Simultaneous Estimation of Surface Motion, Depth and Slopes Under Changing Illumination

Tobias Schuchert and Hanno Scharr

ICG III, Research Center Jülich, 52425 Jülich, Germany
{T.Schuchert,H.Scharr}@fz-juelich.de

Abstract. In this paper we extend a multi-camera model for simultaneous estimation of 3d position, normals, and 3d motion of surface patches [17] to be able to handle brightness changes coming from changing illumination. In the target application only surface orientation and 3d motion are of interest. Thus color related surface properties like bidirectional reflectance distribution function do not need to be reconstructed. Consequently we characterize only changes of the brightness using a second-order power series. We test two new models within a total least squares estimation framework using synthetic data with ground truth available. Motion estimation results improve severely with respect to the brightness constancy model when brightness changes are present in the data.

1 Introduction

Our target application is plant leaf growth analysis at a time range of minutes and spatial resolution of several micrometers. Growth is the divergence of the motion vector field projected onto the leaf surface, thus we need very accurate subpixel motion estimates. As temporal resolution is not 30Hz 'real-time' but minutes, image acquisition at multiple camera positions may be done by a single camera mounted on a moving stage as long as overall acquisition time for one 'time instance' is only a few seconds. Thus we can use elaborate camera setups at low cost like e.g. a 5×5 camera grid with grid-spacing even smaller than physical camera dimensions instead of really using 25 cameras (as e.g. in [13]).

In our experiments we illuminate the scene using 880nm light emitting diodes resulting in a directed, but not completely homogeneous illumination. We are restricted to this, as plants react on visible light. While this is no issue for 3d reconstruction, it is a major problem when measuring motion using a brightness constancy assumption. When plant leaves grow, they change their position and surface orientation with respect to the stationary illumination. Even if brightness changes due position change could be suppressed by optimally homogeneous illumination, brightness changes due to surface orientation change remain significant. These changes depend on the bidirectional reflectance distribution function (BRDF) of the leaf surface, thus there is no way to suppress this change experimentally without disturbing the plant.

Related Work. Estimating parameters of dynamic scenes like 3d surface position and orientation as well as motion of objects is a problem central to computer

F.A. Hamprecht, C. Schnörr, and B. Jähne (Eds.): DAGM 2007, LNCS 4713, pp. 184–193, 2007.

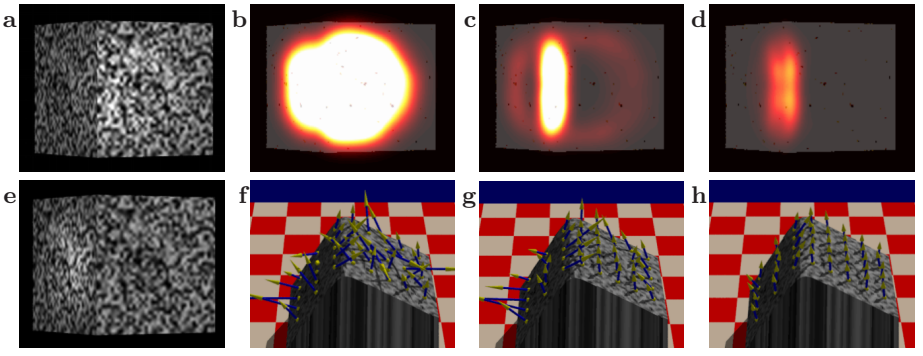

Fig. 1. Motion estimation of cube moving towards camera with spot light moving around cube center. (**a**, **e**): first and last image taken with central camera. (**b**–**d**): color coded model errors (projected on contrast reduced cube) for models without (**b**), constant temporal (**c**), and spatially varying temporal brightness change (**d**). Below the model errors, scaled motion estimates for the models are depicted, respectively (**f**–**h**).

vision research. For subpixel motion estimation as needed here, as well as stereo reconstruction optical flow techniques are applied successfully since many years [10,12] and became more and more accurate [1,9,2,15]. More complex models like affine motion [5,6], scene flow [20] and physics-based brightness changes [4,7] have been proposed. Stereo reconstruction extensions to curved surfaces [11] and depth estimation via optical flow and epipolar geometry [18] have been presented recently. Simultaneous motion and stereo analysis are addressed in [19,21,3,17]. The currently richest optical-flow-like model for local scene reconstruction [17] handles translational motion of slanted surfaces. There the basic idea is to interpret the camera position (s_x, s_y) as additional data dimensions. Hence all image sequences (x-y-t data blocks) acquired by a 2d camera grid are interpreted as a 5d-Volume in x-y-s_x-s_y-t-space. The scene model boils down to be an affine optical flow model with 3 dimensions (s_x, s_y, t) behaving like time dimension in an usual affine optical flow model.

Our Contribution. Two extensions of that model [17] need to be addressed in order to be applicable to our application: rotational motion and brightness changes. In the current paper we only deal with brightness changes. Therefore we closely follow [17] in the derivation of the geometrical part of the model, including all approximations, even though we will have to change them for rotational motion in future work. Thus, we will not look at rotating objects under nonmoving illumination in this paper, but all our tests use translating objects and rotating illumination. This means that only synthetic sequences fully fulfill the model presented here. Thus we are restricted to synthetic data for now. Without modeling brightness changes motion estimates are corrupted by illumination changes, cmp. Fig. 1**b**,**f**. Being an optical-flow-like model we follow [7] for physics-based brightness changes. The models derived there assume spatially constant brightness change parameters

leading to severe inaccuracies when illumination changes spatially (Fig. 1**c**,**g**). We therefore also model spatial changes of temporal changes leading to more accurate motion estimates (Fig. 1**d**,**h**).

Paper organization. We derive the differential model including brightness changes in Sec. 2 followed by a description of parameter estimation and disentangling of parameters (Sec. 3). We then present experiments showing the performance of the known and new models for various brightness changes (Sec. 4).

2 Derivation of the Model Equation

This section derives the constraint equation describing local changes in data acquired with a camera grid, following [17]. It combines a 3d object/motion model, a camera model and a brightness change model. For completeness we briefly present the full derivation, but focus on the brightness change model.

The dynamic surface patch is modeled by its geometry $\boldsymbol{X}$, which can be described by its initial world coordinate position (X_0, Y_0, Z_0), velocity (U_x, U_y, U_z) and X- and Y-slopes Z_x and Z_y (i.e. surface normal $(-Z_x, -Z_y, 1)$)

$$\boldsymbol{X}(\Delta X, \Delta Y, t) = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X_0 + U_x t + \Delta X \\ Y_0 + U_y t + \Delta Y \\ Z_0 + U_z t + Z_x \Delta X + Z_y \Delta Y \end{pmatrix} \tag{1}$$

with time t and local world coordinates $(\Delta X, \Delta Y)$. It is projected into the images by pinhole cameras at world coordinates $(s_x, s_y, 0)$, looking in Z-direction

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} X - s_x \\ Y - s_y \end{pmatrix} \tag{2}$$

A camera grid samples camera position space equidistantly. The cameras convert light intensity L into image intensities I (i.e. gray values). In order to derive a model for $\mathrm{d}I/\mathrm{d}t$, the temporal changes visible in the data, we look into the dependencies of L. In this paper, a translating surface patch is illuminated by a spatially smoothly varying, translating and rotating light source (see Sec. 1). Direction $\boldsymbol{n}_i$ of incident irradiance E may vary smoothly with time and space but reflectance direction $\boldsymbol{n}_r$ is kept constant.[1] Visible light intensity i.e. reflected radiance L depends on incident irradiance E and on the patch's bidirectional reflectance distribution function (BRDF) B (cmp. e.g. [8]) according to

$$L(\boldsymbol{X}(\Delta X, \Delta Y, t), t, \boldsymbol{n}_r) = B(\boldsymbol{X}(\Delta X, \Delta Y, t), \boldsymbol{n}_i(t), \boldsymbol{n}_r) E(\Delta X, \Delta Y, t, \boldsymbol{n}_i(t)) \tag{3}$$

and the BRDF depends on the material and hence on the position on the surface patch as well as the directions of incidence $\boldsymbol{n}_i$ and reflectance $\boldsymbol{n}_r$. We assume that the material does not change with time and therefore

$$B(\boldsymbol{X}(\Delta X, \Delta Y, t), \boldsymbol{n}_i(t), \boldsymbol{n}_r) = B(\boldsymbol{X}(\Delta X, \Delta Y, 0), \boldsymbol{n}_i(t), \boldsymbol{n}_r) \tag{4}$$

[1] Reflectance direction $\boldsymbol{n}_r$ obviously also varies with pixel position in the cameras, but for this paper we do not use this extra information.

If the BRDF is smooth enough, which is typically given at sufficient angular distance from specularities, changes due to smoothly changing incidence direction $\boldsymbol{n}_i(t)$ can be modeled using a smooth function $h_B(t)$ with $h_B(0) = 1$

$$B(\boldsymbol{X}(\Delta X, \Delta Y, t), \boldsymbol{n}_i(t), \boldsymbol{n}_r) = B(\boldsymbol{X}(\Delta X, \Delta Y, 0), \boldsymbol{n}_i(0), \boldsymbol{n}_r) h_B(t) \tag{5}$$

Being spatially inhomogeneous the moving irradiance E changes not only by a time dependent factor, but by a factor also varying smoothly in space

$$E(\Delta X, \Delta Y, t, \boldsymbol{n}_i(t)) = E(\Delta X, \Delta Y, 0, \boldsymbol{n}_i(0)) h_E(\Delta X, \Delta Y, t) \tag{6}$$

Here again $h_E(\Delta X, \Delta Y, t)$ is a smooth function with $h_E(\Delta X, \Delta Y, 0) \equiv 1$. Plugging Eq. 5 and Eq. 6 in Eq. 3 the reflected radiance L becomes

$$L(\boldsymbol{X}(\Delta X, \Delta Y, t), t) = L(\boldsymbol{X}(\Delta X, \Delta Y, 0), 0) h_B(t) h_E(\Delta X, \Delta Y, t) \tag{7}$$

We assume image intensities I to be proportional to the radiance L, i.e. the characteristic curve of the used camera to be linear, and therefore

$$I(\boldsymbol{X}(\Delta X, \Delta Y, t), t, s_x, s_y) = I(\boldsymbol{X}(\Delta X, \Delta Y, 0), 0, s_x, s_y) \exp(h_I(\Delta X, \Delta Y, t)) \tag{8}$$

where $h_I(\Delta X, \Delta Y, t) := ln(h_B(t) h_E(\Delta X, \Delta Y, t))$. The sought for temporal derivative of Eq. 8 is thus

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t} I &= I(\boldsymbol{X}(\Delta X, \Delta Y, 0), 0, s_x, s_y) \exp(h_I(\Delta X, \Delta Y, t)) \tfrac{\mathrm{d}}{\mathrm{d}t} h_I(\Delta X, \Delta Y, t) \\ &= I(\boldsymbol{X}(\Delta X, \Delta Y, t), t, s_x, s_y) \tfrac{\mathrm{d}}{\mathrm{d}t} h_I(\Delta X, \Delta Y, t) \end{aligned} \tag{9}$$

The most common assumption in optical-flow-like approaches is brightness constancy, boiling down to $h_I(\Delta X, \Delta Y, t) \equiv 0$. Haussecker and Fleet [7] derive models for changing surface orientation and a moving illumination envelope approximating h_I as a second order power series respecting temporal changes only

$$h_I(\Delta X, \Delta Y, t) \approx h_{HF}(t, \boldsymbol{a}) := \sum_{i=1}^{2} a_i t^i \tag{10}$$

where a_1 and a_2 are treated as local constants in the estimation process. Looking at Fig. 1**f** and **g** we observe that for highest accuracy this is not sufficient. Therefore we introduce a more accurate approximation of h_I explicitly modeling spatial variations still respecting $h_I(\Delta X, \Delta Y, 0) \equiv 0$

$$h_I(\Delta X, \Delta Y, t) \approx h(\Delta X, \Delta Y, t, \boldsymbol{a}) := \sum_{i=1}^{2} (a_i + a_{i,x} \Delta X + a_{i,y} \Delta Y)\, t^i \tag{11}$$

The temporal derivative of h is then

$$f(\Delta X, \Delta Y, t, \boldsymbol{a}) := \frac{\mathrm{d}}{\mathrm{d}t} h(\Delta X, \Delta Y, t, \boldsymbol{a}) = \sum_{i=1}^{2} i\, (a_i + a_{i,x} \Delta X + a_{i,y} \Delta Y)\, t^{i-1} \tag{12}$$

using the notation $\boldsymbol{a} = (a_1, a_2, a_{1,x}, a_{1,y}, a_{2,x}, a_{2,y})$. Following [17] the brightness change model is finally formulated as total differential $\mathrm{d}I = I f \mathrm{d}t$ or

$$I_x \mathrm{d}x + I_y \mathrm{d}y + I_{s_x} \mathrm{d}s_x + I_{s_y} \mathrm{d}s_y + I_t \mathrm{d}t = I f \mathrm{d}t \tag{13}$$

where lower indices at I indicate partial derivatives, e.g. $I_x = \partial I / \partial x$.

2.1 Combination of Patch-, Camera-, and Brightness-Models

We will now briefly summarize how to combine the dynamic surface patch (Eq. 1), camera model (Eq. 2) and the brightness change model (Eq. 13). A more detailed and comprehensive derivation can be found in [17].

Points (X, Y, Z) of a surface element (Eq. 1) are projected onto the camera chip at pixel position (x, y) via Eq. 2

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} X_0 + U_x t + \Delta X - s_x \\ Y_0 + U_y t + \Delta Y - s_y \end{pmatrix} \tag{14}$$

At fixed surface locations with constant ΔX and ΔY differentials $\mathrm{d}x$ and $\mathrm{d}y$ are

$$\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} (U_x - U_z \frac{x}{f})\mathrm{d}t - \mathrm{d}s_x \\ (U_y - U_z \frac{y}{f})\mathrm{d}t - \mathrm{d}s_y \end{pmatrix} \tag{15}$$

being nonlinear in U_z as $Z = Z_0 + U_z t + Z_x \Delta X + Z_y \Delta Y$. Using image-based expressions 3d optical flow, disparity, local pixel coordinates, and projected slopes

$$\begin{aligned} u_x &= \tfrac{f}{Z_0} U_x, \quad & u_y &= \tfrac{f}{Z_0} U_y, \quad & x &= x_0 + \Delta x, \quad & \Delta x &= \tfrac{f(1 - Z_x \frac{x}{f})}{Z_0} \Delta X, \quad & z_x &= \tfrac{Z_x}{Z_0(1 - Z_x \frac{x}{f})} \\ u_z &= -\tfrac{1}{Z_0} U_z, \quad & v &= -\tfrac{f}{Z_0}, \quad & y &= y_0 + \Delta y, \quad & \Delta y &= \tfrac{f(1 - Z_y \frac{y}{f})}{Z_0} \Delta Y, \quad & z_y &= \tfrac{Z_y}{Z_0(1 - Z_y \frac{y}{f})} \end{aligned} \tag{16}$$

omitting $U_z t$ in Z by the assumption $|Z_0| \gg |U_z t|$ and linearizing f/Z by

$$\frac{-f}{Z_0 + Z_x \Delta X + Z_y \Delta Y} \approx v + z_x \Delta x + z_y \Delta y \tag{17}$$

we get an affine-optical-flow-like model [6] when plugging all this into Eq. 13

$$\begin{aligned} &\begin{pmatrix} I_x \\ I_y \end{pmatrix} \left[\begin{pmatrix} v\mathrm{d}s_x + (u_x + x_0 u_z)\,\mathrm{d}t \\ v\mathrm{d}s_y + (u_y + y_0 u_z)\,\mathrm{d}t \end{pmatrix} + \begin{pmatrix} z_x \mathrm{d}s_x + u_z \mathrm{d}t & z_y \mathrm{d}s_x \\ z_x \mathrm{d}s_y & z_y \mathrm{d}s_y + u_z \mathrm{d}t \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right] \\ &+ I_{s_x} \mathrm{d}s_x + I_{s_y} \mathrm{d}s_y + I_t \mathrm{d}t - I f \mathrm{d}t = 0 \end{aligned} \tag{18}$$

where all nonlinear terms coming from multiplications with $z_x \Delta x$ and $z_y \Delta y$ are suppressed. We decompose Eq. 18 into data vector $\boldsymbol{d}$ and parameter vector $\boldsymbol{p}$:

$$\begin{aligned} \boldsymbol{d} &= (I_x,\, I_y,\, I_x \Delta x,\, I_x \Delta y,\, I_y \Delta y,\, I_y \Delta x, \\ &\quad\; I_{s_x},\, I_{s_y},\, I_t,\, I,\, I\Delta x,\, I\Delta y,\, It,\, It\Delta x,\, It\Delta y)^T \\ \boldsymbol{p} &= (\, v\mathrm{d}s_x + (u_x + x_0 u_z)\,\mathrm{d}t,\, v\mathrm{d}s_y + (u_y + y_0 u_z)\,\mathrm{d}t, \\ &\quad\; z_x \mathrm{d}s_x + u_z \mathrm{d}t,\; z_y \mathrm{d}s_x,\, z_y \mathrm{d}s_y + u_z \mathrm{d}t,\, z_x \mathrm{d}s_y, \\ &\quad\; \mathrm{d}s_x,\, \mathrm{d}s_y,\, \mathrm{d}t,\, b_1 \mathrm{d}t,\, b_{1,x} \mathrm{d}t,\, b_{1,y} \mathrm{d}t,\, b_2 \mathrm{d}t,\, b_{2,x} \mathrm{d}t,\, b_{2,y} \mathrm{d}t)^T \end{aligned} \tag{19}$$

where f has been substituted by the novel brightness change model from Eq. 12 and the brightness change parameters are

$$\begin{aligned} b_1 &= -a_1 \qquad & b_{1,x} &= -a_{1,x} \tfrac{Z_0}{f(1 - Z_x \frac{x}{f})} \qquad & b_{1,y} &= -a_{1,y} \tfrac{Z_0}{f(1 - Z_y \frac{y}{f})} \\ b_2 &= -a_2 \qquad & b_{2,x} &= -a_{2,x} \tfrac{Z_0}{f(1 - Z_x \frac{x}{f})} \qquad & b_{2,y} &= -a_{2,y} \tfrac{Z_0}{f(1 - Z_y \frac{y}{f})} \end{aligned} \tag{20}$$

For simpler brightness models or when a 1d camera grid is used (i.e. $\mathrm{d}s_y = 0$) terms with non-existing parameters are simply omitted. Eq. 18 is a model equation of the form $\boldsymbol{d}^T \boldsymbol{p} = 0$ (cmp. Eq. 19).

3 Parameter Estimation

Also for total least squares parameter estimation we closely follow [17]. For every 5d-pixel a constraint equation of the form $\boldsymbol{d}^{\mathrm{T}}\boldsymbol{p} = 0$ is given. To get an over-determined system of equations, we assume that all equations within a local neighborhood Ω are solved by the same parameter vector, i.e. $\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{p} = e_i$ for all pixels i in Ω, with errors e_i. The errors are minimized in weighted L_2-norm

$$||\boldsymbol{e}|| = ||\mathbf{D}\boldsymbol{p}|| = \boldsymbol{p}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{W}\mathbf{D}\boldsymbol{p} =: \boldsymbol{p}^{\mathrm{T}}\mathbf{J}\boldsymbol{p} \stackrel{!}{=} \min \tag{21}$$

with $\mathbf{D}_{ij} = (\boldsymbol{d}_i)_j$ and a diagonal matrix $\mathbf{W}$ containing the weights. As in [17] Gaussian weights with variance σ^2 are used (see Sec. 4). The matrix $\mathbf{J}$ is called structure tensor. For a 2d camera grid the space of solutions $\tilde{\boldsymbol{p}}$ is spanned by the 3 eigenvectors to the smallest eigenvalues of $\mathbf{J}$. From these eigenvectors the sought for parameters are derived by linear combination of the eigenvectors $\tilde{\boldsymbol{p}}$ such that all but exactly one component of $\{\mathrm{d}s_x, \mathrm{d}s_y, \mathrm{d}t\}$ vanish. From the linear combination with $\mathrm{d}t \neq 0$ and $\mathrm{d}s_x = \mathrm{d}s_y = 0$, we calculate motion and brightness change components. First u_z is derived and then used to calculate u_x and u_y from the first and second component of this linear combination. From the other 2 eigenvector combinations with $\mathrm{d}t = 0$ we derive depth and normals. The parameters v, z_x, z_y and u_z occur twice in the model (Eq. 18) and therefore can be estimated independently from different components and/or different linear combinations of the eigenvectors. The estimates for these parameters can be combined according to their error estimates, provided that their covariance matrix (see [14]) is diagonal. This is made sure by suitable coordinate transformations in x-y-space (for u_z) or s_x-s_y-space (for v, z_x, and z_y).

4 Experiments

We show a systematic error analysis using sinusoidal patterns, and a reconstruction of a cube with a high contrast noise texture raytraced with *povray* [16].

4.1 Sinusoidal Pattern

Sinusoidal pattern data is used to evaluate systematic errors and noise dependence of the estimation process. In Fig. 2 two images of such a sequence are shown. The wavelengths are 8 pixel in x- and 80 pixel in y-direction and amplitude changing according to Eq. 11. We generated data sets for different values of brightness change parameters a_1, $a_{1,x}$, a_2, and $a_{2,x}$, but not for $a_{1,y}$ and $a_{2,y}$ as they work like the respective x-parameters. The other parameters are $U_X = U_Y = Z_X = Z_Y = 0$, $Z_0 = 100$, $f = 10$ and $U_Z = 0.1$. As performance measure for a parameter Q we use the mean absolute value either of the relative error if $Q_{ref} \neq 0$ or of the absolute error if $Q_{ref} = 0$

$$Q_{rel} = \frac{1}{N}\sum_i^N \frac{|Q_i - Q_{ref}|}{|Q_{ref}|} \qquad Q_{abs} = \frac{1}{N}\sum_i^N |Q_i - Q_{ref}| \tag{22}$$

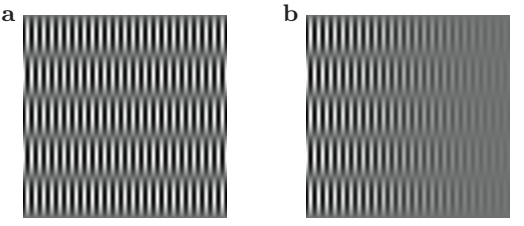

Fig. 2. Sinusoidal pattern data. (**a**, **b**): first and last image taken at central camera position, $a_1 = a_{1,x} = 0$, $a_2 = -0.2$ and $a_{2,x} = -0.002$ (cmp. Eq. 11).

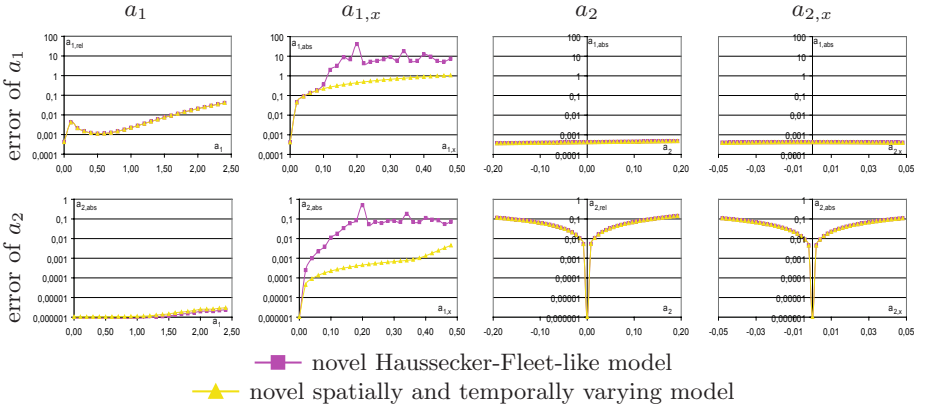

Fig. 3. Mean absolute value of relative or absolute error of brightness change parameters a_1 (top) and a_2 (bottom) versus the brightness change parameters a_1, $a_{1,x}$, a_2, and $a_{2,x}$. Noise free data.

where the sum runs over all pixels not suffering from border effects and the lower indices *rel* stand for 'relative error', *abs* for 'absolute error' and *ref* for 'reference'. Parameter estimation was done according to Sec. 3, with weighting matrix $\mathbf{W}$ implemented via a 65-tab Gaussian with standard deviation $\sigma = 16$.

The first experiment evaluates systematic error of and cross talk between brightness change parameters. In Fig. 3 errors of a_1 and a_2 versus brightness change parameters a_1, $a_{1,x}$, a_2, and $a_{2,x}$ are shown. We observe that the relative error of a_1 is well below 0.5% if $a_1 < 1$ and then moderately raises. This is due to the fact that temporal derivatives of the data I_t become less and less accurate when exponential behavior of the data becomes more and more prominent. The same explanation holds for the linear error increase of a_2 with increasing a_2. And as local brightness changes due to $a_{1,x}$ come close to changes due to a_1 if $a_1 = a_{1,x}\Delta X$ for the same local patch, we expect and observe severe cross talk between $a_{1,x}$ and a_1, more severe for the model not containing $a_{1,x}$. This is also true for $a_{2,x}$ and a_2, but there the cross talk is the same for both models, thus modeling $a_{2,x}$ is of no advantage here. Further a_1 is almost independent of a_2

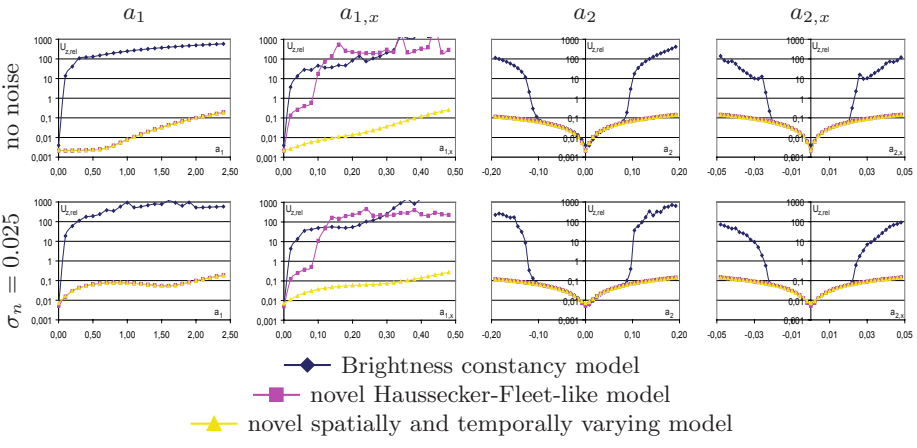

Fig. 4. Mean absolute value of relative error of U_Z versus the brightness change parameters a_1, $a_{1,x}$, a_2, and $a_{2,x}$. Noise free (top row) and noisy data (bottom).

and $a_{2,x}$, as well as a_2 of a_1. But while a_2 does depend on $a_{1,x}$ if $a_{1,x}$ is not modeled, the error of a_2 is about 1 to 2 orders of magnitude smaller if $a_{1,x}$ is modeled. The positive effect on accuracy of the method if $a_{1,x}$ is modeled is even higher for U_Z, as we see next.

In Fig. 4 results for $U_{Z,rel}$ versus brightness change parameters are shown, using noise free data and data with Gaussian noise of standard deviation $\sigma_n = 0.025$ being 2.5% of the amplitude of the signal at $t = 0$. As before all parameters except the one on the ordinate have been kept fix. U_Z is the most relevant motion parameter, because errors in U_Z directly also influence U_X and U_Y (see the first 2 components of the parameter vector $\boldsymbol{p}$ in Eq. 19). Let us first look at the noise free case. As soon as a_1 is significantly larger than 0 the brightness constancy model immediately breaks down, errors get unacceptably high. For the two other models U_Z does not react on small a_1 and only weak for larger values of a_1. When brightness changes due to $a_{1,x}$ are present only the model containing spatial changes remains stable, brightness constancy and Haussecker-Fleet-like models have severe problems. Looking at $U_{Z,rel}$ with changes due to a_2 or $a_{2,x}$ we observe that a_2 and $a_{2,x}$ cause similar errors in U_Z. This is in complete consistency with our earlier observation in Fig. 3. While all models behave the same for small absolute value of a_2 or $a_{2,x}$, the brightness constancy model rapidly breaks down at $|a_2| \approx 0.1$ or $|a_{2,x}| \approx 0.02$. Comparing errors of U_Z for noise free and noisy data sets, we see only a small effect when a_2 or $a_{2,x}$ are close to 0. For larger a_2 or $a_{2,x}$ the plots for noisy and noise free data look almost identical. Also for large a_1 and $a_{1,x}$ errors remain unchanged. But for smaller a_1 and $a_{1,x}$ the influence of noise can be quite high. We observe that errors increase from well below $U_{Z,rel} = 0.01$ up to nearly $U_{Z,rel} = 0.1$.

We conclude that modeling $a_{1,x}$ is worth the effort while $a_{2,x}$ does not really help. Noise may be an issue, thus it has to be kept as low as possible.

4.2 Synthetic Cube

Temporal image sequences with 9 images were created at 25 positions of a 2d 5×5 camera grid using *povray* [16]. For the whole cube ground truth is $U_X = U_Y =$ 0mm/frame, $U_Z = 2$mm/frame, and $Z_Y = 0$. At the left side $Z_X \approx 1.73 \hat{=} 60°$ and on the right $Z_X \approx 0.577 \hat{=} 30°$. As one can see in Fig. 1**a** and **e**, a noise texture with high contrast is mapped on the sides of the cube and in addition to the ambient illumination a spot light rotates around the center of the cube such that it moves from right to left. In Fig. 1**b-d** the numerical model error, i.e. the largest of the 3 smallest eigenvalues of the structure tensor is depicted as color overlay on the central input image. For the brightness constancy model (Fig. 1**b**) error is highest. Modeling spatially constant brightness changes (Fig. 1**c**) errors reduce, but at the edge of the cube and at the border of the spotlight they are still high. With spatially varying temporal changes errors again become smaller, visible only at the edge of the cube. The components U_X and U_Y of the motion vectors shown in Fig. 1**f-h** are scaled by a factor 135 relatively to U_Z in order to visualize estimation errors (U_X and U_Y should be 0). Even with this large accentuation of errors motion vectors estimated with the richest model point in the correct direction almost everywhere. The other models yield much less accurate vector fields.

5 Summary and Outlook

In this paper we extended the brightness constancy model presented in [17] by brightness change parameters. They are derived as a power series approximation of the changes in reflected radiance due to (1) changes of illumination direction and (2) changes in incoming light intensity caused by moving inhomogeneous incident irradiance. While the first effect may be modeled by spatially constant temporal changes, the latter one causes spatially variant temporal changes. The sinusoidal pattern experiments reveal that modeling spatial variations of brightness changes result in increased motion estimation accuracy with respect to $a_{1,x}$, but not with $a_{2,x}$ (cmp. Eq. 11). Motion vector fields of a translating cube illuminated by a moving spotlight have been estimated using brightness constancy assumption and brightness change model with or without spatial changes. The richest model yields significantly better results than the other ones.

In future work we will extend this model to be able to handle rotating objects. Rotation leads to divergence visible in the image data, currently used for the estimation of U_Z, leading to erroneous motion estimates.

References

1. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. IJCV 12(1), 43–77 (1994)
2. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. IJCV 61(3), 211–231 (2005)
3. Carceroni, R., Kutulakos, K.: Multi-view 3d shape and motion recovery on the spatio-temporal curve manifold. In: ICCV (1), pp. 520–527 (1999)

4. Denney Jr., T.S., Prince, J.L.: Optimal brightness functions for optical flow estimation of deformable motion. IEEE Trans. Im. Proc. 3(2), 178–191 (1994)
5. Farnebäck, G.: Fast and accurate motion est. using orient. tensors and param. motion models. In: ICPR, pp. 135–139 (2000)
6. Fleet, D.J., Black, M.J., Yacoob, Y., Jepson, A.D.: Design and use of linear models for image motion analysis. IJCV 36(3), 171–193 (2000)
7. Haußecker, H., Fleet, D.J.: Computing optical flow with physical models of brightness variation. PAMI 23(6), 661–673 (2001)
8. Haussecker, H.: Interaction of radiation with matter. In: Handbook of Computer Vision and Applications, vol. 1, pp. 37–62. Academic Press, London (1999)
9. Haußecker, H., Spies, H.: Motion. In: Jähne, B., Haußecker, H., Geißler, P. (eds.) Handbook of Computer Vision and Applications, Academic Press, London (1999)
10. Horn, B.K., Schunck, B.G.: Determining optical flow. Art. Int. 17, 185–204 (1981)
11. Li, G., Zucker, S.W.: Differential geometric consistency extends stereo to curved surfaces. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 44–57. Springer, Heidelberg (2006)
12. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: DARPA Im. Underst. Workshop, pp. 121–130 (1981)
13. Nakamura, Y., Matsuura, T., Satoh, K., Ohta, Y.: Occlusion detectable stereo–occlusion patterns in camera matrix. In: CVPR, pp. 371–378 (1996)
14. Nestares, O., Fleet, D.J., Heeger, D.J.: Likelihood functions and confidence bounds for total-least-squares problems. In: CVPR, pp. 523–530 (2000)
15. Papenberg, N., Bruhn, A., Brox, T., Didas, S., Weickert, J.: Highly accurate optic flow computation with theoretically justified warping. IJCV 67(2), 141–158 (2006)
16. Pov-ray 3.6, `http://www.povray.org`
17. Scharr, H., Schuchert, T.: Simultaneous motion, depth and slope estimationwith a camera-grid. In: Vision, Modeling and Visualization 2006, pp. 81–88 (2006)
18. Slesareva, N., Bruhn, A., Weickert, J.: Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps. In: Kropatsch, W.G., Sablatnig, R., Hanbury, A. (eds.) Pattern Recognition. LNCS, vol. 3663, pp. 33–40. Springer, Heidelberg (2005)
19. Szeliski, R.: A multi-view approach to motion and stereo. In: CVPR (1999)
20. Vedula, S., Baker, S., Rander, P., Collins, R., Kanade, T.: Threedimensional scene flow. In: ICCV 1999, pp. 722–729 (1999)
21. Vedula, S., Baker, S., Seitz, S., Collins, R., Kanade, T.: Shape and motion carving in 6d. In: CVPR 2000, pp. 592–598 (2000)