Object Tracking and Reconstruction with a Quanta Image Sensor

Istvan Gyongy[1], Tarek Al Abbas[1], Neale A.W. Dutton[2], Robert K. Henderson[1]
[1]The University of Edinburgh, Institute for Integrated Micro and Nano Systems, Edinburgh, U.K.
[2]STMicroelectronics, Imaging Division, Tanfield, Edinburgh, UK, EH3 5DA
Istvan.Gyongy@ed.ac.uk Tel: +44 131 650 5568

Abstract—Quanta Image Sensor devices offer single-photon sensitivity, coupled with high frame rates (>100kfps), making them ideal for the tracking of high-speed objects in low light conditions. However, motion artefacts emerge as bit-planes are aggregated in time to increase bit depth (and image detail). This work explores a scheme for object tracking, based on a sequence of binary output fields, followed by a reconstruction of the object, through summing transformed versions of the bit-planes.

I. Introduction

Quanta Image Sensors (QIS), and photon counting sensors in general, are enjoying considerable research interest, thanks to ongoing developments in enabling solid-state technologies [1]. Deep-sub-electron read noise (DSERN) CMOS image sensor pixels have been devised [2], in addition to Single Photon Avalanched Diode (SPAD) image sensors attaining DSERN [3,4]. On the application side, algorithms have been proposed for low-light object classification and tracking based on streams of photon counts [5], as well as image reconstruction from low photon count QIS data [6].

The basic output of a (single-bit) QIS is a binary bit-plane, each constituent photodetector giving a value of 0 (for no photon detected) or 1 (at least one photon detected), with negligible read noise. These binary states ("jots" in QIS terminology) are then summed in space and/or time to form a spatio-temporally oversampled greyscale image frame. The flexibility in aggregating jots represents a distinct advantage in QIS imagers. For example, it has been recently shown that through signal-only aggregation, the effective sensitivity of a SPAD QIS device can be increased significantly in single molecule localisation microscopy [7].

Another potential application of QIS is in high-speed vision cameras. Indeed, it has been previously proposed that output bit-planes (which, taken individually, may present limited detail) could be shifted prior to summation so as to avoid motion blur [8]. The present paper gives the first demonstration of this idea, using image sequences taken by a SPAD implementation of QIS.

Existing schemes for removing motion blur typically rely on estimating the point spread function of the distortion, either using a sequence of (blurred) images [9] or using an additional sensor (an inertial sensor or a secondary image sensor [10]). De-convolution is then applied to the blurred images in post-processing. An alternative approach used in digital image stabilisation is to take a sequence of sub-exposures, then select, re-align and combine the sharpest exposures. This is complemented by the use of gyroscope-based optical image stabilisation in many cameras [11].

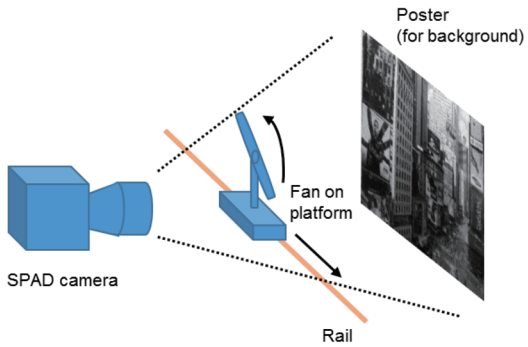

Figure 1. Experimental setup for imaging fan undergoing both linear and rotational motion.

The key attributes of the approach in this paper is that it is capable of removing motion blur from the whole field of view as well as tracking individual moving objects. Furthermore, it requires no secondary sensors (the scheme is wholly image processing-based), and uses a camera with back-to-back exposures, all of which are used to form the final image sequence (important for low-light conditions).

II. Algorithm

We demonstrate the algorithm using data captured by a 320×240, 10kfps SPAD camera [12] of a 5000rpm fan moving at 2m/s on a rail (Fig. 1.). An example raw bit-plane is shown in Fig. 2. The amount of detail that may be seen on individual bit-planes is limited, especially if the spatial resolution is relatively low, as is the case here. Due to the high speed of the fan, carrying out standard aggregation in time to produce video-rate image frames leads to significant motion blur, making the fan unrecognisable (Fig. 3.). The steps in the tracking-reconstruction algorithm are outlined in Fig. 4. A kernel (or "cubicle") of size (N_x,N_y,N_t), designed to be small enough to capture spatial and temporal variations in the image scene, is used to compose aggregated "test" frames (Fig. 5.). Assuming uniform light intensity across a cubicle, and following on from the analysis of [13], the value of each resulting pixel will be a binomial count with

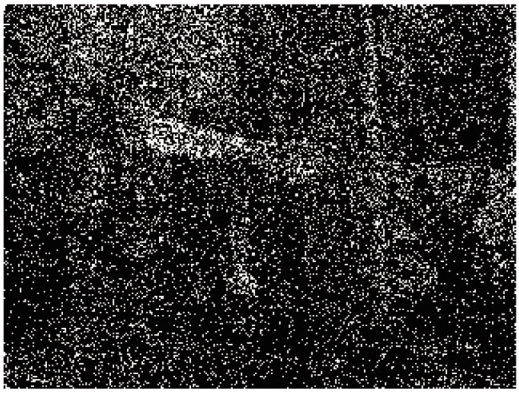

Figure 2. Single raw bit-plane of fan (exposure=2μs)

Figure 3. Sum of N=256 bit-planes (still frame from video rate image sequence)

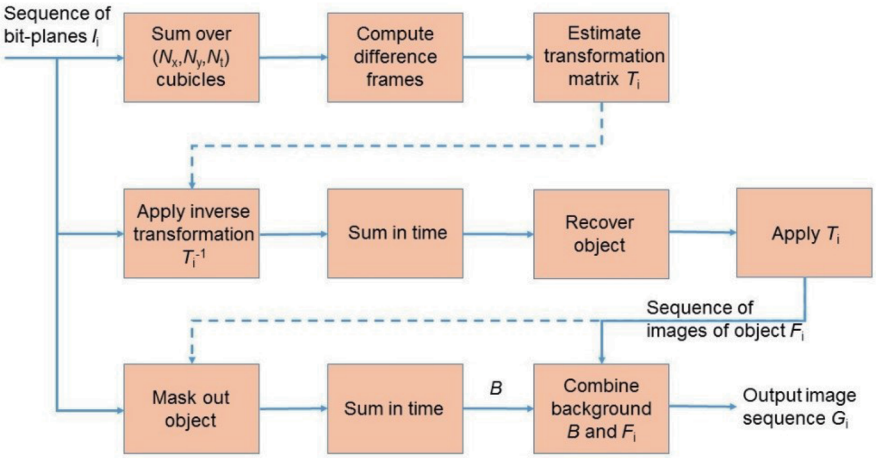

Figure 4. Block diagram indicating the steps in the tracking-reconstruction algorithm. The input is a sequence of bit-planes I_i capturing a high-speed object. The algorithm tracks the motion of this object (as defined by T_i) and outputs a higher bit-depth image sequence G_i.

a certain success probability $P_{x,y,t}$. Thus confidence bounds can be attached to each pixel, as to the "true" underlying photon flux, and "difference frames" generated, mapping statistically significant changes in pixel values in between consecutive "test" frames (Fig. 6.). We can thus detect moving objects, and under the assumption that these may be modelled as planar objects in 3D space, we estimate the transformation between "difference frames" to quantify the motion (as indicated in Fig. 6.; the estimated trajectory of the fan is shown in Fig. 7.). Note that whilst we have assumed a rigid transformation in this example (which can account for linear motion and rotation), we could have chosen a similarity transformation (also including scaling, i.e. the object moving closer to/further away from the camera) or a projective transformation (if there is a change in perspective).

The next step is to apply the inverse of this transformation to the original sequence of bit-planes, and carry out aggregation in time, which recovers the shape of the moving object (Fig. 7.). Based on the object outline, and its estimated motion, we carry out an additional sum of the bit-planes, untransformed but with the object masked out, so as to obtain the background. We can then combine the enhanced image of the object and that of the background, to compose high bit-depth images, at the native resolution of the camera, with no apparent motion artefacts (Fig. 8.).

The scheme can be readily extended to the case of multiple moving objects in the field of view. Fig. 9. shows an example bit-plane capturing a 1:43 scale toy car (moving with the rail) and a rotating fan. The illumination of the scene was adjusted to 10lux to simulate reasonably dark conditions, with the camera set to take back-to-back rolling shutter exposures to obtain a suitable signal level. As in the previous example, aggregating bit-planes so as to get a video-rate image sequence leads to substantial motion blur (Fig. 10.). Following the same algorithm as before and computing difference frames enables the motion of both the car and the fan blade to be established (Fig. 11). Carrying out sums in the respective frames of reference recovers the two objects, which are then combined with the re-constructed background for an effectively blur-free image sequence (Fig. 12.).

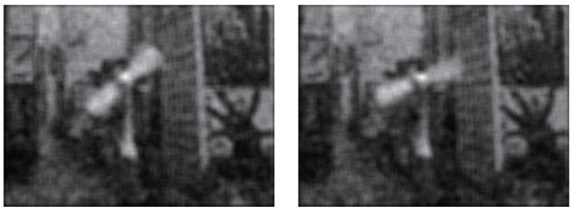

Figure 5. Consecutive "test" frames, created by aggregating bit-planes using a kernel of size (N_x,N_y,N_t)=(8,8,8) (overlapping in space but not time)

Figure 6. Consecutive "difference" frames. A value of '1' indicates "test frame" pixel values that have non-overlapping 95% confidence intervals in consecutive frames.

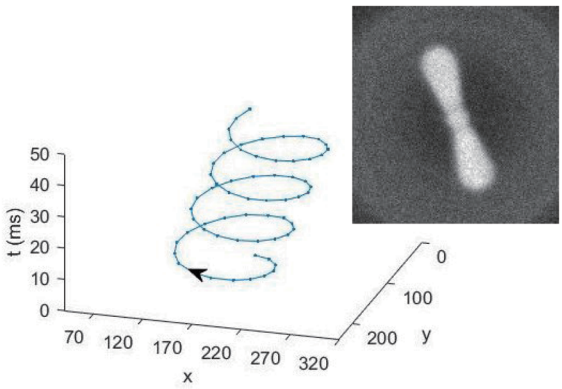

Figure 7. Reconstructed fan blade and its estimated trajectory

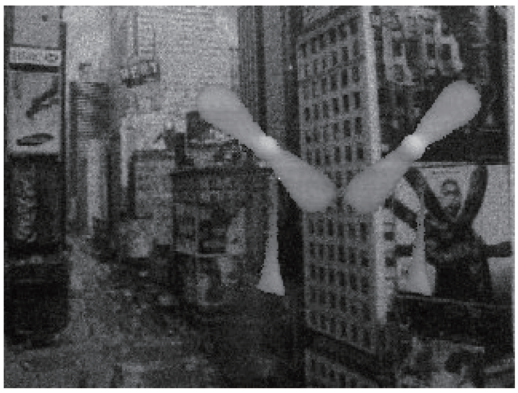

Figure 8. Two (overlaid) frames from output image sequence (75ms apart), both generated using N=256 bit-planes. We note the superior sharpness compared with Fig. 3., but same bit-depth and resolution.

We have also considered the case of global motion in the captured scene, for instance due to camera shake, and applied a similar approach as above for compensating for this motion. More specifically, the procedure is as follows. A sequence of "test" frames are created by aggregating bit-planes in time. The transformations that align these frames are computed (we used Matlab's `imregtform` function, and assumed a rigid transformation for this preliminary analysis). We then interpolate between the transformations associated with consecutive "test" frames to estimate the required alignment at the bit-plane level. Transforming the bit-planes accordingly, and aggregating for a second time, gives a motion-compensated image sequence. As an example, Fig 13. plots the extracted motion when the camera was intentionally shaken (by hand) in front of the test rig, whilst Fig 14. compares image frames before and after compensation. Image sharpness is seen to improve noticeably.

III. Conclusions

We have demonstrated methods for extracting and compensating for motion in a QIS camera. The methods lack the sophistication of tracking schemes with elaborate segmentation [14] or dynamic modelling (e.g. via a Kalman filter) [5], but nevertheless give an indication of the image quality improvement that can be achieved through appropriate processing of QIS bit-planes. Whilst currently implemented in post-processing software, an appropriate hardware implementation could potentially have real-time, low-light vision applications. In this context, an important benefit of composing images from a set of short sub-exposures is that rolling shutter can be used (giving high temporal aperture), whilst distortion effects are minimized.

Acknowledgments

This research was funded by the ERC TOTALPHOTON grant (grant no. 339747). The authors appreciate the support of STMicroelectronics who fabricated the device.

References

[1] Fossum, E.R. et al. "The Quanta Image Sensor: Every Photon Counts" Sensors 2016, 16(8), 1260.
[2] Ma, J. et al. "Characterization of quanta image sensor pump-gate jots with deep sub-electron read noise" IEEE J. EDS 2015, 3(6), 472-480.
[3] Antolovic, I.M. et al. "Photon-Counting Arrays for Time-Resolved Imaging" Sensors 2016, 16(7), 1005.
[4] Dutton, N.W.A, et al. "Single Photon Counting Performance and Noise Analysis of CMOS SPAD-Based Image Sensors" Sensors 2016, 16(7), 1122.
[5] Chen, B., et al. "Vision without the image" Sensors 2016, 16(4), 484.
[6] Chan, S.H., et al. "Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors" Sensors 2016, 16(11), 1961.
[7] Gyongy, I. et al. "Smart-aggregation imaging for single molecule localisation with SPAD cameras" Scientific Reports 2016, 6.
[8] Aull, B. "Geiger-Mode Avalanche Photodiode Arrays Integrated to All-Digital CMOS Circuits" Sensors 2016, 16(4), 495.
[9] Bascle, B. et al. "Motion deblurring and super-resolution from an image sequence" Computer Vision—ECCV'96, 571-582.
[10] Nayar, S.K. et. al. "Motion-based motion deblurring" IEEE trans. on pattern analysis and machine intelligence 2004, 26(6), 689-698.
[11] La Rosa, F. et al. "Optical Image Stabilization (OIS)" STMicroelectronics white paper 2015.
[12] Dutton, N.A.W. et al. "320x240 Oversampled Digital Single Photon Counting Image Sensor" IEEE VLSI Sym. 2014.
[13] Fossum, E.R. "Modeling the Performance of Single-Bit and Multi-Bit Quanta Image Sensors" IEEE J. EDS 2013, 1(9), 166-174.
[14] Bascle, B. et al. "Region tracking through image sequences" In IEEE Computer Vision 1995, 302-307.

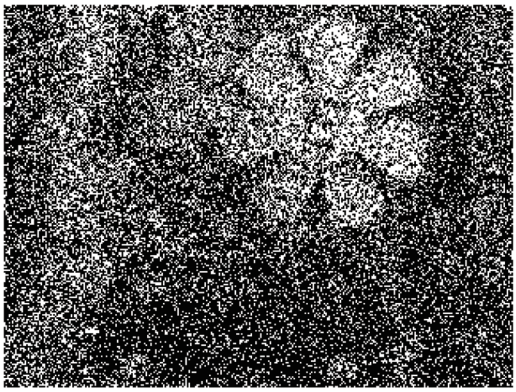

Figure 9. Single raw bit-plane of fan and car (exposure=100μs)

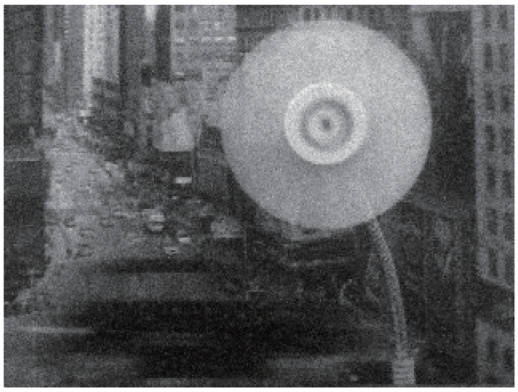

Figure 10. Sum of N=250 bit-planes (fan and car test). Interpolation has been applied over hot pixels.

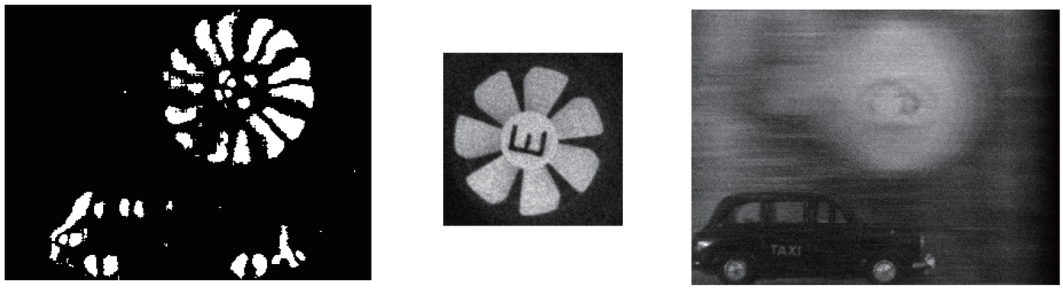

Figure 11. Detected motion in scene (left); recovery of fan blade (middle) and car (right) using N=250 bit-planes.

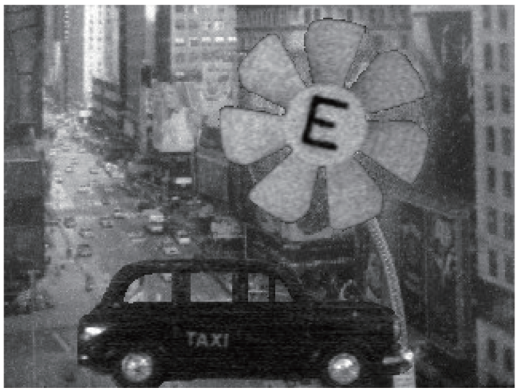

Figure 12. Frame from output image sequence for fan and car test

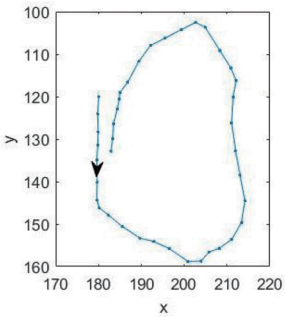

Figure 13. Extracted camera motion in camera shake test (indicated points are 25ms apart)

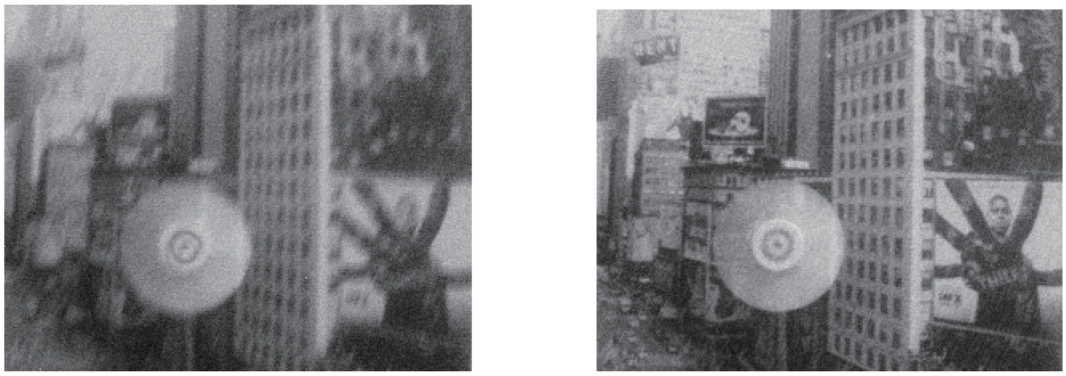

Figure 14. Image frames from camera shake test, original (left) and with compensation (right). N=250 in both cases.