# Research Statement

Robert LiKamWa, roblkw@rice.edu

My research interest is to understand and build energy-efficient mobile systems, especially architecture and operating systems support for a future of efficient personal computing. Over the past five years, I have been driving toward a vision of *continuous mobile vision* services, which invoke frequent vision sensing, computation, and offload to understand a user's real-world environment, providing vision-based services that relieve users' memory and attention. Continuous mobile vision will revolutionize personal computing: immersive interaction will empower consumer applications; sight assistance will aid the memory- and vision-impaired; hazard detection will alert military units; and visual localization will guide robotic drones.

However, systems hosting continuous mobile vision face a fundamental challenge: *energy efficiency*. Despite recent advances in vision algorithms, current systems are severely limited by the substantial energy consumption of always-on vision sensing and processing. Conventional vision systems drain a small wearable battery in 40 minutes and raise the device's surface temperature to over $55\,°C$ [APSys '14].

This efficiency challenge is symptomatic of a key fact: conventional mobile systems are provisioned for on-demand user interaction, not for continuous service. Systems must be redesigned at all levels to meet the demands of periodically and/or sporadically interpreting visual information. My research takes on this challenge by innovating mobile system support, using an experimental, prototype-driven approach that combines domain knowledge from software systems, hardware architecture, and machine learning.

Dissertation Work: Provisioning Mobile Systems for Continuous Mobile Vision

My dissertation research ventures through multiple levels of the vision system stack, designing solutions in: (*i*) application support, (*ii*) operating systems, and (*iii*) sensor hardware, as shown in the figure. The principal objective has been to enable *energy-proportionality*; energy consumption should be proportional to the quantity and quality of the capture and compute needed to complete a set of tasks.
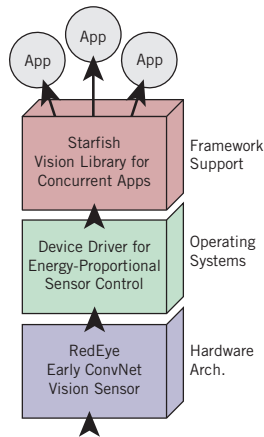
My envisioned future of continuous mobile vision assumes that multiple vision applications will concurrently run in the background. In today's systems, this challenges system performance and efficiency with substantial computational demand. I relieve this burden by providing application support to concurrent vision applications through Starfish, a split-process library system to reduce redundant computations [MobiSys '15]. Starfish is inspired by two key observations: (*i*) applications use identical vision primitives; frames, points, descriptors, etc., are derived through common streams of processing; (*ii*) developers leverage common libraries, e.g., OpenCV, to implement vision algorithms. Using these observations, my work transforms an existing vision library into a Starfish Library, which intercepts and shares common library processing workloads to eliminate redundant computation and memory use. Built on efficient shared memory and function cache structures to reduce redundancy, Starfish ensures that computational energy consumption is proportional to the volume of features computed across the system. With Starfish, a system running 10 concurrent applications can run with less than 1% additional computational overhead than a system running one application.

The necessary input to vision processing, *image capture*, is itself known to be energy-expensive. Spending a summer at Microsoft Research, I characterized the power consumption of the imaging subsystem, surprisingly finding that image capture is not energy-proportional [MobiSys '13a]. That is, though the sensor is provisioned for high pixel counts and frame rates, I was surprised that we could not reduce energy consumption simply by reducing the resolution or frame rate of the image capture. I instead found that sensor efficiency is limited by a high static power, and devised two device driver mechanisms to circumvent this limit: (*i*) Aggressive standby: activating sensor registers to opportunistically place the sensor into a low-power mode; (*ii*) Pixel clock

frequency scaling: leveraging the time-power tradeoff afforded by frequency scaling to reduce energy consumption. These driver mechanisms proportionally scale energy-per-frame with the spatiotemporal resolution of the image capture. This gives developers the flexibility to use low-power, low-resolution modes and high-power, high-resolution modes as needed.

My characterization study also found a bottleneck to energy-efficient sensing: sensors consume high dynamic power, due to sending raw images through analog-to-digital conversion, i.e., readout. I designed a novel mixed-signal sensor architecture that shifts vision processing into the analog domain to reduce readout [Under Review]. With this shift, sensor output consists of low-bandwidth vision features. I found that early stages of vision ConvNets map well to the analog domain for many reasons: (*i*) Vision is patch-based, operating on local data; (*ii*) Vision is robust to noise; (*iii*) ConvNets are repetitive in nature, applying layers of convolutional instructions to image patches. My architecture exploits these findings to enable deep analog processing with fixed complexity. Through cyclic modular reuse of a column-based topology, I limit chip design complexity while enabling iterative execution before readout. My design uses capacitance-based tuning to admit signal noise for increased efficiency, providing energy-proportionality to vision fidelity.

Holistically, my dissertation is a cross-layer investigation, targeting a redesign of the vision system. In designing a library system for concurrent applications, driver mechanisms for energy-proportionality, and analog processing to reduce readout, my research substantially reduces vision system energy consumption.

Ahead: Ensuring Efficiency, Privacy, and Usability for Continuous Mobile Vision

My dissertation work is just the first step toward a complete rethinking of the vision system stack for efficiency. Moving forward, I target two roadblocks that prevent widespread adoption: *programmability* and *privacy*.

Near-term agenda: Operating System Support for Architecture-Aware Continuous Mobile Vision

My prior investigations reveal that sacrificing data fidelity and reducing performance will substantially reduce energy consumption. However, designing, executing, and tuning optimizations in complicated vision applications will overwhelm developers and systems resources, especially over mixed-signal domains. This necessitates novel operating system support for application development and runtime management.

While optimizing a vision application is imperative to reduce energy consumption, mobile application developers should not be forced to track the varying noise, energy, memory, and timing implications of analog processing sensors, digital accelerators, microprocessors, and network-based servers. Instead, my research will define *programming support for mixed-signal vision systems*, allowing developers to specify high-level application logic, timing, and fidelity constraints for mixed-signal computation. The resulting code a compiler and runtime design to manage complex task allocation and optimal power management on the heterogeneous system. This line of work presents an opportunity to work with programming language experts. Furthermore, in collaboration with machine learning researchers, I will use my research tools to design novel *architecture-aware vision algorithms* that fully leverage the mixed-signal system stack for unprecedented efficiency and performance, enabling new classes of vision applications.

My research will also pursue *efficient runtime management of vision tasks*. For energy-proportionality, a heterogeneous system, e.g., a mobile System-on-Chip, should shift always-on workloads to low-power resources, such that the system can aggressively gate high-energy power domains. To satisfy this, executing a complex vision taskgraph demands a separation between always-on management of low-complexity control states and isolated, bursty access to high-complexity data states. These states are currently intertwined in application processes, system services, and device drivers. To support a separation in memory access, I will pursue novel strategies in dynamic namespace partitioning and heterogeneous virtual memory management. With such division, the system can offload management tasks to a low-power core [ASPLOS '12], despite its limited cache size. Furthermore, as inspired by works to power-manage CPU cores [1] and devices [2], I will investigate support to efficiently migrate tasks before wake-up and sleep of cores and devices, reducing the critical path of power management latency.

Long-term vision: Privacy and Usability of Continuous Mobile Vision

Thus far, I have launched into a complete redesign of the vision system stack and will continue to investigate operating systems support for continuous mobile vision. Moving towards the future, I must collaborate with experts from many other domains to establish continuous mobile vision as the future of personal computing.

Real-world adoption of continuous mobile vision faces a strong social barrier: *privacy*. Applications using the vision system have the potential to violate the privacy of human users and subjects in a sensed environment. My research will investigate low-level mechanisms on which useful privacy policies can be efficiently implemented for higher layers of the system stack. For example, irreversibly executing a vision workload in the analog domain, and discarding the raw image in the sensor would provide strong, yet efficient, privacy guarantees to users and subjects for continuous vision applications. While privacy policies will be developed in collaboration with security, privacy, and human-computer interaction, my research will focus on hardware and software support for privacy policy implementations.

As a secure, efficient platform for continuous mobile vision emerges from my research, I will collaborate with ubiquitous computing and programming languages researchers to grow the application space of continuous mobile vision. Beyond mobile computing, a diverse set of fields, from autonomous vehicles to medical devices, would benefit from efficient continuous mobile vision research. Because of its practical social and economic impact, systems research on continuous mobile vision will capitalize upon funding opportunities in cyber-physical systems and Internet-of-Things. I will also collaborate with industry partners to create opportunities for technology transfer, building towards a future of efficient context-aware computing.

Continuous mobile vision is an exciting – yet deeply challenging – research pursuit, with many avenues for collaborative systems research. My research will continue to cross boundaries between software systems, hardware architecture, and machine learning research to influence the future design of mobile systems.

References

[1] Sankaralingam Panneerselvam, Michael M Swift, and Nam Sung Kim. Bolt: faster reconfiguration in operating systems. In *Proc. USENIX Annual Technical Conference*, pages 511–516. USENIX Association, 2015.

[2] Chao Xu, Felix Xiaozhu Lin, Yuyang Wang, and Lin Zhong. Automated OS-level device runtime power management. In *Proc. Int'l Conf. on Architectural Support for Programming Languages and Operating Systems*, pages 239–252. ACM, 2015.