# Coded Exposure HDR Light-Field Video Recording

David C. Schedl[1], Clemens Birklbauer[1] and Oliver Bimber[1]
Johannes Kepler University Linz

SIGGRAPH2013

JOHANNES KEPLER UNIVERSITY LINZ | JKU

Institute of Computer Graphics

## Abstract

**Introduction and Motivation**

Capturing exposure sequences for computing high-dynamic range (HDR) images causes motion blur in case of camera movements. This also applies for light-field cameras: Frames rendered from multiple blurred HDR light-field perspectives are also blurred. While the recording times of exposure sequences for a single-sensor camera cannot be reduced, we demonstrate how to achieve this for a camera array. Thus, we increase capturing time and reduce motion blur for HDR light-field video recording.

**Coded Exposure HDR Light-Field Video Recording**

We apply a spatio-temporal exposure pattern (Figures 1c,d), while capturing frames with a camera array (Figure 1a). This reduces the overall recording time and enables the estimation of camera movement within one light-field video frame. By estimating a depth map and local point spread functions (PSFs) from multiple perspectives with the same exposure (exposure perspectives), regional motion deblurring can be supported in each exposure. Missing exposures at various perspectives are interpolated. A general outline of our method is shown in Figure 2.

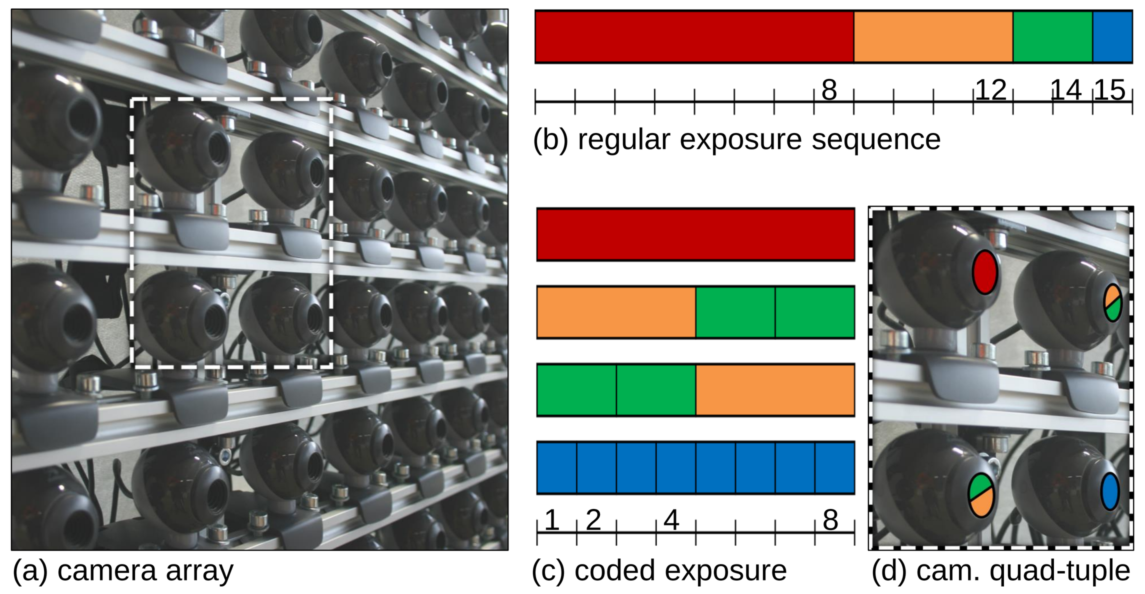


**Figure 1:** Our camera array (a), corresponding recording intervals for capturing a regular exposure sequence (b), our coded exposure approach (c). The exposure pattern is repeated for each camera quad-tuple of the array (d).

## Related Work

**HDR Light-Field Cameras**

To our knowledge there is no previous work on capturing HDR light fields. In [1], a camera array was used to record either HDR images or LDR light fields - yet not HDR light fields (neither still frames nor video frames). For recording HDR images, neutral density filters in front of the micro-lenses of a focused plenoptic camera have been used [2]. We reduce motion blur caused by long exposure times to enable the recording of HDR light-field video frames with moving camera arrays.

**Deblurring Methods**

Image deblurring has been explored strongly in recent years. However, methods that track PSFs for deblurring often require additional hardware, such as a low-resolution high-frame-rate cameras [3] or inertia sensors attached to the main camera [4]. Some techniques use depth maps to allow blind deconvolution on depth layers, where the depth is estimated from blurry stereo-pairs [5]. We do not require additional sensors or specialized cameras, and do not rely on blind deconvolution. Conventional camera arrays together with our exposure coding allow the simultaneous recording of HDR light-field frames and the estimation of local PSFs needed for regional motion deblurring.

## Our Method

**Capturing**

We encode four exposure times in each repeating camera quad-tuple of a camera array. Instead of recording all exposures sequentially for all perspective cameras (Figure 1b) we apply a spatio-temporal exposure pattern (Figure 1c).

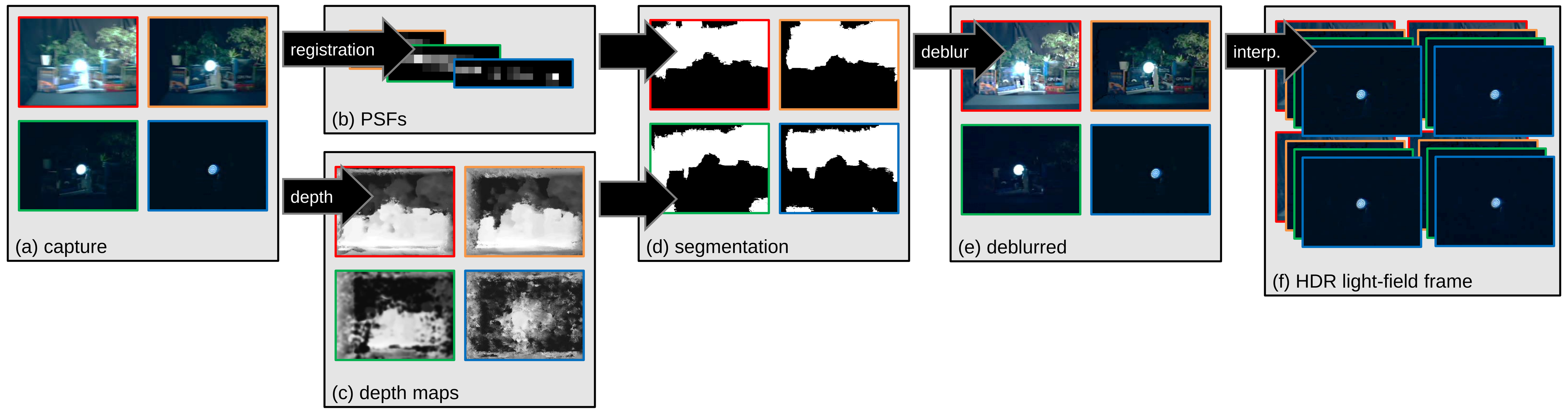


**Figure 2:** An outline of our technique showing four captured perspectives of a camera array (a), local PSFs computed by registration (b), depth maps for each exposure perspective (c), segmentations based on PSFs and depths (d), the deblurred exposure perspectives (e), and the final HDR light-field frame with interpolated exposure perspectives (f).

This reduces the overall recording interval, from 15 to 8 time steps (assuming full stop exposure spacing), where one time step is the shortest exposure time.

Within each camera quad-tuple, the shortest and longest exposures are captured only from one perspective each and two intermediate exposures are recorded time-interleaved from the two remaining perspectives (Figure 1d). Although, this hinders us from recording all exposures at all perspectives, it allows to capture subframes at camera positions with shorter exposure times (i.e. up to 8 subframes for the shortest exposure). Figure 2a illustrates the captured exposure perspectives for one camera quad-tuple. While the longest exposure contains the strongest motion blur, the subframes of the shortest exposure are nearly blur free.

**Registration**

We calculate SURF features for each exposure perspective and match those features within each subframe of the same exposure, thus retrieving 3D coordinates of those features. This is repeated for all subframes, except for the longest exposures. The resulting 3D features are matched within a subframe sequence of the same exposure, and used for inter-frame registrations. Registration is done iteratively, starting with the long-exposed subframes, where the previous registration is used as initial guess. PSFs are computed for each exposure perspective by transforming and projecting each 3D feature with its best fitting registration (Figure 2b). The lower PSF samplings at the longer exposures are upsampled.

**Depth**

For each of the four different exposures that are recorded at varying perspectives, we compute a depth map (Figure 2c) and interpolate missing depth perspectives. Since these depth maps vary locally in quality (due to interpolation, motion blur, and low SNR), we compile them to a single composite depth map per perspective, based on a normalized quality metric (coherence across perspectives).

**Segmentation**

We perform clustering on the depths of all estimated PSFs to derive discrete depth layers in the composite depth maps. For each layer, dense (per pixel) PSF maps are computed. Segmenting them results in raw clusters that are further refined by matting [6]. This reduces noise and inaccuracies due to the interpolation (Figure 2d).

**Deblurring**

For each cluster (Figures 3a,e) we calculate a single PSF by averaging all least upsampled PSFs (i.e., PSFs of which the best fitting registration has a high sampling rate). Since the final PSFs might suffer from low sampling rates, we apply a blind deconvolution [7] and use our PSF estimations (Figures 3b,f) as initial guess. This leads to refined PSFs (Figures 3c,g) and deblurred clusters (Figures 3d,h). Finally, the clusters are merged by alpha-blending, resulting in deblurred subframes for all camera perspectives (Figure 2e).

By shifting the PSFs before deconvolution, it becomes possible to estimate two additional frames for a recording interval at time steps 4 and 8 (Figure 1c). Thus, for one HDR light-field video frame we obtain one exposure subframe for the shortest, one for the longest exposure perspective, and two exposure subframes for the two intermediate exposure perspectives. Compared to classical exposure sequencing (Figure 1 b), this leads to a 3.75 times higher frame rate.

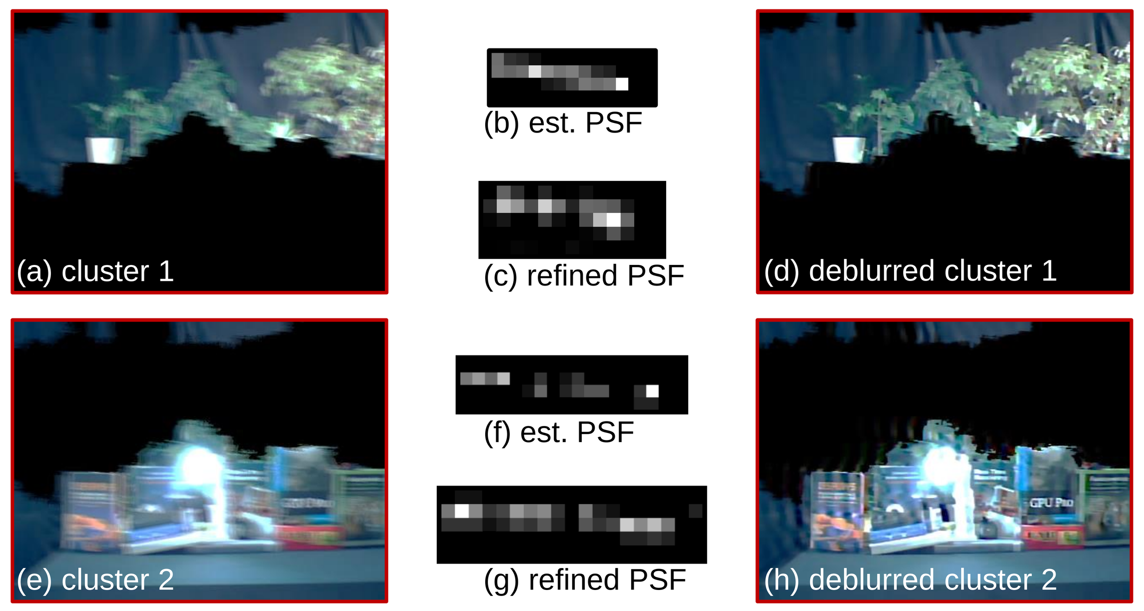


**Figure 3:** Two clusters of one longest exposed perspective (a,e), the PSFs computed for each cluster (b,f), the refined PSFs (c,g), and the deblurred clusters (d,h) after deconvolution with [7].

**Interpolation**

With the recorded and deblurred exposure subframes we finally compute an enhanced composite depth map to interpolate deblurred exposure images for all camera perspectives that have not been directly recorded (Figure 2f). With these, we can derive an HDR image for each camera position.

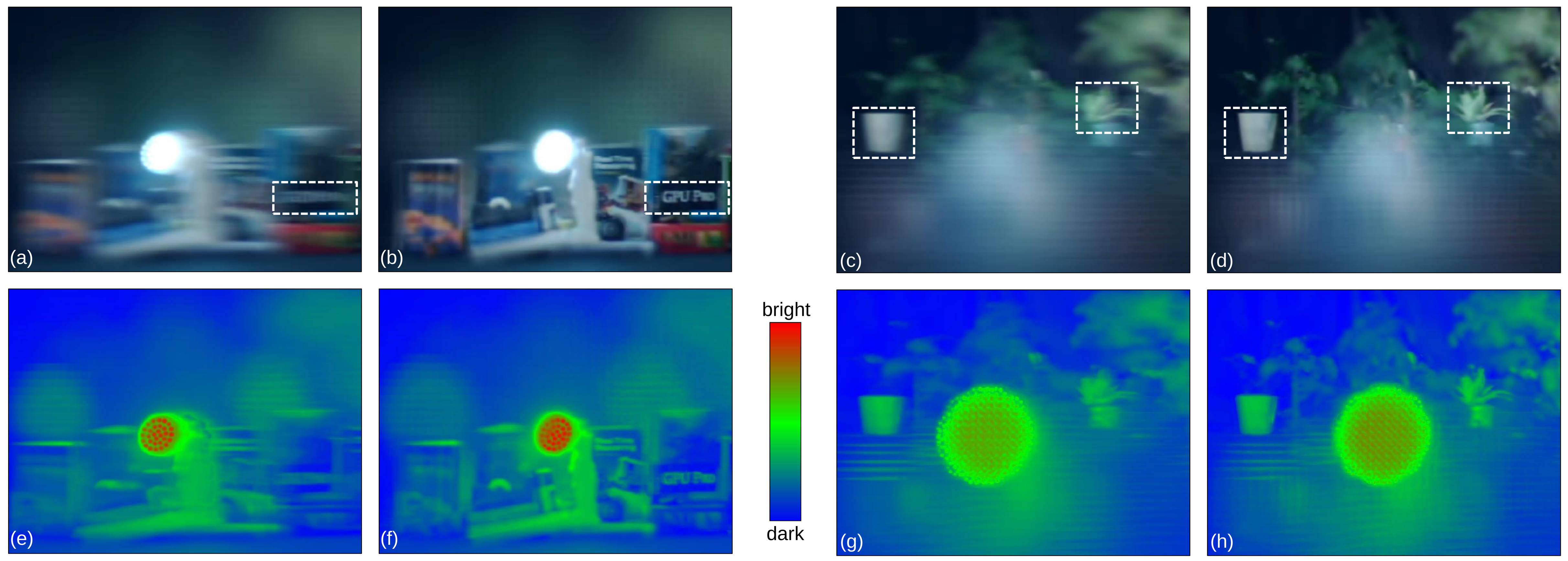


**Figure 4:** Light-field rendering of an HDR light field recorded with regular exposure sequencing (a,c,e,g) and our coded expsoure approach (b,d,f,h), where (a,b,c,d) are tone-mapped and (e,f,g,h) are heat-mapped to illustrate the dynamic range of the scene. The synthetic focus of the light-field renderings is set to the foreground (a,b,e,f) and to the background (c,d,g,h).

## Results

Figures 4a,b,c,d illustrate tone-mapped, wide-synthetic-aperture (i.e., shallow depth-of-field) images rendered from a light field that was recorded during camera motion. Figures 4a,c show the results at different focus settings with regular exposure sequences for each camera perspective. Figures 4b,d display the same images recorded with our coded exposures and computed with our method.

Figures 4e,f,g,h show heat-mapped versions of the recorded light fields. Figures 4e,g present regular exposure sequencing, while Figures 4f,h illustrate the results with our approach. Note, that while the motion blur is vastly reduced with our method, there is no reduction in the dynamic range compared to regular exposure sequencing.

## Limitations and Future Work

Our method is currently limited to static scenes, while camera motion is supported. One field of future investigation is therefore the detection and treatment of local object motion. Currently our camera array cannot capture a light-field video at an acceptable frame rate, due to bandwidth constraints. A setup with several computers [1] avoids bandwidth limitations, but reduces mobility. Our technique is limited to arrays with individually controllable cameras (i.e. exposure settings). Future per-pixel controllable sensors might make our method applicable for other light-field camera designs, such as ones based on micro-lenses.

## References

1. Wilburn, Bennett, Joshi, Neel, Vaish, Vaibhav et al. High performance imaging using large camera arrays. SIGGRAPH (2005).
2. Georgiev, Todor and Lumsdaine, Andrew. Rich image capture with plenoptic cameras. ICCP (2010).
3. Tai, Yu-Wing, Du, Hao, Brown, Michael S., and Lin, Stephen. Image/Video Deblurring using a Hybrid Camera. CVPR (2008).
4. Joshi, Neel, Kang, Sing Bing, Zitnick, C. Lawrence, and Szeliski, Richard. Image Deblurring using Inertial Measurement Sensors. SIGGRAPH (2010).
5. Xu, Li and Jia, Jiaya. Depth-Aware Motion Deblurring. ICCP (2012).
6. Levin, Anat, Lischinski, Dani, and Weiss, Yair. A Closed Form Solution to Natural Image Matting. CVPR (2006).
7. Levin, Anat, Weiss, Yair, Durand, Frédo, and Freeman, William T. Efficient Marginal Likelihood Optimization in Blind Deconvolution. CVPR (2011).

[1] {firstname.lastname}@jku.at

More information at: www.jku.at/cg