

TGV-Fusion

Thomas Pock¹, Lukas Zebedin², and Horst Bischof¹

¹ Inst. for Computer Graphics and Vision,
Graz University of Technology

{pock,bischof}@icg.tugraz.at

² Microsoft Photogrammetry
lucas.zebedin@microsoft.com

Abstract. Location awareness on the Internet and 3D models of our habitat (as produced by Microsoft (Bing) or Google (Google Earth)) are a major driving force for creating 3D models from image data. A key factor for these models are highly accurate and fully automated stereo matching pipelines producing highly accurate 3D point clouds that are possible due to the fact that we can produce images with high redundancy (i.e., a single point is projected in many images). Especially this high redundancy makes fully automatic processing pipelines possible. Highly overlapping images yield also highly redundant range images. This paper proposes a novel method to fuse these range images. The proposed method is based on the recently introduced total generalized variation method (*TGV*). The second order variant of this functional is ideally suited for piece-wise affine surfaces and therefore an ideal case for buildings which can be well approximated by piece-wise planar surfaces. In this paper we first present the functional consisting of a robust data term based on the Huber- L^1 norm and the *TGV* regularization term. We derive a numerical algorithm based on a primal dual formulation that can be efficiently implemented on the *GPU*. We present experimental results on synthetic data as well as on a city scale data set, where we compare the method to other methods.

1 Introduction

Recovery of realistic models of urban environments from aerial imagery has been an important research topic in computer vision for more than 20 years. Recently, large scale efforts from large enterprises like Microsoft are underway, aiming at building a virtual equivalent of our planet including realistic models of thousands of cities worldwide. The enormous scale of such a project, encompassing tens of thousands of the largest cities worldwide, makes any manual attempt to solve the arising problems prohibitively expensive. The only way to tackle such an undertaking is by a fully automatic processing pipeline. Such a workflow would start with the raw images collected by a digital aerial camera mounted on an aircraft and would automatically produce a virtual, digital representation of the covered area. For a recent overview on producing 3D models and other information from aerial images see [1]. The other papers in this special issue cover other

aspects of producing 3D models of our virtual habitat. Using digital cameras, we can produce highly overlapping and thus redundant images of no additional cost. This high redundancy enables fully automatic methods for high quality 3D reconstruction (e.g. [2]). An important step towards the 3D reconstruction is computing the range images. Due to the high redundancy in the input data, also the range images are highly redundant. The robust fusion of several range images to a single high quality range image - also called digital surface model (*DSM*) - is the main topic of this paper.

Basically, the fusion can be done in full 3D (e.g. [3]) or in $2\frac{1}{2}D$ (e.g. [4]). Our target application is 3D modeling of buildings from very large aerial images where it is simply not possible (and also not necessary) to have a full 3D representation of the buildings. Hence, we will consider the $2\frac{1}{2}D$ case in this paper.

The range image fusion approach we are going to present in this paper is based on variational approaches for image regularization [5,6,7]. The basic idea of the variational approach is that the solution of the model is given by the minimizer of an energy-functional. The energy functional usually is composed of two basic terms. A regularization term, which reflects the a-priori assumption about the smoothness properties of the solution and a data term which forces the solution to be similar to the input data. Clearly, different choices of the regularization and data terms will lead to different solutions. Variational models can be further divided into two fundamental classes: Convex and non-convex problems. The obvious advantage of convex problems over non-convex problems is that one has the guarantee to find a global optimum. This means that the quality of the solution of a convex functional solely depends on the accuracy of the variational model. On the other hand, for non-convex problems, the quality of the solution is subject to both the model and the optimization algorithm, since in general only a local minimizer can be computed. We will therefore restrict ourselves to convex models in this paper.

State-of-the-art variational image regularization techniques [6,7] are typically based on a first-order smoothness assumption. While this assumption might be reasonable for general images, it is less useful for regularizing range images. The reason is that first-order smoothness leads to a preference of piecewise constant solutions (*a.k.a.* staircase effect), which is not appropriate to regularize range images containing slanted surfaces such as roofs. Total generalized variation (*TGV*) as recently introduced in [8] has some desired properties for regularizing range images. The key property of TGV^k regularization is, that it favors piecewise polynomial signals of order $k - 1$ (e.g. TGV^2 favors piecewise affine functions). Therefore, TGV^2 regularization is perfectly suited to regularize range images. The main contributions of this paper are:

- A new convex variational model for robust fusion of range images, which combines TGV^2 regularization with a robust Huber- L^1 data term.
- A novel efficient numerical algorithm based on a new first-order primal-dual algorithm which can be efficiently parallelized on graphics processing units.

- The demonstration of the favorable properties of the proposed method on a large scale city dataset.

The paper is organized as follows. In section 2 we review convex variational methods for image regularization and highlight possible advantages and disadvantages. As most existing methods are formulated to deal with only one observation, we will also consider (straight-forward) generalizations of these models to multiple observations. In section 3 we present the proposed TGV^2 model for range image fusion. For minimization we present an efficient first-order primal-dual algorithm, which will be detailed in section 4. In section 5 we show experimental results using synthetic and real images. In the last section we draw conclusions and show directions for future research.

2 Related Work

In this section we review some basic variational image denoising models. While in their original formulations, the models include a data term which takes only one observation into account, we will already consider the case with multiple observations.

2.1 Quadratic Model

The quadratic model (or Tikhonov model) [5] is one of the earliest - dating back to 1943 - and simplest regularization method used for ill-posed problems. It is defined as the quadratic variational problem

$$\min_u \left\{ \alpha \int_{\Omega} |\nabla u|^2 dx + \sum_{l=1}^K \int_{\Omega} (u - f_l)^2 dx \right\}, \quad (1)$$

where $\Omega \subset \mathbb{R}^2$ is the image domain, K is the number of observed range images, $f_l : \Omega \rightarrow \mathbb{R}$ denotes a single observation, and $u : \Omega \rightarrow \mathbb{R}$ is the sought solution. The free parameter $\alpha \geq 0$ is used to control the amount of smoothing in u . The left term is the regularization term which reflects the smoothness assumption. The right term measures the distance of the solution to the observed data. Note that it is very common for range images that the number of observations may vary over the image domain (e.g. invalid ranges due to occlusions). Since image locations with a larger number of observations imply a higher confidence in the input data, we do not normalize the data term with respect to the number of observations in our models.

In view of our range image fusion problem, the quadratic model tries to find a smooth solution u which minimizes the squared distance to the single observations f_l . Being quadratic in u , the Tikhonov model poses a simple optimization problem, but it does not perform very well for our purpose. The main reason is that the quadratic regularization term leads to an oversmoothing of edges and the quadratic data term is not robust against strong outliers in the observed data.

2.2 ROF Model

L^1 estimation procedures have shown to be an effective tool for many different problems. The first who introduced L^1 estimation methods for image restoration were Rudin, Osher and Fatemi (*ROF*) in their seminal paper on edge preserving image denoising [9]. In its unconstrained form, the *ROF* model is defined as the variational model

$$\min_u \left\{ \alpha \int_{\Omega} |\nabla u| dx + \frac{1}{2} \sum_{l=1}^K \int_{\Omega} (u - f_l)^2 dx \right\} . \tag{2}$$

The first term is the so called total variation semi-norm of u . We point out that in this definition, the *TV* norm is only valid for sufficiently smooth function u (e.g. $u \in C^1(\Omega)$). Fortunately, there exists also a duality based formulation of the total variation which enables a valid definition for any integrable function $u \in L^1(\Omega)$.

$$TV^\alpha(u) = \sup \left\{ - \int_{\Omega} u \operatorname{div} \varphi dx : \varphi \in \mathcal{C}_c^1(\Omega, \mathbb{R}^2), \|\varphi\|_\infty \leq \alpha \right\} , \tag{3}$$

where $\mathcal{C}_c^1(\Omega, \mathbb{R}^2)$ is the space of continuously differentiable functions with compact support in Ω . Note that this formulation also allows to overcome the non-differentiability of the *TV* norm. Besides its convexity, the main advantage of the *ROF* model of the quadratic model is, that it allows for sharp discontinuities in the solution. This is an important feature, since we assume to have sharp depth discontinuities in our range images. However, due to the quadratic data term the *ROF* model is still very sensitive to strong outliers (e.g. caused by occlusion) in the observed range images.

2.3 TV- L^1 Model

The *TV- L^1* model [10,7,11] is obtained from the *ROF* model by replacing the L^2 norm in the data term with the L^1 norm.

$$\min_u \left\{ \alpha \int_{\Omega} |\nabla u| dx + \sum_{l=1}^K \int_{\Omega} |u - f_l| dx \right\} . \tag{4}$$

Due to the L^1 norm in the data term, it turns out that the *TV- L^1* model is much more effective than the *ROF* model at removing strong outliers [7].

2.4 Huber Model

Although the *TV- L^1* model has the nice properties to preserve sharp discontinuities in the solution and to be robust against outliers, it has two problems: First, it suffers from the so-called staircasing effect - an effect which describes the formation of artificial discontinuities in the solution [12]. Second, the L^1 norm is

not the optimal choice for the expected noise. In real range images the observed noise is the sum of Gaussian noise and outliers.

It turns out that the Huber norm helps in reducing the staircasing effect and also better reflects the noise model of real range images. The Huber norm [13] is defined as

$$|x|_\gamma = \begin{cases} \frac{|x|^2}{2\gamma} & \text{if } |x| \leq \gamma \\ |x| - \frac{\gamma}{2} & \text{if } |x| > \gamma \end{cases}, \tag{5}$$

where γ is a free parameter that defines the threshold between quadratic and L^1 penalization. Using (5) for both the regularization term and the data term in (4), we obtain the Huber model

$$\min_u \left\{ \alpha \int_\Omega |\nabla u|_\varepsilon dx + \sum_{l=1}^K \int_\Omega |u - f_l|_\delta dx \right\}, \tag{6}$$

where ε , and δ are the parameters of the respective Huber norms. Note that by appropriate choices of ε and δ , the Huber model unifies the quadratic model, the *ROF* model and *TV-L¹* model into a single model. However, while the Huber norm reduces the staircasing effect to some extent, it still favors flat solutions. In the next section we will see that the incorporation of higher-order derivatives plays a key role in obtaining a model which does not suffer from this problem.

3 Total Generalized Variation

In [8] Bredies, Kunisch and Pock proposed the mathematical model of total generalized variation (*TGV*). The main property of *TGV* regularization is that it allows to reconstruct piecewise polynomial functions of arbitrary order (piecewise constant, piecewise affine, piecewise quadratic, ...) . As the *TV* regularizer, the *TGV* regularizer has the nice property of being convex. This allows to compute a globally optimal solution.

The *TGV* semi-norm of order $k \geq 1$ with regularization parameters $\alpha = (\alpha_0, \dots, \alpha_{k-1}) > 0$ is defined as

$$TGV_k^\alpha(u) = \sup \left\{ \int_\Omega u \operatorname{div}^k \varphi dx : \varphi \in \mathcal{C}_c^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^2)), \|\operatorname{div}^l \varphi\|_\infty \leq \alpha_l, l = 0, \dots, k - 1 \right\}, \tag{7}$$

where $\mathcal{C}_c^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^2))$ denotes the space of continuously differentiable symmetric k -tensors with compact support in Ω . It is obvious that for $k = 1$, (7) corresponds to the dual definition of the total variation semi-norm (3), i.e. *TGV* is indeed a generalization of the *TV* regularizer. The definition of the total generalized variation restricts the function v to be in a complicated convex set. This leads to computationally complex minimization algorithms [8]. Using Legendre

- Fenchel duality, we can transform the dual problem (7) to a primal formulation [8]:

$$TGV_k^\alpha(u) = \inf_{\substack{u_l \in \mathcal{C}^{k-l}(\Omega, \text{Sym}^l(\mathbb{R}^2)) \\ l=1, \dots, k-1, u_0=u, u_k=0}} \sum_{l=1}^k \alpha_{k-l} \int_{\Omega} |\mathcal{E}(u_{l-1}) - u_l| dx, \quad (8)$$

where $\mathcal{E}(u)$ denotes the symmetrized gradient operator

$$\mathcal{E}(u) = \frac{\nabla u + \nabla u^T}{2}.$$

Note that this representation has converted functional (7), which depends on higher order derivatives to a functional of recursive expressions depending only on first order derivatives. Using this representation one can intuitively assess how the total generalized variation is working. Before measuring the L^1 norm of the expression $\mathcal{E}(u_{l-1})$ a vector field u_l is subtracted which itself should have low variation. That is, if low variations are present in some areas of u_{l-1} (e.g. smooth gradients), the vector field u_l will cover these areas, and therefore will decrease the L^1 norm of $\mathcal{E}(u_{l-1})$ in these areas. Hence, $TGV_k^\alpha(u)$ automatically balances the first and higher order derivatives instead of using any fixed combination. For the application to range images from our habitat, it turns out that TGV regularization of second order ($k = 2$) is sufficient since buildings can be well approximated by piecewise planar surfaces.

3.1 The Proposed Model

We are now ready to state the proposed variational model for range image fusion.

$$\min_{u,v} \left\{ \alpha_1 \int_{\Omega} |\nabla u - v| dx + \alpha_0 \int_{\Omega} |\mathcal{E}(v)| dx + \sum_{l=1}^K \int_{\Omega} |u - f_l|_{\delta} dx \right\}. \quad (9)$$

This model combines TGV regularization of second order (8) with the Huber- L^1 norm in the data term. Note that it significantly differs from the TGV^2 denoising model proposed in [8], where a quadratic data term is used. Therefore we can not make use of the minimization algorithm proposed in [8]. In addition, the original algorithm used in [8] is quite slow, since it is based on the dual formulation (7) which requires an additional inner loop to project onto a very complex set. The algorithm we are going to describe in the next section is based on the primal formulation (8) which does not require such an inner loop.

4 Numerical Algorithm

In order to implement an algorithm to minimize (9) on a digital computer, we have to introduce the discrete version of (9).

4.1 Discretization

We consider a regular Cartesian grid of size $M \times N$:

$$\Omega^h = \{(x_i, y_j) = (ih, jh) : 1 \leq i \leq N, 1 \leq j \leq M\} ,$$

where h denotes the size of the spacing and (i, j) denote the indices of the discrete locations $(ih, jh) \in \Omega^h$. From now on, we will denote the quantities of the discrete setting by the superscript h . Let $U^h = \mathbb{R}^{MN}$ and $V^h = \mathbb{R}^{2MN}$ be finite dimensional vector spaces equipped with scalar products

$$\hat{u}^h, \bar{u}^h \in U^h : \langle \hat{u}^h, \bar{u}^h \rangle = \sum_{i,j} \hat{u}_{i,j}^h \bar{u}_{i,j}^h ,$$

$$\hat{v}^h, \bar{v}^h \in V^h : \langle \hat{v}^h, \bar{v}^h \rangle = \sum_{i,j} (\hat{v}_1^h)_{i,j} (\bar{v}_1^h)_{i,j} + (\hat{v}_2^h)_{i,j} (\bar{v}_2^h)_{i,j} .$$

Furthermore, let $u^h \in U^h$ and $v^h = (v_1^h, v_2^h) \in V^h$ be discrete versions of the unknown functions u and v in (9) and let $f^h = (f_1^h, \dots, f_K^h)^T \in \mathbb{R}^{KMN}$, $l = 1 \dots K$ be the discrete observations. For discretization of the gradient operator ∇^h we use standard finite differences with Neumann boundary conditions

$$(\nabla^h u^h)_{i,j} = \begin{pmatrix} (\delta_x^h u^h)_{i,j} \\ (\delta_y^h u^h)_{i,j} \end{pmatrix} ,$$

where

$$(\delta_x^h u^h)_{i,j} = \begin{cases} \frac{u_{i+1,j}^h - u_{i,j}^h}{h} & \text{if } 0 < i < M \\ 0 & \text{if } i = M \end{cases} , \quad (\delta_y^h u^h)_{i,j} = \begin{cases} \frac{u_{i,j+1}^h - u_{i,j}^h}{h} & \text{if } 0 < j < N \\ 0 & \text{if } j = N \end{cases} ,$$

are standard first order finite differences. Similarly, the discrete variant of the symmetrized gradient operator is defined as

$$(\mathcal{E}^h v^h)_{i,j} = \begin{pmatrix} (\delta_x^h v_1^h)_{i,j} & \frac{(\delta_y^h v_1^h)_{i,j} + (\delta_x^h v_2^h)_{i,j}}{2} \\ \frac{(\delta_y^h v_1^h)_{i,j} + (\delta_x^h v_2^h)_{i,j}}{2} & (\delta_y^h v_2^h)_{i,j} \end{pmatrix} . \tag{10}$$

The discrete variant of (9) can now be written as

$$\min_{u^h, v^h} \left\{ \alpha_1 \|\nabla^h u^h - v^h\|_1 + \alpha_0 \|\mathcal{E}^h v^h\|_1 + \sum_{l=1}^K \sum_{i,j} |u_{i,j}^h - (f_l^h)_{i,j}|_\delta \right\} . \tag{11}$$

This minimization problem poses a large scale non-smooth optimization problem. Hence one can not expect that any black box solver will find the solution in a reasonable time. Instead we will make use of first-order primal-dual algorithms which have been shown to be a good choice for large-scale convex optimization problems [14] in imaging.

4.2 Primal-Dual Formulation

Let us rewrite minimization problem (11) as a convex-concave saddlepoint problem [14]. By applying the Legendre-Fenchel transform to (11) we obtain

$$\begin{aligned} \min_{u^h, v^h} \max_{p^h, q^h, r^h} & \left\{ \langle \nabla^h u^h - v^h, p^h \rangle + \langle \mathcal{E}^h v^h, q^h \rangle + \sum_{l=1}^K \langle u^h - f_l^h, r_l^h \rangle - \frac{\delta}{2} \|r_l^h\|^2 \right\} \\ \text{subject to} & \quad \|p^h\|_\infty \leq \alpha_1, \quad \|q^h\|_\infty \leq \alpha_0, \quad \|r_l^h\|_\infty \leq 1, \end{aligned} \quad (12)$$

where p^h , q^h and r_l^h are the dual variables. The obvious advantage of the primal-dual formulation (12) over the primal formulation (11) is, that the non-differentiable L^1 terms have been transformed to linear terms with simple ball constraints on the dual variables. Note that for clarity of the presentation, we did not exploit the obvious symmetry in (10). In any practical implementation, one can easily make use of the symmetry to reduce the memory footprint of the algorithm.

4.3 Numerical Algorithm

Our algorithm is based on the first-order primal-dual algorithm introduced in [14], where also convergence of the algorithm is proven. Before we start to describe the algorithm, let us define some useful quantities. The convex sets P^h , Q^h and R^h are defined as

$$\begin{aligned} P^h &= \{p^h \in \mathbb{R}^{2MN} : \|p^h\|_\infty \leq \alpha_1\}, \\ Q^h &= \{q^h \in \mathbb{R}^{4MN} : \|q^h\|_\infty \leq \alpha_0\}, \\ R^h &= \{r^h \in \mathbb{R}^{MN} : \|r^h\|_\infty \leq 1\}. \end{aligned} \quad (13)$$

Furthermore, let the convex function $F(r_l^h)$ be defined as

$$F(r_l^h) = \mathcal{I}_{R^h}(r_l^h) + \frac{\delta}{2} \|r_l^h\|^2, \quad \mathcal{I}_{R^h}(r_l^h) = \begin{cases} 0 & \text{if } r_l^h \in R^h \\ \infty & \text{else} \end{cases}.$$

The primal-dual algorithm is now as follows. We choose the primal and dual steps widths $\tau > 0$, $\sigma > 0$. We let $(u^h)^0 \in U^h$, $(v^h)^0 \in V^h$, $(p^h)^0 \in P^h$, $(q^h)^0 \in Q^h$ and $(r_l^h)^0 \in R_l^h$, $l = 1, \dots, K$. Furthermore we let $(\bar{u}^h)^0 = (u^h)^0$ and $(\bar{v}^h)^0 = (v^h)^0$. Then for any $n \geq 0$ the iterations are given by

$$\left\{ \begin{aligned} (p^h)^{n+1} &= \mathcal{P}_{P^h}((p^h)^n + \sigma \nabla^h(\bar{u}^h)^n) \\ (q^h)^{n+1} &= \mathcal{P}_{Q^h}((q^h)^n + \sigma \mathcal{E}^h(\bar{v}^h)^n) \\ (r_l^h)^{n+1} &= (I + \sigma \partial F)^{-1}((r_l^h)^n + \sigma((\bar{u}^h)^n - f_l^h)), \quad l = 1 \dots K \\ (u^h)^{n+1} &= (u^h)^n - \tau \left((\nabla^h)^T(p^h)^{n+1} + \sum_{l=1}^K (r_l^h)^{n+1} \right) \\ (v^h)^{n+1} &= (v^h)^n - \tau \left((\mathcal{E}^h)^T(q^h)^{n+1} \right) \\ (\bar{u}^h)^{n+1} &= 2(u^h)^{n+1} - (u^h)^n \\ (\bar{v}^h)^{n+1} &= 2(v^h)^{n+1} - (v^h)^n \end{aligned} \right. \quad (14)$$

The Euclidean projectors $\mathcal{P}_{P^h}(\hat{p}^h)$ and $\mathcal{P}_{Q^h}(\hat{q}^h)$ are given by

$$(\mathcal{P}_{P^h}(\hat{p}^h))_{i,j} = \frac{\hat{p}_{i,j}^h}{\max\{1, \|\hat{p}_{i,j}^h\|/\alpha_1\}}, \quad (\mathcal{P}_{Q^h}(\hat{q}^h))_{i,j} = \frac{\hat{q}_{i,j}^h}{\max\{1, \|\hat{q}_{i,j}^h\|/\alpha_0\}}.$$

The solution of the resolvent operator $(I + \sigma\partial F)^{-1}(\hat{r}_l^h)$ is defined through the minimization problem

$$(I + \sigma\partial F)^{-1}(\hat{r}_l^h) = \arg \min_{\rho^h} \frac{\|\rho^h - \hat{r}_l^h\|^2}{2\sigma} + F(\rho^h). \quad (15)$$

This problem poses a simple quadratic problem with pointwise ball constraints. The solution of (15) is given by

$$((I + \sigma\partial F)^{-1}(\hat{r}_l^h))_{i,j} = \frac{(\hat{r}_l^h)_{i,j}/(1 + \sigma\delta)}{\max\{1, |(\hat{r}_l^h)_{i,j}/(1 + \sigma\delta)|\}}.$$

Note that the basic iterations of the algorithm are extremely simple. Indeed, the main computational effort of each iteration consists of computing the discrete gradient operations which are resembled by matrix vector products with very sparse matrices. The algorithm is therefore easy to implement and can efficiently be accelerated on parallel hardware such as graphics processing units. Finally, note that by using $K = 1$ and δ large enough our model reduces to a TGV^2 model with quadratic data term and only one input image. Hence, the proposed primal-dual algorithm can be used to efficiently minimize the original TGV^2 -based denoising model of [8].

5 Results

In this section we present experimental results for the proposed TGV^2 -based range image fusion model. The robustness of the method is evaluated on synthetic data and its practical applicability is justified by applying the proposed method to a large scale city dataset.

5.1 Synthetic Data

In our first experiment we apply the proposed fusion model to synthetically generated data to study the effects of noise and in particular the impact of redundancy. The synthetic dataset consists of an idealized building block with roof shapes, fairly common in urban environments. On the left hand side there is a gabled roof, whereas on the right hand side a hip roof is modeled. The dynamic range of this synthetic signal is $[50, 200]$, in the experiments this signal is sampled on a regular grid of 256×256 pixels. Real world range images usually contain a small amount of additive noise but contain a large number of gross outliers. In order to simulate this noise characteristics we use the following procedure to generate the degraded images. First, we add zero mean Gaussian noise with a

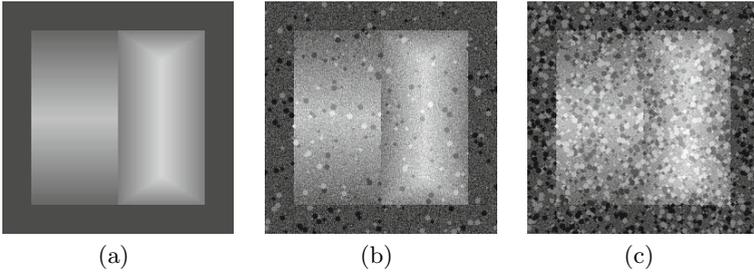


Fig. 1. Illustration of the synthetic dataset used to evaluate the performance of the proposed TGV^2 -based fusion model. Image (a) depicts the ground truth, (b) and (c) show degraded observations samples by adding Gaussian noise of $\sigma = 10$ and 10% and 50% outliers, respectively.

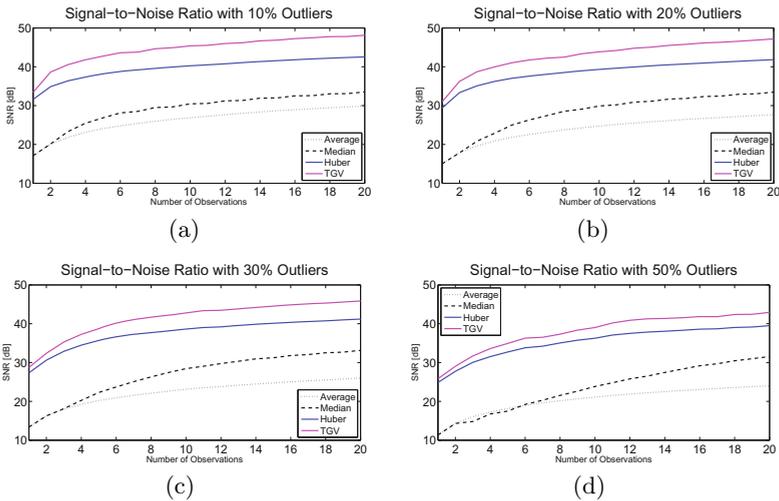


Fig. 2. Signal-to-noise ratio (SNR) for different degrees of degradation and a different number of observations. While for all methods the SNR increases with an increasing number of observations, the proposed TGV^2 model performs best in all situations.

standard deviation of 10. Second, a certain amount of pixel groups with variable sizes are replaced with outliers having an offset of ± 50 . Fig. 1 shows the ground truth range image and the degraded versions with an increasing number of outliers. Fig. 3 (a) shows a 3D rendering of the ground truth range image.

We compared our method against three methods: The first two methods are two very simple methods - computing the point-wise average and computing the point-wise median. The third method is the Huber model (6) which serves as the baseline model. For minimization of the Huber model we use an adapted version of the proposed primal-dual algorithm (14). For each method and experiment, we determined the optimal parameter settings. Note that for the Huber model, this implies that also the quadratic, ROF and $TV-L^1$ models are included in the

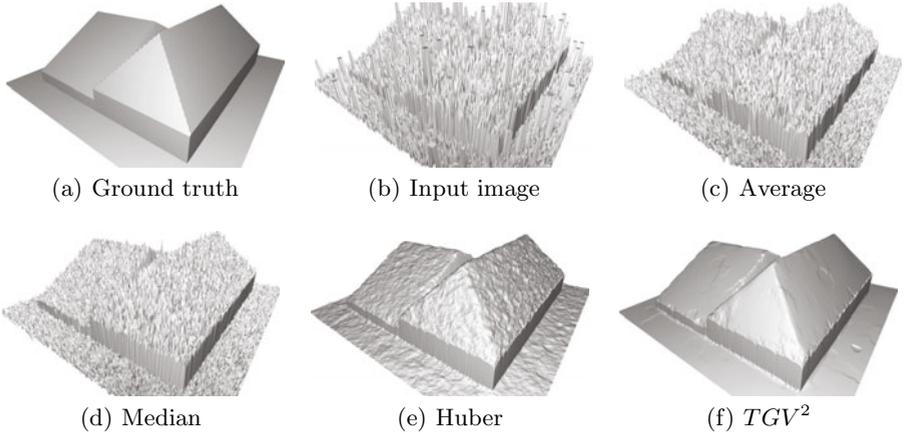


Fig. 3. *3D* renderings of an experiment with 10% outliers and 5 observations. The proposed TGV^2 -based fusion model significantly outperforms the Huber model. It very well captures the piecewise affine nature of the *3D* model and hence leads to a natural reconstruction of the *3D* building. Also note that the simple methods (Average and Median) completely fail to reconstruct the *3D* building.

parameter optimization. Performance of each method is evaluated with respect to both the degree of noise and the number of observations.

Fig. (2) shows the results of the performance evaluation. For evaluation purpose, we computed the signal-to-noise ratio (SNR)

$$SNR = 10 \log_{10} \frac{\|g^h\|^2}{\|u^h - g^h\|^2},$$

between the ground truth image g^h and the solution of the method u^h . Note that the proposed TGV^2 model performs best in all situations. Interestingly, the advantage of the TGV^2 model over the Huber model increases with an increasing number of observations. From Fig. 3, one can further see that the TGV^2 model captures the piecewise affine nature of the synthetic *3D* building very well. On the other hand, the Huber model inherently imposes a bias towards piecewise constant solutions and hence exhibit small staircases on the roof.

5.2 Real Data

In our second experiment we apply the proposed TGV^2 model to a large real-world dataset of Graz. The data set consist of 155 aerial images each of 11500×7500 pixels covering an area of 7.6 km^2 . The images have an average along-strip overlap of 85% and across-strip overlap of 65% at a ground sampling distance of 8 cm which amounts to about twenty observations per ground pixel on average. The input range images were produced by applying a state-of-the-art dense image matching algorithm based on scanline optimization and backmatching similar to [15]. For each image a pixel-synchronous range image is computed by stereo

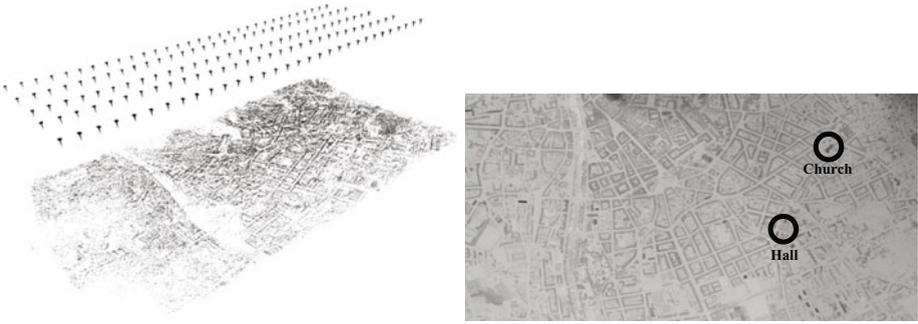


Fig. 4. Large aerial data set of Graz (a) shows a sparse reconstruction together with the camera positions. (b) shows the final digital surface model. The black circles mark two buildings for which detailed views are presented in Fig. 6 and Fig. 5.

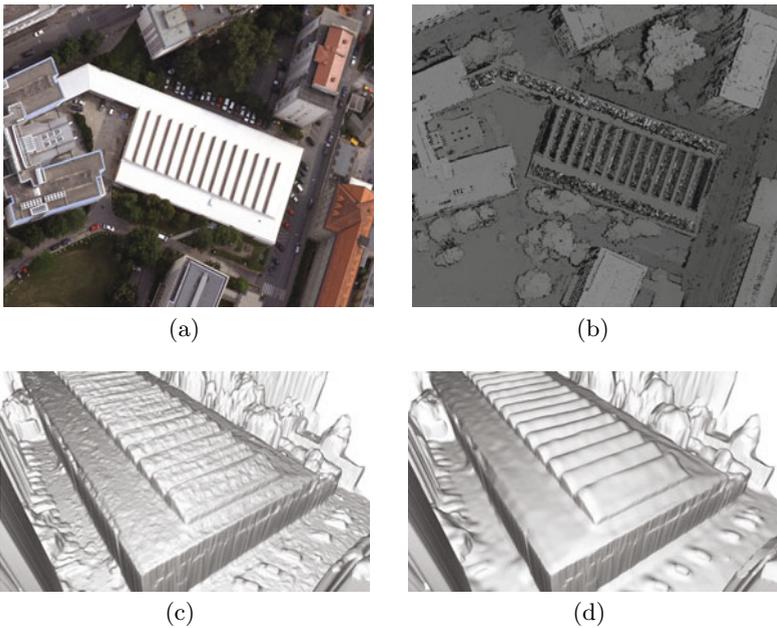


Fig. 5. Hall: (a) shows one input image, (b) shows one corresponding range image (black pixels denote invalid ranges), (c) shows the result of range image fusion using the Huber model and (d) shows the result using the TGV^2 model

matching of two adjacent images. See Fig. 6 (a)-(b) and Fig. 5 (a)-(b) for some sample images. The range images are then converted into a point cloud, which is projected onto the ground plane, thus giving multiple observations per pixel on the ground. Fig. 4 (a) shows a sparse reconstruction of the city together with the regular pattern, the images have been taken from the plane.

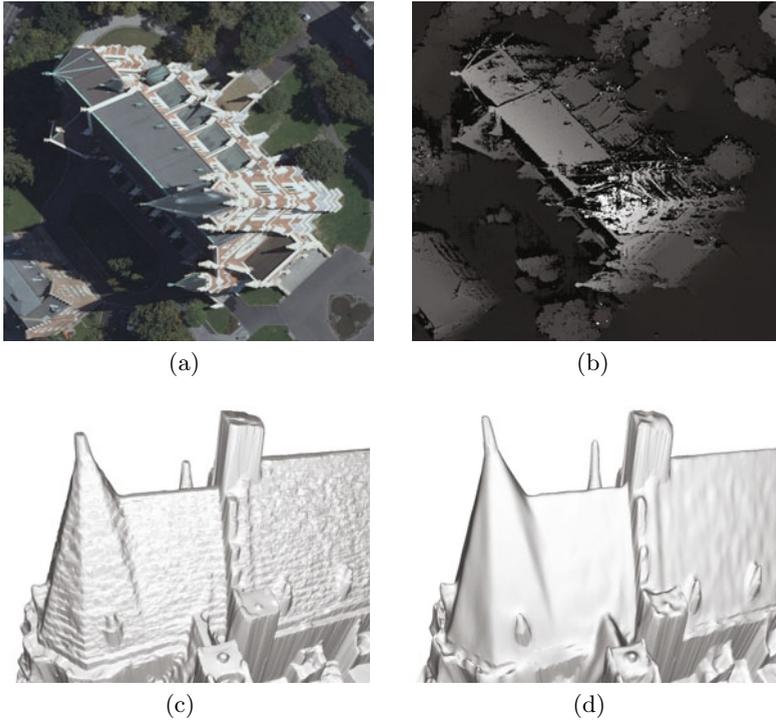


Fig. 6. Church: (a) shows one input image, (b) shows one corresponding range image (black pixels denote invalid ranges), (c) shows the result of range image fusion using the Huber model and (d) shows the result using the TGV^2 model

We applied the proposed TGV^2 -based range image fusion method to compute one single digital surface model (*DSM*) out of the range images. Clearly, the complete data does not fit the memory of any graphics card. Thus we divided the complete image domain into 282 tiles each of 2048×2048 pixels and applied the algorithm to each of the tiles. In order to eliminate boundary effects we used a sufficiently large overlap between the tiles. Thanks to a GPU-implementation of the primal-dual algorithm and a workstation equipped with 4 Nvidia Tesla GPUs, we were able to process 4 tiles at once. The overall processing time for the city data set was about 45 minutes for the stereo matching and 18 minutes for the fusion.

Fig. 4 (b) depicts the final *DSM* model. The black circles mark two interesting buildings, for which we provide detailed 3*D* renderings. Fig. 5 shows a large hall with a roof that consist of a regular pattern of planes. Note that due to reflections, the range images are very sparse in these areas. The proposed TGV^2 model leads to much smoother results and keeps the details of the 3*D* building. Fig. 6 shows a detailed 3*D* view of a church with very steep roofs. One can see that the TGV^2 model leads to much smoother results while keeping the important details in the reconstruction. On the other hand, the Huber model exhibits strong staircasing on slanted surfaces such as the roof.

6 Conclusion

In this paper we proposed a convex variational approach for the fusion of several range images to a single high quality digital surface model. The variational model is based on total generalized variation regularization of second order and a robust Huber- L^1 data term. For minimization we proposed an simple first-order primal-dual algorithm, which can be efficiently parallelized on graphics processing units. In experimental results on synthetic data we showed the robustness of the method against noise and - most importantly - showed that the method efficiently exploits redundancy in the data. Results on real data, showed that the proposed method can be applied to large scale data sets. Future work will concentrate on studying total generalized variation of an order larger than two. The proposed solution is another step towards realistic 3D models of our habitat.

References

1. Leberl, F., Bischof, H., Pock, T., Irschara, A., Kluckner, S.: Aerial computer vision for a 3d virtual habitat. *Computer* 43, 24–31 (2010)
2. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building rome in a day. In: *International Conference on Computer Vision, ICCV (2009)*
3. Zach, C., Pock, T., Bischof, H.: A globally optimal algorithm for robust TV- L^1 range image integration. In: *Proceedings of the 11th International Conference Computer Vision, Rio de Janeiro, Brazil*, pp. 1–8 (2007)
4. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: *Proceedings of SIGGRAPH 1996*, pp. 303–312 (1996)
5. Tikhonov, A.N.: On the stability of inverse problems. *Dokl. Akad. Nauk SSSR* 5, 195–198 (1943)
6. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259–268 (1992)
7. Nikolova, M.: A variational approach to remove outliers and impulse noise. *J. Math. Imaging and Vision* 20, 99–120 (2004)
8. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. Technical report, Institute for Computer Graphics and Vision (2010)
9. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259–268 (1992)
10. Aujol, J.F., Gilboa, G., Chan, T., Osher, S.: Structure-texture image decomposition—modeling, algorithms, and parameter selection. *Int. J. Comp. Vision* 67, 111–136 (2006)
11. Chan, T., Esedoglu, S.: Aspects of total variation regularized L^1 function approximation. *SIAM J. Appl. Math.* 65, 1817–1837 (2004)
12. Chan, T., Esedoglu, S., Park, F., Yip, A.: Total Variation Image Restoration: Overview and Recent Developments. In: *Mathematical Models in Computer Vision*. Springer, Heidelberg (2005)
13. Huber, P.: *Robust Statistics*. Wiley, New York (1981)
14. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging (2010), <http://hal.archives-ouvertes.fr/hal-00490826>
15. Hirschmüller, H.: Stereo vision in structured environments by consistent semi-global matching. In: *Conference on Computer Vision and Pattern Recognition*, pp. 328–341 (2006)