A LOCAL ADAPTIVE APPROACH FOR DENSE STEREO MATCHING IN ARCHITECTURAL SCENE RECONSTRUCTION

C. Stentoumis[1], L. Grammatikopoulos[2], I. Kalisperakis[2], E. Petsa[2], G. Karras[1]

(cstent@mail.ntua.gr, lazaros@teiath.gr, ikal@teiath.gr, petsa@teiath.gr, gkarras@central.ntua.gr)
[1] Laboratory of Photogrammetry, Department of Surveying,
National Technical University of Athens, GR-15780 Athens, Greece
[2] Laboratory of Photogrammetry, Department of Surveying,
Technological Educational Institute of Athens, GR-12210 Athens, Greece

KEY WORDS: Matching, Correlation, Surface, Reconstruction, Point Cloud

ABSTRACT

In recent years, a demand for 3D models of various scales and precisions has been growing for a wide range of applications; among them, cultural heritage recording is a particularly important and challenging field. We outline an automatic 3D reconstruction pipeline, mainly focusing on dense stereo-matching which relies on a hierarchical, local optimization scheme. Our matching framework consists of a combination of robust cost measures, extracted via an intuitive cost aggregation support area and set within a coarse-to-fine strategy. The cost function is formulated by combining three individual costs: a cost computed on an extended census transformation of the images; the absolute difference cost, taking into account information from colour channels; and a cost based on the principal image derivatives. An efficient adaptive method of aggregating matching cost for each pixel is then applied, relying on linearly expanded cross skeleton support regions. Aggregated cost is smoothed via a 3D Gaussian function. Finally, a simple 'winner-takes-all' approach extracts the disparity value with minimum cost. This keeps algorithmic complexity and system computational requirements acceptably low for high resolution images (or real-time applications), when compared to complex matching functions of global formulations. The stereo algorithm adopts a hierarchical scheme to accommodate high-resolution images and complex scenes. In a last step, a robust post-processing work-flow is applied to enhance the disparity map and, consequently, the geometric quality of the reconstructed scene. Successful results from our implementation, which combines pre-existing algorithms and novel considerations, are presented and evaluated on the Middlebury platform.

1. INTRODUCTION

Generation of dense 3D information using calibrated images is a central part of most applications in the field of photogrammetry and computer vision (3D reconstruction, DSM production, novel view synthesis, automatic navigation). Two distinct processes are the core of automating 3D scene reconstruction: establishment of correspondences among images for camera calibration and dense 3D surface reconstruction. In this contribution a stereo matching algorithm for dense reconstruction is presented, based on epipolar images.

In recent years a significant number of efficient algorithms have been proposed for creating accurate disparity maps (storage of x-parallax for all pixels) from single stereo-pairs. The effectiveness of such algorithms has been extensively evaluated in surveys (Brown et al., 2003; Dhond & Aggarwal, 1989); Scharstein & Szeliski, 2002, Hirschmüller & Scharstein 2009) and a dedicated web site, which has also been used for evaluating our algorithm (http://vision.middlebury.edu/stereo/). After Scharstein & Szeliski (2002), dense stereo correspondence algorithms may be broken down into four main steps.

• ***Matching cost computation,*** where for every individual pixel a cost value is assigned to all possible disparities. Costs robust to radiometric differences, texture-less areas and regions with proximity to occlusion borders are needed. Klaus et al. (2006) have used a weighted sum of colour intensities and image gradients; Hirschmüller (2008) has employed the 'mutual information' approach in a semi-global context, while Mei et al. (2011) have combined absolute colour differences with a non-parametric image transformation (census).
• ***Cost aggregation***. The assumption is made that neighbouring pixels share the same disparity, thus a summation (aggregation) of initial pixel-wise matching costs is carried out over a support region around each pixel. Rectangular windows of fixed or variable size, fixed windows with varying weights (Yoon & Kweon, 2006), as well as regions of arbitrary shape (Zhang et al., 2009), have been proposed.
• ***Disparity optimization***. Here an optimal disparity value is selected for each pixel. Local methods (region-based) usually employ a 'winner-takes-all' strategy, namely the disparity with the lowest aggregated cost is chosen. Global methods, on the other hand, optimize an energy function defined over all image pixels by simultaneously imposing a smoothness constraint. Regarding the latter, various approaches have been implemented based on partial differential equations (Faugeras & Keriven, 1998; Strecha et al., 2004), dynamic programming (Veksler, 2005), simulated annealing (Barnard, 1986), belief propagation (Sun et al., 2003) and graph-cuts (Kolmogorov & Zabih, 2001).
• ***Disparity refinement***, which aims at correcting inaccurate disparity values and handling occlusion areas. Commonly used approaches include scan-line optimization, median filtering, sub-pixel estimation, region voting, peak removal or occluded and mismatched area detection as well as interpolation (Hirschmüller, 2008; Zhang et al., 2009; Mei et al., 2011).

An approach for stereo-matching efficient in complex scenes is presented here. State-of-the-art techniques are integrated, with certain improvements being proposed for the cost computation and cost aggregation processes. In particular, the matching cost combines, via an exponential function, three individual costs: a cost computed on an extended census transformation of images; the absolute difference cost, by taking into account information from colour channels; and a cost based on the principal image derivatives. Costs of all pixels of the reference image over all

possible disparities are stored in the form of a 'disparity space image' (Bobick & Intille, 1999). An aggregated cost volume is computed by linearly expanded cross skeleton support regions, similar to Zhang et al. (2009) and Mei et al. (2011). The aggregated cost volume is smoothed via a 3D Gaussian function, and the disparity map is estimated using a 'winner-takes-all' selection. The above steps are integrated into a hierarchical scheme. As a last step, a robust post-processing work-flow is applied to enhance the disparity map.

The algorithm presented here incorporates significant improvements compared to that recently published by Sentoumis et al. (2012). For instance, the cost function is re-established since the census-based cost is improved, and a new cost term based on gradients in added. Also, a complete post-processing procedure is proposed and applied, resulting in refined reconstructions. Finally, a hierarchical scheme is composed in order to accommodate high-resolution images and complex scenes.

2. MATCHING COST FUNCTION

2.1 Census on intensity principal derivatives

Census (T_C) is a non-parametric image transformation (Zabih & Woodfill, 1994). For a support neighbourhood N(m×n) of pixel **p**, a binary vector forms a map of neighbouring pixels with intensities smaller than that of **p**. A binary vector **I** of length m×n is then assigned to each pixel. Thus, if **q** is a neighbour of **p**:

$$c\left(p,q\right)=\begin{cases}0 & I\left(p\right)\le I\left(q\right)\\ 1 & I\left(p\right)>I\left(q\right)\end{cases} \tag{1}$$

$$T_C\left(p\right)=\bigotimes_{q\in N_p}\left(p,q\right) \tag{2}$$

In case m×n < 255 a 128 bit-string can store the descriptive vector of each pixel. Census-transformed image T_C depends on how a pixel relates to its environment within the image patch. It is, therefore, robust against linear changes in brightness/contrast, i.e. radiometric distortions not modifying the ordering of intensity values. Moreover, in this binary approach the actual values of individual pixel intensities do not affect the overall measure, but only a specific bit of the binary descriptor of **p**.

Unlike other approaches (Banks & Corke, 2001; Hirschmüller & Scharstein, 2009), in the present algorithm the transformation is performed not on gray-scale image intensity function ***I*** but on its principal derivatives $\partial\boldsymbol{I}/\partial$x, $\partial\boldsymbol{I}/\partial$y. Image derivatives relate to characteristic structural image features (points, edges) and are, of course, extensively exploited as a rich source of information in image processing and computer vision. In the case of image matching they are used e.g. in gradient-based methods in global optimization formulations, feature-based methods and local stereo (Scharstein 1994; Brown et al. 2003; Klaus et al. 2006).

The present approach provides an extended binary vector

$$T_C\left(p\right)=\bigotimes_{p\in\left\{\frac{\partial I}{\partial x},\frac{\partial I}{\partial y}\right\}}\ \bigotimes_{q\in N_p}c\left(p,q\right) \tag{3}$$

which strengthens the original transformation (Eq. 2). In Eq. (3) ⊗ denotes the act of concatenation, following the original definition of T_C.

Direct introduction of gradients in two image directions doubles the size of vector *Tc*, thus exploiting the representational potential of image gradients. The matching cost between pixel **p** of the reference image and its homologue pixel **p'** in the matching image is, then, calculated as the Hamming distance (Hamming, 1950). Census transformation based on image gradients appears as being less sensitive to radiometric differences and repetitive patterns for the purposes of stereo matching, while the discriminative capability of the binary descriptor increases, leading to results of higher accuracy.

Fig. 1 shows the performance of the proposed enhanced census matching function on the *Tsukuba* (above right) and *Teddy* stereo pairs (below right) of the Middlebury data. For comparison, disparity maps obtained from matching with the original census transformation are also seen (left). According to results supplied by the Middlebury evaluation platform, the new disparity maps after the aggregation step (see section 0) have up to 2.5% less erroneous pixels if evaluating for wrong disparities over the 1 pixel threshold. For sub-pixel accuracy (0.75 pixels) results for *Tsukuba* from the same platform are in fact improved by 5%.

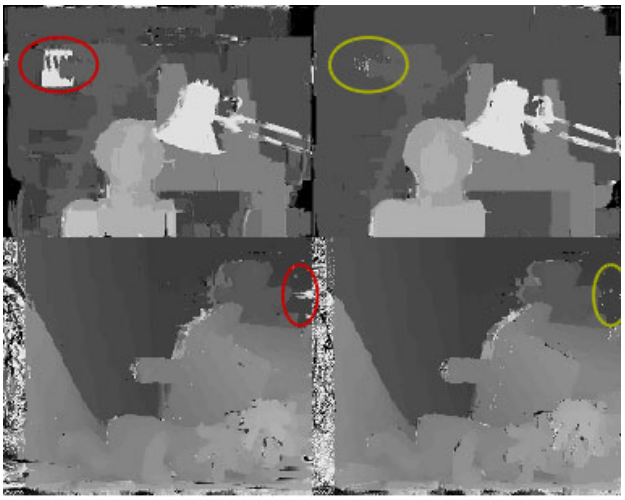

Figure 1. Disparity maps for the *Tsukuba* (above) and *Teddy* (below) stereo pairs obtained via the default census transformation (left) and census on gradients (right) after the aggregation step. Indicated are examples for areas of improvement.

2.2 Absolute difference on image colour

The absolute difference on colour channels (ADc), or on intensity, is a simple and easily implementable measure, widely used in matching (L_1 correlation: sum of absolute differences). Although sensitive to radiometric differences, it has been proven as an effective measure when combined with flexible aggregation areas and referring to combination of all colour layers. The cost term C_{ADc} is the average AD value of all three channels:

$$C_{ADc}\left(p,d\right)=\frac{1}{n}\sum_i\left(I_i^{ref}\left(p\left(x,y\right)\right)-I_i^{mat}\left(p\left(x,y\right),d\right)\right) \tag{4}$$

$$\forall i\in\left\{r,g,b\right\},\quad n\text{: number of image channels}$$

which turns out to improve results compared to matching on separate channels or gray-scale (Yoon & Kweon, 2006).

2.3 Absolute difference on image principal gradients

Here, the image intensity derivatives in the two principal directions are extracted, and the sum of absolute differences of each derivative value in the *x*, *y* directions is used as a cost measure. Use of directional derivatives (Eq. 5) separately, i.e. before they are summed up to a single measure ADg (Eq. 6), introduces into the cost measure the directional information for each derivative.

$$\nabla I\left(x,y\right)=\left(dI/dx,\,dI/dy\right) \tag{5}$$

$$C_{ADg}\left(p,d\right)=\sum_{x,y}\left(\nabla I^{ref}\left(p\left(x,y\right)\right)-\nabla I^{mat}\left(p\left(x,y\right),d\right)\right) \tag{6}$$

A mild Gaussian filter (3x3, σ = 0.5) is applied on the greyscale images before calculating partial derivatives for reducing noise and smoothing around image edges.

2.4 Total matching cost

The final matching cost C is derived by merging three different costs: Census transformation on image gradients (expressed via the Hamming distance), absolute difference in colour (or intensity) values and absolute difference on principal image gradients. A robust exponential function for cost combination (Yoon & Kweon, 2006; Mei et al., 2011) has been preferred, which resembles a Laplacian kernel (Eq. 7):

$$\rho\left(C,\lambda\right)=1-\exp\left(-\frac{C}{\lambda}\right) \tag{7}$$

$$C\left(x,y,d\right)=1-\exp\left(-\frac{C_c}{\lambda_c}\right)+1-\exp\left(-\frac{C_{ADc}}{\lambda_{ADc}}\right)+1-\exp\left(-\frac{C_{ADg}}{\lambda_{ADg}}\right) \tag{8}$$

The values of each cost should be normalized by λ to ensure equal contribution to the final cost, or tuned differently to accordingly adjust their impact on cost. Tests on the Middlebury data for stereo-matching are presented in Fig. 2 (see next page).

3. COST AGGREGATION

3.1 Support region formation

Here, a modification of the cross-based support region approach is used. Such cross-based support regions are constructed by expanding around each pixel **p** a cross-shaped skeleton to create 4 segments defining two sets of pixels H(**p**), V(**p**) in the horizontal and vertical directions, as seen in Fig. 3 (Zhang et al., 2009; Mei et al., 2011). The support skeleton (H(**p**) ∪ V(**p**)) is expanded on the basis of thresholds in spatial and colour space.

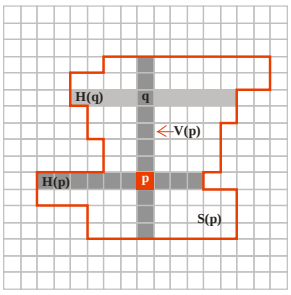

Figure 2. Expansion of the cross-based support region S(**p**) driven by the skeleton of each pixel. The skeleton pixels for V(**p**) and H(**p**) sets are calculated only once per pixel. When pixel **q** belongs to V(**p**), the corresponding horizontal arm H(**q**) is added to S(**p**). S(**p**) consists of the union of H(**q**) for all pixels **q** which participate in V(**p**). [After Zhang et al., 2009.]

For a more detailed description of cost aggregation, reference is made to Stentoumis et al. (2012). Thus only the main principles of this approach and our new contributions are mentioned here.

A linear threshold is adopted for the cross-skeleton expansion:

$$\tau\left(l_q\right)=-\frac{\tau_{max}}{L_{max}}\times l_q+\tau_{max} \tag{9}$$

This linear threshold in colour similarity is a function of the distance between neighbouring pixels **p** and **q**. The user-defined arguments in Eq. (9) express the largest semi-dimension L_{max} of the window size and the largest colour disimilarity τ_{max} between **p** and **q**. Typical support regions created according to the above considerations are presented in Fig. 4.

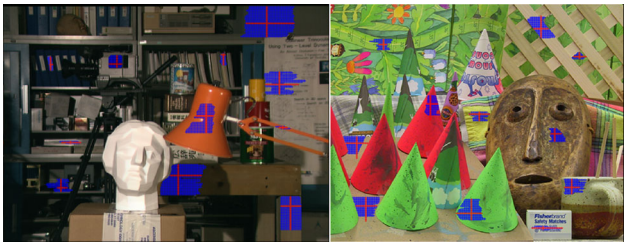

Figure 3. Examples of regions formed with the linear approach for the generation of cross-based windows.

We note that generally a 3×3 median filter is applied for cross-skeleton determination to ensure that image noise will not prevent skeleton expansion. In case of defocused (or blurred) imagery, adaptive histogram equalization is applied before median filtering. Moreover, the minimum length of all cross-segments is 1 pixel to ensure a minimum support region S of 9 pixels.

3.2 Aggregation step

Aggregation is applied on cost-disparity volume using the combined support region S (i.e. the intersection of the support regions which are formed based on the left I^{ref} and right I^{mat} images per each disparity). The aggregated pixel costs C_{aggr} are normalized by the number of pixels in the support region to ensure that costs per pixel have the same scale (Eq. 10):

$$C\left(p,d\right)=\frac{C_{aggr}}{\left\|S\left(p,d\right)\right\|} \tag{10}$$

At this point of the algorithm, a 3D Gaussian function is applied for smoothing the aggregated cost volume. This improves coherence of neighbouring cost values and removes noise from the cost. Aggregation is implemented through integral images for reasons of efficiency (Viola & Jones, 2001). Disparity is estimated in the 'winner-takes-all' mode, i.e. by simply selecting the disparity label with the lowest cost.

4. HIERARCHICAL APPROACH

An increasing variety of high resolution image acquisition hardware is available today, both to the general public and professionals. Hence, algorithms which are capable of handling large data are needed. An extension to high resolution images of the method reported in the preceding sections is inevitably based on scaled representations of the stereo pair. The aim is to limit the disparity search space to a computationally feasible range, and besides to guide matched disparities in a coarse-to-fine context.

This method also reveals structures in different layers of image pyramids, which lead from a rough but robust 3D surface to fine detail as one "climbs up" the image pyramid. Typical Gaussian pyramids are employed here, with a 3×3 ($\sigma = 0.5$) filter for all scales and subsequent down-sampling by a factor of 2.

The disparity map is expanded to the next finer level by propagating disparities via bilinear interpolation and smoothing them with a Gaussian filter for removing spikes. The initial disparity map can be regarded as a zero-map, and the search space can be very roughly bounded between zero and the width of the lowest pyramid layer. At this point, the already defined cross windows are introduced to restrict the field of values of the disparity function per pixel, extended by a number of pixels (here 2 pixels).

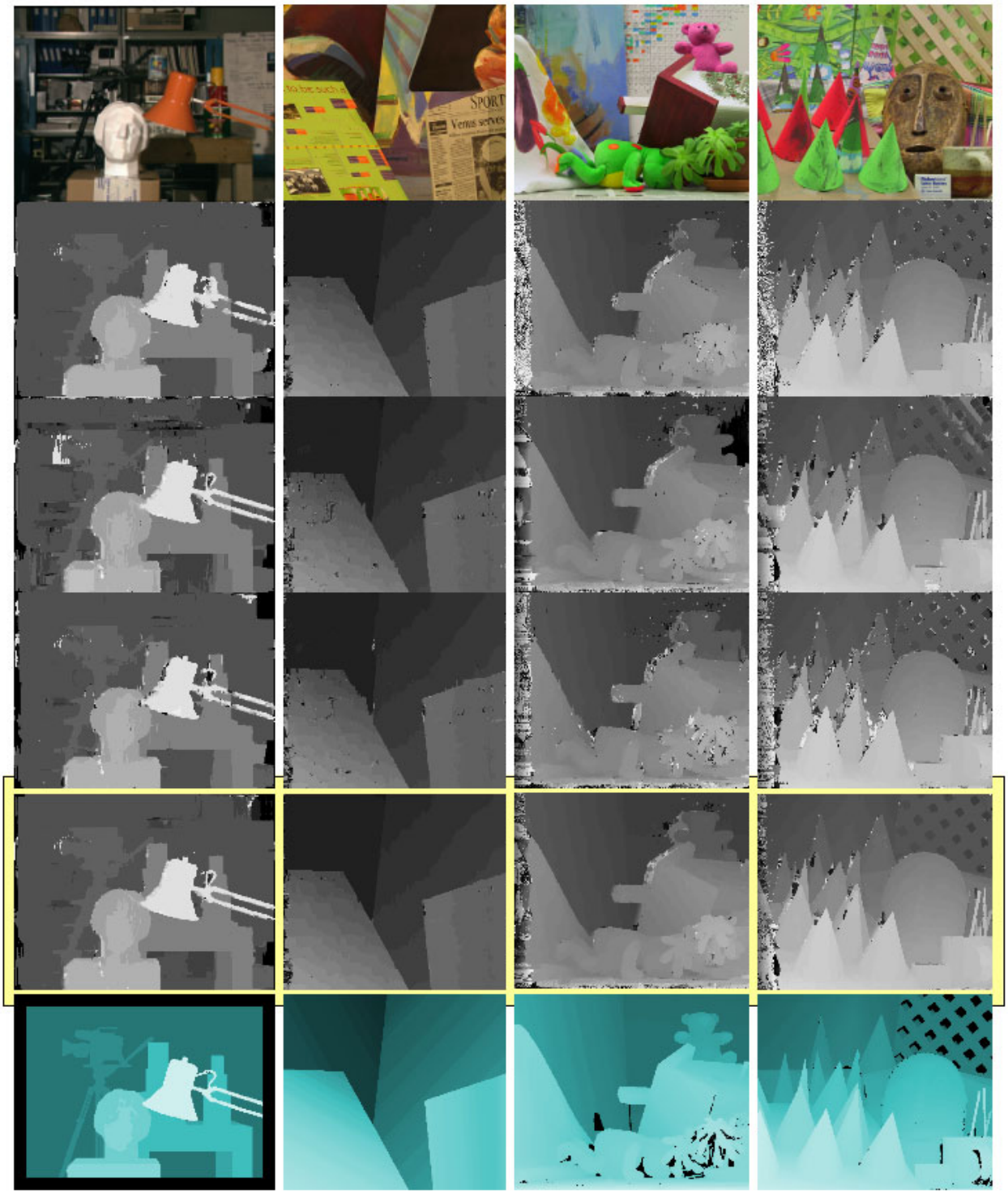

Figure 4. Comparison of different matching cost functions for four Middlebury stereo pairs. Disparity maps are presented from individual and the overall matching function after the aggregation step. *Top to bottom*: left image; proposed modified census transformation; AD on image channels; AD in principal images gradients; combined cost; and reference disparity maps. The refinement steps described in section 5 have not been used here, in order to illustrate individual results and the improvement achieved by fusing the three costs.

In particular, it is accepted that neighbouring pixels ***q***, i.e. pixels belonging to the support region $\mathbf{S}_{\boldsymbol{p}}$ of ***p*** as defined in the cost aggregation step (section 3.1), can adequately define the range via their approximate disparity $d_p^{(s-1)}$ computed in the coarser layer:

$$d_p^s \in \left\{ min\, d_q^{(s-1)}, \ldots max\, d_q^{(s-1)} \right\}, \forall q \in S_p \qquad (11)$$

In this way mismatches near edges, where abrupt "jumps" in the disparity map occur, are addressed. The search interval is adequately wide to accommodate edge misplacement in the disparity map of lower resolution layers, while at the same time being restricted through the support region.

5. POST-PROCESSING

5.1 Left-right match constraint

Matching consistency ('cross-checking') between reference and match images is a common reliable tool for evaluating the quality of a disparity map (Banks, 2001; Brown et al., 2003). It is both easy to implement in local stereo algorithms and efficient for validating matched pixels, although it does not distinguish among outliers from different origins. Pixel **p** is characterized as valid (inlier) if the following constraint holds for the disparity maps D_{map} of the reference and matching images:

$$D_{map}^{base}(p) = D_{map}^{match}\left(p - \left[D_{map}^{base}(p), 0\right]\right) \qquad (12)$$

5.2 Outlier cross-based filtering

The cross-based support regions provide a robust description of pixel neighbourhoods. For this reason cross-windows can be exploited to correct localized outliers in a disparity map (Lu et al., 2008; K. Zhang et al., 2009; Mei et al., 2011). Thus, the valid disparities belonging to the support region $\mathbf{S}_{\boldsymbol{p}}$ for an outlying pixel ***p*** are retrieved. In this paper, the median value of inliers in the support region is selected and attributed to the mismatched pixel. The method is iterative since it is a 'region closing' technique for large areas of outliers, such as occlusion areas, which are progressively filled with neighbouring disparities.

5.3 Occlusion/mismatch labelling

Remaining outliers are re-estimated via an image match consistency test. Outliers may stem from different origins. Two main cases, however, are of particular importance for developing an interpolation strategy, namely occlusions and mismatches. In Bobick & Intille (1999) and Brown et al. (2003) interesting approaches are found; yet, an efficient technique is that suggested by Hirschmüller (2008). For rejected pixels characterized as mismatches, a new disparity value is given using median interpolation in a small patch around them; occluded pixels, on the other hand, are given the second lowest disparity value in their neighbourhood as their new disparity value. Fig. 5 presents an example to illustrate the improvement of matching results after this refining treatment of mismatches and occlusions.

5.4 Sub-pixel estimation

Finally, estimation at the sub-pixel level is made by interpolating a 2nd order curve to the cost volume C(d). This curve is defined by the disparities of the preceding and following pixels of the 'winner-takes-all' solution and their corresponding cost values. Optimal sub-pixel disparity value d_{opt} is determined by the minimum cost position through a closed form solution for the 3 curve points ($d_{opt} = \text{argmin}_d(C(d))$).

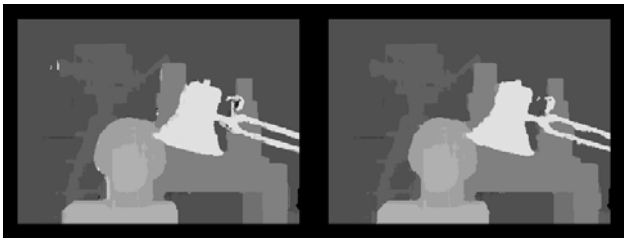

Figure 5. Disparity map of the left *Tsukuba* image before (left) and after (right) occlusions and mismatches have been handled.

5.5 Disparity map smoothing

A final median filter is applied on the disparity map after sub-pixel estimation to reject spikes. Moreover, as in the case of the *Herz-Jesu-K7* stereo-pair (see following section), a bilateral filter (Tomasi & Manduchi, 1998) is applied on the disparity map to improve the quality of the reconstructed point cloud without disturbing disparity, i.e. object edges. Fig.6. gives an example illustrating the effect of the overall post-processing refinement.

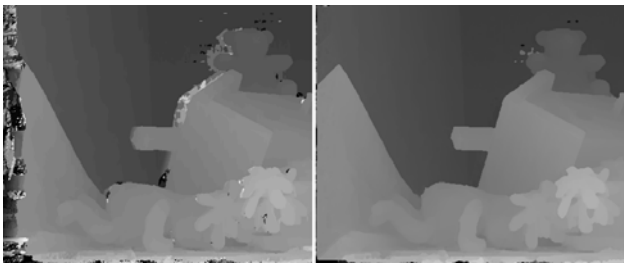

Figure 6. Disparity map of the *Teddy* reference image before (left) and after (right) the overall refinement.

Finally, Fig.7 presents the improvement achieved from each refinement step for the 0.75 pixel threshold. The average error for all four Middlebury stereo-pairs has been taken into account.

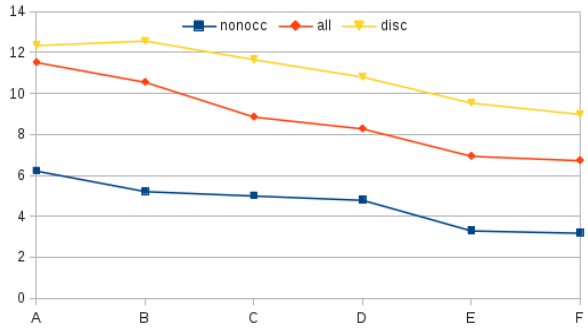

Figure 7. Performance of each refinement procedure regarding disparity map accuracy for the 0.75 pixel threshold, seen as average error for four Middlebury datasets (% of wrong pixels). A: initial disparity map, B: cost smoothing, C: outlier cross-based filtering, D: remaining occlusion/mismatch handling, E: sub-pixel estimation, F: median smoothing (*nonocc*: non-occluded pixels, *disc*: discontinuities, *all*: all pixels).

6. RESULTS

The presented algorithm has been evaluated on the Middlebury on-line platform and also tested using EPFL multi-view datasets (Strecha et al., 2008). Images acquired with the same camera, or

under the same conditions, are expected to be handled with the same parameter set. Thus, parameter values were kept constant for all tests (Table 1), although results from the EPFL dataset may improve significantly if parameters are tuned.

Table 1. Parameters values used in all dataset evaluations.

census transformation	m	11	lambda λ_c	45
	n	9	lambda λ_{ADc}	5
length threshold L_{max}		31	lambda λ_{ADg}	18
colour threshold τ_{max}		24		

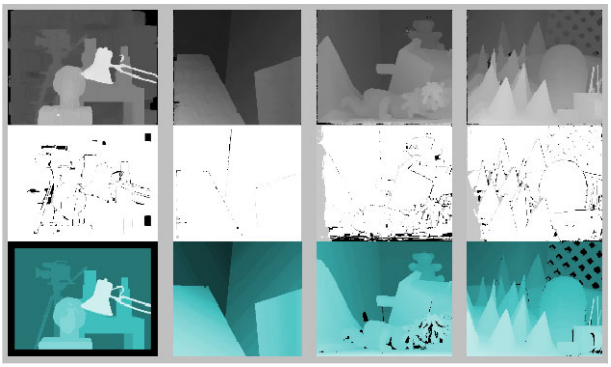

Figure 8. *Top*: final resulting disparity maps after full matching procedure using the four stereo pairs of Middlebury evaluation platform. *Middle*: "bad" pixels of the produced disparity maps evaluated for the 0.75 pixel error threshold. Mismatched pixels in occluded areas are indicated by gray colour, non-occluded areas in black. *Bottom*: True disparity maps.

Final results of the complete method, along with erroneous pixels, are seen in Fig. 8. Its performance is indeed encouraging, as it is rated among top algorithms using the Middlebury evaluation stereo dataset. When comparing for the >1 pixel error tolerance, our approach scores worse than *ADCensus* (Mei et al., 2011) which is currently at the 2nd position of this platform rating and uses a similar aggregation support region based on support skeletons (Zhang et al., 2009). On the other hand, a major improvement is observed when comparing for sub-pixel accuracy. The pipeline proposed in this paper results in a high performance under the error thresholds of 0.5 and 0.75 pixels in disparity values disposition. In fact, it rates 2nd in the 0.75 pixels threshold comparison, while most top performing methods for the 1 pixel threshold give bad results for sub-pixel testing, regardless of optimization method (local, global). It is also noted that the only method outperforming our algorithm for the 0.75 pixel threshold compares poorly for the 1 pixel threshold. Average % error in our case is 6.15 (Fig. 9). Good sub-pixel accuracy is important for several applications since it significantly improves the quality of 3D reconstruction and of the triangular meshes subsequently produced (lack of sub-pixel accuracy results in poor 3D meshes because of the discrete values of depth information).

Finally, Fig. 10 shows a *Herz-Jesu-K7* stereo pair (6 Megapixel images: 0006.png, 0007.png) and the created disparity map. An indication for the accuracy of reconstruction has been gained by registering the generated point cloud onto the ground truth data via the ICP algorithm (Fig. 11). Mismatch is represented by an average distance of 10 mm and a standard deviation of 19 mm. If reduced to mean image scale, these values correspond to ~1.8 and ~3.4 pixels, which are considered as quite satisfactory.

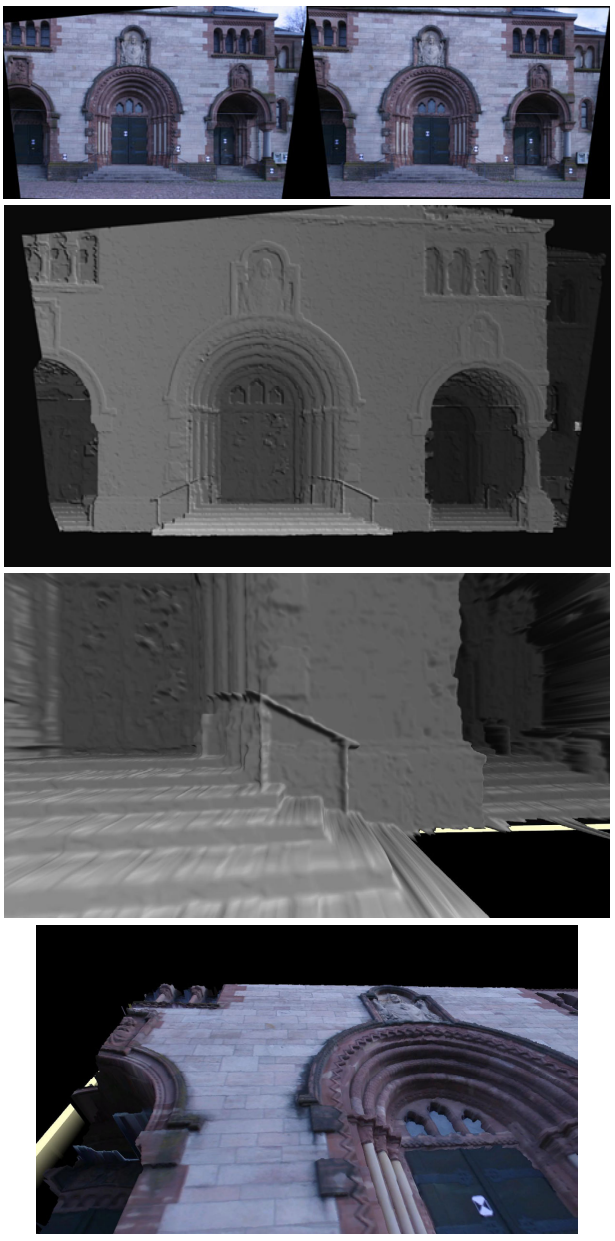

Figure 9. From top to bottom: epipolar *Herz-Jesu-K7* stereo-pair; disparity map; detail of the 3D representation of the disparity map; textured 3D representation of the disparity map.

SubPixSearch [127]	6.0	4.64 4	5.28 4	**8.19** 1	0.27 6	0.67 15	3.26 13	4.87 5	7.90 3	13.1 5	2.74 5	7.87 6	7.67 5	5.53
YOUR METHOD	11.1	4.10 3	4.92 2	8.66 2	0.30 8	0.74 20	3.82 21	5.48 11	12.0 25	14.9 13	2.61 4	9.14 21	7.14 3	6.15
SRW [139]	14.0	5.93 11	6.70 10	13.1 10	0.31 9	0.85 26	3.17 11	4.80 4	11.4 19	11.7 3	3.33 19	9.20 22	9.45 24	6.66
PMF [138]	17.2	11.0 42	11.4 40	16.0 30	0.45 22	0.63 12	4.91 50	**3.12** 1	7.26 2	**10.1** 1	2.34 2	7.29 3	6.81 2	6.78

Figure 10. Results from the Middlebury evaluation platform for the 0.75 pixel threshold. Columns record from left to right: method; average rank; errors for the *Tsukuba*, *Venus*, *Teddy*, *Cones* stereo pairs; and average percent of bad pixels (errors are recorded for cases of non-occluded pixels, all pixels, discontinuities). Date of evaluation: January 30, 2013.

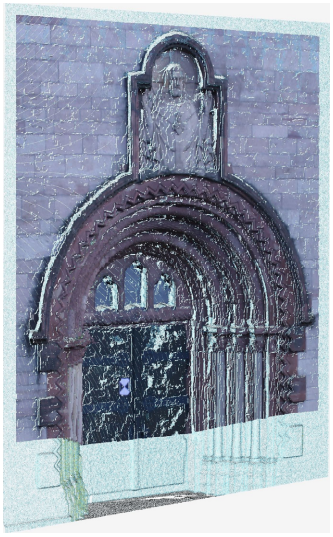

Figure 11. Registration of the reconstructed point cloud onto the ground truth data (*Herz-Jesu-K7* stereo-pair).

7. CONCLUDING REMARKS

The stereo matching algorithm, presented here as part of a 3D reconstruction pipeline, is based on a local hierarchical scheme. As illustrated in the preceding section, architectural scenes, for instance, may be accurately reconstructed using standard commercial cameras without participation of the user. Although global optimization methods turn out to be more accurate over the years, they face limitations in memory and speed, especially in cases of large images. On the other hand, developments on cost aggregation support regions exploited in local methods appear to produce competitive (if not better) results; such approaches, moreover, are usually easier to implement, computationally feasible and sufficiently fast, even for real-time tasks. Inherent limitations (i.e. favouring front parallel surfaces and unconstrained relations among neighbouring pixels) can be gradually dealt with. Of course, the reconstruction of a full 3D model requires combination of more views, whereby in this case multiple depth values for a voxel in space will be available, thus providing additional information for removing erroneous disparities. In this sense, future research topics include fusion of more views for each scene, but also improvements in the matching algorithm itself. In particular, a better adaptive combination of individual cost measures in the final cost function through an image segmentation scheme is expected to improve performance, particularly for high resolution imagery.

8. REFERENCES

Banks J., Corke P., 2001. Quantitative evaluation of matching methods and validity measures for stereo vision. *International Journal of Robotics Research*, 20(7):512–532.

Barnard S.T., 1986. A stochastic approach to stereo vision. *5th National Conference on Artificial Intelligence*, pp. 676-689.

Bobick A.F., Intille S.S., 1999. Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200.

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

Dhond U.R., Aggarwal J.K., 1989. Structure from stereo – a review. *IEEE Transactions on Systems, Man & Cybernetics*, 19(6): 1489–1510.

Faugeras O., Keriven R., 1998. Variational principles, surface evolution, PDE's, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344.

Hamming R.W., 1950. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160.

Hirschmüller H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341.

Hirschmüller H., Scharstein D., 2009. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1582–1599.

Klaus A., Sormann M., Karner K., 2006. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. *Proc. IEEE International Conference on Pattern Recognition*, pp. 15–18.

Kolmogorov V., Zabih R., 2001. Computing visual correspondence with occlusions using graph cuts. *IEEE International Conference on Computer Vision*, vol. 2, pp. 508–515.

Lu J., Lafruit G., Catthoor F., 2008. Anisotropic local high-confidence voting for accurate stereo correspondence. I*mage Processing: Algorithms and Systems VI*, Proc. SPIE, vol. 6812, pp. 1–12.

Mei X., Sun X., Zhou M., Jiao S., Wang H., Zhang X., 2011. On building an accurate stereo matching system on graphics hardware. *ICCV Workshop on GPU in Computer Vision Applications*, pp. 467–474.

Scharstein D., 1994. Matching images by comparing their gradient fields. *International Conference on Pattern Recognition*, vol. 1, pp. 572–575.

Scharstein D., Szeliski R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1), pp. 7–42.

Stentoumis C., Grammatikopoulos L., Kalisperakis I., Karras G., 2012. Implementing an adaptive approach for dense stereo-matching. *International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 39, pp. 309–314.

Strecha C., Fransens R., van Gool L., 2004. A probabilistic approach to large displacement optical flow and occlusion detection. *Statistical Methods in Video Processing*, Lecture Notes in Computer Science, vol. 3247, Springer, pp. 25–45.

Strecha C., von Hansen W., van Gool L., Fua P., Thoennessen U., 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8.

Sun J., Zheng N.-N., Shum H.-Y., 2003. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, 25(7):787–800.

Tomasi C., Manduchi R., 1998. Bilateral filtering for gray and color images. *IEEE International Conference on Computer Vision*, pp. 839–846.

Veksler O., 2005. Stereo correspondence by dynamic programming on a tree. *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 384–390.

Viola P., Jones M., 2001. Rapid object detection using a boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518.

Yoon K.J., Kweon I.S., 2006. Adaptive support-weight approach for correspondence search. I*EEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656.

Zabih R., Woodfill J., 1994. Non-parametric local transforms for computing visual correspondence. *European Conference in Computer Vision*, pp. 151–158.

Zhang K., Lu J., Lafruit G., 2009. Cross-based local stereo matching using orthogonal integral images. *IEEE Transactions on Circuits & Systems for Video Technology*, 19(7):1073–1079.