Globally optimal robust DSM fusion

Roland Perko and Christopher Zach

1Joanneum Research, Graz, AT
2Toshiba Research Europe, Cambridge, UK
*Corresponding author, e-mail address: roland.perko@joanneum.at

Abstract
This work presents the mathematical formulation of a novel globally optimal robust digital surface model (DSM) fusion method, that can be used to combine several 2.5D DSMs extracted from airborne or spaceborne stereo images. The main novelty is the definition of a convex energy functional with a \( \beta \)-Lipschitz continuous gradient that allows a trivial solution of the posed minimization problem, where the robustness is achieved by incorporating the Huber norm into the energy functional. All according mathematical proofs are derived within this work. The experiments are based on two different minimization schemes and are applied on airborne optical, on spaceborne optical and on spaceborne synthetic aperture radar (SAR) images. The resulting fused 2.5D DSMs are rich in detail and are of higher quality than results of other local fusion methods.

Keywords: DSM fusion, robust fusion, multiple view geometry, energy functional, global minimization, gradient descent.

Introduction
When acquiring multiple images from different view points over arbitrary objects for mapping purposes in the field of remote sensing, the standard procedure is to perform a pure stereo processing. This procedure yields one 2.5D model, i.e. a digital surface model (DSM) or elevation map, for each possible stereo pair. Such a 2.5D model is typically stored as an image raster with equidistant spacing where each pixel represents the elevation of the covered terrain area. It represents a top view and thus only encodes the highest elevation points and e.g. not the façades of buildings.

The individual DSMs have to be fused into one final 2.5D DSM, which should have a better quality than the individual ones. In addition, undefined regions, which are not reconstructed, e.g. due to occlusions in the single stereo pairs, should be filled with height information from other pairs. This general issue of fusion is omnipresent when it comes to 3D reconstruction from multiple view images.

In literature the mentioned principal philosophy of 2.5D DSM fusion is applied on very different imagery, e.g. on images from a hand held camera [Grab et al., 2011], on images from airborne cameras [Hirschmüller, 2008; Rothermel et al., 2012], on remote sensing image from optical stereo satellites [Perko et al., 2014] or even on SAR satellite data.
Globally Optimal Robust DSM Fusion

Perko and Zach

This work focuses on fusing multiple 2.5D elevation maps into one final 2.5D elevation map. However, it should be noted that in general such a fusion could be also done in complete 3D. Examples in literature for 3D fusion are e.g. global methods [Zach et al., 2007] or local methods [Fuhrmann and Goesele; 2014, Kuhn et al., 2014]. However, in this work the term fusion always represents a 2.5D elevation map fusion.

In this work a global formulation for DSM fusion with a convex energy functional employing the Huber norm is presented. This convex energy functional consists of a regularization term and a data term. While the first enforces a smooth surface including sharp 3D breaklines (the robustness to outliers comes from the Huber norm which penalizes larger elevation jumps only linearly like the $L_1$ norm), the second drags the solution towards the given input DSMs. Defining an energy functional which uses the Huber norm is a common practice to tackle optimization problems in computer vision. Similar formulations have e.g. used for fusing aerial images [Kluckner et al., 2010], super-resolution techniques [Unger et al., 2010], dense stereo matching [Heise et al., 2013], optical flow estimation [Werlberger et al., 2009], image deconvolution and denoising [Chambolle and Pock, 2011] or DSM fusion [Pock et al., 2011]. The main contribution of this work is the definition of a convex energy functional that is differentiable and has a $\beta$-Lipschitz continuous gradient (similar to [Pock et al., 2011]). Moreover, the convexity of the functional and the Lipschitz continuity is mathematically proven. This allows to cast the minimization problem on a trivial gradient descent (GD) optimizer (cf. [Combettes and Pesquet, 2011]; a very comprehensive collection on all mathematical terms used in the work, like $\beta$-Lipschitz continuity or Huber norm, can be found in [Vandenberghe, 2014]). In contrast, other works mostly use a variational method for optimization. In addition, the GD optimizer is compared with the fast iterative shrinkage-thresholding algorithm (FISTA) [Beck and Teboulle, 2009], which yields faster convergence.

In order to show the generality of the approach, the proposed algorithms are evaluated on a synthetic data set, on a photogrammetric airborne data set (UltraCam), on a photogrammetric spaceborne data set (Pléiades) and on a radargrammetric SAR spaceborne data set (TerraSAR-X Staring Spotlight mode). The results are compared to local methods, that take the mean or median value over one DSM cell or the method by [Rumpler et al., 2013] which determines a probability density function (pdf) for a local pixel neighborhood and extracts the mode of this pdf.

Other global solutions for 2.5D DSM fusion, like [Graber et al., 2011; Pock et al., 2011; Kuschk and d’Angelo, 2013], are very related to the proposed formulation. Those works often use total variation based methods for minimizing the energy functional such that similar results could be expected. The main difference to the presented work is the used optimization technique.

**Mathematical Formulation**

As a first step an energy functional $E(u)$ has to be defined that should be minimized, where is the resulting fused DSM. Second, it has to be proved that $E(u)$ is a convex functional that is differentiable and has a $\beta$-Lipschitz continuous gradient. These are the prerequisites that a simple gradient descent algorithm will converge and is thus able to find the global, not necessarily unique, minimum $u$. Third, two optimization algorithms that can be used for solving the presented optimization problem are given.
**Formulation of the Energy Functional**

The energy $E(u)$ to be minimized is defined in Equation [1]. The energy functional is composed of two terms, the regularization term and the data term. The first ensures a smooth solution while the second forces the solution to be close to the individual input DSMs, called $f_i$. This specific formulation of the two terms is used, since the Huber norm [Huber, 1973] is a *convex approximation of the truncated quadratic* (actually the Huber norm is the Moreau-envelope of $\gamma \| \cdot \|_2$ [Parikh and Boyd, 2013] or equivalently a smoothed version of the $L_2$ norm which makes it somewhat robust to outliers while still leading to a stronger smoothing effect) and favors piecewise smooth solutions (due to the regularization) and is robust to outliers (in the data term).

$$E(u) = \int_{\Omega} \left\{ \alpha \| \nabla u \|_2^2 + \lambda \sum_{i=1}^{k} \omega_i \| u - f_i \|_2 \right\} dxdy \quad [1]$$

with:

$$\nabla u = \left( \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y} \right) = (u_x, u_y) \quad [2]$$

and:

$$|a|_\gamma = \begin{cases} 
\frac{a^2}{2\gamma} & \text{if } |a| \leq \gamma \\
|a| - \frac{\gamma}{2} & \text{if } |a| > \gamma
\end{cases} \quad [3]$$

with:
- $E(u)$ energy to be minimized \( E : \mathbb{R}^{m \times n} \rightarrow \mathbb{R} \)
- $u(x, y)$ final 2D elevation map, i.e. the desired fused DSM with $m \times n$ pixels $\in \mathbb{R}^{m \times n}$
- $f_i(x, y)$ input DSMs $\left( i \in [1, k] \subset \mathbb{N} \right)$ with each pixels $m \times n$ pixels $\in \mathbb{R}^{m \times n}$
- $\nabla u$ partial derivatives of $u \in \left( \mathbb{R}^{m \times n}, \mathbb{R}^{m \times n} \right)$, in particular the forward differences $u_x$ and $u_y$
- $\alpha$ total variation regularization weight $\in \mathbb{R}^+$
- $\lambda$ data fidelity weight $\in \mathbb{R}^+$
- $\omega_i$ weights of input measurements with $\sum_i \omega_i = 1$, each $\in \mathbb{R}^+$
- $\Omega$ image plane, i.e. the image region $\in \mathbb{R}^{m \times n}$
- $|a|_\gamma$ Huber norm of $\in \mathbb{R}^N$ with parameter $\gamma \in \mathbb{R}$ and $|a|_\gamma : \mathbb{R}^N \rightarrow \mathbb{R}$
The first and second derivatives of the Huber function are:

\[
\frac{d |a|}{da} = \begin{cases} 
  a / \gamma & \text{if } |a| \leq \gamma \\
  \text{sgn}(a) & \text{if } |a| > \gamma
\end{cases} \tag{4}
\]

and

\[
\frac{d^2 |a|}{da^2} = \begin{cases} 
  1 / \gamma & \text{if } |a| \leq \gamma \\
  0 & \text{if } |a| > \gamma
\end{cases} \tag{5}
\]

Obviously, the regularization weight \( \alpha \) can be increased to get a smoother solution, while increasing the data fidelity weight \( \lambda \) drags the fused DSM towards the input DSMs. A larger parameter \( \xi \) of the regularizer’s Huber norm ensures a more detailed result, while a large \( \zeta \) yields a smoother solution. Overall, the two parameter sets are highly correlated which might lead to an over-parametrization. The real difference however is that \( \alpha \) and \( \lambda \) are just weights to balance the regularization and the data term, whereas \( \xi \) and \( \zeta \) are metric Huber parameters given in meters. A visualization of the influences of those four parameters on a real example will be presented in the results section in Figure 10.

It should be noted that using only forward differences (cf. Eq. [8]) as discrete derivatives is better than using central differences, as in this case the 2D field \( u(x, y) \) would get decoupled like a checkerboard field. A better solution would be a staggered grid [Harlow and Welch, 1965] which is however not the main point of this work.

**Convexity of the Energy Functional**

Next, it has to be shown that the energy \( E : \mathbb{R}^{m \times n} \rightarrow \mathbb{R} \) is a convex function which means that:

\[
\forall u, v \in \mathbb{R}^{m \times n}, \forall t \in [0,1] \subseteq \mathbb{R}: E(tu + (1-t)v) \leq tE(u) + (1-t)E(v) \tag{6}
\]

This is obviously the case, as the Huber norm is a convex function (cf. Fig. 1), sums of convex functions are convex, and the composition of convex function and linear term is convex (\( \nabla_u \) is linear in \( u \)). That is actually one reason why the Huber norm was chosen within the energy functional. \( E(u) \) is also differentiable as the Huber norm is differentiable.
**Lipschitz Continuity of the Energy Functional Gradient**

Next, it is shown that $E(u)$ has a $\beta$-Lipschitz continuous gradient $\nabla E$, which means that a positive Lipschitz constant $\beta$ exists with:

$$\forall u, v \in \mathbb{R}^{m \times n} : \| \nabla E(u) - \nabla E(v) \| \leq \beta \| u - v \| \quad [7]$$

This constraint basically enforces that the gradient of the functional $E(u)$ is bounded and thus a gradient descent algorithm will converge. Now consider the energy, given in Equation [1] and its discretized variant:

$$E(u) = \sum_{s \in N^2 \cap \Omega} \left( \alpha \left| u_s - u_{s+(1,0)} \right|^2 + \alpha \left| u_s - u_{s+(0,1)} \right|^2 + \lambda \sum_{i} \alpha_i \left| u_s - f_{s,i} \right|_\xi \right) \quad [8]$$

In order to apply a gradient descent algorithm, an estimate of the respective Lipschitz value $\beta$ of $\nabla E(u)$ is needed. Thus, a new set of unknowns $y = Au$ (with three times the number of elements in $u$) is introduced, such that

$$y_{s,1} = u_s - u_{s+(1,0)^T} \quad [9]$$
\[ y_{s,2} = u_s - u_{s+(0,1)^T} \]  \[ y_{s,3} = u_s \]

and the energy is expressed in terms of \( y \),

\[
\tilde{E}(y) = \sum_{s \in \mathbb{N}^2 \cap \Omega} \left( \alpha |y_{s,1}|_\zeta + \alpha |y_{s,2}|_\zeta + \lambda \sum_i \omega_i |y_{s,3} - f_{s,i}|_\zeta \right)
\]

Then \( \tilde{E}(Au) = E(u) \), and by the chain rule, \( \nabla_u \tilde{E}(Au) = A^T \nabla_y \tilde{E}(y) \|_{y=u} \). For a differentiable function \( \tilde{E} \) and a matrix \( A \), the upper bound of the Lipschitz constant \( L \) of \( \nabla \tilde{E}(Au) \) is given by \( L \leq \|A\|_2^2 \|\tilde{E}\|_2 \), where \( \|\tilde{E}\|_2 \) is the respective operator norm of the matrix \( A \). Consequently, the Lipschitz constant of \( \nabla E(u) \) can be bounded if \( \|A\|_2 \) and \( \|\tilde{E}\|_2 \) is known. Since \( \tilde{E}(y) \) is a sum of terms over independent unknowns, \( \tilde{E} \) is bounded by the largest Lipschitz value of each terms, i.e. \( \|\tilde{E}\|_2 \leq \max \{\alpha / \xi, \lambda / \zeta\} \) (also cf. Eq. [5]). This can be seen as follows: Let \( h(x, y) = f(x) + g(y) \), where \( f \) and \( g \) have Lipschitz gradient with values \( L_f \) and \( L_g \), respectively. Then:

\[
\left\| \nabla f(x) - \nabla f(x') \right\|^2 + \left\| \nabla g(y) - \nabla g(y') \right\|^2 \\
\leq L_f^2 (x-x')^2 + L_g^2 (y-y')^2 \\
\leq \max \{L_f^2, L_g^2\} \left\| x - x' \right\|_2^2 \left\| y - y' \right\|_2^2
\]

It still remains to estimate the operator norm of \( A \). Thus, the fact that \( \|A\|_2 \leq \|A\|_1 \) is used (actually a special case of the Hölder’s inequality). Note that \( \|A\|_1 \) is the maximum absolute column sum, and \( \|A\|_\infty \) is the maximum absolute row sum. The columns of \( A \) are indexed by the (original) unknowns \( (u_s) \), and the rows of \( A \) correspond to the variables \( y_{s,k} \) \((k=1,2,3)\). Since all occurrences of \( u_s \) have a +1 or –1 coefficient, it is sufficient to just count the occurrences of each variable. Since at most 2 variables appear in one term (rows corresponding to e.g. \( y_{s,1} = u_s - u_{s+(1,0)^T} \)), it yields \( \|A\|_\infty = 2 \). \( u_s \) for a particular \( s \) appears in terms (in the expression for \( y_{s,k} \) for \( k=1,2,3 \)), but also in the ones for the left and upward neighbors \( y_{s-(0,1)^T,1} \) and \( y_{s-(0,1)^T,2} \), and, thus, \( \|A\|_1 = 5 \). Overall the bound is \( \|A\|_2 \leq 2 \cdot 5 = 10 \). Combining everything, the Lipschitz constant \( \beta \) for \( \nabla E \) is at most \( \max \{\alpha / \xi, \lambda / \zeta\} \).

**Minimization Schemes for the Energy Functional**

Using the findings above, the minimum of the function can be found employing a gradient
descent method (cf. [Combettes and Pesquet, 2011]), as sketched below (the step size $\beta^{-1}$ is actually $1 / 10 \max\{\alpha / \xi, \lambda / \zeta\}$, employing the previously calculated Lipschitz constant):

| Algorithm 1 | Gradient Descent (GD) |
|-------------|-----------------------|
| 1: for $n = 1, 2, \ldots$ do |
| 2: $x_n = x_{n-1} - \beta^{-1}\nabla E(x_{n-1})$ |
| 3: end for |

There are many other methods to optimize a convex function (e.g. conjugate gradient, Euler-Lagrange, accelerated gradient), where the fast iterative shrinkage-thresholding algorithm (FISTA) [Beck and Teboulle, 2009] is presented, actually an accelerated gradient method, with the definition from [Vandenberghe, 2014]:

| Algorithm 2 | Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) |
|-------------|-------------------------------------------------------|
| 1: for $n = 1, 2, \ldots$ do |
| 2: $y = x_{n-1} + \frac{n-1}{n+2}(x_{n-1} - x_{n-2})$ |
| 3: $x_n = y - \beta^{-1}\nabla E(y)$ |
| 4: end for |

FISTA is chosen for comparison to GD due to its simplicity and due to its significantly improved convergence behavior (the error rate of GD is at $O(1/k)$ at $k$ iterations while FISTA allows an optimal error rate of $O(1/k^2)$). The main difference between FISTA and GD is that it is not employed on the previous point $x_n$, but rather at the point $y$ which uses a very specific linear combination of the previous points $x_{n-1}$ and $x_{n-2}$. The computationally expensive of GD and FISTA is basically the same, namely the computation of , respectively . Since both GD and FISTA need a starting DSM, namely $x_0$, the median based fusion is taken and the remaining invalid regions are interpolated by means of linear hole filling. For FISTA $x_{-1}$ is additionally set to $x_0$.

**Numerical Discrete Solution for the Partial Derivatives**

The only missing part is a numerical discrete solution for calculating the partial derivatives $\nabla E(u)$, which can be derived as (cf. Eq. [8]):

$$
\frac{\partial E(u)}{\partial u_{x,y}} = \alpha \left( \frac{d|u_{x,y} - u_{x+1,y}|}{du_{x,y}} + \frac{d|u_{x,y} - u_{x,y+1}|}{du_{x,y}} \right) + \lambda \sum_i \omega_i \frac{d|u_{x,y} - f_i|}{du_{x,y}} \quad [14]
$$

E.g. one has:
\[
\frac{d|u_{x,y} - u_{x+1,y}|_\xi}{du_{x,y}} = \begin{cases} 
\frac{1}{\xi} (u_{x,y} - u_{x+1,y}) & \text{if } |u_{x,y} - u_{x+1,y}| \leq \xi \\
\text{sgn}(u_{x,y} - u_{x+1,y}) & \text{otherwise}
\end{cases}
\]  

[15]

\[
\frac{d|u_{x,y} - f_i|_\xi}{du_{x,y}} = \begin{cases} 
\frac{1}{\xi} (u_{x,y} - f_i) & \text{if } |u_{x,y} - f_i| \leq \xi \\
\text{sgn}(u_{x,y} - f_i) & \text{otherwise}
\end{cases}
\]  

[16]

**Additional Notes**

Details on the weights \( \omega_i \) (influencing the individual DSMs): The default setting would be to handle all DSMs equally \( \omega_i = 1/k \), like done in all tests presented in this paper. However, since the DSM quality is dependent on the stereo intersection angle \( \delta_i \), it would be a good idea to set \( \omega_i = \tan(\delta_i) \) and \( \omega_i = \sum \omega_i \) to model the sensitivity w.r.t. the intersection angle. Smaller angles yield lower 3D accuracy and thus a smaller weight in the DSM fusion and vice versa. Also the degree of texturedness at a pixel location influence the DSM quality and could be modelled in the fusion process. In this specific case the weight function would become two-dimensional \( \omega_i(x, y) \).

Also note that the input DSMs \( f_i \) may contain invalid values, i.e. undefined pixels visualized in black in Figures 7, 11, 14. Those pixels are not considered when evaluating \( \nabla E(u) \) in Algorithm 1 and Algorithm 2. Thus, the data term of \( E(u) \) may become zero if no height measurements are available at this pixel coordinate. In this case the regularization term is the only force on that pixel location and ensures a smooth (interpolated) surface.

The proposed energy functional was chosen in the special way, that a trivial gradient descent algorithm will converge to a globally optimal, not necessarily unique, solution. On the contrary, in literature researcher often approximate a non-convex energy functional with a convex quadratic function, e.g. using second-order Taylor expansion (see [Werlberger et al.; 2010; Ranftl et al., 2014] for optical flow estimation). In this case the resulting solution is however not a minimum of the original non-convex formulation.

**Test Data and Evaluation Strategy**

**Test Data**

In order to show the generality of the proposed method, tests were conducted on four very different data sets described below. The photogrammetric respectively radargrammetric processing was done by employing the Remote Sensing Software Graz (RSG) (http://www.remotesensing.at/en/remote-sensing-software.html) developed by the authors.

**Synthetic Data**

Constructed from a building block with roof shapes, fairly common in urban environments with 256×256 pixels and a height dynamic of 50 to 200 digital numbers (see Fig. 2a), similar to the one proposed in [Pock et al., 2011]. To simulate more realistic examples, five
input DSMs were generated by randomly introducing gross outliers (a common artifact from image matching) and adding Gaussian noise. Examples are given in Figure 2b-c.

**Airborne Optical Data**
The data set Vaihingen in Germany from the EuroSDR benchmark (http://www.ifp.uni-stuttgart.de/ISPRS-EuroSDR/ImageMatching/index.en.html) on High Density Image Matching for DSM Computation [Haala, 2013] consists of 36 UltraCam-X images with a ground sampling distance (GSD) of 20 cm. Using the photogrammetric workflow mentioned above DSMs with 20 cm spacing were extracted for each stereo pair. For this test a representative area of 451×401 pixels holding the castle Kaltenstein is chosen which is covered by 20 individual stereo DSMs with a height dynamic of 280 to 350 meters. Six of those input DSMs are shown in Figure 7.

**Spaceborne Optical Data**
This data set is composed of a Pléiades tri-stereo satellite acquisition (panchromatic; 50 cm GSD) over the city of Innsbruck in Austria. From each of the adjacent stereo pairs DSMs were generated using the workflow from [Perko et al., 2014], resulting in four input DSMs with 1 m spacing shown in Figure 11. A subset of 701×401 pixels is chosen in the city center with a height dynamic of 610 to 690 meters.

**Spaceborne SAR Data**
This data set holds six TerraSAR-X Staring Spotlight images that cover the region of Burgau in Austria from ascending and descending orbit with three different look angles per orbit. For this specific processing multi look ground range detected (MGD) images, that represent the magnitude of the radar backscatter, were employed with a GSD down to 25 cm. The radargrammetric processing suite [Gutjahr et al., 2014] was employed to extract the 12 individual stereo DSMs with 1 m spacing. A representative area of 1501×1001 pixels is chosen covering agricultural fields and a forest with a height dynamic of 275 to 365 meters. Six of those DSMs are shown in Figure 14. It should be noted that radargrammetric processing of TerraSAR-X Staring Spotlight data is a very young topic and first results were recently published by the authors in [Hennig et al., 2015; Gutjahr et al., 2015].

**Evaluation Strategy**
To evaluate the results statistical values are extracted by comparing the resulting DSM with a given ground truth DSM $h_{gt}$. Calculated are minimal (min) and maximal (max) error, standard deviation of error (std), mean absolute error (mae) [Haala et al., 2010], normalized median absolute deviation (nmad) [Rousseeuw and Croux, 1993; Leys et al., 2013], and signal-to-noise ratio (snr) [Pock et al., 2011]. Let $\Delta h = h_{gt} - h_r$, then those measures are defined as:

\[
\text{mae}(\Delta h) = \text{mean}(|\Delta h|) \quad [17]
\]

\[
\text{nmad}(\Delta h) = 1.4826 \text{median}(|\Delta h - \text{median} \Delta h|) \quad [18]
\]
For the synthetic dataset this evaluation is simple, as the ground truth DSM is available. For the optical airborne dataset it becomes more difficult. The problem here is that the existing airborne laser scanner (ALS) reference DSM is of lower quality than the resulting photogrammetric DSMs w.r.t. point density and overall accuracy. Therefore, the ALS data was excluded from the EuroSDR benchmark [Haala et al., 2010]. However, the organizers calculated a median DSM, holding the median value for each DSM pixel of all DSMs that where send to the contest [Haala, 2014]. This median DSM is now taken as the ground truth in the presented evaluation. However, especially at 3D breaklines this median DSM is inaccurate such that the evaluation is still disputable. Similar difficulties in the evaluation where also reported in Kuschk and d’Angelo [2013]. For the spaceborne optical and SAR data sets no ground truths are available.

Results
For each test set the presented function $E(u)$ is minimized employing the simple gradient descend method and also the FISTA. The convergence of those algorithms is evaluated by plotting the residual error $|E(x_n) - E(x_*)|/|E(x_*)|$ with $x_*$ being the solution after 1000 iterations for each iteration $n$. This error metric is proposed in Vandenberghe [2014]. In addition to this numerical evaluation several DSMs are shown in 3D, where always the version after 1000 iterations is given. Results are visually compared to the mean and median fusion, which is defined by taking the statistical value from the pool of values for each DSM cell. In addition the results are compared to the probabilistic DSM fusion methodology [Rumpler et al., 2013] which determines a probability density function (pdf) for a local pixel neighborhood and extracts the mode of this pdf employing the mean shift procedure [Fukunaga and Hostetler, 1975].

All tests were performed using the same parameters of the function, which were chosen by tuning the fusion result of the airborne optical data set: $\alpha=1$, $\lambda=1$, $\xi=10$, $\zeta=0.1$, $\omega_i=1/k$, with $k$ being the number of input DSMs.

Synthetic Data
Figure 2 shows the noise-free DSM in (a) and two examples of the noisy input DSMs (b-c) to be fused. Figure 3 gives 3D views of the noise free DSM (a), an input noisy version (b), the mean based fusion (c), the median based fusion (d), the probabilistic fusion (e), and the proposed fusion (f). It is clearly visible that the globally optimal robust fusion is able to get rid of the gross outliers due to the global optimization and due to the smoothness constraint, while the local methods are not able to remove those outliers. From the set of local solutions the probabilistic fusion [Rumpler et al., 2013] performs best. The residual error for each iteration is given in Figure 4. While the simple GD optimizer needs a lot of iterations to get a small residual error, FISTA could be stopped after 50 iterations.

As a proof-of-concept the fusion is also performed with a zero solution, i.e. starting with...
a DSM $x_0$ holding zeros. Comparison of the final, mostly identical, solutions are given in Figure 5, while the convergence is plotted in Figure 6. Obviously, more iterations are necessary for convergence from a zero solution. Thus, in practice starting from a robust, e.g. the median, solution results in a significant speed-up. It is also visible that FISTA is no strict descent method, since the error e.g. increases after iteration 92 and decreases again after iteration 105 (actually a known effect [Beck and Teboulle, 2009; Vandenberghe, 2014]). Table 1 gives the statistical evaluation of errors. For this example the proposed method outperforms all other tested methods and obviously the fused models are better than the individual input models.

![Figure 2 - Synthetic data set: (a) Two houses and (b-c) noisy versions of (a).](image)

![Figure 3 - Fusion of synthetic data. Shown are 3D views of (a) the noise free input DSM, (b) an input noisy version, (c) the mean based fusion, (d) the median based fusion, (e) the probabilistic fusion, and (f) the proposed fusion.](image)
Figure 4 - Convergence of the proposed optimization schemes for synthetic data.

Figure 5 - Results of the synthetic data fusion. Shown are 3D views of the fusion starting from (a) the median solution, (b) the zero solution.

Figure 6 - Convergence of the proposed optimization schemes for synthetic data starting from the zero solution.
Table 1 - Statistical results of the synthetic data fusion. Best results are shown in bold face.

|      | min   | max   | std  | mae  | nmad | snr  |
|------|-------|-------|------|------|------|------|
| Input 1 | -88.00 | 88.00  | 24.37 | 16.40 | 13.34 | 13.07 |
| Input 2 | -87.00 | 90.00  | 24.22 | 16.29 | 13.34 | 13.13 |
| Input 3 | -89.00 | 83.00  | 24.19 | 16.27 | 13.34 | 13.14 |
| Input 4 | -89.00 | 88.00  | 24.28 | 16.37 | 13.34 | 13.11 |
| Input 5 | -85.00 | 87.00  | 24.36 | 16.42 | 13.34 | 13.08 |
| Mean  | -50.40 | 52.40  | 10.90 | 8.55  | 10.67 | 35.11 |
| Median| -65.00 | 65.00  | 9.01  | 6.16  | 7.41  | 38.10 |
| Mode  | -101.44 | 98.44  | 7.25  | 3.64  | 3.60  | 38.64 |
| Proposed | **-9.85** | **32.80** | **1.64** | **1.20** | **1.34** | **42.65** |

**Airborne Optical Data**

For visualization purposes six randomly selected DSMs out of the 20 input DSMs are given in Figure 7. Of course all 20 DSMs were used in the calculations. Shown are DSMs derived by the photogrammetric workflow described in Perko et al. [2014]. Due to occlusions and mismatches several regions of the DSMs do not have a height measurement and are shown in black. Figure 8 gives 3D views of two input DSMs (a-b), the mean based fusion (c), the median based fusion (d), the probabilistic fusion (e), and the proposed fusion (f). In (a-b) invalid pixels are set to the lowest height value just for visualization. Like in the synthetic example gross outliers are removed in the fused DSMs, e.g. the spikes right of the tower. For this data set the probabilistic fusion also works very well. However, it tends to smooth the DSM too much since it always takes the fixed neighborhood. The residual errors for each iteration are given in Figure 9. Again FISTA converges faster than GD. E.g. the residual error that FISTA achieves at iteration 50 is lower than the error of GD at iteration 250.

Table 2 gives the statistical evaluation of errors. For this example the proposed method outperforms all other tested methods in 5 out of 6 statistical values. It is not the best regarding the signal-to-noise ratio. Here it is assumed that this issue relates to the reference data which is not a real ground truth as discussed in the evaluation strategy section. Additionally, the fused models are better than the individual input models. The table also gives the percentage of valid pixels for each model. While the input DSMs have valid values in the range of 59 to 80 %, the fused models only contain valid pixels even in the case when all input DSMs are invalid at one position.

On this test set the influence of the parameters $\alpha$, $\lambda$, $\xi$ and $\zeta$ within the energy functional given in Equation [1] are visualized in Figure 10 by first setting them to $\alpha=1$, $\lambda=1$, $\xi=1$, $\zeta=1$ and then altering one parameter at a time. As stated before, increasing of $\alpha$ or $\zeta$ results in a smoother solution, while increasing $\lambda$ or $\xi$ yields a more detailed reconstruction. A useful trade-off for this data set are the parameters $\alpha=1$, $\lambda=1$, $\xi=10$, $\zeta=0.1$, which were thus used in all other tests as the standard parameter set.
Figure 7 - Airborne optical data set: Six of 20 input DSMs derived via semi-global image matching.

Figure 8 - Results of the airborne optical data fusion. Shown are 3D views of (a-b) two input DSMs, (c) the mean based fusion, (d) the median based fusion, (e) the probabilistic fusion, and (f) the proposed fusion.
Figure 9 - Convergence of the proposed optimization schemes for airborne optical data.

Table 2 - Statistical results of the airborne optical data fusion. Best results are shown in bold face.

| Input  | min    | max    | std   | mae   | nmad  | snr   | valid [%] |
|--------|--------|--------|-------|-------|-------|-------|-----------|
| 1      | -17.65 | 24.70  | 1.76  | 0.68  | 0.35  | 45.20 | 64.83     |
| 2      | -17.29 | 24.79  | 1.73  | 0.67  | 0.35  | 45.32 | 69.07     |
| 3      | -18.60 | 23.12  | 1.65  | 0.70  | 0.35  | 45.73 | 69.55     |
| 4      | -17.70 | 30.05  | 1.36  | 0.60  | 0.32  | 47.38 | 67.05     |
| 5      | -19.50 | 23.01  | 1.64  | 0.69  | 0.35  | 45.79 | 65.81     |
| 6      | -22.73 | 23.97  | 1.81  | 0.72  | 0.34  | 44.88 | 61.97     |
| 7      | -21.68 | 17.81  | 1.81  | 0.73  | 0.34  | 44.90 | 65.59     |
| 8      | -21.24 | 16.26  | 1.47  | 0.58  | 0.32  | 46.74 | 73.28     |
| 9      | -21.28 | 16.16  | 1.63  | 0.67  | 0.35  | 45.86 | 65.91     |
| 10     | -17.78 | 20.94  | 1.37  | 0.61  | 0.32  | 47.30 | 73.13     |
| 11     | -21.12 | 16.13  | 1.51  | 0.60  | 0.33  | 46.52 | 80.02     |
| 12     | -16.45 | 21.55  | 1.56  | 0.62  | 0.35  | 46.25 | 77.10     |
| 13     | -21.30 | 16.02  | 1.77  | 0.72  | 0.35  | 45.14 | 72.13     |
| 14     | -16.28 | 21.16  | 1.51  | 0.60  | 0.34  | 46.52 | 69.42     |
| 15     | -20.11 | 16.78  | 1.64  | 0.66  | 0.35  | 45.85 | 58.97     |
| 16     | -21.43 | 16.17  | 1.75  | 0.69  | 0.34  | 45.23 | 63.73     |
| 17     | -14.59 | 18.28  | 1.62  | 0.67  | 0.37  | 45.92 | 62.31     |
| 18     | -22.43 | 23.50  | 1.64  | 0.68  | 0.37  | 45.79 | 64.30     |
| 19     | -22.39 | 15.80  | 1.65  | 0.67  | 0.35  | 45.76 | 59.99     |
| 20     | -22.56 | 24.36  | 1.64  | 0.68  | 0.37  | 45.77 | 60.92     |
| Mean   | -20.06 | 13.50  | 1.27  | 0.59  | 0.27  | 58.06 | 100.00    |
| Median | -21.31 | 15.06  | 1.44  | 0.58  | 0.26  | 57.57 | 100.00    |
| Mode   | -21.22 | 14.89  | 1.54  | 0.59  | 0.25  | 56.82 | 100.00    |
| Proposed | -16.87 | 13.30  | 1.12  | 0.47  | 0.24  | 57.68 | 100.00    |
Figure 10 - Visualization of the influence of parameters $\alpha$, $\lambda$, $\xi$, and $\zeta$ in the fusion process. Parameters are set to $\alpha=1$, $\lambda=1$, $\xi=1$, $\zeta=1$, and then for each row one parameter is altered.

**Spaceborne Optical Data**

All four input DSMs are given in Figure 11. The holes in the DSMs mostly correspond to occluded regions, which result from off-nadir along and across track view angles. At first glance the DSMs look very similar. However, due to the varying acquisition angles the occluded regions are different (well visible next to the large building in the upper left corner). Also the trees in the park below that building are differently reconstructed and pose a challenge for the fusion process. Figure 12 gives 3D views of two input DSMs (a-b), the mean based fusion (c), the median based fusion (d), the probabilistic fusion (e), and the proposed fusion (f). In this example the local fusion methods are able to extract rather nice
fusion results, as the four input DSM are quite coherent and do not contain much clutter. The globally optimal DSM fusion generates a smoother solution, which can be seen as shiny surfaces in Figure 12, and it removes incorrect spikes. The residual error for each iteration is given in Figure 13. Again, FISTA converges faster than GD.

---

**Figure 11** - Spaceborne optical data set: All 4 input DSMs (includes material of Pléiades © CNES 2014, Distribution Airbus DS).

**Figure 12** - Results of the spaceborne optical data fusion. Shown are 3D views of (a-b) two input DSMs, (c) the mean based fusion, (d) the median based fusion, (e) the probabilistic fusion, and (f) the proposed fusion (includes material of Pléiades © CNES 2014, Distribution Airbus DS).
Figure 13 - Convergence of the proposed optimization schemes for spaceborne optical data.

**Spaceborne SAR Data**

Six of the 12 input DSMs are given in Figure 14. The denser DSMs stem from stereo pairs with a small intersection angle, while those with sparse height values are from pairs with large (probably too large) stereo intersection angles. Figure 15 gives 3D views of two input DSMs (a-b), the mean based fusion (c), the median based fusion (d), the probabilistic fusion (e), and the proposed fusion (f). Even though the input height values from the different DSMs are inconsistent, the proposed fusion is able to generate a more or less realistic surface of the given forest region. The local methods are not able to achieve relatively smooth solutions. The residual error for each iteration is given in Figure 16. Again, FISTA converges faster than GD.

Figure 14 - Spaceborne SAR data set: 6 of 12 input DSMs.
**Conclusions**

This In this work the mathematical formulation of a novel DSM fusion method was presented, that can be used to combine several DSMs extracted from airborne or spaceborne stereo images. The main novelty is the definition of a special convex energy functional with a $\beta$-Lipschitz continuous gradient that allows a trivial solution of the posed minimization problem. Because of this, the globally optimal solution of the optimization problem will always be found. The according mathematical proofs were derived within this work. In order to show the generality of the approach, experiments were based on two different
minimization schemes and were applied on airborne optical, on spaceborne optical and on spaceborne SAR images. The resulting fused DSMs were rich in detail and were of higher quality than results of local methods.

The comparison of the three local methods, namely mean or median fusion of each DSM cell and the probabilistic fusion incorporating a $3 \times 3$ pixel neighborhood, can be summarized as follows:

- First, simple mean or median fusion using only one DSM cell always leads to noisy DSMs;
- Second, the probabilistic fusion performs better than the simple mean or median fusion;
- Third, when the individual input DSMs are of rather good quality the probabilistic fusion achieves very good results, comparable to the presented method (e.g. for UltraCam or Pléiades based DSMs). With lower quality of the input (e.g. SAR radargrammetry based DSMs) no local method is able to reconstruct a rather realistic smooth surface like the proposed fusion does;
- Fourth, the proposed method easily allows to define the smoothness of the resulting fused DSM by altering the regularization parameters, which is not possible for the local methods.

In the future the following issues will be investigated in detail: First, it would be really nice to get test data with better ground truth height information such that the different fusion methods could be compared in terms of quantitative residual statistics. Second, it would be of interest to exchange the Huber norm with the so-called pseudo Huber norm [Hartley and Zisserman, 2004, p. 619] and evaluate the change. Third, according weights that define the influence of the individual DSMs, as discussed in the mathematical formulation section, will be considered. Fourth, a comparison to the total generalized variation method [Pock et al., 2011; Kuschk and d’Angelo, 2013] and to the multi-baseline fusion [Hirschmüller, 2008] is envisaged.

Acknowledgements
This research was partly funded by BMVIT/BMWFW under COMET programme, project nr. 836630, by “Land Steiermark” through SFG under project nr. 1000033937, and by the Vienna Business Agency. The authors would like to thank all institutions for providing the test data sets. In particular, we acknowledge the organizers of the EuroSDR benchmark providing the UltraCam images; GeoVille Information Systems GmbH, Austria for providing the Pléiades tri-stereo images; and Airbus Defence and Space, Geo-Intelligence, Germany for providing the TerraSAR-X Staring Spotlight images. Many thanks to Norbert Haala for discussions on the accuracy analysis.

References
Beck A., Teboulle M. (2009) - *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*. SIAM Journal on Imaging Sciences, 2 (1):183-202. doi: http://dx.doi.org/10.1137/080716542.

Chambolle A., Pock T. (2011) - *A first-order primal-dual algorithm for convex problems with applications to imaging*. Journal of Mathematical Imaging and Vision, 40 (1):120-
145. doi: http://dx.doi.org/10.1007/s10851-010-0251-1.

Combettes P.L., Pesquet J.-C. (2011) - Proximal splitting methods in signal processing. In: Fixed-point algorithms for inverse problems in science and engineering. Science and Engineering, Springer Optimization and its Applications, 49: 185-212 Springer New York.

Fuhrmann S., Goesele M. (2014) - Floating scale surface reconstruction. ACM Transactions on Graphics (TOG), 33 (4): 46. doi: http://dx.doi.org/10.1145/2601097.2601163.

Fukunaga K., Hostetler L.D. (1975) - The estimation of the gradient of a density function with applications in pattern recognition. In IEEE Transactions on Information Theory, 21 (1): 32-40. doi: http://dx.doi.org/10.1109/TIT.1975.1055330.

Graber G., Pock T., Bischof H. (2011) - Online 3D reconstruction using convex optimization. In IEEE International Conference on Computer Vision Workshops, pp. 708-711. doi: http://dx.doi.org/10.1109/iccvw.2011.6130318.

Gutjahr K., Perko R., Raggam H., Schardt M. (2014) - The epipolarity constraint in stereoradargrammetric DEM generation. IEEE Transactions on Geoscience and Remote Sensing, 52 (8): 5014-5022. doi: http://dx.doi.org/10.1109/TGRS.2013.2286409.

Gutjahr K., Perko R., Raggam H., Schardt M. (2015) - 3D-mapping from TerraSAR-X Staring Spotlight data - First results. In IEEE International Geoscience and Remote Sensing Symposium, pp. 1817-1820.

Haala N. (2013) - The landscape of dense image matching algorithms. In Photogrammetric Week, Stuttgart, pp. 271-284.

Haala N. (2014) - Dense image matching final report. EuroSDR Publication Series, (64):115-145.

Haala N., Hastedt H., Wolf K., Ressl C., Baltrusch S. (2010) - Digital photogrammetric camera evaluationgeneration of digital elevation models. Photogrammetrie-Fernerkundung-Geoinformation, 2010 (2): 99-115. doi: http://dx.doi.org/10.1127/1432-8364/2010/0043.

Harlow F.H., Welch J.E. (1965) - Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. Physics of fluids, 8 (12): 2182. doi: http://dx.doi.org/10.1063/1.1761178.

Hartley R., Zisserman A. (2004) - Multiple View Geometry in Computer Vision. Cambridge University Press, Second Edition. doi: http://dx.doi.org/10.1017/CBO9780511811685.

Heise P., Klose S., Jensen B., Knoll A. (2013) - PM-Huber: PatchMatch with Huber regularization for stereo matching. IEEE International Conference on Computer Vision, pp. 2360-2367. doi: http://dx.doi.org/10.1109/iccv.2013.293.

Hennig S.D., Koppe W., Gutjahr K., Perko R., Raggam H. (2015) - Evaluation of radargrammetry DEMs based on TerraSAR-X Staring Spotlight imagery. In: 35. Wissenschaftlich-Technische Jahrestagung der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation, 24: 1-8, Cologne, Germany.

Hirschmüller H. (2008) - Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30 (2): 328-341. doi: http://dx.doi.org/10.1109/TPAMI.2007.1166.

Huber P.J. (1973) - Robust regression: Asymptotics, conjectures and Monte Carlo. The Annals of Statistics, pp. 799-821. doi: http://dx.doi.org/10.1214/aos/1176342503.

Kluckner S., Pock T., Bischof H. (2010) - Exploiting redundancy for aerial image fusion.
using convex optimization. In Pattern Recognition, pp. 303-312. doi: http://dx.doi.org/10.1007/978-3-642-15986-2_31.

Kuhn A., Mayer H., Hirschmüller H., Scharstein D. (2014) - A TV prior for high-quality local multi-view stereo reconstruction. IEEE 2nd International Conference on 3D Vision, 1: 65-72. doi: http://dx.doi.org/10.1109/3dv.2014.76.

Kuschk G., d’Angelo P. (2013) - Fusion of multi-resolution digital surface models. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1 (3): 247-251. doi: http://dx.doi.org/10.5194/isprsarchives-XL-1-W3-247-2013.

Leys C., Ley C., Klein O., Bernard P., Licata L. (2013) - Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology, 49 (4): 764-766. doi: http://dx.doi.org/10.1016/j.jesp.2013.03.013.

Parikh N., Boyd S. (2013) - Proximal algorithms. Foundations and Trends in Optimization, 1 (3): 123-231.

Perko R., Raggam H., Gutjahr K., Schardt M. (2014) - Assessment of the mapping potential of Pléiades stereo and triplet data. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, II (3): 103-109, Zurich, Switzerland.

Pock T., Zebedin L., Bischof H. (2011) - Rainbow of computer science. chapter TGV-fusion, pp. 245-258, Springer-Verlag, Berlin, Heidelberg. doi: http://dx.doi.org/10.1007/978-3-642-19391-0_18.

Ranftl R., Bredies K., Pock T. (2014) - Non-local total generalized variation for optical flow estimation. In: European Conference on Computer Vision, II (3): 439-454, Zurich, Switzerland.

Rothermel M., Wenzel K., Fritsch D., Haala N. (2012) - SURE: Photogrammetric surface reconstruction from imagery. Proceedings LC3D Workshop, Berlin.

Rousseeuw P.J., Croux C. (1993) - Alternatives to the median absolute deviation. Journal of the American Statistical Association, 88 (424): 1273-1283. doi: http://dx.doi.org/10.1080/01621459.1993.10476408.

Rumpler M., Wendel A., Bischof H. (2013) - Probabilistic range image integration for DSM and true-orthophoto generation. In: Scandinavian Conference on Image Analysis, 7944: 533-544, Kämäräinen J.-K., Koskela M. (Eds.) Springer LNCS, Espoo, Finland. doi: http://dx.doi.org/10.1007/978-3-642-38886-6_50.

Unger M., Pock T., Werlberger M., Bischof H. (2010) - A convex approach for variational super-resolution. Lecture Notes in Computer Science, 6376: 313-322. doi: http://dx.doi.org/10.1007/978-3-642-15986-2_32.

Vandenbergh L. (2014) - EE236C - Optimization methods for large-scale systems. Lecture notes, 9. Available online at: http://www.seas.ucla.edu/~vandenbe/ee236c.html. (Accessed 8 July 2016).

Werlberger M., Pock T., Bischof H. (2010) - Motion estimation with non-local total variation regularization. Computer Vision and Pattern Recognition, pp. 2464-2471, San Francisco, CA, USA. doi: http://dx.doi.org/10.1109/cvpr.2010.5539945.

Werlberger M., Trobin W., Pock T., Wedel A., Cremers D., Bischof H. (2009) - Anisotropic Huber-L1 optical flow. Proceedings of the British Machine Vision Conference, London, UK. doi: http://dx.doi.org/10.5244/c.23.108.
Zach C., Pock T., Bischof H. (2007) - *A globally optimal algorithm for robust TV-L1 range image integration*. IEEE International Conference on Computer Vision, pp. 1-8.

© 2016 by the authors; licensee Italian Society of Remote Sensing (AIT). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).