Global Solutions of Variational Models with Convex Regularization[*]

Thomas Pock[†], Daniel Cremers[‡], Horst Bischof[†], and Antonin Chambolle[§]

Abstract. We propose an algorithmic framework for computing global solutions of variational models with convex regularity terms that permit quite arbitrary data terms. While the minimization of variational problems with convex data and regularity terms is straightforward (using, for example, gradient descent), this is no longer trivial for functionals with nonconvex data terms. Using the theoretical framework of calibrations, the original variational problem can be written as the maximum flux of a particular vector field going through the boundary of the subgraph of the unknown function. Upon relaxation this formulation turns the problem into a convex problem, although in a higher dimension. In order to solve this problem, we propose a fast primal-dual algorithm which significantly outperforms existing algorithms. In experimental results we show the application of our method to outlier filtering of range images and disparity estimation in stereo images using a variety of convex regularity terms.

1. Introduction. Energy minimization methods have had great success for a number of computer vision problems [22, 31, 39]. The basic idea of energy minimization methods is that the solution of a problem corresponds to the minimizer of a so-called energy functional. The success of energy minimization methods is therefore subject to two lines of research: first, the design of appropriate objective functions to model the characteristics of the problem and second, the development of efficient optimization algorithms to compute minimizers of respective energy functionals.

Basically, energy minimization problems can be divided into two fundamentally different classes: convex and nonconvex problems. The main advantage of convex problems is that a global optimum can be computed, generally with good precision and in a reasonable time. This means that the quality of the solution depends solely on the appropriateness of the model which gives rise to the energy functional. On the other hand, for nonconvex problems,

the quality of the solution is subject to both the model and the optimization algorithm: In general only a local minimizer can be computed, the quality of which typically depends on the initialization and choice of tuning parameters.

In computer vision there are two different philosophies in treating images. In the spatially discrete approach, image pixels are assumed to be discrete entities, whereas in the continuous approach, images are defined as functions on a continuous domain. While much work in computer vision has been done in the discrete setting [16], one of the first advances in the continuous setting was taken by Mumford and Shah [31].

1.1. Discrete setting. In the discrete approach, the theory of Markov random fields (MRFs) provides a mathematically consistent way to describe the graph structure which is used to represent the grid points and connections between the grid points [26]. In the MRF approach image pixels can adopt a finite number of states, also called labels. Given a graph with node set $\mathcal{V}$ and edge set $\mathcal{E}$ and a label set $L \subset \mathbb{Z}$, the typical task is to find an optimal labeling $l \in L^{\mathcal{V}}$ for an energy of the form

$$(1.1) \qquad \min_l \left\{ \sum_{(u,v) \in \mathcal{E}} P(l(u), l(v)) + \sum_{v \in \mathcal{V}} U(l(v)) \right\}.$$

Such a labeling problem combines a pairwise regularity term $P(\cdot, \cdot)$ with a unary data term $U(\cdot)$.

If the set of possible labels is binary, and the pairwise terms are submodular, combinatorial algorithms such as graph cuts can be used to compute the global minimizer [20, 29]. On the other hand, multilabel problems generally cannot be globally minimized. In general, they can only be solved approximately, for example, by transforming the problem into a sequence of binary labeling problems [7, 40], by linear programming (LP) relaxations [6, 41], or via roof duality relaxation [21], which has recently attracted renewed interest in the MRF community (cf. [37]).

A notable exception is the work of Ishikawa and Geiger [25, 24], who showed that exact solutions for certain multilabel problems can be computed in polynomial time. Provided that the pairwise interaction terms are convex functions of the differences $l(u) - l(v)$ with respect to a linearly ordered label set, respective problems can be solved globally as binary cuts of a higher-dimensional graph.

If both the prior term and the data term are submodular, (1.1) reduces to the convex cost tension problem [28], which can be solved in the original problem domain.

1.2. Spatially continuous setting. The continuous counterpart to discrete labeling problems is the variational approach. Similar to the discrete labeling problem, the aim of the variational approach is to find minimizers for energy functionals $F : L^1(\Omega) \to [0, \infty]$ of the form

$$(1.2) \qquad \min_u \left\{ F(u) = \int_\Omega f(x, u(x), \nabla u(x)) \, dx \right\},$$

where Ω is a d-dimensional bounded open subset of $\mathbb{R}^d$ and $u : \Omega \to \mathbb{R}$ is an unknown scalar function. For $d = 2$, Ω is usually assumed to be a rectangular image domain. The Lagrangian

$f(x, t, p^x)$ is the "core" of the energy functional and is used to model the characteristics of the energy functional.

The calculus of variations provides a framework for finding the solution of a continuous minimization problem such as (1.2). A local minimizer (if it exists) can be computed by solving its associated Euler–Lagrange partial differential equation. However, a global minimizer can be computed only if (1.2) is a convex function of u.

In this work we focus on a specific class $F(u)$ of energy functionals (1.2) which are continuous in x, t and convex in p^x of $f(x, t, p^x)$. This is the spatially continuous analogue of the class addressed by Ishikawa and Geiger in the discrete setting. Typical applications for this type of energy include disparity estimation and image restoration. In general, this class of energy functionals cannot be solved globally due to the lack of convexity in t. Nevertheless, we show that a global minimizer can be computed by representing the original variational problem in higher dimensions. Our approach builds on works of Alberti, Bouchitté, and Dal Maso [1] and Chambolle [8]. It extends our previous work [34] to a more general class of convex regularity terms.

1.3. Contributions. Our work has strong connections to the works, in the discrete setting, of Ishikawa and Geiger [25, 24] and of Roy and Cox [38]. It is based on the same idea of increasing the dimensionality of the problem. Nevertheless, the proposed spatially continuous formulation offers the following advantages:

- For general convex regularizers (e.g., quadratic), Ishikawa and Geiger's approach requires the introduction of additional long range edges in the graph structure. This increases the density of the graph structure, reducing the efficiency of standard graph cut algorithms. On the other hand, the proposed approach, which can be tackled using standard finite-difference methods, merely requires local couplings for any convex regularizer.
- Our method is largely independent from grid bias, also known as metrication error. This leads to more accurate approximations of the continuous solution and allows for subpixel accurate solutions. A quantitative and qualitative comparison between the proposed method and Ishikawa's method was already presented in [34] for the particular case of total variation regularization. An experimental comparison of discrete and continuous shape optimization was presented in [27].
- Our method is based on simple and efficient primal-dual optimization techniques which can be easily accelerated on parallel architectures such as graphics processing units (GPUs). On the other hand, an efficient parallelization of max–flow-type algorithms is still an open problem [19]. Furthermore, it requires considerably less memory. Thisalso makes our method applicable for quite large practical problems.

The remainder of the paper is organized as follows. In section 2 we review the approach of [34] and show connections to the work presented in this paper. In section 3 we present the theoretical framework underlying the proposed approach. We show how the originally nonconvex energy functional can be transformed into a convex optimization problem. In section 4 we give details of the proposed framework for four different convex regularity terms. In particular, we study quadratic, total variation, Huber, and Lipschitz regularization since these are of particular interest in a number of computer vision applications. The algorithmic

framework for computing minimizers to the convex problem is presented in section 5. Specifically, we propose a provably convergent primal-dual algorithm which clearly outperforms existing algorithms. In section 6 we show the application of our approach to outlier filtering in industrial range images and stereo.

2. Related work. In [34], an approach is presented to solve minimization problems of the form

$$(2.1) \qquad \min_u \left\{ \int_\Omega |\nabla u| dx + \int_\Omega g(x, u(x)) dx \right\},$$

on a domain $\Omega \subset \mathbb{R}^2$, where $u : \Omega \to \Delta$ takes on values in a range $\Delta = [t_0, t_1] \subset \mathbb{R}$. The left term denotes the total variation of u, which is a popular regularizer in a number of imaging problems. The right term is a pointwise generally nonconvex data term. In this sense (2.1) is a special case of the more general class of problems (1.2) that we consider in this paper.

The idea of [34] is then to rewrite (2.1) by means of the upper level sets of the function u, which results in an anisotropic minimal surface problem of the form

$$(2.2) \qquad \min_{\hat{v} \in \mathcal{C}} \left\{ \int_{\Omega \times \Delta} |\nabla_x \hat{v}| + g(x, t)|\nabla_t \hat{v}| \, dxdt \right\},$$

where

$$(2.3) \qquad \mathcal{C} = \left\{ \hat{v}(x, t) : \Omega \times \Delta \to \{0, 1\}, \, \hat{v}(\cdot, t_0) = 1, \, \hat{v}(\cdot, t_1) = 0 \right\}.$$

Using duality, (2.2) is then transformed to a problem of the form

$$(2.4) \qquad \min_{\hat{v} \in \mathcal{C}} \left\{ \sup_{\phi \in \mathcal{K}} \int_{\Omega \times \Delta} \phi \cdot \nabla \hat{v} \, dxdt \right\},$$

with the dual variable ϕ constrained to the convex set

$$(2.5) \quad \mathcal{K} = \left\{ \phi(x, t) = (\phi^x(x, t), \phi^t(x, t)) : \Omega \times \Delta \to \mathbb{R}^3, \, |\phi^x(x, t)| \le 1, \, |\phi^t(x, t)| \le g(x, t) \right\}.$$

After relaxation of the binary functions $\hat{v}$ to continuous functions (i.e., the convex hull of the set $\mathcal{C}$), the resulting convex problem is solved using an Arrow–Hurwicz-type algorithm [3].

In the next section, we will describe the theoretical framework for the convex representation of functionals of the form (1.2). It builds upon the same basic idea of [34]—rewriting the functional by means of its upper level sets—but while the approach of [34] can be applied only to total variation regularization, the theoretical framework proposed in this paper extends to the general case of arbitrary convex regularizers.

3. Convex representation. Our approach is based on a general theoretical framework which is quite classical in the calculus of variations, yet less known in the optimization community. The basic concept is the idea of *Cartesian currents* [17, 18], which consists of taking the whole graph $(x, u(x))$ of a function as the *object* to optimize upon, rather than the function u itself. It is related to the so-called theory of *calibrations*, which was recently brought back
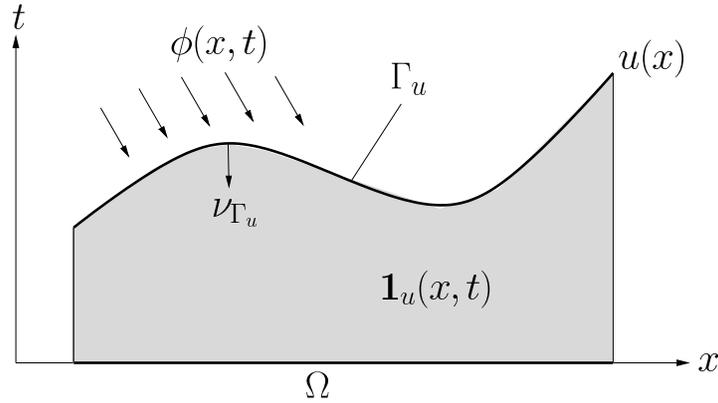
Figure 1. *The central idea of the proposed approach is to determine functions $u(x)$ implicitly by means of their higher-dimensional subgraph $\mathbf{1}_u$, which is equal to 1 in the shaded area and 0 otherwise. To this end, respective functionals on u are expressed as the flux of a vector field $\phi(x,t)$ through the membrane Γ_u representing the graph of u, where different functionals are encoded by corresponding convex constraints on $\phi(x,t)$. Subsequent relaxation gives rise to a convex optimization problem.*

to light by Alberti, Bouchitté, and Dal Maso in [1] as an approach to characterizing the minimizers of the Mumford–Shah functional [31] by an implicit (and novel) convex representation. Their approach allows one to actually characterize (some) minimizers of the Mumford–Shah functional by means of divergence-free vector fields in higher dimensions. A similar general framework was introduced in [8], where it was observed that, roughly speaking, many functionals of a scalar function in L^1 could be minimized by finding the solution of a convex functional in higher dimension. We summarize in this section some known results that lead to an interesting representation which allows us to tackle numerically a class of nonconvex problems of the form (1.2).

Let us start by considering the subgraph of the function $u(x)$, which is the collection of all points lying below the function value $u(x)$. (Figure 1 shows an example for a one-dimensional function $u(x)$, where the subgraph is represented as the gray area.) We also introduce the function $\mathbf{1}_u(x,t) : \Omega \times \mathbb{R} \to \{0,1\}$ which is the characteristic function of the subgraph of $u(x)$:

$$(3.1) \qquad \mathbf{1}_u(x,t) = \begin{cases} 1 & \text{if} \quad u(x) > t, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, let us denote by Γ_u the boundary of $\mathbf{1}_u(x,t)$. For the sake of simplicity, we assume first that u is smooth: In this case, Γ_u is nothing but the graph $\{(x,u(x)) : x \in \Omega\}$. The key idea is that the energy in (1.2) can be seen as an *interfacial* energy of the boundary Γ_u. Let ν_{Γ_u} denote the inner unit normal to Γ_u, which is given by

$$(3.2) \qquad \nu_{\Gamma_u} = \frac{1}{\sqrt{1 + |\nabla u(x)|^2}} \begin{pmatrix} \nabla u(x) \\ -1 \end{pmatrix}.$$

By interfacial energy, we mean an energy of the form

$$(3.3) \qquad \int_{\Gamma_u} h(x,t,\nu_{\Gamma_u}(x)) d\mathcal{H}^d(x),$$

where we integrate over the graph of u a Lagrangian h that may depend on the point (x, t), with $t = u(x)$ and the normal ν_{Γ_u}. Here $\mathcal{H}^d$ denotes the d-dimensional Hausdorff measure. It can be shown that (3.3) defines a lower semicontinuous (l.s.c.) energy (with respect to L^1 convergence of the characteristics $\mathbf{1}_u$) as soon as h is continuous in (x, t) and convex, one-homogeneous with respect to the last argument (the normal vector).

Let us for the moment assume that u is sufficiently smooth; i.e., u is in the Sobolev space $W^{1,1}(\Omega; \mathbb{R})$—the space of functions whose weak derivatives up to order 1 have finite L^1 norm [4]. Then, by a simple change of variable, one can express the integral on Γ_u in (3.3) as an integral over Ω, and the Jacobian of the mapping $\Gamma_u \ni (x, t) \mapsto x$ is precisely $\sqrt{1 + |\nabla u|^2}$ so that

$$\int_{\Gamma_u} h(x, t, \nu_{\Gamma_u}(x)) d\mathcal{H}^d(x) = \int_\Omega h(x, u(x), (\nabla u(x), -1)) \, dx,$$

where we have used (3.2) and the one-homogeneity of h with respect to its last argument. Hence, the representation holds as soon as $h(x, t, (p^x, -1)) = f(x, t, p^x)$. This leads us to introduce the *Lagrangian* $h(x, t, p)$, defined for $(x, t, p) \in \Omega \times \mathbb{R} \times \mathbb{R}^{d+1}$, with $p = (p^x, p^t) \in \mathbb{R}^d \times \mathbb{R}$, as

$$(3.4) \qquad h(x, t, p) = \begin{cases} |p^t| f(x, t, p^x/|p^t|) & \text{if } p^t < 0, \\ f^\infty(x, t, p^x) & \text{if } p^t = 0, \\ +\infty & \text{if } p^t > 0, \end{cases}$$

where $f^\infty(x, t, p^x) := \lim_{\lambda \to +\infty} f(x, t, \lambda p^x)/\lambda$ is the *recession function* of f. This function h is shown to be l.s.c., convex, and one-homogeneous with respect to its last argument $p = (p^x, p^t)$; see, for instance, [12, 18].

In general, when f has superlinear growth in p^x, energy (1.2) is (weakly) coercive and l.s.c. on $W^{1,1}(\Omega)$, and hence we will have $f^\infty(x, t, p^x) = +\infty$ (except at $p^x = 0$ where it is always 0). However, many interesting cases include situations where f has linear growth in p^x. For instance, the case of total variation regularization with some additional data term, where $f(x, t, p^x) = |p^x| + g(x, t)$. Then, (1.2) has to be properly extended to functions in $BV(\Omega)$, the space of functions of bounded variation. In contrast to the Sobolev space $W^{1,1}(\Omega)$ we considered above, the space $BV(\Omega)$ also allows for functions having sharp discontinuities [4]. In this case, we have $f^\infty(x, t, p^x) = |p^x|$, and it turns out that the equality

$$\int_{\Gamma_u} h(x, t, \nu_{\Gamma_u}(x)) d\mathcal{H}^d(x) = \int_\Omega f(x, u(x), Du)$$

also holds when $u \in BV(\Omega)$ has jumps, which correspond to *vertical* parts in the graph Γ_u. Here, Du denotes the distributional derivative of u, which in an integral sense is also well defined for discontinuous functions (e.g., characteristic functions). Since h is convex and one-homogeneous with respect to its last argument p, we can introduce the convex, one-homogeneous functional

$$(3.5) \qquad \mathcal{F}(v) = \int_{\Omega \times \mathbb{R}} h(x, t, Dv) \, ,$$

defined for any $v \in BV_{\text{loc}}(\Omega \times \mathbb{R})$. Now, for $v = \mathbf{1}_u$, one has

$$(3.6) \qquad D\mathbf{1}_u = \nu_{\Gamma_u}(x,t)\mathcal{H}^d \llcorner \Gamma_u(x,t) \,.$$

It follows that for any $u \in W^{1,1}(\Omega)$,

$$(3.7) \qquad \mathcal{F}(\mathbf{1}_u) = \int_{\Gamma_u} h(x,t,\nu_{\Gamma_u}(x))d\mathcal{H}^d(x) = \int_\Omega f(x,u(x),\nabla u(x))\,dx \,.$$

In summary, the solution of (1.2) is equivalent to the minimization of the convex functional $\mathcal{F}$,

$$(3.8) \qquad \min_{u \in W^{1,1}(\Omega)} \mathcal{F}(\mathbf{1}_u) \,,$$

however, over a nonconvex set of binary functions.

3.1. Convex relaxation. We now show that we can we replace the function $\mathbf{1}_u$ in (3.8) by a more general function $v \in \mathcal{C}$, where the convex set $\mathcal{C}$ is given by

$$(3.9) \qquad \mathcal{C} = \left\{ v \in BV(\Omega \times \mathbb{R}; [0,1]) \;:\; \lim_{t \to -\infty} v(x,t) = 1, \; \lim_{t \to +\infty} v(x,t) = 0 \right\}.$$

This set is related to the convexification of $\{\mathbf{1}_u : u \in W^{1,1}(\Omega)\}$; however, in the latter, the functions are allowed to decrease only in the vertical direction t. In our problem, though, $\mathcal{F}$ is easily seen to be $+\infty$ for functions in $\mathcal{C}$ which increase in the vertical direction t in some set with positive measure. We now consider the relaxed problem

$$(3.10) \qquad \min_{v \in \mathcal{C}} \mathcal{F}(v),$$

which is equivalent to minimizing $\mathcal{F}$ over the convex hull of the binary functions $\mathbf{1}_u$. Note that this is analogous to the concept of LP relaxation in discrete optimization. Our intention is still to solve the binary problem. Hence, the question remains: In which sense the minimizers of (3.10) and (3.8) are related? In fact, one checks that a simple thresholding of the solution of the relaxed problem produces a solution of the binary problem, in analogy to what has been observed in past years [11, 9].

Theorem 3.1. *Let v^* be a global minimizer of (3.10). Then for any $s \in [0,1)$ the characteristic function $\mathbf{1}_{\{v^*>s\}}$ is also a global minimizer of (3.8). Furthermore, the function $\mathbf{1}_{\{v^*>s\}}$ is the characteristic of the subgraph of a minimizer of (1.2).*

Proof. As we shall show below, the functional $\mathcal{F}$ satisfies the generalized coarea formula:

$$(3.11) \qquad \mathcal{F}(v) = \int_{-\infty}^{+\infty} \mathcal{F}(\mathbf{1}_{\{v>s\}})ds.$$

Applying this to $v^* \in \mathcal{C}$, we have

$$\mathcal{F}(v^*) = \int_0^1 \mathcal{F}(\mathbf{1}_{\{v^*>s\}})ds;$$

i.e., the energy of v^* can be expressed in terms of the energies of its upper level sets. From the minimality of v^*, $\mathcal{F}(\mathbf{1}_{\{v^*>s\}}) \geq \mathcal{F}(v^*)$ for any s, it follows that for a.e. $s \in [0,1]$, also $\mathbf{1}_{\{v^*>s\}}$ is a minimizer of (3.10) or, equivalently, of (3.8). As a consequence, for any $s \in [0,1)$, one can find a sequence $(s_n)_{n\geq1}$ with $s_n > s$, $s_n \to s$, and such that $\mathbf{1}_{\{v^*>s_n\}}$ is a minimizer of (3.8). Passing to the limit, we deduce that also $\mathbf{1}_{\{v^*>s\}}$ is a minimizer.

It remains to show (3.11). In fact, it follows from the standard coarea formula for BV functions [15, 14, 43, 2]. For any v, let $\nu_v = Dv/|Dv|$ be the Besicovitch derivative of the measure Dv with respect to its variation $|Dv|$. We then have

$$\mathcal{F}(v) = \int_{\Omega \times \mathbb{R}} h(x,t,\nu_v(x,t))|Dv| = \int_{-\infty}^{+\infty} \int_{\Omega \times \mathbb{R}} h(x,t,\nu_v(x,t))|D\mathbf{1}_{\{v>s\}}|\,ds$$

$$= \int_{-\infty}^{+\infty} \int_{\Omega \times \mathbb{R}} h(x,t,D\mathbf{1}_{\{v>s\}}/|D\mathbf{1}_{\{v>s\}}|)|D\mathbf{1}_{\{v>s\}}|\,ds = \int_{-\infty}^{+\infty} \mathcal{F}(\mathbf{1}_{\{v>s\}})\,ds,$$

where we have used the fact that $\mathcal{H}^{d-1}$-a.e. on the boundary of $\{v > s\}$, $\nu_v = \nu_{\{v>s\}} = D\mathbf{1}_{\{v>s\}}/|D\mathbf{1}_{\{v>s\}}|$; that is, the gradient of v is normal to its level lines. ∎

3.2. Duality. Our approach to tackling problem (3.10) is through duality. Indeed, the Lagrangian h is in general very singular, while it turns out that its dual expression involves the projection on a rather simple set, which can often be performed with good precision by simple algorithms.

The key idea is now to consider the flux of a dual vector field $\phi = (\phi^x, \phi^t) : \Omega \times \mathbb{R} \to \mathbb{R}^d \times \mathbb{R}$ through the boundary Γ_u:

$$(3.12) \qquad\qquad \Phi = \int_{\Gamma_u} \phi \cdot \nu_{\Gamma_u} d\mathcal{H}^d;$$

see Figure 1. Using (3.6), the flux can also be written as

$$(3.13) \qquad\qquad \Phi = \int_{\Gamma_u} \phi \cdot \nu_{\Gamma_u} d\mathcal{H}^d = \int_{\Omega \times \mathbb{R}} \phi \cdot D\mathbf{1}_u.$$

The following states that $F(u) = \mathcal{F}(\mathbf{1}_u)$ can be expressed as the maximal flux of ϕ through Γ_u for ϕ in some class depending on the Lagrangian f.

Theorem 3.2. *For any function $u \in W^{1,1}(\Omega; \mathbb{R})$, we have*

$$(3.14) \qquad\qquad F(u) = \mathcal{F}(\mathbf{1}_u) = \sup_{\phi \in \mathcal{K}} \int_{\Omega \times \mathbb{R}} \phi \cdot D\mathbf{1}_u,$$

where the convex set $\mathcal{K}$ is given by

$$(3.15) \quad \mathcal{K} = \Big\{ \phi = (\phi^x, \phi^t) \in C_0\big(\Omega \times \mathbb{R}; \mathbb{R}^d \times \mathbb{R}\big) \ :$$

$$\phi^t(x,t) \geq f^*(x,t,\phi^x(x,t)), \ \forall x,t \in \Omega \times \mathbb{R} \Big\} .$$

Here, $f^*(x, t, \phi^x)$ denotes the Legendre–Fenchel conjugate (or convex conjugate) of $f(x, t, p^x)$ with respect to p^x. It is defined as

$$(3.16) \qquad f^*(x, t, \phi^x) = \sup_{p^x} \left\{ \phi^x \cdot p^x - f(x, t, p^x) \right\} .$$

In analogy, the biconjugate (the convex conjugate of the convex conjugate) is defined as

$$(3.17) \qquad f^{**}(x, t, p^x) = \sup_{\phi^x} \left\{ \phi^x \cdot p^x - f^*(x, t, \phi^x) \right\} .$$

It follows from classical arguments based on the convex separation theorem that for any (x, t), $f^{**}(x, t, \cdot)$ is the convex, l.s.c. envelope of $f(x, t, \cdot)$. In particular, when $f(x, t, p^x)$ is convex and l.s.c. in p^x, as is always assumed in this work,

$$(3.18) \qquad f^{**}(x, t, p^x) = f(x, t, p^x).$$

For more details on convex analysis we refer the reader to [36]. We now sketch the proof of Theorem 3.2 which is based on [1].

Proof. We first check that for any $\phi \in \mathcal{K}$, we have

$$(3.19) \qquad F(u) \geq \int_{\Omega \times \mathbb{R}} \phi \cdot D\mathbf{1}_u.$$

Indeed, using (3.13) and the definition of the inner unit normal (3.2), the flux can be rewritten as

$$(3.20) \quad \int_{\Omega \times \mathbb{R}} \phi \cdot D\mathbf{1}_u = \int_{\Gamma_u} \phi(x, t) \cdot \begin{pmatrix} \nabla u(x) \\ -1 \end{pmatrix} \frac{d\mathcal{H}^d(x, t)}{\sqrt{1 + |\nabla u(x)|^2}}$$

$$= \int_{\Omega} \phi^x(x, u(x)) \cdot \nabla u(x) - \phi^t(x, u(x)) \, dx,$$

where we have used as before that $\sqrt{1 + |\nabla u(x)|^2}$ is the Jacobian of the change of variable $\Gamma_u \ni (x, t) \mapsto x \in \Omega$. Since $\phi \in \mathcal{K}$, it follows that

$$\int_{\Omega \times \mathbb{R}} \phi \cdot D\mathbf{1}_u \leq \int_{\Omega} \phi^x(x, u(x)) \cdot \nabla u(x) - f^*(x, t, \phi^x(x, u(x))) \, dx,$$

which is less than or equal to $F(u)$ by definition of the convex conjugate f^*. This shows (3.19).

The proof that the supremum is actually $F(u) = \mathcal{F}(\mathbf{1}_u)$, that is, the proof of (3.14), is more technical, and we just give the main lines. Essentially, at each point $(x, u(x))$ one needs to choose $\phi^x(x, u(x)) = \nabla_p f(x, u(x), \nabla u(x))$ and $\phi^t(x, u(x)) = f^*(x, t, \phi^x(x, u(x)))$. If f, u are sufficiently smooth (essentially, C^1), then such a choice can be performed. In other cases, it is shown that (thanks to the lower semicontinuity properties of f^*) one can construct a continuous field $\phi \in \mathcal{K}$ such that its flux (3.13) is arbitrarily close to $F(u)$. ∎

Remark 3.3. In fact, the result still holds for $u \in BV(\Omega)$, a bounded variation function, and a Lagrangian $f(x, t, p^x)$ with linear growth in p at ∞, with a similar proof. It can also be

extended to Lagrangians which take the value $+\infty$, such as illustrated in Figure 2(d), with some additional regularity assumptions in (x,t).

Remark 3.4. An alternative way to show this result would be to show that for any $(x,t) \in \Omega \times \mathbb{R}$, the function $h(x,t,\cdot)$ is given by

$$(3.21) \qquad h(x,t,p) \;=\; \sup\{p \cdot q \,:\, q = (q^x, q^t) \in \mathbb{R}^d \times \mathbb{R}, q^t \geq f^*(x,t,q^x)\};$$

that is, it is the convex conjugate of the indicator function of the closed, convex set $\{q^t \geq f^*(x,t,q^x)\}$, which has the value 0 on this set and $+\infty$ elsewhere. Then, standard results of convex analysis will imply that $\mathcal{F}$ is the convex conjugate of the indicator function of $\mathcal{K}$.

From (3.14) and the coarea formula or, alternatively, from Remark 3.4, it follows that for any $v \in BV_{\mathrm{loc}}(\Omega \times \mathbb{R})$,

$$(3.22) \qquad \mathcal{F}(v) \;=\; \sup_{\phi \in \mathcal{K}} \int_{\Omega \times \mathbb{R}} \phi \cdot Dv.$$

We will actually use this form for minimizing (3.10) by a primal-dual algorithm which will be introduced in section 5. The advantage of this technique is that we never need to compute a subgradient of the singular Lagrangian h, but only need to perform orthogonal projections onto the convex set $\mathcal{K}$, which are often relatively easy to compute.

4. Convex regularity terms. In the last section we showed that the nonconvex problem (1.2) can be solved by computing the minimizer of problem (3.22). In this section we will now study explicit examples of (1.2) by concentrating on Lagrangians of the form

$$(4.1) \qquad f(x,t,p^x) = g(x,t) + h(p^x) \,,$$

where $g(x,t)$ is a general potential function which corresponds to a pointwise data term in the energy formulation, and $h(p^x)$ is a convex potential function which is used to realize different types of regularizers. We will show the application of Theorem 3.2 to four different functions: quadratic, linear (total variation), Huber, and a Lipschitz constraint, as these functions are interesting for a number of different problems in computer vision and image analysis. See also Figure 2 for an illustration of these functions. According to Theorem 3.2, the core of our approach is to derive the convex set

$$(4.2) \qquad \mathcal{K} = \{\phi = (\phi^x, \phi^t) \,:\, \phi^t(x,t) \geq f^*(x,t,\phi^x(x,t)) \forall (x,t) \in \Omega \times \mathbb{R}\} \,.$$

By exploiting the local structure of the Lagrangian $f(x,t,p^x)$, one has $f^*(x,t,\phi^x) = h^*(\phi^x) - g(x,t)$ so that the constraint on $\phi(x,t)$ in (4.2) boils down to $\phi^t(x,t) + g(x,t) \geq h^*(\phi^x(x,t))$. Hence, it will be sufficient to consider the pointwise constraints

$$(4.3) \qquad K(x,t) = \{\phi(x,t) \in \mathbb{R}^{d+1} \,:\, \phi^t(x,t) + g(x,t) \geq h^*(\phi^x(x,t))\} \,.$$

The substantial advantage of the resulting algorithm is the fact that we can efficiently compute the projection of the $(d+1)$-dimensional vector $\phi(x,t)$ onto $K(x,t)$.

(a) Quadratic function

(b) Linear function

(c) Huber function

(d) Lipschitz constraint
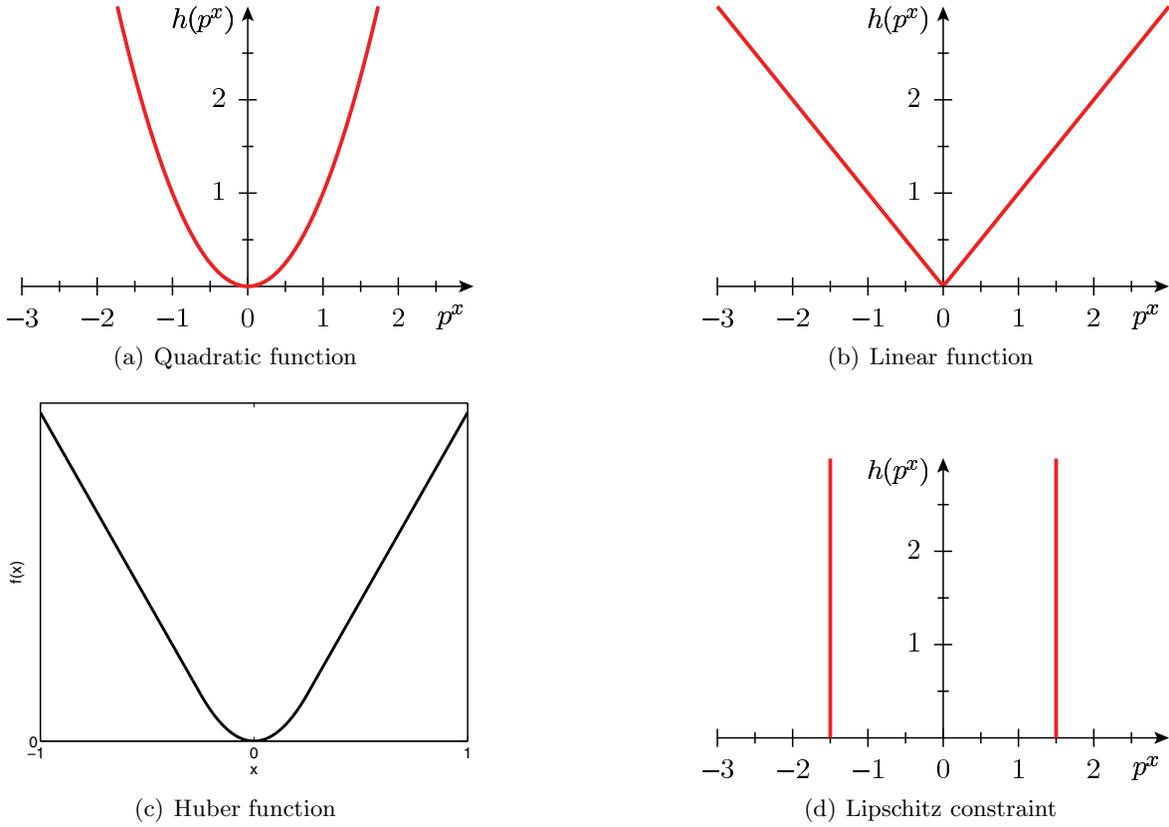
Figure 2. *Several convex potential functions.*

4.1. Quadratic regularization. Quadratic regularization can be realized by choosing

$$(4.4) \qquad h(p^x) = \frac{|p^x|^2}{2}.$$

It is well known that quadratic regularization leads to an oversmoothing of image edges. Therefore, quadratic regularization is useful only in cases where the solution is expected to be a smooth function. However, quadratic regularization is one of the most basic regularization techniques and will later be utilized in Huber regularization. According to (3.16) the convex conjugate of h is given by

$$(4.5) \qquad h^*(\phi^x) = \sup_{p^x} \left\{ \phi^x \cdot p^x - \frac{|p^x|^2}{2} \right\} = \frac{|\phi^x|^2}{2}.$$

Hence the local convex $K_q(x,t)$ which realizes quadratic regularization in (4.1) is given by

$$(4.6) \qquad K_q(x,t) = \left\{ \phi(x,t) \in \mathbb{R}^{d+1} \ : \ \phi^t(x,t) + g(x,t) \geq \frac{|\phi^x(x,t)|^2}{2} \right\}.$$

(a) Convex set K_q

(b) Convex set K_t

(c) Convex set K_h
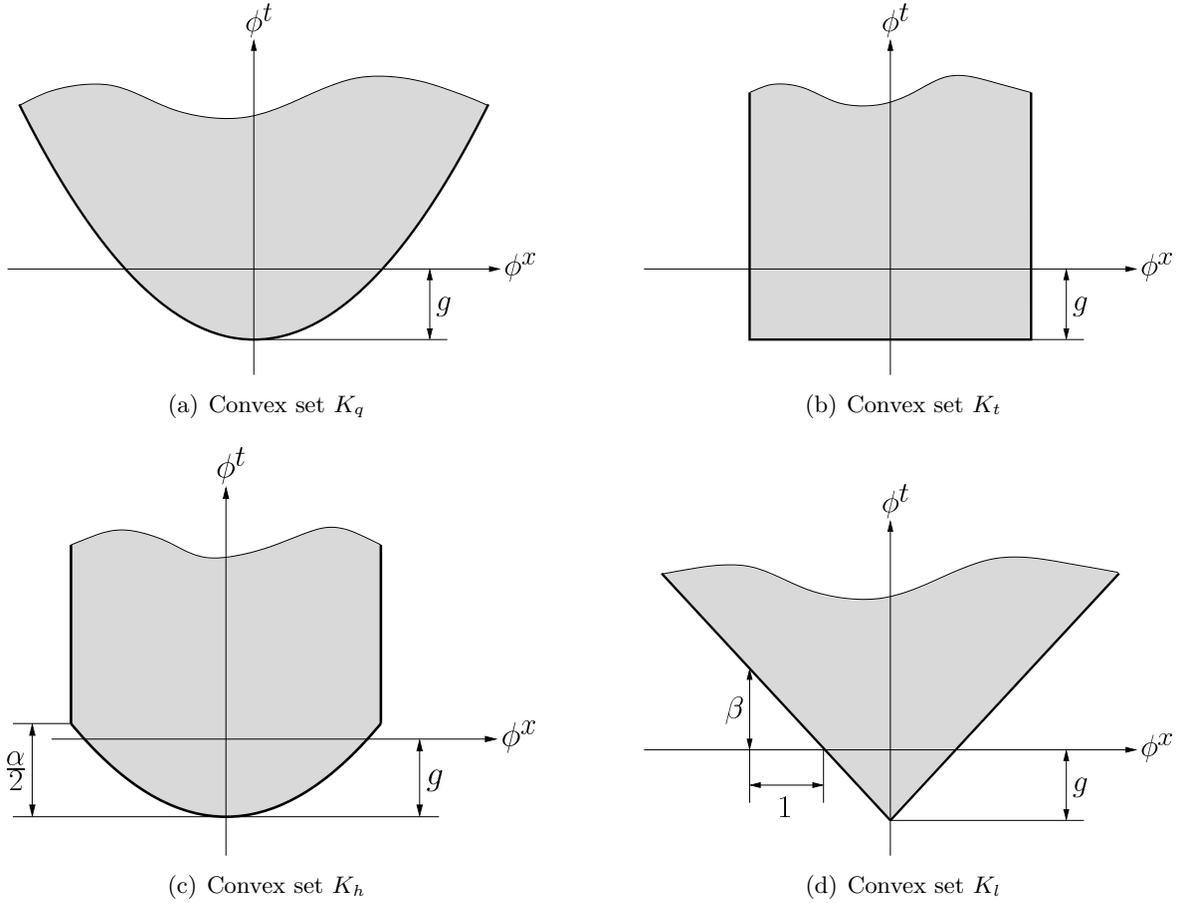
(d) Convex set K_l

Figure 3. *Illustration of the convex sets for* (a) *quadratic,* (b) *total variation,* (c) *Huber, and* (d) *Lipschitz regularization.*

Figure 3(a) illustrates the convex set $K_q(x, t)$. One can see that $K_q(x, t)$ is a paraboloid with a vertical offset of $(-g(x, t))$ from the origin. Note that the constraints obtained from quadratic regularization also appear in the calibration method for the Mumford–Shah functional [1].

4.2. Total variation regularization. Total variation regularization is obtained by choosing a linear potential function

$$(4.7) \qquad\qquad h(p^x) = |p^x| .$$

Unlike quadratic regularization, total variation regularization has the desirable property of preserving sharp discontinuities in the solution. Total variation regularization was first applied to computer vision by Rudin, Osher, and Fatemi in the seminal work on nonlinear image denoising [39]. We again compute the convex conjugate of h with respect to p^x, yielding

$$(4.8) \qquad h^*(\phi^x) = \sup_{p^x}\{\phi^x \cdot p^x - |p^x|\} = \delta_{\{|\phi^x| \leq 1\}} ,$$

where $\delta_\Sigma(\sigma)$ denotes the indicator function of the set Σ, that is,

$$(4.9) \qquad \delta_\Sigma(\sigma) = \begin{cases} 0 & \text{if} \quad \sigma \in \Sigma, \\ \infty & \text{else.} \end{cases}$$

This leads to the following characterization of the local constraint:

$$(4.10) \qquad K_t(x,t) = \left\{ \phi(x,t) \in \mathbb{R}^{d+1} \; : \; \phi^t(x,t) + g(x,t) \geq 0, \; |\phi^x(x,t)| \leq 1 \right\}.$$

Note that $K_t(x,t)$ basically consists of two simple orthogonal constraints. From an implementation point of view this is especially appealing because it will make it very easy to compute projections onto $K_t(x,t)$. Figure 3(b) shows an illustration of the set $K_t(x,t)$ which is basically a translated cylinder.

Note that the constraint (4.10) differs from the constraint used in [34]. As also seen from (2.5), the basic difference is that in [34] it is not assumed that the function v is decreasing in t. However, the equivalence of both approaches for the case of total variation regularization is shown in [10].

4.3. Huber regularization. Besides the advantage of total variation regularization in allowing for sharp discontinuities in the solution, it suffers from the so-called staircasing effect, a phenomenon which stems from the fact that total variation regularization has a tendency to produce piecewise constant solutions. A quite simple but effective method for reducing the staircasing effect is to use quadratic penalization for small values of the image gradient and to use linear penalization for larger values. This is essentially the Huber norm, which has its origin in robust statistics [23]. In our framework we have

$$(4.11) \qquad h(p^x) = |p^x|_\alpha,$$

where

$$(4.12) \qquad |p^x|_\alpha = \begin{cases} \dfrac{|p^x|^2}{2\alpha} & \text{if } |p^x| \leq \alpha, \\ |p^x| - \dfrac{\alpha}{2} & \text{if } |p^x| > \alpha \end{cases}$$

defines the Huber function, which is quadratic for small values of t and linear for large values of t, and α is a tuning parameter. Again we have to compute the convex conjugate of h with respect to p^x which gives

$$(4.13) \qquad h^*(\phi^x) = \sup_{p^x} \left\{ \phi^x \cdot p^x - |p^x|_\alpha \right\} = \delta_{\{|\phi^x| \leq 1\}} + \frac{\alpha}{2} |\phi^x|^2.$$

Then we can define the feasible set of vectors $\phi(x,t)$ which realize the Huber regularization in our convex framework:

$$(4.14) \qquad K_h(x,t) = \left\{ \phi(x,t) \in \mathbb{R}^{d+1} \; : \; \phi^t(x,t) + g(x,t) \geq \frac{\alpha}{2} |\phi^x(x,t)|^2, \; |\phi^x(x,t)| \leq 1 \right\}.$$

Figure 3(c) shows the shape of the convex set $K_h(x,t)$. It is interesting to note that $K_h(x,t)$ is the intersection of the set $K_q(x,t)$ scaled by parameter α and the set $K_t(x,t)$.

4.4. Lipschitz regularization. In some computer vision tasks one has precise a priori information about the object to be reconstructed. For example, one knows that the local surface slope is limited by some maximal value. In such cases it is useful to enforce a so-called Lipschitz constraint, which means that no regularization is performed as long as $|\nabla u| \leq \beta$, where β is the Lipschitz bound, while higher gradients are penalized with the value $+\infty$, and hence forbidden. A Lipschitz constraint on ∇u is obtained by choosing

$$(4.15) \qquad\qquad\qquad\qquad h(p^x) = \delta_{B(0,\beta)}(p^x) \ ,$$

where $\delta_{B(0,\beta)}$ denotes the indicator function of a d-dimensional ball centered around 0 and radius β. This situation is exactly dual to the total variation regularization, and the convex conjugate of h with respect to p^x is simply

$$(4.16) \qquad\qquad\qquad h^*(\phi^x) = \sup_{|p^x| \leq \beta} \left\{ \phi^x \cdot p^x \right\} = \beta |\phi^x|.$$

The feasible set of vectors $\phi(x,t)$ is then

$$(4.17) \qquad\qquad K_l(x,t) = \left\{ \phi(x,t) \in \mathbb{R}^{d+1} \ : \phi^t(x,t) + g(x,t) \geq \beta |\phi^x(x,t)| \right\} \ .$$

Note that $K_l(x,t)$ is essentially—pointwise—a translated cone, and the projection onto it will be straightforward to implement. Figure 3(d) shows an illustration of the set $K_l(x,t)$.

5. Numerical algorithms. In this section we present numerical algorithms in order to compute the solution of the convex problem (3.22). Before detailing the algorithms, we will introduce the discrete setting.

5.1. Discretization. We consider only the case of two-dimensional images; i.e., $d = 2$. Hence, we consider a three-dimensional (3D) Cartesian grid G^h of size $N_x \times N_y \times N_t$,

$$(5.1) \qquad\qquad G^h = \left\{ (ih_x, jh_y, kh_t) : (0,0,0) \leq (i,j,k) < (N_x, N_y, N_t) \right\} \ ,$$

where h_x, h_y, and h_t denote the discretization widths and (i,j,k) denotes the discrete locations on the grid. In the following we will use the superscript h to indicate the discrete setting. Next, let us introduce $v^h \in \mathcal{C}^h$, where

$$(5.2) \qquad\qquad \mathcal{C}^h = \left\{ v^h \in [0,1]^{N_x N_y N_t} : \ v^h_{i,j,0} = 1, v^h_{i,j,N_t-1} = 0 \right\} \ ,$$

and $\phi^h \in \mathcal{K}^h$, where

$$(5.3) \qquad\qquad \mathcal{K}^h = \left\{ \phi^h = \left(\phi^h_x, \phi^h_y, \phi^h_t \right) \in \left(\mathbb{R}^3 \right)^{N_x N_y N_t} : \ (\phi^h)_{i,j,k} \in K \right\} \ ,$$

where K is the pointwise defined convex set which reflects the type of regularization.

In the following, we will treat v^h as the primal and ϕ^h as the respective dual variable. The discrete version of (3.22) is the following convex-concave saddle-point problem

$$(5.4) \qquad\qquad\qquad \min_{v^h \in \mathcal{C}^h} \max_{\phi^h \in \mathcal{K}^h} \left\langle \nabla^h v^h, \phi^h \right\rangle \ ,$$

where the linear operator $\nabla^h : \mathbb{R}^{N_x N_y N_t} \to \left(\mathbb{R}^3\right)^{N_x N_y N_t}$ is a discrete version of the gradient operator in (3.22). We approximate the gradient using simple finite differences and impose Neumann boundary conditions on the boundary of our grid G^h. More specifically, at a certain point (i, j, k) the evaluation of the linear operator ∇^h yields

$$
(5.5) \qquad (\nabla^h v^h)_{i,j,k} = \begin{pmatrix} (\delta_x^h v^h)_{i,j,k} \\ (\delta_y^h v^h)_{i,j,k} \\ (\delta_t^h v^h)_{i,j,k} \end{pmatrix},
$$

where the finite differences are given by

$$
(5.6) \qquad (\delta_x^h v^h)_{i,j,k} = \begin{cases} (v_{i+1,j,k}^h - v_{i,j,k}^h)/h_x & \text{if} \quad i < N_x - 1, \\ 0 & \text{else,} \end{cases}
$$

$$
(5.7) \qquad (\delta_y^h v^h)_{i,j,k} = \begin{cases} (v_{i,j+1,k}^h - v_{i,j,k}^h)/h_y & \text{if} \quad j < N_y - 1, \\ 0 & \text{else,} \end{cases}
$$

$$
(5.8) \qquad (\delta_t^h v^h)_{i,j,k} = \begin{cases} (v_{i,j,k+1}^h - v_{i,j,k}^h)/h_t & \text{if} \quad k < N_t - 1. \\ 0 & \text{else.} \end{cases}
$$

5.2. Primal-dual algorithm. The optimization problem (5.4) poses a classical saddle-point problem [3]. We use the primal-dual algorithm that we recently proposed in [33]. This algorithm is related to the algorithms proposed by Korpelevich [30] and Popov [35]. While these algorithms requires the computation of two leading points, the algorithm of [33] require only one leading point. This leads to a reduction in the number of projections and needs less memory. Our algorithm also shares connections to the simple primal-dual projected subgradient scheme [42], but which requires more restrictive assumptions on the step widths in order to ensure convergence [13].

Our algorithm is as follows: We choose $((v^h)^0, (\phi^h)^0) \in \mathcal{C}^h \times \mathcal{K}^h$ and let $(\bar{v}^h)^0 = (v^h)^0$. We choose two time-steps $\tau, \sigma > 0$. Then, for each $n \geq 0$ we let

$$
(5.9) \qquad \begin{cases} (\phi^h)^{n+1} &= \operatorname{proj}_{\mathcal{K}^h}((\phi^h)^n + \sigma(\nabla^h \bar{v}^n)), \\ (v^h)^{n+1} &= \operatorname{proj}_{\mathcal{C}^h}((v^h)^n - \tau(\operatorname{div}^h(\phi^h)^{n+1})), \\ (\bar{v}^h)^{n+1} &= 2(v^h)^{n+1} - (v^h)^n, \end{cases}
$$

where the operator div^h is chosen to be adjoint to ∇^h. Then, we have the following convergence result.

Theorem 5.1. *Choose τ, σ such that $\tau \sigma L^2 < 1$, with norm $L = \|\nabla^h\|$. Then, as $n \to \infty$, $((v^h)^n, (\phi^h)^n) \to ((v^h)^*, (\phi^h)^*)$, which solves (5.4).*

Proof. The general proof of convergence of the algorithm is presented in [33]. However, in order to ensure convergence of our algorithm we have to compute L. The operator norm is defined as

$$
(5.10) \qquad L = \|\nabla^h\| = \sup \frac{\|\nabla^h v^h\|}{\|v^h\|}.
$$
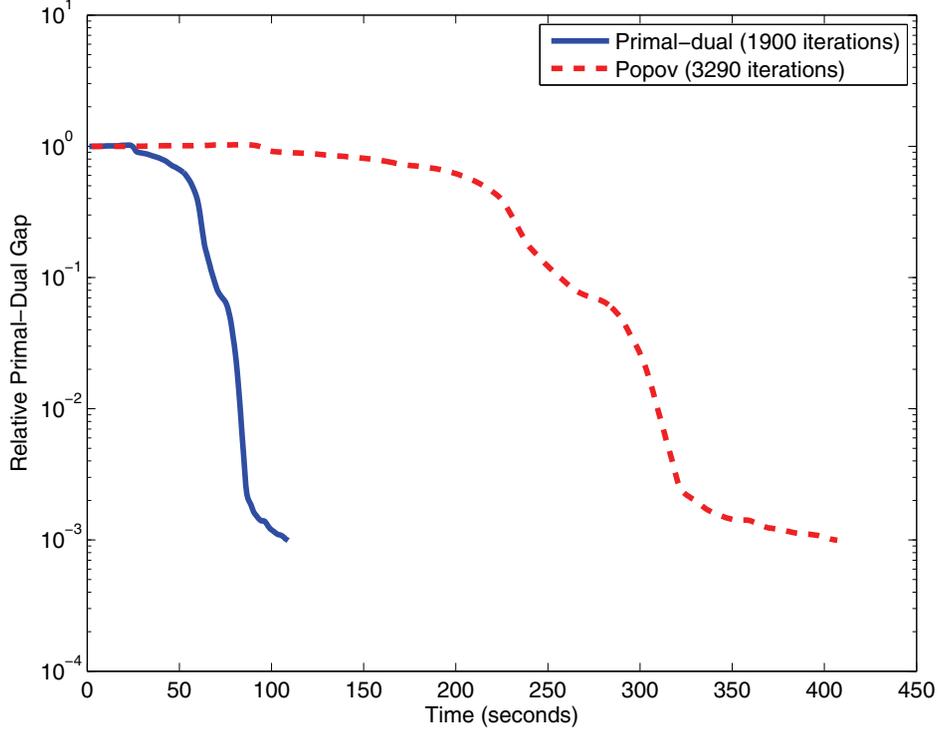
Figure 4. *Comparison of the convergence of the proposed primal-dual algorithm with the algorithm of Popov* [35] *using a total variation prior.*

Hence, in order to compute L, we need to find an estimate on $\|\nabla^h v^h\|$. First, we have the following estimate on $\|\nabla^h v^h\|^2$:

$$(5.11) \quad \|\nabla^h v^h\|^2 = \sum_{i,j,k} \frac{(v_{i+1,j,k}^h - v_{i,j,k}^h)^2}{(h_x)^2} + \frac{(v_{i,j+1,k}^h - v_{i,j,k}^h)^2}{(h_y)^2} + \frac{(v_{i,j,k+1}^h - v_{i,j,k}^h)^2}{(h_t)^2}$$

$$\leq \sum_{i,j,k} \frac{2\{(v_{i+1,j,k}^h)^2 + (v_{i,j,k}^h)^2\}}{(h_x)^2} + \frac{2\{(v_{i,j+1,k}^h)^2 + (v_{i,j,k}^h)^2\}}{(h_y)^2} + \frac{2\{(v_{i,j,k+1}^h)^2 + (v_{i,j,k}^h)^2\}}{(h_t)^2}$$

$$\leq \left(\frac{4}{(h_x)^2} + \frac{4}{(h_y)^2} + \frac{4}{(h_t)^2} \right) \|v^h\|^2 .$$

Then, by taking the square root on both sides of (5.11) and substituting into (5.10) we get

$$(5.12) \qquad\qquad L = \sqrt{\frac{4}{(h_x)^2} + \frac{4}{(h_y)^2} + \frac{4}{(h_t)^2}} \ .$$

Note that this bound becomes sharp as $N_x, N_y, N_t \to \infty$. ∎

Figure 4 shows a comparison of the proposed primal-dual algorithm to the algorithm of Popov [35]. The example was computed using total variation regularization of the input

image of Figure 5. We assume that $h_x = h_y = h_t = 1$, which implies $L = \sqrt{12}$, and hence we set $\sigma = \tau = \sqrt{1/12}$. The plot shows the convergence of the relative primal-dual gap over the number of iterations. Both algorithms were computed on a Tesla C1060 GPU and the iterations were stopped after the relative primal-dual gap dropped below a threshold of 10^{-3}. Note that in our case, computing the relative gap makes sense, since it is invariant to the absolute value of the energy. The primal-dual algorithm took 1900 iterations and 109 seconds, whereas Popov's method needed 3290 iterations and 407 seconds. Hence the proposed primal-dual algorithm is about four times faster than Popov's algorithm. As mentioned above, the proposed algorithm is also more efficient in terms of memory (more than a factor of two), allowing our method to compute larger problems.

5.3. Computing the projections. In this section, we give details about how to compute the projections used in the proposed primal-dual algorithm (5.9).

The projection of the primal variable v^h onto the convex set $\mathcal{C}^h$ is very easy and can be done by simple pointwise clamping operations of v^h to the interval $[0, 1]$.

Next, we detail the projections of the dual variable ϕ^h onto the convex sets $\mathcal{K}^h$ we have introduced in order to realize quadratic, total variation, Huber, and Lipschitz regularization. Recall that we have to consider only pointwise projections of the form $\text{proj}_K(\phi^h_{i,j,k})$. For notational simplicity we will make use of the following convention. We first compute projections $\hat{p} = \text{proj}_K(q)$ of a translated vector $q = (\phi^h_x, \phi^h_t + g^h)_{i,j,k}$, where g^h is the discrete version of the data term. Then the final projection is given by a back-translation, i.e., $\text{proj}_K(\phi^h_{i,j,k}) = (\hat{p}^x, \hat{p}^t - g^h_{i,j,k})$.

For quadratic regularization, one has to ensure that the constraint $q \in K_q$ is fulfilled. This is achieved by projecting q onto the paraboloid illustrated in Figure 3(a). In other words, one has to solve the following constraint optimization problem:

$$(5.13) \qquad \hat{p} = \arg\min_{p \in K_q} \left\{ \frac{|p - q|^2}{2} \right\} .$$

If the constraint $q \in K_q$ is already fulfilled, then the solution is trivially $\hat{p} = q$. If q does not satisfy the constraint, i.e., $q^t < \frac{|q^x|^2}{2}$, then q has to be projected onto the paraboloid $q^t = \frac{|q^x|^2}{2}$. In this case, (5.13) can be expressed as the following unconstrained optimization problem:

$$(5.14) \qquad \hat{p} = \arg\min_p \left\{ \frac{|p - q|^2}{2} - \lambda \left(p^t - \frac{|p^x|^2}{2} \right) \right\} ,$$

where λ is a Lagrange multiplier for the equality constraint $p^t - \frac{|p^x|^2}{2} = 0$. The optimality conditions of (5.14) are given by

$$
\begin{aligned}
p^x - q^x + \lambda p^x &= 0, \\
p^t - q^t - \lambda &= 0, \\
p^t - \frac{|p^x|^2}{2} &= 0.
\end{aligned}
$$

$$(5.15)$$

After eliminating p^t and p^x, we arrive at the following cubic equation for the solution of λ:

$$(5.16) \qquad \lambda^3 + \lambda^2(q^t + 2) + \lambda(2q^t + 1) + q^t - \frac{|q^x|^2}{2} = 0.$$

Instead of using a direct cubic solver for (5.16) we utilize Newton's method. We choose a starting point $\lambda^0 = \max\{0, -(2q^t + 1)/3)\} + 1$ and let for each $n \geq 0$

$$(5.17) \qquad \lambda^{n+1} = \lambda^n - \frac{(\lambda^n)^3 + (\lambda^n)^2(q^t + 2) + \lambda^n(2q^t + 1) + q^t - \frac{|q^x|^2}{2}}{3(\lambda^n)^2 + 2\lambda^n(q^t + 2) + 2q^t + 1} .$$

We stop the iterations of Newton's method when the difference between two subsequent iterations is below a certain threshold. We found this scheme to have a quite fast convergence (10–20 iterations). Then, after computing the solution of (5.16), the solution of the projection is given by

$$(5.18) \qquad \hat{p} = \left(\frac{q^x}{1 + \lambda}, q^t + \lambda \right) .$$

In case of total variation regularization, we need to compute the orthogonal projection of q onto the cylinder K_t shown in Figure 3(b). This is very easy, since K_t consists of two independent constraints in p^x and p^t. The solution of this projection is given by

$$(5.19) \qquad \hat{p} = \left(\frac{q^x}{\max\{1, |q^x|\}}, \max\{0, q^t\} \right) .$$

Huber regularization is achieved by projecting onto K_h, which essentially is a combination of quadratic and total variation regularization. If $q \in K_h$, we trivially obtain $p = q$. Otherwise, we have to compute the projection onto K_h. According to Figure 3(c), we can distinguish three different cases. We have the paraboloidal part, the cylindrical part, and the interface where the paraboloid intersects with the cylinder:

$$(5.20) \qquad \hat{p} = \begin{cases} \left(\frac{q^x}{\max\{1, |q^x|\}}, q^t \right) & \text{if} \quad \frac{\alpha}{2} \leq q^t, \\ \left(\frac{q^x}{\max\{1, |q^x|\}}, \frac{\alpha}{2} \right) & \text{if} \quad \frac{\alpha}{2} - \frac{1}{\alpha}(|q^x| - 1) \leq q^t < \frac{\alpha}{2}, \\ \left(\frac{q^x}{1 + \lambda}, q^t + \lambda \right) & \text{if} \quad q^t < \frac{\alpha}{2} - \frac{1}{\alpha}(|q^x| - 1), \end{cases}$$

where λ is the solution of the cubic equation (5.16).

The Lipschitz constraint is realized by projecting q onto the cone defined by K_l (see also Figure 3(d)). The projection is straightforward and is given by

$$(5.21) \qquad \hat{p} = \left(\mu \frac{q^x}{|q^x|}, \beta\mu \right) ,$$

where μ is given by

$$(5.22) \qquad \mu = \frac{\max\{0, |q^x| + \beta q^t\}}{1 + \beta^2} .$$

6. Experimental results. In this section we first show experimental results of our convex framework on two applications: outlier filtering of illustrative range images and disparity estimation of stereo images. In both cases, we will show the application of quadratic, total variation, Huber, and Lipschitz regularization.

(a) Range image
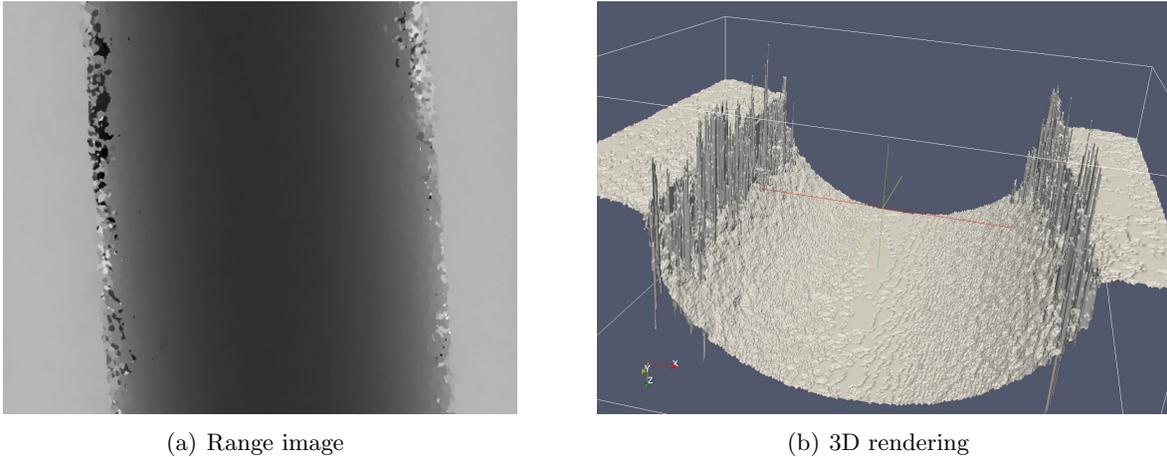
(b) 3D rendering

Figure 5. *Range image and* 3*D rendering of an industrial part obtained by an optical measurement system. Note the erroneous measurements (outliers) in the region of the steepest slope.*

6.1. Range image filtering. In the first experiment, we apply our convex framework to the task of outlier removal of range images. The range image has been obtained by an optical measurement system used for industrial applications. The goal of the filtering is that the outliers should be removed without modifying the right measurements. Figure 5 shows the range image of an industrial part and the corresponding 3D rendering. The size of the range image was $N_x \times N_y = 443 \times 370$, and the function $u(x)$ was discretized using $N_t = 256$ levels. Furthermore, we set $h_x = h_y = h_t = 1$ and hence $L = \sqrt{12}$. The data term was computed using truncated quadratic differences, i.e., $g^h_{i,j,k} = \mu \min\{(I_{i,j}-k)^2, \nu\}$, where $I_{i,j} \in [0, 255]$ was the input image, $\mu = 0.05$ was a scaling parameter, and $\nu = 100$ was the truncation threshold. For Huber regularization we set $\alpha = 5$, and for Lipschitz regularization the maximal slope was set to $\beta = 2$

Figure 7 shows the result of outlier filtering of the range image depicted in Figure 5. Clearly, quadratic regularization leads to an oversmoothing of the image, and, hence, the small scale structures, which are quite important in this application, also are destroyed. Total variation regularization shows the ability to preserve sharp discontinuities but leads to the well-known staircasing effect. As one can see, Huber regularization does not suffer from the staircasing effect, but, similar to quadratic regularization, the fine scale structures of the surface are destroyed. Finally, Lipschitz regularization shows the best results for this application, since it does not destroy the fine scale structures of the surface while outliers are effectively removed.

6.2. Disparity estimation. In our second experiment we apply our convex framework to disparity estimation of a stereo image pair taken from [5]. Figure 6 shows the rectified color input images $I_l = (I_l^r, I_l^g, I_l^b), I_r = (I_r^r, I_r^g, I_r^b)$ and the ground truth disparity. The size of the input images was $N_x \times N_y = 443 \times 370$ and the function $u(x)$ was discretized using $N_t = 141$ levels. Again we set $h_x = h_y = h_t = 1$ and hence $L = \sqrt{12}$. The data term was the pixelwise absolute differences of the RGB values of the input images, that is, $g^h_{i,j,k} =$

| (a) Left input image | (b) Right input image | (c) True disparity |
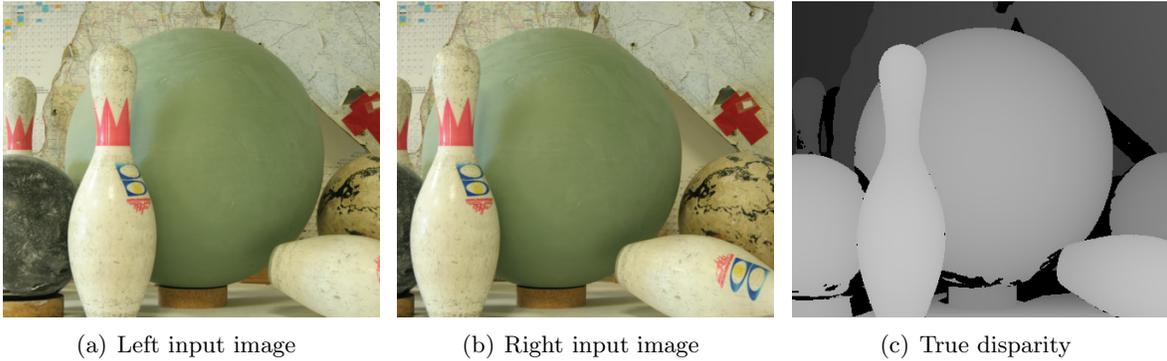
Figure 6. *Rectified stereo image pair and the ground truth disparity. Light gray pixels indicate structures near to the camera, and black pixels correspond to unknown disparity values.*

$\mu \sum_{i \in \{r,g,b\}} |(I_l^i)_{i,j} - (I_r^i)_{i,j+k}|$, and μ is a scaling parameter. For quadratic regularization we used $\mu = 100$, for total variation and Huber regularization we used $\mu = 50$. Note that in the case of Lipschitz regularization, the result was independent of μ. The Huber parameter was set to $\alpha = 5$, and the maximum slope in Lipschitz regularization was set to $\beta = 2$.

Figure 8 shows the results of using different convex regularity terms. As expected, quadratic regularization leads to an oversmoothing of depth discontinuities. Total variation regularization has the ability to preserve sharp depth discontinuities but exhibits some staircasing in flat regions. Huber regularization gives the best result here, since it leads to piecewise smooth disparities. Finally, one can see that Lipschitz regularization is (obviously) not a good choice in this case. While the smooth surface parts still exhibit severe noise, large depth discontinuities are oversmoothed due to the Lipschitz constraint.

7. Conclusion. In this work we have presented a theoretical framework for computing global solutions of variational models with convex regularization but which can have quite general data terms. Our approach builds upon the same idea as the MRF-based discrete method of Ishikawa [24], namely, to increase the dimensionality of the problem. Unlike the approach of Ishikawa, which requires long-range interactions in the graph structure in order to implement general convex priors (e.g., quadratic), our approach can handle quite arbitrary convex regularizers by just projecting onto different convex sets.

We also have given details on several different instances of convex regularity terms including quadratic, total variation, Huber, and Lipschitz regularization. For solution of the resulting convex optimization problem, we proposed a novel primal-dual algorithm which is memory efficient and can be effectively accelerated on state-of-the-art graphics devices.

In experimental results we have demonstrated the applicability of the proposed method using different convex regularizers to typical computer vision problems such as image filtering and stereo.

While the framework developed in this paper allows one to compute globally optimal solutions of functionals with convex regularizers, the same relaxation concept can be extended to cases of nonconvex regularizers; see [32, 33].

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
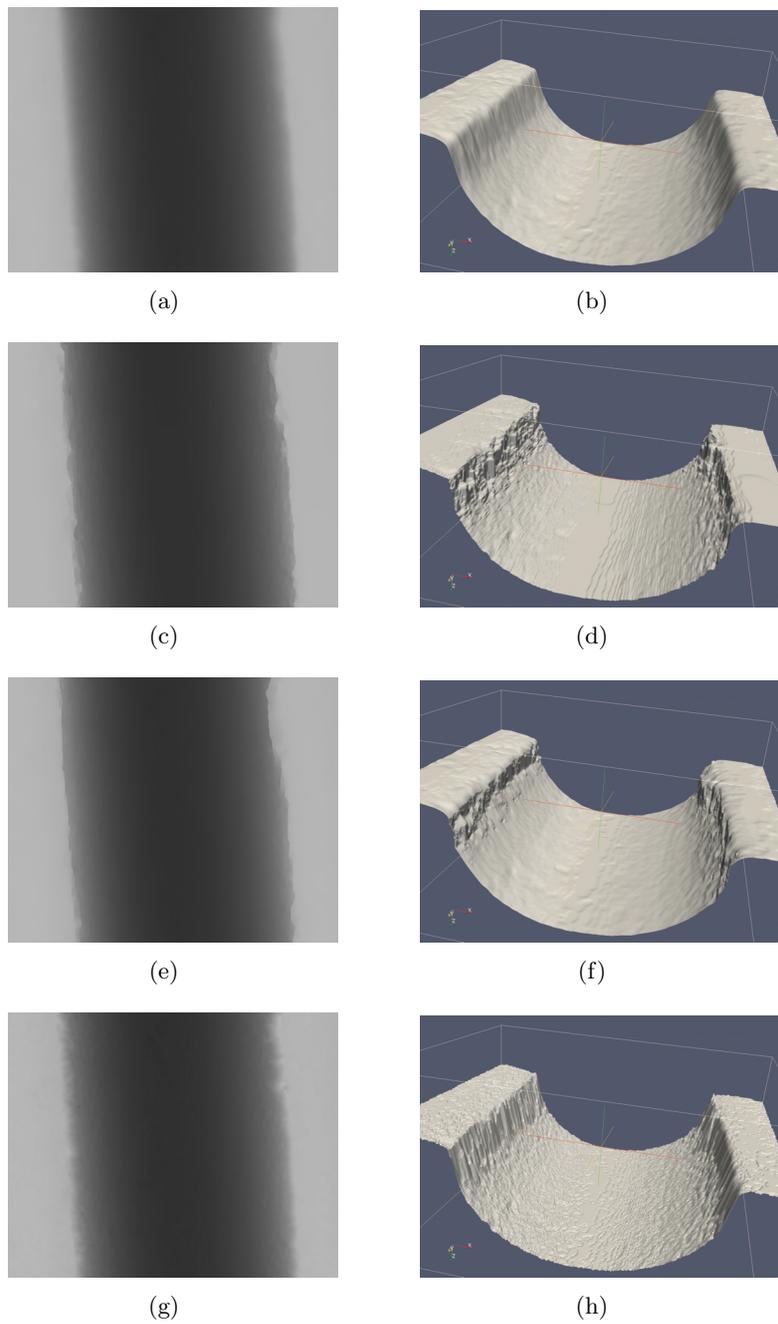
Figure 7. *Different convex regularity terms applied to outlier filtering of a range image. First row: quadratic regularization; second row: total variation regularization; third row: Huber regularization, and fourth row: Lipschitz regularization.*

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
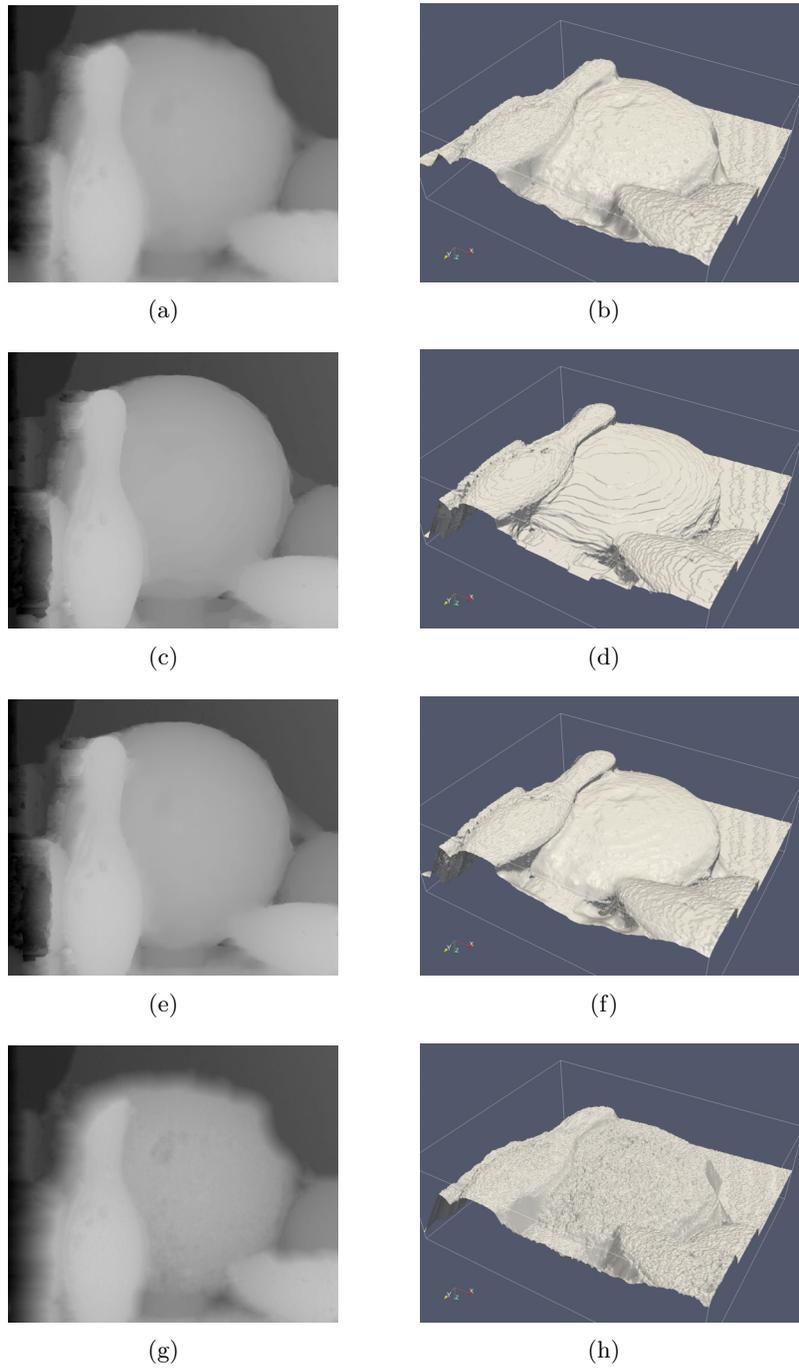
Figure 8. *Application of different convex regularity terms applied to disparity estimation. First row: quadratic regularization; second row: total variation regularization; third row: Huber regularization; and fourth row: Lipschitz regularization.*

REFERENCES

[1] G. Alberti, G. Bouchitté, and D. Dal Maso, *The calibration method for the Mumford-Shah functional and free-discontinuity problems*, Calc. Var. Partial Differential Equations, 16 (2003), pp. 299–333.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Math. Monogr., The Clarendon Press, Oxford University Press, New York, 2000.

[3] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*, Stanford Mathematical Studies in the Social Sciences, Vol. II., Stanford University Press, Stanford, CA, 1958.

[4] H. Attouch, G. Buttazzo, and G. Michaille, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*, MPS/SIAM Ser. Optim. 6, SIAM, Philadelphia, 2006.

[5] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. Black, and R. Szeliski, *A database and evaluation methodology for optical flow*, in Proceedingsof the IEEE International Conference on Computer Vision (ICCV), 2007, pp. 1–8.

[6] D. Bertismas and J. Tsitsikilis, *Introduction to Linear Optimization*, Athena Scientific, Nashua, NH, 1997.

[7] Y. Boykov, O. Veksler, and R. Zabih, *Fast approximate energy minimization via graph cuts*, IEEE Trans. Pattern Anal. Mach. Intell., 23 (2001), pp. 1222–1239.

[8] A. Chambolle, *Convex representation for lower semicontinuous envelopes of functionals in L^1*, J. Convex. Anal., 1 (2001), pp. 149–170.

[9] A. Chambolle, *An algorithm for mean curvature motion*, Interfaces Free Bound., 6 (2004), pp. 195–218.

[10] A. Chambolle, D. Cremers, and T. Pock, *A Convex Approach for Computing Minimal Partitions*, Technical report, Ecole Polytechnique, Paris, France, 2008.

[11] T. F. Chan, S. Esedoğlu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM J. Appl. Math., 66 (2006), pp. 1632–1648.

[12] G. Dal Maso, *Integral representation on $BV(\Omega)$ of Γ-limits of variational integrals*, Manuscripta Math., 30 (1979/80), pp. 387–416.

[13] E. Esser, X. Zhang, and T. Chan, *A General Framework for a Class of First Order Primal-dual Algorithms for TV Minimization*, CAM Report 09-67, Center for Applied Mathematics, UCLA, Los Angeles, CA, 2009.

[14] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, FL, 1992.

[15] H. Federer, *Geometric Measure Theory*, Springer-Verlag, New York, 1969.

[16] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell., 6 (1984), pp. 721–741.

[17] M. Giaquinta, G. Modica, and J. Souček, *Cartesian Currents in the Calculus of Variations.* I. Cartesian Currents, Ergeb. Math. Grenzgeb. 37, 3. Folge, Springer-Verlag, Berlin, 1998.

[18] M. Giaquinta, G. Modica, and J. Souček, *Cartesian Currents in the Calculus of Variations.* II. Variational Integrals, Ergeb. Math. Grenzgeb. 38, 3. Folge, Springer-Verlag, Berlin, 1998.

[19] L. M. Goldschlager, R. A. Shaw, and J. Staples, *The maximum flow problem is log space complete for p*, Theoret. Comput. Sci., 21 (1982), pp. 105–111.

[20] D. M. Greig, B. T. Porteous, and A. H. Seheult, *Exact maximum a posteriori estimation for binary images*, J. Roy. Statist. Soc. Ser. B, 51 (1989), pp. 271–279.

[21] P. L. Hammer, P. Hansen, and B. Simeone, *Roof duality, complementation and persistence in quadratic 0-1 optimization*, Math. Programming, 28 (1984), pp. 121–155.

[22] B. Horn and B Schunck, *Determining optical flow*, Artificial Intelligence, 17 (1981), pp. 185–203.

[23] P. J. Huber, *Robust Statistics*, John Wiley & Sons, New York, 1981.

[24] H. Ishikawa, *Exact optimization for Markov random fields with convex priors*, IEEE Trans. Pattern Anal. Mach. Intell., 25 (2003), pp. 1333–1336.

[25] H. Ishikawa and D. Geiger, *Segmentation by grouping junctions*, in Proceedings of the IEEE Computer

Society Conference on Computer Vision and Pattern Recognition (CVPR), 1998, pp. 125–131.

[26] R. KINDERMANN AND J. L. SNELL, *Markov Random Fields and Their Applications*, American Mathematical Society, Providence, RI, 1980.

[27] M. KLODT, T. SCHOENEMANN, K. KOLEV, M. SCHIKORA, AND D. CREMERS, *An experimental comparison of discrete and continuous shape optimization methods*, in Proceedings of the 10th European Conference on Computer Vision, 2008, 332–345.

[28] V. KOLMOGOROV AND A. SHIOURA, *New algorithms for convex cost tension problem with application to computer vision*, Discrete Optim., 6 (2009), pp. 378–393.

[29] V. KOLMOGOROV AND R. ZABIH, *What energy functions can be minimized via graph cuts?*, IEEE Trans. Pattern Anal. Mach. Intell., 26 (2004), pp. 147–159.

[30] G. M. KORPELEVICH, *Extrapolational gradient methods and their connection with modified Lagrangians*, Ehkon. Mat. Metody, 19 (1983), pp. 694–703 (in Russian).

[31] D. MUMFORD AND J. SHAH, *Optimal approximation by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–685.

[32] T. POCK, A. CHAMBOLLE, D. CREMERS, AND H. BISCHOF, *A convex relaxation approach for computing minimal partitions*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 810–817.

[33] T. POCK, D. CREMERS, H. BISCHOF, AND A. CHAMBOLLE, *An algorithm for minimizing the Mumford-Shah functional*, in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2009.

[34] T. POCK, T. SCHOENEMANN, G. GRABER, H. BISCHOF, AND D. CREMERS, *A convex formulation of continuous multi-label problems*, in European Conference on Computer Vision (ECCV), 2008.

[35] L. POPOV, *A modification of the Arrow–Hurwicz method for search for saddle points*, Math. Notes, 28 (1980), pp. 845–848.

[36] R. T. ROCKAFELLAR, *Convex analysis*, Princeton Landmarks Math., Princeton Paperbacks, Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original.

[37] C. ROTHER, V. KOLMOGOROV, V. LEMPITSKY, AND M. SZUMMER, *Optimizing binary MRFs via extended roof duality*, in Proceedings of the IEEE Computer Society, Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8.

[38] S. ROY AND I. J. COX, *A maximum-flow formulation of the n-camera stereo correspondence problem*, in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1998, pp. 492–502.

[39] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[40] D. SCHLESINGER AND B. FLACH, *Transforming an Arbitrary Minsum Problem into a Binary One*, Technical report TUD-FI06-01, Dresden University of Technology, Dresden, Germany, 2006.

[41] M. J. WAINWRIGHT, T. S. JAAKKOLA, AND A. S. WILLSKY, *Map estimation via agreement on trees: Message-passing and linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 3697–3717.

[42] M. ZHU AND T. CHAN, *An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration*, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.

[43] W. P. ZIEMER, *Weakly Differentiable Functions. Sobolev Spaces and Functions of Bounded Variation*, Springer-Verlag, New York, 1989.