GLOBALLY OPTIMAL DSM FUSION

Roland Perko$^{1}$ and Christopher Zach$^{2}$

1. Joanneum Research, DIGITAL, Graz, AT; roland.perko@joanneum.at
2. Toshiba Research Europe, Cambridge, UK; christopher.m.zach@gmail.com

ABSTRACT

This work presents the mathematical formulation of a novel globally optimal DSM fusion method that can be used to combine several 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. The experiments are based on two different minimization schemes and are applied on airborne optical, on spaceborne optical and on spaceborne SAR images. The resulting fused DSMs are rich in detail and are of higher quality than results of other local methods.

INTRODUCTION

When acquiring multiple images from different viewpoints over arbitrary objects, the standard procedure is to perform a pure stereo processing. This procedure yields one 2.5D digital surface model (DSM) for all possible stereo pairs. The individual DSMs have to be fused into one final DSM, which should have a better quality than the individual ones. In addition, undefined regions, which are not reconstructed in the single stereo pairs, e.g. due to occlusions, should be filled with height information from other pairs. This general issue of DSM fusion is omnipresent when it comes to 3D reconstruction from multiple view images. The mentioned principal philosophy is applied on very different imagery, e.g. on images from a hand held camera [3], on images from airborne cameras [9, 10], on remote sensing image from optical stereo satellites [7] or even on SAR satellite data [4].

Some solutions are based on local methods, like just taking the mean or median value over one DSM cell or the method by Rumpler et al. [10], which determines a probability density function (pdf) for a local 3x3 pixel neighborhood and extracts the mode of this pdf. Other solutions are based on global methods [12,8,3] which in general use non-convex energy functionals and are therefore rather difficult to solve. We present a global formulation for DSM fusion with a convex energy functional that is differentiable and has a $\beta$-Lipschitz continuous gradient. Therefore, the minimization problem can be cast on a trivial gradient descent (GD) optimizer (cf. [2]).

In this paper the mathematical formulation of the novel globally optimal DSM fusion methodology is described in detail. The convex energy functional consists of a regularization term and a data term. While the first forces a smooth surface including sharp 3D breaklines, the second drags the solution towards the given input DSMs. In addition, we compare the trivial GD optimizer with the fast iterative shrinkage-thresholding algorithm (FISTA) [1], 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 processed airborne data set (UltraCam), on a photogrammetric processed spaceborne data set (Pléiades) and on a radargrammetric processed SAR spaceborne data set (TerraSAR-X Staring Spotlight mode).

Beside this specific application the proposed methodology could also be used to fuse DSMs from different sources, e.g. to combine LiDAR height models with photogrammetric DSMs.

Even though the mathematical formulations seem complicated on the first glance, the resulting optimization algorithm can be implemented in a few lines of code. Thus, the formulation scores with its simplicity and its beauty.
MATHEMATICAL FORMULATION

As a first step we have to define an energy functional \( E(u) \) that should be minimized whereas \( u \) is the resulting fused DSM. Second, we have to proof that \( E(u) \) is a convex functional that is differentiable and has a \( \beta \)-Lipschitz continuous gradient. These are the prerequisites that a trivial 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 presented.

Formulation of the Energy Functional

We define the energy \( E(u) \) to be minimized in Eq. (1). The energy functional is composed 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 is used, since the Huber-norm is a convex approximation of the truncated quadratic and favors piecewise smooth (due to the regularization) and outlier robust (in the data term) solutions.

\[
E(u) = \int_{\Omega} \left( \alpha |\nabla u|_\xi + \frac{\lambda}{k} \sum_{i=1}^{k} |u - f_i|_\zeta \right) dx dy
\]

With \( u(x, y) \) the desired fused DSM, \( f_i(x, y) \) input DSMs, \( \nabla u \) partial derivatives of \( u \), \( \alpha \) total variation regularization weight, \( \lambda \) data fidelity weight, \( \Omega \) image plane, \( |x|_\gamma \) Huber norm of \( x \) w.r.t. \( \gamma \).

Convexity of the Energy Functional and Lipschitz Continuity of its Gradient

It can be shown that the functional \( E(u) \) is convex (since the Huber norm is a convex function, sums of convex functions are convex and the composition of convex function and linear term is convex (\( \nabla u \) is linear in \( u \))). With more effort it can also be shown that \( E(u) \) has a \( \beta \)-Lipschitz continuous gradient and that its respective \( \beta \) is at most \( 10 \max\{\alpha / \xi, \lambda / \zeta\} \). The according rather difficult proof will be given in an extended version of this paper.

Minimization Schemes for the Energy Functional

Using the findings above, the minimum of our function can be found employing a trivial gradient descent method (cf. [2]), as sketched below:

Algorithm 1 Gradient Descent (GD)

1: for \( n = 0, 1, \ldots \) do
2: \( x_{n+1} = x_n - \beta^{-1} \nabla E(x_n) \)
3: end for

There are many other methods to optimize a convex function (e.g. conjugate gradient, Euler-Lagrange, accelerated gradient), where we present the fast iterative shrinkage-thresholding algorithm (FISTA) [1], actually an accelerated gradient method, with the definition from [11]:

Algorithm 2 Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)

1: for \( n = 1, 2, \ldots \) do
2: \( y = x_{n-1} + \frac{n-1}{n} (x_{n-1} - x_{n-2}) \)
3: \( x_n = y - \beta^{-1} \nabla E(y) \)
4: end for

Since both GD and FISTA need a starting DSM, namely \( x_0 \), we take the median based fusion and interpolate the remaining invalid regions by means of linear hole filling.

The numerical discrete solution for calculating the partial derivatives \( \nabla E(u) \) can be derived as

\[
\frac{\partial E(u)}{\partial u_{x,y}} = \alpha \left( \frac{d[u_{x,y} - u_{x+1,y}]}{du_{x,y}} \right)_\xi + \frac{d[u_{x,y} - u_{x,y+1}]}{du_{x,y}} \right)_\xi + \frac{\lambda}{k} \sum_{i=1}^{k} \frac{d[u_{x,y} - f_i]}{du_{x,y}} \right)_\zeta
\]

(2)
TEST DATA

In order to show the generality of the proposed method, tests were conducted on four very different data sets:

- **Synthetic Data**: Constructed from a building block with roof shapes, fairly common in urban environments with 256x256 pixels and a height dynamic of 50 to 200 digital numbers (see Figure 1), similar to the one proposed in [8]. To simulate more realistic examples five input DSMs were generated by randomly introducing gross outliers (a common artefact from image matching) and additional Gaussian noise. Examples are also given in Figure 1.

- **Airborne Optical Data**: The data set Vaihingen in Germany from the EuroSDR benchmark on *High Density Image Matching for DSM Computation* consists of 36 UltraCam-X images with a ground sampling distance (GSD) of 20 cm. Using the photogrammetric workflow of JOANNEUM RESEARCH DSMs with 20 cm spacing were extracted for each stereo pair. For this test we chose a representative area of 451x401 pixels holding the castle Kaltenstein which is covered by 20 individual stereo DSMs with a height dynamic of 280 to 350 meters. Three of those input DSMs are shown in Figure 3.

- **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 [7], resulting in 4 input DSMs with 1 m spacing. A subset of 701x401 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 from descending orbit with three different look angles per orbit. For this specific processing multi look ground range detected (MGD) images, that represents the magnitude of the radar backscatter were employed with a GSD down to 16 cm. The radargrammetric processing suite available at JOANNEUM RESEARCH [4] was employed to extract the 12 individual stereo DSMs with 1 m spacing. A representative area of 1501x1001 pixels is chosen covering agricultural fields and a forest with a height dynamic of 250 to 380 meters. 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 [6,5].

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 $|E(u_n) - E(u_*)|/|E(u_*)|$ for each iteration $n$ (cf. [11]) with $u_*$ being the solution after 1000 iterations. In addition to this numerical evaluation several DSMs are shown in 3D, where always the version after 1000 iterations is given. Our results are visually compared to the median fusion, which is defined by taking the median from the pool of values for each DSM cell. All tests were performed using the same parameters of our function $E(u)$, which were chosen by tuning the fusion result of the airborne optical data set: $\alpha = 1, \lambda = 1, \xi = 10, \zeta = 0.1$.

Figure 1 shows that the proposed method is able to get rid of gross outliers and clearly outperforms the local approach. The plot of the residual errors (Figure 2) visualizes that FISTA converges much faster than GD and it could be stopped after 50 iterations. Same conclusions hold for the airborne data set in Figure 3 and 4. Again the proposed fusion is able to remove outliers, e.g. the spikes right of the tower and FISTA converges faster than GD. Figure 5 shows the results of fusing the spaceborne optical Pléiades DSMs. Since the input DSMs are of high quality the results of DSM fusion look similar. A closer look reveals that the proposed fusion results in a smoother more realistic DSM. In the SAR case (Figure 6) the proposed fusion is able to extract a

---

more or less realistic surface of the given forest region even though the input height values are inconsistent. The median fusion just reduces the noise but does not reconstruct the forest.

![Figure 1: Synthetic data set: Top row: The original and 2 out of 5 noisy versions of the building block. Bottom row: 3D views of the noise free input DSM, median fusion and the proposed fusion.](image)

![Figure 2: Convergence of the proposed optimization schemes for synthetic data.](image)

![Figure 3: Airborne optical data set: Top row: 3 of 20 input DSMs. Bottom row: 3D views of one input DSM, median fusion and the proposed fusion.](image)
CONCLUSIONS

In this work the mathematical formulation of a novel globally optimal DSM fusion method was presented, that can be used to combine several DSMs extracted from airborne or spaceborne stereo images. The main novelty was the definition of a special convex energy functional with a $\beta$-Lipschitz continuous gradient that allowed a trivial solution of the posed minimization problem. 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 other local methods.

When comparing the proposed results to the local median fusion methods the following became obvious: First, simple median fusion using only one DSM cell always leads to noisy DSMs. Second, when the individual input DSMs are of rather good quality the median fusion is able to achieve acceptable results, comparable to the proposed method (e.g. for Pléiades based DSMs). With lower quality of the input (e.g. SAR radargrammetry based DSMs) the local method is not able to reconstruct a realistic smooth surface like the proposed fusion does. Third, the proposed
method easily allows to define the smoothness of the resulting fused DSM by altering the regularization parameters, which is not possible by employing the local method.

In future the following issues will be investigated in detail: First, it would be really nice to get test data with 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 and evaluate the change due to this alteration. Third, according weights that define the influence of the individual DSMs will be incorporated in the energy functional. Fourth, a detail analysis of the influence of the parameters $\alpha, \lambda, \xi$ and $\zeta$ will be conducted.

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 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.

REFERENCES