ISSN PRINT 2319 1775 Online 2320 7876

Research paper **Volume 11, Sp.Iss 6 , 2022**

Recognizing sign language from many camera angles using 2D video skeleton data

Ch.Raghava Prasad

Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Deemed to be University, Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India chrp@kluniversity.in.

DOI : 10.48047/IJFANS/11/S6/004

Abstract

In this study, we suggest a method for view-oriented feature fusion (VOFF) in multi stream CNNs. Here, we use nine different perspectives to train a CNN model. Each of the nine perspectives can be further broken down into thirds based on the camera's position relative to the action: middle, far left, and far right. Following their training by the dense networks, these three subsets are fused together using a common set of features. Once all 3 softmax layers' scores have been accumulated, a prediction is made. As demonstrated by the results, a strong discriminative view feature vector may be constructed by fusing spatial features across different viewpoints. This fusion technique generates good view feature distribution but fails to distinguish between signs that are visually identical across several views. To address the aforementioned difficulty, researchers have investigated using a contrastive network with triple loss embedding (CNTLE). Within this framework, the viewpoints are coupled as a support set consisting of pro-class and anti-class perspectives. During training, CNN networks are subjected to global cross entropy losses and view-specific triplet losses. The model's shortcomings have been mitigated thanks to the solution, which pairs perceptions of inter-class homogeneous physical appearance with unfavorable evaluations. Because of this, the model was able to provide view invariant features for classification that were satisfactory.

1.Introduction

The power to communicate visually through sign language has been a game-changer for the deaf community. It's a visual language, so you have to use your eyes instead of your ears to understand what's being said. The use of sign language can be traced back to the 17th century. throughout 70 million people throughout the world rely on sign language as their major means of communication right now. The number of Indians who rely on sign language to communicate is estimated at three million. One person does the talking and the other person does the listening in speech communication. However, with sign language, the speaker postures (gestures) with their hands, body, and face, and the listener visualizes what they

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved, Volume 11, Sp.Iss 6 , 2022**

mean. The listener's mind creates meaning from these hand gestures based on their position in relation to the speaker's body and face. Humanity has successfully developed solutions to the first basic challenge of the hearing-impaired: learning sign language. The second most difficult challenge is interacting with "normal" people, who rely mostly on verbal exchanges. Human interpreters have been the go-to remedy for the second issue before the advent of widespread digitalization and widespread internet access. Human interpreters are native English speakers who have studied sign language. An individual with this skill becomes a multilingual translator, able to interpret spoken information into visual cues and vice versa. Human knowledge shared with others in the form of a service will, at some point, need to be compensated for. Because of this, hearing-impaired people face even more challenges than they had before, which limits their potential for personal development.

The purpose of this work is to investigate the potential for identifying 2D skeletal video data in a multi view system. K.L.University's Biomechanics and vision computing research center is the source of the 3D motion capture data used to create the 2D skeletal video data for sign language. The upper body was represented in 3D by data captured from a motion capture system's 57-joint skeleton. Research into the Indian Sign Language has shown that 57 joints may account for 97% of the signs used there. The goal of this study is to train a deep learning model to recognize multi-view sign language from 2D skeletal video. Figure.1 displays the processed skeletal images.

Fig. 1: Camera Positions used in the entire thesis for capturing skeletal sign language across 15 views

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 6 , 2022**

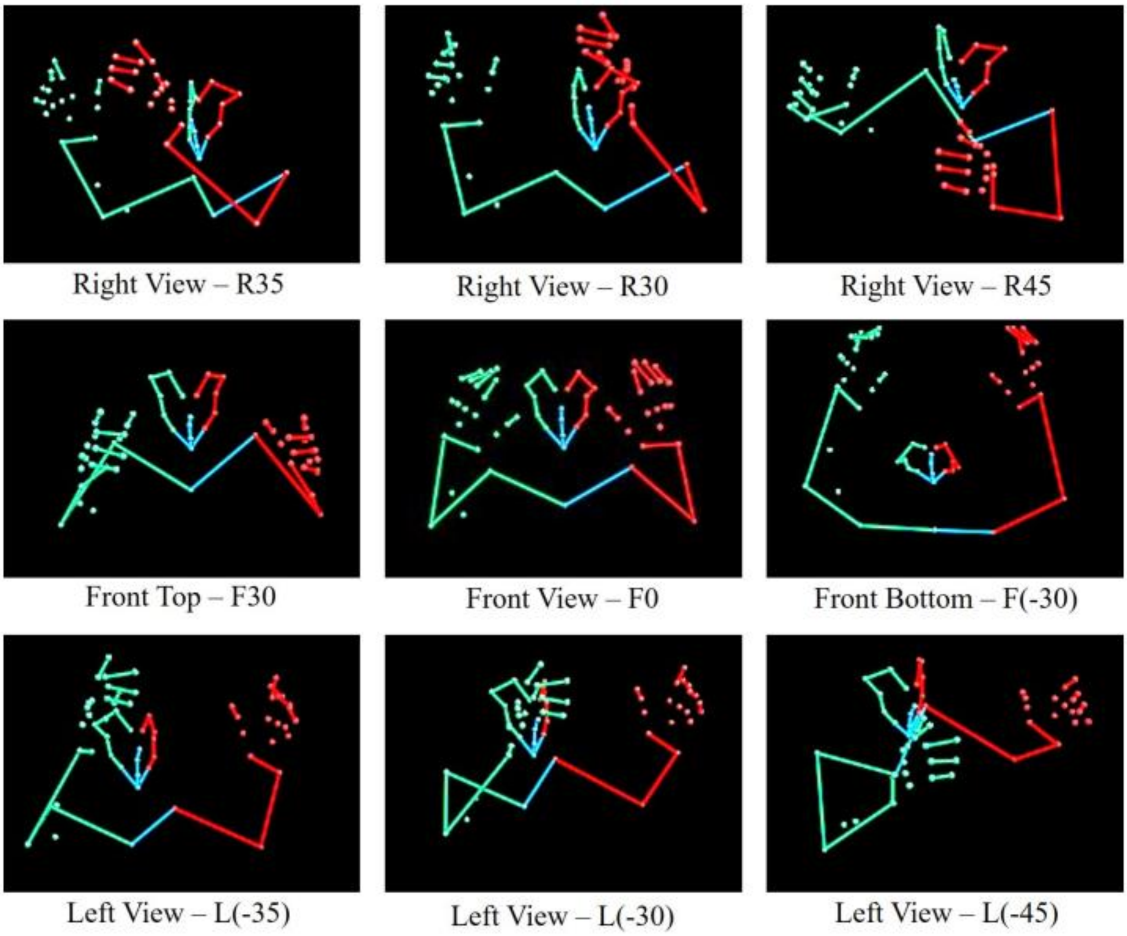

Fig. 2: Captured Training Frames in 9 views of the sign 'Basket Ball' in the KLSkel2DVidSL dataset.

Figure.2 displays the training video frames collected from these 9 angles. Due to the presence of self-occlusions from hands during the signature process, it is evident that a multi view processing strategy is required.

2.CNN Design with Multiple Views

After much deliberation and attempting several permutations of layers, we have settled on the experimental network architecture depicted in figure.3. In Figure.3, we can see a 9-stream shallow CNN model being used to learn view-specific characteristics from the 9-sequences of skeletal video used in the training phase.

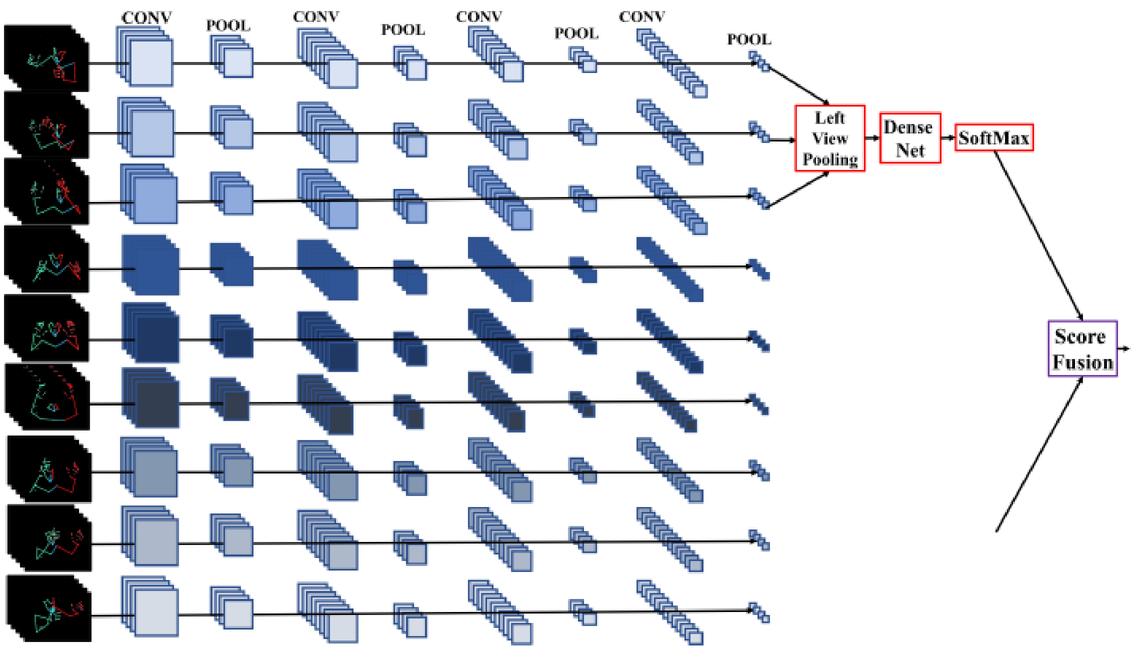

Fig. 3: Proposed CNN architecture for multi view 2D skeletal sign language recognition.

3.conclusion

The suggested model outperforms competing multi view CNN techniques on action datasets collected under a wide range of situations. The polling network, which pools features in categories according to the view profiling, is responsible for the improved performance of the proposed multi view CNN. A stronger feature vector was generated as a result of this procedure, one that included both shared features associated with a certain perspective and highly discriminant features that are specific to that perspective. The proposed multi view CNN is then compared in terms of computing complexity and ablation studies.

References:

1.D. Wang, W. Ouyang, W. Li, and D. Xu, “Dividing and aggregating network for multi-view action recognition,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 451–467.

2. E. P. Ijjina and K. M. Chalavadi, “Human action recognition in rgb-d videos using motion sequence information and deep learning,” Pattern Recognition, vol. 72, pp. 504–516, 2017.

3. C. Li, P. Wang, S. Wang, Y. Hou, and W. Li, “Skeleton-based action recognition using lstm and cnn,” in 2017 IEEE International conference on multimedia & expo workshops (ICMEW). IEEE, 2017, pp. 585–590.

4. F. Zhu, L. Shao, and M. Lin, "Multi-view action recognition using local similarity random forests and sensor fusion," Pattern recognition letters, vol. 34, no. 1, pp. 20–24, 2013.

5. A. Iosifidis, A. Tefas, and I. Pitas, "Multi-view action recognition based on ac tion volumes, fuzzy distances and cluster discriminant analysis," Signal Processing, vol. 93, no. 6, pp. 1445–1457, 2013.

6. Y. Yan, G. Liu, E. Ricci, and N. Sebe, "Multi-task linear discriminant analysis for multi-view action recignition," in Int'l Conference on Image Processing, 2013.

7. A. A. Chaaraoui, P. Climent-P´erez, and F. Fl´orez-Revuelta, "An efficient approach for multi-view human action recognition based on bag-of-key-poses," in Human Behavior Understanding: Third International Workshop, HBU 2012, Vilamoura, Portugal, October 7, 2012. Proceedings 3. Springer, 2012, pp. 29–40.

8. Sundereshan G., Krishna K.V., Vineela R., Vyshnavi N. (2019), 'Dual band monopole antenna for UWB applications with added GSM and bluetooth bands', International Journal of Innovative Technology and Exploring Engineering, 8(6), PP.212-216.

9. D.-O. Kim, N.-I. Jo, H.-A. Jang, and C.-Y. Kim, "Design of the Ultrawideband Antenna with a Quadruple-Band Rejection Characteristics Using a Combination of the Complementary Split Ring Resonators," Progress In Electromagnetics Research, Vol. 112, 93-107, 2011. doi:10.2528/PIER10111607.

10. Dimitris E. Anagnotou, Michael T. Chryssomallis, "Reconfigurable UWB antenna with RF-MEMs for on-demand WLAN Rejection", IEEE Transactions on Antennas and Propagation, Vol. 62, No. 2, Feb. 2014.

11. S. A. Aghdam, "Reconfigurable Antenna With a Diversity Filtering Band Feature Utilizing Active Devices for Communication Systems," in IEEE Transactions on Antennas and Propagation, vol. 61, no. 10, pp. 5223-5228, Oct. 2013.

12. N. Tasouji, J. Nourinia, C. Ghobadi and F. Tofigh, "A Novel Printed UWB Slot Antenna With Reconfigurable Band-Notch Characteristics," in IEEE Antennas and Wireless Propagation Letters, vol. 12, pp. 922-925, 2013.