Advances in Robotics Research, *Vol. 2, No. 2 (2018) 129-140*
DOI: https://doi.org/10.12989/arr.2018.2.2.129

Robust 2D human upper-body pose estimation with fully convolutional network

Seunghee Lee[1a], Jungmo Koo[1b], Jinki Kim[1c] and Hyun Myung[*1,2]

[1]*Department of Civil and Environmental Engineering, Korean Advanced Institute for Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea*
[2]*Robotics Program, Korean Advanced Institute for Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea*

(Received April 28, 2018, Revised May 10, 2018, Accepted May 11, 2018)

Abstract. With the increasing demand for the development of human pose estimation, such as human-computer interaction and human activity recognition, there have been numerous approaches to detect the 2D poses of people in images more efficiently. Despite many years of human pose estimation research, the estimation of human poses with images remains difficult to produce satisfactory results. In this study, we propose a robust 2D human body pose estimation method using an RGB camera sensor. Our pose estimation method is efficient and cost-effective since the use of RGB camera sensor is economically beneficial compared to more commonly used high-priced sensors. For the estimation of upper-body joint positions, semantic segmentation with a fully convolutional network was exploited. From acquired RGB images, joint heatmaps accurately estimate the coordinates of the location of each joint. The network architecture was designed to learn and detect the locations of joints via the sequential prediction processing method. Our proposed method was tested and validated for efficient estimation of the human upper-body pose. The obtained results reveal the potential of a simple RGB camera sensor for human pose estimation applications.

Keywords: human pose estimation; skeleton extraction; fully convolutional network; semantic segmentation; upper-body joint segmentation

1. Introduction

Human body pose estimation is one of the most important techniques that has been studied for decades. There have been extensive efforts to efficiently estimate human body poses along with reliable skeleton extraction results. Such technology allows a higher level of human-computer interaction and the recognition of human activities for various applications (Aggarwal *et al.* 1997, Moeslund *et al.* 2006). Pose estimation is mainly aimed at recognizing the gestures of humans in action; the recognition of human gestures may be adapted for the development of body language or

*Corresponding author, Professor, E-mail: hmyung@kaist.ac.kr
[a]Ph.D. Student, E-mail: seunghee.lee@kaist.ac.kr
[b]Ph.D. Student, E-mail: jungmokoo@kaist.ac.kr
[c]Master Student, E-mail: rlawlsrl@kaist.ac.kr

 ISSN: 2287-4976 (Print), 2287-4984 (Online)

Ramanan, D. (2007), "Learning to parse images of articulated bodies", *Proceedings of the Advances in Neural Information Processing Systems* (*NIPS*), Vancouver, British Columbia, Canada.

Roweis, S.T. and Saul, L.K. (2000), "Nonlinear dimensionality reduction by locally linear embedding", *Science*, **290**(5500), 2323-2326.

Sapp, B. and Taskar, B. (2013), "Modec: Multimodal decomposable models for human pose estimation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Portland, Oregon, U.S.A., June.

Schwarz, L.A., Mkhitaryan, A., Mateus, D. and Navab, N. (2012), "Human skeleton tracking from depth data using geodesic distances and optical flow", *Image Vis. Comput.*, **30**(3), 217-226.

Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A. and Blake, A. (2013), "Efficient human pose estimation from single depth images", *IEEE Trans. Pattern Anal. Mach. Intell.*, **35**(12), 2821-2840.

Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M. and Moore, R. (2013), "Real-time human pose recognition in parts from single depth images", *Commun. ACM*, **56**(1), 116-124.

Simonyan, K. and Zisserman, A. (2014), *Very Deep Convolutional Networks for Large-Scale Image Recognition*, arXiv preprint arXiv:1409.1556.

Straka, M., Hauswiesner, S., Rüther, M. and Bischof, H. (2011), "Skeletal graph based human pose estimation in real-time", *Proceedings of the British Machine Vision Conference* (*BMVC*), Dundee, Scotland, U.K., August-September.

Takahashi, K., Uemura, T. and Ohya, J. (2000), "Neural-network-based real-time human body posture estimation", *Proceedings of the 2000 IEEE Signal Processing Society Workshop*, Lafayette, Louisiana U.S.A.

Tompson, J., Goroshin, R., Jain, A., LeCun, Y. and Bregler, C. (2015), "Efficient object localization using convolutional networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, Massachusetts, U.S.A., June.

Tompson, J.J., Jain, A., LeCun, Y. and Bregler, C. (2014), "Joint training of a convolutional network and a graphical model for human pose estimation", *Proceedings of the Advances in Neural Information Processing Systems* (*NIPS*), Montreal, Canada, December.

Toshev, A. and Szegedy, C. (2014), "DeepPose: Human pose estimation via deep neural networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Columbus, Ohio, U.S.A., June.

Wei, S.E., Ramakrishna, V., Kanade, T. and Sheikh, Y. (2016), "Convolutional pose machines", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, Nevada, U.S.A., June-July.

Xia, F., Wang, P., Chen, X. and Yuille, A. (2017), *Joint Multi-Person Pose Estimation and Semantic Part Segmentation*, arXiv preprint arXiv:1708.03383.

Yang, W., Ouyang, W., Li, H. and Wang, X. (2016), "End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, Nevada, U.S.A., June-July.

Zhang, Z., Seah, H.S., Quah, C.K. and Sun, J. (2013), "GPU-accelerated real-time tracking of full-body motion with multi-layer search", *IEEE Trans. Multimedia*, **15**(1), 106-119.

HM