

Graphs in Computer Vision then and now: how Deep Learning has reinvigorated Structural Pattern Recognition

Donatello Conte

Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT - EA 6300) - Université de Tours
64 Avenue Jean Portalis, Tours, France donatello.conte@univ-tours.fr

Abstract

Computer Vision Problems, such as object detection, object tracking, action recognition and so on, have been, in the past, usually addressed through Statistical Pattern Recognition techniques. SVM, Regression or Neural Networks, are some examples of classical statistical techniques that have been used, quite effectively, in many application contexts of computer vision.

Nevertheless, some attempts have been proposed using more complex data structures (notably graphs) for solving Computer Vision Tasks. However, in terms of performances, their use did not have the same success as techniques based on vector representations. First part of this talk will present some of these proposals, in the context of object tracking (Conte, Foggia, Jolion, and Vento (2006)), people re-identification (Mahboubi, Brun, and Conte (2018)) and action recognition (Cortés, Conte, and Cardot (2018)). An graph representation is proposed in Conte et al. (2006) to deal with occlusion problem. The representation is based on a graph pyramid, namely, each moving region is represented at different levels of resolution using a graph for each level. The algorithm compares the topmost levels of each pyramid in the association phase between moving objects in two consecutive frames. If the comparison outcome is sufficient to assign a label to each node the tracking algorithm stops. Instead, if some ambiguities arise (as it is the case when two objects overlap), the algorithm is repeated using the next levels of the pyramids, until either a consistent labelling is found. The purpose of re-identification (re-id) is to identify people coming back into the field of view of a camera or to recognize an individual through different cameras in a distributed network. At the heart of the process there is a comparison between signatures given probe and gallery sets. In Mahboubi et al. (2018) graphs are used to represent people appearance and comparison is done by means of Graph Kernels. Finally, action recognition is a classification problem in which each video representing an action has to be classified with the correct action label. In Cortés et al. (2018) we proposed to represent videos using graph sequences and proposed a model inspired from bag-of-words techniques to classify a sequence.

Recently, graphs have gained a lot of attention in the Computer Vision community thanks to the use of this kind of data within deep learning techniques. Graph Neural Networks have demonstrated their effectiveness in solving Computer Vision problems, and in some cases recent proposals have bridged the gap between statistical and structural pattern recognition. Second part of the talk will be devoted to illustrate some of these examples (Rangesh et al. (2021); Wu et al. (2020); Yan, Xiong, and Lin (2018)). Starting from the already mentioned applications in Computer Vision (object tracking, action recognition), we will discuss the new proposals based on Deep Learning with graphs and the open problems in this context.

Keywords: Computer Vision; Structural Pattern Recognition; Graph Neural Networks

References

- Conte, D., Foggia, P., Jolion, J.-M., & Vento, M. (2006). A graph-based, multi-resolution algorithm for tracking objects in presence of occlusions. *Pattern Recognition*, 39(4), 562–572.
- Cortés, X., Conte, D., & Cardot, H. (2018). Bags of graphs for human action recognition. In *Joint iapr international workshops on statistical techniques in pattern recognition (spr) and structural and syntactic pattern recognition (sspr)* (pp. 429–438).
- Mahboubi, A., Brun, L., & Conte, D. (2018). A structural approach to person re-identification problem. In *2018 24th international conference on pattern recognition (icpr)* (pp. 1616–1621).

- Rangesh, A., Maheshwari, P., Gebre, M., Mhatre, S., Ramezani, V., & Trivedi, M. M. (2021). Trackmpnn: A message passing graph neural architecture for multi-object tracking. *arXiv preprint arXiv:2101.04206*.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1), 4–24.
- Yan, S., Xiong, Y., & Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Thirty-second aai conference on artificial intelligence*.