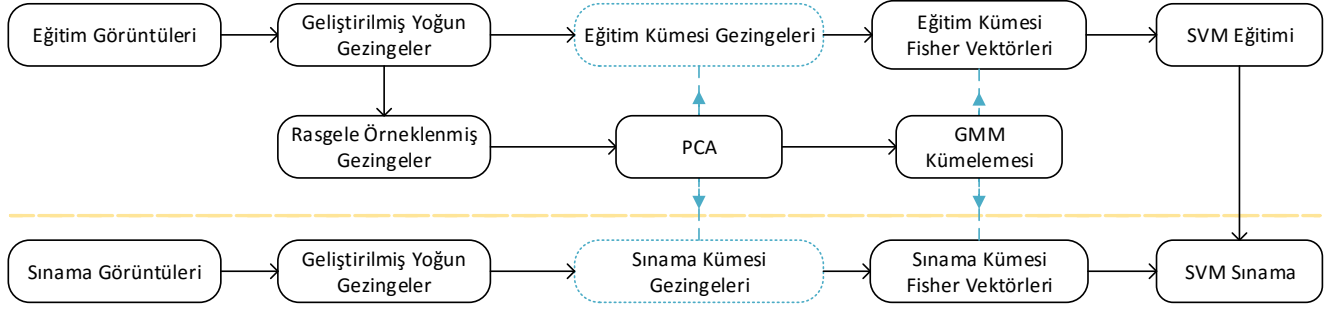
Hesaplanan fisher vektörler doğrusal olarak iyi ayrılabilir olduğundan, eğitim örneklerine ait betimleyiciler kullanılarak çok sınıflı doğrusal Destek Vektör Makineleri (Support Vector Machine - SVM) eğitilmiştir. Eğitilen SVM'ler sınıma kümesine ait Fisher vektörleri sınıflandırmak için kullanılmaktadır.

## IV. DENEYSEL SONUÇLAR

### A. Kullanılan Veri Kümesi

Önerilen yöntemin performansını değerlendirebilmek için, İşaret dili tanıma için Kinect İşaret Dili Çalışma Grubu tarafından toplanmış olan Çin İşaret Dili kelimelerini barındıran DEVISIGN veri seti üzerinde deneyler yapılmıştır<sup>1</sup>. Bu veri seti

<sup>1</sup><http://vipl.ict.ac.cn/homepage/KSL/home.html>



Şekil 1: Önerilen yöntemde kullanılan akışın şeması

genel kullanıma açık olup DEVISIGN-G, DEVISIGN-D ve DEVISIGN-L olmak üzere üç adet alt veri kümesine sahiptir. Bu çalışmada kullanılan DEVISIGN-L alt veri kümesinde, sekiz kullanıcı (dört kullanıcı iki tekrarlı) ve her bir kullanıcı için 2000 adet işaret görüntüsü bulunmaktadır.

Yapılan deneylerde DEVISIGN-L veri kümesinden 200 işaretlik bir alt küme alınarak Geliştirilmiş Yoğun Gezinge öznelikleri çıkartılmıştır. Özneliklerin görselleştirilmiş örnekleri Şekil 2’de görülmektedir. Çıkarılan öznelikler önerilen yöntemde kullanılarak sınıflandırmadaki başarımı ölçülmüştür.

### B. Deneysel Düzen ve Sonuçlar

Geliştirilmiş Yoğun Gezingerlerin bileşenleri olan HOG, HOF ve MBH’in yedi değişik kombinasyonu kullanılarak deneyler yapılmıştır. Önerilen yöntem kullanılarak 200 işaretlik alt veri kümesi sekiz kullanıcı için her seferinde bir kullanıcı dışarıda bırakılarak (Leave One Out) sekiz defa doğrusal SVM ile eğitilmiş ve sınanmıştır. Her kattan önce alt veri kümesi  $N = 1000$  örnekleme oranı ile rastgele beş defa örneklendirilip,  $k = 64$  küme boyutu ile kümelendirildikten sonra bütün alt veri kümesi üzerinden Fisher vektörler elde edilmiştir.

Doğrusal SVM parametreleri, ızgaralama (Grid Search) ve çapraz doğrulama (cross-validation) yöntemleri kullanılarak  $C = \{2^{-3}, 2^{-1}, 2^1, 2^3, 2^5, 2^7, 2^9\}$  ve beş kat için öğrenildikten sonra en iyi sonuç veren parametreler seçilmiştir.

Her bir deney düzeneği üzerinden örneklendirilen bütün veriler ile sonuç alınmıştır. Her bir bileşen kombinasyonu için alınan sonuçların ortalaması ve standart sapması hesaplanmıştır. Farklı katların performansları, Tablo 1’de verilmiştir. Yedinci kullanıcının dışarıda bırakıldığı katın performansının, diğerlerinden önemli derecede düşük olduğu görülmektedir. Bu kullanıcının işaretleri hatalı yaptığı saptanmıştır. Buna dair bir örnek Şekil 2’de görülmektedir. Bu nedenle, tüm katların ortalamasının yanısıra, DBK-7 dışında diğer katların ortalaması da verilmiştir.

Ortalama değerlere bakıldığında Geliştirilmiş Yoğun Gezingerlerin bileşenlerinden MBH’in en yüksek başarımı (%80, 34) verdiği; HOF’un yakın ancak daha düşük (%75, 43) başarımı gösterdiği, HOG’un ise performansının düşük

(%52, 53) olduğu görülmüştür. MBH ile HOF’un beraber kullanılması halinde, başarımlar %80, 43’e çıkmakta; ancak HOG’un kullanımı, başarımları düşürmektedir.

## V. VARGILAR

Görüntü bazlı işaret dili tanıma alanında yapılan önceki çalışmalarda bu problemi zaman serisi olarak modelleyerek çözmek için Saklı Markov Modelleri, Dinamik Zaman Bükmesi veya Saklı Koşullu Rastlantısal Alanlar gibi yöntemler kullanılmıştır.

Bu çalışmada büyük çaplı hareket görüntülerini tanımak için kullanılan Geliştirilmiş Yoğun Gezingerler yöntemi kullanılarak DEVISIGN-L veri kümesi üzerinde gezingerlerin farklı bileşenlerini kullanarak deneyler yapılmıştır.

Yapılan deneylerde alınan sonuçlarda HOF ve MBH tanımlayıcılarının işaret dili tanımadaki performansı arttırdığı görülürken, HOG tanımlayıcısının başarımlar oranını düşürdüğü görülmüştür.

Kullanılan yöntemin özelliği, tüm görüntü için tek bir betimleyici çıkarılmasıdır. Bu nedenle, zamansal değişimlere karşı gürbüz değildir. İlerideki çalışmalarda, Geliştirilmiş Yoğun Gezingerlerin değişken görüntü parçalarına uygulanması ve diğer zamansal modelleme yöntemleriyle birleştirilmesi düşünülmektedir.

## KAYNAKÇA

- [1] N. C. Camgöz, “Human-computer interaction platform for the hearing impaired in healthcare and finance applications,” Master’s thesis, Boğaziçi University, 2016.
- [2] X. Chai, G. Li, Y. Lin, Z. Xu, Y. Tang, X. Chen, and M. Zhou, “Sign language recognition and translation with kinect,” in *IEEE Conference on Automatic Face and Gesture Recognition*, 2013.
- [3] M. M. Süzgun, H. Özdemir, N. C. Camgöz, A. A. Kindroğlu, D. Başaran, C. Togay, and L. Akarun, “Hospisign: An interactive sign language platform for hearing impaired,” *Deniz Bilimleri ve Mühendisliği Dergisi*, vol. 11, no. 3, 2015.
- [4] S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, and S. Abbott, “Tessa, a system to aid communication with deaf people,” in *Fifth International ACM Conference on Assistive Technologies*, pp. 205–212, 2002.
- [5] C. Wang, W. Gao, and S. Shan, “An approach based on phonemes to large vocabulary chinese sign language recognition,” in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pp. 411–416, IEEE, 2002.



Şekil 2: Geliştirilmiş Yoğun Gezintilerin kullanılan veri kümesi üzerinden görselleştirilmiş örnekleri

Tablo I: 7 farklı bileşen kurulumu ve 8 tekrarlı bir kullanıcıyı dışarıda bırakan çapraz doğrulama deneylerinin performansı (Dışarıda Bırakılan Kullanıcı - DBK)

	HOG	HOF	MBH	HOG + HOF	HOG + MBH	HOF + MBH	HOG + HOF + MBH
DBK-1	46,55 ± 2,04	73,45 ± 0,74	<b>78,15 ± 0,74</b>	77,75 ± 0,25	77,85 ± 1,51	77,10 ± 0,76	77,30 ± 1,25
DBK-2	46,55 ± 2,23	76,00 ± 1,02	<b>81,95 ± 1,10</b>	77,45 ± 0,48	71,65 ± 2,86	81,85 ± 0,42	78,25 ± 1,22
DBK-3	60,35 ± 2,00	83,55 ± 0,38	<b>88,10 ± 0,65</b>	86,25 ± 0,94	85,20 ± 1,63	87,55 ± 0,33	87,45 ± 1,24
DBK-4	31,85 ± 0,99	72,00 ± 0,88	75,85 ± 0,22	71,65 ± 1,23	61,70 ± 4,02	<b>78,00 ± 1,10</b>	74,80 ± 1,88
DBK-5	64,00 ± 0,61	74,90 ± 1,64	<b>81,90 ± 1,52</b>	78,30 ± 0,76	78,40 ± 1,39	81,50 ± 1,54	81,50 ± 0,79
DBK-6	62,70 ± 3,68	71,20 ± 2,17	<b>77,80 ± 1,04</b>	77,10 ± 0,65	74,90 ± 2,56	77,10 ± 1,56	77,50 ± 2,67
DBK-7	31,10 ± 3,03	48,40 ± 1,78	51,40 ± 0,96	<b>55,50 ± 1,41</b>	46,30 ± 2,80	52,10 ± 3,54	52,40 ± 4,08
DBK-8	55,70 ± 0,97	77,10 ± 0,42	78,60 ± 1,24	79,00 ± 2,18	77,90 ± 1,14	<b>79,90 ± 1,08</b>	76,40 ± 2,04
Ortalama	49,85 ± 1,95	72,05 ± 1,13	<b>76,72 ± 0,93</b>	75,38 ± 0,99	71,74 ± 2,24	<b>76,89 ± 1,29</b>	75,70 ± 1,90
7.ci kullanıcı hariç	52,53 ± 1,79	75,43 ± 1,03	<b>80,34 ± 0,93</b>	78,21 ± 0,93	75,37 ± 2,16	<b>80,43 ± 0,97</b>	79,03 ± 1,59

- [6] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, pp. 116–124, 2013.
- [7] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, "American sign language recognition with the kinect," in *Proceedings of the 13th international conference on multimodal interfaces*, pp. 279–286, ACM, 2011.
- [8] S. Lang, M. Block, and R. Rojas, "Sign language recognition using kinect," in *Artificial Intelligence and Soft Computing*, pp. 394–402, Springer, 2012.
- [9] C. Keskin, A. T. Cemgil, and L. Akarun, "Dtw based clustering to improve hand gesture recognition," in *Human Behavior Understanding*, pp. 72–81, Springer, 2011.
- [10] A. Farhadi, D. Forsyth, and R. White, "Transfer learning in sign language," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2007.
- [11] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, "Action recognition by dense trajectories," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 3169–3176, IEEE, 2011.
- [12] X. Peng, L. Wang, Z. Cai, and Y. Qiao, "Action and gesture temporal spotting with super vector representation," in *Computer Vision-ECCV 2014 Workshops*, pp. 518–527, Springer, 2014.
- [13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.
- [14] J. Perš, V. Sulić, M. Kristan, M. Perše, K. Polanec, and S. Kovačič, "Histograms of optical flow for efficient representation of body motion," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1369–1376, 2010.
- [15] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Computer Vision-ECCV 2006*, pp. 428–441, Springer, 2006.
- [16] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, pp. 3551–3558, IEEE, 2013.
- [17] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision-ECCV 2006*, pp. 404–417, Springer, 2006.
- [18] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [19] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *International journal of computer vision*, vol. 105, no. 3, pp. 222–245, 2013.