Master of Science Topics

Human Tracking and Human Activity Recognition

Title: User detection and tracking

Coordinators: *Prof. Adina Magda Florea (adina.florea@upb.ro)*
Assist. Prof. Mihai Trascau (mihai.trascau@upb.ro)
drd. ing. Ștefania Ghiță (stefania.a.ghita@upb.ro)

Description:

People tracking, unlike other recognition and interpretation tasks, is difficult both from the point of view of recognition and prediction of the trajectory, and from the one of the identification of the ground truth. Partially visible, occluded, or cropped targets, reflections in mirrors or windows, and objects that very closely resemble targets, all impose intrinsic ambiguities, such that even humans may not agree on one particular ideal solution. Moreover, establishing evaluation metrics with free parameters and ambiguous definitions often lead to conflicting quantitative results [Lea et.al. 2017].

People detection from videos or images is subsumed by the problem of object detection. During the last years, the problem has received a lot of attention and many robust solutions exist, both based on traditional computer vision processing and based on deep neural networks. Even in this case, the problem is not yet entirely solved if we are to consider real-life situations, fast processing time and limited resources.

On the other hand, people tracking from videos is a more difficult problem and is currently a challenge. Some years ago, the trend on people detection from video sequences was to find strong, preferably optimal methods to solve the data association problem. Linking detections in a set of consistent trajectories (matching two detections based on either simple distances or weak appearance models) was solved by various methods such as DP NMS, by Conditional Random Fields or as a variational Bayesian model; performances were not very good.

Recently, the focus of people tracking from videos is on building robust pairwise similarity costs, mostly based on strong appearance cues, leading to better tracker performances and more complex scenarios. Some good approaches use sparse appearance models or integral channel feature appearance models or aggregated local flow of long-term interest point trajectories to improve detection affinity. Still, most of the available tracking approaches do not include a learning algorithm to determine the set of model parameters for a dataset. Some recent approaches tried to use deep learning, such as Recurrent Neural Networks to encode appearance, motion, and interactions or deep matching to improve the affinity measure. These approaches are rather few and results are promising, however they do not surpass other approaches for the time being.

The research topic implies to use both traditional tracking methods (e.g. Kalman filters) but also deep learning models to achieve better performances than recent approaches, and evaluating the results on the MOTT Challenge [Lea et.al. 2017].

https://towardsdatascience.com/people-tracking-using-deep-learning-5c90d43774be

[Lea et.al. 2017] Laura Leal-Taixe, Anton Milan, Konrad Schindler, Daniel Cremers, Ian Reid, Stefan Roth, (2017) Tracking the Trackers: An Analysis of the State of the Art in Multiple Object Tracking, https://arxiv.org/abs/1704.02781

Title: Human activity recognition from video sequences

Coordinators: *Prof. Adina Magda Florea (adina.florea@upb.ro)*
Assist. Prof. Mihai Trascau (mihai.trascau@upb.ro)
drd. ing. Mihai Nan (mihai.nan.cti@gmail.com)

Description:

The problem of recognizing human activity using the RGB image extracted from video is one of the first problems that computer vision has tried to solve. From the beginning, this problem was considered a challenging one, because there are many variables like that: the height of the person, the scene in which the action takes place, the brightness, the angle from which it is viewed, the fact that an action can be executed in a different manner from one person to another. The information provided by the video cameras has been extensively studied and analyzed as input for systems capable of identifying and recognizing human actions. Models applied initially for action classification used 2D images as features and classifiers such as Support Vector Machines (SVM) and Hidden Markov Models. Recently deep learning methods have emerged as a good candidate for human activity recognition.

The research topic implies developing different deep models for user activity recognition based on RGB images extracted from video sequences.

https://escholarship.org/uc/item/2mr798mn http://blog.qure.ai/notes/deep-learning-for-videos-action-recognition-review

Title: Human pose estimation

Coordinators: *dr. ing. Mihai Trăscău (mihai.trascau@upb.ro)*
drd. ing. Mihai Nan (mihai.nan.cti@gmail.com)

Description:

Human pose estimation is defined as the problem of localizing and identifying anatomical key-points of the human body and it is considered a fundamental and challenging task in Computer Vision. Poses can serve as base features in other vision problems like activity recognition, virtual and augmented reality, human reidentification or in human-computer interaction. This task can be very useful for solving the problem of recognizing human actions using a skeleton-based approach.

This research project aims to create a module capable of extracting human skeleton coordinates from RGB images. There are many challenges which need to be taken into account when inferring human pose. First, the number of people in the image is unknown and usually varies. Tackling this gives the option of either iteratively processing each person instance in the image or to attempt to obtain all human poses simultaneously. Moreover, the image may contain people who are in contact (e.g. shaking hands, carrying one another) making it difficult to assign the correct correspondence between person and key-point or body part. Occlusion, be it of actual key-points or not, also affects accuracy quite drastically.

[Zhe Cao et al.] “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”. In: CoRR abs/1812.08008 (2018). arXiv: 1812.08008. URL : http://arxiv.org/abs/1812.08008.
[Rıza Alp Güler, Natalia Neverova, and Iasonas Kokkinos.] “Densepose: Dense human pose estimation in the wild”. In: arXiv preprint arXiv:1802.00434 (2018).
[Zhe Cao et al.] “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”. In: CVPR. 2017.
[Shih-En Wei et al.] “Convolutional pose machines”. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, pp. 4724–4732.

Title: Human action prediction

Coordinators: *dr. ing. Mihai Trăscău (mihai.trascau@upb.ro)*
drd. ing. Mihai Nan (mihai.nan.cti@gmail.com)

Description:

Future human action prediction is a probabilistic process that aims to identify ongoing action from video only containing the beginning part of the action. For this task, the goal is to allow early recognition of unfinished action from temporally incomplete video data. In contrast to the task of human action recognition, human action prediction is a before-the-fact video understanding task and is focusing on the future state. This aspect makes this task very useful for many real-life scenarios where various critical situations could be avoided if they could be predicted quickly enough.

The purpose of this research project is to implement a framework for future human action prediction, integrating several modules (e.g., human pose estimation, object detection) and analyzing various approaches.