HAL open science

Distortion driven variational multi-view reconstruction

Patricio Alejandro Galindo, Rhaleb Zayer

To cite this version:

Patricio Alejandro Galindo, Rhaleb Zayer. Distortion driven variational multi-view reconstruction. Proceedings on International Conference in 3D Vision (3DV), Dec 2014, Tokyo, Japan. hal-01088428

HAL Id: hal-01088428

https://inria.hal.science/hal-01088428v1

Submitted on 27 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.

Distortion driven variational multi-view reconstruction

Patricio Galindo
INRIA
Nancy, France
patricio.galindo@inria.fr

Rhaleb Zayer
INRIA
Nancy, France
rhaleb.zayer@inria.fr

***Abstract*—This paper revisits variational multi-view stereo and identifies two issues pertaining to matching and view merging: i) regions with low visibility and relatively high depth variation are only resolved by the sole regularizer contribution. This often induces wrong matches which tend to bleed into neighboring regions, and more importantly distort nearby features. ii) small matching errors can lead to overlapping surface layers which cannot be easily addressed by standard outlier removal techniques. In both scenarios, we rely on the analysis of the distortion of spatial and planar maps in order to improve the quality of the reconstruction. At the matching level, an anisotropic diffusion driven by spatial grid distortion is proposed to steer grid lines away from those problematic regions. At the merging level, advantage is taken of Lambert's cosine law to favor contributions from image areas where the cosine angle between the surface normal and the line of sight is maximal. Tests on standard benchmarks suggest a good blend between computational efficiency, ease of implementation, and reconstruction quality.**

I. Introduction

The classical variational matching, e.g. [1], [2], [3], [4] provides a simple and straightforward mean for multi-view reconstruction. However, the global nature of the solution raises several challenges at the data exploitation and post processing. It is therefore not surprising that on benchmarks such as [5], none of the top performing methods, e.g. [6], [7], [8], [9], [10], [11] is purely variational.

The contributions of this paper to the classical variational setting are twofold. First, a modification of the variational objective function to account for weakly resolved regions is proposed. Second, a filtering approach which allows selecting the best relative contribution of image pairs and triplets is proposed, thus reducing data redundancy and noise effects without sacrificing completeness. The proposed methods are motivated by two observations:

i) At the matching level, we observe that the quality of the results deteriorates near large discontinuities where images do not provide enough information. Typical examples are shown in figure 1. In fact, the classical variational formulation comprises a data term and a robust smoothing regularizer. So when there is little image information only the smoothing term accounts for the results, as for instance, the basin of the fountain (figure 1). While such data points can be filtered out in a post-processing step, the global nature of the solution causes these wrong matches to bleed into neighboring areas, e.g. the rim of the fountain in figure 1.

It can be argued that one could possibly tweak the parameters for individual cases to improve the results around the mentioned regions or, use a total variation formulation with an $L1$ norm (e.g. [12]), in which case the results in other regions would be negatively affected by the staircaising effect commonly present in results that use this method [13]. Alternatively, occlusion detection (e.g. [14], [15]) and confidence measures (e.g. [16], [17]) can be included in the variational matching. These approaches have shown improvements for scenes with small discontinuities (e.g. [14], [15], [16]) and complete occlusions, as in [17]. In the later work, occluded regions are excluded from further computations as soon as they are detected. As a general remark, most of these methods rely only on 2D information without accounting for the resulting spatial configuration. Furthermore, we are not aware of any special treatments for fixing mismatches located next to largely sheared or occluded regions. The method proposed herein targets the more general problem of resolving regions with highly crammed correspondences regardless if they stem from occlusions or tilts. **ii) At the merging level**, we note that due to the dense nature of the solution, the size of the data can become very large, encompassing tens or hundreds of millions of points. The processing of such large spatial data is generally a daunting task and raises many challenges. Furthermore, the variational nature of the approach leads to over- and inter-layered data and none of the available methods are specifically tailored to take advantage of its nature. A typical example of over and interlaying data is shown in figure 2.

Certainly, ideas outlined in local matching approaches e.g. [6], [10] and the more general [7], [18] can help filter out the resulting point cloud by means of spatial visibility constraints. In particular [18], formulates the problem as quasi-dense matching followed by a variational refinement and a later global optimization where the number of intersections of the line of sight with the surface is minimized. Other multi-view approaches, which work on the fusion of depth maps to provide an effective multi-view solution (e.g. [19]), have proven successful when used in a small number of images and/or at low resolutions (e.g. [20]); but do not scale well due to the multiple volumetric data structures needed

Figure 1. Closeups on reconstruction results obtained using classical variational matching across a pair of views. The Fountain-P11 [5] (top) and our face1 data set (bottom). Please note, the deterioration along the rim of the fountain and the tip of the nose.

at all times. As a general remark methods which operate globally e.g. [7], [19] have a significant memory footprint which can be challenging when dealing with large datasets or high-resolution images.

In order to address the issues raised at both levels, we rely on the analysis of the distortion of spatial and planar maps. We characterize the distortion induced by these maps in order to guide the variational matching and the merging of the results. At the matching stage, we observe that regions with large discontinuities present important shearing in the resulting 3D grid. In section (II), we show how this shearing information can be used to identify problematic regions and how it can be injected into the variational formulation to drive the optimization process adaptively. At the merging stage, described in section (III), we regard the matching as a mapping of a regular grid from one view to another and we study the evolution of certain local geometric measures. While we are well aware that thresholds on triangle quality have been commonly used in removing possible outliers, e.g. [7], we emphasize that a naive thresholding indiscriminately removes data and can negatively affect completeness. Moreover, it does not necessarily reduce the over- and interlaying in the data shown, e.g. figure 2.

Most closely related to our work is the method presented in [21] which maintains a constant error while merging multiple views; however, our goal is to define what is best viewed in an image triplet or in an image pair relative to their neighboring views. By recalling that the amount of a viewed area decreases as it is tilted with respect to the view point (see section III), our approach favors matches from regions where the cosine of the surface normal and line of sight is large. Certainly, our merging method does not aim to provide a globally optimal solution, rather, it aims to locally and adaptively select the best contributions from pairs and triplets of views in order to avoid the layering problem without sacrificing completeness or modifying the data.

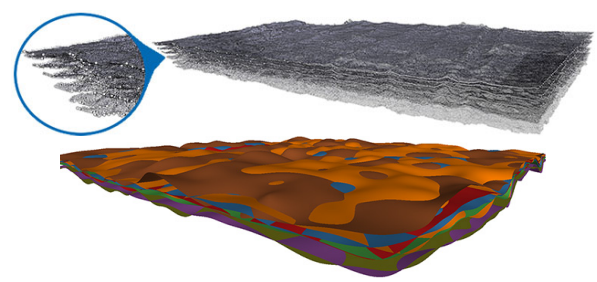

Figure 2. Two slices through the fountain back-wall (see figure 1) obtained using the classical formulation of variational multi-view matching. Besides from the overwhelming data size, slight matching errors can lead to overlaying (top) and interlaying (bottom) data.

II. Geometry driven variational matching

A. Classical variational matching

The variational stereo pair matching formalism emerged from the related optical flow problem which has received extensive attention in the literature [1], [2], [22]. Within this formalism several issues can be addressed including large displacements, illumination variations, and strong and large discontinuities. It aims to minimize the following objective function:

$$E(u,v) = \int_\Gamma \Psi_d((I_1 - I_0^w)^2) + \alpha\Psi_d((\nabla I_1 - \nabla I_0^w)^2) + \beta\Psi_s(|\nabla u|^2 + |\nabla v|^2) \quad (1)$$

The first two terms in equation (1) are commonly referred to as the data terms. They quantify the change in intensities and gradients between the warped source image I_0^w and the target image I_1, with $I_0^w = I_0(x+u, y+v)$. The differences in image gradients are a reliable way to match when changes in illumination occur. The warping helps reduce the displacement within the coarse-to-fine (pyramidal) formulation. The smoothness of the resulting mapping is controlled through the spatial gradient. The robust functions Ψ_d and Ψ_s are introduced to alleviate problems related to outliers, noise and occlusions in the data term and, sharpness in the smoothness term. For both cases we use $\Psi(r^2) = \sqrt{(r^2 + \epsilon^2)}$, where ϵ is a small (in our experiments $\epsilon = 10^{-4}$) term for stabilizing the function when r gets close to zero.

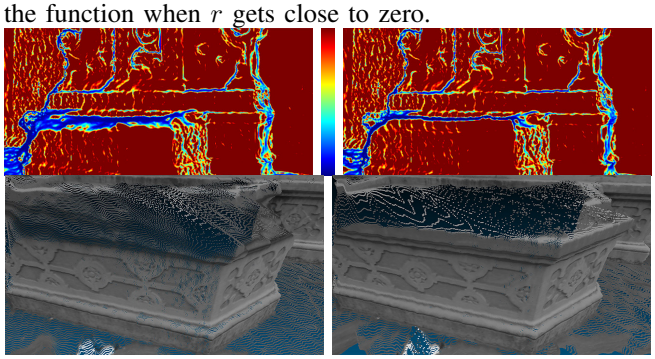

Figure 3. Color coded visualization of the quality of reconstructed triangles (projected into the image plane), using the classical formulation (top-left), and our proposed method (top-right). Red and blue represent best and worst quality resp. Closeups to reconstructions of the fountain's base are shown at the bottom. Please note the crisper rim to the right.

B. Distortion driven variational matching

Commonly, the robust regularizer is used to avoid smoothing across discontinuities which can range from small details to large surface jumps. While it definitely helps curb down smoothing effects at discontinuities, it falls short in many cases, e.g. the bleeding problem in figure 3. To address this shortcoming, we introduce spatial geometric information to steer grid lines away from largely sheared locations towards more meaningful regions. In this manner, the bleeding problem is addressed indirectly by means of a simple geometric characterization, without requiring any intricate problem reformulation. The idea is to sequentially measure the qualities of the 3D reconstructed triangular elements and then re-inject these measurements to drive the matching computation.

For this purpose, we need first to define a geometric characterization of the distortion of each quadrilateral cell. Since triangles represent the simplest element for describing a map distortion, we decompose the quadrilateral cell into two triangles as shown in figure 4. We found in our experiments that the condition number c of the jacobian matrix $\mathbf{J}$ of the transform which maps the planar triangle to its spatial counterpart works well in general. Figure 3-top illustrates how this measure clearly captures the regions of interest, in this case, regions with low visibility and relatively high depth variation (shown in blue).

In the current setting, the condition number of the jacobian

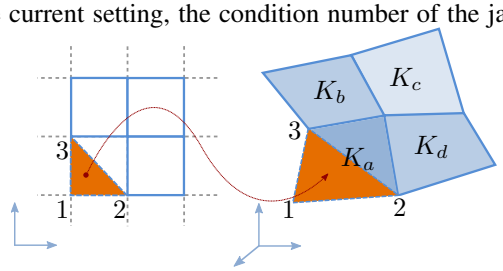

Figure 4. We decompose the quadrilateral cells into triangles to measure their distortion.

of the transformation can be explicitly expressed according to [23] as $c = c(\mathbf{J}) = (l_{12}^2 + l_{13}^2)/2A$, where l_{12} and l_{13} are the lengths of the reconstructed edges $\{1, 2\}$ and $\{1, 3\}$ resp. and A is the reconstructed triangle area (see figure 4). We define the distortion measure k for a given cell as the sum of the distortion of its two sub-triangles weighted by their respective inverse areas. That is $k = \frac{c_1}{A_1} + \frac{c_2}{A_2}$.

In order to drive the optimization process, we propose to replace the standard Laplacian operator $div(grad)$ within the robust function in the smoothing term by an anisotropic operator, namely the more general quasi-harmonic operator $div(k\ grad)$. The discretization of the new operator remains similar to the discretization of the smoothing term in equation 1, but with small modifications. For instance, for the central vertex in figure 4, the discrete contribution of u in the regularization term becomes $-\beta(r_S u_{y_S} - r_N u_{y_N} + r_E u_{x_E} - r_W u_{x_W})$; where the weights r_N, r_S, r_E, r_W are associated to the derivatives at each side (North, South, East, West) of the vertices in the classical way, see e.g. [13]. In detail, r_S is given by $r_S = (k_a + k_b)\Psi'_S$, with $\Psi'_S = \Psi'_S(x, y) = (\Psi'(x, y) + \Psi'(x, y + 1))/2$. The terms r_N, r_E and r_W are defined in a similar manner. The contribution of v for the same term can be derived in the same way as for u. In order to maintain a close check on problematic regions the modified flow formulation needs to intervene at each level of the coarse to fine optimization. So at each level of the pyramid we compute the classical solution, measure the induced spatial distortion and then re-inject it into the modified anisotropic formulation. The resulting solution is then transferred to the next level. This leads to the following variational matching algorithm 1

A typical result using the proposed algorithm is shown in figure 3-right. The effect of the bleeding problem is clearly reduced to a thin area around the surface jump as substantiated by the projected quality of the triangles (figure 3-top-right), and the crisper fountain rim (figure 3-bottom-right).

Algorithm 1 Coarse to fine variational matching of images I_1 and I_2 using their corresponding projection matrices.

Start at the coarsest level of the pyramid with $w_g \leftarrow 0$
while finest level of the pyramid is not yet solved **do**
1. Using w_g as initial state, find a solution w_o using equation 1
2. Compute the 3D positions for the current solution
3. Obtain the values of k
4. Obtain w_g using the variational matching with the distortion driven smoothing operator (using w_o as initial state)
5. Move to a finer level in the pyramid
end while

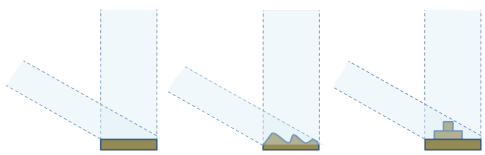

Figure 5. Illustration of the area change across two different views in the case of a flat surface (left). This effect becomes more pronounced for smooth surfaces with visible curvature change (middle), and non-smooth surfaces (right).

III. Distortion driven multiple-view merging

Our goal, when merging the matches from multiple views, is to select what is best viewed in a given subset of images (triplets and pairs). For this purpose, we rely on Lambert's cosine law which indicates that the image area of a captured scene region increases when the line of sight is closer to a perpendicular configuration. This is illustrated on figure 5 for the cases of flat surfaces (left), smooth surfaces with visible curvature change (middle), and for non-smooth surfaces (right). For the last two cases this area variation becomes more pronounced.

In practice, we look at the induced distortion when mapping across different views. A simple measure is the local change of the signed triangle area (a negative area tells when the mapping exhibits triangle flips). While this can be a sufficient measure in many cases, it fails to capture deformations which do not affect triangle areas (authalic maps). In general, a planar transformation S which maps a triangle $\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}\}$ onto $\{\mathbf{q_1}, \mathbf{q_2}, \mathbf{q_3}\}$ can be characterized by its Jacobian $\mathbf{J} = \left(\frac{\partial S}{\partial x}, \frac{\partial S}{\partial y}\right)$, where the partial derivatives are obtained by means of the divergence theorem:

$$\frac{\partial S}{\partial x} = \frac{1}{A}\int_A (\frac{\partial S}{\partial x}) dA \simeq \frac{\mathbf{q}_1(y_3 - y_2) + \mathbf{q}_2(y_1 - y_3) + \mathbf{q}_3(y_2 - y_1)}{A}; \quad (2)$$

where A is the triangle area and $\frac{\partial S}{\partial y}$ can be obtained similarly.

One way to characterize the deformation of the triangle is by looking at the square roots of the eigenvalues λ_1 and λ_2 of $T = J^T J$, which describe the well known elongation along the principal axis when mapping a unit circle onto an ellipse, see figure 6. This eigen-ratio ($\sqrt{\lambda_1/\lambda_2}$) indicates the amount of induced shearing. Combining this measurement with the signed area, we can characterize the overall distortion of the triangles.

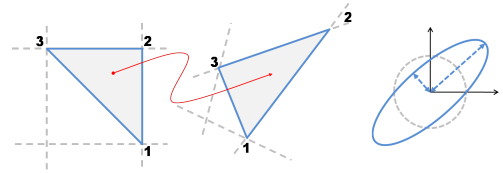

Figure 6. Mapping of a (triangulated) regular grid into a deformed configuration(left). Visualization of the deformation by means of its action on a circle (right).

Using the characterization just described and standard tools from multi-view geometry [24], our merging approach proceeds in two stages. In a first stage, we extract matches which are best viewed in triplets of neighboring views. In a second stage we use best viewed matches in pairs of views to recover regions which were not reliably captured by triplets. For the sake of clarity, let us consider the following scenario, where cameras are ordered in a daisy chain fashion, see figure 7. This setting is not a restrictive as it can always be arranged for by measuring the angle between the cameras' principal directions [10].

A. Triplet contributions

Without loss of generality, let us consider an image triplet $\{I_b, I_c, I_d\}$ and its direct neighbors I_a and I_e, see figure 7. First, we compute the dense variational matching from I_c to I_b and from I_c to I_d. In both computations, the grid corresponding the middle image I_c remains fixed. This comes in handy as it yields a direct correspondence between the three images. Next, the matches $\{I_c, I_b\}$ are transferred to I_d using the trifocal tensor [24] and then the distances between the transferred points and the ones obtained directly from the matching $\{I_c, I_d\}$ are measured. As the computation is not symmetric, we perform the same operation in the other direction as shown in figure 7. Only matches with an error below a threshold ϵ_{trif} are kept. In all our experiments we found that setting this parameter to 1 (pixel) gave reasonable results. We perform this operation on all triplets (defined by neighboring views) in the image set.

This validation can yield very large data sets. For instance the fountain data set [5] results in more than 60M points. To obtain reliable information out of such large number of points, we observe that errors in the matching manifest themselves as overlaying and interlaying surface sheets, as shown in figure 2. The offset between surface sheets can be

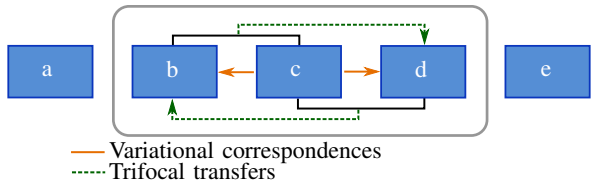

Figure 7. Triplet matches are validated based on trifocal transfer (dotted green arrows) within the triplet.

very tight at places and disturbingly large at others. In order to address this problem, we take advantage of the distortion measure we described previously. We mark a triangle as best viewed in an image triplet if its area decreases and its eigen-ratio deteriorates in neighboring views. To do so, we transfer the matches $\{I_c, I_b\}$ and $\{I_c, I_d\}$ to neighboring images I_a and I_e resp. as shown in figure 8. By inspecting triangles whose areas decay, we can identify which triangles are better viewed in the triplet $\{I_b, I_c, I_d\}$.

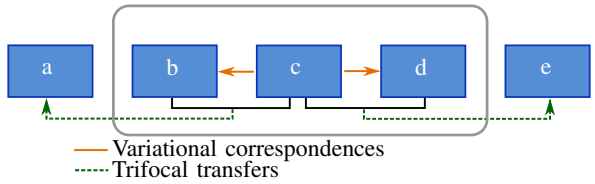

Figure 8. The selection of best viewed regions in a triplet is performed by transferring the grid to neighboring views (dotted green arrows) and analyzing its distortion.

Additionally, triangles which pass the area test, need also to satisfy the low distortion measure. We impose $\sqrt{\lambda_1/\lambda_2} > 1/r_{eig}$ and $\sqrt{\lambda_1/\lambda_2} < r_{eig}$ as indicators for acceptable distortion and proceed similarly to the signed area case. In all our experiments we found that setting r_{eig} to 2 works well in practice (the optimal eigenvalue ratio is 1).

Lastly, we need to account for points which disappear outside of the triplet as we move to neighboring views. This can be done by simply checking for points which get out the image range when transferred to I_a and I_e or triangles whose areas get close to zero. Such regions are marked in yellow in figure 10.

It could be argued that the trifocal constraints can be directly incorporated within the matching formulation (e.g. [25]) and, although, this can improve the results in some regions, it turns out to be more problematic for areas seen only by two views. Because of this, a lot of information is not recovered at particularly interesting regions and, this shortcoming becomes stronger as the baseline increases.

B. Pair contribution

Certainly, all points are not necessarily visible in triplets or do not score high enough during triplet validation. Therefore, we still need to account for points visible mainly in pairs of views so as not to miss some important features. In a similar manner to triplets, we need to validate the matches. We have two measures at hand. The first is the distance to the epipolar line, and the second is by means of the forward backward map [1], see figure 9.

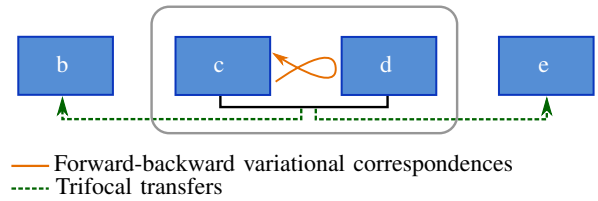

Figure 9. Pair matches are validated by epipolar distance and a forward-backward map. The selection of best viewed regions in a pair is performed by transferring the grid to neighboring views and analyzing its distortion. Only regions which where not covered by triplets are considered.

In order to handle visibility for a pair of images, we keep only triangles which exhibit an area decrease and satisfy the eigen-ratio criteria when transferred to neighboring views. Furthermore, we include points which disappear and triangles whose areas get close to zero. Despite these checks, some outlier triangles can pass all the filters. A simple but reliable way for sorting out these cases, is to transfer the triangles to a slightly faraway view and then measure their distortion, e.g. the pair I_a, I_b can be transferred to I_e.

Figure 10. Processing best viewed regions in the Fountain data set [5]. Each view represents the central image of a triplet (other two images not shown). The red-colored regions represent areas best viewed in the triplet. Yellow regions represent regions that are only visible in the triplet and therefore are included even if they do not comply with the best view requirement.

IV. Results and discussion

We tested our method on several data sets comprising standard benchmarks as well as in-house acquired data. For the reconstruction, either the original pointclouds are shown or the textured reconstructions using [26]. Typical results of our approach are shown for the full sized Fountain-P11 (figure 11), and for the Herz-Jesu-P8 data sets (figure 12). Both data sets are available from [5]. We benchmarked our results following the method detailed in [5]. In figures 11 and 12 we show our results obtained and their corresponding error distribution histograms. In these figures the red color represents locations where no result were obtained or locations where result are farther than 30σ. In green are represented regions where no results can be obtained; dark gray represents larger errors (smaller than 30σ).

In the following table we compare the level of completeness and relative errors of our results with those of Furukawa and Ponce [6] and Keriven et al [7] (Only completeness is available at the time of writing).

	Fountain			Herz-Jesu		
	Ours	[6]	[7]	Ours	[6]	[7]
Relative error	1.76	2.04	NA	2.11	2.98	NA
Complete (%)	85.0	79.6	90.8	81.8	80.4	91.1

For the Fountain-P11 the variational matching computation requires about 10 minutes for a pair. For the same example, which contains eleven $3,072 \times 2,048$ images, the merging of the correspondences required only 215s in a Matlab implementation.

We also tested our approach on our own datasets. Results on a face dataset composed of six 1.3 mega-pixels images are shown in figure 15. Our approach clearly fares better than the classical formulation, especially at problematic regions such as the hair, ears, chin and neck. Figure 13 shows our results on a different face dataset comprising nine 8 mega-pixels images. In this example, our method does a particularly better job reconstructing important features such as the nose and the ears. Lastly, in figure 14 we show results for the reconstruction of a boot. This dataset consists of five 4 mega-pixels images. We can observe the chipped boot collar when our distortion driven matching method is not used.

As our approach operates on local neighboring views, its memory requirements are low. In fact, we do not require loading all the spatial data at once for processing as it's generally done in related work [6], [7]. In all our experiments, we observed that our merging strategy allows reducing the raw variational matching data by up to a 75%.

We are aware that the proposed merging method is sub-optimal, however we do not see this as a limitation. The locality and low computational requirements of this approach allows reducing the data at hand significantly at a fraction of the computational cost of methods based on global optimization (e.g. [7], [12]). In particular, the results of our approach can be directly used in [7].

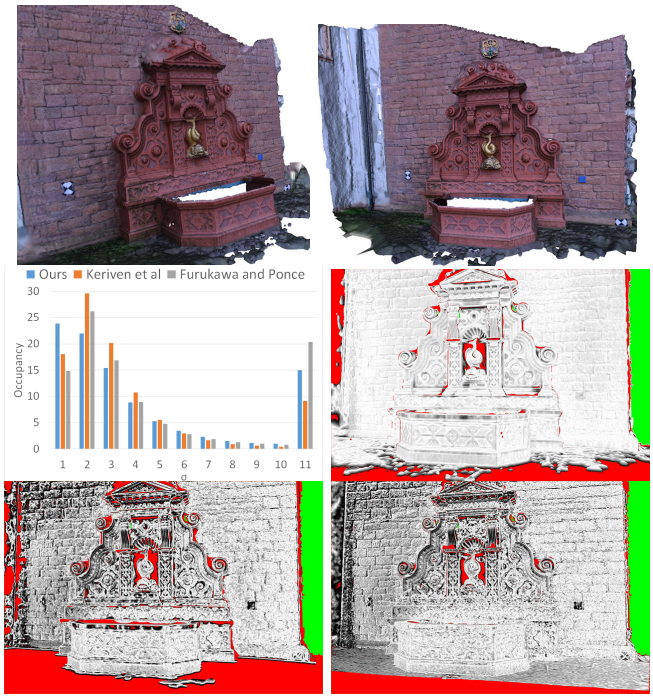

Figure 11. Top row, result of our approach on the Fountain-P11 data set [5]. Middle-left, a histogram of the error accumulated for all the views obtained with our method, [6] and [7]. Middle-right, our depth estimation error for view five. At the bottom, same evaluations using [6] and [7].

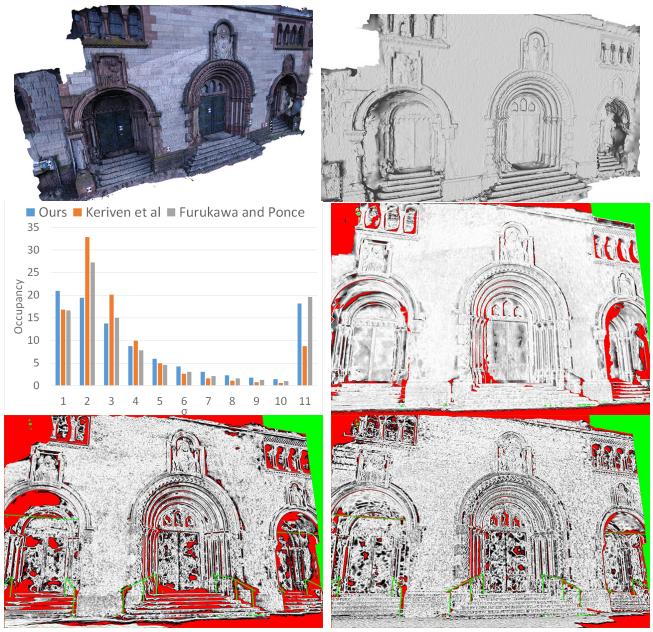

Figure 12. Top row, result of our approach on the Herz-Jesu-P8 data set [5]. Middle-left, a histogram of the error accumulated for all the views obtained with our method, [6] and [7]. Middle-right, our depth estimation error for view four. At the bottom, same evaluations using [6] and [7].

V. Conclusions

In this paper, we proposed two systematic enhancements to the classical variational scene reconstruction. For the variational matching, an adaptive approach allows recapturing lost details by means of an anisotropic diffusion driven by geometric distortion. At the merging level, a selective technique for obtaining the best contributions across neighboring views allows reducing data redundancy. Unlike most of related work, the approach resolves a large number of outliers cases prior to estimating the final spatial point cloud.

All of our contributions are achieved by means of a simple, yet principled, characterization of geometric deformations and, can be easily reused in other methods within the variational context. Our approach is fairly straightforward, simple to implement, and our results can be easily reproduced.

In future work we can explore the idea of having adaptive thresholding values for the merging stage described in section III since they are currently defined globally.

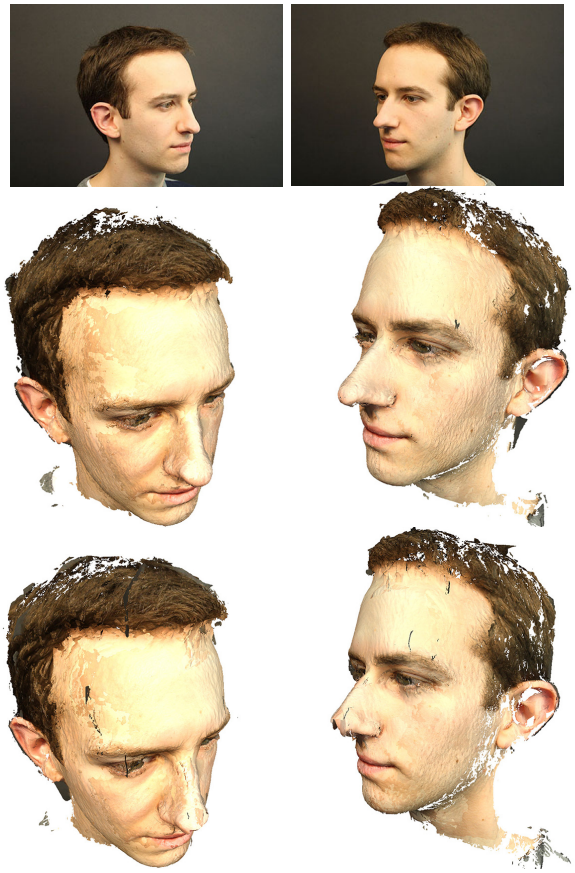

Figure 13. Two images from face1 dataset-9×8 MP-(top) reconstructed using our approach (middle) and without the use of our distortion driven matching (bottom). Note the problems in the nose and around neck and ears for the second case.

Figure 14. Two images from our boot dataset-5×4 MP- (top) reconstructed using our approach (middle) and without the use of our distortion driven matching (bottom). Note the problem near the boot collar for the second case.

Figure 15. Two images from face2 dataset-6×1.3 MP-(top) reconstructed using our approach (middle) and using variational matching code provided by the authors of [22].

Acknowledgment

The authors would like to thank the anonymous reviewers for their feedback on the paper, Christoph Strecha for his Multi-View datasets and evaluations and Samuel Hornus for the face2 dataset. This work was funded by the ANR (Agence Nationale de la Recherche) under grant (PhysiGrafix ANR-09-CEXC-014-01).

References

[1] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert, "Dense disparity map estimation respecting image discontinuities: A {PDE} and scale-space based approach," *Journal of Visual Communication and Image Representation*, vol. 13, no. 1 - 2, pp. 3 – 21, 2002.

[2] H. Zimmer, A. Bruhn, L. Valgaerts, M. Breuß, J. Weickert, B. Rosenhahn, and H.-P. Seidel, "PDE-based anisotropic disparity-driven stereo vision," in *VMV*, 2008.

[3] C. Strecha and L. Van Gool, "PDE-based multi-view depth estimation," in *Proc. 3D Data Processing Visualization and Transmission*, 2002, pp. 416–425.

[4] C. Strecha, T. Tuytelaars, and L. Van Gool, "Dense matching of multiple wide-baseline views," in *ICCV*, 2003, pp. 1194–1201 vol.2.

[5] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *CVPR*, 2008, pp. 1–8. [Online]. Available: http://cvlabwww.epfl.ch/ strecha/multiview/denseMVS.html

[6] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in *CVPR*, june 2007, pp. 1 –8.

[7] V. H. Hiep, R. Keriven, P. Labatut, and J.-P. Pons, "Towards high-resolution large-scale multi-view stereo," in *CVPR*, 2009, pp. 1430–1437.

[8] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz, "Multi-view stereo for community photo collections," in *ICCV*, 2007, pp. 1–8.

[9] R. Tylecek and R. Sara, "Depth map fusion with camera position refinement," *Proc. CVWW*, pp. 59–66, 2009.

[10] E. Tola, C. Strecha, and P. Fua, "Efficient large-scale multi-view stereo for ultra high-resolution image sets," *Mach. Vision Appl.*, vol. 23, no. 5, pp. 903–920, Sep. 2012.

[11] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, 2008.

[12] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l 1 optical flow," in *Pattern Recognition*. Springer, 2007, pp. 214–223.

[13] J. Sanchez, N. Monzon, and A. Salgado, "Robust Optical Flow Estimation," *Image Processing On Line*, vol. 3, pp. 252–270, 2013.

[14] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sánchez, "Symmetrical dense optical flow estimation with occlusions detection," in *Computer VisionECCV 2002*. Springer, 2002, pp. 721–735.

[15] J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi, "Bilateral filtering-based optical flow estimation with occlusion detection," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 211–224.

[16] C. Kondermann, D. Kondermann, B. Jähne, and C. Garbe, *An adaptive confidence measure for optical flows based on linear subspace projections*. Springer, 2007.

[17] O. Mac Aodha, A. Humayun, M. Pollefeys, and G. J. Brostow, "Learning a confidence measure for optical flow," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 5, pp. 1107–1120, 2013.

[18] H.-H. Vu, P. Labatut, J.-P. Pons, and R. Keriven, "High accuracy and visibility-consistent dense multiview stereo," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 5, pp. 889–901, 2012.

[19] C. Zach, T. Pock, and H. Bischof, "A globally optimal algorithm for robust tv-l 1 range image integration," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[20] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *CVPR*, 2006.

[21] D. Gallup, J.-M. Frahm, P. Mordohai, and M. Pollefeys, "Variable baseline/resolution stereo," in *CVPR*. IEEE, 2008, pp. 1–8.

[22] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *ECCV*, 2004.

[23] P. M. Knupp, "Algebraic mesh quality metrics for unstructured initial meshes," *Finite Elements in Analysis and Design*, vol. 39, no. 3, pp. 217 – 241, 2003.

[24] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[25] Z. Sun, "A three-frame approach to constraint-consistent motion estimation," in *ICPR*, vol. 1, 2006, pp. 35–38.

[26] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Eurographics symposium on Geometry processing*, 2006.