Joint Recursive Monocular Filtering of Camera Motion and Disparity Map

Johannes Berger and Christoph Schnörr

Image & Pattern Analysis Group, Heidelberg University
{johannes.berger@iwr.uni-heidelberg.de,schnoerr@math.uni-heidelberg.de}
ipa.math.uni-heidelberg.de

Abstract. Monocular scene reconstruction is essential for modern applications such as robotics or autonomous driving. Although stereo methods usually result in better accuracy than monocular methods, they are more expensive and more difficult to calibrate. In this work, we present a novel second order optimal *minimum energy filter* that *jointly* estimates the camera motion, the disparity map and also higher order kinematics recursively on a product Lie group containing a novel *disparity group*. This mathematical framework enables to cope with non-Euclidean state spaces, non-linear observations and high dimensions which is infeasible for most classical filters. To be robust against outliers, we use a *generalized Charbonnier energy function* in this framework rather than a quadratic energy function as proposed in related work. Experiments confirm that our method enables accurate reconstructions on-par with state-of-the-art.

Keywords: minimum energy filter, monocular reconstruction, camera motion estimation, Lie groups.

1 Introduction

1.1 Overview

Reconstruction of the scene structure of images and videos is a fundamental building block in computer vision and is required for plenty of applications, e.g. autonomous driving, robot vision and augmented reality. Although stereo methods usually lead to exact reconstruction and work fast, they require calibration of the camera setup and, due to the second camera, these systems are more expensive than single camera systems. Therefore, in this work, we will focus on the monocular approach that consists of reconstructing the scene structure based on the data gained by a *single* moving camera. In contrast to the stereo setting, this problem is ill-posed because of the unknown motion parallax. On the other hand, monocular approaches enable cheaper hardware costs.

To increase accuracy and robustness of the monocular reconstruction, we want to use temporal information for smoothing and propagation. Thus, we will introduce a mathematical framework based on *non-linear* filtering equations which describe the behavior of latent variables and the dependency between latent variables and observations. Since, in this scenario, the state variables, e.g.

camera motion, do not evolve on an Euclidean space but a more general Lie group, we cannot use classical filters, such as *extended Kalman* filters [14]. Moreover, other state-of-the-art non-linear filters, such as *particle filters* [11], that can be applied to specific Lie groups [17], cannot be easily extended to high dimensional problems [8]. Due to these mathematical problems we will use the novel *minimum energy filter* on compact Lie groups [25] that minimizes a quadratic energy function to penalize deviations of the filtering equations by means of optimal control theory. This filter was shown to be superior to extended Kalman filters on the low dimensional Lie group SE_3 [3]. We will demonstrate that this approach can also be successfully applied to high dimensional problems, enabling *joint* optimization of camera motion and disparity map. As in [3], we will also incorporate higher order kinematics of the camera motion. To be robust against outliers, we will extend the approach of [25] from quadratic energy function to a generalized Charbonnier energy function.

1.2 Related Work

Plenty of methods for depth or disparity map estimation were published during the last decade. We distinguish between stereo methods (that benefit from the additional information gained from the calibrated camera setup) and monocular methods. Recognized stereo methods include [16,22,27] that use the known distance of the cameras (baseline) for accurate triangulation of the scene. These methods also enable reducing the computational effort by using epipolar geometry and by combining local and global optimization schemes. Monocular methods [9,1,13,12,20,19,5] benefit from less calibration effort in comparison to stereo methods, but suffer from a peculiarity of the mathematical setup that prevents to reconstruct the scale of the scene uniquely. To increase the robustness and the accuracy of the reconstruction, modern methods incorporate multiple consecutive frames into the optimization procedure. Well-known is bundle adjustment [26] which optimizes a whole trajectory but cannot be used in online approaches such as sliding window [2] or filtering methods [1,5]. Filtering methods usually require a suitable modeling of the unknown *a posteriori* distribution. However, they suffer from the drawback that the definition of probability densities on non-Euclidean spaces, such as Lie groups, is complicated, although successful strategies to find a solution to this problem have been developed [6,7,17]. Zamani et al. [29] introduces so-called *minimum energy filters* for linear filtering problems for compact Lie groups based on optimal control theory and the recursive filtering principle of Mortensen [18]. This approach was generalized to (non-)compact Lie groups in [25] and applied to a *non-linear* filtering problem on SE_3 for camera motion estimation [4].

1.3 Contributions

Our contributions in this paper add up

- to provide a mathematical filtering framework for *joint* monocular camera motion and disparity map estimation including higher order kinematics,

- to introduce a *novel disparity Lie group for inverse depth maps* which avoids additional positive depth constraints such as barrier functions,
- to solve the corresponding challenging *non-linear* and *high-dimensional* filtering problem on a *product Lie group* by using novel *minimum energy filters*,
- to provide a *generalized Charbonnier* energy function instead of a quadratic energy function [25], which results in robustness against outliers .

1.4 Notation

We use the following spaces: real vector space $\mathbb{R}^n$, special orthogonal/Euclidean group $\mathrm{SO}_3, \mathrm{SE}_3$, with their corresponding Lie algebras $\mathfrak{so}_3, \mathfrak{se}_3$, as well as $\mathcal{G}$ for a general Lie group with Lie algebra $\mathfrak{g}$. Tangent spaces at a point x of $\mathcal{G}$ are denoted by $T_x\mathcal{G}$. A tangent vector $\eta \in T_x\mathcal{G}$ can be expressed in terms of a tangent vector ξ on the Lie algebra $\mathfrak{g}$ by using the tangent map of the left translation L_x evaluated at the identity element Id of the Lie group, denoted by $\eta = T_{\mathrm{Id}}L_x\xi$. We also use the shorthand $x\xi := T_{\mathrm{Id}}L_x\xi$. We use the $*-$symbol to indicate dual spaces and operators with respect to the Riemannian metric that can be defined by the tangent map as $\langle x\eta, x\xi\rangle_x := \langle \eta, \xi\rangle_{\mathrm{Id}}$ for $\eta, \xi \in \mathfrak{g}$. The dual of the tangent map is $T_{\mathrm{Id}}L_x^*\eta =: x^{-1}\eta$. We denote by $\mathrm{vec}_{\mathfrak{g}} : \mathfrak{g} \to \mathbb{R}^n, \mathrm{mat}_{\mathfrak{g}} : \mathbb{R}^n \to \mathfrak{g}$ the vectorization and its inverse operation, respectively, where the underlying Lie group $\mathcal{G}$ has dimension n. These operations allow representing the Lie algebra $\mathfrak{g}$ in a compact form. $\mathbf{D}f$ denotes the differential of a function, whereas $\mathbf{D}f(x)[\eta] := \langle \mathbf{D}f(x), \eta\rangle_x$ indicates the directional derivative for a specific direction η. For compactness, we write $\mathbf{D}_i f(x, y, z)$ for the differential of the function f respective the $i-$th component, whereas $\mathbf{D}_y f(x, y, z)$ directly addresses a specific variable. $\mathrm{Hess}\, f[\cdot]$ stands for the Riemanian Hessian on the considered Lie group; the calculation of the latter requires the Riemannian connection ∇ that can be expressed in terms of a connection function ω on the Lie algebra $\mathfrak{g}$.

2 Model

In this section we will introduce the mathematical framework of joint monocular camera motion and disparity map estimation from the point of view of (stochastic) filtering. Note, that we will use the notion *disparity map* for the inverse of the depth map in this work without using the baseline that is required in stereo settings. In classical filtering theory one wants to determine the most likely state of an unknown process $x = x(t)$ modeled by a perturbed differential equation $\dot{x}(t) = f(x(t)) + \delta(t)$ based on prior perturbed observations $y(s) = h(x(s)) + \epsilon(s)$ for $s \leq t$, which results in a *maximum a posteriori* problem. In this work, we require the state space of x to be a Lie group $\mathcal{G}$ which we need to describe non-Euclidean expressions such as camera motions. Using the expressions $\delta = \delta(t)$ and $\epsilon = \epsilon(t)$ to represent model noise and observations noise, respectively, the resulting filtering equations can be written as

$$\dot{x}(t) = x(t)\big(f(x(t)) + \delta(t)\big)\,, \quad x(t_0) = x_0\,, \tag{1}$$

$$y(t) = h(x(t)) + \epsilon(t)\,. \tag{2}$$

The state equation (1) is modeled on a Lie Group $\mathcal{G}$ by means of the tangent map of the left translation at identity and functions $f, \delta \in \mathfrak{g}$ such that $\dot{x}(t) \in T_x\mathcal{G}$. In the following sections we will introduce the state space of x, the propagation functions f and the observation function h.

2.1 State Space

The camera motion is modeled on the Special Euclidean group $\mathrm{SE}_3 := \{\left(\begin{smallmatrix} R & w \\ 0 & 1 \end{smallmatrix}\right) | R \in \mathrm{SO}_3, w \in \mathbb{R}^3\}$, and we also use a higher order kinematics (e.g. acceleration of camera) modeled by a vector $v \in \mathbb{R}^6$. The disparity map can be represented by a large vector $d_i \in \mathbb{R}^{|\Omega|}$, resulting in an own dimension for each pixel in the image. However, the depth must always be positive and we want to avoid additional constraints within our optimization. Therefore, we introduce a novel Lie group for the inverse of the depth, denoted by $(0,1)^{|\Omega|}$ which is defined as follows:

Definition 1 (Lie group $(0,1)^n$ (Disparity group)). *By denoting $d_i(z,t) := \frac{1}{d(z,t)} \in (0,1)$ the inverse of the depth we define the Lie group $(0,1)^n$ with group action for $x, y \in (0,1)^n$ as*

$$x \circ y \mapsto \big((x^{-1} - \mathbf{1}) \cdot (y^{-1} - \mathbf{1}) + \mathbf{1}\big)^{-1} = \frac{xy}{\mathbf{1} - x - y + 2xy}\,.$$

The (Lie group inverse) can be computed as $i(x) := \mathbf{1} - x$. This results in the identity element $\mathrm{Id} = \frac{\mathbf{1}}{\mathbf{2}}$, i.e. a vector full of $1/2$. The exponential map $\mathrm{Exp}_{(0,1)^n} : \mathbb{R}^n \to (0,1)^n$ and the logarithmic map $\mathrm{Log}_{(0,1)^n} : \mathbb{R}^n \to (0,1)^n$ are given through $x \mapsto \frac{e^{\mathbf{4}x}}{\mathbf{1}+e^{\mathbf{4}x}}$ and $x \mapsto \frac{1}{4}\log\left(\frac{x}{\mathbf{1}-x}\right)$, respectively. All operations apply componentwise to the vectors involved.

Using SE_3 for the camera motion, $\mathbb{R}^6$ for the acceleration of the camera and the Lie group given through definition 1 for the disparity map, we find the product Lie group $\mathcal{G}$ for our state space, i.e.

$$\mathcal{G} := \mathrm{SE}_3 \times \mathbb{R}^6 \times (0,1)^{|\Omega|}\,. \tag{3}$$

2.2 Propagation of the Camera Motion

For propagation of the camera we will use a second order kinematic model that can be expressed as second order differential equation on SE_3 as in [3], which is

$$\begin{aligned} \dot{E}(t) &= E(t)\,\mathrm{mat}_{\mathfrak{se}}(v(t)), \quad E(t_0) = E_0\,, \\ \dot{v}(t) &= \mathbf{0}, \quad v(t_0) = v_0\,, \end{aligned} \tag{4}$$

where $E = E(t) \in \mathrm{SE}_3$ and $v = v(t) \in \mathbb{R}^6$.

Remark 1. Since E describes the local camera motion from frame to frame, a first order model $\dot{E}(t) = \mathbf{0}$ corresponds to a constantly moving camera, i.e. with constant velocity. Thus, the model (4) describes a constant acceleration in the global camera frame.

2.3 Discrete Propagation of the Disparity Map

The propagation consists of mapping the image grid forward by an estimate of the motion $\hat{E} = (\hat{R}, \hat{w})$, by cubic interpolation of the depth on the irregular grid and back-projection of the resulting scene points. This leads to the following algorithm, which is also depicted in Figure 1.

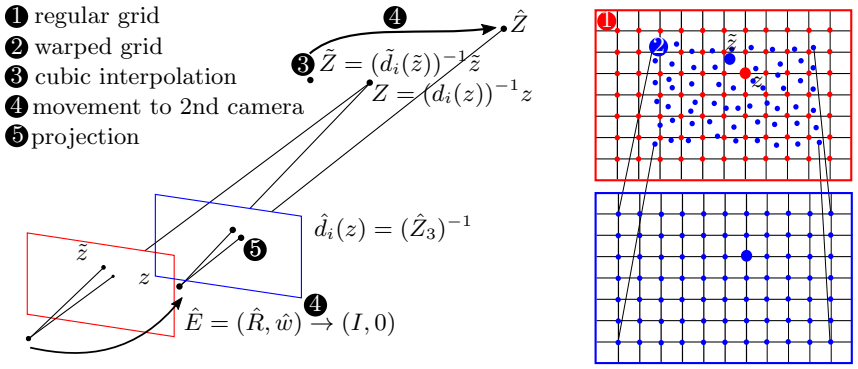

Fig. 1. Discrete propagation of the disparity map

1. Start with the disparity map d_i on regular image grid in camera $(I, 0)$.
2. Warp the image grid forward into next image (camera estimate $\hat{E}(t)$) by using current disparity map d_i to get a grid with points $\tilde{z} = \pi(\hat{R}\binom{z}{1}(d_i(z))^{-1} + \hat{w})$, where $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ is given through $(z_1, z_2, z_3)^\top \mapsto (z_3)^{-1}(z_1, z_2)^\top$.
3. Perform cubic interpolation on the warped grid $\tilde{z}$ given the values $(z, d_i(z))$ which gives the new depth map $(\tilde{z}, \tilde{d}_i(\tilde{z}))$ in frame $(I, 0)$.
4. Move $\tilde{Z} = \binom{\tilde{z}}{1}(\tilde{d}_i(\tilde{z}))^{-1}$ to second camera to obtain $\hat{Z} = \hat{R}^\top(\tilde{Z} - \hat{w})$.
5. Recognize the propagated disparity map as third component, $\hat{d}_i(z) = (\hat{Z}_3)^{-1}$.

2.4 Camera Motion and Disparity Map induced Optical Flow

Since the state space $\mathcal{G}$ consists of the camera motion $E(t)$ and the disparity map $d_i(\cdot, t)$ (inverse of depth map), we require observations that depend on both variables. It is well-known that from a given disparity map and a given camera motion the correspondences between a pair of consecutive images expressed as *optical flow* can be uniquely determined if the scene is static. To be precise, the dependency between the optical flow vector $u(z, t)$ at a position $z \in \Omega$ can be expressed with the following *non-linear* relation, where we denote by $R(t) \in \mathrm{SO}_3$ and $w(t) \in \mathbb{R}^3$ the rotational and translational component of the camera motion $E(t) = (R(t), w(t)) \in \mathrm{SE}_3$, respectively. For details see [1, Eq. (6)].

$$u(z, t) + z = \pi\Big(R(t)\binom{z}{1}(d_i(z, t))^{-1} + w(t)\Big). \tag{5}$$

2.5 Overall Filtering Model

The function $f(x(t)) : \mathcal{G} \to \mathfrak{g}$ in (1) can now be defined as follows:

$$f(x(t)) := (f_E(x(t)), f_v(x(t)), f_{d_i}(x(t)))\,, \tag{6}$$

with component functions $f_E(x(t)) := \mathrm{mat}_{\mathfrak{se}}(v(t))$, and $f_v(x(t)) := \mathbf{0_6}$ as in (4) as well as $f_{d_i}(x(t)) := \mathbf{0}_{|\Omega|}$. Beside this continuous propagation step we also incorporate discrete updates of the disparities as described in section 2.3.

By setting $y_z(t) := u(z,t) - z$ and $h_z : \mathcal{G} \to \mathbb{R}^2$, as the right hand side of (5), we find the following observation equations by adding noise $\epsilon_z(t) \in \mathbb{R}^2$ for all $z \in \Omega$.

$$y_z(t) = h_z(x(t)) + \epsilon_z(t)\,, \qquad z \in \Omega\,. \tag{7}$$

2.6 Objective Function

Minimum energy filtering requires to define an energy function that penalizes the model and observation noise. In contrast to [24], that we will follow in this work, we will not use quadratic energy functions but an energy function that is a smooth approximation of the L^1-norm. The reason is that we want to reduce the influence of outliers in the observations that may cause numerical problems because the gradient grows linearly. The norm of the gradient of the proposed L^1 penalty function is bounded. A smooth approximation to the non-differentiable L^1-norm is the generalized charbonnier penalty function that is smooth (C^∞) and has linear growth, such that we use it for ϕ, i.e. $\phi(x) := (x+\nu)^\beta - \nu^\beta$. With this notation and the shorthand $\|x\|_Q^2 := x^\top Q x$ the energy function reads

$$\mathcal{J}(\delta,\epsilon,x;t) := \tfrac{1}{2}\|x - x_0\|^2_{R_0^{-1}} + \int_{t_0}^{t} \Big(\tfrac{1}{2}\|\mathrm{vec}_{\mathfrak{g}}(\delta(\tau))\|^2_{R^{-1}} + \sum_{z\in\Omega} \phi(\tfrac{1}{2}\|\epsilon_z(\tau)\|^2_{Q_z^{-1}}) \Big) d\tau, \tag{8}$$

where Q_z, R_0 and R are symmetric and positive definite matrices.

2.7 Optimal Control Problem

After replacing the observation noise $\epsilon_z(t)$ by the residual $\epsilon_z(t) = \epsilon_z(x(t),t) := y_z(t) - h_z(x,t)$ in (8) we want to minimize the energy function $\mathcal{J}(\delta,x,x(t_0);t) = \mathcal{J}(\delta,\epsilon(x),x(t_0);t)$ regarding the model noise $\delta(t)$ with respect to the differential equation (1) yielding the *value function*

$$\mathcal{V}(x(t),t,x(t_0)) := \min_{\delta|_{[t_0,t]}} \mathcal{J}(\delta,x;t) \quad \text{subject to (1).} \tag{9}$$

Calculation of the value function requires to introduce the time-varying (left-trivialized) Hamiltonian function $\tilde{\mathcal{H}} : \mathcal{G} \times \mathfrak{g}^* \times \mathfrak{g} \times \mathbb{R} \to \mathbb{R}$ that is given through

$$\begin{aligned} \tilde{\mathcal{H}}(x,\mu,\delta,t) := \big(\tfrac{1}{2}\|\mathrm{vec}_{\mathfrak{g}}(\delta(t))\|^2_{R^{-1}} + \sum_{z\in\Omega} \phi(\tfrac{1}{2}\|y_z(t) - h_z(x(t)))\|^2_{Q_z} \big) \\ - \langle \mu, f(x(t)) + \delta(t) \rangle_{\mathrm{Id}}\,. \end{aligned} \tag{10}$$

Owing to the *Pontryagin minimum principle* [21] we find the minimizing argument of the value function (9) by minimizing the Hamiltonian $\tilde{\mathcal{H}}$ with respect to δ. Since the Hamiltonian is convex in δ we obtain a unique minimum $\delta^* = \mathrm{mat}_{\mathfrak{g}}\left(R\,\mathrm{vec}_{\mathfrak{g}}(\mu)\right)$ resulting in the optimal Hamiltonian $\mathcal{H}(x,\mu,t) : \mathcal{G}\times\mathfrak{g}^*\times\mathbb{R}\to\mathbb{R}$ given through $\mathcal{H}(x,\mu,t) := \tilde{\mathcal{H}}(x,\mu,\delta^*,t)$ such that

$$\mathcal{H}(x,\mu,t) = -\left\langle \mu, f(x(t)\right\rangle_{\mathrm{Id}} - \tfrac{1}{2}\|\mathrm{vec}_{\mathfrak{g}}(\mu)\|_R^2 + \sum_{z\in\Omega}\left(\phi\left(\tfrac{1}{2}\|y_z(t) - h_z(x(t))\|_{Q_z}^2\right)\right).$$

In the case of a linear-quadratic control problem this optimal Hamiltonian satisfies the (left-trivialized) Hamilton-Jacobi-Bellman equation, i.e.

$$\frac{\partial}{\partial t}\mathcal{V}(x,t) - \mathcal{H}(x, x^{-1}\mathbf{D}_1\mathcal{V}(x,t), t) = 0\,. \tag{11}$$

Here, $\mathbf{D}_1\mathcal{V}(x,t) \in T_x^*\mathcal{G}$ is an element of the cotangent space.

Remark 2. Note that our control problem has neither *linear control dynamics* nor a *quadratic* energy function. Thus, we have no guarantee that the HJB equation is a necessary *and sufficient* condition for optimality. Instead we require a good initialization to gain an optimal reconstruction. However, we will show that a fairly general initialization will lead to good reconstructions.

2.8 Recursive Filtering Principle and Truncation

Computation of the total time derivative of the necessary condition

$$\mathbf{D}_1\mathcal{V}(x,t,x(t_0)) = \mathbf{0}\,,$$

and insertion of the HJB equation (11) leads to the following lemma that gives a recursive description of the optimal state $x^* = x^*(t)$ (cf. [25, Eq. (37)]).

Lemma 1. *The evolution equation of the optimal x^* state is given through*

$$\dot{x}^*(t) = x(t)\Big(f(x^*(t)) - \hat{Z}(x^*(t),t)^{-1}\circ x^{-1}(\mathbf{D}_1\mathcal{H}(x^*(t),\mathbf{0},t))\Big)\,, \tag{12}$$

where $\hat{Z} : \mathfrak{g}\to\mathfrak{g}^*$ is the left-trivialized Hessian of the value function given through

$$\hat{Z}(x^*,t)\circ\eta = (x^*)^{-1}\,\mathrm{Hess}\,\mathcal{V}(x^*(t),t,x(t_0))[x^*\eta]\,,\quad \eta\in\mathfrak{g}\,. \tag{13}$$

Because the non-linear filtering problem is infinite dimensional we will replace the exact operator $\hat{Z}$ by an approximation $Z : \mathfrak{g}\to\mathfrak{g}^*$ which can be obtained by truncation of the full evolution equation of Z. But still the operator $Z(x^*,t)$ on $\mathfrak{g}$ is complicated such that we introduce a matrix representation $P(t)$ that is defined through the relation $\mathrm{vec}_{\mathfrak{g}}(Z(x^*,t)^{-1}\circ\eta) =: P(t)\,\mathrm{vec}_{\mathfrak{g}}(\eta)$.

Lemma 2. *The matrix representation of the approximation of the operator $\hat{Z}$ evolves regarding the following matrix Riccati equation*

$$\dot{P}(t) = R + C(x^*, t)P(t) + P(t)C(x^*, t)^\top - P(t)H(x^*, t)P(t), \tag{14}$$

where the matrix R is the weighting matrix in the energy function (8) *and the matrices C and H are given for $\eta \in \mathfrak{g}$ through*

$$\begin{aligned} C(x^*, t)P(t)\,\mathrm{vec}_{\mathfrak{g}}(\eta) &:= \mathrm{vec}_{\mathfrak{g}}\big((x^*)^{-1}\mathbf{D}_2(\mathbf{D}_1\mathcal{H}(x^*, \mathbf{0}, t))[Z(x^*, t) \circ \eta]\big) \\ &\quad + \mathrm{vec}_{\mathfrak{g}}(\omega^{\rightleftharpoons *}_{\mathbf{D}_2\mathcal{H}(x^*,\mathbf{0},t)} \circ Z(x^*, t) \circ \eta) + \mathrm{vec}_{\mathfrak{g}}(\omega^*_{(x^*)^{-1}\dot{x^*}} \circ Z(x^*, t) \circ \eta)\,, \\ H(x^*, t)\,\mathrm{vec}_{\mathfrak{g}}(\eta) &:= \mathrm{vec}_{\mathfrak{g}}((x^*)^{-1}\,\mathrm{Hess}_1\,\mathcal{H}(x^*, \mathbf{0}, t)[x\eta])\,. \end{aligned}$$

Here, $x\omega_\xi\eta := \nabla_{x\xi}x\eta$ denotes the connection function on the Lie algebra $\mathfrak{g}$ of the Levi-Civita connection $\nabla_{\cdot}\cdot$ for $\xi, \eta \in \mathfrak{g}$ and $x \in \mathcal{G}$, and $\omega^{\rightleftharpoons *}_\xi$ is the dual of the "swaped" connection function $\omega^{\rightleftharpoons}_\xi\eta := \omega_\eta\xi$ (cf. [24]).

By insertion of the expression P into (12) and by evaluation of the expressions in Lemma 1 and 2 we obtain the final minimum energy filter that consists of continuous propagation of the states with a discrete update of the disparity map.

Theorem 1. *The second order minimum energy filter with additional discrete propagation step for the disparity map is given through the following evolution equations of the optimal state $x^* \in \mathcal{G}$ as well as the second order operator $P \in \mathbb{R}^{(12+|\Omega|)\times(12+|\Omega|)}$.*

$$\dot{x}^*(t) = x^*(t)\big(f(x^*(t)) - \mathrm{mat}_{\mathfrak{g}}(P(t)\,\mathrm{vec}_{\mathfrak{g}}(G(x^*(t), t)))\big), \tag{15}$$

$$\dot{P}(t) = R + C(x^*, t)P(t) + P(t)C(x^*, t)^\top - P(t)H(x^*, t)P(t), \tag{16}$$

with initial conditions $x^*(t_0) = x_0$ and $P(t_0) = R_0$, where R_0 is the matrix in (8)*. $G(x^*, t) = (G_E(x^*), \mathbf{0}, G_{d_i}(x^*)) \in \mathfrak{g}$ denotes the Riemannian gradient of the Hamiltonian in* (12) *with components G_E and G_{d_i}.*

The numerical integration of these equations between the time steps t_{k-1} and t_k correspond to the update step of a filter, where the updates are assumed to be piecewise constant. After each update step the disparity map is propagated forward using the procedure in Fig. 1 that result in the final filter.

Remark 3. The expressions for $C(x^*, t), H(x^*, t)$ and $G(x^*, t)$ can be calculated explicitly but require matrix calculus and differential geometry. The resulting expressions become involved such that we refer the interested reader to the supplemental material[1].

Remark 4. The optimal state can be calculated by geometric numerical integration of the ordinary differential equations (15) and (16), e.g. Crouch-Grossman methods (cf. [15]). During numerical integration it is important to keep the matrix P sparse, therefore we set the off-diagonal entries of the lower right part of P (that addresses the disparities) after each iteration to zero.

[1] `http://hciweb.iwr.uni-heidelberg.de/people/johannesberger`

3 Experiments

Preprocessing As stated above, our method requires precise optical flow as input. Since we propose a monocular method we also demand that the optical flow is computed from two consecutive image frames without stereo information. For this reason we used the well-known *EpicFlow* approach [23]. The matches are computed with *Deep Matching* [28]; the required edges are from [10].

Choice of the weighting matrices Monocular methods suffer from the fact that observations that appear close to the epipole (focus of expansion) are orthogonal to the camera motion such that these regions cannot be reconstructed correctly. Therefore we use the weighting term from [1, Eq. (14)] for the weighting matrix Q that decreases the influence of the data term in regions close to the epipole.

Outlier detection To remove outliers, we computed the backward flow from frame i to $i+1$ as well as the forward flow from frame $i+1$ to i. In regions where these flows are not consistent with each other, we decreased the weight of the term R such that the filter has less ability to fit to the data and the discrete disparity map propagation from section 2.3 reduces the error.

Scale correction As monocular approaches cannot estimate the scale of a scene without prior knowledge about invariants in the scene, we corrected the scale by calculating of the pixel-wise quotient of the disparities and taking its median as scale $s := \text{median}\{d_i^{\text{gt}}(z,t)/d_i^{\text{est}}(z,t) | z \in \Omega^*\}$, where Ω^* denotes the image domain without points which are close to the epipole (< 50 pixel distance).

3.1 Qualitative Results

In Fig. 2 we compared the reconstruction of the disparity map of our method with the results from [1] and the ground truth. One can recognize that our method preserves small details and depth discontinuities better than [1] and returns sharper edges.

3.2 Quantitative Results

We evaluated the mean amount of pixels in Ω^* with a disparity error larger than three pixel for both occluded and not occluded scenarios in Table 1. We are slightly inferior towards Becker et al. [1]; however, unlike [1] we do not have spatial regularization within our optimization which explains the differences.

Table 1. Evaluation of the mean disparity errors.

	p_{3px}[%] (occ)	p_{5px}[%] (occ)	p_{3px}[%] (noc)	p_{5px}[%] (noc)
Becker et al. [1]	17.74	10.82	17.63	10.72
our approach	19.24	10.69	19.14	10.59

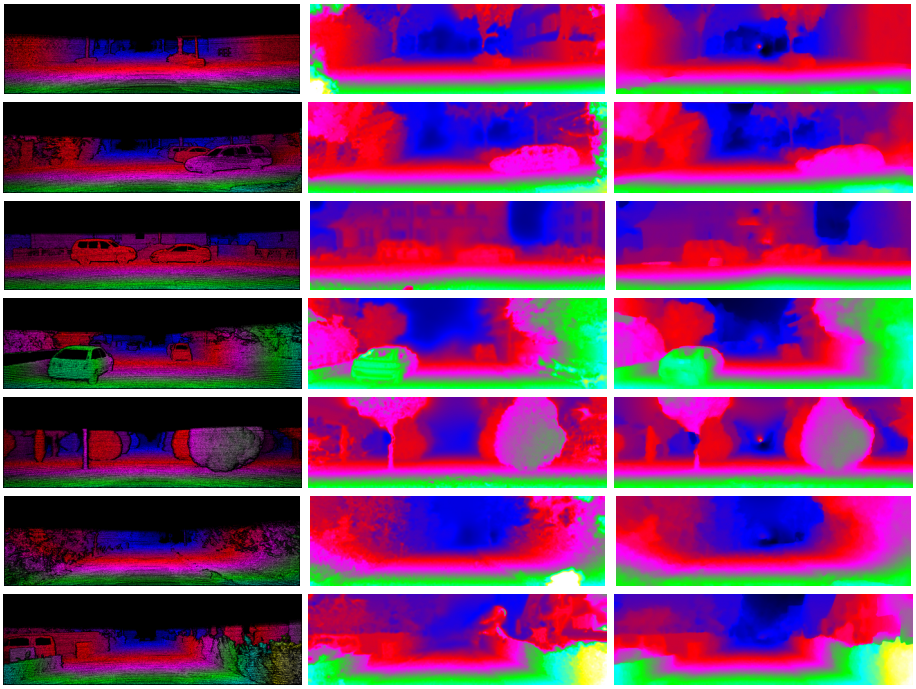

Fig. 2. *Best viewed in color.* Reconstruction of the disparity maps; left column: ground truth from the KITTI stereo benchmark, middle column: monocular method of Becker et al. [1], right column: reconstruction with our monocular method. Although in the quantitative evaluation both methods perform equally, one can recognize that our method results in sharper corners. Due to spatial regularization [1] reconstructs regions close to the epipole better.

4 Conclusion

We provided a sound mathematical filtering framework for monocular scene reconstruction based on novel minimum energy filters, extending the classical quadratic energy function from Saccon et al. [25] to a generalized Charbonnier energy function. We demonstrated that the proposed filter copes with challenging mathematical issues, such as a non-Euclidean state space, non-linear filtering equations based on projections, as well as high dimensions; in fact, these difficulties are infeasible for most classical stochastic filters. The introduced *disparity group* enables filtering without additional constraints making the model relatively compact. Our experiments confirmed that the proposed filter is almost as accurate as other state-of-the-art monocular and recursive methods without having an own regularization within the model.

References

1. F. Becker, F. Lenzen, J. H. Kappes, and C. Schnörr. Variational Recursive Joint Estimation of Dense Scene Structure and Camera Motion from Monocular High Speed Traffic Sequences. *IJCV*, 105:269–297, 2013.
2. F. Bellavia, M. Fanfani, F. Pazzaglia, and C. Colombo. Robust Selective Stereo SLAM without Loop Closure and Bundle Adjustment. In *Image Analysis and Processing–ICIAP 2013*, pages 462–471. Springer, 2013.
3. J. Berger, F. Lenzen, F. Becker, A. Neufeld, and C. Schnörr. Second-Order Recursive Filtering on the Rigid-Motion Lie Group SE(3) Based on Nonlinear Observations, 2015. ArXiv, preprint.
4. J. Berger, A. Neufeld, F. Becker, F. Lenzen, and C. Schnörr. Second Order Minimum Energy Filtering on SE(3) with Nonlinear Measurement Equations. In *SSVM*, pages 397–409. Springer, 2015.
5. G. Bourmaud and R. Mégret. Robust Large Scale Monocular Visual SLAM. In *CVPR*, pages 1638–1647, 2015.
6. G. Bourmaud, R. Mégret, M. Arnaudon, and A. Giremus. Continuous-Discrete Extended Kalman Filter on Matrix Lie Groups Using Concentrated Gaussian Distributions. *Journal of Mathematical Imaging and Vision*, 51(1):209–228, 2015.
7. Y. Chikuse. *Statistics on Special Manifolds*, volume 174. Springer Science & Business Media, 2012.
8. F. Daum and J. Huang. Curse of Dimensionality and Particle Filters. In *Aerospace Conference*, 2003.
9. A. J Davison, I. D Reid, N. D Molton, and O. Stasse. MonoSLAM: Real-Time Single Camera SLAM. *PAMI*, 29(6):1052–1067, 2007.
10. Piotr Dollár. Piotr's Computer Vision Matlab Toolbox (PMT). `http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html`.
11. A. Doucet, N. Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*, chapter An Introduction to Sequential Monte Carlo Methods, pages 3–14. Springer New York, New York, NY, 2001.
12. J. Engel, T. Schöps, and D. Cremers. LSD-SLAM: Large-Scale Direct Monocular SLAM. In *ECCV*, pages 834–849. Springer, 2014.
13. J. Engel, J. Sturm, and D. Cremers. Semi-Dense Visual Odometry for a Monocular Camera. In *ICCV*, pages 1449–1456. IEEE, 2013.
14. P. Frogerais, J. Bellanger, and L. Senhadji. Various Ways to Compute the Continuous-Discrete Extended Kalman Filter. *Automatic Control, IEEE Transactions on*, 57:1000–1004, 2012.
15. E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*, volume 31. Springer Science & Business Media, 2006.
16. H. Hirschmüller. Stereo Processing by Semiglobal Matching and Mutual Information. *PAMI*, 30(2):328–341, 2008.
17. J. Kwon, M. Choi, Frank C. Park, and C. Chun. Particle Filtering on the Euclidean Group: Framework and Applications. *Robotica*, 25(6):725–737, 2007.
18. R. E. Mortensen. Maximum-Likelihood Recursive Nonlinear Filtering. *J. Opt. Theory Appl.*, 2(6):386–394, 1968.
19. A. Neufeld, J. Berger, F. Becker, F. Lenzen, and C. Schnörr. Estimating Vehicle Ego-Motion and Piecewise Planar Scene Structure from Optical Flow in a Continuous Framework. In *GCPR*, 2015.
20. M. Pizzoli, C. Forster, and D. Scaramuzza. REMODE: Probabilistic, Monocular Dense Reconstruction in Real Time. In *ICRA*, pages 2609–2616. IEEE, 2014.

21. L. S. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko. The Mathematical Theory of Optimal Processes. *Interscience Publishers, Inc.*, 1962.
22. E. T Psota, J. Kowalczuk, M. Mittek, and L. C Perez. MAP Disparity Estimation Using Hidden Markov Trees. In *ICCV*, pages 2219–2227, 2015.
23. J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow. In *CVPR*, 2015.
24. A. Saccon, J. Trumpf, R. Mahony, and A P. Aguiar. Second-Order-Optimal Filters on Lie groups. In *CDC*, 2013.
25. A. Saccon, J. Trumpf, R. Mahony, and A. P. Aguiar. Second-Order-Optimal Minimum-Energy Filters on Lie Groups. *IEEE TAC*, PP(99):1–1, 2015.
26. B. Triggs, P. F McLauchlan, R. I Hartley, and A. W Fitzgibbon. Bundle Adjustment – A Modern Synthesis. In *Vision Algorithms: Theory and Practice*, pages 298–372. Springer, 2000.
27. C. Vogel, K. Schindler, and S. Roth. 3D Scene Flow Estimation with a Piecewise Rigid Scene Model. *IJCV*, 115(1):1–28, 2015.
28. P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. Deepflow: Large Displacement optical Flow with Deep Matching. In *ICCV*, pages 1385–1392, 2013.
29. M. Zamani, J. Trumpf, and M. Mahoney. A Second Order Minimum-Energy Filter on the Special Orthogonal Group. In *Proc. ACC*, 2012.