Recovery and Tracking of Continuous 3D Surfaces from Stereo Data Using A Deformable Dual-Mesh

Yusuf Sinan Akgul and Chandra Kambhamettu

Video/Image Modeling and Synthesis(VIMS) Lab
Department of Computer and Information Sciences
University of Delaware
Newark, Delaware 19716
{akgul|chandra}@cis.udel.edu
http://www.cis.udel.edu/~vims

Abstract

We propose a novel method for continuous 3D depth recovery and tracking using calibrated stereo. The method integrates stereo correspondence, surface reconstruction and tracking by using a new single deformable dual mesh optimization, resulting in simplicity, robustness and efficiency. In order to combine stereo correspondence and structure recovery, the method introduces an external energy function defined for a 3D volume based on cross-correlation between the stereo pairs. The internal energy functional of the deformable dual mesh imposes smoothness on the surfaces and it serves as a communication tool between the two meshes. Under the forces produced by the energy terms, the dual mesh deforms to recover and track the 3D surface. The newly introduced dual-mesh model, which is one of the main contributions of this paper, makes the system robust against local minima and yet it is efficient. A coarse-to-fine minimization approach makes the system even more efficient. Tracking is achieved by using the recovered surface as an initial position for the next time frame. Although the system can effectively utilize initial surface positions and disparity data, they are not needed for a successful operation, which makes this system applicable to a wide range of areas. We present the results of a number of experiments on stereo human face and cloud images, which proves that our new method is very effective.

1 Introduction

The first step of a traditional stereo analysis system is to extract a disparity map from the stereo image pair. The subsequent analysis steps use this disparity to perform their tasks. A few of the many examples of this approach are by Terzopoulos[17] and Blake and Zisserman[3], who fit a surface to a readily available depth information by minimizing a spline function, by Wildes[18] and Devernay and Faugeras[7], who calculate local differential properties from the disparity map. The common problem with these methods is that the main analysis, such as surface reconstruction and extraction of differential properties, is done separately from the extraction of 3D data, i.e., extraction of disparity map. This results in sequential systems where erroneous or noisy results of the first step have to be used in subsequent steps. In addition, the subsequent steps can not help the first step by feeding back additional constraints, such as smoothness, that can be useful in producing better results from the first step.

In order to address the above problem, there have been a number of proposals to integrate the main analysis phase with the extraction of 3D data. Hoff and Ahuja[11] and Fua[9] combine the steps of stereo matching and surface reconstruction. Kambhamettu et. al.[13] couple motion estimation analysis with stereo matching problem. Faugeras and Keriven[8] pose the stereo problem as a variational problem to drive partial differential equations, which are solved by level-set methods in a single step.

Following the research trend on integration of stereo correspondence and surface reconstruction, in this paper we present a novel method to unify stereo correspondence, continuous surface reconstruction and tracking at the same step using a deformable dual mesh. Although we find Hoff and Ahuja[11], Fua[9], and Faugeras and Keriven[8] closest to our work, our system is fundamentally very different in the assumptions and in the basic methods used.

The basic similarity between the above three systems and our system is that all four of them are formulated as an optimization framework. Hoff and Ahuja's surface reconstruction is based on fitting planar and quadratic patches after a matching process between the stereo pair. The results of matching are used as initial positions for

the planar and quadratic patches. Similarly, Fua's work has an initialization phase where a correlation process is used for stereo matching to determine initial local surface positions. In contrast, our system does not perform any explicit matching between the stereo pair and it does not need any initial mesh positions to start optimization. Another fundamental difference is the optimization methods used. Hoff and Ahuja use Hough transforms and standard least-square fitting as their main optimization tools. Fua uses conjugate gradient as the optimization tool, which requires derivatives of the objective function. On the contrary, we use a novel optimization method that is based on energy information flow between the two meshes, which is robust against local minima and is computationally efficient. In spirit, the system of Faugeras and Keriven[8] is the most similar one to our system. Both systems are formulated as one single step optimization without any initialization steps. In addition, both systems recover a smooth surface after this optimization. However, the optimization methods are fundamentally very different as will be explained in later sections.

Although we assume a continuous surface, we note that the ideas presented in this paper can be applied to discontinuous surfaces with proper modifications and additions, such as detecting the discontinuities first and applying our method on several 3D surfaces. Further details of the differences and contributions of our work will be mentioned in later sections.

Our system first forms a 3D array with cells representing the 3D spatial locations. This array is filled with correlation values as explained in section 2.2. The resulting 3D array is used by the potential energy in the deformable mesh energy formulation. Then we form two deformable meshes parallel to each other; one is at the nearest depth position (camera side in Figure 2) and the other is at the farthest depth position. The internal energy of the deformable mesh is used to impose smoothness and as a communication tool between the two meshes. The deformation occurs under the internal forces as well as the external forces produced by the 3D array. Our new minimization method guarantees that the dual mesh finds the same position by introducing additional forces that push the dual mesh towards each other. When the two meshes find the same position, we take the mesh element positions as the assigned depth values to the 3D surface. The system does not use or extract any explicit disparity data. However, the disparity data can be obtained at the end of the deformation without any computational overhead. This whole process is done in a coarse to fine scheme as explained in Section 3.2.

Tracking of the deformation of the surface in time is tackled by utilizing the usual deformable model tracking proposed by Kass *et. al.*[14]. We take the recovered surface from a time frame and use it as the initial surface position in the next time frame. Using our minimization method, the dual mesh is allowed to deform, thus recovering and tracking the 3D surface. Although our tracking does not achieve point correspondences between the tracked surfaces, there are advantages of using this method as explained in Section 4.

2 The Deformable Mesh

In this section, we give the formal definition and the energy functional for the deformable mesh. A deformable mesh, M, with l columns and k rows is a set of horizontally and vertically connected points in 3D space in a mesh form.

$$M = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1l} \\ m_{21} & m_{22} & \dots & m_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ m_{k1} & m_{k2} & \dots & m_{kl} \end{bmatrix}$$

where m_{kl} represents the element at the k^{th} row and l^{th} column. Each element of the mesh represents a point in the 3D space. The column and row of a mesh element will represent x and z positions. In addition, each mesh element will hold a depth value as its y dimension. See Figure 1 for the orientation of our coordinate system.

The energy functional of a mesh is written in terms of spatial positions of the two meshes, as our system is based on a dual mesh formulation. Given the two meshes N and M, the energy associated with the deformable mesh M is written as

$$E_{Mesh}(M,N) = \sum_{i=1}^{k}\sum_{j=1}^{l} \alpha E_{Smo}(m_{ij}) + \beta E_{Comm}(m_{ij}, n_{ij}) + \gamma E_{Ext}(m_{ij}) \quad (1)$$

where α, β, and γ are the weighting parameters, and n_{ij} is an element of the other mesh N.

2.1 Internal Energy

Internal energy of Equation(2) is the weighted sum of smoothness term and the communication term. We do not measure the continuity on the mesh because it is already assumed by the definition of the deformable mesh.

The smoothness energy term $E_{Smo}(m_{ij})$ is based on the summation of 3D dot vector products in both horizontal and vertical mesh directions and it is proportional to the angles between these 3D vectors. Formally,

$$E_{Smo}(m_{ij}) = \left(1 - \frac{\overrightarrow{m_{i-1j}m_{ij}} \cdot \overrightarrow{m_{ij}m_{i+1j}}}{|\overrightarrow{m_{i-1j}m_{ij}}|\,|\overrightarrow{m_{ij}m_{i+1j}}|}\right) + \left(1 - \frac{\overrightarrow{m_{ij-1}m_{ij}} \cdot \overrightarrow{m_{ij}m_{ij+1}}}{|\overrightarrow{m_{ij-1}m_{ij}}|\,|\overrightarrow{m_{ij}m_{ij+1}}|}\right).$$

For the mesh elements where the above formula is not valid, e.g, elements at the edge of the mesh, we use the smoothness term of the nearest mesh element that has a valid smoothness value.

The communication term $E_{Comm}(m_{ij}, n_{ij})$ is not always active. It is activated only if one or both of the meshes stop deforming without finding the same 3D surface. This is a mechanism to address attraction by local minima which is a serious problem in optimization methods. The details of this term will be explained in Section 3.1.

2.2 External Energy

The external energy is the only mechanism that links the deformation of the dual mesh to the stereo image pair. Given a mesh element position m_{ij} in 3D space, a smaller value of $E_{Ext}(m_{ij})$ indicates that m_{ij} is likely to be on a 3D surface.

$$E_{Ext}(m_{ij}) = 1 - V(x, y, z) \tag{2}$$

where x, y, z are the positions of m_{ij} on the main axes of 3D space. V is a 3D array that holds the correlation values that are computed as described below. Any

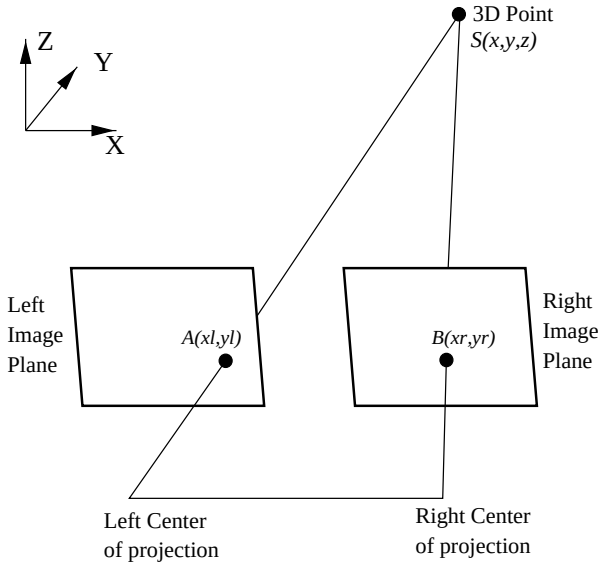

Figure 1: Projection of 3D scene points on the image planes

point $S(x, y, z)$ in 3D space that is visible in both calibrated left and right cameras will be projected on left and right image planes, producing two image intensity pixels $A(xl, yl)$ and $B(xr, yr)$ (Figure 1). If the point $S(x, y, z)$ lies on a physical surface and it is visible by both cameras, then classical assumption of stereo analysis states that the 2D regions around the projection points A and B should produce a high correlation value. Using this principle, for each 3D point $S(x, y, z)$ we calculate the image locations $A(xl, yl)$ and $B(xr, yr)$ and we run the following normalized mean and variance correlation on the regions centered around 2D points A and B and assign the resulting value to $V(x, y, z)$

$$V(x,y,z) = \frac{\sum_{i,j}(A_{ij} - \overline{A})(B_{ij} - \overline{B})}{\left[\sum_{i,j}(A_{ij} - \overline{A})^2\right]^{\frac{1}{2}} \left[\sum_{i,j}(B_{ij} - \overline{B})^2\right]^{\frac{1}{2}}} \tag{3}$$

where $\overline{A}$ and $\overline{B}$ are the mean values of the regions centered around the points A and B, and A_{ij} and B_{ij} are the elements of these regions. Since the above correlation cannot be larger than 1.0, which is the perfect matching case, the external energy term (Equation 2) cannot be smaller than 0.0.

The process of filling the correlation values of V may seem to be similar to what Fua[9] does for filling the 3D buckets in the initialization phase. What we fill in the 3D array is the result of correlation values without any interpretation of whether the 3D point is part of a surface or not. On the contrary, Fua calculates some initial surface points from the correlation values and these points are filled into the 3D buckets. The rest of Fua's algorithm depends on these initial surface points. The filling process of V resembles the construction of the initial volume before the space carving begins in Kutulakos and Seitz[15] and the filling of "u-v-d" volume before the extraction of disparity surface in Chen and Medioni[5]. Kutulakos and Seitz[15] has a number of pointers for systems utilizing scene-space algorithms. In the above two methods, the surface reconstruction do not have any explicit surface model. On the other hand, our system extracts the underlying surface by using a deformable surface model which imposes more constraints on the extraction process to make it more robust and efficient.

As apparent from Figure 1 and the filling process of V, our system does not need the images to be rectified because filling V does not involve any search processes that require a rectified stereo pair. Since the external energy is the only mechanism that relates the optimization process to the stereo data, our system does not need rectified images. We only need the precise positions of the cameras used, that is we need a calibrated stereo pair. However, using a rectified stereo pair makes the formulations simpler and it is relatively easier to calculate the dimensions of the 3D volume that is visible from both cameras. For the sake of a clear presentation, we will assume rectified stereo pairs in this paper.

It is trivial to extend this system to work with more than two views by modifying the filling process of the 3D array V. If we are using trinocular images, then for a 3D point $S(x, y, z)$ we will have three image points that we can run correlations in combinations. In other words, we can take the correlation values between points one and two, between points two and three, and between

points one and three. Finally, we can put the weighted sum of these correlation results to the 3D array position $V(x, y, z)$. Another possibility is to keep only the maximum of these correlation values into the 3D array.

One may argue that our calculation of the external energy is computationally expensive because it involves many unnecessary correlation operations to fill the 3D array V. However, this is not the case. First, producing a correlation value for a given 3D point $S(x, y, z)$ corresponds to the epipolar constraint, which prevents many unnecessary correlation operations. Second, when we are running the filling process, we know the depth of the point $S(x, y, z)$. If this depth value is not possible to be a part of 3D surface, then we simply do not calculate the correlation and we do not assign any values to $V(x, y, z)$, which is initialized with a negative number to push the mesh away. This constraint corresponds to the search window constraint on the epipolar conjugate in the rectified stereo pair. Finally, our algorithm uses a coarse to fine scheme, which was used in many computer vision systems including stereo analysis. First we start with a coarse 3D array, which requires much lesser number of correlation calculations. Then we minimize the mesh energy functional using this coarse 3D array. After the minimization, we calculate a finer 3D array around the minimized mesh locations and the process continues on. At the coarser level, we use larger image templates for correlation around image points A and B and the meshes are also coarse. At finer levels we use smaller templates and finer meshes, which greatly increases efficiency. As a result of the above reasons, we argue that our calculation of external energy is very efficient.

Figure 2-(b) shows the visualization of the 3D array V produced from the rectified stereo pair shown in Figure 2-(a). Only three perpendicular slices of the volume are shown for visibility. The camera side of the volume is marked on the figure. Red color represents high correlation areas. For visualization purposes, we filled in every element of the array with the corresponding correlation values. In usual filling process, most of the array elements are not filled due to the coarse to fine scheme and the constraints we use, as shown in Figure 3.

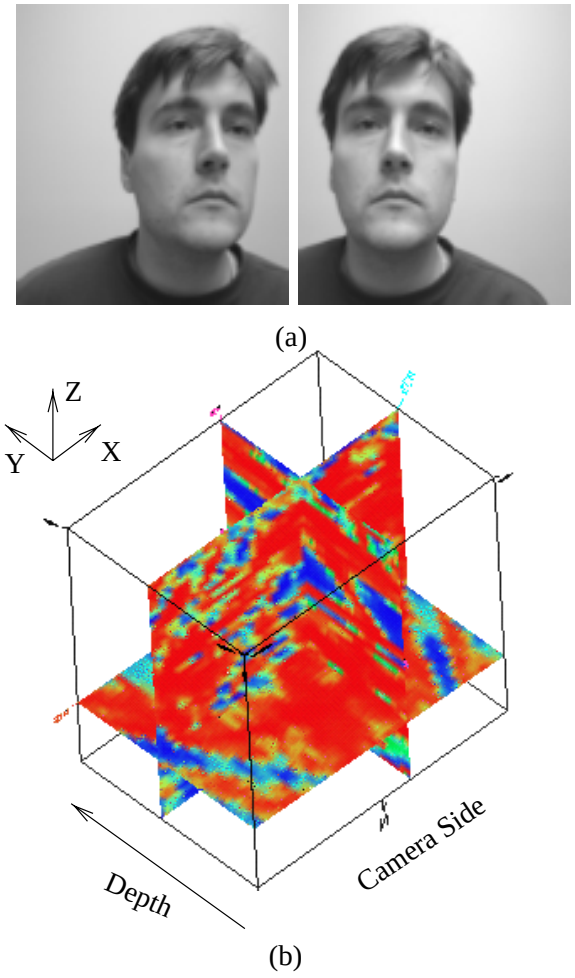

Figure 2: (a) A rectified stereo pair (b) Visualization of the 3D array V filled by correlation values. Red represents higher correlation values.

3 Surface Recovery and Deformable Model Optimization

One may consider using the 3D array V shown in Figure 2-(b) as a real 3D data such as Magnetic Resonance Imaging (MRI) data or Computed Tomography (CT) data. This gives us the possibility of using a volume segmentation algorithm such as the one used by Cohen and Cohen [6] or by McInerney and Terzopoulos [16], which are based on deformable models. However, the assumption of considering the 3D array V as an MRI image is not valid because inherently V holds only the depth information of the image points of the stereo pair and it is not a real 3D data. Therefore, direct use of volume segmentation algorithms is not possible. This is due to the definition of the stereo analysis, which is the problem of calculating the distance of scene points relative to the camera position or the viewer [12, 11].

We propose using a deformable mesh to recover the surface inside a volume, such as the one shown in Figure 2. The mesh will be large enough so that the projections of the mesh elements will cover all the image pixels on both images. The mesh deforms under the energy forces produced by Equation (2) with a restriction that the mesh elements can move in the depth dimension only so that it is always visible from all the image points during the deformations. At the end of the deformation, we will have the assigned depth values to the stereo image pixels by projecting mesh points on both image planes and the deformed mesh will give us the structure of the surface recovered. Our work should be distinguished from

the methods that use snakes to perform the stereo correspondence such as Cham and Cipolla[4] and the original snake formulation[14].

The deformations described above will place the mesh into a position that minimizes the mesh energy defined by Equation(2). This kind of minimization can be performed using a gradient descent optimization technique, that uses energy functional gradient information to guide the search process. However, this approach is sensitive to initial mesh positions and local minima is a serious problem especially with noisy data. The external energy contains a large amount of noise because of the inherent stereo analysis problems such as noise, occlusions, inaccuracies in the correlation process, etc. Moreover, a good initial position is not always available for most applications. Alternatively, an exhaustive enumeration that iterates on all possible mesh element positions would be computationally prohibitive. Exhaustive mesh minimization methods based on dynamic programming with a polynomial running time, such as the one proposed by [1], might be feasible. However, they are still not as efficient as we need. As a result, we needed to develop a new minimization technique to fit the needs of our application to minimize Equation(2).

3.1 Dual Mesh Optimization

A greedy minimization algorithm is an iterative approach that checks the possible locations for a single mesh element at a time and chooses the location that produces lowest energy. It performs this process for all mesh elements iteratively and the process continues until there is no improvement in the mesh energy. Due to its greedy nature, it is very fast. However, it is extremely sensitive to local minima for the same reason. Williams and Shah [19] use this approach to minimize a deformable contour energy for a 2D application. They argue that it is fast and the extracted contours are reliable. Applying such a minimization algorithm to the minimization of Equation(2) seriously suffers from the local minima due to the noise in the external energy mentioned in the previous section.

In order to address the local minima problem, we propose to use a dual mesh instead of just one mesh as our deformable model. We initialize one of the meshes at the smallest possible depth location, and we call it the near mesh, NM. The initial position of this mesh assigns the smallest depth values to all the pixels of stereo images. The other mesh, the far mesh FM, is initialized at the greatest depth location. Similar to NM, initially FM assigns greatest possible depth values to all of the pixels of the stereo pair(Figure 3). The greatest and smallest possible depth values are generally available in stereo analysis problems. The size of the search window in a classical stereo system assumes the same information as in the form of minimum and maximum possible disparity. If greatest and smallest possible depth values are not available, we use zero as the smallest and a large value as the greatest.

After the initialization phase the optimization continues by the following steps.

1. Let the initial dual mesh deform independently and simultaneously under the forces produced by Equation(2) ignoring the communication term, $E_{Comm}(m_{ij}, n_{ij})$. Due to their greedy nature, the dual mesh will be attracted by local minima and will stop deforming.

2. After the deformations stop, the energy of each mesh element is compared with the corresponding element in the other mesh ignoring the $E_{Comm}(m_{ij}, n_{ij})$ value. That is, the energy of NM_{ij} is compared with FM_{ij} for all i, j. If one of the mesh element energies is lower, that mesh element is more likely to find the global minima and the higher energy mesh element should be pushed towards the other mesh. This is accomplished by turning on the internal energy term $E_{Comm}(m_{ij}, n_{ij})$ for the high energy element. If m_{ij} is the high energy mesh element then

$$E_{Comm}(m_{ij}, n_{ij}) = |depth(m_{ij}) - depth(n_{ij})|.$$

   E_{Comm} returns zero for the lower energy mesh element n_{ij} in this case. The value of E_{Comm} is updated after each greedy iteration so that it reflects the change when m_{ij} gets closer to n_{ij}.

3. After the energy comparison and E_{Comm} setting step, we again let the dual mesh deform under the energy forces, this time with E_{Comm} turned on, which pushes the meshes together.

4. When they stop deforming, we check if they found the same location. If they did not, we turn the E_{Comm} forces off and let the dual mesh deform again until they stop deforming.

5. We repeat the above energy comparison in step 2, turn E_{Comm} on and repeat the whole process until both of the meshes find the same position. We also increase the forces that push the meshes together at each iteration step. This guarantees that they will find the same position after a finite number of steps.

We found that the dual mesh approach is robust against local minima. This is because when one of the meshes is caught by a minima, we immediately know whether it is a local minima by comparing the mesh energies. If it is a local minima, we know which way to push the mesh in order to beat the local minima. We also argue that the dual mesh minimization does better than

many gradient descent optimization methods for our application. In a gradient descent algorithm, it is not trivial to know if the current minima is local or global because the gradient vectors do not show which direction to move in that case.

We also found our approach very efficient because it is based on a greedy algorithm and both meshes are guaranteed to find the same position.

Our work on dual mesh is motivated by our previous work [2], which in turn was motivated by Gunn and Nixon[10]. Gunn and Nixon use a dual active contour to extract contours in noisy images robustly. The user has to supply the initial positions of the dual contour because they use a gradient descent algorithm in their minimization. The minimization stops when the dual contour finds the same position. Although the basic idea is to overcome the local minima, our approach is fundamentally different because it works on meshes and it does not need initial mesh positions. Another important difference is that, in our method we always know the correspondences between the two meshes. This gives us the possibility of comparing the local energies and changing the communication energy locally, which will result in robustness in capturing the global minima. On the other hand, Gunn and Nixon do not describe a way of establishing correspondences between the contours. As a result, the whole contour has to be pushed towards the other lower energy contour. This may cause the system to miss the global minima because the high energy contour may locate the global minima contour position partially and pushing it towards the other contour may move the contour from this partial global minima position.

3.2 Coarse to Fine Scheme

In order to make the system more efficient, we use a coarse to fine scheme that starts with a coarse 3D array V and a coarse dual mesh. Figure 3 shows a coarsely initialized 3D array V. The blue plane shows the initial position of the near mesh and the red plane shows the initial position of the far mesh.

Figure 4-(a) shows the two meshes initialized at the coarsest level before the minimization to recover the surface of the subject's face shown in Figure 2-(a). After the initialization, the dual mesh starts deforming under the energy forces. Figure 4-(b,c) shows the two planes deforming at the coarsest level. The final position of the dual mesh at the coarsest level (Figure 4-(d)) is used to determine the initial positions of the near mesh and the far mesh at the finer level by shifting it in positive and negative directions along the depth dimension (Figure 4-(e)). When the optimization ends at this fine level(Figure 4-(f)), the process is repeated in the same way at a finer level. The final recovered surface (Figure 5), which is the result of the optimization at the finest level, shows the

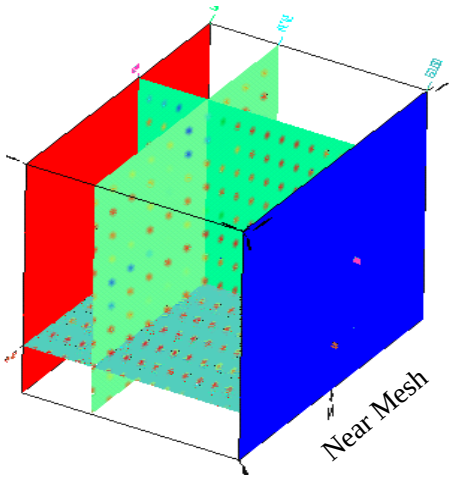

Figure 3: A coarsely initialized 3D array V. The blue and red planes show the initial positions of near and far mesh, respectively.

accuracy of our system, especially at the occluded areas near the sides of the subject's nose. There was a considerable amount of noisy correlation values at the occluded areas which affected the external energy. However, the internal forces were able to estimate a smooth surface for those areas. We found the dual mesh mechanism to be very successful at these situations in order to come up with the optimal result. The viewing angle of the face in Figure 5 is intentionally chosen to show the robustness of the system near the occluded areas. The recovery of the surface shown takes about 300 seconds on an SGI Octane.

4 Tracking

In order to handle the tracking problem, we use the usual tracking method of deformable models, which was first proposed by Kass et. al.[14]. We take the recovered surface from a time frame, use it as an initial estimate for the next time frame and let the mesh deform under a new potential energy. Although this kind of tracking will not give us point correspondences on the 3D surfaces between time frames, it is very useful in a number of ways. First, it is efficient because we know what part of the 3D space we should perform the search for global minima. Second, the recovered surfaces will be more accurate because the deformations will be much less than what it would be without initial positions. As a result, it is more likely to find the global minima.

Figure 6 shows the tracking results of Hurricane Luis by our system for three frames. We cannot show more frames due to space limitation. Although there are small discontinuities, we assumed that the cloud surfaces are continuous in order to test our system's robustness

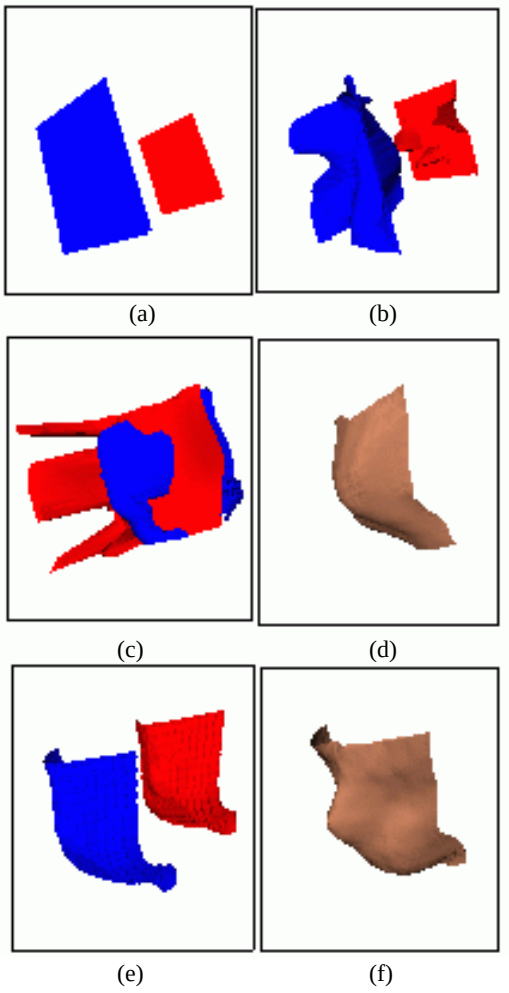

Figure 4: The deformation process of the dual mesh using a coarse to fine scheme. The initial stereo pair is shown in Figure 2. See the text in Section 3.2 for explanations.

against these kinds of surfaces. We confirmed the results by numerically comparing the computed disparity values with the ground truth which was obtained by infrared (IR) images; IR is a good approximation for the opaque cloud-top heights. The experiments show that our system is consistently within less than 1 pixel disparity error range. We also confirmed the results by manually measuring the disparity on the cloud images for a number of pixels.

As expected, the tracking of a single cloud pair takes about 20% less CPU time compared to surface recovery where no initial frame is supplied. From the experiments, we found that the performance of our system is very promising.

5 Conclusions

We presented a new method for integrating stereo correspondence, continuous 3D surface recovery and tracking in a single deformable dual mesh optimization process. The system introduces a number of novel ideas that would be valuable for stereo reconstruction, tracking and deformable model optimization research. The stereo correspondence and the surface recovery is unified by using the external energy mechanism of the deformable models in a novel way. The optimization of the deformable model is performed by a newly introduced dual mesh method, which showed that it is robust against local minima and efficient by a number of surface recovery and tracking experiments. By combining main stereo analysis steps in an efficient deformable optimization process, we achieved an easy to implement and simple system without sacrificing performance. Although we assumed a continuous surface, the system can be conveniently adapted for a system that can handle discontinuities.

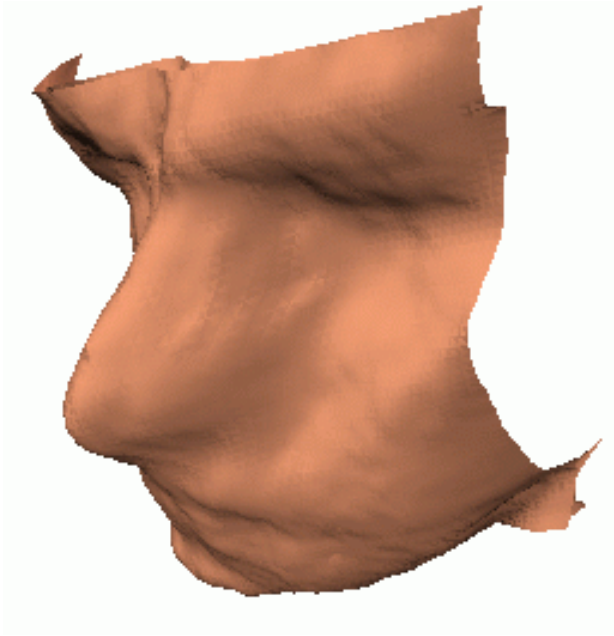

Figure 5: The recovered surface from the finest level (The mesh is shaded for better visualization)

There are many practical applications for this work. We are working on a framework that uses this system to analyze human facial expressions by tracking 3D features. We will embed constraints from the facial expression analysis into our deformable model to increase the system performance. For example, the nose can be assumed to be less rigid than the cheek. This information can be embedded into the optimization process by imposing more rigidity for the nose area mesh elements, resulting in a more robust and efficient stereo analysis system.

Acknowledgments

This work is supported by NSF CISE Research Infrastructure Grant CDA-9703088 and NSF Grant No. IRI 961924.

References

[1] Yusuf Sinan Akgul, Chandra Kambhamettu, and Maureen Stone. Analysis of the tongue surface movement using a spatiotemporally coherent deformable model. In *Proc. IEEE Workshop on Applications of Computer Vision*, pages 109–114, 1998.

[2] Yusuf Sinan Akgul, Chandra Kambhamettu, and Maureen Stone. Extraction and tracking of the tongue surface from ultrasound image sequences. In *CVPR*, pages 298–303, 1998.

[3] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[4] T.J. Cham and R. Cipolla. Stereo coupled active contours. In *CVPR*, pages 1094–1099, 1997.

[5] Q. Chen and G. Medioni. A volumetric stereo matching method: Application to image-based modeling. In *CVPR99*, pages I:29–34, 1999.

[6] L.D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-D and 3-D images. *PAMI*, 15(11):1131–1147, November 1993.

[7] F. Devernay and O.D. Faugeras. Computing differential properties of 3d shapes from stereoscopic images without 3d models. In *IEEE Computer Vision and Pattern Recognition*, pages 208–213, 1994.

[8] O.D. Faugeras and R. Keriven. Variational-principles, surface evolution, pdes, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, March 1998.

[9] P. Fua. From multiple stereo views to multiple 3-D surfaces. *International Journal of Computer Vision*, 24(1):19–35, August 1997.

[10] Steve R. Gunn and Mark S. Nixon. A robust snake implementation; a dual active contour. *PAMI*, 19(1):63–68, January 1997.

[11] W. Hoff and N. Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. *PAMI*, 11(2):121–136, February 1989.

[12] R.C. Jain, R. Kasturi, and B.G. Schunck. *Machine Vision*. McGraw-Hill, 1995.

[13] C. Kambhamettu, K. Palaniappan, and A.F. Hasler. Coupled, multi-resolution stereo and motion analysis. In *Symposium on Computer Vision*, pages 43–48, 1995.

[14] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *ICCV*, pages 259–269, 1987.

[15] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. Technical Report 692, University of Rochester CS, May 1998.

[16] T. McInerney and D. Terzopoulos. A finite element model for 3d shape reconstruction and nonrigid motion tracking. In *ICCV93*, pages 518–523, 1993.

[17] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *PAMI*, 8(4):413–424, July 1986.

[18] R.P. Wildes. Direct recovery of three-dimensional scene geometry from binocular stereo disparity. *PAMI*, 13(8):761–774, August 1991.

[19] D. J. Williams and M. Shah. A fast algorithm for active contours and curvature estimation. *Computer Vision, Graphics, Image Processing*, 55:14–26, 1992.

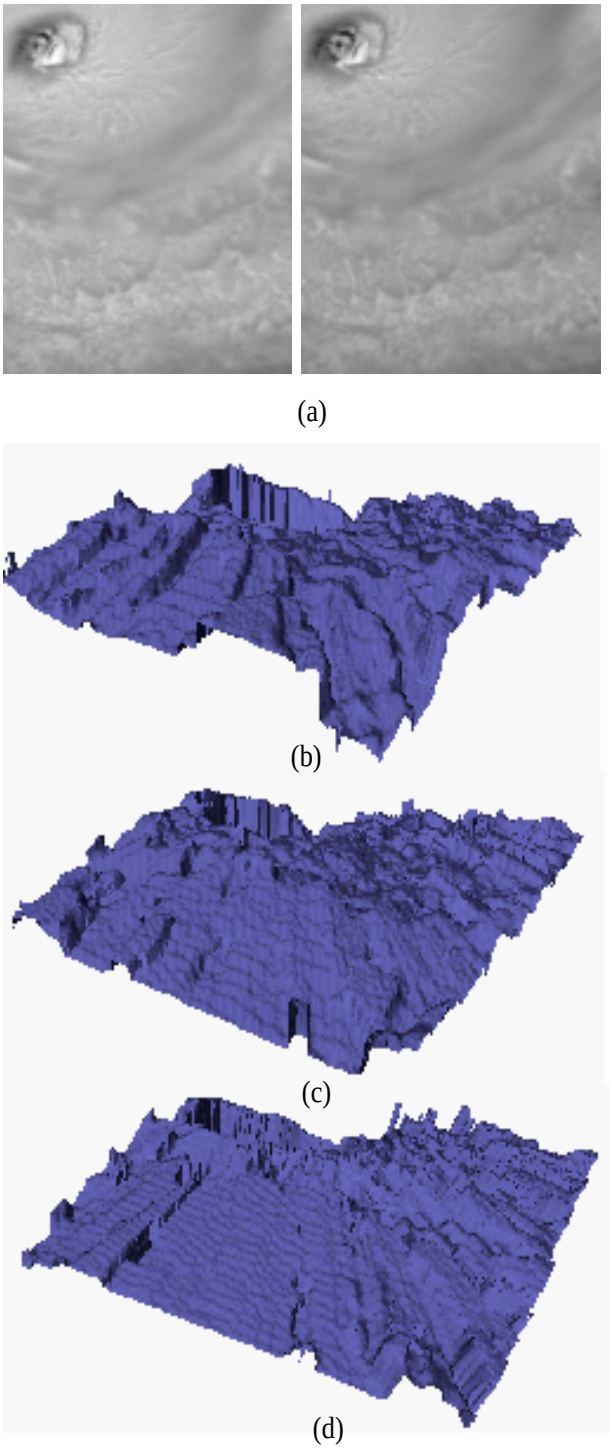

Figure 6: Tracked cloud surfaces. (a) Stereo image pair of the first frame (b) Recovered surface for the first frame (c-d) Tracked surfaces for the second and third frame