A Unified Framework for Multi-Sensor HDR Video Reconstruction

Joel Kronander[a,*], Stefan Gustavson[a], Gerhard Bonnet[b], Anders Ynnerman[a], Jonas Unger[a]

[a]*C-Research, Department of Media and Information Technology, Linköping University*
[b]*AG Spheron VR*

Abstract

One of the most successful methods for modern high quality HDR-video capture is to use camera setups with multiple sensors imaging the scene through a common optical system [1, 2, 3, 4]. However, such systems pose several challenges for HDR reconstruction algorithms. Previous reconstruction techniques have considered debayering, denoising, resampling (alignment) and exposure fusion as separate problems. We instead present a unifying approach, performing HDR assembly directly from raw sensor data. Our framework includes a camera noise model adapted to HDR video and an algorithm for spatially adaptive HDR reconstruction based on fitting local polynomial approximations to observed sensor data. The method is easy to implement and allows reconstruction to an arbitrary resolution and output mapping. We present an implementation in CUDA and show real-time performance for an experimental 4 Mpixel multi-sensor HDR video system. We further show that our algorithm has clear advantages over existing methods, both in terms of flexibility and reconstruction quality.

Keywords: HDR Video Capture, HDR Reconstruction, Local Polynomial Approximation, Camera Noise

1. Introduction

High dynamic range (HDR) video is an emerging field of technology, by many considered to be one of the key components of future imaging, that will enable a wide range of new applications in image processing, computer vision, computer graphics and cinematography. A few prototype HDR-video camera systems have recently been presented, showing that the computational power and bandwidth is now high enough to handle high resolution HDR-video processing and storage. Current CMOS and CCD imaging sensors are still unable to accurately capture the full dynamic range in general scenes. There are sensors with logarithmic response to incident light, that exhibit significantly higher dynamic range compared to standard CMOS and CCD sensors [5]. These are, however, in most applications still not accurate enough due to problems with image noise and that the entire dynamic range usually is quantized to 10-12 bit output. To date, the most successful approach for high quality HDR-video capture has instead been to use camera setups with multiple sensors imaging the scene through a common optical system [1, 2, 3, 4]. The idea is similar to exposure bracketing [6], but instead of varying the exposure time, the optical setup is designed to make the sensors capture different exposures of the scene by the use of different neutral density (ND) filters. Temporal artifacts such as ghosting and incorrect motion blur inherent to techniques where the exposure time is varied,

e.g. [7, 8], are avoided by capturing all exposures simultaneously. Although multi-sensor systems currently provide the (arguably) best alternative for high quality HDR video capture, there is a lack of a formalized framework for efficient and accurate image reconstruction for general sensor configurations.

In order to reconstruct HDR images from multi-sensor systems it is necessary to perform: *demosaicing* of color filter array (CFA) sampled data, *resampling* to correct for geometric misalignments between the sensors, *HDR assembly* to fuse the input low dynamic range (LDR) images into the output HDR image, and *denoising* to reduce image and color noise.

Most existing HDR reconstruction techniques, e.g. [6, 9, 10, 3], have considered this as separate steps in pipeline fashion, and performed demosaicing and realignment either before or after HDR assembly. This pipeline approach introduces several problems. Demosaicing before HDR assembly as in [6], causes problems with bad or missing data around saturated pixels. This is especially problematic as color channels usually saturate at different levels. Demosaicing after HDR assembly as in [9, 3] causes problems with blur and ghosting unless the sensors are perfectly aligned. For multi-sensor systems using high resolution sensors, it is problematic, costly, and sometimes even impossible to match the sensors so that the CFA patterns align correctly. Tocci et al. [3] report misalignment errors

*Corresponding Author

Email address: `joel.kronander@liu.se` (Joel Kronander)

around the size of a pixel despite considerable alignment efforts. This means that overlapping pixels from different sensors might not have the same spectral response. Also, treating demosaicing, resampling and HDR assembly in separate steps makes it difficult to analyze and incorporate sensor noise in the HDR reconstruction in a coherent way.

The main contribution of this article is a filtering framework that address all of these steps simultaneously in a formalized way. Our reconstruction method takes into account the entire set of non-saturated LDR samples available around the output HDR pixel, and performs demosaicing, resampling, HDR assembly and denoising in a single operation. The framework is based on spatially adaptive cross-sensor filtering, and uses a noise aware local polynomial approximation (LPA) approach adapted to include heterogeneous noise in the reconstruction. Using our framework, we present two flexible and robust algorithms for high quality HDR reconstructions using both isotropic filtering supports as well as adaptive anisotropic filtering supports taking correlations between the color channels into account. A key feature of our framework is it's computational speed, making it applicable to real-time HDR reconstruction for video applications. Using isotropic filtering, our CUDA implementation of the reconstruction algorithm runs at a sustained rate of 25 fps for input data from a 4Mpixel system with four sensors. As a part of our framework, we have also developed a new sensor noise model.

To demonstrate the usefulness of our approach, we evaluate the algorithms and noise model using both simulated data for a range of possible multi-sensor setups for HDR-imaging as well as on data from an experimental multi-sensor HDR-video camera prototype.

2. Related Work

To the authors knowledge, no previous approach has considered HDR reconstruction with joint realignment, noise reduction and CFA interpolation. While our method is not designed to provide optimal results with respect to denoising, demosaicing or super-resolution alone, we take a unified approach based on a model of sensor noise and provide a flexible framework which is straightforward to implement and use. Our method is also well suited to the needs of a versatile high dynamic range video system, by allowing an arbitrary resolution for the reconstruction and fast parallel processing. In comparison to our previous work [4], where we only consider isotropic filtering supports, we use adaptive anisotropic filtering supports to adapt the reconstruction to the local image structure. This also enables us to incorporate correlations between the color channels in the reconstruction, similar to modern demosaicing methods [11].

The algorithms presented in this paper is related to a large body of previous work, ranging from HDR capture and reconstruction, e.g. [12, 13], to theory and algorithms for accurate image reconstruction and image fusion, e.g. [14]. In this section, we give an overview of the previous work most closely related to the methods proposed in this paper.

2.1. HDR Video Capture

Nayar and Mitsunaga [15] proposed a HDR imaging system suitable for video capture, based on placing a neutral density (ND) filter array over the image sensor introducing a tradeoff between spatial resolution and dynamic range. While early methods only considered grayscale images the approach was also later generalized to include spectral sampling using a spatially varying exposure and color filter array [16, 17]. Solutions to the sampling and resolution enhancement of image dimensions such as spatial, spectral, temporal and intensity information have also previously been solved using multiple sensors for the same optical viewpoint. A simple example is the tri-color ccd system, where a beam splitter separates the color spectrum into three components (R, G, B) which is then measured individually on three different sensors. Aggarwal and Ahuja [1] present a system for HDR video capture using a mirror pyramid aligned with the optical axis to split incoming light and consider different exposure times or ND-filters for the sensors, or they considers mirrors with an offset from the optical axis to effectively distribute light unevenly over the sensors. Mcguire et al. [2] presented a framework for optimally selecting components for optical splitting trees given specific target goals, such as dynamic range, spectral sensitivity and cost budgets. Tocci et al. [3] recently presented a light efficient multi-sensor HDR video system utilizing up to 99.96% of the incident light, which have spawned a renewed interest in these systems. Hybrid approaches could also use a multi-sensor setup where one of the sensors includes a spatially varying ND-filter array. Our reconstruction method is general and applies to all of these general setups.

2.2. HDR fusion

The most common method for HDR reconstruction from a set of differently exposed Low Dynamic Range (LDR) images is to compute a per-pixel weighted average of the LDR measurements. The weights, often based on heuristics, are chosen to suppress image noise and remove saturated values from processing [25, 6, 26] . Weight functions can also be based on more sophisticated camera noise models. Mitsunaga and Nayar [27] derived a weight function that maximizes SNR assuming signal-independent additive noise. Kirk and Andersen [28] derived a weight function inversely proportional to the temporal variance of the digital LDR values. Granados et al. [10] later extended this approach to include both spatial and temporal camera noise. While most previous methods consider only a single pixel at a time from each LDR exposure, Tocci et al. [3] presented an algorithm that incorporates a neighbourhood of LDR samples in the reconstruction. This method

helps smooth the transition regions between sensors and performs well for sensors with large exposure differences, over 3 f-stops apart.

2.3. Non-parametric Image Processing

The last decade has seen an increased popularity of image processing operations using locally adaptive filter weights, for applications in e.g. interpolation, denoising and upsampling. Examples include normalized convolution [18], the bilateral filter [19], and moving least squares [20]. Recently, deep connections have been shown [21, 22] between these methods and traditional non-parametric statistics [23]. In this paper, we fit Local Polynomial Approximations (LPA) [14] to irregularly distributed samples around output pixels using a localized maximum likelihood estimation [24] to incorporate the heterogeneous noise of the samples.

2.4. Image Fusion

When multiple images of the same scene are captured within sub-pixel displacements of each other, super-resolution reconstruction methods can be applied. A class of computationally simple super-resolution techniques are those based on interpolation in a common high resolution grid [29]. These methods typically first use some registration to put observed images in a common high resolution reference frame. The high-resolution image is then estimated in a least squares sense given a local neighbourhood of observed pixels [30, 31]. As noted in previous work, super-resolution and HDR imaging complement each other. Super-resolution increases the resolution in the spatial domain, and HDR increases the radiometric range and resolution. Several methods have also been proposed to incorporate both enhancements in the same processing framework [32, 33].

2.5. Demosaicing

Most modern camera systems acquire color images using a sensor covered by a Color Filter Array (CFA), and estimates the final color vector for each pixel from neighboring color samples. This process is known as demosaicing. The most frequently used CFA of today is the RGB Bayer pattern [34], sampling the G channel on a quincunx grid and the R and B images on a rectangular at half the resolution compared to the G channel. Several demosaicing methods have been developed through the years. A conperhensive overview can be found in Gunturk *et al.* [11]. Joint denoising and demosaicing have been considered in previous work [35, 36]. Using a global maximum a posteriori optimisation with priors that enforce correlation between color channels, Farsiu et al. [37] proposed a framework for joint super-resolution and demosaicing. A similar approach was taken in [38] for HDR imaging with joint demosaicing for sensors with non-destructive readout (ie perfect alignment between consecutive readouts/frames). While a global MAP optimization could also be used for joint HDR
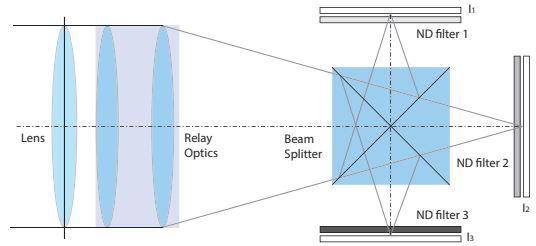

Figure 1: Sketch of a general multi-sensor setup. The incident light is first propagated through relay optics to a beamsplitter, splitting the optical path into S components. Each sensor, I_s receives a fraction, n_s of the incident light depending on the beamsplitter and ND-filter configuration.
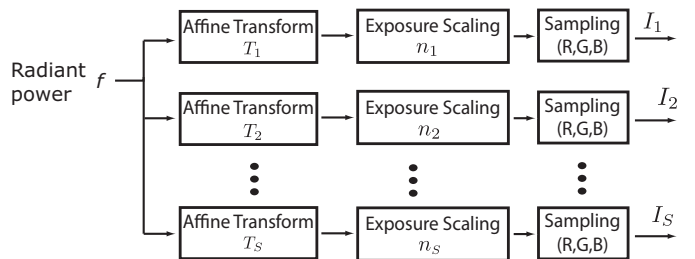

Figure 2: The image formation model used to describe the HDR capture.

fusion, realignment and demosaicing, it's computationally demanding and does not straightforwardly allow for parallel processing, making it less suitable for the HDR video applications considered in this paper.

3. Problem formulation

The goal of the algorithm presented in this paper is to generate a video stream of output HDR images based on input data from multi-sensor imaging systems, that for each output frame capture a set of N differently exposed raw CFA images, I_s where $s = 1...N_s$. The algorithm is designed for multi-sensor systems where the sensors image the scene through a common front lens, and where the light path is split onto the sensors by a beam splitter arrangement. An example multi-sensor setup is displayed in Figure 1. We assume that misalignments between the sensors can be described as 2-D transformations, T_s, so that the images can be registered to the same frame of reference, and that the difference in exposure between the sensors, created by e.g. neutral density (ND) filters placed in the ray path, for each sensor can be described as an exposure scaling coefficient n_s. Each pixel in the input images, I_s, provides a measurement of the incident radiant power at the pixel location, scaled by an unknown factor due to the quantum efficiency of the sensor, [39]. In this work we assume that the sensors have the same quantum efficiency or that it can be measured separately for each sensor so that
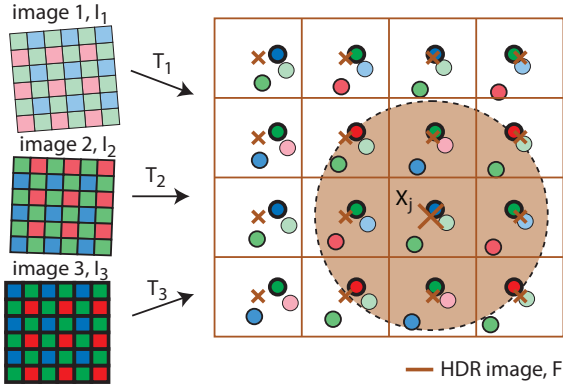
Figure 3: HDR reconstruction is performed to a virtual reference system with arbitrary resolution and mapping. The sensor samples are mapped to the reconstruction grid using the transformations T_s. The pixel at position X_j in the HDR image is estimated using a local polynomial approximation of nearby samples (within the shaded circle). Saturated and defective pixels are discarded, and samples near the black level, local outliers and samples farther from the reconstruction point contribute less to the radiance estimate.

it can be incorporated into the exposure scaling coefficient n_s.

This yields the image formation model described in Figure 2. Each sensor image, I_s, samples the incident radiant power, f, at a set of discrete pixel locations. Using a linear index i for pixels in each sensor image, we define the measured digital sample value at a pixel i in image I_s as $y_{s,i}$. The samples, $y_{s,i}$, contain measurement noise that is dependent on the incident radiant power, $f_{s,i}$, as well as the sensor characteristics.

During capture, we assume that the sensors are synchronized or that the scene is static, so that the motion blur characteristics are the same for all sensors. The parameters that are allowed to vary between the sensors are: the exposure time t_s, the sensor gain setting g_s, and the exposure scaling coefficient, n_s.

4. Calibration

Before estimating the HDR frames, we perform radiometric and geometric calibration of the sensors.

4.1. Radiometric Calibration

The (noisy) digital pixel values $y_{s,i}$ in the input images, I_s, are first converted to radiant power estimates $\hat{f}_{s,i}$. This is carried out by subtracting a bias frame and taking into account the exposure settings (exposure time, exposure scaling, global gain) and per-pixel gain variations. The variance of the radiant power, $\sigma^2_{\hat{f}_{s,i}}$, can then be estimated using a calibrated radiometric camera noise model. Similarly to [10], we assume the noise to follow a Gaussian

distribution. This assumption is described in detail in Section 6, where we give an in depth overview of our radiometric noise model adapted for HDR video. It should be noted that our framework can use any radiometric noise model that assumes an approximately Gaussian noise distribution.

4.2. Geometric Alignment

The HDR reconstruction is performed in a virtual reference space, corresponding to a virtual sensor placed somewhere in the focal plane. The virtual sensor dimensions are chosen to reflect the desired output frame, unconstrained by the resolution of the input frames. Using a geometric calibration procedure, an affine transform, T_s, is established for each sensor which maps sensor pixel coordinates, $x_{s,i}$, to the coordinates of the reference output coordinate system, $X_{s,i} = T_s(x_{s,i})$. In general, the transformed input pixel coordinates, $X_{s,i}$, are not integer-valued in the output image coordinates, and for general affine transformations the sample locations $X_{s,i}$ will become irregularly positioned. An example for three sensor images is shown in Figure 3.

5. HDR Reconstruction

To reconstruct the HDR output frame we estimate the relative radiant power for each pixel in the virtual sensor separately in a non-parametric fashion using the transformed samples $\hat{f}_{s,i}(X_{s,i})$. For each output pixel j with integer valued, regularly spaced coordinates X_j, we fit a local polynomial model to observed samples $\hat{f}_{s,i}(X_{s,i})$ in a local neighborhood, see Figure 3. We first discuss how a single color channel, $c = \{R||G||B\}$, can be reconstructed independently of the other channels using LPA with isotropic filtering supports. In section 5.4 we then show how the first LPA estimate can be improved by using adaptive anisotropic filtering supports. Using adaptive filtering supports also enables us to utilize cross-correlation between color channels to improve the estimate. This is discussed in section 5.5.

5.1. Local Polynomial Model

To estimate the radiant power, $f(x)$, at an output pixel we use a generic local polynomial expansion of the radiant power around the output pixel location $X_j = [x_1, x_2]^T$. Assuming that the radiant power $f(x)$ is a smooth function in a local neighborhood around the output location X_j we can use an *M-th* order Taylor series expansion to predict the radiant power of a point X_i close to X_j as:

$$\widetilde{f}(X_i) = C_0 + C_1(X_i - X_j) \\ + C_2 tril\{(X_i - X_j)(X_i - X_j)^T\} \quad ... \quad (1)$$

where *tril* lexicographically vectorizes the lower triangular part of a symmetric matrix and where

$$C_0 = f(X_j) \qquad (2)$$

$$C_1 = \nabla f(X_j) = \left[\frac{\partial f(X_j)}{\partial x_1}, \frac{\partial f(X_j)}{\partial x_2} \right] \qquad (3)$$

$$C_2 = \frac{1}{2} \left[\frac{\partial^2 f(X_j)}{\partial x_1^2}, 2 \frac{\partial^2 f(X_j)}{\partial x_1 \partial x_2}, \frac{\partial^2 f(X_j)}{\partial x_2^2} \right] \qquad (4)$$

Given the fitted polynomial coefficients, $C_{1:M}$, we can thus predict the radiant power at the output location X_j by $C_0 = f(X_j)$, and the first order gradients by C_1. In this work we only consider local expansions of order $M \leq 2$.

5.2. Maximum Localized Likelihood Fitting

To estimate the coefficients, C, of the local polynomial model, we maximize a localized likelihood function [24] defined using a smoothing window centered around the output location X_j:

$$\mathcal{W}_H(X) = \frac{1}{det(H)} \mathcal{W}(H^{-1}X) \qquad (5)$$

where H is a 2×2 smoothing matrix that determines the shape and size of the window. In section 5.3.1, we discuss how the shape and size of the window function and smoothing can be selected. To lessen the notational burden, the observed samples in the neighborhood, $\{\hat{f}_{s,i}(X_{s,i}) : X_{s,i} \in \mathrm{supp}\left(\mathcal{W}_H(X)\right)\}$, are denoted below by f_k with a linear index $k = 1...K$. Using the assumption of normally distributed radiant power estimates f_k, see Section 6, the log of the localized likelihood for the polynomial expansion centered at X_j is given by

$$L(X_j, C) = \sum_k \log(N(f_k | \widetilde{f}(X_k), \hat{\sigma}_{f_k}^2) \mathcal{W}_H(X_k))$$

$$= \sum_k \left[f_k - C_0 - C_1(X_i - X_j) \right.$$

$$\left. - C_2^T tril\{(X_i - X_j)(X_i - X_j)^T\} - ... \right]^2 \frac{\mathcal{W}_H(X_k)}{\hat{\sigma}_{f_k}} + R$$

where R represents terms independent of C. The polynomial coefficients, $\tilde{C}$, maximizing the localized likelihood function is found by the weighted least squares estimate

$$\tilde{C} = \underset{C \in \mathcal{R}^M}{\mathrm{argmax}}(L(X_j, C))$$
$$= (X_x^T W X_x)^{-1} X_x^T W \bar{f} \qquad (6)$$

where

$$\bar{f} = [f_1, f_2, ... f_K]^T$$

$$W = diag[\frac{\mathcal{W}_H(X_1)}{\hat{\sigma}_{f_1}}, \frac{\mathcal{W}_H(X_2)}{\hat{\sigma}_{f_2}}, ..., \frac{\mathcal{W}_H(X_K)}{\hat{\sigma}_{f_K}}]$$

$$X_x = \begin{bmatrix} 1 & (X_1 - X_j) & tril^T\{(X_1 - X_j)(X_1 - X_j)^T\} & ... \\ 1 & (X_2 - X_j) & tril^T\{(X_2 - X_j)(X_2 - X_j)^T\} & ... \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (X_K - X_j) & tril^T\{(X_K - X_j)(X_K - X_j)^T\} & ... \end{bmatrix}$$
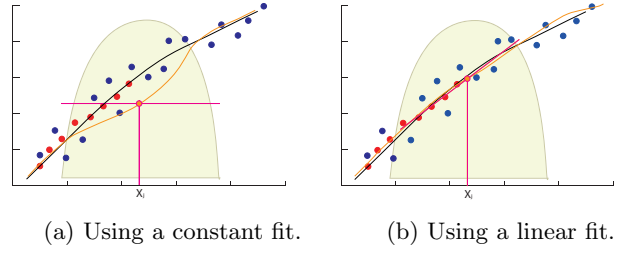

(a) Using a constant fit. (b) Using a linear fit.

Figure 4: 1D example showing the effect of the polynomial order, M, in areas affected by sensor saturation. The red circles represents a sensor saturating at lower values than the sensor represented by the blue circles. a) Using a constant fit, $M = 0$, introduces significant bias as more unsaturated measurements are available to the left of the reconstruction point. b) Using a linear fit, $M = 1$, the bias is reduced drastically.

Note that similarly to the kernel regression framework [21], using equivalent kernels, the solution to the weighted least squares problem can also be formulated as a locally adaptive linear filtering.

5.3. Parameters

The expected mean square error of the reconstructed image depends on a trade-off between bias and variance of the estimate. This trade-off is determined by: the order of the polynomial basis M, the window function $\mathcal{W}$, and the smoothing matrix H.

Using a piecewise constant polynomial, $M = 0$, the estimator corresponds to an ordinary locally weighted average of neighboring sensor observations

$$\hat{f}(X_j) = \frac{\sum_k \frac{w_i}{\hat{\sigma}_{f_k}} f_k}{\sum_k \frac{w_i}{\hat{\sigma}_{f_k}}} \qquad (7)$$

However, locally weighted averages can exhibit severe bias at image boundaries and in regions around sensor saturation, due to the asymmetry of the number of available sensor measurements in these locations. By instead fitting a linear polynomial (a plane, $M = 1$), the bias can be reduced significantly, see Figure 4. Introducing higher order polynomials is possible, but may lead to increased variance in the estimates. In this work, we only consider $M \leq 2$.

5.3.1. Window and scale selection

In this paper we only show results using a Gaussian window function

$$\mathcal{W}_H(X_k) = \frac{1}{2\pi det(H)} exp\left\{ -(X_k - X_j)^T H^{-1}(X_k - X_j) \right\} \qquad (8)$$

There are also other possible choices of symmetric smooth window functions, e.g. Epanechnikov or Tricube windows [14].

5

The 2×2 smoothing matrix, H, affects the shape of the gaussian. It is important that the support of the smoothing window is extended to incorporate enough samples into the estimate in Equation (6) so that $(X_x^T W X_x)$ is invertible (full rank). The simplest choice of the smoothing matrix is $H = hI$, where h is a global scale parameter and I is the identity matrix. This corresponds to an isotropic filter support. The choice of the scale parameter, h, is dependent on the sensor characteristics (noise) and the scene. We therefore treat h as a user parameter. To reconstruct sharp images in low noise conditions, we generally choose the smallest possible scale parameter, h, such that $(X_x^T W X_x)$ is invertible. However, if the captured images exhibit severe noise, a larger scale parameter may be beneficial. For Bayer CFA sensors, we consider different scale parameters for the green and red/blue color channels, $h_G = \frac{h_{R,B}}{\sqrt{2}}$, as there are more green samples per unit area. For the remainder of this paper we will refer to the reconstruction method using a global smoothing matrix, $H = hI$, as *Local Polynomial Approximation LPA*.

5.4. Adaptive LPA Reconstruction

The truncated local polynomial expansion assumes that the radiant power, $f(x)$, is a smooth function (up to order M) in a neighborhood of the reconstruction point X_j. In many image regions these assumptions are not valid. For example, if an output pixel is located near a sharp edge, it cannot be represented accurately with a finite polynomial expansion. It is therefore desirable to adapt the window support so that it only includes observations on the same side of the edge.

To adapt the support of the local polynomial model to the data, we use a two step approach: First, we use LPA with $M \geq 1$ and isotropic window supports to compute an initial estimation of the local gradients, $\left[\frac{\partial f(X_i)}{\partial x_1}, \frac{\partial f(X_i)}{\partial x_2} \right]$. Using the estimated gradients, we then locally adapt the smoothing matrix, H_j, to the local structure of the signal at each output pixel X_j.

We define an oriented, anisotropic smoothing matrix as $H_j^s = hC_j$ where C_j represents the local gradient covariance of the signal estimated from the data. To estimate C_j, we use a parametric approach similar to Takeda et al. [21]. Specifically, we consider a decomposition of the covariance matrix as:

$$C_j = \gamma_j U_{\theta_j} \Lambda_j U_{\theta_j} \tag{9}$$

$$U_{\theta_j} = \begin{bmatrix} cos(\theta_j) & sin(\theta_j) \\ -sin(\theta_j) & cos(\theta_j) \end{bmatrix} \tag{10}$$

$$\Delta_j = \begin{bmatrix} \sigma_j & 0 \\ 0 & \frac{1}{\sigma_j} \end{bmatrix} \tag{11}$$

This corresponds to describing the covariance matrix as a function of three parameters: σ_j describing a elongation along the principal directions, θ_i describing a rotation angle and γ_i describing a overall scaling. The elongation,

rotation and scaling parameters are estimated from a truncated singular value decomposition of weighted gradients in a local neighborhood around the output location X_j:

$$G = \begin{bmatrix} \vdots & \vdots \\ f^{x_1}(X_i) & f^{x_2}(X_i) \\ \vdots & \vdots \end{bmatrix} = U_j S_j V_j \quad \forall X_i \in supp\{w_i\} \tag{12}$$

where $f^{x_1}(X_i)$ and $f^{x_2}(X_i)$ are the estimated gradients along x_1 and x_2 at the nearby point X_i, and w_i represents a local window function. The singular values s_1 and s_2 along the diagonal of S_j represent the energy along the dominant directions in the neighborhood defined by the local window function. Thus, the orientation angle is set so that the window spreads orthogonal to the dominant gradient direction, i.e. from the second column of V_j, $v = [v_1, v_2]^T$ as

$$\theta_j = arctan\left(\frac{v_1}{v_2}\right) \tag{13}$$

The elongation parameter, σ_j is then set according to:

$$\sigma_j = \frac{s_1 + \lambda_1}{s_2 + \lambda_1}, \lambda_1 \geq 0 \tag{14}$$

where λ_1 is a regularization parameter. The scaling parameter, γ_j, is finally set according to

$$\gamma_i = \left(\frac{s_1 s_2 + \lambda_2}{M}\right)^{\alpha} \tag{15}$$

where λ_2 is another regularization parameter, α is the structure sensitivity parameter, and M is the number of samples in the local neighborhood of the gradient analysis window w. The intuition for the shape of the elongation and scaling parameters is that the window support should be: circular and relatively large in flat areas, elongated along edges, and small in textured areas to preserve detail. For all results presented in this paper we fix the regularization parameters to $\lambda_1 = 1$ and $\lambda_2 = 0.001$. The structure sensitivity parameter, α, is treated as a user specified parameter,representing a tradeoff between denoising and detail preservation. For a more detailed analysis of the choice of parameters, we refer the reader to Takeda et al. [21] and Milanfar [22]. Note that as we are primarily interested in irregularly sampled data we choose to adapt the window functions instead of the support of each observation (kernel) as in the work of Takeda *et al.*.

5.5. Color Channel Correlation

Modern demosaicing methods consider the correlation between color channels to improve CFA interpolation. Commonly used heuristics are that edges should match between color channels and that interpolation should be performed along and not across edges. For Bayer pattern CFA arrays, in which the G channel is sampled more

densely than the R and B channels, a common method is to first reconstruct the G channel, then use the gradients of the G channel when reconstructing the R and B channel.

To take into account the correlation between color channels in our reconstruction, we adapt this idea to the irregularly sampled input data from multi-sensor camera systems. First we estimate the gradients for the G channel using LPA with $M \geq 1$ and isotropic window supports. We then use the G channel gradients to locally adapt the the smoothing matrix H_j as described in the previous section. Using the adapted smoothing matrix H_j, we then compute the radiant power estimates for all color channels, (R, G, B). This effectively forces the interpolation to be performed along and not across edges and preserves their location between the color channels. We refer to this reconstruction method as *Color Adaptive Local Polynomial Approximation (CALPA)*.

6. Sensor noise model

In this section we describe a radiometric camera noise model adapted to HDR video. We assume a linear digital response and model the noise using a radiometric camera model taking into account the exposure times, gain settings and ND-filter coefficients for the different sensors. Our noise model is inspired by previous methods considering radiometric camera calibration [40, 41] and optimal HDR capture methods [10, 39].

Each observed digital pixel value, $y_{s,i}$, is obtained using an exposure time, t_s, and an exposure scaling, n_s that is assumed constant over the image I_s. The pixel response is a measurement of the radiant power reaching the image sensor, which we for convenience express as the number of photo-induced electrons collected per unit time, $f_{s,i}$. The accumulated number of photoelectrons, $e_{s,i}$, collected at the pixel during the exposure time, t_s, follows a Poisson distribution with expected value

$$E[e_{s,i}] = t_s(a_{s,i}n_sf_{s,i}) \qquad (16)$$

and variance $Var[e_{s,i}] = E[e_{s,i}]$ where $a_{s,i}$ is a per pixel factor due to non-uniform photo-response. The recorded digital value, $y_{s,i}$, is also affected by the sensor gain and signal independent readout noise,

$$y_{s,i} = g_s(e_{s,i}) + r_{s,i}(g) \qquad (17)$$

where g_s is the analog amplifier/sensor gain (proportional to the native ISO settings on modern cameras) and $r_{s,i}(g)$ is the readout noise, that generally depends on the gain setting, g_s, of the sensor. As we focus on video applications with frame rates of 25 fps or more, the dark current noise is neglected. For most modern camera sensors dark current noise has negligible effect for exposures less than a second [39]. In contrast to previous work [39?] we don't assume a simple parametric model for the readout noise dependence on the gain, as we have found that such

models don't generally provide a satisfactory fit to measured readout noise at different gain settings for modern camera sensors. To handle sensors with different gain settings, we instead calibrate the readout noise, $r_{s,i}(g)$ for each separate gain/ISO setting considered.

Before reconstruction of the HDR output frame, each digital pixel value, $y_{s,i}$, is independently transformed to an estimate of the number of photoelectrons reaching the pixel per unit time, $\hat{f}_{s,i}$. To compensate for the readout noise bias (blacklevel), $E[d_{s,i}(g,t)]$, we subtract a *bias frame*, $b_{s,i}$, from each observation. The bias frame is computed as the average of a large set of black images captured with the same camera settings as the observations but with the lens covered, so that no photons reach the sensor. The radiant power $f_{s,i}$ can then be estimated as

$$\hat{f}_{s,i} = \frac{y_{s,i} - b_{s,i}}{g_s t_s a_{s,i} n_s} \qquad (18)$$

6.1. Variance estimate

We assume $\hat{f}_{s,i}$ to follow a normal distribution with mean $f_{s,i}$ and standard deviation $\sigma_{\hat{f}}$. For low light levels, photon shot noise is generally dominated by signal independent readout noise approximately following a normal distribution, and for brighter areas, with a high photon count, the Poisson distributed shot noise is well approximated by a normal distribution. Assuming no saturated or clipped pixels values, the variance of $\hat{f}_{s,i}$ is given by

$$Var[\hat{f}_{s,i}] = \sigma_{\hat{f}}^2 = \frac{g_s^2(Var[e_{s,i}]) + Var[r_{s,i}(g,t)]}{g_s^2 t_s^2 a_{s,i}^2 n_s^2} \qquad (19)$$

Due to the photon shot noise, to estimate, $Var[e_{s,i}]$, we should ideally know the true value of $f_{s,i}$, however in practice we are instead forced to use the (noisy) estimate, $\hat{f}_{s,i}$, obtained from equation (18). Thus we form an approximate variance estimate, $\hat{\sigma}_{\hat{f}_{s,i}}^2$, as

$$\hat{\sigma}_{\hat{f}_{s,i}}^2 = \frac{g_s^2 t_s a_{s,i} n_s \hat{f} + Var[r_{s,i}(g,t)]}{g_s^2 t_s^2 a_{s,i}^2 n_s^2} \qquad (20)$$

The above analysis assumes non-saturated pixels. In practice, we exclude saturated pixels by comparing each digital pixel value, $y_{s,i}$, to a *saturation frame*. The model does not include the effect of pixel cross-talk, and the variances, $\sigma_{\hat{f}_{s,i}}^2$, are assumed to be independent of each other. More complicated models of sensor noise also include the effects of saturation on the digital value variance, $Var[y_{s,i}]$, see e.g. [42]. However such models assume $y_{s,i}$ to follow a clipped normal distribution which requires an inpractical non-linear estimation when computing the coefficients for the local polynomial approximations, discussed in section 5.

6.2. Parameter calibration

The exposure scaling, n_s, and per-pixel non uniformity, $a_{s,i}$, can in practice be estimated using a flat field image

computed as the average over a large sequence of images, running the sensors at different exposure times ensures well exposed pixels for all sensors. Choosing one sensor as reference, the exposure scaling, n_s, can be estimated as the spatial mean of the ratio, computed pixel wise between the radiant power estimates, Eq. 18, of the sensors. The non uniformity, $a_{s,i}$, can then be computed as the per pixel deviation from the exposure scaling, n_s.

The variance of the readout noise, $Var[r_{s,i}(g,t)]$ can be estimated, similarly to the bias frame, from a set of black images, captured with the same camera settings, (g, t), as the observations but with the lens covered, so that no photons reach the sensor.

The sensor gain, g_s, can be calibrated using the relation,

$$\frac{Var[y_{s,i}] - Var[b_{s,i}]}{E[y_{s,i}] - E[b_{s,i}]} = \frac{g_s^2 Var[e_{s,i}]}{g_s E[e_{s,i}]} = g_s \qquad (21)$$

where the second equality follows from $e_{s,i}$ being Poisson distributed shot noise with $E[e_{s,i}] = Var[e_{s,i}]$. $E[y_{s,i}]$ and $E[b_{s,i}]$ can be estimated by averaged flat fields and the bias frame respectively, and $Var[b_{s,i}]$ as described above.

7. Algorithm evaluation

To evaluate our algorithm, we compare its performance against several reconstruction methods. As most previously proposed algorithms do not generalize to setups with arbitrary misalignments between the sensors, we include evaluations on setups where the sensors are assumed to be perfectly aligned as well as setups where the sensors only have a translational offset from each other.

To evaluate our algorithm, we compare its performance against the follow methods:

- *Tocci et al.* : The recent real-time multi-sensor reconstruction algorithm presented by Tocci et al. [3].

- *Debayer-first* : A method performing demosaicing before HDR assembly, using heuristic linear radiometric weights.

- *Debayer-last:* A method performing demosaicing after HDR assembly, using radiometric weights inversely proportional to the variance estimate $\hat{\sigma}_{\hat{f}}^2$ given by our sensor noise model 6.

- *Granados et al.* : The iterative method proposed by Granados et al. [10] for high-quality HDR assembly of perfectly aligned data.

For the above methods we consider demosaicing using both bilinear interpolation and the gradient adaptive method presented by Malvar et al. [43].

We base the majority of our comparisons on HDR image data generated from a camera simulation framework. The camera simulation framework allows us to generate images of arbitrary bit depth, sensor misalignments, exposure settings and noise characteristics according to the


(a) LPA, $M = 1$, $h_{R/B} = 0.7$ (b) Granados *et al.* [10]

Figure 6: Reconstructed results from four perfectly aligned exposures 4 f-stops apart captured with a Canon 40D camera. (a) LPA Reconstruction and (b) the iterative method of Granados *et al.* [10].

model described in Section 6. In the simulator, the scene is represented by a virtually noise free high resolution HDR image, meticulously generated from a set of exposures captured 1 *f*-stop apart covering the dynamic range of the real scene. Each exposure is computed as an average of 100 frames, in this case captured with a Canon 5D SLR camera. For our experiments, we simulate different multi-sensor setups with different sensor misalignments, T_s, and different noise characteristics, given the simulation parameters $g_s, n_s, t_s, Var[r_{s,i}]$. We focus our evaluation on simulated sensors with three to four f-stops difference between the sensor exposures as this is the most common setup for multi-sensor HDR video systems in practice [3].

7.1. Perfect alignment

In Figure 5 the local polynomial approximation (LPA) and color adaptive local polynomial approximation (CALPA) methods are compared to the *Debayer-first* method and the *Debayer-last* method. The virtual sensors were perfectly aligned, captured with exposure scalings $n_1 = 1, n_2 = 2^{-4}, n_3 = 2^{-8}$ and set to simulate the noise properties of a Canon 5D sensor, with parameters: $g_s = 0.23$, and $Var[r_{s,i}] = 6.5$ as reported by [10]. The Debayer-first method perform poor for high frequency detail, introducing significant color artifacts. The Debayer-last method perform slightly better for Bilinear CFA interpolation, and using the gradient adaptive method of Malvar et al. [43] effectively suppresses color moire artifacts in low contrast regions, however in high contrast regions, such as close to the light bulb filament, significant color artifacts are introduced. Compared to the debayer first method, the LPA reconstruction reduces color artifacts by performing resampling and CFA interpolation jointly with reconstruction. LPA reconstruction also gracefully handles transitional regions around sensor saturation points, as a smoothing spatial filtering and a cross-sensor blending are both inherent in the reconstruction. This has the benefit of implicitly performing denoising. Using anisotropic window supports, correlated between the color channels, CALPA produces even smoother results in high contrast regions, e.g. close to the filament. CALPA also produces less color artifacts than LPA in regions with high spatial frequency.
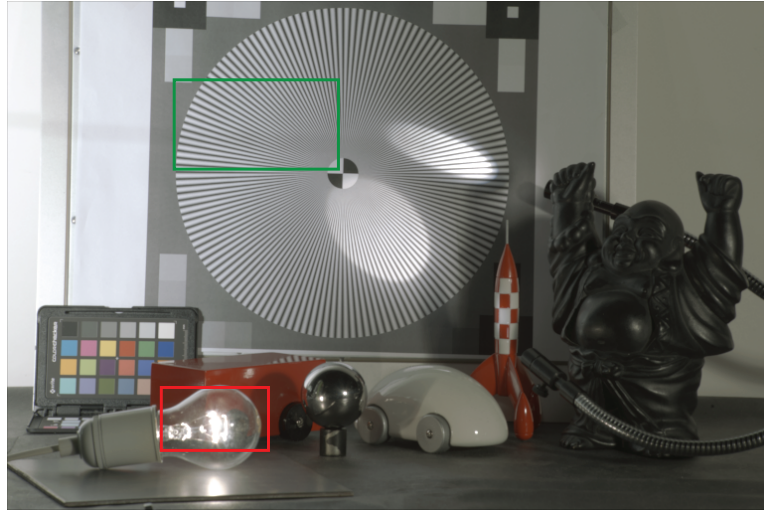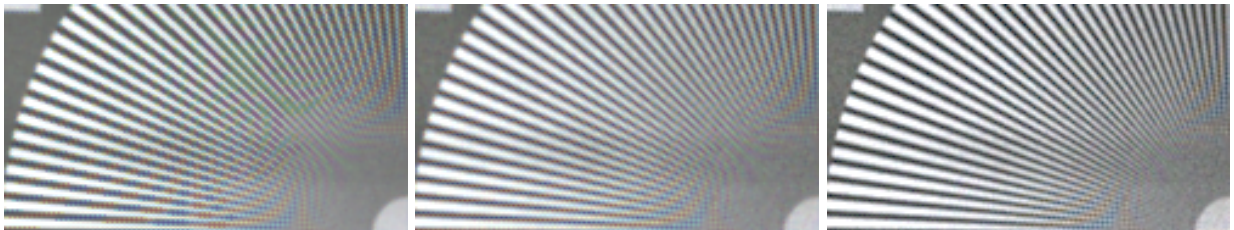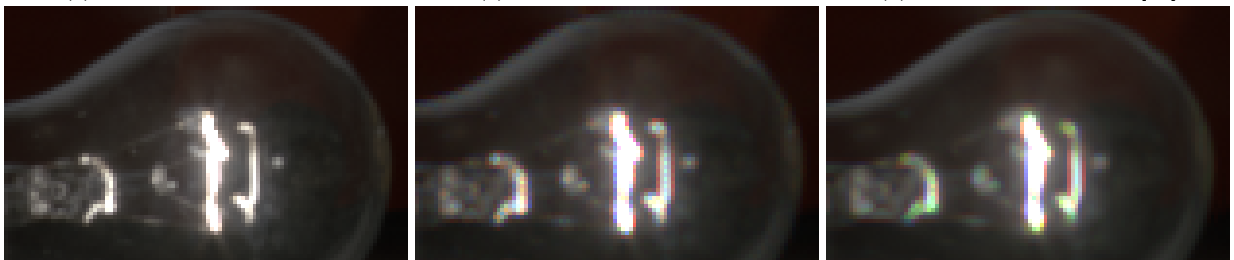
(a) HR Reference
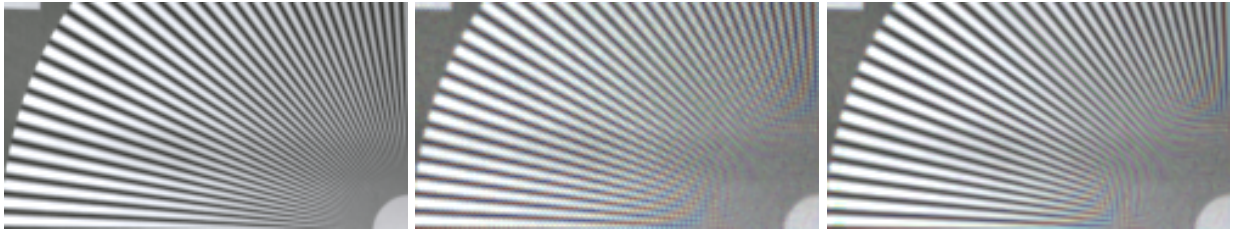
(b) Reference

(d) LPA

(f) Color Adaptive LPA

(c) Debayer First, Bilinear

(e) Debayer Last, Bilinear

(g) Debayer Last, using [43]

(h) Reference

(j) LPA

(l) Color Adaptive LPA

(i) Debayer First, Bilinear

(k) Debayer Last, Bilinear

(m) Debayer Last, using [43]

Figure 5: Reconstructed results from 3 perfectly aligned virtual Canon 5D sensors captured 3 f-stops apart.

To compare our method to the algorithm proposed by Granados *et al.* [10] we captured four exposures four f-stops apart with a Canon 40D camera, with calibrated gain 1.24 (ISO 400), and readout noise variance 64.2 (14 bit sensor). For a fair comparison to our method we perform demosacing (bilinear CFA interpolation) after HDR-assembly for Granados method, in contrast to the publicly available code that performs a crude nearest neighbour demosacing before HDR assembly. In Figure 6 it is visible that that Granados method performs slightly better, although the difference is not significant. Note that in contrast to the method of Granados *et al.*our method handles arbitrary misalignments between the sensors.

7.2. Sensor Missalignment

In Figure 7, LPA and CALPA is compared to the real-time multi-sensor reconstruction method recently proposed by Tocci et al. [3]. The reconstructions were generated from 3 virtual exposures of a simulated KANON KAI-04050 sensor (used in our experimental HDR video system) captured 5 f-stops apart. The lowest exposed and the highest exposed sensors are perfectly aligned, while the middle sensor has a translational offset of $0.4, 0.45$ in pixel units. This is similar to the alignment errors reported by Tocci *et al.*for their camera system. The method of Tocci *et al.*copes rather well with translational misalignments, because each reconstructed pixel uses information from only one sensor. However, the reconstruction suffers from noise. This is expected as the method of Tocci et al. only considers observations from the highest exposed non-saturated sensor at a time. In contrast LPA and CALPA is based on a maximum likelihood approach incorporating information from all non-saturated sensor observations. CALPA effectively denoises the image by using larger filtering supports in flat image areas. Compared to the other methods, CALPA also reduces unwanted color artifacts in areas of high spatial frequency.

A robust multi-sensor reconstruction algorithm requires handling general sub-pixel misalignments between the sensors, including rotational errors. Both LPA and CALPA gracefully handles any geometric misalignments between the sensors, as can be see Figure 9, where the reconstructions were based on 3 virtual KANON KAI-04050 sensors simulated with 6 degree rotational errors between them. The Debayer-first method can also handle general misalignment errors including rotations. however, methods performing demosaicing before HDR assembly often suffers from color artifacts in regions of high spatial frequency, e.g. seen as color fringes in Figure 5. Methods performing demosaicing after resampling (alignment), e.g. Tocci *et al.* [3], cannot handle general misalignments, as the color filters no longer necessarily overlap after resampling.

LPA parameters - Varying the polynomial order and the scale parameter for the LPA method imposes a trade-off between image sharpness, noise reduction and processing speed. Figure 8b and 8c show reconstructions from three aligned virtual Kodak KAI-04050 sensors 4 f-stops

apart and a resolution of 512×341 pixels. The close-ups show how the image fidelity varies for different choices of polynomial order the M and the scale parameter h described in Section 5.3. Using a constant fit, $M = 0$, leads to unwanted block artifacts and color fringes, especially at edge features and in regions were one sensor is close to saturation. The difference between order $M = 1$ and $M = 2$ is more subtle, with only small improvements in visual quality. A low value for the scale parameter, h, increases image sharpness, while a large h reduces image noise.

8. Experimental validation

To evaluate the real world performance of our approach, we have implemented the LPA algorithm in CUDA to enable GPU processing of the output from a custom built multi-sensor HDR camera system.

The camera system used in our experiments is based on four high quality Kodak KAI-04050 CCD sensors with a resolution of 2336×1752 pixels and RGB Bayer pattern CFA sampling. The sensors receive different amounts of light from an optical system with a common lens, a four-way beam splitter and four different ND filters, see Figure 10, top. The relay lens was custom built for this setup, and extends the optical path length to make room for the beam splitters. All other optics are standard, off-the-shelf components. The sensors have 12 bits linear A/D conversion, and the exposure scalings cover a range of $1 : 2^{12}$, yielding a dynamic range equivalent to $12 + 12 = 24$ bits of linear resolution, commonly referred to as "24 f-stops" of dynamic range. The dynamic range can be extended further by varying the exposure times between the sensors. The sensors are connected to a host computer through a CameraLink interface. The system allows for capture, processing and off-line storage of up to 32 frames per second at 4 Mpixels resolution, amounting to around 1 GiB/s per second of raw data.

The misalignments between the sensors were measured by imaging a checkerboard calibration target and computing the cross-sensor correlation using OpenCV 2.2. The misalignments could be characterized as translations of a few pixels and pure 2D rotations in the order of fractions of a degree. The sensor transformation matrices could be estimated with an accuracy in the order of 0.1 pixels. The system was radiometrically calibrated as described in Section 6.2. The gain parameter was estimated for each pixel separately, and then spatially averaged to find the sensor gain $g_s = 0.27DV/e$. Readout noise was estimated for each pixel, with a cross-sensor spatial mean of $\hat{\mu}_{s,i} = 72$ e, and a standard deviation of $\hat{\sigma}_{s,i} \approx 11.8$ e

Our software implementation of the LPA algorithm required several seconds for processing of reasonably sized frames. To enable real time processing for high resolution video data, we implemented the LPA algorithm in CUDA. Two CUDA kernels are executed per frame. First,

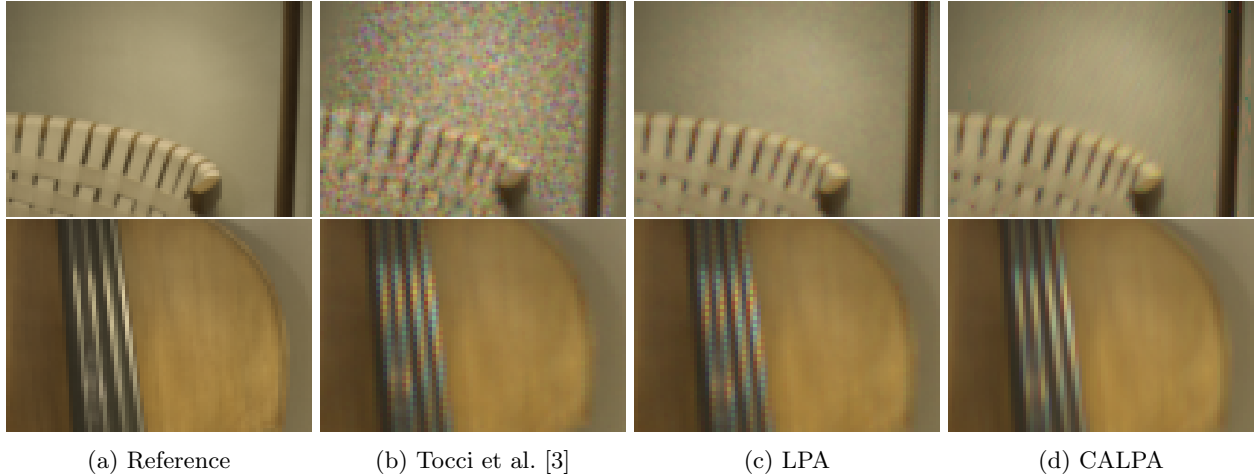(a) Reference   (b) Tocci et al. [3]   (c) LPA   (d) CALPA

Figure 7: Magnifications from reconstructions of a scene virtually exposed with 3 KODAK KAI-04050 sensors 5 f-stops apart. a) Tonemapped HDR reference images. b) Reconstructions by the method proposed by Tocci et al. [3], c)LPA $M = 1$, $h = 0.7$, d)CALPA, $M = 1$, $h = 0.7$, $\alpha = 0.005$. In areas of sensor transitions, the method by Tocci *et al.*exhibits severe noise as only one sensor is used at a time, as can be seen in the magnification on the top row. In contrast LPA handles sensor transitions gracefully due to the incorporation of a sensor noise model and a statistical maximal likelihood fit to all sensor data. CALPA show ever better results due to the inherent denoising using adaptive filtering supports. Using CALPA also help in reducing unwanted color artifacts.
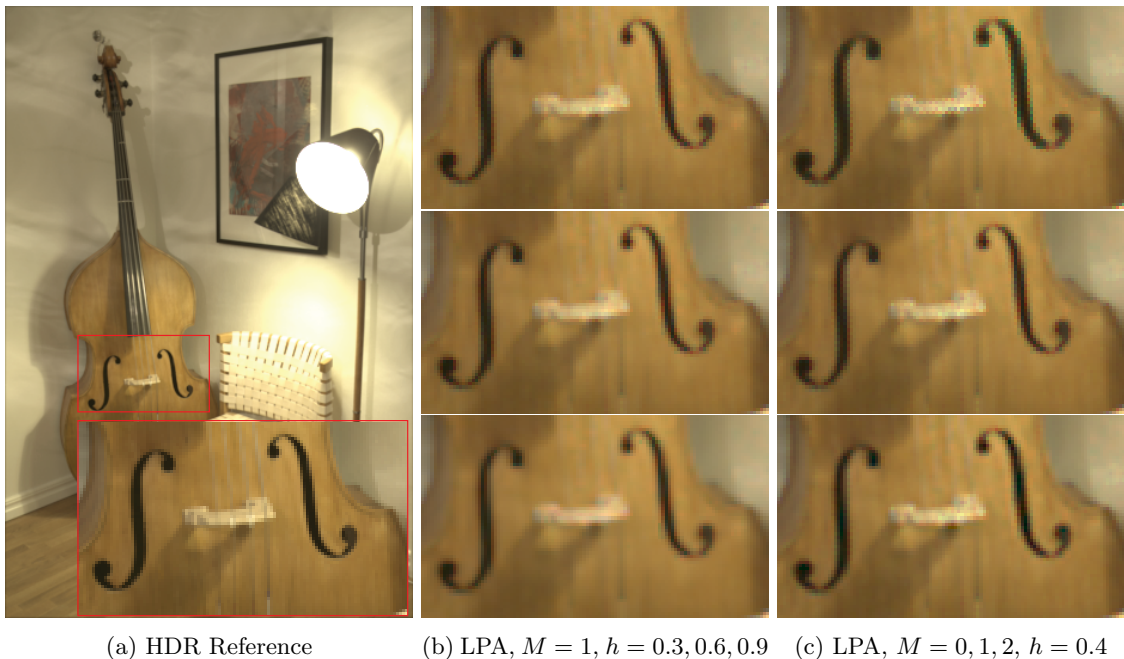

(a) HDR Reference   (b) LPA, $M = 1$, $h = 0.3, 0.6, 0.9$ (c) LPA, $M = 0, 1, 2$, $h = 0.4$

Figure 8: a) Tonemapped HDR reference image. b) LPA reconstructions, $M = 1$, for different scale parameters, $h_{R,B} = 0.3, 0.6, 0.9$. Using a low scale parameter leads to sharp reconstructions. If severe image noise is present, a larger scale parameter may be beneficial. c)LPA reconstructions of different polynomial order $M = 0, 1, 2$ with a fixed scale parameter, $h_G = 0.2$, $h_{R,B} = 0.4$. Using a constant polynomial ($M = 0$), blocky artifacts and color fringes are observed close to sensor saturation and sharp edges.

the raw sensor images are converted to a half-float radiant power estimate, $\hat{f}_{s,i}$, and a variance estimate, $\hat{\sigma}_{\hat{f}_{s,i}}$, is computed. Secondly, for each pixel in the output HDR image, all samples overlapping the observation window are read from global GPU memory. The weight corresponding is computed and used to computed the weighted least

(a) HDR Reference   (b) LPA, $M = 1$ (c) CALPA, $M = 1$, $\alpha = 0.006$
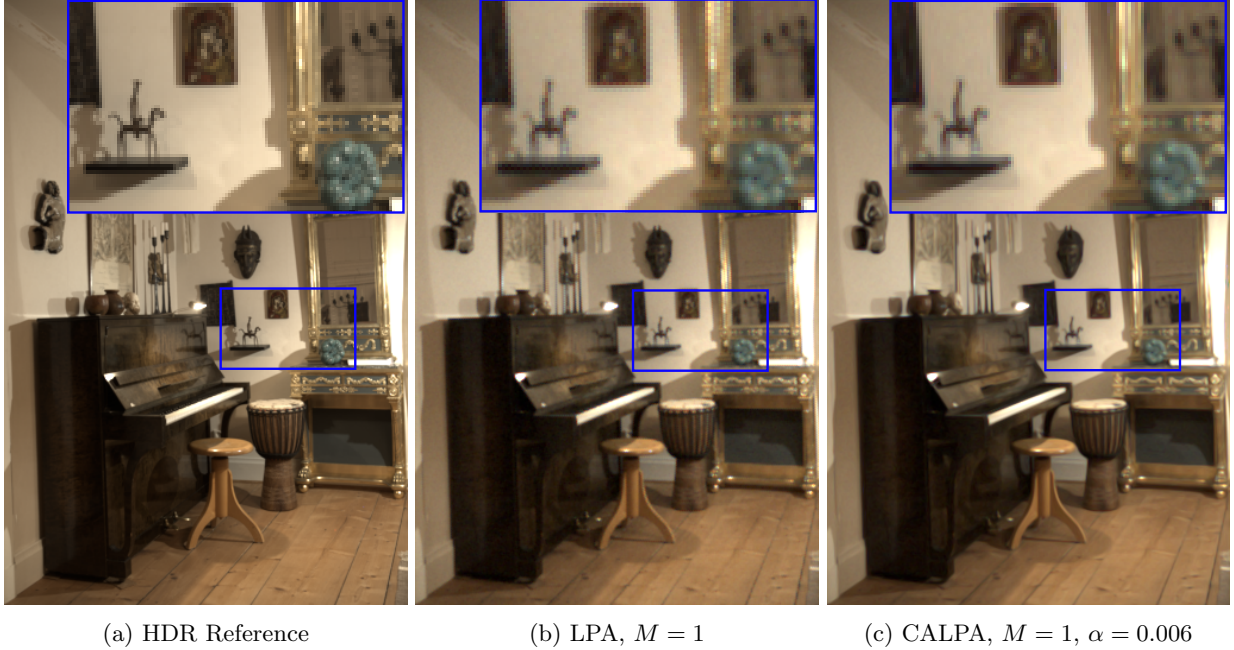
Figure 9: Reconstructed results from 3 virtual sensors with 6 degrees rotational misalignment between the sensors. a) Tonemapped HDR reference image (no bayer pattern). b) LPA, $M = 1$ c) CALPA, $M = 1$, $\alpha = 0.006$. Results are best viewed close up in digital form.
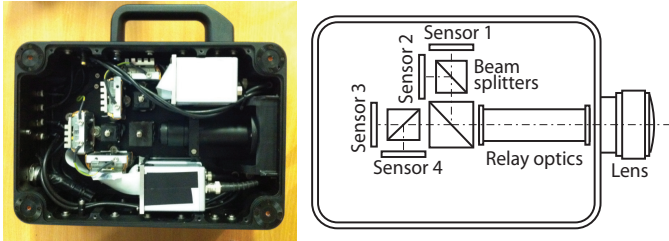

Figure 10: Top: An overview of our experimental HDR video camera. Bottom: a full resolution locally tone mapped frame from a video sequence captured with the HDR video camera. The image is reconstructed from four 4 Mpixel sensors at 26 fps. Bottom-right: Four gamma mapped images 2 f-stops apart from the same HDR-image.

squares estimate described in Equation (6). Since memory transfer is a bottle-neck, we use page-locked and write-combined memory to improve efficiency. The data is transferred to the GPU for kernel execution and display, and then back to the CPU for disk storage. To enable simultaneous data transfer and kernel execution we utilize two CUDA streams.

If the sensors are aligned or related by purely translational transforms, and if the output resolution matches the input resolution, the spatial arrangement of samples in each observation window will be invariant across the image, and the window weights w_i can be precomputed once. This results in a 2-3× speedup for a local constant fit $M = 0$.

Figure 10 shows an example frame captured using our experimental HDR-video system. Using our CUDA implementation with $M = 0$, and an output resolution of 2336×1752 pixels, the four input frames are processed at 26 fps using pre-computed weights on an Nvidia GForce 680 GPU. As can be seen in the tone-mapped image, which is in full resolution to enable zooming in during on-screen viewing, the small rotational misalignments present in the hardware setup are not visible even though the weights are pre-computed. This is partly due to the smoothing inherent in the LPA method. Reconstructing the same frame without pre-computed weights achieves an interactive speed of 7 fps. The performance scales linearly with the number of pixels, so the corresponding figures for reconstructing to HD resolution frames of 1920×1080 are 51 fps and 14 fps respectively.

9. Conclusions and Future Work

This paper presented a novel filtering framework for HDR reconstruction from multi-sensor images that performs all steps in the traditional HDR imaging pipeline in a single step. The reconstruction is based on spatially adaptive filtering and noise-aware local likelihood estimation using isotropic and anisotropic filter supports taking into account the correlation between color channels. The method uses a novel sensor noise model adapted to multi-sensor systems. We also presented an overview of a novel HDR video camera and showed how our algorithm achieves real-time performance for isotropic filter supports. It should be noted that the framework is general in that the algorithms apply equally well to still image HDR reconstruction from e.g. exposure bracketing. The only requirement is that the sensor transformations can be accurately estimated.

As future work, we would like to extend the filter supports from 2D to 3D to include both the spatial and temporal domain in the reconstruction.

10. Acknowledgements

[1] M. Aggarwal, N. Ahuja, Split aperture imaging for high dynamic range, International Journal of Computer Vision 58 (2004) 7–17.

[2] M. Mcguire, J. F. Hughes, Optical Splitting Trees for High-Precision Monocular Imaging, IEEE Computer Graphics And Applications 27 (2007) 32–42.

[3] M. D. Tocci, C. Kiser, N. Tocci, P. Sen, A Versatile HDR Video Production System, ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH 2011) 30 (2011).

[4] J. Kronander, S. Gustavson, G. Bonnet, J. Unger, Unified hdr reconstruction from raw cfa data, in: Proc. of ICCP 2013.

[5] N. New Imaging Technologies, Magic technology - scene contrast indexed image sensing with wdr, White paper, 2008.

[6] P. Debevec, J. Malik, Recovering high dynamic range radiance maps from photographs, in: Proc. of ACM SIGGRAPH, ACM, ACM, 1997, pp. 369–378.

[7] S. Kang, M. Uyttendaele, S. Winder, R. Szeliski, High dynamic range video, ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH 2003) 22 (2003) 319–325.

[8] J. Unger, S. Gustavson, High-dynamic-range video for photometric measurement of illumination, Proc. of SPIE 6501 (2007) 65010E.

[9] B. Ajdin, M. B. Hullin, C. Fuchs, H.-P. Seidel, H. P. A. Lensch, Demosaicing by smoothing along 1-d features, in: Proc. of CVPR 2008.

[10] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. Seidel, H. Lensch, Optimal HDR reconstruction with linear digital cameras, in: in Proc. of CVPR, IEEE, IEEE, 2010.

[11] B. Gunturk, J. Glotzbach, Y. Altunbasak, R. Schafer, R. Mersereau, Demosaicking: color filter array interpolation, IEEE Signal Processing Magazine 22 (2005) 44–54.

[12] E. Reinhard, W. Heidrich, S. Pattanaik, P. Debevec, G. Ward, K. Myszkowski, High dynamic range imaging: acquisition, display, and image-based lighting, Morgan Kaufmann, 2010.

[13] K. Myszkowski, R. Mantiuk, G. Krawczyk, High Dynamic Range Video, Morgan & Claypool, 2008.

[14] J. Astola, V. Katkovnik, K. Egiazarian, Local Approximation Techniques in Signal and Image Processing, SPIE Publications, 2006.

[15] S. Nayar, T. Mitsunaga, High dynamic range imaging: Spatially varying pixel exposures, in: Proc. of CVPR, pp. 472 – 479.

[16] S. G. Narasimhan, S. K. Nayar, Enhancing Resolution Along Multiple Imaging Dimensions Using Assorted Pixels, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 518–530.

[17] F. Yasuma, T. Mitsunaga, D. Iso, S. K. Nayar, Generalized Assorted Pixel Camera : Postcapture Control of Resolution , Dynamic Range , and Spectrum, IEEE Transactions on Image Processing 19 (2010) 2241–2253.

[18] H. Knutsson, C.-F. Westin, Normalized and differential convolution, in: Proc. of CVPR, pp. 515–523.

[19] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Sixth International Conference on Computer Vision (ICCV), Narosa Publishing House, 1998, pp. 839–846.

[20] P. Lancaster, K. Salkauskasr, Surfaces generated by moving least squares methods, Mathematics of Computation 87 (1981) 141–158.

[21] H. Takeda, S. Farsiu, P. Milanfar, Kernel regression for image processing and reconstruction, IEEE Transactions On Image Processing 16 (2007) 349–366.

[22] P. Milanfar, A tour of modern image filtering, to appear in IEEE Signal Processing Magazine (2011).

[23] C. Loader, Local regression and likelihood, New York: Springer-Verlag, 1999.

[24] R. Tibshirani, T. Hastie, Local likelihood estimation, Journal of the American Statistical Association 82 (1987) pp. 559–567.

[25] S. Mann, R. W. Picard, On Being 'undigital' With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures, in: Proc. of IS&T Ann. Conf.

[26] A. O. Akyüz, E. Reinhard, Noise reduction in high dynamic range imaging, Journal of Visual Communication and Image Representation 18 (2007) 366 – 376.

[27] T. Mitsunaga, S. K. Nayar, Radiometric Self Calibration, in: Proc. of CVPR, volume 1, pp. 374–380.

[28] K. Kirk, H. Andersen, Noise characterization of weighting schemes for combination of multiple exposures, in: Proc. British Machine Vision Conference (BMVC), volume 3, pp. 1129–1138.

[29] P. Milanfar (Ed.), Super-resolution Imaging, CRC Press, 2010.

[30] S. Farsiu, M. D. Robinson, M. Elad, P. Milanfar, Fast and robust multi-frame super-resolution, IEEE Transactions on Image Processing 13 (2004) 1327–1344.

[31] R. Hardie, A fast image super-resolution algorithm using an adaptive Wiener filter., IEEE Transactions on Image Processing 16 (2007) 2953–64.

[32] B. K. Gunturk, M. Gevrekci, High-Resolution Image Reconstruction From Multiple Differently Exposed Images, IEEE Signal Processing Letters 13 (2006) 197–200.

[33] H. Zimmer, A. Bruhn, J. Weickert, Freehand HDR imaging of moving scenes with simultaneous resolution enhancement, Computer Graphics Forum (Proceedings of Eurographics) 30 (2011) 405–414.

[34] B. Bayer, Color imaging array, US Patent 3 971 065, 1976.

[35] T. P. K. Hirakawa, Joint demosaicing and denoising, IEEE Trans. Image Processing (2006).

[36] K. Hirakawa, Color filter array image analysis for joint denoising and demosaicking, in: R. Lukac (Ed.), Single-Sensor Imaging: Methods and Applications for Digital Cameras, CRC Press, 2008.

[37] S. Farsiu, M. Elad, P. Milanfar, Multiframe demosaicing and super-resolution of color images, Image Processing, IEEE Transactions on 15 (2006) 141–159.

[38] R. v. S. Anton Kachatou, Dynamic range enhancement algorithms for cmos sensors with non-destructive readout, in: IEEE International Worshop on Imaging Systems and Techniques.

[39] S. Hasinoff, F. Durand, W. Freeman, Noise-optimal capture for high dynamic range photography, in: in Proc. CVPR, pp. 553 –560.

[40] Y. Reibel, M. Jung, M. Bouhifd, B. Cunin, C. Draman, Ccd or cmos camera noise characterisation, The European Physical Journal - Applied Physics 21 (2003) 75–80.

[41] A. Foi, M. Trimeche, V. Katkovnik, K. Egiazarian, Practical poissonian-gaussian noise modeling and fitting for single-image raw-data, IEEE Transactions on Image Processing 17 (2008) 1737–1754.

[42] A. Foi, Clipped noisy images: heteroskedastic modeling and practical denoising, Signal Processing (2009) 2609–2629.

[43] H. S. Malvar, L. wei He, R. Cutler, High-quality linear interpolation for demosaicing of bayer-patterned color images, in: Proceedings of the IEEE International Conference on Speech, Acoustics, and Signal Processing.

[44] R. Clark, Digital camera sensor performance summary, http://www.clarkvision.com/imagedetail/digital.sensor. performance.summary/, 2009.