Research Statement

of Peter Gusev

1 Summary

My primary research interests are developing novel methods for decentralized, edge-supported collaborative mapping and 3D-reconstruction that support many (on the order of tens) mobile agents simultaneously and scale robustly. With the increasing capabilities of mobile devices it has became evident that the need for such methods extends well beyond the domain of robotics. My previous work at the UCLA Center for Research in Engineering, Media and Performance (REMAP) has gradually evolved over 7 years from designing low-latency data dissemination algorithms for data-centric networks into building collaborative mixed reality systems. It has generated a number of publications and experimental production projects at the UCLA Theater Film and Television school, and, most importantly, motivated me to pursue research in distributed computer vision algorithms.

2 Previous Research Experience

The research I initially set out to do at UCLA REMAP was an original work on designing a low-latency streaming algorithm for the NSF-supported future internet architecture "Named Data Networking" (NDN). Existing connection-oriented networking protocols such as IP are challenged by newly emerging usage patterns and use cases such as smart cities, autonomous vehicles and multimedia-rich interactive systems (mixed reality, telepresence, remote work, immersive learning). NDN, a specific architecture of a broader concept of Information-Centric Networking (ICN), promises an elegant, universal and reliable solution for these new types of applications, by providing secure and robust communication [1]. My work, published in several conferences ([2], [3], [4]), has been instrumental for NDN community in designing architecture and congestion control schemes, and load-testing the network. Moreover, it shifted my research interests towards emerging use-cases in collaborative interactive mixed reality systems. In the ICE-AR project, I focused my research efforts on designing a proof-of-concept for a system where edge networking nodes support augmented reality (AR) devices in obtaining deeper knowledge about their environments. This work has highlighted the opportunities of robust and scalable networking for decentralized AR, offered by ICN [5]. The ICN benefits of accessing data at any level of granularity (stream, frame, field-of-view, feature point, detected object) and on-demand data processing, allow power-constrained AR devices to offload heavy processing (object recognition and scene understanding) onto the edge infrastructure of ML-enabled nodes. The resulting "deep context" of the surroundings can be shared between clients and utilized for improved quality of AR experience. My subsequent experimental work on supporting infrastructure and mobile device relocalization for immersive AR has demonstrated the need for further research in collaborative 3D-reconstruction and mapping in instrumented environments.

3 Current Projects

The immersive AR theatre domain presents a plethora of challenges for state-of-the-art computer vision and machine learning. At the center of my work is the typical scenario of multiple (15 or more) heterogeneous clients (mobile AR, headsets, desktop VR, etc.) that share the same physical and virtual environments, operate in low-lit dynamic theatrical spaces and are supported by a local network of computing nodes and static cameras. My involvement as a lead developer drove me to research practical solutions for multi-device relocalization and resulted in a prototype, which registers mobile devices in physical spaces, equipped with multi-camera RGB-D vision system OpenPTrack ([6], [7]). While there is still work required for improving the accuracy, the integration allows to align tracking spaces between OpenPTrack and mobile AR, and share data, otherwise not available to devices (people and object tracks, movement patterns and clustering, scene graphs). Even though the events of the 2020 pandemic have shifted my work agenda towards addressing the immediate need to build shared virtual production environments for faculty and students, it has not affected my long-standing interest in designing computer vision and 3D reconstruction algorithms for collaborative interactive mixed reality systems. Over the past two years, my experimental work in this domain has resulted in my building an underlying multiplayer AR infrastructure for an immersive theatrical performance [8], several virtual production projects and an ongoing effort to build a hybrid AR/VR space for a scalable, immersive learning environment at the UCLA Anderson school.

4 Future Work

In the post-pandemic world the long-standing need for distributed media-rich systems will become even more evident. They represent the next-generation multi-user mixed reality environments that will find themselves permeating many domains ranging from education to entertainment, to telepresence and remote work, to smart communities and cities. With the advent of consumer-grade LiDARs and specialized deep learning processors, the proliferation of mobile AR ([9]) brings us closer to a future where cooperative visual reconstruction and mapping becomes ubiquitous. My future research is aimed at synthesizing novel methods for hybrid collaborative SLAM and 3D-reconstruction. The "hybridity" of these methods means that they must allow mobile agents to operate independently or ad hoc, while leveraging local computational power whenever available, with little to no dependency on the cloud. Reliable communication is of paramount importance for decentralized mapping approaches. The state-of-the-art systems ([10], [11]) try to find a delicate balance between network bandwidth and mapping accuracy. In these approaches only the results of localization are unique to each agent, while mapping and reconstruction data ultimately represent same physical locations shared by many. This brings forward an idea of data sharing and computation resource pooling among available nodes. What if localization measurements are accessed and iterated upon without any assumptions about the network topology or physical adjacency of the agents? In my future work, I'd like to explore the benefits of applying data-centric approaches to agent-to-agent and agent-to-infrastructure connectivity. This will help to leverage such affordances of ICN as intrinsic multicast, in-network storage and processing, as well as channel-agnostic data security [5].

As the first step of my work, I intend to perform a careful study of the state-of-the-art influential works in the area of collaborative SLAM and 3D-reconstruction ([12], [13], [11]) and identify major scalability bottlenecks. This will lay the groundwork for the second step where the prototype will be developed in an iterative fashion with the "running code" deliverables in the form of open-sourced libraries and components. The intention is to use OpenPTrack as the underlying platform for edge computing infrastructure, however the final decision will be highly dependent on the result of the first step. Finally, as soon as the prototype reaches an alpha stage, a continuous evaluation of the system accross three criteria will be performed. These are: a) accuracy of relocalization and mapping for dynamic scenes in comparison to the state of the art, b) robustness in the presence of intermittent connectivity and high client mobility, and c) scalability to tens (over 30) of mobile agents. I welcome any constructive critique and insights on my future work and am looking forward to discussing it in more depth with the research faculty during the interview or, if feasible, a campus visit. I also aim to leverage my connections at UCLA REMAP and the UCLA Internet Research Lab to collaborate on the possible applications of my future work and share intermediate results accross disciplines.

References

[1] A. Afanasyev, J. Burke, T. Refaei, L. Wang, B. Zhang, and L. Zhang, "A Brief Introduction to Named Data Networking," in *Proceedings - IEEE Military Communications Conference MILCOM*, 2019.

[2] P. **Gusev** and J. Burke, "NDN-RTC: Real-time videoconferencing over Named Data Networking," in *Proceedings of 2nd International Conference on Information-Centric Networking (ICN'15)*, September 2015.

[3] P. **Gusev**, Z. Wang, J. Burke, L. Zhang, E. Muramoto, R. Ohnishi, and T. Yoneda, "Real-time streaming data delivery over Named Data Networking," *IEICE Transactions*, May 2016.

[4] S. Mastorakis, P. **Gusev**, A. Afanasyev, and L. Zhang, "Real-time data retrieval in named data networking," in *2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN)*. IEEE, 2018, pp. 61–66.

[5] P. **Gusev**, J. Thompson, and J. Burke, "Data-centric video for mixed reality," in *2019 28th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2019, pp. 1–11.

[6] M. Munaro, A. Horn, R. Illum, J. Burke, and R. B. Rusu, "OpenPTrack: People Tracking for Heterogeneous Networks of Color-Depth Cameras," in *IAS-13 Workshop Proceedings: 1st Intl. Workshop on 3D Robot Perception with Point Cloud Library*, 2014.

[7] M. Munaro, F. Basso, and E. Menegatti, "OpenPTrack: Open source multi-camera calibration and people tracking for RGB-D camera networks," *Robotics and Autonomous Systems*, 2016.

[8] "Ucla tft future storytelling summer institute 2018." [Online]. Available: https://remap.ucla.edu/future-storytelling-summer-institute-2018/

[9] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui, "Mobile Augmented Reality Survey: From Where We Are to Where We Go," *IEEE Access*, vol. 5, pp. 6917–6950, 2017.

[10] T. Cieslewski, S. Choudhary, and D. Scaramuzza, "Data-efficient decentralized visual SLAM," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 2466–2473, 2018.

[11] A. Cunningham, V. Indelman, and F. Dellaert, "DDF-SAM 2.0: Consistent distributed smoothing and mapping," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2013.

[12] A. Dai, M. Nießner, M. Zollhöfer, S. Izadi, and C. Theobalt, "BundleFusion: Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Re-integration," *ACM Transactions on Graphics*, apr 2016. [Online]. Available: http://arxiv.org/abs/1604.01093

[13] S. Golodetz, T. Cavallari, N. A. Lord, V. A. Prisacariu, D. W. Murray, and P. H. Torr, "Collaborative large-scale dense 3D reconstruction with online inter-agent pose optimisation," *IEEE Transactions on Visualization and Computer Graphics*, 2018.