Figure 1. Overview of our method. Four coding patterns are applied to aperture plane during single exposure of image frame. Image and events are jointly used to reconstruct light field through convolutional neural network (CNN).

show that our method can reconstruct light fields more accurately than several other imaging methods with a single exposure, and our method works successfully with our prototype camera in capturing real 3-D scenes with convincing visual quality.

To the best of our knowledge, we are the first to investigate the combination of coded-aperture imaging and events in the context of computational light-field imaging. Our contribution is not limited to shortening the measurement time for a light field, but it enables us to *go beyond the limitation of the frame-rate* of image sensors; Our method can better utilize the time resource during a single exposure, and obtain more information per unit time (i.e., being time-efficient) than the baseline coded-aperture imaging method. Our method is also distinctive in the sense that events are induced actively by the camera's optics rather than the moving objects or ego-motion of the camera.

2. Related Works

The most straight-forward approach to light-field acquisition is to construct an array of cameras [8, 40, 51], which involves costly and bulky hardware. Lens-array-based cameras [1, 2, 31, 32] gained popularity because a light field can be captured in a single shot. However, this type of camera has an intrinsic trade-off between the number of views and the spatial resolution of each view. Mask-based coded-imaging methods [10, 14, 22, 26, 29, 30, 45, 47] have been developed to increase the efficiency of light-field acquisition. With coded-aperture imaging, two to four images, taken from a stationary camera with different aperture-coding patterns, are sufficient to computationally reconstruct a light field with 8×8 views in full-sensor resolution [10, 14, 42]. However, since several coded images need to be acquired in sequence, the lengthy measurement time remains an issue. Joint aperture-exposure

coding [28, 41, 42, 46] enables more flexible coding patterns during a single exposure but comes with complicated hardware implementation; as far as we know, only Mizuno et al. [28] reported a working prototype for this method but with awkward hardware restrictions for the non-commercialized image sensor. Our method also applies several aperture-coding patterns during a single exposure, but we combine them with an off-the-shelf event camera to achieve time-efficient and accurate light-field acquisition.

Event cameras [3, 9] are bio-inspired sensors that can record intensity changes asynchronously at each pixel with a very low latency. Compared with ordinary frame-based cameras, event cameras can capture more fine-grained temporal information in a higher dynamic range, which opens up many applications such as optical-flow estimation, deblurring, video interpolation, and camera pose estimation. Event cameras have also been used extensively for 3-D reconstruction [17, 24, 35, 37, 56, 57]. In these studies, however, the events were usually caused either by the moving objects or ego-motion of the camera. Our method can be regarded as a new application of an event camera; the events are induced actively by the camera's optics in the framework of computational imaging.

Single-view view synthesis (SVVS) [5–7, 19, 20, 34, 43, 44, 53, 55] is used to reconstruct a 3-D scene from a single image. Since this is geometrically an ill-posed problem, SVVS methods rely on implicit prior knowledge learned from the training dataset rather than physical cues. These methods are not necessarily designed to be physically accurate but to hallucinate a visually-plausible 3-D scene. Our method takes an orthogonal approach to SVVS; we use not only a single image but also a coded aperture and events to obtain solid physical cues from the target 3D scene.

Various computational imaging methods have been developed on the basis of deep-optics [13, 14, 21, 28, 33,