Three Dimensional Ultrastructure from Transmission Electron Microscope Tilt Series

Motilal Agrawal*† David Harwood* Ramani Duraiswami* Larry Davis*
Paul Luther‡

Abstract

We describe the application of stereo techniques to the recovery of three-dimensional ultrastructure from Transmision Electron Microscope(TEM) tilt series images. Algorithms for both calibration and reconstruction from a pair of TEM images are presented. We assume orthographic projection, and use a feature-based matching technique that uses correlation windows centered around the features. A novel nonlinear method of warping the window to get better measures of similarity is proposed and implemented, which is then used to extract the matchable features from the images. Use of a two stage optimization process using Dynamic Programming makes the matching accurate. In the first stage, matching is done along epipolar rows. Several candidate solutions obtained from the first stage are then input to a second stage which finds the solutions with maximum edge connectivity.

1 Introduction

This paper presents stereo techniques to recover the three dimensional structure of the neuro-muscular postsynaptic cytoskeleton, imaged using the transmission electron microscope(TEM). Transmission electron microscope images of thin biological specimens are obtained by rotary replication of the specimen with platinum atoms which is then mounted on a stage that can be tilted about two axes. A series of images are then taken by tilting the stage and subjecting it to electron beams. In this paper we will consider the recovery of ultrastructure from a pair of images. Stereo correspondence techniques from computer vision can then be applied to these stereo pairs in order to find corresponding points in the images to reconstruct the structure of the specimen.

Classical stereo algorithms work reasonably well provided the scene being imaged is smooth and does not contain many discontinuities. However, the present TEM images are very complex and have significant number of discontinuities. Figure 1 shows the left image of the stereo pair of TEM images. Note the presence of tubular structures criss-crossing each other. This paper describes specialized techniques for matching which take into account the complex nature of TEM images.

The basis of any stereo correspondence algorithm is a matching cost function that measures (dis)-similarity of two locations. This cost can be defined either locally or over an area of support. The absolute difference in intensity is the most common example of a locally defined cost [1] [2]. Costs defined over an area of support consider windows centered around the points and measures of their similarity such as correlation or the normalized cross-correlation [3].

Stereo correspondence algorithms can be either *feature based* [3] or *area based* [4] [5]. In feature based matching, only points with a certain amount of local information (such as intensity edges) are matched, resulting in a sparse disparity map. In area based methods, small patches of the image are matched giving a dense disparity map. There is yet another method which has become popular and it is based on *pixel based* [1] [6] matching. Our stereo matching is a feature based matching. We use correlation windows centered on these features. One of the contributions of our paper is a nonlinear warping technique for the windows. This warping technique accounts for the fact that the underlying surface which is being matched is a 3D surface and thereby induces projective distortions in the images [4]. Thus this warping technique gives us a better measure of similarity which is used as the cost function in our matching.

Since ours is a feature based matching technique, we need to select the features in the left and the right images. Our selection of features in the left image is based on intensity edges. However using edges to select features in the right image also may result in either inclusion of extra features which were not in-

[*] The authors are with UMIACS, University of Maryland, College Park, MD

[†] Email address for correspondence: mla@cs.umd.edu

[‡] The author is with Department of Physiology, University of Maryland, Baltimore, MD

cluded in the left image or exclusion of features which were selected in the left image. This may result in many mismatches. To deal with this, we present a new method of selecting features in the right image based on the discriminatory power of the cost function itself.

This cost function is used in a two stage optimisation to perform the matching. In the first stage, the matching is done along single rows. Under suitable assumptions, this optimisation can be done using Dynamic Programming. In the second stage, we take into account the inter row constraint that connected features in the left image should match to connected features in the right as well. The inter and intra row constraint has been used previously in [7] and [8]. However their technique matches edge segments and therefore force edge continuity. Also extraction of edge segments by edge linking so as to preserve the topology in the left and right images is not an easy task for the complex TEM images.

We begin by describing the calibration and rectification for orthographic cameras in Section 2. Section 3 describes our warping technique. The cost function for matching is described in Section 4. Our feature selection method is described in Section 5. Section 6 discusses single row matching and Section 7 describes the second level of optimisation for multiple rows. Finally we conclude with results and describe the ongoing and future work.

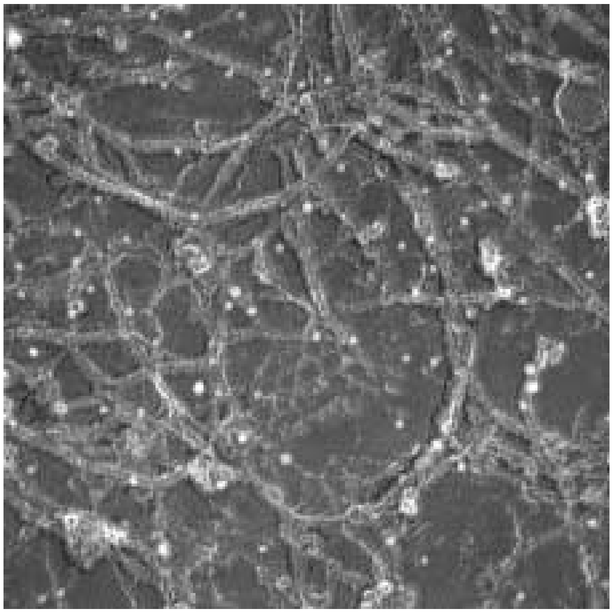

Figure 1: Left image of the stereo pair

2 Calibration and Rectification

The imaging in transmission electron microscope can be modelled as an orthographic projection. Although nonlinearities in the form of distortions [9] exist, they come into prominence in thick sections when the angle of tilt is large.
For orthographic projection, a 3D point (X, Y, Z) is projected to (x, y) in the image. In homogeneous coordinates this can be written as

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} m_x & m_y & m_z & t_x \\ n_x & n_y & n_z & t_y \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

and $\begin{pmatrix} m_x & m_y & m_z \end{pmatrix}^t$ is orthogonal to $\begin{pmatrix} n_x & n_y & n_z \end{pmatrix}^t$. Calibration is done using the factorization approach of Tomasi and Kanade [10]. Colloidal gold particles of two different sizes (10 and 100 nm) are introduced on both sides of the film as shown in Figure 2. The centers of the gold beads are manually matched and serve as input for the factorization based method. These gold beads can be seen in Figure 1.

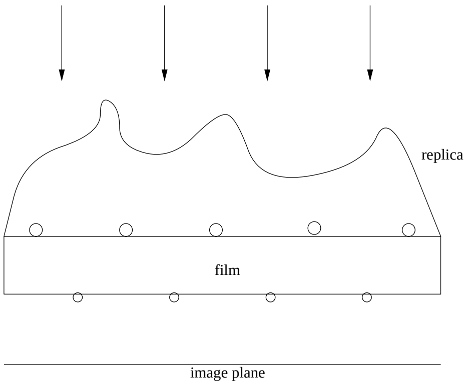

Figure 2: Calibration setup

2.1 Rectification for orthographic cameras

For our tilt series we would like to have the corresponding features to lie on the same rows in both the images i.e; the epipolar lines must be horizontal. This may not hold true for the set of projection matrices calculated from the factorization approach. Therefore the images have to be rectified to make the epipolar lines horizontal.

Consider two cameras $P^i = \begin{pmatrix} m_i^t & t_x^i \\ n_i^t & t_y^i \\ 0^t & 1 \end{pmatrix}$ and

$P^j = \begin{pmatrix} m_j^t & t_x^j \\ n_j^t & t_y^j \\ 0^t & 1 \end{pmatrix}$ generating images I and J. The two points Q_1 and $Q_2 = Q_1 + \lambda(m_i \times n_i)$ in $3D$ both project to the same point in image I, hence $(m_i \times n_i)$ represents the direction of projection in image I. Also the direction of the epipolar line in J for camera I is given by $\left(\begin{bmatrix} m_j & m_i & n_i \end{bmatrix} \quad \begin{bmatrix} n_j & m_i & n_i \end{bmatrix} \right)^t$ where $\begin{bmatrix} x & y & z \end{bmatrix}$ denotes the scalar triple product. Therefore, for the epipolar line in J to be horizontal $\begin{bmatrix} n_j & m_i & n_i \end{bmatrix}$ has to be zero.

This is done by multiplying by two homographies H_i and H_j in images I and J respectively. Consider the projection equation, $x = PX$. Premultiplying both sides by a 3×3 affine matrix H gives $Hx = HPX$. Therefore we have to find affine matrices H_i and H_j such that the resultant projection matrices $\hat{P}^i = H_i P^i$ and $\hat{P}^j = H_j P^j$ have their epipolar lines horizontal.

Let

$$H_i = \begin{pmatrix} h_{i1} & h_{i2} & h_{i3} \\ h_{i4} & h_{i5} & h_{i6} \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$H_j = \begin{pmatrix} h_{j1} & h_{j2} & h_{j3} \\ h_{j4} & h_{j5} & h_{j6} \\ 0 & 0 & 1 \end{pmatrix}$$

For epipolar lines in I and J to be horizontal $\begin{bmatrix} \hat{n}_j & \hat{m}_i & \hat{n}_i \end{bmatrix} = 0$ and $\begin{bmatrix} \hat{n}_i & \hat{m}_j & \hat{n}_j \end{bmatrix} = 0$. Plugging in the values and simplification leads to

$$\frac{h_{j4}}{h_{j5}} = -\frac{\begin{bmatrix} n_j & m_i & n_i \end{bmatrix}}{\begin{bmatrix} m_j & m_i & n_i \end{bmatrix}}$$

$$\frac{h_{i4}}{h_{i5}} = -\frac{\begin{bmatrix} n_i & m_j & n_j \end{bmatrix}}{\begin{bmatrix} m_i & m_j & n_j \end{bmatrix}}$$

It is clear that there are infinite number of (invertible) H matrices that could be taken to perform rectification as long as they meet the above constraint on the ratio of h_4/h_5. Also we would not only like to make the epipolar lines horizontal, but we would also like to have the corresponding points in the two images on the same row as well. This introduces one additional constraint for the H_i and H_j matrices. The best H matrix is found by taking that H matrix which introduces least deformation. In our case, we take the H matrix to act like a simple 2D rotation.

$$H_i = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & h_{i3} \\ \sin(\theta) & \cos(\theta) & h_{i6} \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$H_j = \begin{pmatrix} \cos(\phi) & -\sin(\phi) & h_{j3} \\ \sin(\phi) & \cos(\phi) & h_{j6} \\ 0 & 0 & 1 \end{pmatrix}$$

where θ and ϕ are determined by the above constraints on the ratios and the translational parameters are obtained from the constraint that corresponding epipolar lines should be on the same rows.

3 Warping

Given a feature in the left image and its candidate match in the right image, we would like to know how likely is it that they are the projections of the same point in the 3D. This can be done by correlating windows centered around the feature points. In general, a window in the left image will undergo a warping in order for it to perfectly correlate with the feature in the right image. This mapping is induced by the 3D structure and hence in general, will be a non-linear mapping. Therefore, a simple one to one correspondence is not enough. Even an affine mapping cannot perfectly account for the gray level variation because an affine map would imply that the underlying surface is a plane. Therefore we have come up with a nonlinear warping to fit the left and the right windows. This warping can be calculated efficiently and gives a better measure of the goodness of the match.

We consider windows centered around the left and the right feature points. The top and bottom portions are treated independently. The center row is kept fixed. Each row from the center is allowed to undergo a horizontal shift of one pixel relative to the previous row. This can be thought of as a kind of smoothness constraint. The goal then is to find the horizontal shifts for each row which results in the maximum gray level cross correlation between the rows of the warped left and right image. This optimisation can be done very efficiently using dynamic programming.

As a result of the warping process, we get two measures of matching. The first is the amount of warp that was required for the best fit. This is simply the sum of the shifts required for each row. The second measure is the residual error in the gray level between the left warped region and the right region. For a NxN window this is calculated as

$$error = \frac{\sum_i \sum_j \left(l'_{ij} - r_{ij} \right)^2 / N^2}{\text{variance}(l'_{ij} + r_{ij})}$$

Here l'_{ij} and r_{ij} are the intensities at (i,j) in the warped left and the right image respectively.

4 Maximum likelihood cost function

The likelihood ratio of the i^{th} feature matching a potential candidate j is defined as

$$\text{likelihood ratio}(lr_{i,j}) = \frac{\text{Prob.(true match)}}{\text{Prob.(false match)}}$$

In our case these probabilities would depend on the warp required to align the left window with the right and the residual error after warping. Qualitatively, for correct matches, the amount of warp required would be less and the residual error in the gray level correlation after warping would also be small. On the other hand for incorrect matches, one or both of warp and the gray level error is expected to be large. However, in order to use the likelihood ratio we have to determine quantitative measures of the conditional probabilities $P(\text{true match/error, warp})$ and also $P(\text{false match/error, warp})$. Modelling these distributions is not straightforward as these depend on a variety of factors including the structure of the scene, which we do not know. This is further complicated by the fact that the warps and the residual error are not independent. In our approach, we try to estimate these probability distributions from the images themselves. In order to do this, for each feature point in the left, we look for the feature in the right (within the disparity bounds) which results in a minimum residual error after warping. We assume that this is the 'correct' match for that feature and all other matches are 'incorrect'. Based on this, we build a histogram for the 'correct' and 'incorrect' features and normalize it to produce a probability distribution function. Figure 3 shows the true and the false distribution for the residual error and Figure 4 shows the same for the warp. This distribution can be later refined when we have correctly matched the feature points by our two stage optimisation process. Thus this pdf can be progressively refined to get more accurate estimation. But in our experiments this did not produce substantial improvements.

5 Feature selection

Since ours is a feature based matching technique, we need to select features in the left as well as right image. In order to aid this feature selection process we perform some low level image processing operations. To begin with, we perform a simple gray level normalization of the left and right image by making their mean and standard deviation same. This accounts for any relative change in the brightness between the left and right image. Since our approach is based on matching edge features, we apply an image enhancement operation by applying SNN [11] to the images.

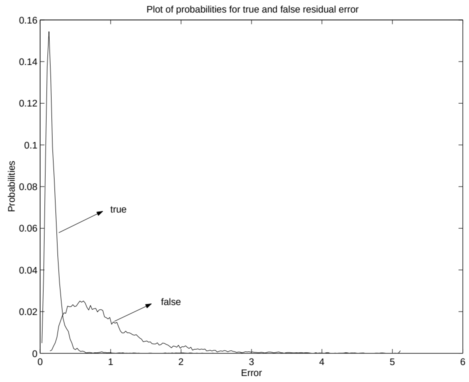

Figure 3: Distribution of true and false residual error

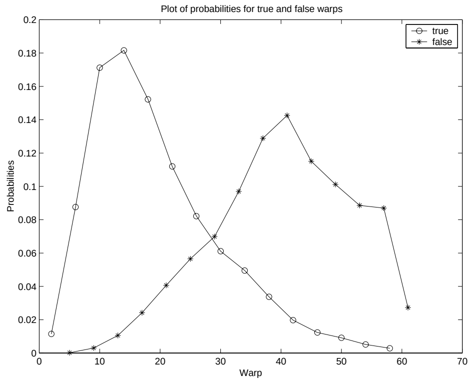

Figure 4: Distribution of true and false warp

Application of SNN operators not only deblurs edges but also reduces local interior variation, without distorting edges. Since we are matching rows in the left image to corresponding rows in the right image, there is an inherent ambiguity in matching horizontal features. Therefore we select the edges in the left image which run vertically. This can be done by applying a gradient operator along the horizontal direction. Once the features in the left image are selected, we need to select features in the right image with which they are to be matched. Selection of features in the right which do not have a corresponding feature selected in the left or failure to select a feature in the right whose left has already been selected can both lead to ambiguity and therefore subsequent errors. Hence a simple edge detection for selecting features in the right image does not yield a good result as it is very sensitive to thresholds. On the other hand, since we are using the likelihood ratio as our cost function to perform matching, it is logical to use that for selecting the features in the right image as well. Features that were not selected by this cost function would perform very poorly in the matching process anyway. For each feature point in the left image, its matches will lie within a certain maximum disparity range. Therefore we calculate the likelihood ratios of these pixels from the warp and gray level error obtained as a result of warping windows centered around each of these pixels. The local maxima in the cost function which corresponds to a likelihood ratio greater than or equal to one are selected as features in the right image. Figures 5 and 6 show the selected features in a small region(upper left) of the left and the right image respectively.

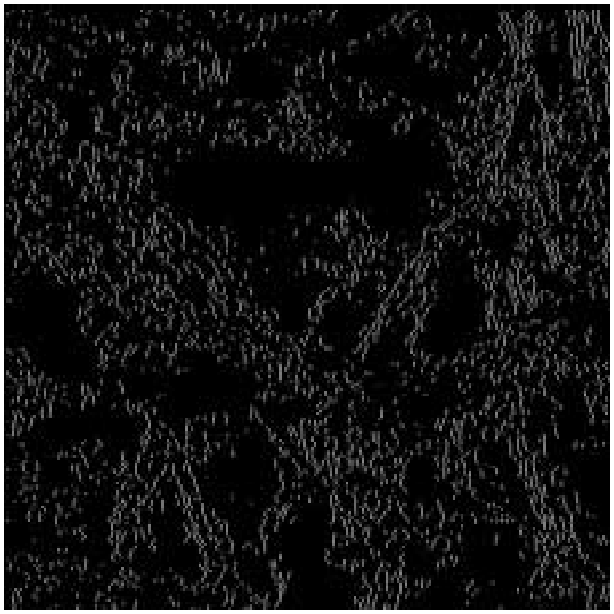

Figure 5: Features selected in the left image of the stereo pair

6 Single row matching

Once the feature points in the left and the right image are selected, for each feature in the left image we have a list of candidate features in the right image which match to it. Each candidate has a likelihood ratio associated with it, which tells us how good a match that particular candidate is. Next we consider an entire row of feature points in the left image and match it to the features points in the corresponding row of the right image. This is done under the additional assumptions

1. *uniqueness constraint*: each feature in the left image has a unique match in the right image and vice versa.
2. *ordering constraint*: If feature A is to left of B in the left then the order will be preserved for their corresponding matches in the right image.

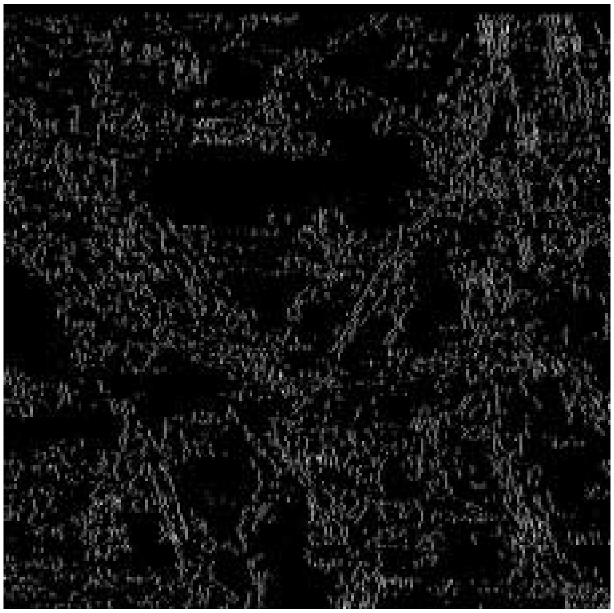

Figure 6: Features selected in the right image

Obviously, depending on the structure of the scene, there will be feature points which violate these assumptions. To account for this we have a default occlusion probability (β). This corresponds to a feature which has no corresponding match(null matched). For all our experiments we have set the default occlusion probability as 1/1000, which implies that to be matched, the probability that its a true match should be 1000 times or more than the probability that it is a false match.

Under these constraints we want to maximize the overall likelihood ratio of all the features in that row. This problem is suitable for Dynamic Prograaming as described in [1]. Omitting the details, if there are N left feature points and M right feature points, then we setup a NxM grid. From each point in the grid we can either go right, bottom or bottom right, thereby ensuring that the above two constraints are preserved. Each such choice has a different cost associated with it. The cost at any point is the minimum of the choices that lead to that point:

$$C_{i+1,j+1} = min\left(C_{i,j} + lr_{i,j}, C_(i+1,j) + \beta, C_{i,j+1} + \beta\right)$$

At each point we also keep track of the choice that leads to that point. The goal is to find the least cost path from (0,0) to (M,N). Figure 7 illustrates the grid and shows two valid paths for the grid.

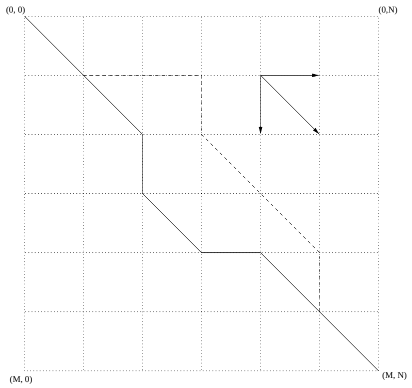

Figure 7: The algorithm for dynamic programming showing two best paths

7 Combining multiple rows - the edge continuity constraint

Matching a single row by DP ignores the constraint that edges which are connected in the left have matches which are connected in the right as well. In general, this is a very strong constraint and can potentially correct for any mismatches in the single row matching. To make use of this constraint, in our single row matching instead of generating the single best solution we generate the K best solutions. In our experiments we have found that a K value of 100 is sufficient.

An important observation to be made is that the second best path will be simply a diversion from the minimum cost path obtained. Therefore to find the second best path, as we backtrace from (M,N) to (0,0) we note down the least penalizing choice for each point along this path. In Figure 8 shown below, the best path to reach $C_{i+1,j+1}$ is from $C_{i,j}$. However if instead we choose the horizontal path from $C_{i+1,j}$ we would incur a penalty resulting in an increase in the cost of the path.

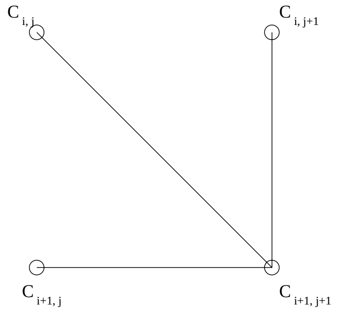

Figure 8: Alternate paths and penalty

The penalty incurred would be $C_{i+1,j} + \beta - (C_{i,j} + lr_{i,j})$ and this would be the increase in the cost of the alternate path. Figure 7 above shows the second best path as a dashed line which results from a diversion of the best path at (M-1, N-1). Similarly the third best path will be produced as a result of diversion in the first and the second best path, depending upon which of them will result in the least increase in the cost. Continuing this way we can find the K best paths for each row of the image.

In the second level of optimization out of the K best paths for each row we find those paths which results in maximum edge continuity. Features $f_{i,j}$ and $f_{i+1,k}$ in two consecutive rows are said to be connected if their horizontal distance is less than or equal to two. ($|j-k| \leq 2$). If their corresponding matches $f^{'}_{i,j}$ and $f^{'}_{i+1,k}$ are also connected then we say that the edge continuity is preserved, otherwise violated. Let $P_{preserve}$ and $P_{violate}$ be the probability that edges preserve and violate continuity respectively. The costs for these, $C_{preserve}$ and $C_{violate}$, are obtained by taking $-\log$ of $P_{preserve}$ and $P_{violate}$ respectively. For our experiments, we have taken $P_{preserve} = 0.7$ and $P_{violate} = 0.3$. Let A and B be two paths obtained from the K best dynamic programming for two consecutive rows. Then we define the edge continuity cost($EC_{A,B}$) between A and B as the sum of the edge preserving and violating costs for the features which are connected.

$$EC_{A,B} = \sum C_{preserve} + \sum C_{violate}$$

The total cost(TC) of selecting the path A and B in the second level of optimization is $TC = C_A + C_B + EC_{A,B}$ where C_A and C_B are the costs of these paths (obtained by K best DP). The goal is to minimize the total cost by selecting those paths which result in minimum total cost, thereby ensuring maximum edge continuity. This optimization is also well suited for DP as follows. If there are N rows in the left image, each row having K solutions, where each solution represents a match set for the points, then we construct an NxK grid. From each point we can move to the K corresponding points on the next row. The cost for each point on the grid($Cost_{i,j}$) is

$$Cost_{i,j} = min_{k=1}^{K} \left(C_j + EC_{k,j} + Cost_{i-1,k}\right)$$

As in the previous optimization we store the choice which lead to the minimum cost at each point. Finally we find the $min_i \quad Cost_{N,i}$ and backtrack to find the best solutions for each row.

8 Results and conclusion

We have applied our techniques to several TEM stereo pairs and also standard test images for stereo. Since our disparity map is sparse, we have not included the depth maps. But a movie clip showing the reconstruction of the TEM specimen can be seen at our web site `http://www.umiacs.umd.edu/~mla/tem`. We are working on a detailed analysis of the results of matching. However, visually looking at the matches have shown that about 90% of the matches are correct. Currently we are also working on surface fitting and depth interpolation of the sparse depth map obtained from the matching.

In this paper we have presented stereo reconstruction techniques as applicable to TEM images. We have come up with a new measure of similarity which not only includes a gray level correlation but also a warp measure. We have also shown how the traditional Dynamic Programming algorithm can be adapted to produce 'K' best solutions. This when coupled with the edge continuity constraint increases the accuracy of the matches in the second level of our optimisation.

Acknowledgements

The financial support by the National Institute of Mental Health under the Human Brain Project – Grant 5 R01 LM06326-02 is gratefully acknowledged.

References

[1] I.J. Cox, S. Hingorani, B.M. Maggs, and S.B. Rao, "A maximum likelihood stereo algorithm," *Computer Vision and Image Understanding*, vol. 63, no. 3, pp. 542–567, May 1996.

[2] T. Kanade, "Development of a video rate stereo machine," in *Image Understanding Workshop*, Monterey, CA, November 1994, pp. 549–557, Morgan Kaufmann Publishers.

[3] O. Faugeras, B. Hotz, H. Mathieu, T. Vieville, Z. Zhang, P. Fua, E. Theron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy, "Real-time correlation-based stereo: algorithm, implementation and applications," Tech. Rep. RR-2013, INRIA Sophia-Antipolis, 1993.

[4] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 9, pp. 920–932, September 1994.

[5] A. Fusiello, V. Roberto, and E. Trucco, "Efficient stereo with multiple windowing," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, June 1997, pp. 858–863, IEEE Computer Society Press.

[6] Stan Birchfield and Carlo Tomasi, "Depth discontinuities by pixel-to-pixel stereo," in *Proceedings Sixth IEEE International Conference on Computer Vision*, Mumbai, India, January 1998, pp. 1073–1080.

[7] Y. Ohta and T. Kanade, "Stereo by intra and inter-scanline search using dynamic programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 139–154, 1985.

[8] Phillip David, David Harwood, and Larry Davis, "A probablistic, local feature, stereo correspondence algorithm," Tech. Rep. CAR-TR-540, CFAR, University of Maryland, College Park, MD, USA, March 1991.

[9] G. Y. Fan, S. J. Young, Philip Miller, and Mark H. Ellisman, "Conditions for electron tomographics data acquisition," *Journal of Electron Microscopy*, vol. 44, pp. 15–21, 1995.

[10] C. Tomasi and T. Kanade, "Shape from motion from image streams under orthography: A factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, November 1992.

[11] D. Harwood, M. Subbarao, H. Hakalahti, and L. Davis, "A new class of edge-preserving smoothing filters," *Pattern Recognition Letters*, vol. 6, pp. 155–162, 1987.