A COMBINED MATCHING AND TENSOR METHOD TO OBTAIN A HIGH FIDELITY VELOCITY FIELDS FROM IMAGE SEQUENCES OF THE NON-RIGID MOTION OF THE GROWTH OF A PLANT ROOT

Hai S. Jiang 1, K. Palaniappan 1*, Tobias I Baskin 2

1 Computer Engineering & Computer Science, University of Missouri-Columbia
201 Engineering Building West, Columbia, MO 65211
Email: {hai, palani}@meru.cecs.missouri.edu

2 Biology Science, University of Missouri-Columbia,
105 Tucker Hall, Columbia, MO65211
Email: baskint@missouri.edu

Abstract
This paper addresses the problem of estimating the non-rigid growth motion of plant roots. Several assumptions and constraints about the motion estimation are made and a min-cost objective function for robust matching method is constructed. To expedite the matching method, a tensor-based method to initialize the searching parameters is provided. The tensor confidence field is used as a mask to indicate the sites where coherent motion occurred. Consistency between the forward matching and the backward matching is used as a confidence test. Other confidence testimonies include the sharpness of the peak/valley of the objective function in the robust estimation procedure.

Key Words:
motion estimation, robust matching, tensor optical flow

1. Introduction
In recently years biology scientists became interested in accurately measuring the behavior of plant growth. In 1998, Schmundt et al. [7] used a structure tensor-based method in computing the dense displacement vector field of the growth rate of dicot leaf with high spatial and temporal resolution. Beemster and Tobias [9] analyzed the cell division and elongation in root growth experiment in which the measurement of root growth rate was done manually. A further computer-based measurement of the root growth is required to verify the conclusion of [9]. Plant root growth is a dynamic process with non-rigid motion. The motion is non-uniformed, discontinuous and involved with local deformations. Actually the growth is a mixed multi-scale motion. In particular, in addition to the motion caused by the expansion of cells (i.e., growth) there is the potential motion caused by the bulk translation of the root (large scale) and the movement of the internal contents of the cell (cytoplasmic streaming). All of these impose great problems on accurately estimation the growth rate of the root.

Given an image sequence of a root, we want to measure the parallel motion in the root elongation direction and the motion perpendicular to the root. Following [9], the perpendicular direction growth rate of the root is very small and approximately constant, we only want to analysis the parallel growth rate along the root. Actually in the experiment we validated the assumption of small perpendicular growth.

Imagery Characteristics
In this experiment we used seedlings of the species Arabidopsis thaliana L, grown and handled as described by [9]. We chose roots about 2-3mm long and 0.1-0.2mm wide. The roots were imaged when growing under a fixed CCD camera with a setting of imagery at 1.13pixel/um. We took effort to keep constant illumination, but there was still 2% possible variance of illumination between images within one-hour period. The root images were taken every 30 seconds for a sequence of 10 images. The time accuracy of imaging interval is ± 1 sec.

Because of limitation of the scope of the microscope the whole root can not be seen in one image, while several microscopes can not work together for one root, 10 concatenated segments of the root were imaged separately at different time periods and finally got 10 image sequences. Because each segment was imaged at different time, to concatenate the six segments into a whole root image and do a single image sequence analysis is not appropriate. However, to illustrate the root, Figure 3 shows the concatenated root.

The signal-to-noise ratio of the root images was not high because of noise of air vibration and the nutrient water film around the root which even appeared as a slightly darker rim covering the root. In addition we noticed that the growth processes reflected by the images were involved with remarkable deformations, while the root was bending and growing with a little 3D twisting along the center axis of the roots.

Related work and methodology
Currently the approaches for motion estimation include differentiation method, frequency domain method, matching method and tensor-based method and all of these methods are application-oriented. We were
encouraged by the success of the motion estimation of leaf growth by B. Jahne’s group [7][8]. Although the leaf experiment provided excellent results of velocity estimation for the leaf growth using tensor-based method, the root growth estimation in our experiment unfortunately is not exactly the same case. One reason is that the tensor method is only suitable for the estimation of low motion. When the velocity is bigger than 3.0 pixel/frame, the confidence of the velocity from the tensor method is low.

From 1992, many scientists have done much work in fields of phase-based optical flow and robust motion estimation. Currently many works focus on multiple motion and motion discontinuity. Black et. al. [1] proposed a practical framework for modeling and computing motion including techniques of dealing with motion discontinuities. Beauchemin et. al. [4] modeled the structure of occlusions and motion discontinuities in fourier space. In 1999 Black and Fleet [2] proposed a probabilistic framework with two primitive motion models: translation and motion discontinuity. However, it is not until recently that multiple motion and motion deformation can be handled with high reliability to give high-fidelity and dense motion field output.

In this experiment we used correlation-based matching methods with some deformation resilient techniques to get a dense displacement vector field for the root image. Part of the framework of our algorithm is very similar to the framework of robust estimation that Black et. al. [1] had proposed. The difference is that we use image differences within local neighborhood instead of optical-flow gradient summation in the objective function. This is because we want get accurate site-matching between images other than full-density but low-fidelity motion field. We did not do temporal image sequence tracking nor adding temporal coherence component in the objective function because of low temporal support and non-smoothness of local growth along the time line when the time interval is very small.

The rest of our work was focused on the confidence of what motion field we got from robust matching. We used forward matching combined with backward matching to ensure confidential matching should be those matching pairs between two image at time t and t+1 that their corresponding is reciprocal without much deviation and can be repeated in both directions. We assume that these kind of matching is confidential. Combined with other kind of confidence testimonies, we can get high-fidelity matching between sites of two images. The matches actually are displacement field. Then we did averaging over the velocity field generated from each pairs of images in the image sequence to resist noise.

2. Robust Matching of Non-rigid Motion

Let I(x,y,t) be the image intensity at a point (x,y) at time t. For an image of size m x m pixels we define a grid of sites: S = { s_1, s_2, ..., s_m | \forall w, 0 \leq i(s_w), j(s_w) \leq m-1 }, Where (i(s_w), j(s_w)) denotes the pixel coordinates of site s_w. We assume the motion speed is coherent, so we construct the objective function E(u,v, s) for site s as:

\[
E(u,v, s) = \sum_{i,j} \left[ \rho(I(i,j,t) - I(p,q,t+1,\sigma_1)) + \lambda \sum_{s \in G_s} [\rho(u_k - u_s, \sigma_2) + \rho(v_k - v_s, \sigma_2)] \right]
\]

where

1) u,v denote the displacement in x, y axis directions;
2) R is a local neighborhood of site s with the size wxl.
3) ( i, j ) denotes the coordinate of pixel s_k in the neighborhood of s, so s_k \in R(s) \subseteq S(t);
4) ( p, q ) denotes the coordinate of pixel s_k' in the neighborhood of s', while s' is the site at time t+1 to match s at time t, s' \in D. D is the searching domain around (i, j ) at time t+1 with predefined size WxL large enough to find possible s'. So s_k' \in R(s') \subseteq S(t+1);
5) G_s is the 4-connectivity or 8-connectivity of pixel s based on the displacement field at time t;
6) \lambda is a constant to define the importance of the motion coherence in the objective function;
7) \rho(x, \sigma) = \frac{x^2}{\sigma^2 + x^2}
\]

with derivative

\[
\psi(x, \sigma) = \frac{2x\sigma^2}{(\sigma^2 + x^2)^2}
\]

Figure 1 gives the shape of \( \rho(x, \sigma) \) and its derivative. The reasons of choosing this function is that it rejects high errors while keeps the cost for small errors very low.

![Figure 1. Robust error norm (\( \rho \)) and its derivative (\( \psi \))](image)

Although we used tensor-based method is only suitable for estimation of low motion. When the velocity is bigger than 3.0 pixel/frame, the confidence of the velocity from the tensor method is low.

![Figure 2. Diagram of correlation matching](image)
Our object is to find a site \( s' \) in a searching domain \( D \) near site \( s \), that minimize \( E(u,v,s) \) for all sites in \( D \) but the minimum \( E(u,v,s) \) should be less than a certain threshold value. Figure 2 gives the idea of matching between two images and in the experiment we set \( l = w = 15 \), and \( L = 31, W = 81 \).

If the minimum for \( E(u,v,s) \) we found in the matching algorithm is above a threshold, we think the site \( s \) can not get a valid match for current size of the matching block. Some times although in a large scale, a site in an plant image appears deformed comparing with its corresponding site in the latter image, the site may not appear deformed in a smaller scale. This gives hints for the need of multi-scale methods when an outlier occurs.

3. Tensor-based Method to Estimate Motion

The displacement of grey value structures within consecutive images of a sequence yields inclined image structures with respect to the temporal axis of spatiotemporal images. A sequence of infrared images of the plant root has been stacked to build an image cube. The relation between the orientation angle and the optical flow is given by

\[
\mathbf{f} = - \begin{bmatrix} \tan \varphi_x \\ \tan \varphi_y \end{bmatrix}
\tag{4}
\]

where \( \mathbf{f} = (f_x, f_y) \) denotes the optical flow on the image plane and the angles \( \varphi \) defines the angles between the plane normal to the lines of constant grey value and the \( x \) and \( y \) axes, respectively. This basic property of spatiotemporal images allows to estimate the optical flow from a 3D orientation analysis, searching for the direction of constant gray value in \( x \)-space.

In order to determine local orientation Bigun and Granlund [14] proposed a tensor representation of the local grey value distribution. Using directional derivatives, Kass and Witkin [15] came to a solution that turned out to be equivalent to the tensor method. Searching for a general description of local orientation Knutsson [16] concluded that local structure in a \( W \)-dimensional space can be represented by a symmetric \( W \times W \) tensor. Rao [17] used a similar tensor representation for 2D texture analysis. One possible realization of Knutsson's tensor is the structure tensor, \( \mathbf{J} : \)

\[
\mathbf{J}(x) = \int_{-\infty}^{\infty} h(x-x') \mathbf{\nabla} g(x') \mathbf{\nabla}^T g(x') \ d^W x'
\tag{5}
\]

The components of \( \mathbf{J} \) are given by

\[
\mathbf{J}_{pq} = \int_{-\infty}^{\infty} h(x-x') \frac{\partial g(x')}{\partial x_p} \frac{\partial g(x')}{\partial x_q} \ d^W x'
\tag{6}
\]

With \( g(x) \) we denote the spatiotemporal image structure and \( \frac{\partial g(x)}{\partial x_q} \) represents the partial derivation along the direction of the \( x_q \)-axis. The information within a local neighborhood \( U \) around the central point \( x=(x,y,t) \) is weighted by a window-function \( h(x-x') \). In practical applications the size of the local neighborhood \( U \) represents the area and time (in numbers of images) over which the orientation information is averaged.

The structure tensor contains the entire information about the geometrical distribution of gray values within the local spatiotemporal neighborhood. The orientation of iso-gray value lines within a 3D spatiotemporal neighborhood \( U \) can mathematically be formulated as the direction \( \mathbf{n} \) being as much perpendicular to all gray value gradients \( \mathbf{\nabla} g \) in \( U \) as possible, i. e.

\[
S = \int_U (\mathbf{n}^T (x') \mathbf{\nabla} g(x'))^2 \ d^W x' \rightarrow \text{minimum}, \text{ where}
\]

\[
\mathbf{n}^T \mathbf{\nabla} g = n_x \frac{\partial g}{\partial x} + n_y \frac{\partial g}{\partial y} + n_z \frac{\partial g}{\partial z}
\tag{7}
\]

It can be shown that this expression is minimized if the vector \( \mathbf{n} \) is given by the eigenvector of the tensor \( \mathbf{J} \) to the minimum eigenvalue [17]. The search for local orientation therefore reduces to an eigenvalue analysis of the structure tensor \( \mathbf{J} \).

The 3D structure tensor for \( x-y-t \) image volume has the structure

\[
\mathbf{J} = \begin{bmatrix}
J_{xx} & J_{xy} & J_{xz} \\
J_{yx} & J_{yy} & J_{yz} \\
J_{zx} & J_{zy} & J_{zz}
\end{bmatrix}
\tag{8}
\]

The three eigen values of \( \mathbf{J} \) indicate the anisotropy of the volume, while the eigen-vectors indicate the orientation of the anisotropy. We denote the three eigen values as \( \lambda_1, \lambda_2, \) and \( \lambda_3 \), from biggest to smallest, with corresponding eigen vectors \( \mathbf{e}_k = (e_{k,x}, e_{k,y}, e_{k,z})^T \), for \( k = 1,2,3 \). If the motion occurs, the three eigen values will be different and the motion can be indicated by analyzing the ratio of three eigen values.

Same as HauBecker[8] did, we use the following estimators to determine the type of the tensor:

- coherency = \(|(\lambda_1 - \lambda_3)(\lambda_1 + \lambda_3)|^2 \)
- edge = \(|(\lambda_1 - \lambda_2)(\lambda_1 + \lambda_2)|^2 \)
- corner = \(|(\lambda_2 - \lambda_3)(\lambda_2 + \lambda_3)|^2 \)

The rank of the tensor is determined as Table 1 shows.
When the tensor is ranked 0 or 3, no velocity can be extracted. When its rank is 2, which means the motion can be modeled as a moving corner, the real velocity can be extracted as:

\[ \mathbf{v} = (v_x, v_y)^T = (e_{3,1}, e_{3,2})^T / e_{3,3} \]  

When the rank of the tensor is 1, which means the motion can be modeled as a moving line, this method suffers aperture problem. So only the velocity components normal to the edge can be extracted:

\[ \mathbf{v}_{\text{norm}} = -\frac{e_{1,3}}{e_{1,2}^2 + e_{1,1}^2} (e_{1,1}, e_{1,2})^T \]

The tensor-based method has two problems. One is that the rank determination is based on idealization of the motion situation and the values of the coherency or edge indicator are seldom 0 or 1, because the real motion can be the combined motion of the moving line and the moving corner, neither is perfectly ideal. The second problem is that when the motion estimation suffers aperture problem, no real velocity can be extracted. HauBecker [8] proposed a back projection method, trying to solve the real velocity using local structure information and the normal velocity values in the local neighborhood. But in the plant root growth, the aperture situation occurs in a large area, while in these area, there lacks textures other than the moving “edge”, so the back projection method results in no solution when solving the equation set. When we set the local neighborhood larger to get rich texture or structure information, the back projection method still results with no valid solution because in a larger neighborhood the velocity has become non-uniform.

We solve these two problems in two ways. Because we cannot assume the motion to be ideal moving corner or moving line, the output of the tensor-based method can only be viewed as a mixture of the two situations, but the values of ‘edge’ and ‘corner’ indicate the portions of each situation in the mixed motion. First we ‘trust’ only the tensor with big enough ‘coherency’ and ‘corner’, which indicate that the velocity extracted is confidential. In the moving edge situation, the real velocity can be estimated if we know the direction of the real velocity and the \( \mathbf{v}_{\text{norm}} \). We use many methods to determine or initialize the direction of the velocity, including searching in the neighborhood to find confidential velocity results in moving corner situation, and take advantage of the characteristic of the plant root growth that the elongated motion is much bigger than motion perpendicular to the root elongating direction. The other way we use is that we do not fully trust the velocity extracted by the tensor method, we use it as an initialization velocity value in the robust matching method, so that the searching domain of the matching method can be reduced to a much smaller area.

There is another function of the tensor-based method, segmentation automatically when extracting the motion. We can take the value of ‘coherency’ as a decision factor to determine if any motion occurs. We can get a mask if the sites inside the mask indicate their ‘coherency’ is big enough. In the experiment we found that this mask can be automatically used as segmentation of the root, and we only need to compute the velocity inside the mask. The mask can be dense enough if the threshold value of ‘coherency’ is selected carefully.

### 4. Confidence Test

#### Forward-Backward Matching Consistence Test

From section 3 we can get matching for sites in the image at time \( t \) to sites in the image at time \( t+1 \). Then we do the same work to matching for all sites in the image at time \( t+1 \) to sites in image at time \( t \). We define the procedure of matching as a kind of transformation \( \mathbf{M} \) with

\[ I_f(s) \equiv I_{r+1}(\mathbf{M}_{\text{forward}}(s)) = I_{r+1}(s') \]  

where \( I_f(s) \) is the intensity of site \( s \) for time \( t \), and at time \( t+1 \) the matching method can be reduced to a much smaller area.

We examine the multiple-to-one matches using this consistence test by rejecting those matches with very high errors or those matches with errors not very high but the error do not remain within a certain error range for certain times of reciprocal matching.

In Figure 4 (D) and (E), the motion fields before and after the forward-backward consistence test are shown. The density after the consistence test is about 35% with the \( \Gamma = 1 \) pixel in (15).

Although we may not necessarily constrain the valid matching as rigid one-to-one pixel to pixel matching
(E M(s) < Γ = 1 pixel), the result is still reasonable because noise and deformation may deviate the matching location. If the deviation is not too much, we still can accept it. To alleviate the problems of too many multiple-to-one matching and large E M , we do multiple reciprocal matching between two images at time t and t+1:

1) Specify two criteria Γ and Γ I as the accuracy of the reciprocal matching and 0<Γ I<Γ;

2) For each site s ∈ S,

   Do the round trip matching for s, as
   s→s'→s'', compute E M(s) = ||s'' – s||;

   (a) If E M(s) > Γ , the matching is invalid;
   (b) If E M(s) <= Γ , accept this matching.

   (c) Repeat r times( where r ≈ 1/E M(s) )
       Do the round trip matching for s'', as
       s''→s F→s B;
       If E M = ||s – s B|| < Γ
       then s' = s F, s'' = s B;
       Else the matching for s is invalid;

   For those valid matching the displacement(s) = (2s' – s – s'') /2;

Endfor

One thing should be mentioned is that the forward-backward matching consistence test may not detect the mismatch if the robust matching algorithm itself get the same value for displacement in forward and backward direction. The matching site for s actually depends on the objective function we constructed. It is known that robust matching algorithm is robust to get matches without introducing two much error but the matching site may drift. In the experiment we also noticed that the results for some pixels are different for min-square matching method and the robust matching method.

Statistics Confidence Test
For one turn of matching, we use the sharpness of the main peak/valley of the object function near s to determine if the matching is accurate: Recall that the objective function E(u,v,s) in (1). In the searching domain we compute the variance of E, denoted by σ, and the average value for E(u, v, s), denoted by α; If

|α - MIN(E(u, v, s))| > γ • σ

where γ is a constant pre-defined, then we think the sharpness is high enough and the matching is accurate statistically.

5. Practical Implementation and Output
The preprocessing included illumination calibration and image segmentation. Then for each pair of images in the image sequence, for example, image A and B, we want to find matches from image A to image B. For each pixel in image A except those near the border, we defined a local block with size 15x15, which was big enough to contain texture features and to make the objective function non-convex [1]. We also defined a searching domain around the exact pixel site in image B with the size of 81x31. For each pixels in the searching domain with corresponding local block, we computed the objective function for it with the original local block in A. We did a full search in the searching domain to minimize the objective function for robust matching. The practical implementation of statistics confidence test was combined with the robust matching. In practical implementation of the consistence test, we chose Γ =1 pixel which means no drifting is allowed. Because we reject some mismatches and inaccurate matches the density of the displacement field is not full, but we still got more than 35% confidential matches for γ =2 and Γ =1.

Figure 3 and Figure 4 show the results of the algorithm. In the displacement fields of the final results, the intensity stands for the value of displacement.

![Figure 3](image3.png)

Figure 3. The illustration of a plant root, its lateral velocity field and the velocity profile. The upper panel is the plant root, the left part of the root is the tip of the root. The middle panel is the velocity field after the robust-matching and the
Figure 4. The original image sequence of a segment of a plant root and the motion fields after specific processing. A. The original plant root image in a 9-frame image sequence, only frame 5 is shown; B. The motion mask in the tensor motion estimation method, where the gray and red masks are computed by thresholding the coherency in (9) with 0.1 and 0.8,
respectively. C, D, E, F: the motion fields after tensor method, robust matching method, forward-backward consistency test, and interpolation by normalized averaging. Only the horizontal components of the motion are displayed for each step. The vertical components of the motion are not shown here because they are very small and dominated by the noises.

6. Conclusion
This algorithm proposed in this paper heavily relies on the assumptions of constant illumination and constant growth speed for short time period. Fortunately in our experiment these two assumptions were well satisfied. Although this algorithm did not use the time dimension directly in the robust matching, not like the tensor-based method used by the Jahne group [7][8], this method gives high fidelity displacement output because it does not use approximate approaches like normal optical flow methods and also because this method use forward-backward matching consistence test which makes the output confidential.

However, this approach need to be improved because it only measures the similarity between sites in the neighborhood by robust matching without time dimension involved. It fails when serious deformation or multiple motions occur. Because in this experiment we care the accuracy very much, we naturally developed this method and it works well in the experiment.

Reference