

**ECSE 6650 Computer Vision
Final Project**

**Image Change Understanding
Systems: Application to Neural
Implant Design**

BY:

Yousef A. Al-Kofahi
660393893

Dec. 15th 2005

1. Introduction

The goal of this project is to estimate the deformation field of a soft tissue as a result of needle insertion. In this work, three different insertion speeds are presented and the different deformation fields are estimated. Unfortunately, the optical flow estimation methods introduced in class do not work well on these images, because we are dealing with deformable structure in which the intensity constancy constraint may not always hold.

The approach used for estimating the deformation field is based on three important assumptions. The first assumption states that a small set of neighbor pixels have the same deformation (motion) field. Based on this assumption, we can estimate the overall deformation field of the tissue by dividing the tissue into small windows of pixels and then by measuring the deformation in each of these windows. The second assumption states that no one of those small windows has a constant intensity for all of its pixels and hence there should be a point that best represent each window, and this point is called an interest point. The last assumption states that the deformation field of each small window can be approximated by the deformation field of the corresponding interest point.

Based on the presented assumptions, the algorithm used in this work proceeds as follows. First, the image part containing the tissue is divided into 100 small windows. In each window, we search for the best interest points, which are the points with the maximum Harris responses as explained in the next section. After that, each point is tracked independently. The tracking process is performed by implementing Kalman filter algorithm and by using Normalized Cross Correlation (NCC) for measurements. After tracking the interest points over the successive frames, there might be some outliers. The next step is to remove the outliers by using a two-step method.

The last part of this project focuses on analysis of the tracking results. In this part, the tracking results of the interest points for several movies with different insertion speeds are analyzed, and the outcomes are patterns of the estimated deformation fields resulting from the three different speeds.

2. Extracting Interest Points

2.1 Harris Detector

Interest points are points at which the signal changes two dimensionally. A very large set of interest points detectors have been proposed in literature. Most of them assume that interest points are equivalent to corners or more generally speaking to points characterized by a significant gradient amount in more than one direction such as T-junctions and intersections. (refer to [1] for a reference of a good survey of interest points detectors).

In this work, Harris detector is used. This detector was introduced by Harris and Stephen in [2]. The basic idea of this detector is similar to those of all other corner detectors, which is by calculating the autocorrelation matrix M for each pixel from the small window of surrounding pixels W .

$$M = \begin{bmatrix} \sum_w \left(\frac{\partial I}{\partial x} \right)^2 \otimes w & \sum_w \left(\frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \right) \otimes w \\ \sum_w \left(\frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \right) \otimes w & \sum_w \left(\frac{\partial I}{\partial y} \right)^2 \otimes w \end{bmatrix}$$

where the sum is over W , w is a smoothing Gaussian window used to avoid corners due to image noise, and \otimes is the convolution product.

If at a certain point, the two eigen values of M are large, then a small motion in any direction will cause an important change of grey level, and this indicates that this point is a corner. Figure 1 shows the three possible cases for each point.

For a given pixel (i,j) , the corner response function proposed by Harris is given by:

$$R_{i,j} = \det(M_{i,j}) - k(\text{trace}(M_{i,j}))^2$$

where k is a parameter set to 0.04 – 0.06. $R_{i,j}$ is the $(i,j)^{th}$ entry of R , and it represent the Harris response at pixel (i,j) . R is the responses matrix and its size equals the image size, which means that every point (pixel) in the image has its corner response. In general, interest points are points with large responses. The criteria for selecting these points are presented in the next section.

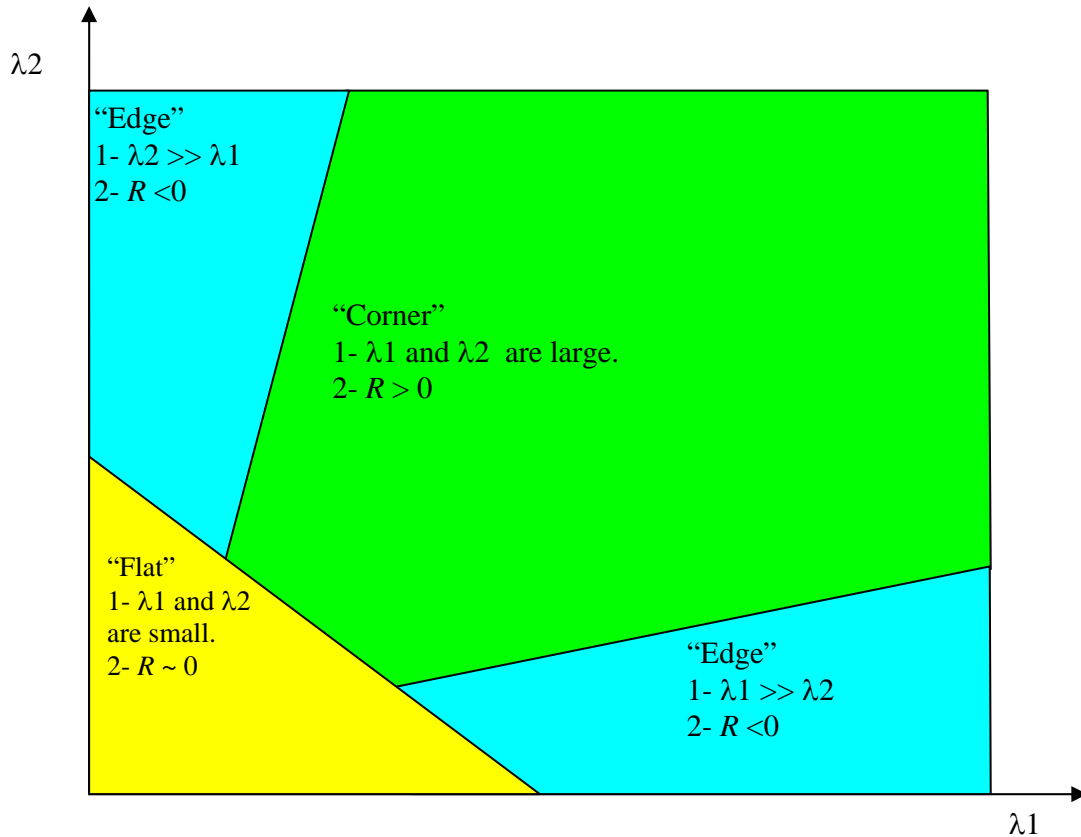


Figure 1: The relationship between R and the eigen values of M . The figure also shows the corresponding classification of a point.

2.2 Choosing interest points

The next step is to choose the interest points based on the calculations done in the previous step. Interest points are chosen as local maxima of the corner responses matrix R . For this purpose, the image was divided into 100 windows, and in each window we look for the point with the maximum Harris response. All the images that we use are of the same size 1040 x 1380. The region of interest is the image region that fall inside the tissue. When this region is divided into 100 windows, each window has the size of 100x100 pixels. In general, in each animation, an area of 1000x1000 in the first frame that fall inside the tissue is selected and divided into 100 small windows of equal sizes. Furthermore, each image is divided horizontally into three parts by drawing two horizontal blue lines. The region between the two lines is the central region, and the others are the upper and the lower regions. These lines are used to make it easier to see

the movement of the needle since several examples showed that the needle goes right and left during insertion. Also, the interest points inside the middle region are marked with yellow while the others are in red. Figure 2 shows an example of a first frame with the initial interest points are selected. Note that the needle is not inserted yet in this image.

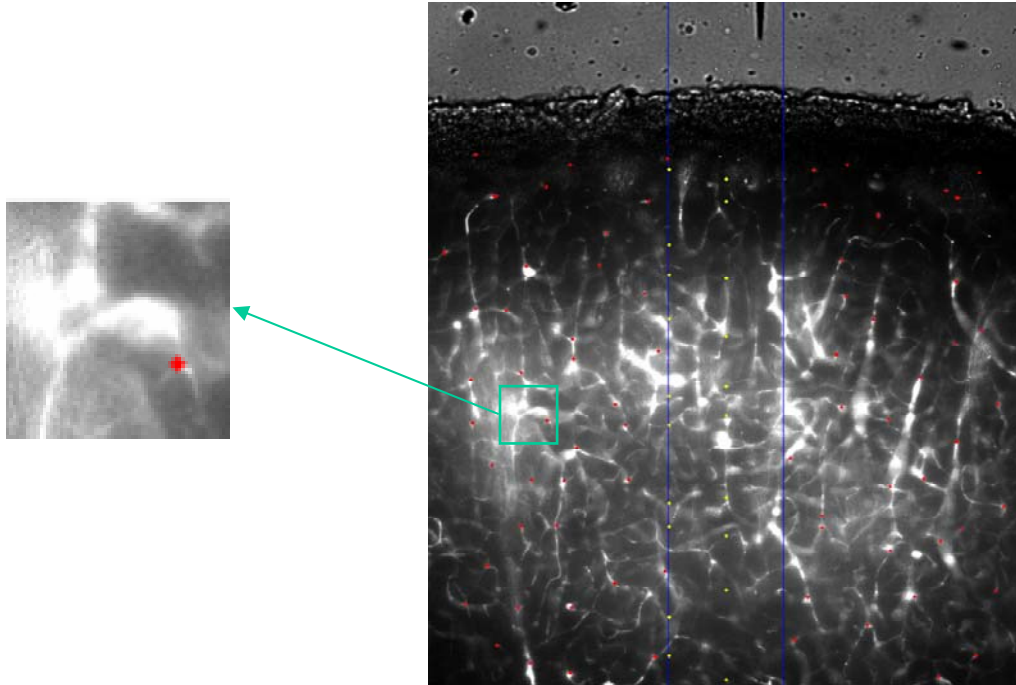


Figure 2: An example of first frame (just before insertion). Interest points are shown in red and yellow. The small image to the left is a zoom-in of one of the 10x10 windows with the best interest point marked in red

In the criteria presented above, each 100x100 window is presented by only one interest point, which is the point with the maximum Harris response. However, the tracking results of some points could be wrong, and hence, the tracking results of the corresponding windows will be wrong. To reduce the probability of having wrong tracking results of a window, three more points are selected in each window resulting in a total of four interest points selected in each window, which have the highest Harris responses. It will be shown in section 4 how the tracking results of each four points are regrouped to obtain the tracking results of the window.

3. Tracking Interest Points

Interest points were assumed to move independently of one another, even though they represent different regions of the same neurovascular network. This assumption was important to permit identification and removal of outliers by comparing tracking results to results from neighboring interest points as discussed later.

Tracking is performed using Kalman filtering. In this way, each point will need one Kalman filter, which is computationally expensive, but since we are not dealing with real time systems, time is not of great concern. As will be discussed in the next subsections, Kalman filter combines its prediction with some measurements (observations). The measurements were obtained by using Normalized Cross-Correlation (NCC).

3.1 Basics in Kalman Filtering

We assume that we have a point $p = (x_t, y_t)$ with velocity $v_t = (v_{xt}, v_{yt})^t$ at a given time instant t . Together, position and velocity are used to express the state of the point at time t such that, $s_t = [x_t, y_t, v_{xt}, v_{yt}]^t$. The goal of tracking turns out to be computing the state vector from frame to frame. Kalman filtering is an approach that linearly relates the state s_{t+1} at the time $(t+1)$ to the previous state s_t by the following state model:

$$s_{t+1} = \Phi s_t + w_t \quad (1)$$

where Φ is the state transition matrix and w_t represents system perturbation which is normally distributed as $w_t \sim N(0, Q)$.

The state model describes the temporal part of the system. If we assume the feature movement between two consecutive frames is small enough, we can consider the motion of feature positions from frame to frame uniform:

$$\begin{cases} x_{t+1} = x_t + v_{xt} \\ y_{t+1} = y_t + v_{yt} \\ v_{x,t+1} = v_{xt} \\ v_{y,t+1} = v_{yt} \end{cases} \Rightarrow \begin{bmatrix} x_{t+1} \\ y_{t+1} \\ v_{x,t+1} \\ v_{y,t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ v_{xt} \\ v_{yt} \end{bmatrix} \Rightarrow s_{t+1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} s_t \quad (2)$$

From (1) and (2), the state transition matrix Φ can be parameterized as

$$\Phi = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

On the other hand, the measurement model of Kalman filtering describes the spatial features of the system:

$$z_t = Hs_t + \theta_t \quad (4)$$

where the matrix H relates the current state to the current measurement, and θ_t represents the measurement uncertainty (or noise) normally distributed as $\theta_t \sim N(0, R)$.

For simplicity, we assume the measurement $z_t = (z_{t,x}, z_{t,y})^t$ to be the measured position of the feature point in an image frame. Then we have

$$z_t = \begin{bmatrix} z_{tx} \\ z_{ty} \end{bmatrix} = \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \theta_t = H \begin{bmatrix} x_t \\ y_t \\ v_{tx} \\ v_{ty} \end{bmatrix} + \theta_t \quad (5)$$

From (4) and (5), we yield

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (6)$$

3.2 Kalman Filtering Algorithm

Kalman filtering consists of two steps; state prediction and state updating. State prediction is performed using the state model while state updating is achieved by using the measurement model.

3.2.1 Prediction

Given the previous state s_t and its error covariance matrix Σ_t , iteration of state prediction involves predicting the current state \bar{s}_{t+1} and estimating the error covariance matrix $\bar{\Sigma}_{t+1}$.

$$\begin{aligned}\bar{s}_{t+1} &= \Phi s_t \\ \bar{\Sigma}_{t+1} &= \Phi \Sigma_t \Phi' + Q\end{aligned}\quad (7)$$

where

$$\bar{s}_{t+1} = [\bar{x}_{t+1}, \bar{y}_{t+1}, \bar{v}_{x,t+1}, \bar{v}_{y,t+1}]^t$$

3.2.2 Updating

Based on the predicted state vector \bar{s}_{t+1} and error covariance matrix $\bar{\Sigma}_{t+1}$, we can compute the Kalman gain matrix K_{t+1} as follow:

$$K_{t+1} = \frac{\bar{\Sigma}_{t+1} H^t}{H \bar{\Sigma}_{t+1} H^t + R} \quad (8)$$

The measurement Z_{t+1} is done by finding the position of the feature point $p_{t+1} = (x_{t+1}, y_{t+1})$ at the time $t+1$. The searching region is determined by the predicted error covariance matrix $\bar{\Sigma}_{t+1}$, and centered by the predicted location $(\bar{x}_{t+1}, \bar{y}_{t+1})$ of the feature point. One way to set the size of the searching area is by using the square roots of the first two diagonal elements of $\bar{\Sigma}_{t+1}$, which will give the standard deviations in x and y . The search is conducted using one of the NCC as shown in the next section.

Once the measurement Z_{t+1} is obtained, we can generate a posteriori state s_{t+1} and error covariance matrix Σ_{t+1} by

$$\begin{aligned}s_{t+1} &= \bar{s}_{t+1} + K_{t+1} (Z_{t+1} - H \bar{s}_{t+1}) \\ \Sigma_{t+1} &= (I - K_{t+1} H) \bar{\Sigma}_{t+1}\end{aligned}\quad (9)$$

3.3 Normalized Cross-Correlation (NCC)

Given an image \mathbf{I} and a template window \mathbf{T} , centered at the point (u, v) in the search window, the normalized cross-correlation (NCC) between the template and its projection on the image (a window centered at (u, v) and with the same size as \mathbf{T}) is given by:

$$NCC_{I,T}(u,v) = \frac{\sum_{x,y} [(I(x,y) - \bar{I}_{u,v})(T(x+u, y+v) - \bar{T})]}{\{\sum_{x,y} [I(x,y) - \bar{I}_{u,v}]^2 \sum_{x,y} [T(x+u, y+v) - \bar{T}]^2\}} \quad (10)$$

This operation is performed over all the points inside the search window, and the maximum value of NCC is at the point that gives the best match. The normalized cross-correlation overcomes some disadvantages of the common cross-correlation operator: (1) Cross-correlation is invariant to a linear (affine) transformation of image intensities (a constant addition and common scaling of intensity), and (2) the range of CC depends on the size of the template window. The solution to these problems is to normalize both the template and image windows to unity. NCC produces a peak with a value of 1.0 at the perfect match between the template window and its corresponding window in the search area. The center of this matching window will be the measurement of the new interest point location. NCC is relatively robust in the presence of noise, changes in scale and gray level, and image deformation [3].

4. Experimental Results and Analysis

4.1 Rejecting Outliers

As outlined in section 2, interest points were treated as independent from one another and now, they are regrouped to reject the outliers by choosing a single point. The portion of the image containing the tissue was initially divided into 100 windows, with the goal of identifying one interest point in each block that represents it in terms of deformation (Fig. 3). In some cases interest points may not reflect overall tissue deformation in a given region; for example, when the point and its vascular counterpart are being dragged by the device during insertion. To reduce the risk of tracking errors, the four best points were tracked. The interest point with the highest response among a group of four is called parent interest point. The deformation occurring on this parent interest point is computed after using the following two-step method for filtering the results:

1. Direction Constraint: Deformation in the direction of device insertion should be greater or equal to the deformation perpendicular to this axis. Any point that does not

pass this test is rejected. Rejection is accomplished by making the deformation on this point equal to the deformation on the nearest neighbor point that passes this test.

2. Local median filtering: Median filtering of every four interest points was performed by taking the median of their x and y deformations, and considering the resultant values as the deformation of their parent interest point.

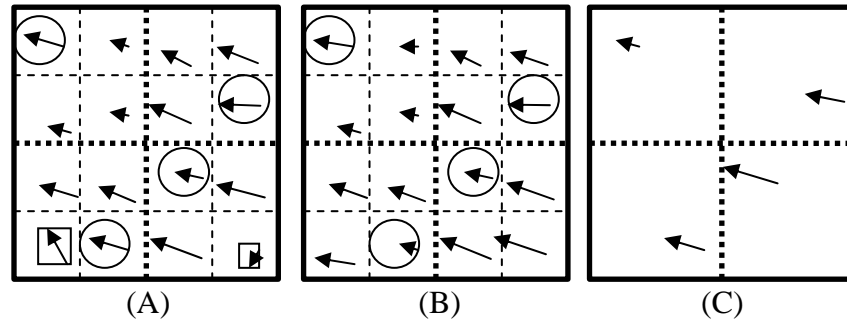


Figure 3: Representation of the filtering process for four windows with four interest points in each one. The arrows shown are the deformation vectors for these points. The circled one is the parent interest point and the points inside rectangles are outliers.

(A): The resulting tracking of the points (B): The tracking results after median after correcting outliers (C): the median of each four deformations is assigned to the parent interest point.

4.2 Tracking Results

In this work, nine animations were tested. The nine animations divided into three groups of three different speeds with three animations in each group. For each animation, the tracking results of the 100 parent interest points are obtained and overlaid on the images. Through the successive frames, interest points locations are marked with either yellow, if it is in the central region identified by two blue lines, or red if the point is outside the central region. The movement paths of the points are shown in green. Figure 4 shows an example of three frames from a slow insertion animation. Full animations will be found on the project's web folder. Also, the tracking results of the points are exported into Excel sheets with each row shows a frame. Figure 5 shows a screen shot of one of those excel sheets. The complete set of excel sheets is also uploaded into the same results folder.

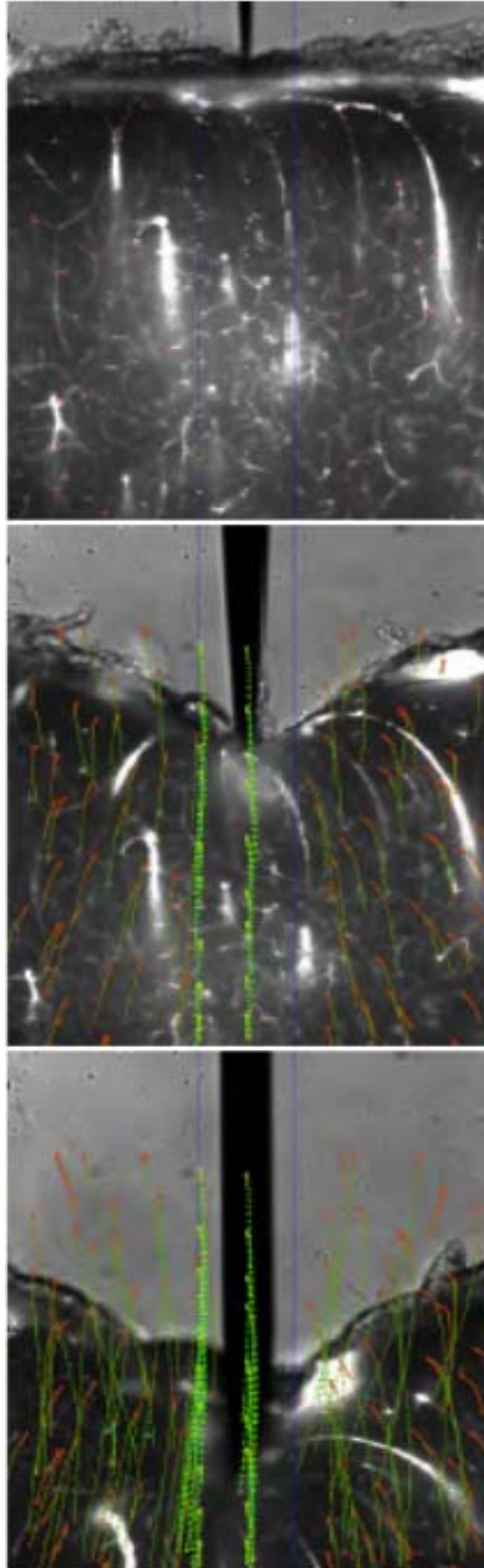


Figure 4: Selected frames from a slow (0.125 mm/s) insertion tracking results. Interest points are marked in red or yellow, and their motion is shown in green lines in subsequent images.

	A	B	C	D	E	F	G	H	I	J	K
1	Frame	row	col	row	col	row	col	row	col	row	col
2	1	117	59	101	169	26	254	83	345	113	563
3	2	117	59	101	168	26	253	83	344	113	562
4	3	117	58	100	166	25	251	82	342	112	560
5	4	117	58	99	163	24	248	81	339	110	557
6	5	117	57	99	160	24	245	81	335	110	553
7	6	116	56	97	156	22	240	79	331	108	549
8	7	116	55	96	151	21	235	77	326	106	543
9	8	116	54	94	146	19	230	75	321	103	537
10	9	115	52	92	140	16	224	72	315	100	531
11	10	115	51	91	134	15	218	71	308	100	524
12	11	114	49	89	128	13	211	69	302	98	516
13	12	114	49	87	123	10	206	66	296	95	510
14	13	114	48	86	119	9	202	65	292	94	505
15	14	114	48	85	116	8	197	65	287	93	499
16	15	114	48	85	116	9	197	65	287	94	499
17	16	114	48	86	116	10	197	66	287	95	499
18	17	113	47	84	114	8	195	64	284	93	497
19	18	113	47	84	111	8	192	64	282	93	494
20	19	113	46	84	108	8	189	64	278	93	489
21	20	113	46	84	106	8	186	65	275	94	485
22	21	113	46	83	104	7	184	63	273	92	484
23	22	113	46	84	102	9	182	65	271	95	480
24	23	113	46	83	100	7	178	63	266	93	476
25	24	113	46	82	99	4	174	60	262	90	470
26	25	113	46	82	98	4	170	60	258	90	465
27	26	113	46	82	96	4	169	60	256	90	463
28	27	113	46	83	96	6	167	63	254	94	460
29	28	113	46	82	95	4	164	61	250	92	455
30	29	113	46	82	93	4	162	60	247	91	450
31	30	113	46	82	92	4	159	60	244	91	446
32	31	113	46	82	91	4	157	60	240	91	441
33	32	113	46	83	90	4	153	61	236	91	435

Figure 5: Screen shot of an excel sheet showing a sample tracing results. Each row represents a frame, and each point is represented by its row and column locations in that frame.

4.3 Experimental Analysis

The last step is to estimate a simple presentation of the deformation field for each case. For this purpose, the tracking results of each group of three animations are combined by taking their averages at each of the 100 points in each of the frames. After that, the tissue area is divided into 9 main sectors representing different regions of the tissue. In each of the nine sectors, the average displacement from the first frame to the last frame is calculated. The results are converted from motion paths of 100 points to 9 motion (deformation) vectors of the 9 sectors mentioned above. Figure 6 shows three examples from three different insertion speeds with the tracking results superimposed with the nine deformation field vectors.

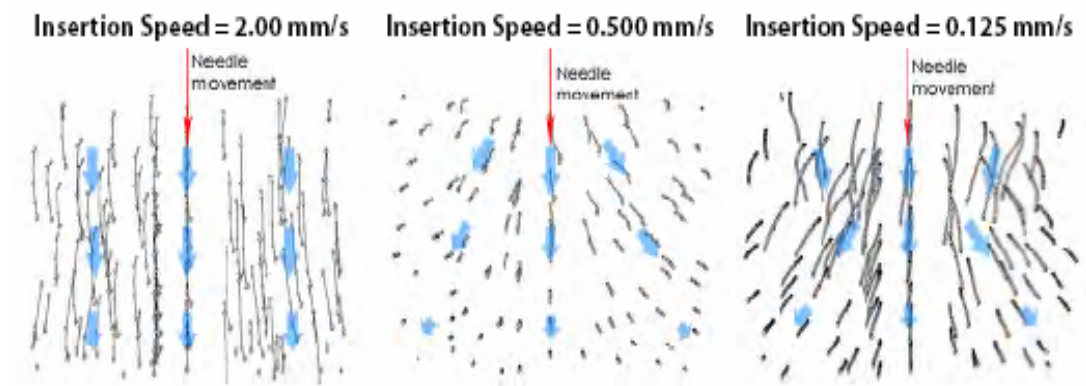


Figure 6: Distinct patterns of tissue deformation are observed at different insertion speeds. Interest points are marked with open circles; movement between consecutive frames is indicated by black lines. The blue arrows define the strain field

5 Summary and Conclusions

Tissue deformation resulting from needle insertion is studied in this project and deformation patterns are estimated for three different insertion speeds. The method was based on Kalman Filtering for tracking. Also, Normalized Cross-Correlation (NCC) was used for measurements. The presented method is expensive in terms of computation time, but since we are not dealing with real time application, computation time was not of high priority. In summary, the presented method provides an objective and fully automated approach to modeling the tissue deformation.

6 References

- [1] Schmid, C., Mohr, R., Bauckhage, C (1998). Comparing and evaluating interest points. *Proceedings of the 6th International Conference on Computer Vision*, Bombay, India.
- [2] Harris, C., Stephens, M. (1988). A combined corner and edge detector. *Proc. of 4th Alvey Vision Conference*, 147-51.
- [3] Martin, J., Crowley, L. (1995). Experimental comparison of correlation techniques. *IAS-4, International Conference on Intelligent Autonomous Systems*, Karlsruhe.