BlockSLAM: Privacy and Security in Spatial Computing for the Gig Economy

Short Whitepaper

Sanjeev J. Koppal
University of Florida
sjkoppal@ece.ufl.edu

Abstract

Spatial Computing encapsulates all the fundamental infrastructure for the next wave of virtual applications. By combining the fast moving realms of augmented reality, 3D vision, action recognition and on-demand services, the future will allow high-skilled gig workers wearing smartglasses to coordinate complex, ordered and time-sensitive tasks, from modern building and construction to fast response to pandemics. In this article, we contend that the future growth of Spatial Computing is constrained by questions of privacy and security. ***This article is a whitepaper that explains how physics-based depth sensors, under the constraints of blockchain processing, can provide privacy and security for Spatial Computing.***

1. Introduction

Many 3D vision systems localize cameras within a scene using simultaneous localization and mapping (SLAM) and structure from motion (SfM). Applying such technologies to smartglasses, body-cams, helmet-cams or wearable sensors can coordinate humans working together in the coming wave of Web 3.0 commerce. These impacts include highly skilled gig workers building complex machinery together, coordinating large infrastructure building, distributing health services across continents and managing scale invariant logistics. The questions of where people were and what they did can be computed by SfM/SLAM and validated by BlockChain in a decentralized manner.

While spatial computing is crucial, there exists a twin privacy and security hole in the current framework. The security hole is due to the fact that 3D point clouds obtained from SfM are enormous, and cannot be stored on the chain. Therefore, chain programs in Ethereum cannot verify computer vision queries (such as recognition of certain safety actions, say, in a decentralized engineering task) to the blockchain certificate — they can only verify that *some* action was performed by an agent earlier at a time claimed by said agent.

Solving this security hole by storing the point clouds at a third party location, and using conventional network security to query the point clouds for certain actions done in a certain order, creates a privacy problem. This is because sparse SfM point clouds retain enough information to reveal scene appearance and compromise privacy. We have previously shown that a privacy attack on SfM data reconstructs color images of the scene from the point cloud [23], using a cascaded U-Net that takes as input, a 2D multi-channel image of the points rendered from a specific viewpoint containing point depth.

In this article, we discuss how to retain privacy and security for spatial computing, that will impact complex group activities. The solution lies in combining two fascinating areas of recent research. The first involves cameras with novel optics that directly capture geometric information for accurate SLAM in a manner that is irreversible, and protects privacy information. The second involves storing activity recognition features in the limited (kB) footprint of a single block, to allow decentralized validation of the activities.

2. Related Work

The combination of novel depth sensors with deep learning has impacted many fields, from video games to autonomous cars [30, 13, 4]. Here we briefly cover some of the related work that is creating an infrastructure widely known as visual computing, where tracking humans in 3D can induce many useful applications.

SfM and SLAM: Augmented reality (AR) and extended version (XR) have spread rapidly, driven by mobile technologies such as ARCore, ARKit, 3D mapping APIs, and new devices such as HoloLens. These devices have set the stage for wearable augmentation of groups of humans working together in professional settings, homes and other sensitive environments. SLAM and SfM allow wearable sensors to estimate their precise pose within the scene. However, this requires persistent storage of sparse 3D point clouds, which we have shown to be surprisingly containing enough information to reconstruct detailed comprehensible images of the scene. This suggests that the persistent point cloud storage poses serious privacy risks that have been widely ignored so far but will become increasingly relevant as localization services are adopted by a larger user community.

Adaptive depth sensors: LIDAR and other novel depth sensors that are flexible and adaptive in the measurements they make are emerging[29, 5]. These use beam steering modulation for intelligent workplaces [25], displays [9] and sensing [17]. MEMS mirrors have been used in scanning LIDARs, for highly reflective fiducials in both fast 3D tracking and VR applications [16, 15]. Galvo mirrors are used with active illumination for light-transport [8] and seeing around corners [20]. Light curtains are used for flexible, structured light reconstruction [2]. We contend that it is possible for adaptive sensors to capture only what is needed for SfM/SLAM.

Depth Completion: Combining sparse depth measurements with color imagery to fill in depths allows dense reconstruction [14, 30, 26]. Benchmarks exist [30] and guided upsampling has been used as a proxy for sensor fusion such as the work that has recently been done for single-photon imagers [12] and flash lidar [7]. In contrast, we measure sparse low-power LIDAR depth measurements and we seek to flexibly change the sensor capture characteristics in order to leverage adaptive neural networks such as [13, 4]. We contend that adaptive sensors can be used such that depth completion does not compromise privacy, and yet allows for accurate SfM/SLAM.

Privacy and security in computer vision. We contend that adaptive depth sensors that allow for SfM/SLAM but do not allow for upsampling inversion can enable privacy preserving localization. The final piece is to integrate these technologies with blockchain based decentralized verification. Prior Privacy and Security work in vision include K-anonymity [28], where stolen keys can only reliably identify k database entries. This idea of deidentification has been applied in images [19] and video [1]. In contrast, our sensors will have significant impact in using smartglasses and wearables in connected health [3] and smart homes [18] to coordinate human group tasks.

3. Learning to BlockSLAM

BlockSLAM consists of two components. The first are physics-based sensors which directly capture geometric features from the scene in such as way that privacy is preserved, by SfM for SLAM is done accurately. The second component is an action recognition protocol done over the chain, within the kB limits of a single block. We now describe each component in detail.

Privacy preserving physics-based SLAM sensors: Conventional SLAM works by using perspective cameras who images are processed by (mostly) hand-trained feature detectors like SIFT to obtain correspondences across views. We have shown that neural networks can inverse such sparse data, breaking privacy protections [23].

Instead, we point to physics-based depth sensors, where the optics and imaging components perform the bulk of the difficult vision processing "off-board". These sensors remove undesirable information prior to image capture, and complement existing hardware and software based approaches to privacy preservation, such as deidentification and cryptography, which can add further levels of protection to these physics-based sensors. We have been a first mover on these sensors, [10] and follow-up work from others has embedded neural networks in optics layers[6] and created sensors that directly obtain geometric entities from measurements, such as blocks and planes[11].

A final step is to train such sensors such that the geometric measurements are better for SLAM and not for recovering imagery. Such features have been discovered for SLAM [27] and we have built a general framework for deep learning-based privacy preserving encoders [21], which can be applied to a series of optical-based privacy sensors that we proposed[24, 22].

Action recognition policy over Ethereum: Etheruem allows for contracts between a taskmaster and clients to make sure (a) tasks are done in the way they were specified and (b) tasks are done in the order they are specified to activate other contracts. Our assumptions here are that a map of the rigid scene where the activities are taking place (factory floor, home, outdoor street, etc.) are stored and shared openly. We also assume that a pretrained action recognition network A is agreed on and an original version A_{orig} is openly available.

The A_{orig} is used to compare features across time, to make sure clients are not using the same video clips to show repeated work. However, the clients may try to fool A_{orig} with slightly modified versions of the same video clip passed off as different work cycles.

To combat this, each taskmaster puts out a block which modifies the mother network with randomly selected weights that are reset to random values, create a modified network A_{rand}. The taskmaster privately retrains an augmented $A_{augment}$ network with a few extra layers to compensate for these randomly selected changed weights. Each client of the taskmaster uses the randomly modified network A_{rand} and generates output at a small number of locations and places a few instantiations of the feature vector on the chain. Optimization is impossible since the client does know the $A_{augment}$ network.

Ethereum contracts are written that check if the activities put forward in this way are consistent with the augmented network $A_{augment}$ and the rigid 3D structure of the scene. Every augmented network has a time limit, and new work cycles have new augmented blocks. We exploit network fragility to make it difficult for clients to optimize their inputs — i.e. input is restricted vanilla video feed. The entire system is decentralized as no one is the ultimate decider on the contracts which have been made secure, *both* with the chain and network fragility.

4. Summary of the Proposed Research Sub-area

Adaptive physics-based sensors for SLAM can allow for reliable and interpretable imaging for spatial computing. We will soon see novel learning driven imaging systems designed in concert with the chain, which we term BlockSLAM. The protection of privacy and security in the coming wave of spatial computing will be an important ethical and social impact of such research.

5. Acknowledgments

The author was partially supported by the National Science Foundation (NSF) through NSF grants 2004544 and 1942444.

References

[1] Agrawal, P., Narayanan, P.: Person de-identification in videos. IEEE Transactions on Circuits and Systems for Video Technology (2011) 2

[2] Bartels, J.R., Wang, J., Whittaker, W., Narasimhan, S.G.: Agile depth sensing using triangulation light curtains. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 7900–7908 (2019) 2

[3] Beach, S., Schulz, R., Downs, J., Matthews, J., Barron, B., Seelman, K.: Disability, age, and informational privacy attitudes in quality of life technology applications: Results from a national web survey. ACM Transactions on Accessible Computing (2009) 2

[4] Bergman, A., Lindell, D., Wetzstein, G.: Deep adaptive lidar: End-to-end optimization of sampling and depth completion at low sampling rates. ICCP (2020) 2

[5] Chan, S., Halimi, A., Zhu, F., Gyongy, I., Henderson, R.K., Bowman, R., McLaughlin, S., Buller, G.S., Leach, J.: Long-range depth imaging using a single-photon detector array and non-local data fusion. Scientific reports **9**(1), 8075 (2019) 2

[6] Chang, J., Sitzmann, V., Dun, X., Heidrich, W., Wetzstein, G.: Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. Scientific reports **8**(1), 12324 (2018) 2

[7] Gruber, T., Julca-Aguilar, F., Bijelic, M., Ritter, W., Dietmayer, K., Heide, F.: Gated2depth: Real-time dense lidar from gated images. arXiv preprint arXiv:1902.04997 (2019) 2

[8] Hawkins, T., Einarsson, P., Debevec, P.E.: A dual light stage. Rendering Techniques **5**, 91–98 (2005) 2

[9] Jones, A., McDowall, I., Yamada, H., Bolas, M., Debevec, P.: Rendering for an interactive 360 light field display. ACM Transactions on Graphics (TOG) **26**(3), 40 (2007) 2

[10] Koppal, S.J., Gkioulekas, I., Young, T., Park, H., Crozier, K.B., Barrows, G.L., Zickler, T.: Toward wide-angle microvision sensors. IEEE transactions on pattern analysis and machine intelligence **35**(12), 2982–2996 (2013) 2

[11] Lee, J., Gupta, M.: Blocks-world cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11412–11422 (2021) 2

[12] Lindell, D.B., O'Toole, M., Wetzstein, G.: Single-photon 3d imaging with deep sensor fusion. ACM Transactions on Graphics (TOG) **37**(4), 113 (2018) 2

[13] Liu, C., Gu, J., Kim, K., Narasimhan, S., Kautz, J.: Neural rgb-to-d sensing: Depth and uncertainty from a video camera. arXiv preprint arXiv:1901.02571 (2019) 2

[14] Lu, J., Forsyth, D.: Sparse depth super resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2245–2253 (2015) 2

[15] Milanović, V., Kasturi, A., Siu, N., Radojičić, M., Su, Y.: "memseye" for optical 3d tracking and imaging applications. In: Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS), 2011 16th International. pp. 1895–1898. IEEE (2011) 2

[16] Milanović, V., Kasturi, A., Yang, J., Hu, F.: A fast single-pixel laser imager for vr/ar headset tracking. In: Proc. of SPIE Vol. vol. 10116, pp. 101160E–1 (2017) 2

[17] Nayar, S.K., Branzoi, V., Boult, T.E.: Programmable imaging using a digital micromirror array. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. vol. 1, pp. I–I. IEEE (2004) 2

[18] Neustaedter, C., Greenberg, S.: Balancing privacy and awareness in home media spaces. UBICOMP (2003) 2

[19] Newton, E., Sweeney, L., Malin, B.: Preserving privacy by de-identifying facial images. CMU Technical Report CMU-CS-03-119 (2003) 2

[20] O ´ Toole, M., Lindell, D.B., Wetzstein, G.: Confocal non-line-of-sight imaging based on the light-cone transform. Nature **555**(7696), 338 (2018) 2

[21] Pittaluga, F., Koppal, S., Chakrabarti, A.: Learning privacy preserving encodings through adversarial training. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 791–799. IEEE (2019) 2

[22] Pittaluga, F., Koppal, S.J.: Privacy preserving optics for miniature vision sensors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 314–324 (2015) 2

[23] Pittaluga, F., Koppal, S.J., Kang, S.B., Sinha, S.N.: Revealing scenes by inverting structure from motion reconstructions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 145–154 (2019) 1, 2

[24] Pittaluga, F., Zivkovic, A., Koppal, S.J.: Sensor-level privacy for thermal cameras. In: 2016 IEEE International Conference on Computational Photography (ICCP). pp. 1–12. IEEE (2016) 2

[25] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., Fuchs, H.: The office of the future: A unified approach to image-based modeling and spatially immersive displays. In: Proceedings of the 25th annual conference on Computer graphics and interactive techniques. pp. 179–188. ACM (1998) 2

[26] Riegler, G., Rüther, M., Bischof, H.: Atgv-net: Accurate depth super-resolution. In: European Conference on Computer Vision. pp. 268–284. Springer (2016) 2

[27] Speciale, P., Schonberger, J.L., Sinha, S.N., Pollefeys, M.: Privacy preserving image queries for camera localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1486–1496 (2019) 2

[28] Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems (2002) 2

[29] Tasneem, Z., Wang, D., Xie, H., Koppal, S.J.: Directionally controlled time-of-flight ranging for mobile sensing platforms. In: Robotics: Science and Systems (2018) 2

[30] Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., Geiger, A.: Sparsity invariant cnns. In: 2017 International Conference on 3D Vision (3DV). pp. 11–20. IEEE (2017) 2