

SAKLI MARKOV MODELLERİ İLE İŞMAR TANIMA İÇİN ALAN UYARLAMA DOMAIN ADAPTATION FOR GESTURE RECOGNITION USING HIDDEN MARKOV MODELS

Necati Cihan Camgöz, A. Alp Kindiroğlu, Lale Akarun
Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi
{cihan.camgoz, alp.kindiroglu, akarun}@boun.edu.tr

Oya Aran
Social Computing Group
Idiap Research Institute
oaran@idiap.ch

Özetçe—Otomatik işmar (el işareti) tanıma, insan bilgisayar etkileşiminde etkin bir alternatif girdi yöntemi olarak yaygınlaşmaktadır. Ancak veri toplamanın ve işaretlemenin zorluğu, verinin kalitesi, değişik kullanıcı ve ortamlardan elde edilen verilerin birbiriyle uyumsuzluğu, bu tarz sistemlerin geliştirilmesini güç kılmaktadır. Bu çalışmanın amacı, alan uyarlama yönteminin işmar tanımaya uyarlanmasıdır. Çalışmamızda az sayıda eğitim örneği ile eğitilen Saklı Markov Modelleri (SMM) tabanlı bir el gezingel tanıma yöntemi, veri toplamanın nispeten kolay olduğu el yazısı verileriyle geliştirilmiştir. Yapılan çalışmalarda kullanılan Öznitelik Artımı yönteminin performansı, işmar tanıma başarısı, değişik işmarların uyarlamaya uygunluğu ve eğitim kümesi boyutunun performans etkisi deneyler yapılarak incelenmiştir.

Anahtar Kelimeler — İşmar Tanıma; Gezingel Tanıma; Derinlik İmgeleri; Saklı Markov Modelleri; Alan Adaptasyonu; Öznitelik Artımı.

Abstract—Gesture recognition is becoming popular as an efficient input method for human computer interaction. However, challenges associated with data collection, data annotation, maintaining standardization, and the high variance of data obtained from different users in different environments make developing such systems a difficult task. The purpose of this study is to integrate domain adaptation methods for the problem of gesture recognition. To achieve this task, domain adaptation is performed from hand written digit trajectory data to hand trajectories obtained from depth cameras. The performance of the applied Feature Augmentation method is evaluated through analysis of recognition performance vs percentage of target class samples in training and through the analysis of the transferability of different gestures.

Keywords — Hand Gesture Recognition; Hand Trajectories; Depth Images; Hidden Markov Models; Domain Adaptation; Feature Augmentation.

I. GİRİŞ

Otomatik işmar tanıma, insan bilgisayar etkileşiminde etkin bir alternatiftir. Ancak işmar tanıma sistemlerinin etkinliği, eğitimde kullanılan verinin kalitesine bağlıdır. Veri toplamanın ve işaretlemenin zorluğu, verinin kalitesi, değişik kullanıcı ve ortamlardan elde edilen verilerin birbiriyle uyumsuzluğu, bu tarz sistemlerin geliştirilmesini güç kılmaktadır. Bu nedenle işmar tanıma alanında yapılan çalışmalar dağarcık boyutu, ortam

bağımsızlığı ve işaretçiden bağımsızlık gibi etkenlere hassas olmakta; yeni kullanıcılar ve yeni ortamlarda yapılan canlı deneylerde de başarısız olmaktadır.

Görsel algılayıcılar ile kaydedilen işmarlar, uzaysal ve zamansal düzleme yayılmış el gezintileri olarak ifade edilebilmektedir. Bu gezintilerin modellenmesinde Saklı Markov Modelleri (Hidden Markov Models - HMM) yaygın olarak kullanılmaktadır [1].

SMM'lerin işmar tanıma için kullanımında çeşitli zorluklar bulunmaktadır [2]. İşmar tanıma modellerini eğitmek için gerekli veri miktarı ve veri standartlaştırma zorluklarından dolayı, eğitilen modeller ancak kısıtlı ortamlar ve kullanıcılarla tanıma yapmaya olanak sağlamaktadır. Özellikle canlı demolarda karşılaşılan yeni bir ortam veya yeni bir kullanıcıyla tanıma yapma sorunsalı, kaynak alana ait az sayıda örnekle eğitime olanak sağlayan alan uyarlama yöntemlerinin kullanımını teşvik etmektedir.

Transfer öğrenimi, bir problemin çözümünde kullanılan bilginin, değişik ama benzer bir problemde kullanılmasıdır. Transfer öğrenimi çerçevesinde, makine öğrenimi yapılacak olan işler ve alanlar (öznitelik uzayları ve/veya dağılımları), yapılacak olan transferin yönüne göre gruplanmaktadır. Aktarımın yapılacağı veri ve alan kümelerine kaynak (D_K), aktarımın yapıldığı kümelere ise hedef (D_H) adı verilmektedir.

Farklı alanlardaki öznitelikleri aynı işi yapmak için kullanan transdüktif alan uyarlama yöntemleri; bir alan için geliştirilen öğrenme yöntemlerinin değişik ama ilgili alanlara uygulanabilmesine olanak sağlamaktadır. Transdüktif transfer öğrenme yöntemlerine, öznitelik artımı [3], Multitask LDA [4] veya Tradaboost [5] çalışmaları örnek olarak gösterilebilir. Bu çalışmalardan Multitask LDA ve Öznitelik Artımı yöntemleri, özniteliklerin ortak bir öznitelik uzayına dönüşümünü sağlamaktadır. Tradaboost yönteminde ise kaynak kümesinin hedef küme sınıflandırmada başarılı olan örnekleri seçilmekte ve sınıflandırma yapılırken bu örneklerin ağırlığı artırılmaktadır.

İşmar tanıma için alan uyarlama yöntemlerinin kullanımına ilk örnek Fahradi v.d. [6] tarafından işaret dili tanıma için yapılan çalışmadır. Çalışmasında sentetik ve gerçek verilerden ortak öznitelikler bulan Fahradi, sentetik videolar kullanarak eğittiği sistem ile gerçek videolarda vücut hareketlerini tanımlamıştır. Benzer bir çalışma Wu v.d. [7] tarafından gerçekleştirilmiş, ve çalışmada derinlik imgelerinde en büyük katsayı ilintisi (Maximum Correlation Coefficient) yöntemiyle benzer işmarların alt birimlerinden tek seferde tanıma için yeni modeller oluşturulmuştur.

Bu çalışmada, SMM tabanlı en aza indirgenmiş sayıda örnekle eğitilmiş el işareti tanıma sistemlerinin, farklı alanlara ait

verileri kullanan alan uyarlama yöntemi ile iyileştirilmesi önerilmektedir. Bölüm 2’de kullanılan alan uyarlama yöntemi anlatılmakta, Bölüm 3’te kullanılan kaynak ve hedef veri kümeleri tanımlanmaktadır. Bölüm 4’de deney kurulumu ve sonuçları aktarılmaktadır. Son olarak Bölüm 5’te yaptığımız çalışma özetlenmekte ve gelecekte yapılması planlanan çalışmalara değinilmektedir.

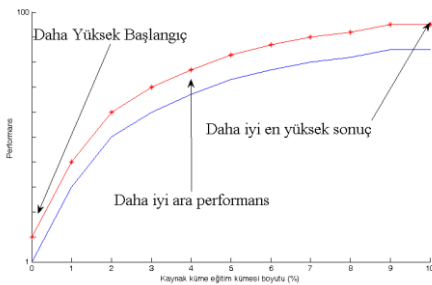
II. YÖNTEM

Alan uyarlama probleminde öznitelikler hedef ve kaynak, x_H^{d1} ve x_K^{d2} , olarak iki ayrı grupta tanımlanmaktadır. Bu öznitelikler benzer öznitelik uzayında oldukları halde farklı $p(x_H)$ ve $p(x_K)$ dağılımlarına sahip oldukları için, klasik yapay öğrenme yöntemleri onları ayırtmakta yetersiz kalmaktadırlar. Bu problemin çözümü için, çalışmamızda Hal Daume [3] tarafından önerilmiş olan öznitelik artımı (feature augmentation) yöntemini kullanılmaktadır.

Öznitelik artımı yönteminde ana fikir, değişik kaynaklardan aynı iş için elde edilen özniteliklerin beraberce sınıflandırma için kullanılabilmesi için bir öznitelik uzayına aktarılmasıdır. Bu yöntemde X_H^d ve X_K^d öznitelikleri kopyalanıp çoğaltılarak yeni ve daha çok boyutlu bir uzaya aktarılmaktadır. Yeni alan uyarlanmış uzay $X_{AUHK}^{3d} = \langle X_H^d, X_K^d, X_{Ortak}^d \rangle$ olarak; hedef veri kümesinden gelen bir örnek $\langle X_H^d, 0, X_H^d \rangle$, kaynak veri kümesinden gelen bir örnek ise $\langle 0, X_K^d, X_K^d \rangle$ olacak şekilde temsil edilmektedir. Bu basit değişiklikle elde edilmesi amaçlanan, değişik kaynaklardan gelen değişik örneklere klasik sınıflandırıcılarda değişik ağırlıklar verilebilmektedir.

Alan uyarlama yöntemini, benzer alanlar yapay öğrenme performansını her durumda arttırmamaktadır. Yapılan uyarlamanın başarısı, transfer edilen alanların benzerliğine bağlı olarak değişmektedir. Birbirine alakasız alanlar arasında yapılan transfer, performans üzerine bir etki sağlamamakta veya tanıma başarısını düşürebilmektedir. Negatif transferi önlemek adına, transfer işlemi yapılacak alanların birbiriyle ilgili olması tercih edilmektedir.

Alan uyarlama problemlerinde başarı ölçümü, kaynak kümesinin değişik oranda elemanlarıyla eğitilen sistemin başarısı karşılaştırılarak değerlendirilmektedir. Şekil 1’de görülen sistemde az sayıda örnekle sağlanan başarı, sistemin performans-örnek grafiğinde daha yüksek bir başlangıç yapmasına olanak sağlamaktadır.



Şekil 1. Alan uyarlama yöntemlerinin yapay öğrenme yöntemlerini geliştirebileceği yollar

III. KULLANILAN VERİ KÜMELERİ

Bu çalışmada kaynak ve hedef olmak üzere iki veri kümesi kullanılmıştır. Hedef veri kümesi, Keskin’in [8] hazırladığı, kullanıcıların elleriyle havaya rakamları çizdikleri derinlik imgeleridir. Derinlik algılayıcıları [9] aracılığıyla kaydedilen veri kümesi, kullanıcıların rakamları çizdiği süreçte, kameranın algıladığı her bir çerçevedeki kullanıcıların eklem noktalarının üç boyutlu koordinatlarını barındırmaktadır. Veri kümesi sağ veya sol ellerini kullanan 13 farklı kullanıcının her bir rakamı 10’ar defa havaya çizdiği örneklerden oluşmaktadır. Veri kümesindeki rakam gezintilerinden örnekler Şekil 2’de incelenebilir.



Şekil 2. İşmar gezintilerinden örnekler

Kaynak veri kümesi olarak Campos v.d tarafından hazırlanmış Chars74k [10] kullanılmıştır. Elektronik kalemle yüzey üzerine yazılarak gezintileri kaydedilen veri kümesinde her rakam için 55 örnek bulunmaktadır.



Şekil 3. Chars74k veri kümesi rakamlarından örnekler

Her iki veri kümesinin de aynı uzayda olması amacıyla iki buçuk boyutlu (RGB-D) hedef veri kümesinin iki boyuttaki iz düşümleri alınmıştır.

Veri kümeleri üzerinde yapılan ilk işlem, farklı hızlarda çizilmiş rakamların gezintinin uzunluklarının normalleştirilip eşitlenmesidir. Bu amaçla, dizilerdeki gezintinin koordinatları kübik yöntemle aradeğerlenerek (interpolate) 20, 50 ve 100 olmak üzere üç farklı boyda eşitlenmiştir. Üç farklı uzunlukta veri kümesi oluşturulmasının nedeni SMM’lerde gözlem serisi uzunluğunun tanıma başarımına etkisinin gözlenmek istenmesidir.

Son olarak iki el koordinatları arasındaki değişim açısı bulunup, bu açının değerlendirilmesi yapılarak, el koordinatları SMM’nin kullanacağı gözlem sembollerine dönüştürülmektedir. Şekil 4’te gezintinin koordinatları arasındaki değişim okları aracılığıyla gösterilmektedir.



Şekil 4. Gezinge koordinatlarının nicelenmesi

IV. DENEYLER VE SONUÇLAR

A. Deneyler

Çalışmada yapılan alan uyarlamanın başarımını sınamak için 4 farklı SMM eğitim kümesi kurulumu ile deneyler yapılmıştır. Eğitim kümeleri kurulumları sırası ile: Kaynak veri kümesi, Hedef veri kümesi, Birleştirilmiş Kaynak ve Hedef veri kümesi ve Alan uyarlanmış Kaynak ve Hedef veri kümeleridir. Tüm kurulumlarda, eğitilen SMM'ler hedef veri kümesinden ayrıştırılmış deneme örnekleri kullanılarak denenmiştir.

Hedef Veri Kümesi ile SMM Eğitimi (D_H): Bu kurulumda eğitim ve deneme kümeleri sadece hedef veri kümesindeki örneklerden oluşturulmuştur. Hedef veri kümesindeki kullanıcıların örnekleri rassal bir biçimde, eğitim kümesinde her rakamın örneklerinin %60'ı ve deneme kümesinde her rakamın geri kalan %40 örneği olacak şekilde ayrıştırılmıştır.

Kaynak Veri Kümesi ile SMM Eğitimi (D_K): Bu kurulumda eğitim kümesi kaynak veri kümesinin, deneme kümesi ise hedef veri kümesinin tüm elemanlarından oluşmaktadır.

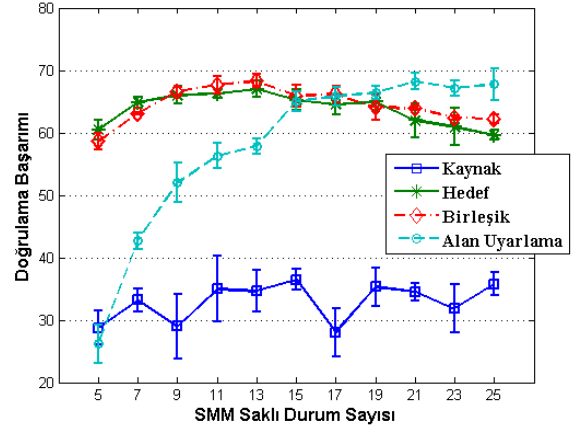
Birleştirilmiş Hedef ve Kaynak Veri Kümeleri ile SMM Eğitimi (D_{BHK}): Bu kurulumda da hedef veri kümesi eğitim ve deneme kümelerine Hedef Veri Kümesi (D_H) kurulumunda izlenen yöntem ile ayrıştırılmıştır. Ek olarak eğitim kümesine kaynak veri kümesindeki örnekler eklenerek eğitim kümesi genişletilmiştir.

Alan Uyarlanmış Hedef ve Kaynak Veri Kümeleri ile SMM Eğitimi (D_{AUHK}): Bu kurulumda kullanılan eğitim ve deneme kümelerine ayrıştırma yöntemi Birleştirilmiş Hedef ve Kaynak Veri Kümeleri (D_{BHK}) kurulumu ile aynıdır. Fakat Hedef ve Kaynak veri kümeleri eğitim ve deneme kümelerine ayrıştırılmadan önce öznitelik artırımı yöntemine tabi tutulmaktadır.

Tüm kurulumlarda gözlemlerin Gauss Karışımlarıyla Modellendiği Saklı Markov Modelleri kullanılmıştır. Saklı Markov Modellerinin en iyi parametrelerinin bulunması için çeşitli durum sayıları ($N = \{1,2,3,\dots,25\}$), gezinme boyutları ($OL = \{20,30,50\}$) ve karışımlar ($M = \{1,2,3,5,7,9,11,13\}$) kullanılarak denemeler yapılmıştır. Şekil 5'te görselleştirilmiş yapılan deneylerin istatistiksel önemini gözlemleyebilmek adına tüm deneyler 5 defa tekrarlanıp sonuçların ortalama ve standart sapma değerleri hesaplanmıştır.

Yapılan deneylerin sonuçlarının hedef veri kümesini oluşturan kullanıcıların özelliklerinden bağımsız olması adına, hedef veri kümesinin eğitim kümesinde kullanıldığı kurulumlarda deneyler, her seferinde bir kullanıcıyı dışarda bırakacak çapraz doğrulamayla yapılmıştır.

Bu çalışma kapsamında yapılan deneylerde Murphy'nin geliştirdiği Saklı Markov Modelleri kütüphanesi kullanılmıştır [11].



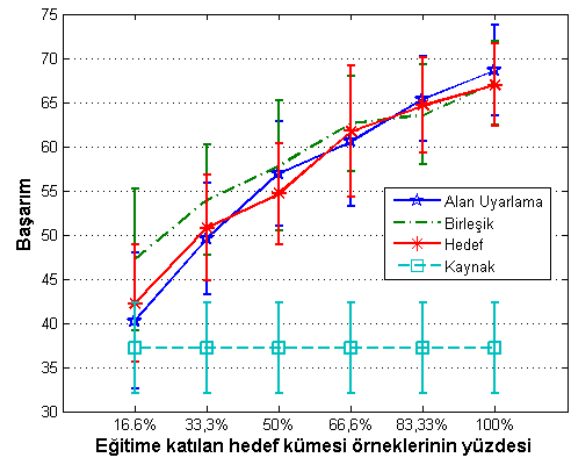
Şekil 5. Saklı Durum Sayısı belirleme için parametre eniyileme

B. Sonuçlar

Çalışmada gezinme tanıma işleminde alan uyarlama performansını analiz etmeyi amaçlayan bir deney kurulumu tasarlanmıştır. Bu kurulumda yapılan her testte kaynak veri kümesinden eğitim kümesine eklenen örnek miktarı sabit tutulurken hedef kümesinden eğitime katılan örnek miktarı değiştirilmiştir. Bu şekilde yapılan testlerde, eğitim kümesi örnek miktarının, eğitim kümesi deneme örneklerinde gezinme tanıma performansına etkisi araştırılmıştır. Tüm sınıflar üzerinde yapılan deneyler sonucunda Tablo 1'de görülen ve Şekil 6'da görselleştirilen sonuçlar elde edilmiştir.

#Örnek	D_{AUHK}	D_{BHK}	D_H	D_K
16,6 %	40,23	7,7	47,19	8,0
33,3 %	49,54	6,3	53,92	6,3
50,0 %	56,92	5,9	57,81	7,4
66,6 %	60,58	7,4	62,58	5,4
83,3 %	65,38	4,8	63,58	5,7
100 %	68,62	5,2	67,18	4,8

Tablo 1. Örnek Sayısı ve Tanıma Başarımı Analizi



Şekil 6. Örnek Sayısı ve Tanıma Başarımı Analizi

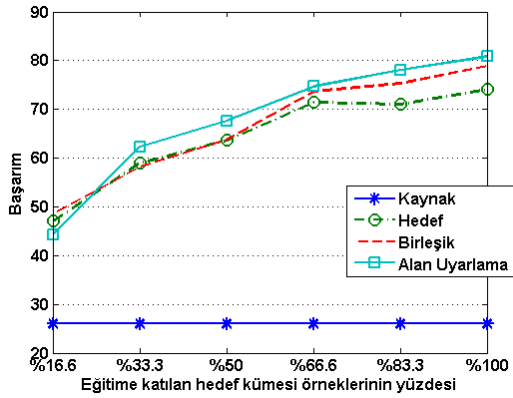
Eğitimde kullanılan hedef veri kümesine ait örnek sayısı ve gezeğe tanıma başarısını karşılaştıran bu testlerde, kaynak kümeden eğitime katılan örnek sayısı arttıkça, alan uyarlama başarısının uyarlama olmayan yöntemlerin başarısını geçtiği görülmektedir.

Ancak alan adaptasyonundan sağlanan başarı, kaynak ve hedef kümenin farklılıkları dolayısıyla düşük örnek sayılarında sadece hedef kümeyle ait testlere göre düşüktür. Bu performans düşüklüğü negatif transferin varlığına işaret etmektedir. Özellikle el yazısı ve el işareti gezinimleri arasındaki uyumsuzluğun, bu performans düşüşüne neden olduğu düşünüldüğünden hedef ve kaynak kümede tüm sınıflar arası modelleme yerine transfer yapılacak sınıfların transfer performansına göre seçilmesi düşünülmüştür. Transferin en iyi sonuç vermesi beklenen 2, 3, 4 ve 6 sınıflarında yapılan testlerin sonuçları Tablo 2’de ve Şekil 7’de verilmiştir.

Bu sonuçlardan, sadece iki alana ait veri kümelerinde örnekleri benzerlik gösteren 2, 3, 4 ve 6 işaretleriye yapılacak transferin, tüm rakamların transferine kıyasla daha iyi alan uyumlama performansı verdiği görülmektedir.

#Örnek	D_{AUHK}	D_{BHK}	D_H	D_K
16,6 %	44,23	48,65	47,02	26,00
33,3 %	62,31	58,17	58,94	26,00
50,0 %	67,60	63,75	63,65	26,00
66,6 %	74,71	73,65	71,44	26,00
83,3 %	78,08	75,29	71,06	26,00
100 %	80,87	78,85	74,04	26,00

Tablo 2. Kaynak ve hedef kümenin benzer gezinimlere sahip sınıflar için transfer sonuçları



Şekil 7. Kaynak ve hedef kümenin benzer gezinimlere sahip sınıflar için transfer sonuçları

V. VARGILAR

Bu çalışmada saklı markov modelleri kullanılarak Keskin’in [8] hazırladığı veri kümesi üzerinde işmar tanıma deneyleri gerçekleştirilmiştir. Problem, alan uyarlama yöntemi kullanılarak yeniden ele alınmış, eğitim kümesi benzer örnekler içeren el yazısı tanıma verileriyle [10] öznelik artımı yöntemi [3] kullanılarak güçlendirilmiştir. Yapılan çalışmalarla öznelik artımı yönteminin, saklı markov modelleri ile işmar tanımak için uygulanabilir olduğu gösterilmiştir.

Eğitim kümesinin en büyük olduğu durumlarda, öznelik artımı yöntemi genel olarak uyumlaştırma olmayan yöntemlerden yüksek başarı sağlamıştır. Çeşitli parametrelere sahip saklı markov modelleri kullanılarak gerçekleştirilen deneylerde en yüksek başarı, kaynak veri kümesi kurulduğunda (D_K) 37%, hedef veri kümesi kurulduğunda (D_H) 67%, öznelik artımlı alan uyarlama kurulduğunda (D_{AUHK}) 68% olmuştur. Öznelik artımı yöntemi, hedef veri kümesinin tamamıyla yapılan testlerde hedef kümeden daha iyi sonuçlar vermiştir.

Eğitim kümesinin eleman sayısı azaldıkça, iki değişik kaynaktaki gezinimler arasındaki farklar, negatif transfer ortaya çıkmasına neden olmuştur. Bu durumda “2”, “3”, “4” ve “6” rakamları için oluşturulan, uyarlanmış modeller, transfer olmayan modellerden daha yüksek sınıf tanıma başarısı elde etmiştir. Bu durum, benzer kaynaklardan yapılan transferin, uyuma olmayan alanlardan yapılan transfere göre daha faydalı olduğunu göstermektedir.

Bu çalışmanın, el işaretleri tanıma sistemleri geliştirmek için bundan sonra yapılacak çalışmalarımıza yol gösterici olması amaçlanmıştır. Bundan sonra yapılacak çalışmalarda mevcut transfer yöntemleri çeşitlendirilecek, değişik kullanıcıları alan uyarlama probleminde değişik kaynaklar olarak alıp kullanıcı bağımsızlığı test etmeye yönelinecek ve sistem vücuda göre el pozisyonu, el şekli ve üst vücut pozisyonu gibi daha karmaşık öznelik kümeleri ve işaretleri kullanılarak test edilecektir.

KAYNAKÇA

- [1] Suarez, J.; Murphy, R.R., "Hand gesture recognition with depth images: A review," RO-MAN, 2012 IEEE, pp.411,417, 9-13 Sept. 2012
- [2] Aran, O. (2008). Vision Based Sign Language Recognition: Modeling and Recognizing Isolated Signs With Manual and Non-manual Components. Doktora Tezi. Bogazici University.
- [3] III. Daume, Hal. (2007). Frustratingly easy domain adaptation. ACL, (June), 256–263.
- [4] Zhang, Y., & Yeung, D. Y. (2011). Multi-Task Learning in Heterogeneous Feature Spaces. In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence.
- [5] Dai, W., Yang, Q., Xue, G. R., & Yu, Y. (2007). Boosting for transfer learning. In Proceedings of the 24th international conference on Machine learning (pp. 193–200). ACM.
- [6] Farhadi, A., Forsyth, D., & White, R. (2007). Transfer Learning in Sign language. Word Journal Of The International Linguistic Association.
- [7] Wu, D., Zhu, F., & Shao, L. (2012). One shot learning gesture recognition from RGBD images. In 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (pp. 7–12). IEEE.
- [8] Cem Keskin, Ali Taylan Cemgil, L. A. (2011). DTW Based Clustering to Improve Hand Gesture Recognition. In A. A. Salah (Ed.), Human Behavior Understanding (pp. 72–81). Springer Berlin Heidelberg.
- [9] Microsoft Kinect, URL: "http://www.microsoft.com/en-us/kinect-for-windows/"
- [10] T.E. de Campos, B.R. Babu and M. Varma, "Character recognition in natural images.", In VISAPP, Lisbon, Portugal, February 2009.
- [11] Kevin Murphy, Hidden Markov Model (HMM) Toolbox for Matlab.