Fast global stereo matching via energy pyramid minimization

B Conejo, S Leprince, F Ayoub, Jean-Philippe Avouac

▶ **To cite this version:**

B Conejo, S Leprince, F Ayoub, Jean-Philippe Avouac. Fast global stereo matching via energy pyramid minimization. Photogrammetric Computer Vision - PCV2014, ISPRS Technical Commission III Midterm Symposium, Sep 2014, Zurich, Switzerland. pp.41 - 48, 10.5194/isprsannals-II-3-41-2014 . hal-01081593

HAL Id: hal-01081593
https://enpc.hal.science/hal-01081593

Submitted on 10 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FAST GLOBAL STEREO MATCHING VIA ENERGY PYRAMID MINIMIZATION

B. CONEJO [a, b*], S. LEPRINCE [a], F. AYOUB [a], J. P. AVOUAC [a]

[a] GPS Division, California Institute of Technology, Pasadena, CA, USA (bconejo@, leprincs@, fayoub@gps., avouac@gps.)caltech.edu
[b] IMAGINE Lab., Universit Paris-Est, LIGM (UMR CNRS 8049), ENPC, F-77455 Marne-la-Valle, France

KEY WORDS: Photogrammetry, Global, Matching, Multiresolution, Algorithms, Estimation, DEM/DTM

ABSTRACT:

We define a global matching framework based on energy pyramid, the Global Matching via Energy Pyramid (GM-EP) algorithm, which estimates the disparity map from a single stereo-pair by solving an energy minimization problem. We efficiently address this minimization by globally optimizing a coarse to fine sequence of sparse Conditional Random Fields (CRF) directly defined on the energy. This global discrete optimization approach guarantees that at each scale we obtain a near optimal solution, and we demonstrate its superiority over state of the art image pyramid approaches through application to real stereo-pairs. We conclude that multiscale approaches should be build on energy pyramids rather than on image pyramids.

1. INTRODUCTION

Precise Digital Surface Models (DSM) are widely employed in urban monitoring, geological surveys, architecture, or archeology. DSM are now mostly generated using remote sensing surveys based on optical stereo-imaging, interferometric Synthetic Aperture Radar (SAR), or Light Detection And Ranging (LiDAR) acquisitions (Li et al., 2005). In this paper, we focus on optical stereo-imaging.

The volume of available optical aerial and satellite images has sky-rocketed during this last decade. Moreover, the resolution and the size of these images have also vastly increased, and it is now common for satellites with push-broom sensors to produce images of more than 35,000 $\times$ 30,000 pixels with resolution up to 50 cm Ground Sampling Distance (GSD) (Worldview, 2014). Aerial imaging with frame camera achieve resolution better than 10 cm GSD with 20,000 $\times$ 13,000 pixels per images (UtraCam, 2014).

DSM with impressive accuracy and resolution can now been computed using multi-stereo pairs (Acute3D, 2014, 123DCatch, 2014, UtraMap, 2014, Pix4d, 2014, Micmac, 2014), such techniques are based on the redundancy of view points as one ground point is viewed by up to hundred of images. We work in a slightly different context where only one stereo-pair is available, which is typical of satellite imagery.

DSM computation from stereo pairs involves a number of steps described for instance in (Hartley and Zisserman, 2004): calibration, aerotriangulation, eventually rectification, and then the estimation of the disparity map; finally, the DSM is computed from the geometry of acquisition and the estimated disparity map. In this paper, we only focus on the estimation of the disparity map in epipolar geometry (Zhang, 1998).

Numerous methods have been designed to estimate disparity maps e.g. (Scharstein et al., 2001, Brown et al., 2003, Ansar et al., 2004, Tombari et al., 2008). State of the art approaches estimate disparity maps by defining an energy to minimize, which commonly enforces a notion of matching and a notion of regularity on the disparity maps. To speed up the computation, the photogrametry community has developed semi-global matching approaches (Hirschmuller, 2005, Pierrot-deseilligny and Paparoditis, 2006) that compute local optimum solutions along different directions and aggregate them together. However, no mathematical guarantee has been given on the aggregation of local solutions to form a global optimum. Under different constraints, the computer vision community has developed more advanced techniques to globally optimize the matching problem (Szeliski et al., 2008, Kappes et al., 2013). These techniques called global matching are mathematically more sound than semi-global matching as they guarantee a near global optimum (Boykov et al., 2001). However, due to their computational complexity, they have only been applied to small images, i.e., less than 1000 $\times$ 1000 pixels, as proof of concept (Middleburry, 2014, Klaus et al., 2006, Kolmogorov and Zabih, 2001) and are not yet scalable to accommodate the large sizes of remote sensing images. Very recently, the discrete optimization community has started to show some interest for variable grouping as a technique to solve global optimization more efficiently, (Bagon and Galun, 2012, Komodakis, 2010, Kim et al., 2011). However, only results from (Middleburry, 2014) have been presented for stereo imagery.

Our contribution is to bring the global matching techniques up to the dimensionality of remote sensing data and to offer the photogrammetry community a sound mathematical framework both in terms of modeling and optimization. In this paper, we describe how to estimate disparity maps based on energy minimization. Then, we propose an algorithm to efficiently solve such optimization problems with a multiscale approach based on an energy pyramid rather than the traditional image pyramid. Finally, we demonstrate the improvements of our approach through an application to real stereo acquisitions.

2. THE STEREO MATCHING PROBLEM

2.1 Probability formulation

Let I_r be a reference image and $G = (\mathcal{V}, \mathcal{E})$ its associated graph. The set of nodes $\mathcal{V}$ consists of the pixels of I_r and the set of edges $\mathcal{E}$ is defined by the 4 connectivity as illustrated in Fig. 1. Let I_t be the target image.

Given I_r and I_t, we need to find the most probable 2D deformation d which associates each pixel of I_r to a pixel of I_t with d being a function of $p \in \mathcal{V} \rightarrow d(p) \in (R \times R)$. Thus, d lives in

*Corresponding author.

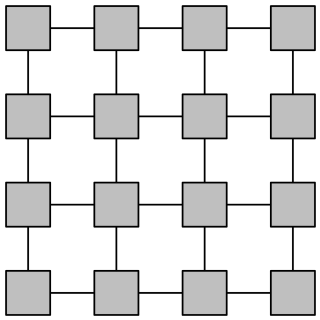


Figure 1: Graph $G = (\mathcal{V}, \mathcal{E})$ of a 4 by 4 image with a 4 connectivity.

$D = (R \times R)^{card(\mathcal{V})}$. We measure how a given d fits the data I_r and I_t by defining:

$$P(d|I_r, I_t). \tag{1}$$

The definition of P is context dependent, but most approaches enforce: (1) a notion of the similarity between I_r and $I_t \circ (id+d)$ (where id refers to the identity operator) by expressing $P_M(d|I_r, I_t)$ and; (2) a notion of regularity for d by expressing $P_R(d|I_r)$. If we suppose that these two probabilities are independent, we can write:

$$P(d|I_r, I_t) = P_M(d|I_r, I_t)P_R(d|I_r). \tag{2}$$

Instead of directly working with probabilities, we prefer using the energy domain as it is easier to define a measure on images. One can simply relate probability density function to energy via the Gibbs measure:

$$P(X = x) = \frac{1}{Z}\exp(-E(x)), \tag{3}$$

With Z being a normalization factor so that the integral of the probability function equal to 1.

2.2 Energy formulation

Through Eq. 3, we relate E, E_M, and E_R to P, P_M, and P_R respectively, which gives the following energy:

$$E(d) = \sum_{p \in \mathcal{V}} E_M(d, p) + \sum_{pq \in \mathcal{E}} E_R(d, p, q). \tag{4}$$

We define a pixel-wise similarity measure based on the similarity function ρ. Commonly used similarity functions are L1 or L2 norms (Birchfield and Tomasi, 1998), Normalized Cross Correlation (NCC) or Zero Normalized Cross Correlation (ZNCC) (Brown, 1992), and the different versions of the Mutual Information (Viola and Wells, 1995, Kim et al., 2003, Hirschmuller, 2005). In any case, the matching energy of a pixel p is defined as:

$$E_M(d, p) = \rho(I_r, I_t \circ (id + d(p))). \tag{5}$$

If the similarity measure is defined on a patch, we apply a rigid translation to the patch. Here, we use the ZNCC coefficient as it is robust to changes of illumination and contrast between I_r and I_t that appear due to specular objects or different acquisition times.

To enforce the regularity of d we choose to penalize the L1-norm of its discretized gradient, modulated by a weight function w. For each edge $pq \in \mathcal{E}$:

$$E_R(d, p, q) = w(p, q)\|d(p) - d(q)\|_1, \tag{6}$$

With :

$$w(p, q) = \lambda_1 + \lambda_2 \exp\left(-\frac{\|I_r(p) - I_r(q)\|^2}{\sigma^2}\right). \tag{7}$$

λ_1, λ_2, and σ are scalar parameters. The L1-norm of the gradient naturally enforces piece-wise constant disparities. The weight function $w(p, q)$ relaxes the regularization on radiometric discontinuities of the reference image as in (Gamble and Poggio, 1987, Boykov et al., 2001). This is an effective heuristic as most of the edges of the disparity maps are also edges of the image I_r. Alternatively, we could use the L1-norm of the laplacian rather than the gradient but this would lead to optimizing second order Conditional Random Fields that despite recent progress in solvers remain intractable in the context of this study (Komodakis and Paragios, 2009, Ishikawa, 2011, Fix et al., 2011).

2.3 Discrete Conditional Random Fields

We have built the measure E_M of how d fits our data and the measure E_R of how d respects the *a priori* knowledge on the disparity map. However, we still need to find the most probable disparity map d^*, i.e., the maximum a posteriori, by minimizing Eq. 8 over $d \in D$:

$$d^* = \min_{d \in D} E(d). \tag{8}$$

Finding d^* means optimizing a continuous Conditional Random Field (CRF), continuous because d belongs to the continuous space D and conditional because P_R depends on the data, i.e., I_r. This task is extremely difficult mainly because d is continuous and non-convex as Fig. 2 illustrates. Instead, we discretize D for each nodes $p \in \mathcal{V}$ to the discrete set $\mathcal{D}_p$, so that we now deal with a discrete CRF that is much easier to solve, with d living in $(\mathcal{D}_p)_p$.

We use the vocabulary of the discrete optimization community and we introduce the notion of graph, unary term, edge cost, distance function, and label set for a CRF.

Our graph is directly the one of the image I_r, i.e., $G = (\mathcal{V}, \mathcal{E})$.

We introduce $\mathcal{D} = \bigcup_{p \in \mathcal{V}} \mathcal{D}_p$, the union of all tested disparities. To each disparity of $\mathcal{D}$ is associated a label, i.e. an index, in the label space $\mathcal{L}$. Hence, each node p has a different label space $\mathcal{L}_p \subset \mathcal{L}$ that relates to the disparity set $\mathcal{D}_p$.

To each node $p \in \mathcal{V}$ and each label $l \in \mathcal{L}$, we associate a unary term corresponding to Eq. 5:

$$c_p(l) = \begin{cases} \rho(I_r, I_t \circ (id + \mathcal{D}(l))) & \text{if } l \in \mathcal{L}_p \\ +\infty & \text{otherwise} \end{cases} \tag{9}$$

If $l \notin \mathcal{L}_p$ the configuration is impossible, and we associate an infinite unary cost.

To each edge $pq \in \mathcal{E}$, we associate an edge cost :

$$w_{pq} = w(p, q). \tag{10}$$

To each pair of labels $(l_1, l_2) \in \mathcal{L}$ we associate a distance function:

$$\delta(l_1, l_2) = \|\mathcal{D}(l_1) - \mathcal{D}(l_2)\|_1. \tag{11}$$

For the sake of completeness, each edge $pq \in \mathcal{E}$ and each pair of labels $(l_p, l_q) \in \mathcal{L}$ defines a pairwise term.

$$pw(pq, l_p, l_q) = w_{pq}\delta(l_p, l_q). \tag{12}$$

We finally optimize the energy of the following discrete CRF:

$$E_{CRF} = \min_{l \in \mathcal{L}} \sum_{p \in \mathcal{V}} c_p(l_p) + \sum_{pq \in \mathcal{E}} w_{pq} \delta(l_p, l_q). \quad (13)$$

After optimizing Eq. 13 we obtain a labelling l^* that relates to the disparity map d^*. For each nodes $p \in \mathcal{V}$, the following holds:

$$d^*(p) = \mathcal{D}(l^*(p)). \quad (14)$$

We note $CRF^1 = [G, c, w, \delta]$ the CRF computed from I_r and I_t and defined by Eq. 13. CRF^1 belongs to the class of first order CRF because for each edges $pq \in \mathcal{E}$, the distance function δ only depends on the two labels l_p and l_q. The size of CRF^1 depends on: (1) its spatial component, i.e., the number of nodes $\mathcal{V}$ and the number of edges $\mathcal{E}$; and (2) its label component i.e., the number of label that relates in our case to the number of disparities per node p to evaluate.

The complexity to solve a first order CRF depends on the relative contribution to the energy of the unary terms and the pairwise terms. The higher the contribution of the pairwise terms, the more complex is the CRF optimization. The nature of the distance functions δ is also important. If δ is issued from a metric function, as in our case, then we have mathematical guarantee to retrieve a solution close to the global optimum while optimizing the CRF (Boykov et al., 2001, Komodakis and Tziritas, 2007).

During the last decade numerous advances have been achieved to optimize this problem (Geman and Geman, 1984, Felzenszwalb and Huttenlocher, 2004, Kolmogorov, 2006, Boykov et al., 2001, Komodakis et al., 2007a, Komodakis, 2010). Our attention was drawn to Fast-PD, (Komodakis and Tziritas, 2007, Komodakis et al., 2007b) for two different reasons: (1) it is the fastest algorithm currently available (Kappes et al., 2013); (2) it has the ability to use an input as initialization. This last property is extremely interesting as our multi-scale approach benefits from a hot-start of the CRF optimization, using the solution found at a previous scale.

3. MULTI-SCALE SCHEME AND SPARSE CRF

3.1 The multi-scale scheme

In our remote sensing context images and disparity to recover are large. Directly minimizing Eq. 13 would be inefficient both in terms of memory consumption and computational speed because the needed discretized disparity space $\mathcal{D}$ will be very large. Instead, we propose a multi-scale approach to efficiently explore the solution space D with a succession of discretized disparity spaces $\mathcal{D}$.

A multi-scale approach is valid because of the particular structure of the disparity map to retrieve. Indeed, while the full disparity range is large at the image scale, one can notice that locally the range is only a mere fraction of the full range. Natural topography is the main contributor to widening the range of disparities, which has a relatively low spatial frequency at the resolution we work at. Moreover, the local range of disparities is mainly due to man-made objects such as buildings, which tend to have higher spatial frequencies than natural features. Hence, the largest low spatial frequency disparities are resolved at the coarsest scales, while the smallest high spatial frequency disparities are resolved at the finest scales.

Our multiscale approach builds coarse to fine sequences of CRF, $CRF^n, ..., CRF^f, ..., CRF^1$, with CRF^1 being defined at the full resolution of the images I_r and I_t, and CRF^f being defined at a downsampling factor f. Our multiscale approach reduces the size of the CRF on both the spatial and label components. To build CRF^f from CRF^1 we can use two different approaches: (1) image pyramid or (2) energy pyramid. We compare the two approaches in 4.1.

For both pyramid approaches, a crucial step is to define the discretized disparity space $(\mathcal{D}_p)_p$ by adequate sampling of D. The key property lays in noticing that due to the interpolation kernel used for computing $I_t \circ (id + d(p))$, the values of the unary terms are continuous with respect to $d(p)$ and have a limited frequency support. Figure 2 represents typical values of unary terms issued from ZNCC coefficients with a very fine sampling along one component of D. Ideally, we would like to sample D such that the unknown minimum of Eq. 8, d^*, is included in the samples. However, we have no prior knowledge of d^*. Therefore, we suppose the image well-sampled and we choose to respect the Shanon sampling theorem. As the ZNCC coefficient is defined on a patch, this doubles the frequency support of the unary terms compare to the frequency support of the images I_r and I_t. Thus, we sample D to half a pixel as in Fig. 2.

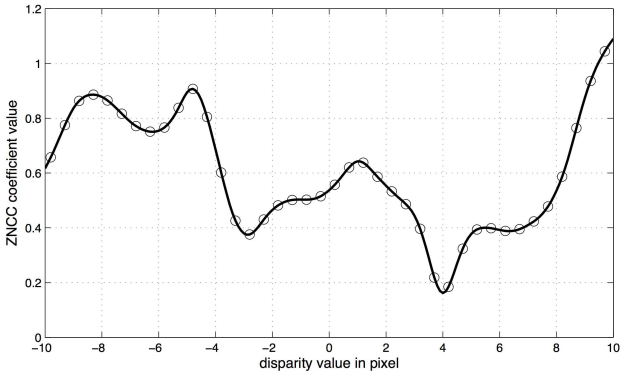


Figure 2: Illustration of the non convexity of unary terms with a 0.5 pixel discretization step to respect the Shanon sampling theorem.

We introduce the notion of energy manifold as the hyper-surface formed by the set of unary terms. The energy manifold lives in a 4 dimensional space. One dimension is linked to each spatial component of the unary terms. One dimension is linked to each disparity component of the unary terms.

3.2 Image pyramid: the GM-IP algorithm

The Global Matching with Image Pyramid (GM-IP) algorithm is an image pyramid approach based on downsampling the images to reduce both components of CRF^1. First, we downsample I_r and I_t to I_r^f and I_t^f. This directly reduces the spatial component of the CRF and it smooths the energy manifold, i.e., the label component of CRF. The smoothing allows to reduce the sampling density for $(\mathcal{D}_p^f)_p$, trimming the label component of the CRF. The image pyramid is defined in Algorithm 1.

As the images have been downsampled by a factor f, it is enough to use a $0.5 \times f$ discretization factor for $(\mathcal{D}_p^f)_p$.

At first glance, the downsampling seems to be beneficial as it reduces both components of the CRF. Unfortunately, the smoothing on the label component destroys the energy manifold as illustrated by the Fig. 3. Disparities that should have been retrieved

Algorithm 1: Image pyramid, the GM-IP algorithm

Data: I_r,I_t
Result: d
Set $d = 0$;
for f = *coarsest to finest downsampling factor* **do**
  Set $d_{init} = d$;
  Downsample I_r, I_t and d_{init} by $f \rightarrow [I_r^f, I_t^f, d_{init}^f]$;
  Define set of disparity to evaluate from $d_{init}^f \rightarrow (\mathcal{D}_p^f)_p$;
  Compute CRF from I_r^f, I_t^f and $\mathcal{D}^f \rightarrow CRF^f$;
  Solve CRF^f with Fast-PD with starting solution $d_{init}^f \rightarrow d^f$;
  if $f \neq 1$ **then**
    Upsample $d^f \rightarrow d$;
  else
    $d^f \rightarrow d$;

at coarse scales might be discarded. This generates, especially at coarse scales, artifacts in the retrieved disparity maps d^f. Some of these artifacts can be corrected at a finest scale only if they are not too important. This phenomenon is well known of the optical flow community (Horn and Schunck, 1981). While different heuristics, such as post-processing filtering between scales have been proposed (Sun et al., 2010), none can guarantee to correct these artifacts. This is the main motivation to build a proper representation of the energy manifold through an energy pyramid.

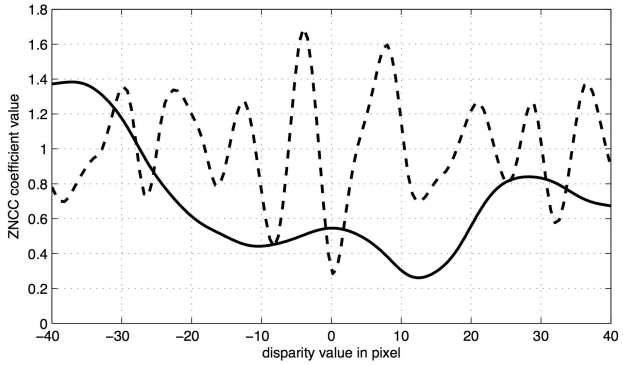


Figure 3: The unary terms (solid line) computed with the image pyramid ($k = 8$) poorly represent the energy manifold (dashed line).

3.3 Enegy pyramid: the GM-EP algorithm

Contrary to the GM-IP algorithm, the Global Matching via Energy Pyramid (GM-EP) algorithm is based on an energy pyramid approach that builds a more faithful representation of the energy as it directly downsamples CRF^1. Our approach is inspired from the work of (Bagon and Galun, 2012, Kim et al., 2011). We explain the downsampling procedure in three steps: (1) we define how to downsample the graph G of CRF^1; (2) we define the coarse edge costs; and (3) given a search space, we define the coarse unary terms.

3.3.1 Downsampling the graph From a downsampling factor f we define G_{Nodes} a function that groups nodes $\mathcal{V}$ of G in packets of $f \times f$ nodes. These packets define a set of nodes $\mathcal{V}^f$. This is illustrated by Fig. 4 as black circled nodes of $\mathcal{V}$ are grouped in a 2×2 fashion, creating the squared nodes of $\mathcal{V}^f$ (f=2).

The grouping of nodes separates the edges of $\mathcal{E}$ in two different kinds: (1) edges that belongs to the same node of $\mathcal{V}^f$, i.e., the

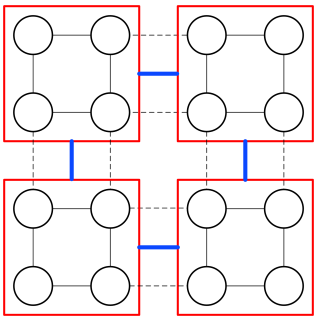


Figure 4: Grouping a set of 4×4 nodes into a set of 2×2 nodes and implications on the edge set.

black thin solid lines in Fig. 4; and (2) edges that belongs to two different nodes of $\mathcal{V}^f$, i.e., the dashed lines in Fig. 4. We discard the first kind of edges as they only contribute to a constant energy term. We define G_{Edges} as the function that groups the second kind of edges together. G_{Edges} defines the new set of edges $\mathcal{E}^f$, i.e., the thick solid lines in Fig. 4.

So far, the grouping procedure defines from $G = [\mathcal{V}, \mathcal{E}]$ a graph $G^f = [\mathcal{V}^f, \mathcal{E}^f]$. G^f is the spatial support of the donwsampled CRF, CRF^f.

3.3.2 Downsampling the edge costs The grouping procedure also naturally defines the edges costs w^f of CRF^f as for each edges $pq^f \in \mathcal{E}^f$ we have:

$$w_{pq}^f = \sum_{pq \in \mathcal{E} | G_{Edges}(pq) = pq^f} w_{pq}. \quad (15)$$

3.3.3 Downsampling the unary terms and defining the distance function The energy pyramid preserves the energy manifold with respect to the label component. As the unary terms are computed from I_r and I_t, we should use a half pixel discretization factor for $\mathcal{D}^f$. However, this would lead to a large label space at the coarsest scales. Instead, we propose to trim further the label component by creating a lower bound approximation of the energy.

Assume that we want each nodes of p^f of CRF^f to investigate a search space of $\mathcal{D}^f(p^f) = [([dMin : k : dMax] \times [dMin : k : dMax]) + d_{init}^f(p)]$, with $k > 0.5$. The unary terms $c_{p^f}^f$ are computed in four steps illustrated by Fig. 5:

1. We compute an intermediate set of unary terms c_p from CRF^1 over the search space $\mathcal{D}(p) = [(dMin - 0.5 \times k/2 : 0.5 : dMax + 0.5 \times k/2] \times [dMin - 0.5 \times k/2 : 0.5 : dMax + 0.5 \times k/2]) + d(p)]$. This is illustrated by the circles of Fig. 5.

2. We apply a morphologic erosion to c_p with a square kernel of size $k \times k$. The erosion creates a lower bound value of c_p, illustrated by the stars in Fig. 5.

3. We decimate the eroded coefficients with a k sampling step, illustrated by the squares in Fig. 5.

4. Finally, using the spatial grouping function, G_{Nodes}, we add all the decimated eroded coefficients to their respected nodes in $\mathcal{V}^f$, defining the unary terms $c_{p^f}^f$ of CRF^f.

The different steps of the GM-EP algorithm are summarized in Algorithm 2.

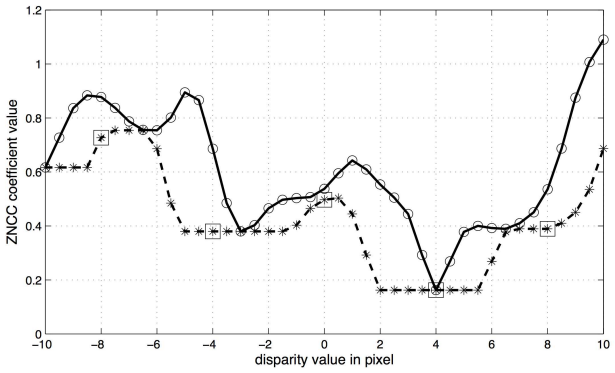


Figure 5: The erosion produces eroded unary terms, the crosses, that are the lowest of the $k = 8$ neighbors intermediate unary terms (the circles). The decimation defines the final unary terms (the squares).

Algorithm 2: Energy pyramid, GM-EP algorithm

Data: I_r,I_t
Result: d
Set $d = 0$
for f = *coarsest to finest downsampling factor* **do**
  Set $d_{init} = d$;
  Downsample d_{init} by $f \rightarrow d^f_{init}$;
  Define the set of disparity to evaluate from $d^f_{init} \rightarrow \mathcal{D}^f$
  Define the grouping from $f \rightarrow G_{Nodes}$ and G_{Edges}
  Apply the grouping on $G \rightarrow G^f = [\mathcal{V}^f, \mathcal{E}^f]$
  Apply grouping on edge cost of $CRF^1 \rightarrow w^f$
  Compute the unary terms of $CRF^f \rightarrow c^f_{pf}$
  Solve CRF^f with Fast-PD with starting solution $d^f_{init} \rightarrow d^f$;
  if $f \neq 1$ **then**
    Upsample $d^f \rightarrow d$;
  else
    $d^f \rightarrow d$;

3.4 Using sparsity to reduce the size of CRF

Thus far, we have defined two pyramid approaches to reduce both the spatial and label components of Eq. 13 that build a sequence of CRF, $[CRF^n, ..., CRF^f, ..., CRF^1]$. While we move through the scales, we always center the search space around the revised solution d. Consequently, due to Eq. 9, many unary terms may have an infinite value as illustrated in Fig. 6. Hence, we know beforehand that some configurations of d^f are impossible. To speed up the optimization process we indicate the solver, Fast-PD, not to evaluate the impossible configurations.

A potential approach is to remove all infinite unary terms, reorganize the CRF with respect to the label space and modify the distance function δ to take into account d_{init}. This leads to a new distance function $\delta^{'}$.

$$\delta^{'}(l_{pf}, l_{qf}) = \|(\mathcal{D}(p^f) + d_{init}(p^f)) - (\mathcal{D}(q^f) + d_{init}(q^f))\|_1 \quad (16)$$

The main drawback is that the new distance function $\delta^{'}$ is not derived from a metric as it does not respect the triangular inequality. Therefore, we are only mathematically guaranteed to find a local optimum during the optimization (Komodakis et al., 2007a).

Instead, we remove all infinite unary terms but we do not reorganize the CRF and we keep δ as a distance function. This leads to a sparse CRF that keeps the metric properties of the distance function δ. Hence, we are mathematically guaranteed to find a near global optimum to the solution. We design our own implementation of Fast-PD to accept sparse CRF. Modifying Fast-PD for sparse CRF input is beyond this paper's scope, but this will be the subject of a future publication.

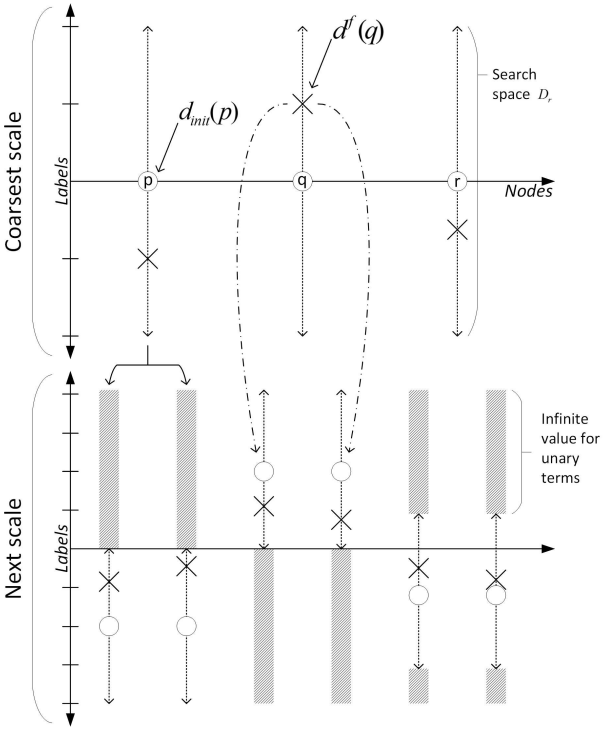


Figure 6: The multiscale approach generates sparsity in the CRF sequence, and only the coarsest CRF is assured to be full. The sparsity is due to defining search space around the previous up sampled scale solution d_{init}

4. EXPERIMENTAL RESULTS

Factory Urban

Figure 7: Image I_r from subsets of a stereo pairs. Each subset is 2500 × 2500 pixels with a GSD of 10 cm.

We use a stereo pair from an aerial survey above an urban environment acquired with (UtraCam, 2014) of 11000 × 20000 pixels. We use our own calibration and aerotriangulation to transform the images into an epipolar geometry limiting the optimization of d to the horizontal direction. We process overlapping tiles of 5000 × 5000 pixels and extract the two subsets presented in Fig. 7 to compare the disparity maps.

4.1 Comparing GM-EP and GM-IP algorithms

We compare the GM-EP and GM-IP algorithms by reporting the disparity maps among each scale with the associated final energy and global computation time. For both algorithms we use 4 scales, $f \in [8, 4, 2, 1]$, with a search range of $[-5 : 0.5 : 5] \times f$ around d_{init}, and we only perform one iteration per scale. We also define a baseline by solving CRF^0 over a large range of disparities, $[-40 : 0.5 : 40]$. Note that the baseline can here be computed because we only focus on a small subset of the entire image. The results are presented in Fig. 8.

The black dots that appear in the upper left part of the disparity maps of the image pyramid and the baseline are due to moving vehicles as the stereo pair is issued from quasi-simultaneous acquisitions. Moreover, the direction of motion is very different from the epipolar lines. Our model does not cope with such motion and we do not account for these artifacts in our comparison.

The coarsest scales, $f = 8$ and $f = 4$, illustrate how the energy pyramid of the GM-EP algorithm performs a better approximation of the energy manifold than the image pyramid of the GM-IP algorithm. Indeed, the GM-IP disparity maps show smoothed factory edges. However, the same edges appear sharp in the GM-EP results, and even the chimney in top right part of the factory is well defined at the scale $f = 4$. At the finest scales $f = 2$ and $f = 1$ the GM-IP is only able to sharpen some of the blurred edges. For instance, the two tubular elements in the center of the images remain blurred while they appear sharp in GM-EP disparity maps. The visual inspection is confirmed by the energy measurements in Table 1 as the final energy of the GM-EP approach is lower than that of GM-IP. The computation time given in Table 1, is however slightly in favor of GM-IP.

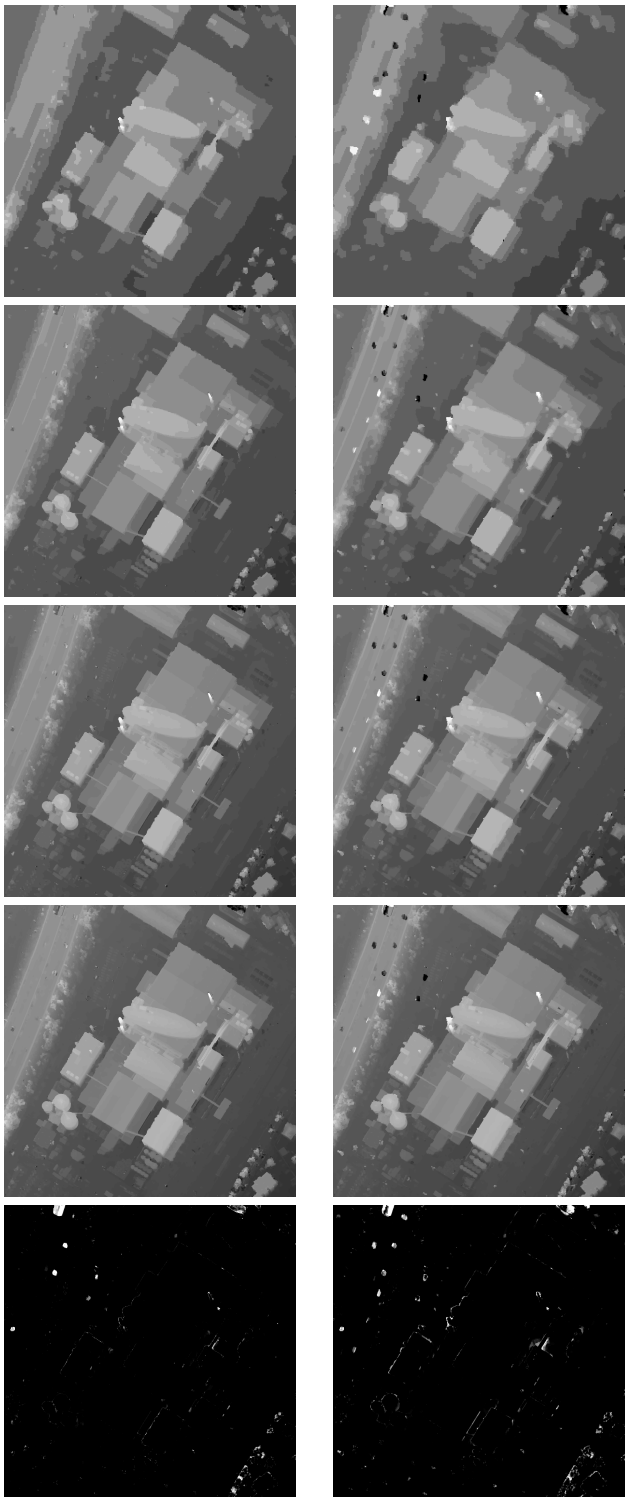
In Fig. 9, we verify how well the GM-EP compares to the baseline. Both results are very close. Nevertheless, little details such as lampposts in the lower left part of image are better represented in the baseline. This is also confirmed in Table 1 as the GM-EP final energy is slightly superior to the baseline energy. However, the computation of the baseline is more than 3 times superior to GM-EP, and this non-pyramidal approach is non-feasible for large images due to exponential memory consumption.

4.2 MicMac comparison

We also compare the GM-EP to (Micmac, 2014), and we only report the final disparity map as the computation time is implementation dependent and the energy is linked to the model, which differs for both algorithms. There are three main differences between the GM-EP and MicMac: (1) the GM-EP model is more complex than the one used in MicMac, i.e., both models seem equivalent if $\lambda_2 = 0$ in Eq. 7; (2), MicMac uses an image pyramid approach while the GM-EP works on an energy pyramid approach; and, (3) MicMac relies on semi-global optimization while GM-EP uses global optimization that produces near optimum solutions.

On both images, MicMac produces disparity maps with noisy areas that correspond to shadows the in stereo-pair, while the GM-EP is unaffected by shadows. This is due to the difference of models where the weight term of Eq. 7 increases the regularization on constant radiometric areas of the reference image.

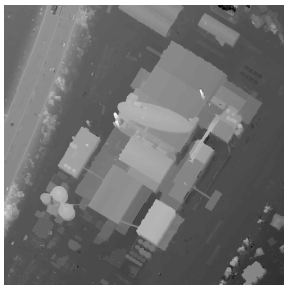
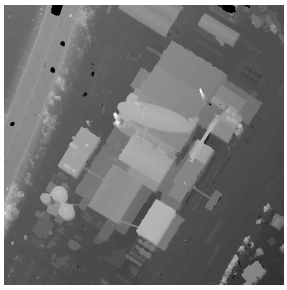
Like the GM-IP, MicMac uses an image pyramid that tends to produce blurred edges when dealing with large disparity range. This is illustrated in the top left image of Fig. 10 by the large chimney, the two tubular objects and the numerous bridges. This also appears on the bell tower in the lower right part of the bottom left image of Fig. 10. The edges produces by the GM-EP are sharper and better defined.



GM-EP GM-IP

Figure 8: Top four rows $f = [8, 4, 2, 1]$, black values represents disparities of -15 pixels or below, and white values represents disparities of 35 pixels or above. Last row absolute difference with respect to the baseline, brighter greys are higher difference.

MicMac suffers from the inherent issues limitations of image pyramid approaches, which are overcome by the GM-EP algorithm. Moreover, as the model of MicMac is less complete than ours, it introduces noisy areas.

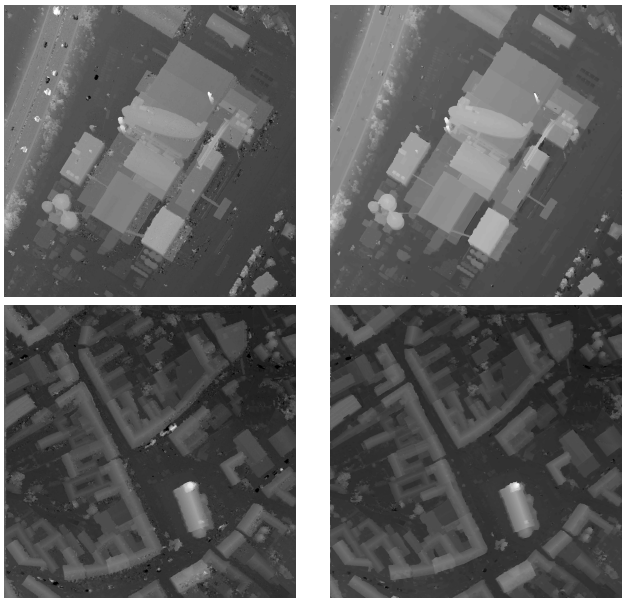



GM-EP Baseline no pyramid

Figure 9: Final disparity maps. Black values represents disparities of -15 pixels or below. White values represents disparities of 35 pixels or above.

	Baseline	GM-EP	GM-IP
Final energy	114.5 10^5	114.9 10^5	117.7 10^5
Computation time ratio	350%	108%	100%

Table 1: The GM-EP algorithm obtains a lower energy than GM-IP for a slightly longer processing time. The Baseline obtains the lowest energy with a significant gap compared to GM-IP, but is more than 3 times slower on an Intel Xeon 8 core CPU computer (the code is not optimized). Therefore, the GM-EP algorithm defines good compromise between effective optimization and computation time.



MicMac GM-EP

Figure 10: Disparity maps on the Factory and Urban subsets.

5. CONCLUSIONS AND EXTENSIONS

We proposed a rigorous yet computationally efficient framework for disparity maps estimation using a single stereo pair. This approach naturally lead to a global optimization problem of a CRF. Instead of relying on a semi-global optimization with no mathematical guarantee on the optimality of the solution, we use global discrete optimization which is mathematically guaranteed to yield a near optimum solution. The global optimization is speed-up by building a truthful representation of the energy manifold with an energy pyramid multiscale approach. We also exploited the sparsity of the resulting sequence of CRF to derive the GM-EP algorithm. From a practical standpoint, we demonstrate through experiments on real stereo-pairs the superiority of the GM-EP algorithm over image pyramid approaches whether they rely on global optimization, such as the GM-IP, or semi-global optimization, such as MicMac.

Future work will extend the experimental evaluation to standard photogrammetric tools to further assess the performance of our approach. Nevertheless, the comparison with MicMac outlines the importance of the energy modeling for disparity map estimations. Therefore, future work will extend our model to a symmetric definition of the energy with respect to the input images. Indeed, the choice of the reference and the target images from a stereo-pair is completely arbitrary and swapping them produces slightly different disparity maps. We think that a symmetric energy might enhance the quality of estimated disparity maps as the problem is by nature symmetric. We also plan to integrate the notion of occlusions and moving objects in our future model.

Finally, instead of using a multiscale approach, future work will focus on a multigrid scheme to further speed-up the optimization. Indeed, an inherent drawback of multiscale approaches is to force all nodes to work at the same scale, which might not be required as some parts of the disparity map might be well-resolved even at coarse scale.

ACKNOWLEDGMENTS

The authors are grateful to Nikos Komodakis for his precious help regarding Fast-PD. The authors would like to thanks the anonymous reviewers for helpful comments. This research was supported by the USGS through the Measurements of surface ruptures produced by continental earthquakes from optical imagery and LiDAR project (USGS Award G13AP00037), the Terrestrial Hazard Observation and Reporting Center of Caltech, the Moore foundation through the Advanced Earth Surface Observation Project (AESOP Grant 2808), and the ANR project STEREO.

REFERENCES

123DCatch, 2014. http://www.123dapp.com/catch.

Acute3D, 2014. http://www.acute3d.com/software/.

Ansar, A., Castano, A. and Matthies, L., 2004. Enhanced real-time stereo using bilateral filtering. In: 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on, pp. 455–462.

Bagon, S. and Galun, M., 2012. A multiscale framework for challenging discrete optimization. CoRR.

Birchfield, S. and Tomasi, C., 1998. A pixel dissimilarity measure that is insensitive to image sampling. Pattern Analysis and Machine Intelligence, IEEE Transactions on 20(4), pp. 401–406.

Boykov, Y., Veksler, O. and Zabih, R., 2001. Fast approximate energy minimization via graph cuts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 23(11), pp. 1222–1239.

Brown, L. G., 1992. A survey of image registration techniques. ACM Computing Surveys 24, pp. 325–376.

Brown, M., Burschka, D. and Hager, G., 2003. Advances in computational stereo. Pattern Analysis and Machine Intelligence, IEEE Transactions on 25(8), pp. 993–1008.

Felzenszwalb, P. and Huttenlocher, D., 2004. Efficient belief propagation for early vision. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 1, pp. I–261–I–268 Vol.1.

Fix, A., Gruber, A., Boros, E. and Zabih, R., 2011. A graph cut algorithm for higher-order markov random fields. In: Computer Vision (ICCV), 2011 IEEE International Conference on, pp. 1020–1027.

Gamble, E. and Poggio, T., 1987. Visual integration and detection of discontinuities: The key role of intensity edges. Technical report, Cambridge, MA, USA.

Geman, S. and Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-6(6), pp. 721–741.

Hartley, R. and Zisserman, A., 2004. Multiple View Geometry in Computer Vision. Cambridge University Press.

Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 2, pp. 807–814 vol. 2.

Horn, B. and Schunck, B., 1981. Determining optical flow. Artificial Intelligence 17(1-3), pp. 185–203.

Ishikawa, H., 2011. Transformation of general binary mrf minimization to the first-order case. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33(6), pp. 1234–1249.

Kappes, J., Andres, B., Hamprecht, F., Schnorr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B., Lellmann, J., Komodakis, N. and Rother, C., 2013. A comparative study of modern inference techniques for discrete energy minimization problems. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp. 1328–1335.

Kim, J., Kolmogorov, V. and Zabih, R., 2003. Visual correspondence using energy minimization and mutual information. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, pp. 1033–1040 vol.2.

Kim, T., Nowozin, S., Kohli, P. and Yoo, C., 2011. Variable grouping for energy minimization. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp. 1913–1920.

Klaus, A., Sormann, M. and Karner, K., 2006. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, Vol. 3, pp. 15–18.

Kolmogorov, V., 2006. Convergent tree-reweighted message passing for energy minimization. Pattern Analysis and Machine Intelligence, IEEE Transactions on 28(10), pp. 1568–1583.

Kolmogorov, V. and Zabih, R., 2001. Computing visual correspondence with occlusions using graph cuts. In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, Vol. 2, pp. 508–515 vol.2.

Komodakis, N., 2010. Towards More Efficient and Effective LP-Based Algorithms for MRF Optimization. Vol. 6312, Springer Berlin / Heidelberg, book section Lecture Notes in Computer Science, pp. 520–534.

Komodakis, N. and Paragios, N., 2009. Beyond pairwise energies: Efficient optimization for higher-order mrfs. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 2985–2992.

Komodakis, N. and Tziritas, G., 2007. Approximate labeling via graph cuts based on linear programming. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29(8), pp. 1436–1453.

Komodakis, N., Paragios, N. and Tziritas, G., 2007a. Mrf optimization via dual decomposition: Message-passing revisited. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pp. 1–8.

Komodakis, N., Tziritas, G. and Paragios, N., 2007b. Fast, approximately optimal solutions for single and dynamic mrfs. In: Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, pp. 1–8.

Li, Z., Zhu, Q. and Gold, C. M., 2005. Digital terrain modeling - principles and methodology. CRC Press.

Micmac, 2014. http://logiciels.ign.fr/?-micmac,3-.

Middleburry, 2014. http://vision.middlebury.edu/stereo/.

Pierrot-deseilligny, M. and Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: An application to surface reconstruction from spot5-hrs stereo imagery. In: In: Proc. of the ISPRS Conference Topographic Mapping From Space (With Special Emphasis on Small Satellites), ISPRS.

Pix4d, 2014. http://pix4d.com/topography/.

Scharstein, D., Szeliski, R. and Zabih, R., 2001. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: Stereo and Multi-Baseline Vision, 2001. (SMBV 2001). Proceedings. IEEE Workshop on, pp. 131–140.

Sun, D., Roth, S. and Black, M., 2010. Secrets of optical flow estimation and their principles. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 2432–2439.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M. and Rother, C., 2008. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30(6), pp. 1068–1080.

Tombari, F., Mattoccia, S., Di Stefano, L. and Addimanda, E., 2008. Classification and evaluation of cost aggregation methods for stereo correspondence. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pp. 1–8.

UtraCam, 2014. http://www.microsoft.com/en-us/ultracam/ultracameagle.aspx.

UtraMap, 2014. http://www.microsoft.com/en-us/ultracam/ultramap.aspx.

Viola, P. and Wells, W., 1995. Alignment by maximization of mutual information. In: Computer Vision, 1995. Proceedings., Fifth International Conference on, pp. 16–23.

Worldview, 2014. http://www.digitalglobe.com/products/information.

Zhang, Z., 1998. Determining the epipolar geometry and its uncertainty: A review. International Journal of Computer Vision 27(2), pp. 161–195.