

İşaret Dili Genel:

İşaret dili, sıklıkla işitme veya konuşma engelli kişilerin iletişim kurmak için kullandığı görsel bir dildir. Sanılanın aksine işaret dilinin konuşma dili ile benzerliği olmayabilir çünkü işaret dilikonuşma dilinden bağımsız olarak yalnızca işaret dilini kullananlar tarafından geliştirilmiştir[1].

Konuşma dillerinde olduğu gibi işaret dilleri de ülkeden ülkeye değişiklik göstermektedir. Bir ülkenin işaret dili ile konuşma dilinin dil bilgisi, cümle yapısı ve kullandığı argolar farklılıklar

göstermektedir[1]. Ayrıca işaret dili kullanımının bir ülkedeki bölgelere göre değişiklik gösterdiği de gözlenmektedir, tıpkı konuşma dilinde şive farklılıkları olduğu gibi[2]. Konuşma diline göre temel fark ise işaret dilinin görsel bir dil olmasında yatmaktadır. Görsel bir işaret dili ise temel olarak şubileşenlerden oluşmaktadır; el şekli, elin vücuda göre pozisyonu, yapılan parmak, el, bilek, kol, vücut hareketleri ve yüz ifadeleri[3]. İşaret dili tanıma problemi, bilgisayarlı görü perspektifinden yaklaşıncı oldukça karmaşık bir problemdir. Bunun nedeni yetersiz veri kümelerinin varlığının yanı sıra eldeki veri kümelerinde yer alan arka plan kirliliği ve ilgili kamera açısının sebep olduğu tanıma zorluklarıdır. Ayrıca bir işaret dilini cümle bazında tanımak ise belli bir sayıda işareti teker teker tespit etmek ve bu işaretleri anlamlandırıp konuşma diline uygun bir yapıya çevirmeyi gerektirmektedir. Bunların dışında konuşma dilinde olduğu gibi işaret dilinde de yer alan bileşenler önceki veya sonraki işaretlere göre birden çok anlam içerebilmektedir[4].

İşaret dili görsel bir dildir ve yapılan eylemlerin tanınarak yazı diline çevrilmesinin gerçekleşmesi amaçlanmaktadır. Yazı diline çevrilme süreci yalnızca kelimelerin tanınması veya cümle bazında tanınma olmak üzere ikiye ayrılabilir. Planlanan bu projede veri kümesi üzerinde veya verilen test görüntüleri üzerinde kelimelerin tanınması gerçekleşmek istenmektedir. Derin öğrenme metotları hareket tanıma problemi için uygulanmaktadır[1, 2, 3]. Halihazırda temel olarak kullanılan yöntemler şu şekilde sıralanabilir: Two-Stream[2], Improved Dense Trajectories[4], Convolutional Neural Networks(CNN) ve Long Short Term Memory(LSTM) yöntemlerinin birleştirilmesi[3] ve ActionVLAD[5]. Günümüzde sıklıkla kullanılan işaret dili veri kümeleri ise şunlardır: RWTH-PHOENIX-Weather(Alman İşaret Dili Veri Kümesi)[6], SIGNUM(Alman İşaret Dili Veri Kümesi)[8], Boston ASL LVD(Amerikan İşaret Dili Veri Kümesi)[9, 10, 11], RWTH-BOSTON-104(Amerikan İşaret Dili Veri Kümesi)[12]. Bu proje kapsamında gözetimli bir derin öğrenme yöntemi kullanılarak işaret dili videoları üzerinde sonuç analizleri yapılması planlanmaktadır.

Kaynaklar

1. Sigurdsson, G. A., Russakovsky, O., & Gupta, A. (2017). What Actions are Needed for Understanding Human Actions in Videos?. arXiv preprint arXiv:1708.02696. 8/10
2. K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, pages 568–576, 2014
3. Yue-Hei Ng, J., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., & Toderici, G. (2015). Beyond short snippets: Deep networks for video classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4694-4702).
4. H. Wang and C. Schmid. Action recognition with improved trajectories. In ICCV, 2013.
5. R. Girdhar, D. Ramanan, A. Gupta, J. Sivic, and B. Russell. Actionvlad: Learning spatio-temporal aggregation for action classification. In CVPR , 2017
6. Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J. H., & Ney, H. (2012, May). RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and

Translation Corpus. In LREC (pp. 3785-3789).

7. The 20BN-jester Dataset. [ONLINE] Available at: <https://www.twentybn.com/datasets/jester>. [Accessed 15 October 2017].

8. Koller, O., Forster, J., & Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141, 108-125.

9. Carol Neidle, Ashwin Thangali and Stan Sclaroff [2012] "Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus," 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012, Istanbul, Turkey, May 27, 2012.

10. Carol Neidle and Christian Vogler [2012] "A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface," Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012, Istanbul, Turkey.

11. Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Yuan, Q., & Thangali, A. (2008, June). The american sign language lexicon video dataset. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on* (pp. 1-8). IEEE.

12. P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney. Speech Recognition Techniques for a Sign Language Recognition System. In *Interspeech*, pages 2513-2516, Antwerp, Belgium, August 2007. ISCA best student paper award Interspeech 2007