The Active Sensing Testbed

Lee Stearns, Neil Fendley, Ashley J. Llorens
The Johns Hopkins Applied Physics Lab (JHU/APL)
{Neil.Fendley; Lee.Stearns; Ashley.Llorens}@jhuapl.edu

Abstract

The Active Sensing Testbed (AST) is a novel framework for research in machine perception and world-view reasoning. The AST supports exploratory development of perception systems that can build internal models of the world by combining multi-view and multi-modal analytics, utilize these models to form hypotheses about a scene, and potentially take action to fill in gaps in knowledge or make predictions about future world states. As a modular software framework, the AST is intended to lower the barrier to entry for researchers and developers in applying state-of-the-art computer vision techniques to real-world problems.

Background and Project Goals

The Active Sensing Testbed (AST) provides a development environment to facilitate novel research in machine perception, particularly research involving active perception where a system can take actions such as changing sensor positions or adjusting parameters to investigate hypotheses, fill in gaps in knowledge, or make predictions about future states. The AST is intended to help advance machine perception research beyond narrowly-trained, single-purpose computer vision algorithms that may provide state-of the-art pattern recognition but can be fragile in handling the complexities of real-world scenes. The AST provides a modular architecture for synthesizing multiple views, multiple sensing modalities, and complementary analytics to help provide robust inference and prediction under real-world conditions.

Framework Architecture

The architecture of the AST centers around a server that receives data feeds from multiple sensors, computes selected analytics and transformations on input data, and then sends analytics and metadata to subscribers for visualization. An example of this architecture is shown in Figure 1. We have created a research testbed around the AST at the Intelligent Systems Center of the Johns Hopkins Applied Physics Laboratory (JHU/APL)[1]. Our testbed includes four ceiling-mounted pan-tilt-zoom cameras to facilitate data collection as well as algorithm design and evaluation. Our AST implementation also includes an operator interface to enable human-machine interaction. Through the operator interface and the associated application programming interface (API), a system operator or other remote user can view data feeds, overlay computed analytics and metadata, and issue commands to the system. We envision that our AST software framework can support similar setups in additional locations for research or other applications.

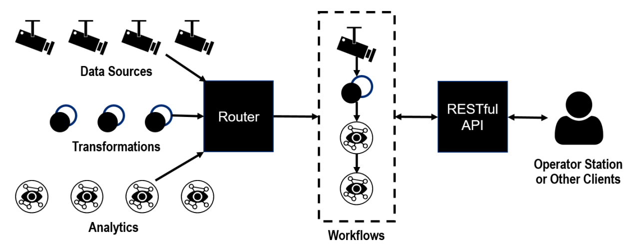

Figure 1: Active Sensing Testbed server architecture.

To allow other projects to easily interface with the AST server functions, we have implemented both a REST API and a Python library. Clients can subscribe to server workflows consisting of user-defined graphs that connect input data sources to sequences of operations. For example, a user may specify a networked camera as a data source, apply a series of transformations to the video from that camera, and then perform analytics on the transformed images. Using these tools allows researchers to build higher-level analytics and reasoning on top of baseline analytics, or to quickly investigate how different transformations affect a scene.

[1] For more information on the Active Sensing Testbed, visit https://www.jhuapl.edu/isc

Baseline Capabilities

Our baseline AST implementation includes a suite of real-time analytics and transformations. Baseline analytics include semantic instance segmentation and human pose estimation. Transformations include traditional operations such as flipping, rotating, color/brightness adjustments, and adding noise, as well as face-swapping, day/night transformations, and dynamic object removal. While many of these algorithms are built upon open-source or published techniques, the novelty of the AST is that it brings these (and potentially many other) state-of-the-art techniques together as an interactive suite of machine perception capabilities.

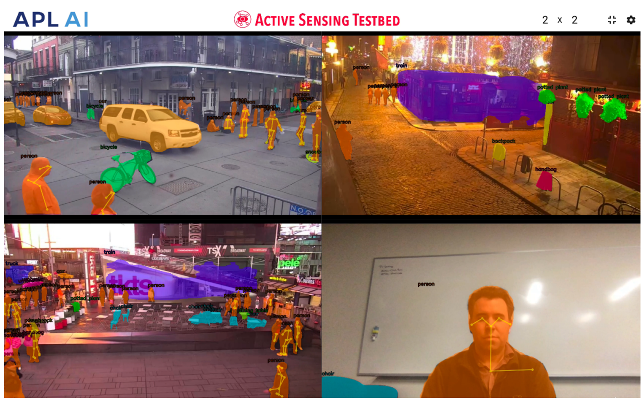

Figure 2: A screen capture demonstrating the Active Sensing Testbed running analytics across multiple feeds

Figure 2 provides a screen capture illustrating baseline AST analytics suite while Figure 3 illustrates a face-swapping transformation based upon (Liu et. al. 2019). We used this technique to create a real-time demonstration that transforms the face of an arbitrary subject to one of a set of preselected faces.

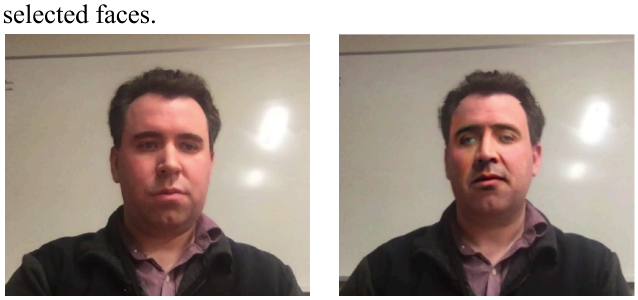

Figure 3: Face swapping demonstration, blending the subject's face with that of the actor Nicolas Cage.

We applied a popular style-transfer technique called CycleGAN (Zhu et. al. 2017) to explore another form of real-time image transformation applied to full images. To test the concept, we trained a simple day/night transformation model and applied it to a real-time video feed from Times Square. While some scenes yielded compelling transformations, future work will focus on creating a more robust capability that works across a variety of scenes and conditions.

To test combining AST analytics and reasoning, we used our baseline semantic instance segmentation algorithm (He et. al. 2017) to detect a target object—in this case a person—and identify the pixels associated with that object. We then replace those pixels with a pretrained adaptive background model and attempt to blend the replaced pixels with the rest of the image so that the person is no longer visible.

Figure 4: A Frisbee being exchanged for a football behind an occluding barrier.

Ultimately, we aim to apply the AST to advance research toward addressing enduring challenges in machine perception. Ideally, a perception algorithm should be able to observe a sequence like the one illustrated in Figure 4, report that a frisbee has been exchanged for a football, and represent the space behind the barrier as a likely location for the missing frisbee. While there are many potential approaches to addressing perception challenges such as this one, we argue that frameworks like AST can help catalyze near-term progress beyond individual narrowly-trained analytics.

Conclusion

The Active Sensing Testbed is a modular framework intended to support and accelerate machine perception research. Our framework greatly simplifies the process of integrating off-the-shelf algorithms for research and application, providing a powerful tool for both novice and experienced AI researchers. We have created a baseline suite of capabilities including state-of-the-art computer vision analytics and image transformation tools to demonstrate the efficacy of our architecture and research value of our approach. We anticipate that the AST will help facilitate AI research across a wide variety of domains and applications.

References

He, K., Gkioxari, G., Dollár, P., & Girshick, R.B. (2017). Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 2980-2988.

Liu, M. Y., Huang, X., Mallya, A., Karras, T., Aila, T., Lehtinen, J., & Kautz, J. (2019). Few-shot unsupervised image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision (pp. 10551-10560).

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).