Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 286

Freehand HDR Imaging of Moving Scenes with Simultaneous Resolution Enhancement

Henning Zimmer, Andrés Bruhn and Joachim Weickert

Saarbrücken 2010

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 286
submitted: December 7, 2010

Freehand HDR Imaging of Moving Scenes with Simultaneous Resolution Enhancement

Henning Zimmer
Mathematical Image Analysis Group
Dept. of Mathematics and Computer Science
Saarland University, Campus E1.1, 66041 Saarbrücken, Germany
zimmer@mia.uni-saarland.de

Andrés Bruhn
Vision and Image Processing Group,
Cluster of Excellence Multimodal Computing and Interaction,
Saarland University, Saarbrücken, Germany
bruhn@mmci.uni-saarland.de

Joachim Weickert
Mathematical Image Analysis Group
Dept. of Mathematics and Computer Science
Saarland University, Campus E1.1, 66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

Freehand HDR Imaging of Moving Scenes with Simultaneous Resolution Enhancemen

Henning Zimmer Andrés Bruhn Joachim Weickert

Abstract

Despite their high popularity, common high dynamic range (HDR) methods are still limited in their practical applicability: They assume that the input images are perfectly aligned, which is often violated in practise. Our paper does not only free the user from this unrealistic limitation, but even turns the missing alignment into an advantage: By exploiting the multiple exposures, we can create a super-resolution image. The alignment step is performed by a modern energy-based optic flow approach that takes into account the varying exposure conditions. Moreover, it produces dense displacement fields with subpixel precision. As a consequence, our approach can handle arbitrary complex motion patterns, caused by severe camera shake and moving objects. Additionally, it benefits from several advantages over existing strategies: (i) It is robust under outliers (noise, occlusions, saturation problems) and allows for sharp discontinuities in the displacement field. (ii) The alignment step neither requires camera calibration nor knowledge of the exposure times. (iii) It can be efficiently implemented on CPU and GPU architectures. After the alignment is performed, we use the obtained subpixel accurate displacement fields as input for an energy-based, joint super-resolution and HDR (SR-HDR) approach. It introduces robust data terms and anisotropic smoothness terms in the SR-HDR literature. Our experiments with challenging real world data demonstrate that these novelties are pivotal for the favourable performance of our approach.

1 Introduction

One of the major obstacles for the practical application of high dynamic range (HDR) imaging is the alignment problem. As common HDR techniques compute a weighted average over irradiances obtained from a given exposure series [19, 6, 28, 27], they rely on perfectly aligned images without displacements in between them. However, this assumption is hardly met under real world conditions: Displacements are caused by moving objects

Figure 1: **Left:** Freehand exposure series. **Right:** Result after alignment plus joint dynamic range and resolution enhancement

(persons, clouds, etc.) and by camera shake, unless the effort is taken to use a tripod. Consequently, there is a strong interest in methods that align the exposure series by compensating the images for the displacements. However, a reliable estimation of the displacements is challenging since the images are taken with varying exposure times that yield severe brightness changes.
In this paper, we show how to adapt a modern energy-based optic flow approach to cope with these brightness changes. This allows for a robust and accurate estimation of dense displacement fields between the input images and yields an alignment method that outperforms existing strategies in challenging real world scenarios. Additionally, our approach neither requires a preceding camera calibration nor knowledge of the exposure times and can be efficiently implemented on CPU and GPU architectures.
As our alignment approach yields displacement fields with subpixel precision, they not only allow for an accurate alignment, but can also be used for increasing the spatial resolution of the result. To this end, one can adapt concepts from super-resolution (SR) approaches [31, 25, 22]. Here, an image with increased spatial resolution is obtained by combining the information from several images with some degree of subpixel displacement between them. This basically allows to fuse different discrete samplings of the same continuous scene.
In this paper, we propose the first energy-based *joint super-resolution and high dynamic range (SR-HDR)* approach that uses a robust data term in combination with an anisotropic smoothness term. Our experiments show that this model gives more appealing results than existing techniques such as [10, 5]. A tone mapped result obtained by our approach is shown in Fig. 1. In this context, we also wish to note that SR and HDR imaging nicely complement each other: While the SR methods increase the resolution of the image domain, HDR techniques increase the resolution of the image co-domain, i.e. the dynamic range.

1.1 Related Work

HDR Alignment. A simple and fast approach for aligning exposure series is to estimate one global transformation per image pair. Ward [32] describes this transformation by a pure translation, whereas later extensions use a translation plus a rotation [9, 13]. To cope with the brightness changes due to the varying exposures, the aforementioned approaches consider mean threshold bitmaps (MTB) obtained by a thresholding at the median of all pixel values. Using a pyramid of these images, the global displacement can then be computed by simple shift and difference operations at each pyramid level. Altough global strategies are thus very efficient, they fail in the presence of independently moving objects or for camera motions such as zooming and tiling. Similar restrictions apply to the method in [30] where a homography is computed from SIFT feature matches [17] that are robust under brightness changes. Such a strategy is implemented in the `align_image_stack` algorithm of the *Hugin* toolkit [1]. Homography-based approaches are known to fail for moving objects or camera motions that are more complex than a pure rotation.

To describe arbitrary camera motions and to handle moving objects in the scene, dense methods are needed that allow to estimate a different displacement vector for each pixel. This can be achieved by multi-step methods such as the global-local alignment strategies [15, 14] that first perform some global alignment and then refine it using a classical, local optic flow approach [18]. As this optic flow approach assumes a similar intensity at corresponding pixels, the measured intensities first need to be transferred to the irradiance domain. This, however, requires a preceding calibration step to estimate the camera response function. Another problem is that local optic flow approaches cannot estimate a displacement in flat image regions (aperture problem) and give blocky artefacts as they assume a constant (or parametric) motion model within a local neighbourhood. A more advanced multi-step method was proposed in [29]. Here, sparse correspondences obtained by feature matching are used to compute a dense displacement field via weighted linear regression. This result is then refined by a local optic flow method [18]. To deal with the brightness changes, a normalisation is proposed that gives a partial invariance to the exposure changes without using the response function.

There also exist dense methods that do not need to apply several processing steps. Menzel and Guthe [20] propose a hierarchical matching of patches based on cross-correlation to ensure robustness under the brightness changes. As no smoothness assumption on the displacements is imposed, this method

[1] `http://hugin.sourceforge.net`

is prone to give noisy displacement fields, leading to artefacts in the alignment. The only approach that imposes an explicit smoothness assumption on the displacements can be found in [16]. Here, a stereo method based on zero-mean normalised cross-correlation is used. This, however, is only possible for static scenes without moving objects and additionally requires a pair of cameras with known epipolar geometry.

SR-HDR Reconstruction. One major challenge for joint SR-HDR approaches is an accurate displacement estimation with subpixel precision. Thus, some methods rely on special camera hardware to facilitate the displacement estimation: While the methods in [24, 12] use multisampled images where the pixels on the image sensor are differently exposed, Nakai et al. [23] influence the displacements by a controlled shift of the image sensor. Even simpler, Choi et al. [5] assume that the displacements are given.
Evidently, it is more convenient to use standard cameras and to take an exposure series with varying viewpoints. This strategy is applied in [26] where the images are first aligned using a frequency domain approach that gives a global translation and rotation, as in [9, 13]. The SR-HDR result is then computed by simply interpolating irradiances obtained from the aligned images. More powerful are approaches that find the SR-HDR image by minimising an energy formulation [10, 5]. These works also prove that a joint SR-HDR reconstruction gives better results than a sequential approach. The main problem of existing energy-based methods is that they use a prior that enforces the result to be close to a mean image obtained by averaging all measurements. Although this prior stabilises the minimisation, it does *not* allow to fill in missing information and to smooth the resulting image. However, our experiments show that an appropriate filling in of information and smoothing of the result is required to obtain favourable reconstructions.

1.2 Our Contributions

This paper presents two main contributions:
(i) We propose the first HDR alignment strategy that is based on a specifically tailored, modern energy-based optic flow approach. The latter gives accurate and dense displacement fields that can describe arbitrary motion patterns caused by complex camera motions and moving objects in the scene. This allows to outperform existing strategies in challenging real world scenarios. To cope with the brightness changes due to the varying exposures, we rely on a robust version of the gradient constancy assumption in the data term. This uses edge information which is hardly perturbed by brightness changes. We additionally perform a normalisation in the data term to

prevent artefacts at image edges. Our smoothness term uses a discontinuity-preserving strategy that gives sharp edges between areas of different motion. A further appealing aspect is that we do not require information on the camera response curve or the exposure times. Thus, our method can also be used to align images for exposure fusion algorithms [21] that do not use the response curve or the exposure times. Concerning run times, efficient implementations allow to align exposure series in about one minute on a CPU and in a few seconds on parallel GPU architectures.
(ii) As we estimate displacements with subpixel precision, they can serve as input for an energy-based, joint SR-HDR method. As novelties, we propose a robust data term and the first smoothness term that allows to smooth the resulting image. Specifically, we design an anisotropic smoothness term that does not only allow to preserve, but even pronounce edges in the image. Experiments show that our proposed model achieves notably better results than existing approaches.

Paper Organisation. We present the energy functions and the minimisation strategies for the HDR alignment and the SR-HDR approach in Sec. 2 and 3, respectively. Experiments that demonstrate the favourable performance of both contributions are shown in Sec. 4. We conclude in Sec. 5 by a brief summary and an outlook on future work.

2 HDR Alignment with Energy-Based Optic Flow

Assume we are given an exposure series $g_k(i,j) : \Omega \to \{0,\dots,255\}$ for $k = 1,\dots,m$. Here, $(i,j) \in \Omega$ denotes a pixel location in the image domain $\Omega = \{1,\dots,n_x\} \times \{1,\dots,n_y\}$. The alignment of such an exposure series comes down to estimating displacement vector fields (u_k, v_k) between images g_k and a reference image g_r, where $r := \lceil m/2 \rceil$ for simplicity.
To ease presentation, we describe our model for greyscale images, but also sketch the extension to colour images.

2.1 Energy Formulation.

We compute the dense displacement fields (u_k, v_k) by minimising an energy function of the general form

$$E_1(u_k, v_k) = \sum_{(i,j)\in\Omega} \Big[D_1(u_k, v_k) + \alpha\, S_1(\nabla u_k, \nabla v_k) \Big] \ . \tag{1}$$

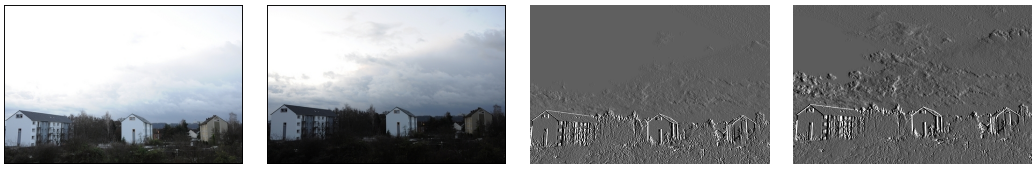

Figure 2: *From left to right:* **(a)** – **(b)** Two images of an exposure series (expsoure times 1/30 and 1/80 s). **(c)** – **(d)** The x-derivatives for the red channel of the images in (a) and (b)

The data term $D_1(u_k, v_k)$ models how well the displacements match to the given images, and the smoothness term $S_1(\nabla u_k, \nabla v_k)$ reflects the assumption of a smooth displacement field. Its influence is steered by a smoothness parameter $\alpha > 0$. The symbol $\nabla := (\mathcal{D}_x, \mathcal{D}_y)^\top$ denotes a discrete version of the gradient operator, with $\mathcal{D}_x$ and $\mathcal{D}_y$ implementing discrete approximations of the x- and y-derivatives, respectively. Note that here and in the remainder of this paper, we intentionally misuse concepts borrowed from continuous models to facilitate notation. Let us now discuss the modelling of the data and the smoothness term.

Data Term. An appropriate design of the data term is mandatory for obtaining reasonable results given images with different exposures. We base our data term on the gradient constancy assumption [3], stating that image gradients remain constant under their displacement, i.e. $\nabla g_r(i+u_k, j+v_k) = \nabla g_k(i, j)$. This results in using edge information for the alignment, see Fig. 2. As noted in [32], edges are not completely invariant under exposure changes, but our experiments will show that they give a sufficient cue when used within a robust energy-based framework.
In our data term, we perform a robust penalisation of the gradient constancy assumption to render our approach robust against outliers caused by saturation problems, noise or occlusions. Furthermore, we follow [33] and normalise the data term to prevent an undesirable overweighting that leads to artefacts in the displacement fields. Incorporating these concepts leads to our data term

$$\begin{aligned} D_1(u_k, v_k) = \Psi_{D_1}\Big(\quad & \theta_x \left|\mathcal{D}_x\, g_r(i+u_k, j+v_k) - \mathcal{D}_x\, g_k(i,j)\right|^2 \\ & + \theta_y \left|\mathcal{D}_y\, g_r(i+u_k, j+v_k) - \mathcal{D}_y\, g_k(i,j)\right|^2 \Big) \ . \end{aligned} \tag{2}$$

For the robust penaliser function Ψ_{D_1} we use the regularised L_1-norm $\Psi_{D_1}\big(s^2\big) = \sqrt{s^2 + 0.001^2}$ as in [3]. The normalisation factors θ_x and θ_y are

defined as

$$\theta_x := \frac{1}{|\nabla(\mathcal{D}_x\, g_k)|^2 + \zeta^2}\ , \quad \theta_y := \frac{1}{|\nabla(\mathcal{D}_y\, g_k)|^2 + \zeta^2}\ , \tag{3}$$

where the parameter $\zeta = 0.1$ avoids division by zero.
If colour images are to be processed, we sum up all channels in the argument of Ψ_{D_1} in (2). This reduces the influence of a pixel as soon as any channel produces an outlier.
In preliminary experiments, we investigated data terms that impose constancy of the *irradiances* to cope with the exposure changes. This, however, requires a preceding camera calibration and knowledge of the exposure times as the irradiances are computed via applying the inverse camera response function and dividing by the exposure time. Since the irradiances make no sense in saturated regions, and due to possible errors in the calibration, working in the irradiance domain alone produced unsatisfactory results. By additionally imposing constancy of the irradiance gradients we could improve the results. However, we could not obtain better results than with our proposed approach, which is simpler and more efficient as it omits calibration and irradiance computations. Note that due to the latter, our approach also allows to align image sets for exposure fusion [21] where no camera response curve or exposure times are given.

Smoothness Term. Unfortunately, the data term alone does not allow to obtain a dense displacement field. For example in flat image regions (that frequently occur in HDR imaging due to saturation problems) the data term gives no information at all. In such regions, a smoothness term is needed to fill in the displacement field. As in [3], we use a Total Variation (TV) regulariser given by

$$S_1(\nabla u_k, \nabla v_k) = \Psi_{S_1}\big(|\nabla u_k|^2 + |\nabla v_k|^2\big)\ , \tag{4}$$

with the same penaliser function as in the data term, i.e. $\Psi_{S_1} = \Psi_{D_1}$. This regulariser has been chosen since it is known to give sharp discontinuities in the displacement field.

2.2 Energy Minimisation

A necessary condition for a minimiser (u_k, v_k) of the energy (1) is given by the equations

$$\partial_{u_k(i,j)} E_1 = 0\,, \quad \text{and} \quad \partial_{v_k(i,j)} E_1 = 0\,, \quad \forall (i,j) \in \Omega\ . \tag{5}$$

Using the abbreviations

$$g_{**} := \mathcal{D}_{**}\, g_r(i+u_k, j+v_k) \ , \tag{6}$$
$$g_{*z} := \mathcal{D}_{*}\ g_r(i+u_k, j+v_k) - \mathcal{D}_{*}\, g_k(i,j) \ , \tag{7}$$

where $**\in\{xx, xy, yy\}$ and $*\in\{x, y\}$, the equations in (5) can be written as

$$\Psi'_{D_1}\big(\theta_x\, g_{xz}^2 + \theta_y\, g_{yz}^2\big)\cdot\big(\theta_x\, g_{xx}\, g_{xz} + \theta_y\, g_{xy}\, g_{yz}\big) \tag{8}$$
$$-\ \alpha\ \mathrm{div}\Big(\Psi'_{S_1}\Big(|\nabla u_k|^2+|\nabla v_k|^2\Big)\nabla u_k\Big) = 0 \ ,$$

$$\Psi'_{D_1}\big(\theta_x\, g_{xz}^2 + \theta_y\, g_{yz}^2\big)\cdot\big(\theta_x\, g_{xy}\, g_{xz} + \theta_y\, g_{yy}\, g_{yz}\big) \tag{9}$$
$$-\ \alpha\ \mathrm{div}\Big(\Psi'_{S_1}\Big(|\nabla u_k|^2+|\nabla v_k|^2\Big)\nabla v_k\Big) = 0 \ ,$$

where Ψ' denotes the derivative of Ψ w.r.t. its argument and $\mathrm{div} := \nabla^\top$ is a discrete variant of the divergence operator.
We solve equations (8)–(9) in a coarse-to-fine warping framework that computes small displacement increments at different levels of a multiscale pyramid [3]. This reduces the chance of getting trapped in a local minimum and allows to handle large displacements. For solving the problem at each pyramid level, we propose two strategies: On sequential CPU architectures, we use a nonlinear multigrid scheme with a Gauß-Seidel type solver [4]. For parallel GPU architectures, we adapt the recent work in [11] that is based on a cascadic, fast explicit gradient descent scheme implemented in CUDA.

3 Energy-based SR-HDR Reconstruction

As the displacement fields resulting from our HDR alignment method are of subpixel precision, we can use them as input for a joint super-resolution and HDR (SR-HDR) method. In addition to the input images g_k, the exposure times t_k and the estimated displacements (u_k, v_k), such a method also requires to specify a zoom factor $z > 1$. From this, we aim at reconstructing a SR-HDR image $F : \Omega_H \to \mathbb{R}^+$, where $\Omega_H = \{1, \ldots, z\cdot n_x\}\times\{1, \ldots, z\cdot n_y\}$ denotes the SR image domain. The codomain of F are the positive real numbers, representing the irradiances.

3.1 Energy Formulation

As before, we find the SR-HDR image F by minimising a suitable energy. This time it takes the form

$$E_2(F) = \sum_{(i,j)\in\Omega_H} \Big[D_2(F) + \lambda\, S_2(\nabla F) \Big] \ , \tag{10}$$

with a smoothness weight $\lambda > 0$.

Data Term. The data term $D_2(F)$ combines common SR and HDR data constraints. For the SR part we adopt a popular observation model, see e.g. [22], which can be written as

$$RBW_k\, G = g_k \ . \tag{11}$$

It models the assumption that the low-resolution images g_k are obtained from the unknown SR image $G : \Omega_H \to \{0, \dots, 255\}$ by a backward warping (W_k), blurring (B) and a downsampling/restriction (R). The warping W_k uses the estimated displacements (u_k, v_k) after upsampling to the SR grid. Note that in the SR-HDR case, the displacements are computed from the images g_k to a reference image g_r. For the HDR alignment in Section 2, they were computed in the opposite direction. The blurring operator B performs a Gaussian convolution with standard deviation σ. In accordance to the sampling theorem, we set $\sigma = z \cdot \sqrt{2}/4$.
For the HDR part, the standard observation model [6, 28] states that the unknown irradiances $f : \Omega \to \mathbb{R}^+$ are obtained as

$$f = \frac{I(g_k)}{t_k} \ , \tag{12}$$

where I is the inverse of the camera response function that describes how image intensities are related to the amount of light arriving at the image sensor. From a set of aligned images, the response function can be estimated by a calibration procedure as in [6, 28]. Thus, we assume the response function to be given for our SR-HDR approach.
Above constraints are combined in our SR-HDR data term

$$D_2(F) = \sum_{k=1}^{m} c_k \cdot \Psi_{D_2}\left(\left(RBW_k\, F - \frac{I(g_k)}{t_k} \right)^2 \right) \ , \tag{13}$$

where c_k is a HDR weighting function that reduces the influence of too dark or bright pixels from the reconstruction, as those give less reliable irradiance information. We use a Gaussian-like weighting function [28] given by

$$c_k = \exp\Big(-s \cdot \big(g_k(i,j) - 127.5 \big)^2 / \left(127.5 \right)^2 \Big) \ , \tag{14}$$

with a scale parameter $s = 8$. We further set $c_k = 0$ if the weight becomes smaller than 0.001 to ignore the influence of less reliable pixels. For the subquadratic penaliser function Ψ_{D_2}, we use as before the regularised L_1-norm, i.e. $\Psi_{D_2} = \Psi_{D_1}$. This renders our approach robust against outliers due to noise or incorrect displacement estimates.
If colour images are given, we again sum up the contributions of each channel in the argument of Ψ_{D_2} in (13).

Smoothness Term. The smoothness term plays a major role in the SR-HDR context as it has to fill in the SR-HDR image in regions where no information is available. This can be due to saturation problems (where $c_k \approx 0$), or in pixels where the warping by the displacements does not provide information.
For our SR-HDR approach, we propose a novel anisotropic smoothness term that adapts its smoothing direction to local image structures and is inspired from [33]. To obtain the required directional information, we consider the structure tensor [8] computed from an image F^0 that is obtained by upsampling a pure HDR reconstruction computed from the aligned input images. This leads to the following structure tensor:

$$J = K_{\rho_1} * \left(\nabla F^0_{\rho_2} \left(\nabla F^0_{\rho_2} \right)^\top \right) , \quad \text{with} \quad F^0_{\rho_2} = K_{\rho_2} * F^0 \ , \tag{15}$$

where $*$ is the convolution operator. The parameters $\rho_1 = 0.4$ and $\rho_2 = 0.2$ serve as neighbourhood and smoothing scale, respectively. By construction, the tensor J possesses two orthonormal eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$, which give the desired direction of local image structures: Whereas $\mathbf{v}_1$ points across image structures, the vector $\mathbf{v}_2$ points along them.
Our smoothness term then penalises the projections of the image gradients onto $\mathbf{v}_1$ and $\mathbf{v}_2$ differently: Along image edges, we perform a quadratic penalisation to obtain a strong smoothing that pronounces the edges. In the orthogonal direction across edges, we use a robust penaliser function that reduces the smoothing and helps to preserve edges. This results in the smoothness term

$$S_2(\nabla F) = \Psi_{S_2}\left(\left(\mathbf{v}_1^\top \nabla F \right)^2 \right) + \left(\mathbf{v}_2^\top \nabla F \right)^2 \ . \tag{16}$$

For the robust penaliser Ψ_{S_2}, we use the Charbonnier function defined as $\Psi_{S_2}\left(s^2\right) = 2\,\mu^2 \sqrt{1 + \left(s^2/\mu^2\right)}$, with a contrast parameter $\mu = 0.1$.
For colour images, we sum up the structure tensors of each channel to obtain a *joint* edge direction. In the smoothness term, we sum up the projections for each channel in F. For the first summand, this is done within the argument of Ψ_{S_2} to reduce the smoothing across joint edges.

3.2 Energy Minimisation

We compute the SR-HDR image F by minimising the energy (10) using a semi-implicit gradient descent scheme given by

$$\begin{aligned}\frac{F^{(l+1)} - F^{(l)}}{\tau} &= -\partial_F E_2 \\ &= \sum_{k=1}^{m} \left(W_k^\top B R^\top c_k \, \Psi'^{(l)}_{D_2}(\cdot) \left(\frac{I(g_k)}{t_k} - RBW_k \, F^{(l)} \right) \right) \\ &\quad + \lambda \, \mathrm{div} \left(T\Big(\mathbf{v}_1, \mathbf{v}_2, \nabla F^{(l)}\Big) \, \nabla F^{(l+1)} \right) \ ,\end{aligned} \tag{17}$$

where we used the abbreviation

$$\Psi'^{(l)}_{D_2}(\cdot) := \Psi'_{D_2} \left(\left(RBW_k \, F^{(l)} - \frac{I(g_k)}{t_k} \right)^2 \right) \ , \tag{18}$$

and where $F^{(l)}$ denotes the result at iteration l. The parameter τ serves as a numerical time step size. Concerning the transposed operators, $W_k^\top$ describes a forward warping, $B = B^\top$ for Gaussian blurring [22], and $R^\top$ results in upsampling. The occurring diffusion tensor T is given by

$$T\big(\mathbf{v}_1, \mathbf{v}_2, \nabla F^{(l)}\big) = \Psi'_{S_2}\Big(\big(\mathbf{v}_1^\top \nabla F^{(l)} \big)^2 \Big) \, \mathbf{v}_1 \mathbf{v}_1^\top + \mathbf{v}_2 \mathbf{v}_2^\top \ . \tag{19}$$

For solving the scheme (17) at iteration l, we perform a single Jacobi iteration. Experimentally, we found that setting $\tau = 0.5$ is feasible, which is more than one order of magnitude larger than the step size used in [22]. Further note that due to the complicated structure of the data term (13), efficient solution strategies as in the optic flow case are more difficult here. As initialisation we set set $F^{(0)} = F^0$, see (15).

4 Experiments

All image data and results presented in this paper are available for download at our supplementary web page `http://www.mia.uni-saarland.de/Research/SR-HDR`. There, we also give additional results for different test scenarios and show limitations of our method.

4.1 Alignment for HDR Imaging

The first experiment compares our proposed HDR alignment method from Sec. 2 to several competing methods. To this end, a real world exposure series (5 images, 512×340 pixels, exposure times between 1/250 to 1/30 seconds) was taken freehand using a standard DSLR camera. Some images of the series are depicted in the first row of Fig. 3. As we can see, the exposure series suffers from severe displacements due to camera shake and moving clouds, which are common problems in HDR imaging.

After alignment, we compute HDR reconstructions following the approach in [28]. For visualising the float-valued HDR data in this and the upcoming experiments, we then apply the tone mapping operator from [7]. Implementations of the before mentioned algorithms can be found in the *pfstools* package [2].

Let us now compare our method to some other alignment strategies. These are *(i)* the global method of Ward [32], *(ii)* the `align_image_stack` algorithm from the *Hugin* toolkit which implements a homography-based approach as in [30], *(iii)* a variant of our method without data term normalisation, and *(iv)* the hierarchical block matching technique of Menzel and Guthe [20].

In Fig. 3 (d) we show the result obtained with the global alignment strategy from [32], where the black border marks pixels warped outside the image domain. It becomes clear that a global translation fails to describe the complex displacements between the exposures. We also tried to apply a local optic flow approach [18] to the globally aligned images after transferring them to the irradiance domain. We found that such a global-local strategy as in [15] leads to unusable results due to the poor global alignment. Also the `align_image_stack` algorithm, see Fig. 3 (e), fails to correctly align the images as the homography computed from feature matches is not expressive enough to describe the present displacements. In Fig. 3 (f) we show a zoom in the result of our approach without data term normalisation. Comparing the latter to the zoom in our final result (Fig. 3 (i)), we realise that the proposed normalisation allows to resolve problems with unpleasant artefacts. Of course, a larger smoothness weight α would also resolve this problem, but creates a too smooth displacement field that cannot capture the discontinuity between the buildings and the clouds anymore. A corresponding zoom in the result obtained with the approach from [20] is shown in Fig. 3 (g). It turns out that the hierarchical block matching technique yields distracting artefacts at the roof of the house since it does not impose smoothness assumptions on the displacement field.

In contrast, our final result in Fig. 3 (h) and (i) shows a favourable HDR

[2] `http://pfstools.sourceforge.net`

reconstruction without disturbing artefacts. This becomes possible due to the robust and accurate displacement estimation of our approach. As an example consider the displacement field in Fig. 3 (j). For visualising the flow vectors, we use the colour code illustrated in the lower left corner of the image. As we can see, the motion of the clouds and the shift of the buildings due to camera shake are nicely discriminated. Aligning the whole exposure series took 75 seconds using our CPU version with unoptimised C code on a 3.2 GHz Intel Pentium 4. An optimised parallel implementation on a GeForce GTX 480 reduced the runtime to less than 3 seconds for the whole series. For computing this and all further results (also the ones on our supplementary web page) we set $\alpha = 2.0$ and presmoothed the images by a Gaussian convolution with standard deviation 0.5. The downsampling factor for the multiscale pyramid was set to 0.95. However, smaller values that lead to a further speed up are often possible.

4.2 SR-HDR Reconstruction

The second experiment concerns our joint super-resolution and HDR (SR-HDR) method described in Sec. 3. As before, a real world exposure series (9 images, 268×178 pixels, exposure times from 1⁄800 to 1⁄8 seconds) was taken freehand. As zoom factor we choose $z = 2$. Some of the input images are shown in the first row of Fig. 4. In the same figure we also compare our result to *(i)* an upsampled HDR result (based on bicubic interpolation) computed from the aligned images, *(ii)* a SR-HDR method similar to [10, 5] using a quadratic data term and a mean prior, and *(iii)* a variant of our method where we use an isotropic TV smoothness term $\Psi_{S_2}(\nabla F) = \sqrt{|\nabla F|^2 + 0.001^2}$, as in our optic flow alignment approach.
Comparing the results in Fig. 4 (d)–(g) and focusing on the zooms in (h)–(k), the following drawbacks of the competing methods become obvious: An upsampled HDR result (Fig. 4 (d) and (h)) looks rather dull and pixelated. Also existing SR-HDR approaches [10, 5] (Fig. 4 (e) and (i)) hardly improve the results, which we attribute to the mean prior that does not smooth the resulting image. Considering our result with a TV smoothness term (Fig. 4 (f) and (j)), we see that a prior that allows to fill in information and that smoothes the result allows to tangibly improve the quality. However, the edges in the image are rather jagged, as can be seen at the planks in the zoom in Fig. 4 (j). Using our proposed anisotropic smoothness term (Fig. 4 (g) and (k)) allows to obtain a better reconstruction of edges due to the pronounced smoothing in edge direction. The reduced smoothing in the orthogonal direction allows to preserve image details, as can be seen at the plant in the zoom in Fig. 4 (k). Concerning parameter settings, we propose

Figure 3: HDR imaging under severe displacements due to moving clouds and camera shake. *First row, from left to right:* **(a)** – **(c)** Image 1, 3 (reference) and 5 of the exposure series. *Second row, from left to right:* **(d)** Tone mapped HDR image after alignment with the method from [32]. **(e)** Same using `align_image_stack`. *Third row, from left to right:* **(f)** Zoom in our result without data term normalisation. **(g)** Same for the method from [20]. **(h)** Same for our final result from (i) *Fourth row:* **(i)** Our result. **(j)** One displacement field.

to only tune the value of the smoothness weight λ to get a visually appealing result. All other parameters can be kept fixed as given in this paper.

5 Conclusions and Outlook

This paper presented two contributions:
(i) We adapted a modern energy-based optic flow approach to reliably align exposure series for HDR imaging. This shows how the HDR community can profit from the intensive research on optic flow within the last three decades. The main advantage of our proposed strategy is that the resulting dense displacement fields can describe arbitrary complex motion patterns, which is indispensable when dealing with complex camera motions or moving objects in the scene. Another attractive aspect is that we do not require knowledge of the camera response curve or the exposure times, which makes our method also applicable to align input data for exposure fusion algorithms [21]. Concerning efficiency, reasonable runtimes can be achieved on sequential CPU architectures, whereas parallel GPU implementations reduce the computation times to a few seconds. We thus hope that our method can be used for an online alignment in portable HDR capturing devices such as the Frankencamera [1].
(ii) As our proposed alignment method yields displacements with subpixel precision, they can serve as input for a joint super-resolution and HDR (SR-HDR) approach, which allows to turn the problem of displacements between the input images into an advantage. In this paper, we presented the first energy-based SR-HDR framework that uses a robust data term in combination with an anisotropic smoothness term. Our experiments showed that especially an appropriate strategy for filling in of missing information and for smoothing the result is a key for obtaining visually appealing results.
Concerning future work, one remaining issue with our optic flow-based alignment strategy is that large displacements of small objects cannot be estimated within the used coarse-to-fine warping framework. An example for this problem is shown on our supplementary web page. Possible solutions could be using a feature-matching prior [2] or applying an anti-ghosting technique [15]. Furthermore, a joint estimation of displacements *and* SR-HDR reconstruction seems interesting, but it is questionable if the more difficult minimisation will pay off in terms of quality.

Acknowledgements. This work has been partially funded by the International Max-Planck Research School and the Cluster of Excellence "Multimodal Computing and Interaction" within the Excellence Initiative of the

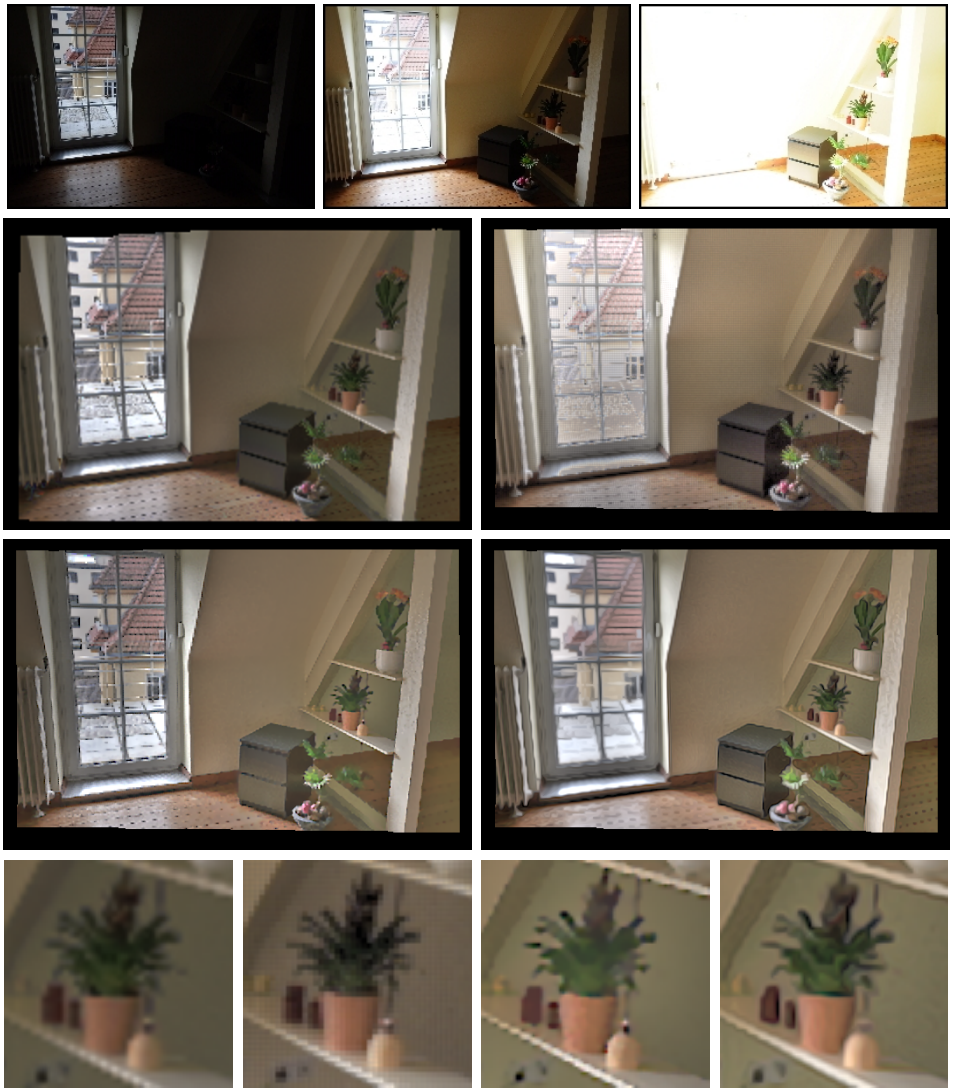

Figure 4: SR-HDR reconstruction on a real world exposure series. *First row, from left to right:* **(a)** – **(c)** Image 1, 5 (reference) and 9 of the exposure series. *Second row, from left to right:* **(d)** Upsampled HDR result (bicubic interpolation), tone mapped. **(e)** Tone mapped SR-HDR result of a method similar to [10, 5] (quadratic data term, mean prior, $\lambda = 5.0$, $\tau = 0.1$, 25 iterations). *Third row, from left to right:* **(f)** Same with our method, but using a TV smoothness term ($\lambda = 0.4, \tau = 0.5, 1000$ iterations). **(g)** Same using our proposed method with an anisotropic smoothness term ($\lambda = 0.3$, other parameters as in (f)). *Fourth row, from left to right:* **(h)** – **(k)** Zooms in (d)–(g)

German Federal Government. We thank Christian Theobalt and Miguel Granados (MPII Saarbrücken) for fruitful discussions.

References

[1] A. Adams, D.E. Jacobs, J. Dolson, M. Tico, K. Pulli, E.-V. Talvala, B. Ajdin, D. Vaquero, H.P.A. Lensch, M. Horowitz, S.H. Park, N. Gelfand, J. Baek, W. Matusik, and M. Levoy. The Frankencamera: an experimental platform for computational photography. In *Proc. ACM SIGGRAPH*, 2010.

[2] T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. In *Proc. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 41–48, Miami, FL, June 2009. IEEE Computer Society Press.

[3] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.

[4] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, 70(3):257–277, December 2006.

[5] J. Choi, M.K. Park, and M.G. Kang. High dynamic range image reconstruction with spatial resolution enhancement. *The Computer Journal*, 52(1):114–125, 2009.

[6] P.E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. ACM SIGGRAPH*, pages 369–378, 1997.

[7] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. In *Proc. ACM SIGGRAPH*, pages 249–256, 2002.

[8] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.

[9] T. Grosch. Fast and robust high dynamic range image generation with camera and object movement. In L. Kobbelt, T.Kuhlen, T. Aach, and R. Westermann, editors, *Proceedings of Vision, Modeling, and Visualization (VMV) 2006*, pages 277–284. AKA Heidelberg, 2006.

[10] B.K. Gunturk and M. Gevrekci. High-resolution image reconstruction from multiple differently exposed images. *IEEE Signal Processing Letters*, 13(4):197–200, 2006.

[11] P. Gwosdek, H. Zimmer, S. Grewenig, A. Bruhn, and J. Weickert. A highly efficient GPU implementation for variational optic flow based on the Euler-Lagrange framework. In *Proc. 2010 ECCV Workshop on Computer Vision with GPUs*, Heraklion, Greece, September 2010.

[12] H.B. Haraldsson, M. Tanaka, and M. Okutomi. Reconstruction of a high dynamic range and high resolution image from a multisampled image sequence. In *Proc. 14th International Conference on Image Analysis and Processing (ICIAP)*, pages 303–310, Modena, Italy, September 2007.

[13] K. Jacobs, C. Loscos, and G. Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, 28:84–93, 2008.

[14] T. Jinno and M. Okuda. Motion blur free HDR image acquisition using multiple exposures. In *Proc. 2008 IEEE International Conference on Image Processing*, pages 1304–1307, San Diego, CA, USA, October 2008.

[15] S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. In *Proc. ACM SIGGRAPH*, pages 319–325, 2003.

[16] S.J. Kim and M. Pollefeys. Radiometric alignment of image sequences. In *Proc. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 645–651, Washington, DC, June 2004. IEEE Computer Society Press.

[17] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[18] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, August 1981.

[19] S. Mann and R.W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. 48th IS&T Annual Conference*, pages 442–448, Washington DC, USA, May 1995.

[20] N. Menzel and M. Guthe. Freehand HDR photography with motion compensation. In H.P.A. Lensch, B. Rosenhahn, H.-P. Seidel, P. Slusallek, and J. Weickert, editors, *Proceedings of Vision, Modeling, and Visualization (VMV)*, pages 127–134. AKA Heidelberg, 2007.

[21] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, 28(1):161–171, 2009.

[22] D. Mitzel, T. Pock, T. Schoenemann, and D. Cremers. Video super resolution using duality based TV-L_1 optical flow. In J. Denzler, G. Notni, and H. Süße, editors, *Pattern Recognition*, volume 5748 of *Lecture Notes in Computer Science*, pages 432–441. Springer, Berlin, 2009.

[23] H. Nakai, S. Yamamoto, Y. Ueda, and Y. Shigeyama. High resolution and high dynamic range image reconstruction from differently exposed images. In G. Bebis and et al., editors, *Advances in Visual Computing*, volume 5359 of *Lecture Notes in Computer Science*, pages 713–722. Springer, Berlin, 2008.

[24] S.G. Narasimhan and S.K. Nayar. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530, 2005.

[25] S. Park, M. Park, and M. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20:21–36, 2003.

[26] A.A. Rad, L. Meylan, P. Vandewalle, and S. Süsstrunk. Multidimensional image enhancement from a set of unregistered differently exposed images. In *Proc. IS&T/SPIE Electronic Imaging: Computational Imaging V*, volume 6498, 2007.

[27] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann Publishers, 2005.

[28] M.A. Robertson, S. Borman, and R.L. Stevenson. Dynamic range improvement through multiple exposures. In *Proc. Sixth International*

Conference on Image Processing, volume III, pages 159–163, Kobe, Japan, October 1999.

[29] P. Sand and S.J. Teller. Video matching. In *Proc. ACM SIGGRAPH*, pages 592–599, 2004.

[30] A. Tomaszewska and R. Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *Proc. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, pages 49–56, Plzen-Bory, CZ, January 2007.

[31] R. Tsai and T. Huang. Multiframe image restoration and registration. In *Advances in Computer Vision and Image Processing*, volume 1, 1984.

[32] G. Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics, GPU, and Game Tools*, 8(2):17–30, 2003.

[33] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H.-P. Seidel. Complementary optic flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin, 2009.