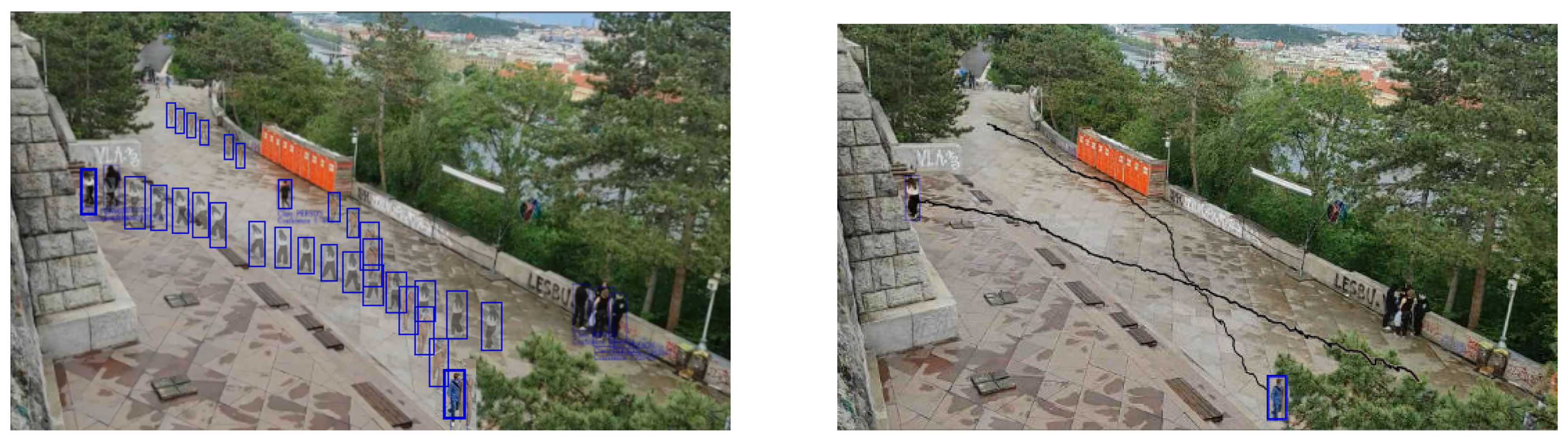
Construction of time-space trajectories from multimodal data

Mgr. Matěj Hrbáček   prof. RNDr. Tomáš Skopal, Ph.D.

Software and Data Engineering   Charles University

Abstract

With the growth of public camera recordings and video streams in recent years, there is an increasing need for automatic processing with limited human input. An essential part of the process is detecting moving objects in the video and grouping individual detections across video frames into trajectories. This thesis presents a set of algorithms for creating trajectories from object detections while using a configurable analytic model. Presented algorithms are based on the clustering of detections, later even simple trajectories into complex trajectories by their features, such as a timestamp (frame), bounding rectangle in the video frame and optionally, image crop defined by the bounding rectangle. To present the usage of the generated trajectories, we introduced in the thesis methods for further analysis and data extraction, such as detection enrichment and semantic description of trajectories.



(a) Detections of objects   (b) Motion curve

Figure 1. Transforming detections to trajectories

Realized solution

In the work, we have designed an algorithm that generates trajectories by clustering detections into trajectories driven by a fully configurable analytical model, and we have tested and proposed different clustering methods. The designed algorithm, as a data input, accepts raw video, detections of objects, and a configuration with information on how to perform clustering for certain types of camera views.

We called this approach $Traged$ (TRAjectories GEnerated from Detections). The general idea of the algorithm is to measure connectivity, as defined later, of each of the two detections and the most connectable connects into a single trajectory. The trajectories must meet real-life conditions, such as that the object is at most in one place at one moment. Thanks to this process, the algorithm creates short and incomplete trajectories, where we can repeat the clustering step as seen in the picture 2.

The algorithm, as described, would be too complex and, because of that, unusable. Thus, we have focused on designing optimisation methods and approximation strategies to reduce the complexity and make the algorithm usable.

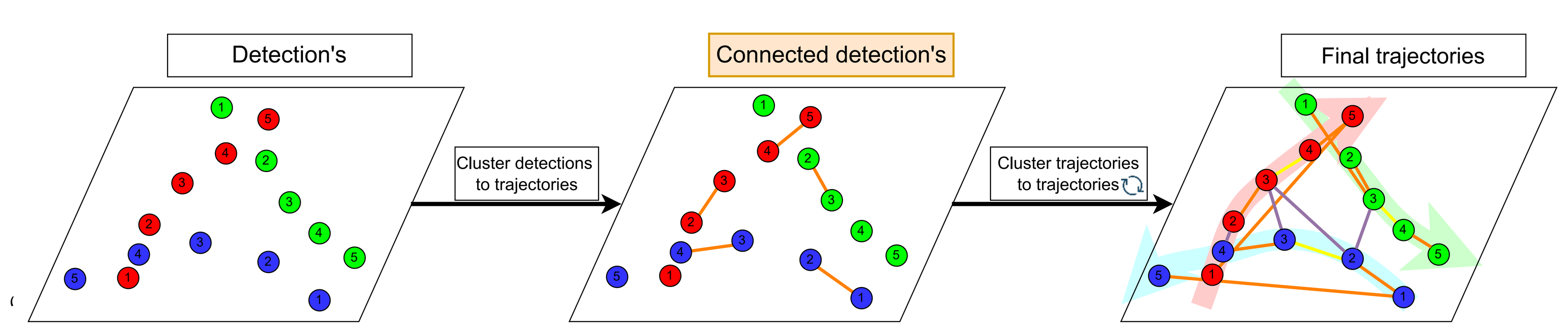


Figure 2. Idea of $Traged$'s clustering flow

The advantage of this approach is resistance to missing detections at the input and no need for a large amount of human-annotated data compared to Deep learning.

To verify the proposed solution, we have implemented the algorithm as a module into project Videolytics[3], where the generated trajectories can be seen.

Features and connectivity

A significant part of the work focused on determining connectivity *(probability that two trajectories/detections belong to the same object)*. In the work, we have analyzed various features of detections and trajectories, such as *colouration*, *size*, *speed* and more. Later, we proposed methods to measure feature connectivity and determine connectivity over various features as illustrated in figure 3.

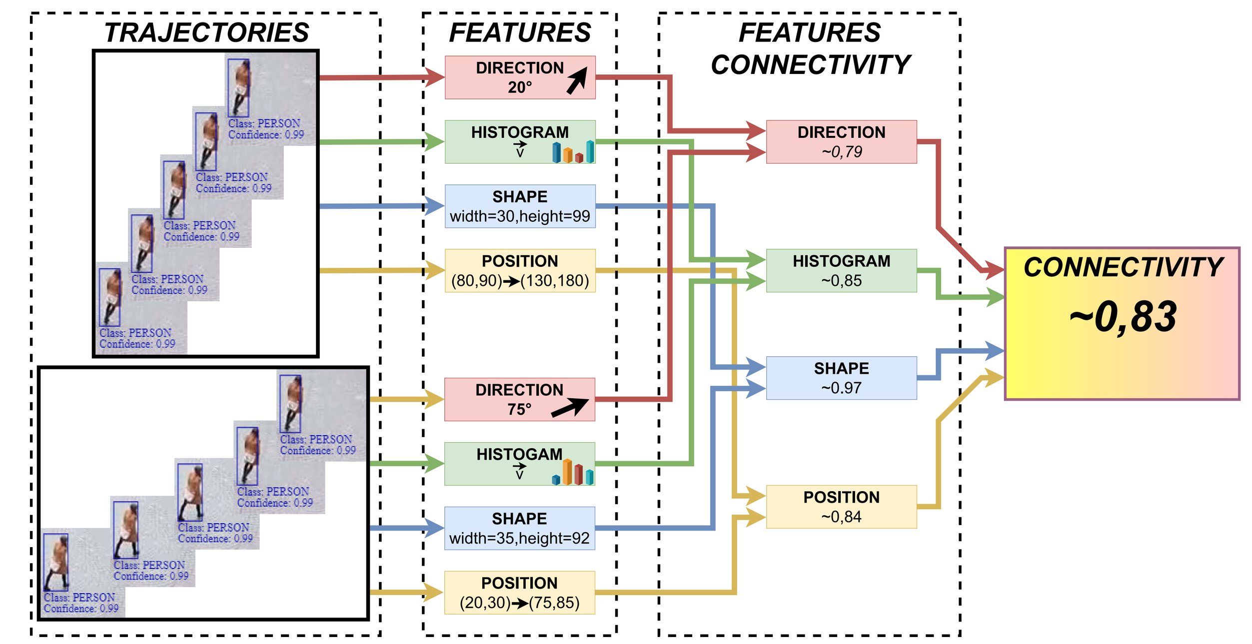
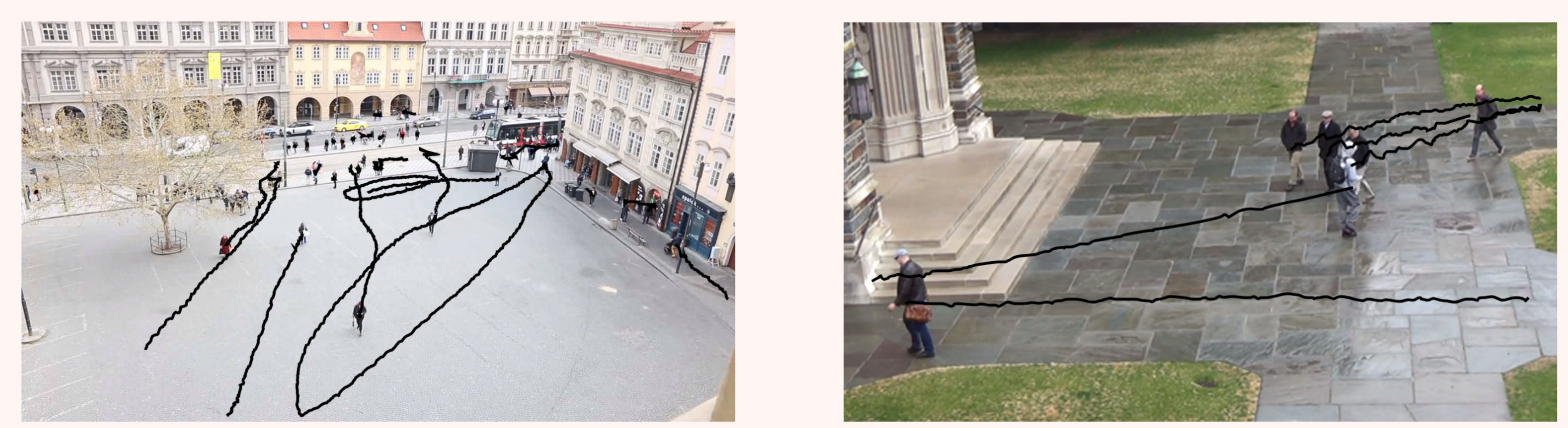


Figure 3. Illustration of connectivity measurement

Generated results

To present the output of our algorithms, we have created a screenshot from the Videolytics page with generated trajectories, as shown in the picture 4.



(a) Close up camera shot   (b) Distant camera shot

Figure 4. Generated trajectories

To demonstrate the usage of generated trajectories, we have created a visualization of a generated semantic description for a single trajectory as seen in the picture 5.

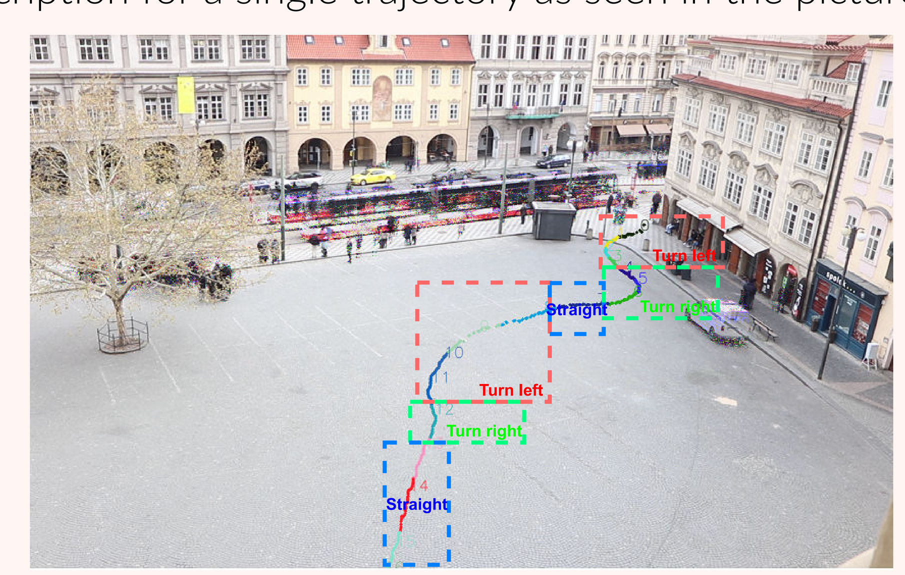


Figure 5. Semantic description of one trajectory

Evaluation

To evaluate the results of our approach, we have decided to create a comparison with OpenCV-CSRT tracker as used for trajectory generation in benchmark[1]. We have annotated a dataset of videos and implemented a set of metrics from HOTA[2] and introduced new ones. The results of the comparison over part of the dataset can be seen in table 1, where the green values represent the best and the red values represent the worst result for a video.

	duke		*kradez_stred*		*rvacka_stred*	
Metric \ Method	TRAGED	CSRT	TRAGED	CSRT	TRAGED	CSRT
TP	1.0	0.938	0.812	0.875	1.0	1.0
FP	0.0	0.25	0.667	0.188	0.48	0.0
TDF	0.0	0.125	0.188	0.125	0.0	0.0
FRAG	0.125	0.0	0.5	0.188	0.4	0.0
SWAP	0.188	0.125	0.167	0.125	0.04	0.0
SWAP-B	0.0	0.0	0.0	0.0	0.0	0.0
SWAP-A	0.188	0.125	0.167	0.125	0.04	0.0
P-AssA	0.944	0.814	0.659	0.727	0.842	0.842
P-DetA	0.994	0.589	0.962	0.64	0.918	0.603
P-HOTA	0.968	0.692	0.796	0.682	0.879	0.713

Table 1. Evaluation Traged X OpenCV-CSRT

As can be seen in the provided table 1, our method provides in some aspects better results than OpenCV-CSRT, but we have to keep in mind that our method is highly specialized on the specific types of videos, whereas OpenCV-CSRT is universal. In the thesis, we have also demonstrated the error rate of our approach in cases when trajectories were created with configuration for different types of videos.

Contribution of the work

- In the thesis, we have designed and described algorithms for video detection processing into trajectories using analytic methods.
- Analyzed features of detections (and trajectories) and researched methods to use features to determine the probability that two detections (or trajectories) belong to the same object.
- Showed usage of trajectories by generating semantic descriptions of them and, with generated trajectories, improved the input detections.
- We have implemented algorithms as a new module into project Videolytics, where generated trajectories can be used as a base for future work.
- We have evaluated algorithms implemented in Videolytics against the OpenCV-CSRT method and created comparisons with multiple metrics.

References

[1] Nadja Dardagan, Adnan Brdjanin, Džemil Džigal, and Amila Akagic. Multiple object trackers in opencv: A benchmark. 10 2021.

[2] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision*, pages 1–31, 2020.

[3] Tomáš Skopal, Dominika Ďurišková, Petr Pechman, Marek Dobranský, and Vladislav Khachaturian. Videolytics- system for data analytics of video streams. *International Conference on Information and Knowledge Management (CIKM 2021)*, pages 4794–4798, 10 2021.

http://videolytics.ms.mff.cuni.cz/stream.html   mateshrbacek@seznam.cz