Divide and conquer: parallel processing in computational imaging

David J. Brady*
University of Arizona, Department of Optical Sciences, Tucson, Arizona, United States

SpeedShot, a dual camera high-speed imaging technology recently demonstrated by Zhang et al.,[1] demonstrates a 32× increase in effective frame rate by leveraging the motion gradient in frames captured by parallel cameras. As interframe-motion estimation is the core element of most video compression algorithms, SpeedShot can be understood as a physical layer implementation of such an algorithm. Here, we seek to explain the context that makes SpeedShot interesting and to consider the roadmap for continuing improvements in camera information capacity.

Optical imaging systems have long been designed to match the human visual system. Cameras see three colors because humans see three colors. Cameras capture 30 frames per second to match human visual persistence. Cameras see megapixels because foveated human vision is satisfied with megapixels. The main reason that cameras have been limited to these ranges, however, is that, until recently, it has been difficult to build cameras that match human vision.

Ambient visible light contains an enormous amount of information. The space-bandwidth product of centimeter-scale apertures and sensors is nearly 10^9, and gigapixel cameras have been demonstrated using such apertures.[2] The time-bandwidth product of the visible spectrum also approaches 10^9 for a 10-microsecond exposure, meaning that at the full data cube, a visual field may exceed exopixels per microsecond. Of course, this data cube is extremely sparsely populated by ambient light, but when one considers that cameras commonly collect 10^{12} to 10^{15} photons per exposure, one understands that conventional photography vastly undersamples the data in the visual field.

The electrical power needed to process, compress, and transmit these data is the primary barrier to improving camera information capacity. Cameras currently may expend 10 to 100 nJ per pixel in this process, which means that cameras collecting 100 exopixels per microsecond may draw 1 to 10 W. Gigapixels or terapixels per second at this energy level are not practical. To address this challenge, computational imaging researchers have long pursued physical layer compression strategies.[3]

Array cameras are the simplest and most flexible mechanism for physical layer coding in photographic systems.[4] The trend toward array solutions is apparent across camera implementations, most notably in the ever-increasing number of camera modules used in mobile and autonomous devices. Although previous studies have utilized array cameras for temporal compression,[5,6] SpeedShot radically expands this work with a computational imaging framework for efficient high-speed video imaging. This framework implements a form of physical layer compression using a novel capture mechanism that significantly reduces the data footprint while preserving crucial motion information.

SpeedShot utilizes a low-speed dual-camera setup that simultaneously captures two temporally coded snapshots. This configuration, compatible with commercial low-speed cameras, offers a versatile and low-bandwidth alternative for various video applications. Instead of capturing individual frames, SpeedShot cross-references these two snapshots to extract a multiplexed temporal gradient image. This process leverages the insight that temporal gradient (TG) images, which capture differences between successive frames, are spatially sparse and primarily contain motion-related signals. By multiplexing several such images, SpeedShot obtains a compact and multiframe motion representation for video reconstruction.

For decoding, Zhang et al. proposed the motion-guided scale-recurrent transformer MSRT, an explicable deep learning model designed specifically for SpeedShot's unique temporal-only modulation model. MSRT exploits cross-scale error maps to bolster cycle consistency between predicted and observed data. It includes innovative components such as motion-guided hybrid enhancement, which extracts motion cues to refine restoration, and an error-aware scale recurrent network for multiscale reconstruction.

SpeedShot's advantages are illustrated in Fig. 1, which is reprinted from the work by Zhang et al.[1] The center images show captured data and an estimated frame from a conventional single-aperture coded exposure compressive video system. The right frame shows the SpeedShot result for the same data.

The design principles of SpeedShot, particularly its reliance on a dual-camera setup and its compatibility with commercial low-speed cameras, make it highly relevant for consumer devices such as smartphones and other devices that increasingly feature array cameras and parallel processing. The use of two cameras in SpeedShot naturally lends itself to parallel processing, where each camera captures its coded snapshot simultaneously, and the computational framework then

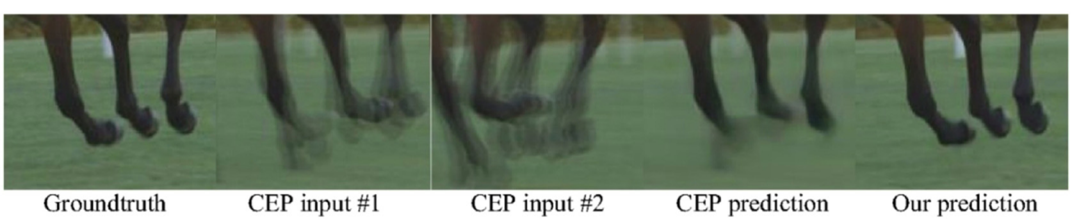

Fig. 1 Comparison of SpeedShot and single-camera coded exposure photography (CEP). Reprinted with permission from Zhang et al.[1]

*Address all correspondence to David J. Brady, djbrady@arizona.edu

 [DOI: 10.1117/1.AP.7.5.050502]

processes these two streams in conjunction to reconstruct the high-speed video. This "divide and conquer" approach, where two standard low-speed cameras collectively capture information that a single, much more expensive and data-intensive high-speed camera would typically acquire, represents a significant step toward democratizing high-speed imaging.

References

1. Y. Zhang et al., "High-speed video imaging via multiplexed temporal gradient snapshot," *Adv. Photon. Nexus* **4**(4), 046017 (2025).
2. D. J. Brady et al., "Multiscale gigapixel photography," *Nature* **486**(7403), 386–389 (2012).
3. D. Healy and D. J. Brady, "Compression at the physical interface," *IEEE Signal Process. Mag.* **25**(2), 67–71 (2008).
4. D. J. Brady et al., "Parallel cameras," *Optica* **5**(2), 127–137 (2018).
5. M. Shankar, N. P. Pitsianis, and D. J. Brady, "Compressive video sensors using multichannel imagers," *Appl. Opt.* **49**(10), B9–B17 (2010).
6. X. Yuan, Y. Sun, and S. Pang, "Compressive video sensing with side information," *Appl. Opt.* **56**(10), 2697–2704 (2017).