One-Two-Pixel Multi-View Image Matching for digital surface modelling

E. Rupnik
M. Pierrot Deseilligny

1 LaSTIG, ENSG, IGN

ewelina.rupnik@ign.fr; marc.pierrot-deseilligny@ensg.eu

Abstract
Generation of photogrammetric Digital Surface Models is typically done with global or semi-global methods. The objective function that minimizes the cost comprises of a term corresponding to likelihood of two pixels having a certain depth (the so-called data term) and a penalty. The state-of-the-art approaches customarily calculate the likelihood on a support domain around a pixel, therefore decrease the precision of the 3D reconstruction by assuming surface planarity within a region. This publication formulates a novel data term as a combination of a single pixel and a two-pixel multi-view similarity measure. We prove superiority of the algorithm by evaluating it on two satellite datasets and an aerial acquisition.

Keywords
dense image matching; multi-view; stereo; cost function; photo-consistency measure;

1 Introduction
Photogrammetrically derived Digital Surface Models (DSM) are typically found by solving an energy functional with global, semi-global (SGM) or local methods. Classically, local methods calculate the best correspondences in three steps: (i) matching cost calculation, (ii) support domain cost aggregation, and (iii) the depth selection by the Winner-Takes-All [YK05], or the belief propagation schemes [BRR11, GLS15]. The local character of the methods makes them computationally efficient, often at the expense of lower 3D precision and completeness. Whether global or SGM approach, the best correspondences are found by minimizing a cost function that is a sum of the data and the regularity term. The data term reflects the likelihood of a disparity (or a depth) to take a certain value, and is expressed as a function of similarity between potentially corresponding pixels in overlapping images. The regularity term implements a shape prior that favours locally similar disparities.

Global methods minimize the energy defined over all pixels in a MRF graph with techniques such as $\alpha$-expansion moves [BVZ01, TMSN17] or message passing [FH06]. Alternatively, definition over a continuous space and sub-pixelar reconstruction precision is is possible with variational methods [RGPB12]. Global minimization is NP-hard, and does not allow for concave regularity terms. The first poses memory issues for large scale reconstructions, while the latter inhibits faithful reconstruction at discontinuities (e.g. buildings). The SGM methods are solved with multi-directional dynamic programming techniques, impose no constraint on the regularity term and the optimization is resolved along independent lines of pixels [PDP06, Hir08]. Therefore, the computational times remain reasonable as the image size grows. Since cost function is cumulative along a number of directions, the found optimal solution is close to the global solution.

Independently of the algorithm, there exist many different measures to describe the similarity (aka photo-consistency) between pixel correspondences. The very first and most straightforward measure compares the pixel intensity values [KKYO95, BT98]. In real-world conditions (i.e. non-Lambertian surfaces, radiometric differences due to illumination change), however, pixel intensity differences turn to be unreliable for matching. In practice, photo-consistency measures defined over a support domain (square regions ordinarily of 3x3, 5x5, etc. size) are preferred [FH15].

Résumé
Le calcul photogrammétrique de modèle numérique de surface s’effectue en général avec une approche globale ou semi-globale. La fonction d’énergie à minimiser intègre un terme d’attache aux données, représentant la vraisemblance que deux pixels soient homologues, et un terme de régularisation qui pénalise les variations du relief et représente l’a-priori. L’état de l’art calcule généralement la vraisemblance sur une région autour de chaque pixel en faisant l’hypothèse que le relief est localement constant, ce qui diminue la précision de reconstruction. Cet article présente une nouvelle formulation de la mesure de similarité qui est la combinaison d’un terme "simple pixel" et d’un terme "deux pixel" qui est calculé en même temps que la régularisation. Nous illustrons l’intérêt de cette méthode en l’évaluant sur de scènes issues d’images satellites et aériennes.

Mots Clef
l’appariement dense ; multi-image ; stereo ; fonction de couts ; mesure de la similarité ;
e.g. zero-mean normalized cross-correlation (NCC). Census, Rank [ZW93], Mutual Information [V97], or 3D planar patches [BRR]. Including more context within the per-pixel matching is advantageous for it increases the uniqueness and the robustness of the matched points, as well as attenuates the noise (caused by the low signal-to-noise ratio of the sensor). The disadvantage is that it privileges fronto-parallel planar surfaces, hence precludes accurate reconstructions of fine 3D structures. The objective of the research presented in this publication was to conceive an enhanced precision dense image matching approach free from the support domain artefacts, while keeping the computational burden "light". Accordingly, the proposed One-Two-Pixel matching is embedded in a SGM optimization framework and the increased precision is accomplished by formulating:

- a robust one pixel multi-view photo-consistency measure, and
- an additional two pixel term.

In section 2 theoretical background of the employed method is laid down, and in section 3 experimental results on two satellites as well as an aerial acquisition are shown.

2 One-two pixel matching

With the objective of excelling the spatial resolution of reconstructed depths, we propose a reformulated energy functional [1].

\[ C(S) = \sum_{k=1}^{N} C^I(e^k_{S(k)}) + C^C(e^k_{S(k)}, e^{k+1}_{S(k+1)}) \]

(1)

where \( N \) is the number of positions (e.g. the sum of all planimetric coordinates), \( e^k_{S(k)} \) corresponds to a candidate depth (i.e. a state) at position \( k \), and \( e^{k+1}_{S(k+1)} \) a candidate depth at position \( k + 1 \), respectively. \( C(S) \) is the minimum cost of the optimal depths set \( S \), calculated across all positions. The minimum solution is found in the following three steps: cost assignment (see Alg. [1]), cost aggregation (see Alg. [2]) and the optimal depth selection. The described methods are implemented in MicMac – the free opensource software for photogrammetry [RDPD].

The data term \( C^I \) is the cost of assigning a particular state to the position \( k \). It is implemented as a truly one-pixel measure which communicates the dissimilarity of two pixels in multiple views (\( NbV \) [2]). It is expressed as a mean of all intensity ratios between the master \( I_0 \) and the slaves, radiometrically calibrated with \( \delta_e \) and weighted by the \( Pds_{pix} \). The radiometric calibration is a function of intensity ratios between the master image and a slave image (see Alg. [1] and Fig. [1]). It is applied to handle the possible global illumination changes caused by a changed viewpoint or changing outside lighting conditions. The value of the function can be pixel-dependent – if non-constant radiometric corrections across the images are necessary (e.g. specular reflections) – or take a global value per an image. When a pixel-dependent case is chosen, the ratios shall be first smoothed by simple averaging or by 2D polynomial function fit. The role of \( Pds_{pix} \) is to control the relative influence of the data term in the cost function. In all our tests we set the value to 1.0.

\[ C^I(e^k_{S(k)}) = \frac{1}{N b V - 1} \cdot Pds_{pix} \cdot \sum_{v=1}^{NbV-1} \delta_e \cdot \frac{I_v(e^k_{S(k)})}{I_0(e^k_{S(k)})} \]

(2)

The new two-pixel term \( C^C \) encodes the dissimilarity of the two-pixel window across multiple views (\( NbV \)) [3]. Ideally, the advantages of the minimum window are twofold: (i) it is more apt to reconstruct fine 3D scene structure than the usual square correlation window, and (ii) is less susceptible to noisy data than as if the single pixel term only was used. For a particular depth assignment, the \( C^C \) cost will depend on the differences of normalised intensity ratios \( r \) at position \( k \) and \( k + 1 \). Hence, within the cost structure, the \( C^C \) costs associate with the arcs (in analogy to \( C^T \)), rather than with the nodes as is the case of \( C^I \) (see Alg. [2] and Fig. [2]). The two-pixel term should not be confused with the regularity term which applies "blind" penalization, without considering the pixel intensity values. \( Pds_{cross} \) plays the same role as \( Pds_{pix} \) and in our tests it was also set to 1.0.

\[ C^C(e^k_j, e^{k+1}_j) = ||r_{e^k_j} - r_{e^{k+1}_j}|| \cdot Pds_{cross} \cdot \frac{1}{N b V - 1} \]

(3)

\[ r_{e^k_j} = I_{0,e^k_j}^{-1} \cdot I_{v,e^k_j} \cdot v \in \{1, NbV\} \]

(4)

The regularizing term \( C^T \) is the cost of transition between two states of the most immediate neighbours \( k \) and \( k + 1 \) (the regularizing term), i.e. at the edges of the cost structure (5) (see Alg. [2] and Fig. [2]). It implements penalization of large disparities. To model sudden jumps at surface discontinuities we employ a concave function controlled by a cost threshold. Discussion of the form of this function is out of the scope of this publication.
Algorithm 1 Cost assignment

\[ V_{Im}[v].ImOrth(X,Y) - \text{intensity in image } v \text{ at position } (X,Y) \]; all images are orthorectified to the geometry of the master

\[
\text{ratio}(\text{int},\text{int}) - \text{calculate ratios} \\
\text{norm}(\text{int}) - \text{normalise to [0-255]} \\
\delta - \text{radiometric calibration} \\
\text{SurfOpt} - \text{optimiser}
\]

\[
\text{Assignment} \\
\text{for } Z = Z_{\text{min}}; Z < Z_{\text{max}}; Z + + \text{ do} \\
\quad \text{for } X = 0; X < Sz.x; X + + \text{ do} \\
\quad \quad \text{for } Y = 0; Y < Sz.y; Y + + \text{ do} \\
\quad \quad \quad V_0 = V_{Im}[0].ImOrth(X,Y) \\
\quad \quad \quad C^I = 0 \\
\quad \quad \quad \text{for } v = 0; v < NbV; v + + \text{ do} \\
\quad \quad \quad \quad V_k = V_{Im}[v].ImOrth(X,Y) \\
\quad \quad \quad \quad Val = \text{ratio}(V_0, V_k) \\
\quad \quad \quad \quad r.push\_back(Val) \\
\quad \quad \quad V_k^{\text{cor}} = \delta \cdot V_k \\
\quad \quad \quad Val^{cor} = \text{ratio}(V_0, V_k^{\text{cor}}) \\
\quad \quad \quad C^I + = \text{Abs}(Val^{cor}) \\
\quad \quad \text{end for} \\
\quad \text{end for} \\
\quad \text{end for} \\
\text{end for}
\]

\[
\text{Communicate the vector of ratios and the data term to the optimiser} \\
\text{SurfOpt}(X,Y,Z) = r \\
\text{SurfOpt}(X,Y,Z) = C^I
\]

3 Experiments

The new dense matching scheme was tested on two satellite and one aerial acquisitions. Qualitative evaluation in form of gray shaded DSMs is provided. Comparisons with the classical SGM (NCC as a similarity measure calculated on a 3 x 3 window), and ground truth data (if available) are also given.

SPOT-7 dataset is a tri-stereo acquired over a rural, densely forested area with the Ground Sampling Distance (GSD) of 1.5m. The results are presented in Figs. 3, 4. Comparison of the One-Two-Pixel matching with the classical SGM result shows significant gains in surface details, visible especially in the vegetated zones. To compensate for the illumination change due to different acquisition viewpoints, radiometric calibration was employed.

To calculate the global correction value, each individual image was first resampled to the terrain (i.e. orthogonal) geometry. Then, the orthoimages were divided by the orthoimage of the master image which resulted in an image per a stereo couple. The ratio image was then blurred and a mean across all pixels was assigned to the \( \delta_v \). A global shift in intensity sufficed to model the illumination changes. No ground truth data was available.

WorldView-3 dataset is a tri-stereo acquired over an urban area with the GSD of 0.3m. The images and the ground truth LiDAR point cloud (20cm raster) are part of the...
Algorithm 2 Cost aggregation

\begin{verbatim}
for each direction + do
    for each direction - do
        for $X = 0; X < Sz.x; X++$ do
            for $Y = 0; Y < Sz.y; Y++$ do
                for $DZ = Z_{\text{min}}; DZ < Z_{\text{max}}; DZ++$ do
                    \[ \text{Cost} = \text{CostOfDZ}(DZ) \]
                    $r_k = \text{SurfOpt.rOfk}(k)$
                    $r_{k+1} = \text{SurfOpt.rOfk}(k + 1)$
                    \[ \text{Cost} += \text{TwoPixCost}(r_k, r_{k+1}) \]
            \end{verbatim}
\end{verbatim}

\begin{verbatim}
end for
end for
end for
end for
\end{verbatim}

\begin{itemize}
\item \textbf{forward}
\item \textbf{backward}
\item \textbf{get the value of concave cost function at } $DZ$
\item \textbf{recover the vectors of ratios}
\item \textbf{calculate the two-pixel cost and add to cost}
\item \textbf{communicate the two-pixel cost to the optimiser}
\end{itemize}

\textit{FIGURE 3} – SPOT-7, DSM gray shading results. (a) the region of interest in RGB; (b) SGM with NCC; (c) One-Two-Pixel matching w/o radiometric calibration; (d) One-Two-Pixel matching with radiometric calibration.

\textit{FIGURE 4} – Zoom on SPOT-7 results in Fig. 3. (a) the region of interest; (b) SGM with NCC; (c) One-Two-Pixel matching w/o radiometric calibration; (d) One-Two-Pixel matching with radiometric calibration.
IARPA’s MVS benchmark and were made available publicly within the Multi-View Stereo 3D Mapping Challenge [Sat16a, Sat16b]. Visual inspection of the results presented in Figs. 5, 6 reveal good performance of the One-Two-Pixel matching as far as restitution of discontinuities and the level of noise are concerned. As no illumination artefacts were observed, the calculation was carried out without the radiometric calibration.

Vaihingen is a set of four aerial images acquired with the UltraCam-X camera at 0.2m GSD. The images are part of the EuroSDR Matching benchmark [Haa14]. The available ground truth data is a surface that is a median of several DSMs furnished by all the participants of the benchmark. The results are presented in Figs. 7, 8, 9. Similarly to the previous examples on satellite imagery, the One-Two-Pixel matching proves capable of recovering fine 3D details both on vegetation and at discontinuities. The calculation was carried out without the radiometric calibration.

4 Conclusions

We proposed a new One-Two-Pixel multi-view image matching method. It is a variant of the SGM where the data term – typically calculated on a support domain (i.e. square window) – is replaced by a multi-view single pixel similarity measure, and a two-pixel window. The introduced modification permits to overcome the commonplace shortcomings of the state-of-the-art approaches such as the bias towards reconstructing fronto-parallel facets or the smoothing artefacts at surface discontinuities. The conducted experiments confirm the superiority of One-Two-Pixel with respect to the classical SGM.

5 Acknowledgement

The authors would like to thank Tristan Faure and Antoine Levenant from IGN DIL department for providing the SPOT-7 imagery.

Références


[Sat16a] Satellite Benchmark JHU/APL. Commercial satellite imagery in the mvs benchmark data set was provided courtesy of digitalglobe. 2016.


Figure 5 – WorldView-3, DSM gray shading results. (Left) SGM with NCC; (middle) One-Two-Pixel matching; (right) LiDAR ground truth.

Figure 6 – Zoom on WorldView-3 results in Fig. 5 (Left) SGM with NCC; (middle) One-Two-Pixel matching; (right) LiDAR ground truth.

Figure 7 – Vaihingen, UltraCAM-X, DSM gray shading results. (a) the region of interest in RGB; (b) EuroSDR ground truth; (c) SGM with NCC; (d) One-Two-Pixel matching.
**Figure 8** – Zoom1 on Vaihingen in Fig. 7. (a) the region of interest in RGB; (b) EuroSDR ground truth; (c) SGM with NCC; (d) One-Two-Pixel matching.

**Figure 9** – Zoom2 on Vaihingen in Fig. 7. (a) the region of interest in RGB; (b) EuroSDR ground truth; (c) SGM with NCC; (d) One-Two-Pixel matching.

