SCALABLE TOPOLOGY ESTIMATION FOR LARGE CAMERA NETWORKS

Key Words video surveillance, multi-camera tracking, activity topology

Motivation Multi-camera video surveillance networks serve as a primary security measure for much of our nation's critical infrastructure facilities. The spread of this technology, however, has far outpaced the ability of human personnel to effectively monitor the information provided by the hundreds or even thousands of active camera views available for a particular airport, embassy, or other wide-area site. Consistent tracking of human or vehicle targets as they move across a facility, especially across gaps in camera coverage, remains a crucial unfulfilled need for site security. Through my research at MIT Lincoln Laboratory, I discovered automated multi-camera tracking to be an exciting blend of my interests in algorithms, networks, and empirical evaluation. Furthermore, the promise of meaningful real-world impact makes this a topic I will eagerly explore in graduate research.

Background Achieving automatic multi-camera tracking requires the successful prediction of a target's camera-to-camera movement. Predicting the next camera view in which a target is likely to appear requires inferring the locations of cameras along the common traffic routes taken by objects traversing the facility. This is known as estimating *activity topology*. A topology estimation procedure should establish for adjacent cameras a transition time distribution as well as a probability of transition. Successfully recovering this information would enable camera-to-camera transition prediction and successful track handoff even for non-overlapping cameras. During my nine weeks as a summer research intern at MIT Lincoln Laboratory, I produced and evaluated a first-pass solution for topology estimation (described in my previous research essay). As a graduate student, I look forward to developing a more thorough, validated procedure.

Previous Work Makris et al. [1] provide a novel algorithm for recovering topology of a non-overlapping camera network. Their method establishes that two cameras view a common route by detecting patterns in the cross-correlation of their respective entry and exit timing signals. When a related camera pair is detected, both a transition probability and a time distribution are recovered. They showed successful detection of ground truth topology in a six camera network for transition gaps as large as 10 seconds. Building on this method, Tieu et al. [2] propose an information-theoretic correspondence detection procedure with several advantages over cross-correlation. The limitation of applying both [1] and [2] in practice, however, remains the computational expense required to detect correspondence for each possible pair of access regions across the hundreds of cameras available in real-world facilities. Addressing this bottleneck, van den Hengel et al. [3] propose an inexpensive topology estimation method explicitly designed to be scalable. The authors demonstrate feasible computational performance for a hundred network camera, though successful estimation of ground truth topology remains unevaluated.

Research Plan Assessing current work in multi-camera topology estimation, I posit that two significant gaps exist. First, a topology inference method that balances scalability with accurate transition prediction is currently unavailable. Second, the feasibility of track handoff given inferred transition probabilities and distributions remains unevaluated. I propose a two stage research project to meet these critical needs.

First, I will build a topology estimator that integrates scalable methods with the sophisticated transition estimation of [2]. Given a large camera network, the system would rapidly weed out improbable links using [3]'s inexpensive exclusion process and then for the remaining links estimate information-theoretic transition time distributions and probabilities. The major contribution would be an algorithm that successfully integrates coarse and fine-grained search while maintaining a balance of computational speed and predictive accuracy.

The second stage of the project requires evaluating my procedure on surveillance footage from several facilities where significant gaps exist between cameras. Significantly, unlike [1], [2], and [3] I intend to assess not only the accuracy of estimated topology compared to ground truth but also the ability of the inferred transition predictors to facilitate track handoff (see previous research). The broader aim here is to help fellow researchers and practitioners understand how variables such as traffic density, scene clutter, and camera gap distance influence the accuracy of camera-to-camera transition prediction for large camera networks.

Broader Impacts Improving topology estimation remains a critical requirement for achieving facilty-wide tracking systems. A reliable multi-camera tracker would revolutionize security at airports and other critical infrastructure sites. Officers could follow active suspects in real-time even across gaps in coverage and could reconstruct past activity to determine point-of-entry and identify collaborators. Automation would dramatically improve speed and accuracy of post-incident investigations and help prevent nascent attacks from materializing.

A major contribution of my project would be seeking partnerships with critical infrastructure facilities. Security experts could provide practical feedback that would guide my research toward the most useful results. Significantly, my research could relate important design considerations to practitioners. For example, motion detection algorithms perform best when objects are is easily distinguishable from the background, suggesting that facility lighting and decor can be tweaked to improve a tracking system’s performance. Establishing this two-way partnership would be a huge step toward achieving meaningful results.

Multi-camera video analytics remains an active area of research across the globe, especially in countries with widespread CCTV monitoring such as the UK. I anticipate my research could inspire collaboration with research groups such as those led by Makris in England or van den Hengel in Australia. My experience collaborating with Chinese teammates at RIPS (see previous research experience) gives me the confidence to initiate this partnership.

Finally, an operational multi-camera tracker could serve as a test bed for deeper analysis of surveillance video. For example, investigators could build a question answering system (“Who left the south exit after 5PM?”) on top of the existing tracker. Access to such a test bed would inspire progress in academic research and could improve facility security as well.

I look forward to a project that could offer dramatic improvements in securing our nation’s critical infrastructure while expanding collaboration with practioners and researchers around the globe. Impact-driven research has always been my ambition. An NSF fellowship award would provide me the credibility and freedom to seek these partnerships and make this project a reality.

References

[1] Makris, Ellis, & Black. Bridging the gaps between cameras. *Computer Vision and Pattern Recognition*, 2004.

[2] Tieu, Dalley, & Grimson. Inference of Non-Overlapping Camera Network Topology by Measuring Statistical Dependence. *International Conference on Computer Vision*, 2005.

[3] van den Hengel, Dick, & Hill. Activity Topology Estimation for Large Networks of Cameras. *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance. 2006.*