PDE Based Surface Estimation for Structure from Motion

H. Aanæs, R. Larsen, and J. A. Bærentzen

Department of Informatics and Mathematical Modelling,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark
{haa, rl, jab}@imm.du.dk
http://www.imm.dtu.dk/image

Abstract

The method of Faugeras and Keriven for solving the multiple view stereo problem by partial differential equations (PDE's) in a level set implementation has been well received. The main reasons are that it produces good results and deals effectively with objects which are, topologically, more complex than a ball. This makes it a good choice for the final step in the usual structure from motion approach, namely making a full 3D surface reconstruction from estimates of camera orientations and 3D point structure.

Here an approach is proposed whereby the 3D point structure, estimated using structure from motion, is used to initialize the method of Faugeras and Keriven. The proposed approach has the advantage of a) considerably improving the run time of the approach, and b) making it more resistant towards noisy data and data with patches of low variability. Both advantages make the approach much more effective on real data.

1 Introduction

The estimation of surface models from multiple images, the so–called multiple view stereo problem, is one of the classical problems within computer vision (e.g. [4,8,10,16]). Some of the best results are achieved by Faugeras and Keriven by posing the problem as a partial differential equation PDE [2,3,4], this has been further developed in [7]. Some of the merits of this approach are, that it in a naturally way employs all the images simultaneously as opposed to the adapted 2 image stereo approaches (e.g. [10,16]) and that it is capable of dealing with objects of arbitrary topology.

Another major problem of computer vision is the reconstruction of structure and motion of a rigid object from an image sequence, the so–called structure from motion problem. Here the usual approach is to first restrict the 3D structure estimation to a few distinct points whereby the structure is represented as a point cloud (e.g. [6]). As a part of this point based structure from motion the camera calibration is also determined – outer and often also inner orientation. But for many applications of structure from motion a full 3D model is needed, and hence a multiple view stereo approach is natural (e.g. [5,14]).

We propose using the PDE based surface estimation approach of Faugeras and Keriven [4] to solve the surface estimation problem from the structure and motion solution based on points. The contribution of this work is altering the approach by using the structure estimates, in the form of the 3D points, to much better initialization of the PDE based surface estimation algorithm. This has the effect of reducing the running time of the algorithm considerably. On a standard 1 GHz PC the order of magnitude is from whole and half days to an hour or often less. Secondly, the proposed approach also renders the algorithm much more resistant to noisy and/or erroneous data, as well as to objects with patches of low variance. Both these issues make the approach more effective with real data.

2   PDE Based Surface Estimation

As a courtesy to the reader and to introduce notation, a short overview of PDE based surface estimation is presented. For a more through introduction the reader is referred to [2,3,4,7].
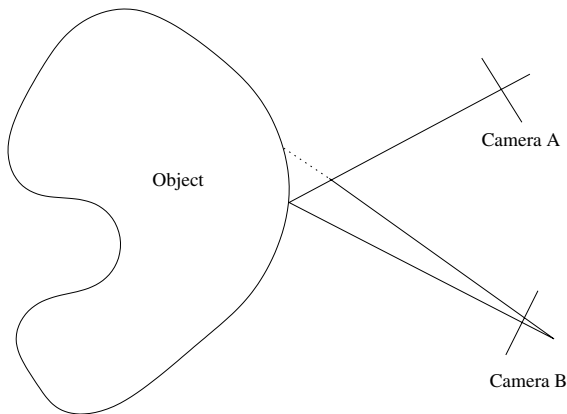

Fig. 1. Based on the Lambertian assumption, the projection of a given surface point, should look the same in different images. A property that will, in general, not hold for image points being projections of different 3D points.

The main idea behind PDE based surface estimation is illustrated in Figure 1, namely that the projection of points on the physical surface should look the same in all images. The implication is that the true physical surface, $\mathcal{S}^*$, is the minimizer of:

$$\mathcal{S}^* = \arg \min_{\mathcal{S}} \int_{\mathcal{S}} \Phi(\mathbf{x}) dA \ , \tag{1}$$

where $\Phi(\mathbf{x})$ is a similarity measure of the projection of $\mathbf{x}$ in all the cameras. The *basic* similarity measure used in [2,3,4,7] and here is; given neighborhoods of the projection

of $\mathbf{x}$ in camera i, denoted $\mathcal{N}_i$. Then

$$\Phi(\mathbf{x}) = \frac{1}{\sum_{i \neq j} v_{ij}(\mathbf{x})} \sum_{i \neq j} \rho(\mathcal{N}_i, \mathcal{N}_j) \cdot v_{ij}(\mathbf{x}) \ , \tag{2}$$

where $\rho(\cdot, \cdot)$ denotes the correlation and $v_{ij}(\mathbf{x})$ is an indicator function denoting whether $\mathbf{x}$ is visible in both image i and j or not.

The PDE determining the gradient flow of $\mathcal{S}$ is then:

$$\frac{\partial \mathcal{S}}{\partial t} = -\nabla \left(\Phi(\mathbf{x}) \right)^T \cdot \mathbf{n} + \alpha \mathcal{C} \ , \tag{3}$$

where $\mathbf{n}$ is the local normal to the surface $\mathcal{S}$, $\mathcal{C}$ is the local curvature, which is basically a second order smoothing term. The constant α determines the amount of smoothing imposed. Since the neighborhood patches, $\mathcal{N}_i$, are warped according to the local orientation of the surface at $\mathbf{x}$, a term dependent on $\mathbf{n}$ should also be present in (3). As noted in [7], this missing term is so similar to the local curvature that it can be dropped[1]. This is done here.

It should be noted, that Jin et al. [7] use the median instead of the mean in (2). Hereby an approach which is robust towards a break–down of the Lambertian assumption is obtained.

2.1   Implementation

The PDE of (3) is used to optimize (1) in a level set framework [11,12,15]. An important issue in this regard, is how to initialize the optimization scheme. The usual initialization scheme [2,3,4,7], is to use a bounding ball or box, which contains the surface with probability 1 (c.f. Figure 2).

An issue with this approach is it speed. The evaluation of a given voxel of the level set grid is very costly, in that each patch, $\mathcal{N}_i$, should be warped in all images, and pairwise correlation between these patches should be estimated. Due to the warping, pre–calculation of intermediate results is infeasible. Hence, if a large change of volume is required, as illustrated in Figure 2, many costly evaluations are needed, and the approach will be very time consuming. In order to address this problem it is proposed [4] to add an inward force to (3). This inward force is set to zero if $\Phi(\mathbf{x}) > \Phi_{\max}$, such that it is 'turned off' when the fit is 'good'. This has the advantage of speeding up the approach, and as far as we can see, it also help avoid local minima, in that at such local minima it can be assumed that $\Phi(\mathbf{x}) < \Phi_{\max}$. Hence (3) is in reality changed to:

$$\frac{\partial \mathcal{S}}{\partial t} = - \left(\nabla \Phi(\mathbf{x}) \right)^T \cdot \mathbf{n} + \alpha \mathcal{C} + \mathcal{G}\left(\Phi(\mathbf{x})\right) \ ,$$

where $\mathcal{G}\left(\Phi(\mathbf{x})\right)$ is the inward force. Even with this inward force, an optimization is not unlikely to take the better part of a day on a standard PC, due to the large change required in $\mathcal{S}$.

---

[1] This slightly changes the interpretation of α.

The inward force $\mathcal{G}(\Phi(\mathbf{x}))$ has some unintended side effects, since $\Phi(\mathbf{x})$ can be small on the true optimal surface $\mathcal{S}^*$. This fact is usually caused by image noise and/or a low variability on part of the object in question. In cases of small $\Phi(\mathbf{x})$ on the true surface, $\mathcal{S}^*$, $\mathcal{G}(\Phi(\mathbf{x}))$ will cause the evolving surface, $\mathcal{S}$, to pass through $\mathcal{S}^*$. Note that

$$\alpha\mathcal{C} + \mathcal{G}(\Phi(\mathbf{x})) \ ,$$

acts as a prior on the surface, and as such determines the behavior of the algorithm when $\nabla\Phi(\mathbf{x})$ is small (c.f. Figure 4).
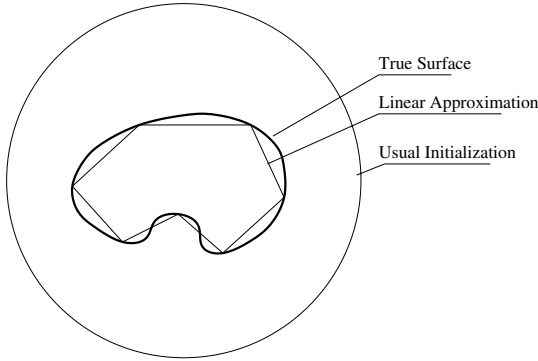

Fig. 2. The true surface and different initializations of the surface before optimization. It is seen, that the Area – or Volume in 3D – to be traversed for convergence is usually much smaller for the linear approximation.

3   Utilizing the 3D Point Structure

In the usual structure from motion setting, the camera orientations and a 3D point structure is estimated first. These estimates are then used as the basis of a full 3D surface model estimation. If the method of [2,3,4,7] was used for this, only the camera orientation would be applied. Hence the information from the already estimated structure is discarded.

We here propose using this 3D point structure by forming a much better initial guess. This is done by applying the method of Morris and Kanade [9], to make an optimal triangulation of the 3D point structure already at hand. This triangulated mesh is then used as an initialization for the level set. Since the elements of the 3D point structure are assumed to lie on the surface, this mesh can be seen as a piecewise linear approximation to the surface (c.f. Figure 2). A practical consideration in this regard is how to convert the mesh to a signed distance field, the latter being an initialization for the level set method. This is done with a modified version of [13] described in [1], whereby the signed distance from each voxel to the mesh is calculated. The sign is positive if the voxel is outside the mesh, and negative if it is inside.
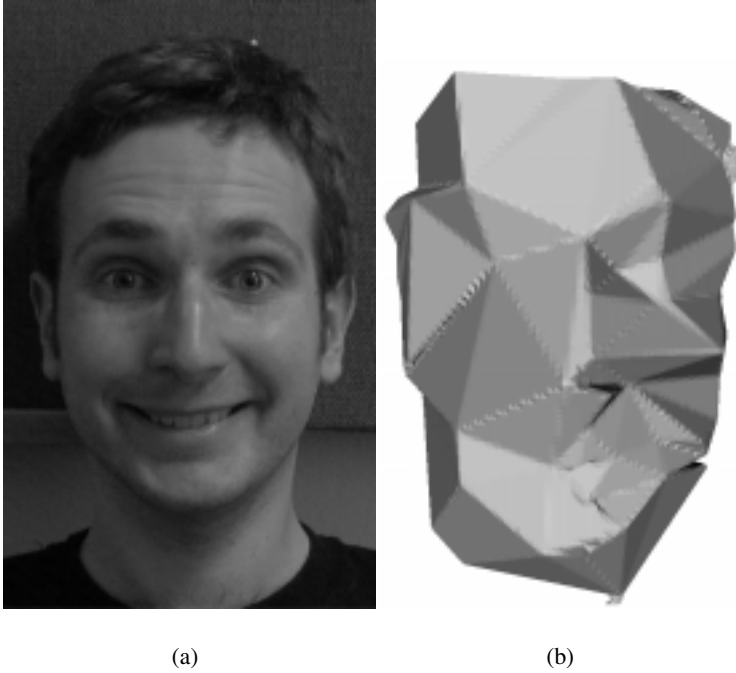
(a)   (b)

Fig. 3. (a) A sub image from the face sequence of 5 images. The subject moved slightly during the taking of the images. (b) The initialization used with the face data set, when the proposed method is applied.

The first advantage of this approach is that the initial guess is very likely to be much closer to the true surface, c.f. Figure 2. Hence the run time is reduced considerably, in that smaller changes in $\mathcal{S}$ is need . In our experiments the order of magnitude of speed up is from whole to half days to approximately 1 hour.

Secondly, the proposed approach has the advantage of not needing the inward force $\mathcal{G}(\Phi(\mathbf{x}))$. Hence the prior on the surface becomes:

$$\alpha \mathcal{C} \quad . \tag{4}$$

The difference is that if the data does impose a force on $\mathcal{S}$, due to the above mentioned reasons, then it should try to smooth out instead of go inward. This makes the approach much more resistant to image noise and surface parts with low variability.

4   Results

To validate the proposed approach, we used 5 images of a face as illustrated in Figure 3(a). This data set was noisy, in that it is unlikely that the subject was completely still. Secondly, there are many patches with very low or no variance. As such this is a
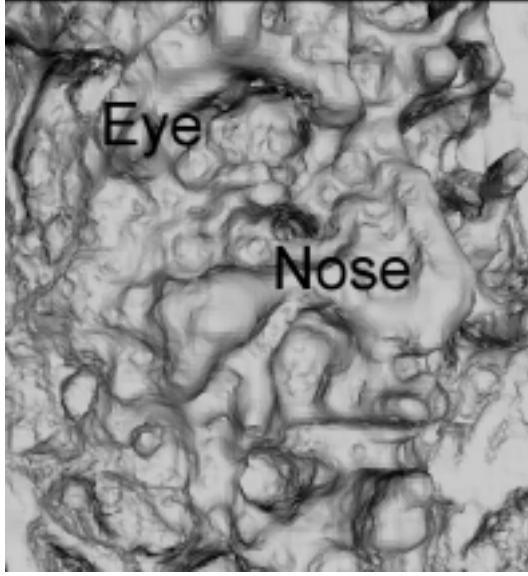
Fig. 4. An intermediate iteration in face sequence, when using an inward force, $\mathcal{G}$. It is seen how the surface has 'gone through' the true surface, $\mathcal{S}^*$. In later iterations the holes will get bigger, and eventually the smoothness constraint of (3) will pull the surface from $\mathcal{S}^*$, whereupon it will collapse under its own curvature.

rather challenging data set for a surface reconstruction algorithm, but it is by no means below the standard of what is expected in a structure and motion setting. The quality of the data also makes a solution with an inward force, $\mathcal{G}$, infeasible as seen in Figure 4. It is seen how the surface has 'gone through' the true surface, $\mathcal{S}^*$. In later iterations the holes get bigger, and eventually the smoothness constraint of (3) will pull the surface from $\mathcal{S}^*$, whereupon it will collapse under its own curvature.

The proposed approach was also applied. It was initialized with the mesh depicted in Figure 3(b). This mesh is optimally triangulated based on a 3D point structure estimated by structure from motion from a series of automatically identified landmarks. The result is seen in Figure 5(a), where it is noted that the algorithm converges to an acceptable result, despite the quality of the data.

To improve the results, the use of more advanced regularization was investigated. It turned out that most of the problematic data was at low variance patches. Hence, it was tried to smooth patches, $\mathcal{N}_i$, with low $\varPhi(\mathbf{x})$ more. From a histogram of $\varPhi(\mathbf{x})$, it was deducted that 0.5 was a good cut off. Hence (3) was modified to

$$\frac{\partial \mathcal{S}}{\partial t} = \begin{cases} -\nabla \varPhi(\mathbf{x}) \cdot \mathbf{n} + \alpha \mathcal{C} & \varPhi(\mathbf{x}) \geq 0.5 \\ -0.5 \cdot \nabla \varPhi(\mathbf{x}) \cdot \mathbf{n} + 1.5 \cdot \alpha \mathcal{C} & \varPhi(\mathbf{x}) < 0.5 \end{cases} . \tag{5}$$

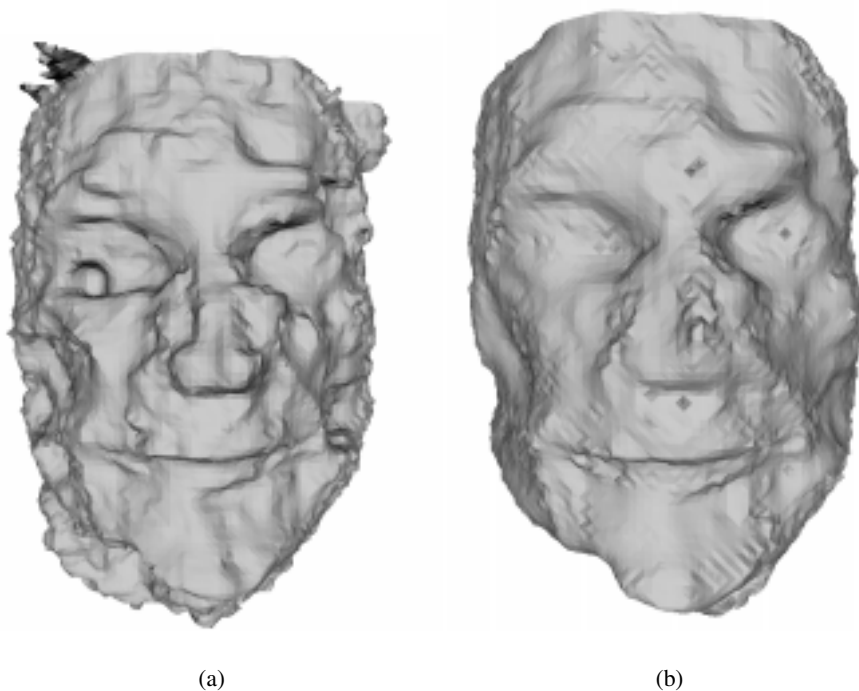<center>(a)   (b)</center>

Fig. 5. (a) Proposed method on the face data set, see Figure 3(a). (b) Proposed method on the face data set, see Figure 3(a). Here the extended regularization of (5) was used.

The result is seen in Figure 5(b). In the results of Figures 5(a) and 5(b) α was set to 0.25. Here it is seen that this extended regularization improves the result, implying that extended regularization is a fruitful path.

5   Discussion

A new approach for PDE based surface estimation has been presented for use in the usual structure from motion framework. This approach uses the estimated 3D point structure to initialize the optimization, whereby a significant speed up and resistance to poor data is achieved. Both these issues are vital when performing structure and motion on real data.

Acknowledgments

References

1. J. A. Bærentzen and H. Aanæs. Computing discrete signed distance fields from triangle meshes. Technical Report 21, IMM, DTU, 2002.
2. O. Faugeras and R. Keriven. Variatinal principles, surface evaluation, pde's, level set methods and the stereo problem. Technical Report 3021, INRIA, October 1996.
3. O. Faugeras and R. Keriven. Complete dense stereovision using level set methods. *Computer Vision - ECCV'98. 5th European Conference on Computer Vision. Proceedings*, pages 379–93 vol.1, 1998.
4. O. Faugeras and R. Keriven. Variational principles, surface evolution, pdes, level set methods, and the stereo problem. *Image Processing, IEEE Transactions on*, 7(3):336 –344, 1998.
5. O. Faugeras, S. Laveau, L. Robert, G. Csurka, and C. Zeller. 3-d reconstruction of urban scenes from sequences of images. Technical Report 2572, INRIA, June 1995.
6. R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, UK, 2000.
7. H. Jin, A. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. Technical Report TR01-0017, UCLA, 2001.
8. K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1:307–314 vol.1, 1999.
9. D. Morris and T. Kanade. Image-consistent surface triangulation. In *IEEE Conf. Computer Vision and Pattern Recognition'2000*, pages 332–338, 2000.
10. M. Okutomi and T. Kanade. A multiple-baseline stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(4):353–363, 1993.
11. S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.
12. S. J. Osher and R. P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer Verlag, 1st edition, November 2002.
13. B. A. Payne and Arthur W. Toga. Distance field manipulation of surface models. *Computer Graphics and Applications*, 12(1), 1992.
14. M. Pollefeys. *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. PhD thesis, K.U.Leuven, 1999.
15. J.A. Sethian. *Level Set Methods and Fast Marching Methods Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, 1999.
16. M. Ziegler, L. Falkenhagen, R. Horst, and D. Kalivas. Evolution of stereoscopic and three-dimensional video. *Image Communication*, 14, 1998.