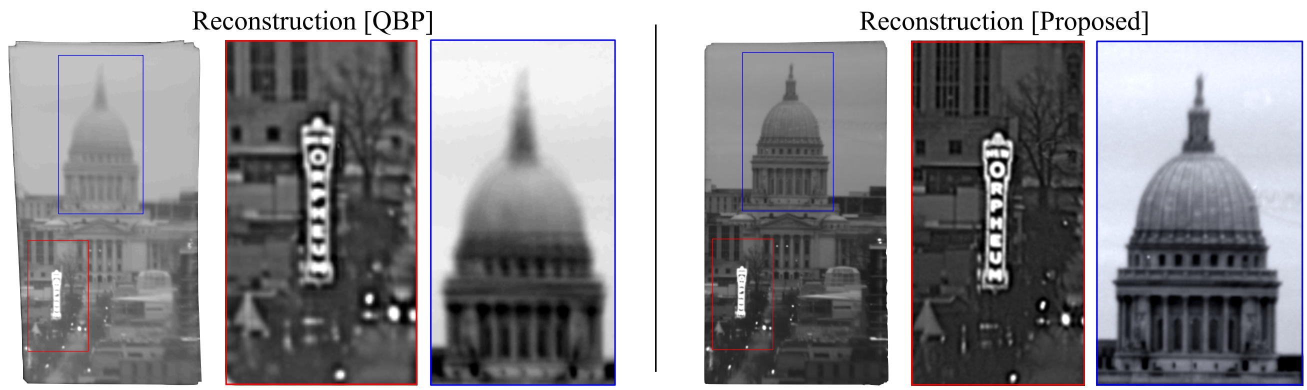
Supplementary Document for “Panoramas from Photons”

Sacha Jungerman, Atul Ingle, Mohit Gupta
sjungerman@wisc.edu, ingle2@pdx.edu, mohitg@cs.wisc.edu

(ICCV 2023)

S.1. Comparison with One-Shot Motion Compensation Methods



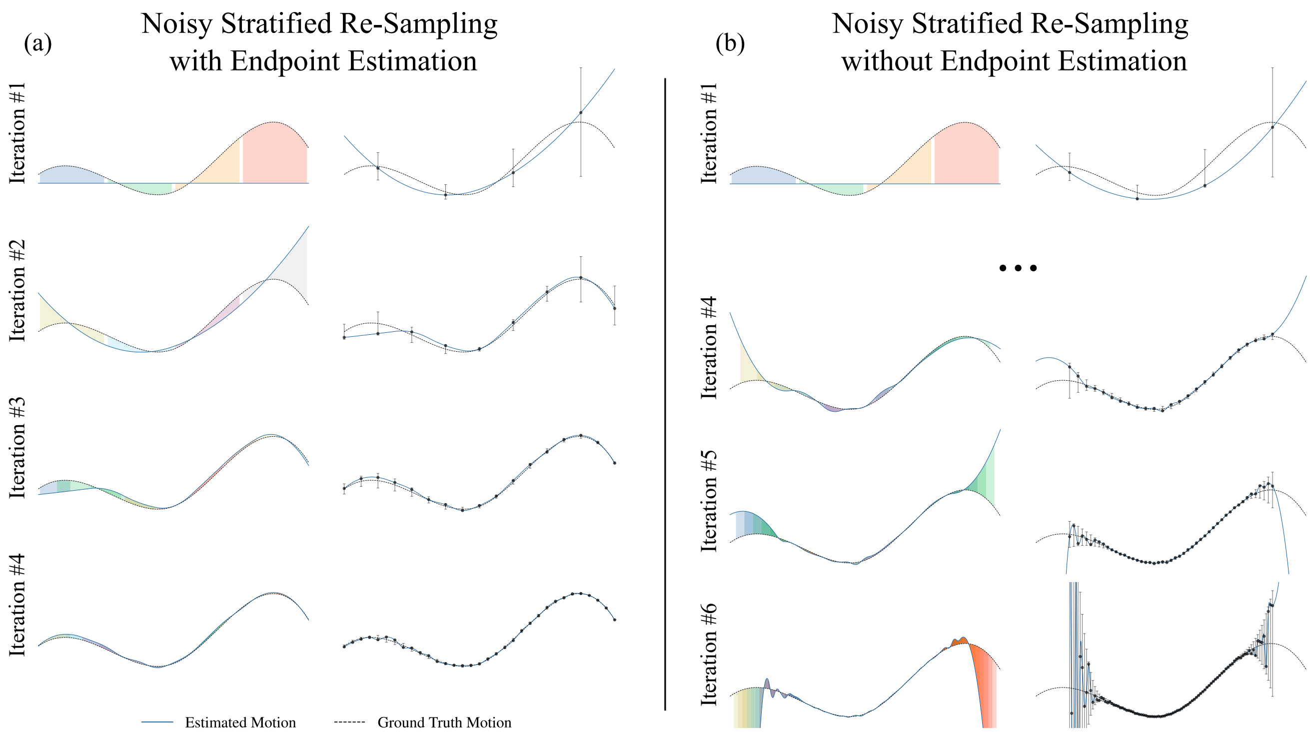
Suppl. Fig. 1. **Comparison with QBP [27]:** The raw photon data was first processed using QBP and then assembled into the reconstruction shown here using traditional stitching methods. QBP cannot fully compensate for the high-frequency camera vibrations causing the reconstruction to be blurry.

Recent work on motion compensation for single-photon camera data [27, 35, 20] focuses on techniques to aggregate binary frames, in a motion-aware manner, by only making a single pass over the set of binary frames. The stratified method proposed in this paper allows motion to be estimated and refined over multiple iterations, which makes direct comparison with existing techniques difficult. That said, the proposed iterative approach could be used with more general motion models, and therefore, is complementary to existing techniques [24]. An important next step is to apply our method synergistically with these prior works.

Quanta burst photography (QBP) [27] is a recently proposed algorithm that uses a more general optical flow-based motion model that locally warps and registers groups of binary frames. A comparison with our method is shown in Suppl. Fig. 1. The QBP image is generated by first creating motion-compensated frames from groups of $m = 1000$ binary frames. These frames are then assembled into the final reconstruction by using a traditional homography estimation technique [5]. This implementation makes a single pass over the binary frames. In contrast, our method assumes a planar motion model and uses two iterations to re-estimate the homography warps used for aligning and merging the raw binary frames. This provides sharper scene details such as the text and smaller features near the top of the dome.[3] Both QBP and our method are agnostic to the sequence length and only sensitive to the group size. In fact, many results in the QBP paper [24] use sequences of 10k frames or less. To ensure fairness of comparison in Suppl. Fig. 1, we use the *same group size* ($m = 1000$) *and the same total number of frames for QBP and our approach*. Also note that while QBP performs robust merging of binary frames (based on Wiener filtering), our approach simply sums binary frames once warped as we assume the estimated warps are sufficiently accurate. Incorporating this robust merging technique could further reconstruction quality.

[3] The QBP image quality is further degraded by quantization artifacts because it operates in the 8-bit quantized sRGB space whereas our method uses linearized intensity images.

S.2. Edge Effects and Pre-warping



Suppl. Fig. 2. **Edge Effects in Stratified Temporal Re-Sampling:** (a) The motion estimate provided by the underlying registration algorithm might be noisy (for simplicity, this noise was omitted in Fig. 3). Despite this noise, the motion trajectory rapidly converges to the ground truth motion. (b) Over the course of a few iterations, localization errors can accumulate at the ends of the trajectory if the endpoints of the trajectory are not estimated. Thankfully this phenomenon can be mitigated easily by either estimating boundary frames or simply stopping the iterative process before it occurs, as most sequences will converge to a satisfactory motion estimate in two or three iterations.

When an off-the-shelf homography estimation algorithm is used over a virtual exposure of duration τ, with respect to which time instant within the exposure duration will the estimated localization be? This is not an issue for small camera movements, but with fast motion, this ambiguity has a compounding effect. A sensible assumption would be to presume that the base model estimates the average location over an exposure time, or perhaps, the location at the center of the exposure. This observation is critical for two reasons: i) if we have already compensated for some motion when creating the aggregate frame, our new estimate will be relative to it, and ii) it allows us to localize with respect to any time instant during the exposure by warping the photon data, before aggregation, such that the time instant of interest is warped by the identity warp instead of the current motion estimate.

Without accounting for the former, any motion estimate would rapidly drift away. In practice, it is beneficial to compensate for this relative offset before aggregation and localization as it helps constrain the size of the aggregate frames and can lead to better matches. The latter enables the localization of off-center time slices of a virtual exposure, enabling precise localization at the boundaries of a captured sequence, which, as seen in Supp. Fig. 2, is necessary for proper convergence of the motion estimate, and finer motion estimation.