

Real-time HDR video reconstruction for multi-sensor systems

Joel Kronander, Stefan Gustavson, Jonas Unger*
Linköping University

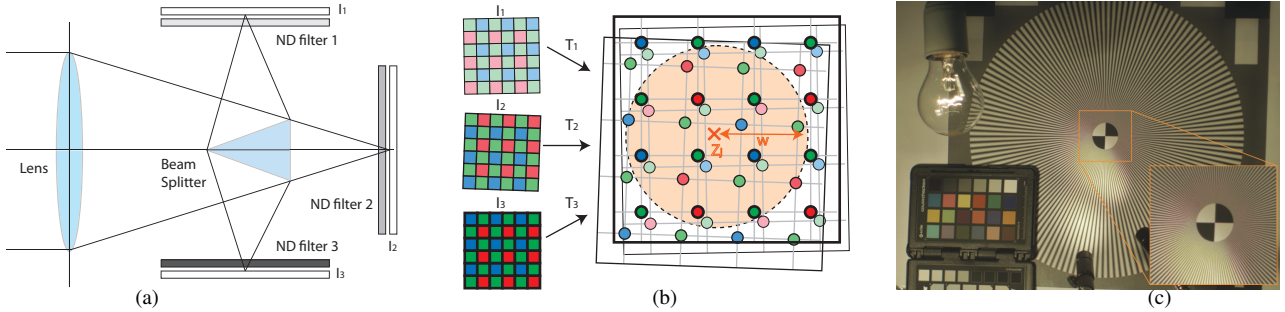


Figure 1: (a) Sketch of a typical multi-sensor HDR capturing system. (b) Assuming a known transformation sensor images can be warped to a common reference coordinate system, note that due to non-perfect pixel alignment, transformed sensor pixels will generally be irregularly distributed in the reference grid. The reconstruction at a point z_j uses measured sensor pixels from all sensors inside the support of a finite window function with radius w . (c) Shows a tonemapped HDR video frame reconstructed at an resolution of 2400x1700 using our method.

1 Introduction

HDR video is an emerging field of technology, with a few camera systems currently in existence [Myszkowski et al. 2008]. Multi-sensor systems [Tocci et al. 2011] have recently proved to be particularly promising due to superior robustness against temporal artifacts, correct motion blur, and high light efficiency. Previous HDR reconstruction methods for multi-sensor systems have assumed pixel perfect alignment of the physical sensors. In contrast to this, our method is specifically designed to handle the case of non-negligible misalignments between the sensors. This is beneficial as perfect alignment is very difficult to achieve in practice and that reflections in beam splitters may make it impossible to match the arrangement of the Bayer filters between sensors. Furthermore, while previous reconstruction techniques have considered HDR assembly, debayering and denoising as separate problems, our method is capable of simultaneous HDR assembly, debayering and smoothing of the data (denoising). The method is also general in that it allows reconstruction to an arbitrary output resolution and mapping. The algorithm is implemented in CUDA, and shows video speed performance for an experimental HDR video platform consisting of four 2400x1700 pixels high quality CCD sensors imaging the scene through a common optical system. ND-filters of different densities are placed in front of the sensors to capture a dynamic range of 24 f -stops.

2 Reconstruction

We treat the reconstruction of each HDR video frame, F , as a separate problem. Using N sensors, the input data for each frame consists of a set of raw images, $I_s(i, j)$ $s = 1 \dots N$, each with a possibly varying exposure time, t_s and ND-filter coefficient, n_s . We assume that there exists an affine transform, T_s relating each image to a virtual reference coordinate system, in which we seek to reconstruct the HDR output image. In practice we find the transforms, T_s , for each sensor matching the detected corners of a chessboard calibration target (implemented in OpenCV 2.1), for our system, we obtain a maximum reprojection error in the order of 0.1 pixels across the

image.

As a first step in the reconstruction, we perform shading correction for all sensor images, linearizing the sensor output to a common scale of reference. We then map all sensor images (pixel measurements) to the output coordinate system. This generally produces an irregular distribution of red, green and blue pixel samples, $y_{s,i,c}$ captured from sensors $s = 1 \dots N$, using a lexicographical ordering index i , see Figure 1(a). We now seek to reconstruct the output image in a regular grid, with spacings corresponding to the desired output resolution. For each output pixel location, z_j , we compute a locally weighted average of nearby pixel measurements, $y_{s,c,i}$, for each color channel $c = R, G, B$.

$$\hat{z}_{j,c} = \sum_s \sum_i \frac{w(j, s, i, c) y_{s,i,c}}{\sum_i w(j, s, i, c)} \quad (1)$$

We compute the weights based on two criteria, $w(j, s, i, c) = w_g(i, j) \cdot w_r(j, s, i, c)$. The first factor $w_g(j, i)$ is a windowing function of finite support, giving higher weights to spatially nearby samples, in our implementation we use a Gaussian function with the euclidean distance as argument. The second factor is a radiometric weight, $w_r(j, s, i, c)$. This weight is set according to a linear function of the raw digital input value with a small offset from the blacklevel level and saturation point. Our method thus automatically removes saturated pixels before reconstruction/debayering by setting their radiometric weight to zero.

Figure 1(b) shows a reconstructed HDR frame captured using our experimental HDR video platform. Using our CUDA implementation, real time video frame rates are achieved (running at a sustained rate of 26 fps) at an output resolution of 1168x876 pixels using a Nvidia GeForce 580 card.

References

- MYSZKOWSKI, K., MANTIUK, R., AND KRAWCZYK, G. 2008. *High Dynamic Range Video*. Morgan & Claypool.
- TOCCI, M. D., KISER, C., TOCCI, N., AND SEN, P. 2011. A Versatile HDR Video Production System. *ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH 2011)* 30, 4.

*e-mail: {joel.kronander, stefan.gustavson, jonas.unger}@liu.se