Hand Gesture Recognition Under Low-Light Conditions

Alex He 贺嘉乐

Abstract

Human gesture recognition has become more and more important in recent years. However, issues such as image lighting interfere with gesture recognition. In this paper, a 3-part model involving image enhancement, keypoint extraction, and gesture recognition is proposed to alleviate the issue of insufficient lighting, and the results are given.

1. Introduction

Being a natural way for humans to express themselves, hand gestures can play a large and important role in human-to-human or human-to-computer communication [1]. This is especially true in noisy places, where speech becomes a less viable option and gestures begin to replace speech as the primary form of communication. The ability for a computer to be able to recognize these gestures in human-to-computer interactions thus becomes rather important, especially under the context of human-computer interfaces (HCI) [2].

Image lighting can interfere with the computer's ability to recognize gestures. When under insufficient lighting, images tend to have lower contrast, brightness, greater noise, and color distortion [3]. As a result, systems designed for normal lighting or high-quality conditions often have poorer performance [4, 5].

Figure 1. Example of Image with Insufficient Lighting from the LOL dataset used by RetinexNet

However, insufficient lighting is common in our daily lives [5], whether it be a lack of light at nighttime or poor indoor lighting. Thus, solving the issue of lighting in regards to hand gesture detection is an important issue.

Many models are dealing with images with poor lighting and the enhancement of said images [3, 4, 5, 6]. Similarly, many models dealing with hand gesture recognition also exist [1, 2, 7, 8]. However, few models simultaneously consider both.

This paper proposes a model combining these two aspects. It first utilizes RetinexNet to enhance low-light images [5] and conjoins it with OpenPose [9, 10, 11, 12] to gather hand keypoints. These keypoints are then fed into a long short-term memory (LSTM) to predict the hand gesture.

2. Proposed Model

2.1 RetinexNet

RetinexNet [5] is a deep network that enhances low-light images. It is based on Retinex theory, which assumes that observed images are composed of reflectance and illumination.

RetinexNet first decomposes input images into reflectance and illumination. Then, it will adjust these two parts, adjusting the illumination to make the image brighter and denoising the reflectance. Finally, it will reconstruct the image with the adjusted illumination and reflectance, resulting in an enhanced image.

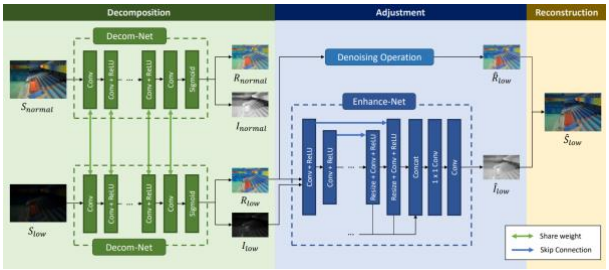

Figure 2. Retinex Framework

2.2 OpenPose

OpenPose is a real-time, multi-person keypoint detection library for the body, face, hands, and feet. The model uses OpenPose to detect only hand keypoints. The enhanced image is passed to OpenPose, which will pass the coordinates of the hand keypoints shown in Figure 3 to the LSTM.

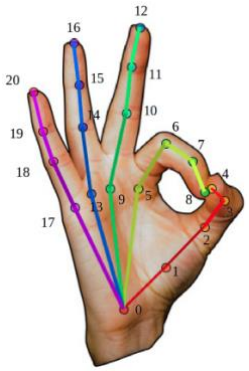

Figure 3. OpenPose hand skeleton keypoints

2.3 LSTM

The LSTM used consists of 2 LSTM layers, 2 Batch Normalization layers, and 3 Dense layers, implemented via Keras. It was trained on a dataset "Image Database for Tiny Hand Gesture Recognition", which includes 7 different hand gestures from 40 people [7].

2.4 Reasoning

This model with OpenPose and an LSTM was used over other methods such as a CNN. This is because CNNs can only give a rough positioning of a hand based on the whole image. The accuracy and precision would need to be improved. Meanwhile, OpenPose is able to extract keypoints based on a skeleton, giving high precision.

3. Experiment

3.1 Results

To evaluate the usefulness of RetinexNet, an ablation study is made. 35 images – 5 images per gesture – were taken with a dark background, and they were fed into OpenPose and the LSTM in two different ways – with and without RetinexNet applied to the image. Confusion matrices were generated, the results of which are in Table 1.

Figure 4. Example of an image used in the ablation study. The top is the original image and the bottom is the image with RetinexNet applied to it.

Gesture	No Retinex			Retinex	
Fist	29	1		30	0
	5	0		5	0
L	30	0		30	0
	5	0		5	0
OK	29	1		30	0
	5	0		5	0
Palm	17	13		22	8
	1	4		5	0
Pointer	30	0		30	0
	5	0		5	0
Thumbs Up	19	11		7	23
	0	5		1	4
Thumbs Down	30	0		30	0
	5	0		5	0

Table 1. Confusion matrices created from the ablation study. Each gesture has a corresponding confusion matrix generated using the original images and the RetinexNet-enhanced ones. The confusion matrix is in the form of true negative, false positive in the first row, false negative, true positive in the second.

3.2 Interpretation

Apart from the thumbs-up gesture, both the original images and the RetinexNet-enhanced images did roughly the same, with the RetinexNet-enhanced images performing slightly better. There is potential for this to be a viable solution.

Overall, both systems performed relatively well (not including the thumbs-up gesture), so, it's likely that there was sufficient lighting for OpenPose to handle. However, it can also be seen that even though performance was relatively well, that was because the model evaluated nearly every image as false and there were simply more false images than true ones for each gesture. A third factor unaccounted for may be causing this error.

In the thumbs-up gesture, a significant drop in accuracy can be observed in the RetinexNet-enhanced side. This was probably a result of too much image noise created when enhanced by RetinexNet – something easily observable. Image denoising techniques may improve these results.

Ultimately, in this scenario, it seems that there is potential for RetinexNet-enhanced images to outperform normal images, but further techniques such as image denoising will need to be used.

4. Conclusion

In this paper, a method to deal with hand gesture recognition under insufficient lighting was proposed. Images would be enhanced via RetinexNet, fed into OpenPose to extract hand keypoints, which will then be fed into an LSTM to perform the hand gesture recognition. Experimentally, it can be seen that the method has potential, but needs more work in aspects such as denoising to be a viable solution.

References

[1] P. Narayana, J. R. Beveridge and B. A. Draper, "Gesture Recognition: Focus on the Hands," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5235-5244, doi: 10.1109/CVPR.2018.00549.

[2] Francis, Jobin & Kadan, Anoop. (2014). Significance of Hand Gesture Recognition Systems in Vehicular Automation-A Survey. International Journal of Computer Applications. 99. 50-55. 10.5120/17389-7931.

[3] Zhang, Y., Di, X., Zhang, B., Li, Q., Yan, S., & Wang, C. (2021). Self-supervised Low Light Image Enhancement and Denoising. *arXiv [cs.CV]*. Opgehaal van http://arxiv.org/abs/2103.00832

[4] Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep Retinex Decomposition for Low-Light Enhancement. *arXiv [cs.CV]*. Opgehaal van http://arxiv.org/abs/1808.04560

[5] Patil, Akshay & Chaudhari, Tejas & Deo, Ketan & Sonawane, Kalpesh & Bora, Rupali. (2020). Low Light Image Enhancement for Dark Images. International Journal of Data Science and Analysis. 6. 99. 10.11648/j.ijdsa.20200604.11.

[6] Wei, X., Zhang, X., Wang, S., Cheng, C., Huang, Y., Yang, K., & Li, Y. (2021). BLNet: A Fast Deep Learning Framework for Low-Light Image Enhancement with Noise Removal and Color Restoration. *arXiv [eess.IV]*. Opgehaal van http://arxiv.org/abs/2106.15953

[7] P. Bao, A. I. Maqueda, C. R. del-Blanco and N. García, "Tiny hand gesture recognition without localization via a deep convolutional network," in IEEE Transactions on Consumer Electronics, vol. 63, no. 3, pp. 251-257, August 2017. doi: 10.1109/TCE.2017.014971

[8] Khan, R. Ibraheem, N. (2012). Hand Gesture Recognition: A Literature Review. International Journal of Artificial Intelligence and Applications, doi: 10.5121/IJAIA.2012.3412

[9] Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., & Sheikh, Y. A. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[10] Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). Hand Keypoint Detection in Single Images using Multiview Bootstrapping. *CVPR*.

[11] Cao, Zhe, Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *CVPR*.

[12] Wei, S.-E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. *CVPR*.