1992

On Elastic Nets and Correspondence in Motion with Homogeneous Deformation

Anupam Joshi

Chia-Hoang Lee

Report Number:
92-087

ON ELASTIC NETS AND CORRESPONDENCE IN MOTION WITH HOMOGENEOUS DEFORMATION

Anupam Joshi
Chia-Hoang Lee

CSD-TR-92-087
November 1992
Revised October 1993

On Elastic nets and correspondence in motion with homogeneous deformation*

Anupam Joshi

Department of Computer Sciences

Purdue University

West Lafayette, IN 47907 USA †

Chia-Hoang Lee

Department of Computer and Information Sciences

National Chiao-Tung University

Hsinchu, Taiwan 30050 R.O.C

Abstract

An important task in feature based dynamic image analysis schemes is establishing correspondence between tokens across image frames. In this work, the authors propose a method to do this between range data sequences when the body in motion also homogeneously deforms. It is possible to establish correspondence when motion parameters

*This work was supported in part by the National Science Council of the R.O.C under grant NSC 81-0408-E-009-565

†This author would like to acknowledge the support of a research fellowship from the Purdue Research Foundation

describing the range data sequence are available. The problem of correspondence thus reduces to one of obtaining these parameters. We cast this problem as an optimisation problem and then apply Elastic net techniques to solve it. Assumptions which underly the method are identified. This formulation is used to obtain the most probable parameters of motion and deformation that describe the image sequence, which are then used to recover correspondence. Results of applying this technique to range data sequences are also presented to establish its veracity .

1 Introduction

Pentland in [28] avers that " To date, almost all research on recovering global, wholebody structure from optical flow has been based on rigid motion". A more general statement along the same lines could be made about the recovery of structure and motion, and would be just as true, whether the underlying mechanism be optical flow based or feature based. Establishing correspondence of feature points across image frames is a prerequisite to obtaining the motion parameters in feature based schemes of motion. It is defined by Ullman [36] as the process by which elements in different views are identified as representing the same object at different times, thereby maintaining the perceptual identity of objects in motion.

While the objects in the real world are three dimensional, research in the area of correspondence has mostly dealt with their two dimensional images and has almost invariably assumed rigidity of motion as one of it's underlying assumptions [31, 15, 37, 2, 19, 30, 40, 20]. However, with the increasing availability of equipment to do range sensing, the problem of establishing correspondence between range data, the three dimensional representation of the object, is also becoming prominent. This issue, again with the assumption of rigidity, has been addressed in some previous work.

Huang and Chen [8] have proposed a scheme that uses preestablished correspondence between three points. If p_1, p_2, p_3 be points from the first frame, and q_1, q_2, q_3 be their corresponding points in the second frame, the correspondence can be established as follows. Tetrahedron p_1, p_2, p_3, p_i is congruent to tetrahedron q_1, q_2, q_3, q_j *iff* points p_i and q_j correspond. In [22] Huang and Lin propose a technique which works very effectively in a noise

free case. They use centroids of the two token sets to obtain two new sets of tokens which are related by rotation only. Let p_1 and p_2 be the two point sets, and let c_1 and c_2 be their centroids, respectively. They obtain token sets q_1 and q_2 as

$$q_{1i} = p_{1i} - c_1$$

and

$$q_{2i} = p_{2i} - c_2 \ ,$$

where the subscript i denotes the i^{th} member of the token set q_1 or q_2.

These new point sets are used to get four candidates for the rotation matrix. Correspondence is obtained from these by chosing the correct R. Another technique, which can tolerate noise better is proposed in [21]. This however, involves obtaining a good initial estimate to the rotation axis, and uses Fourier transforms, making it computationally expensive. Magee and coauthors [24] have used subgraph matching when the objects in the scene are polyhedral or cylindrical to obtain correspondence in range data. They also propose an interesting method to find suitable "feature points" in the object for which range data is obtained. Some other approaches to this problem can be found in [32, 11, 14]. Shuster [32] uses a quadratic loss function to obtain an optimal rotation matrix, and reduces this problem to finding the optimal quaterion. Faugeras and Herbert [11] use a similar technique applied to the vertices and planes of an object, in order to match it with a model by obtaining optimal Translational and Rotational motion parameters which relate the range data with a stored model. This method however is not computationally very efficient. In order to do an image to model match, Grimson and Lozano-Perez[14] use an involved tree pruning approach. Their approach requires knowing the surface normal at each measured

point, and uses distance and angular constraints to do a matching.

There has been some work done in the area of non rigid motion vis a vis registration of images. This class of works essentially involves deforming the boundaries of some image to make it match a model, as in the method proposed by Burr in [4, 5]. Starting from a pair of overlaid images, his mechanism stretches the elastic image towards similar features in the goal image. This process is continued with decreasing stiffness of the image, leading to a sequence of warped images, each a better approximation to the goal. Correspondence can be obtained in such a scheme by tracking a feature point across this image sequence. Kass *et. al.* [18] propose active contour models for this task, which they call snakes. These are essentially energy minimising splines, guided by external forces that pull them towards image features, such as lines and edges. Along similar lines, [34, 35, 33] proposes the use of deformable 3D bodies composed of a simulated elastic material. These are "deformed" under influence of forces that are generated by image features and the motion of the object. Pentland [28] uses the physics underlying nonrigid motion to obtain estimates of object shape and velocity. He also proposes some closed form solutions for recognition of 3D deformable objects[29]. Most such work however assumes the correspondence as a given.

The problem of establishing correspondence can in general be reformulated as one of minimising some suitably constructed energy function. Neural Networks are often used to solve such optimisation problems, and have been used in the case of correspondence, especially in stereo correspondence. Chellappa and Zhou [41] construct an energy function which contains terms for disparity values across regions and smoothness. A three dimensional Hopfield net is then used to minimise it and obtain correspondence across multiple frames. Maeda[23] uses a 3 layer network and a training procedure to learn from examplars.

In [26], the authors use a neural net to extract feature and then use a Hopfield type net to obtain correspondence. More recently, Nasrabadi and Choo[27] propose a Hopfield Network to solve the stereo correspondence problem. Joshi and Lee [16] present a solution to stereo correspondence using elastic nets. The problem with most such approaches is that they try to obtain effects of deformation by individually moving each feature point to optimise the energy criterion. Thus the computation cost increases rapidly with increasing number of feature points.

In this work, we propose an alternate approach to the correspondence problem in range data for non rigid motion. Analogous to some of the work done in the area of correspondence in a rigid motion framework [8, 22, 17, 20], the process involves trying to obtain an estimation of the global motion and deformation parameters that describe the image sequence, and using these to establish correspondence. We use the technique of elastic nets in obtaining the estimates to the motion and deformation parameters. In the sections to follow, we begin by briefly describing elastic nets, and touching upon the Physics and kinematics of deformation. We then formulate the correspondence problem when the underlying motion is not rigid and present our scheme to solve it. Results of simulating the scheme are presented. We also propose a method that can track more than one object in the frame sequence if we are allowed to make the assumption of rigidity.

2 Elastic Nets

Durbin and Willshaw in [10] proposed a novel scheme to solve combinatorial problems that involve geometrical structures and topographical mappings between them. To illustrate

with their example of the Travelling Salesperson problem, consider a small, circular rubber band placed in the middle of the map of the cities. The rubber band is now gradually and non uniformly stretched such that the neighbouring points stay together, and the rubber band passes through all the cities. The path described by the rubber band is thus a valid tour and close to optimal. The forces that stretch this rubber band are obtained as combinations of forces that try to draw points on the rubber band to cities, and forces that keep neighbouring points on the rubber band together. Durbin and Willshaw formulate the energy function of the system as

$$\mathcal{E} = -\alpha K \sum_i \ln \sum_j \phi(d_{ij}, K) + \beta \sum_j |y_{j+1} - y_j|^2 \qquad (1)$$

where $d_{ij} = |x_i - y_j|$ and $\phi(d, K) = \exp(-d^2/2K^2)$

The x_i's represent the coordinates of the cities and y_j's represent the coordinate of the points on the rubber band. They show that if there are more points on the rubber band than there are cities(in their simulation, the ration is 2.5) and in the limit that $K \to 0$, a valid, close to optimal tour is produced. Since $\mathcal{E}$ is bounded from below, it requires that as $K \to 0$,

$$\forall x_i \; \exists y_j \;\; \text{s.t.} \;\; |x_i - y_j| \to 0.$$

This ensures that the tour will pass through all cities. Moreover, as the number of points on the rubber band is increased, the second term in the energy function is minimised by placing all points at equal distances from each other. If $\mathcal{D}$ be the total path length, such a configuration makes the value of the second term $\frac{\mathcal{D}^2}{\text{Number of points}}$, which is obviously minimised by reducing the path length.

To obtain the tour then , we merely need to do gradient descent on the energy surface

defined by $\mathcal{E}$, which is achieved by updating the positions of the points on the rubber band, y_j, by $K\partial\mathcal{E}/\partial y_j$ at each iteration step. Computing this quantity, we obtain Δy_j, the change in value of y_j at a given iteration as

$$\Delta y_j = \alpha \sum_i w_{ij}(x_i - y_j) + \beta K(y_{j+1} - 2y_j + y_{j-1})$$

where

$$w_{ij} = \frac{\phi(d_{ij}, K)}{\sum_l \phi(d_{il}, K)}$$

We may point out here that for large values of K, the energy function of equation (1) is smooth, and has only one minimum. As the value of K is reduced, the function has many minima, corresponding to many possible tours of the city. Their algorithm proceeds by starting with a high value for K, and then slowly reducing is so as to track the global minimum. While this does not guarantee that the minimum eventually achieved will be a global one, it has been observed empirically that the minima are within a few percent of the global. In [9], it is also shown that minimising the energy function is the same as trying to obtain a Bayesian MAP estimate for the tour. We refer the interested readers to [10, 9] for a detailed exposition and analysis of this technique.

Durbin and Willshaw noted that this approach produced better tours than the Hopfield net, and this method scaled better with the number of cities as well. Wilson and Pawley [38] have also extensively documented the problem of scaling in Hopfield nets. In [39], Yuille *et. al.* used methods from statistical physics to analyse the performance of Elastic and Hopfield nets. They showed that the Hopfields nets are not as effective as elastic nets because

- They take into account only the nearest neighbour interactions amongst cities

- The constraints on "valid" final configurations are enforced by adding biases in the energy function, rather than explicitly used in computing the partition function

This is important in light of the fact that most neural approaches to the correspondence problem have used hopfield nets[27, 41].

3 The Physics of Deformation

In order to understand the kinematics of a solid, one must assume that the material of the solid is distributed in some region of space, and that all points in this region are occupied by an element of the solid. The element occupying a particular point is identified by the coordinates of that point. In real bodies the relative position of material particles change when external forces act on them, and consequently, motion is very often accompanied by deformation as well. This is in contrast to ideal rigid bodies, where only the absolute, not the relative position of the points undergoes change.

In general, the equation of motion of a system is defined by

$$F = Md'' + Kd' + Gd. \tag{2}$$

where M is the mass, K is the damping, G is the material stiffness between the points of the body, and F is the external force applied. Here, d denotes the distance moved by each point, d' and d'' are it's first and second derivates w.r.t. time. Assuming that continuos material properties such as mass and stiffness can be discretised at the feature points (using Finite Element Methods), we can come up with a system of equations describing point displacements as a function of the applied force. Pentland in [28, 29] shows how such a system can be diagonalised and solved. The deformations induced in such a case are, in

general, nonlinear, although [28] suggests that they can be accurately approximated by a linear superposition of the various deformation mappings.

A slightly different approach to deformation is to treat it as a linear process. The configuration of the solid at any time can be related by a functional to its original configuration in absence of any forces. In this work, we treat our first frame to be the original configuration of the object. Consider a case where a point X_K, $K = 1 \ldots 3$ in the original configuration moves to a point x_k, $k = 1 \ldots 3$ in the configuration of the second frame. Further, a point with coordinates $X_K + dX_K$ moves to $x_k + dx_k$, then we can establish that

$$dx_k = F_{kK} dX_K.$$

Here, dx_k and dX_K are respectively the elements of order 1 tensors dx and dX that describe the same line element in the two configurations, and the F_{kK}s are the components of a second order cartesian tensor $\mathbf{F}$. This is called the *deformation gradient tensor*. Recall that an order 1 tensor is a 3-vector, and an order 2 tensor is a 3×3 matrix. A similar mechanism can be used to show how $\mathbf{F}$ relates surface and volume elements of the object in the two configurations.

Since the mapping of points between the two configurations is one to one, $\mathbf{F}$ has to be nonsingular. Given its nonsingularity, $\mathbf{F}$ can be uniquely decomposed using the polar decomposition theorem into $\mathbf{VR}$ or $\mathbf{RU}$. The tensor $\mathbf{R}$ is proper orthogonal, and since $|\mathbf{F}|$ is positive, so is $|\mathbf{R}|$. The tensors $|\mathbf{U}|$ or $|\mathbf{V}|$ are positive definite and symmetric. We point out here that $\mathbf{U} = (\mathbf{F}^\mathbf{T}\mathbf{F})^{1/2}$ and $\mathbf{R} = \mathbf{F}\mathbf{U}^{-1}$. This decomposition has a physical meaning. If we chose to split $\mathbf{F}$ as $\mathbf{VR}$, then the net motion can be looked upon as a rigid rotation followed by a deformation, a stretching along the three axes. If $\mathbf{F}$ is split the other way

around, then that refers to a deformation followed by a rigid rotation.

In general, the tensor $\mathbf{F}$ is a function of the position *i.e.* $\mathbf{F}$ is different for each point of the object. If we impose the additional constraint that $\mathbf{F}$ is independent of position, the the object is said to undergo *homogeneous* deformation. For an in depth explanation of the kinematics of deformation, we refer the reader to [3, 25].

4 Problem Formulation

We now examine the problem of non rigid motion. Let A_i' be a point token from the first set and B_i' be a token from the second set. We can represent the correspondence by a permutation σ such that the point $B'_{\sigma(i)}$ from the second frame corresponds to the token A_i' in the first frame. Let $\mathbf{R}$ and $\mathbf{T}$ be the rotation and translation, respectively, that define the motion from the first to the second frame. Assuming that the motion is rigid, we get

$$B'_{\sigma(i)} = \mathbf{R}A_i' + \mathbf{T} \tag{3}$$

To add in nonrigidity in such a formulation, one possible approach is to try and obtain the rigid motion parameters on a patch by patch basis, with some limited provision for nonrigidity. To paraphrase Gibson (as quoted in [28]), however, it does not make much sense to decompose the elastic motion of an object into the rigid motion of it's constituent elementary particles (this, of course, under the newtonian illusion of particles being rigid, solid entities). So we try to obtain a global representation describing deformation. We have shown in the previous section that *homogeneous* deformations can be modelled as a linear transformation defined by a gradient deformation tensor. For non rigid motion under the assumption of homogeneous deformation then, we can relate the feature points in the two

frames as

$$B'_{\sigma(i)} = \mathbf{F}A'_i + \mathbf{T} \tag{4}$$

Since a homogeneous deformation does not affect the centroid, we can translate the origin to the centroid of the point set for each frame and then use a scheme like [21] to obtain new point sets A and B that are only related by $\mathbf{F}$.

$$B_{\sigma(i)} = \mathbf{F}A_i \tag{5}$$

Recall that $\mathbf{F}$ can be decomposed into two matrices $\mathbf{R}$ and $\mathbf{D}$, where $\mathbf{R}$ is orthogonal, and $\mathbf{D}$ is symmetric and positive definite. Thus we can write

$$B_{\sigma(i)} = \mathbf{R}\mathbf{D}A_i \tag{6}$$

We have in (5) and (6) a global description of the motion and deformation parameters describing the image sequence. It also provides us with the basis for our solution to the correspondence problem.

5 Establishing Correspondence

Let us suppose that by some as yet unspecified method, we can obtain the matrix representing the gradient deformation,$\mathbf{F}$. Then, correspondence can be trivially established by observing that if point i corresponds to point j, then $B_j \equiv \mathbf{F}A_i$. If correspondences are unique, then this is a necessary and sufficient condition for establishing them. Suppose that instead of getting $\mathbf{F}$, we get an approximation $\mathbf{L}$ to it. Correspondence can then be

established by observing that $d_{ij} = min_{\ k}\ d_{kj}$ where d_{ij} is the distance between points B_j and $\mathbf{L}A_i$

We will now see how elastic nets can be used to obtain an approximation to the gradient deformation matrix. We consider a system similar to the TSP. In this case, instead of stretching a rubber band, let us instead think of moving the points in the second frame such that they coincide with their corresponding points in the first frame. The important point to remember here is that unlike the TSP, each point in the second frame does not move independently. Instead the movement of all points is described by $\mathbf{L}$.

Let us describe the energy of our system as

$$\mathcal{E} = -\alpha K \sum_i ln \sum_j \phi(d_{ij}, K) \tag{7}$$

where $\phi(d_{ij}, K) = e^{\frac{-d_{ij}^2}{2K^2}}$.

The distance d_{ij} is used slightly differently here. In the original work of Durbin and Willshaw, each point on the contour could move independently of the others, and so distances between the current position of the contour and points representing the cities are computed directly. In our case however, the motion is defined by a global parameter. Hence, we do not individually move points from the first frame to the second. We merely look for the parameter of motion and deformation that would do it. Thus, we define d_{ij} to be $|\mathbf{G}B_j - A_i|$, where $\mathbf{G}$ is $\mathbf{L}^{-1}$. Recall that the deformation tensor is nonsingular, and hence its inverse exists. This is similar in concept to Yuille's "deformable templates"[39]. Observe that (7) is simply the first term of the energy function proposed in [10], with a suitably altered definition of the distance metric. The minimisation of this term ensures that

$$\lim K \to 0, \quad \forall A_i\ , \exists B_j \ \ s.t. \ \ d_{ij} \to 0 \tag{8}$$

In other words, for every point in the first frame, a corresponding point in the second frame exists. The second term of the energy function in [10] is used to minimise the path length of the TSP tour, and is of no consequence in this problem.

Let us now introduce some concepts from statistics which we will use to discuss our method. Note that correspondence can be described by a binary field V, where element V_{ij} has the value one or zero according to whether the feature i in the first image and j in the second correspond. The partition function is defined as the sum of probability distributions over all possible field configurations. It is known [12] that all the mean values of a field can be obtained from its partition function. It is also known that any problem formulated as the minimisation of an energy function can be interpreted probabilistically in terms of Gibbs distributions. Greater discussion on the use of statistical methods in vision problems can be found in [13, 12, 39]. In [13], Geman and Geman present a bayesian method for image restoration. Geiger and Girosi [12] showed that instead of using simulated annealing or other Monte Carlo type methods, a deterministic method could be used to compute the estimates to the mean values of the fields of the kind proposed in [13]. Yuille in [39] gives a good overview of such techniques and relates them to elastic net like approaches. Following his arguments [39], it can be shown that minimising the energy function that we propose will give the correct values for the transformation matrix.

Let us define the energy E as

$$E(V_{ij}, F) = \sum_{ij} V_{ij} |FB_j - A_i|^2 / 2K^2,$$

where F is the transformation matrix (deformation tensor) and V is a binary field denoting matching. The states of this system can be related to a Gibbs distribution $G(V_{ij}, F) =$

$\frac{e^{-E(V_{ij},F)/2K^2}}{Z}$ where Z is the partition function. This can be written as $Z = \sum_{(V_{ij},F)} e^{-E(V_{ij},F)/2K^2}$. The summation thus is over all possible values for V_{ij} and F. This can be rewritten as

$$Z = \sum_{V_{ij}} \sum_{F} e^{\sum_{ij} V_{ij}|FB_j - A_i|^2/2K^2}$$

$$Z = \sum_{V_{ij}} \sum_{F} [\prod_{i} e^{\sum_{j} V_{ij}|FB_j - A_i|^2/2K^2}]$$

We will now factor out the matching field, by computing the sum over $\{V_{ij}\}$s. Notice that ideally the correspondence should satisfy the uniqueness principle, *i.e.*, for a given i, there is only one j such that V_{ij} is one, and for any given j, there is only one i such that V_{ij} is one. Imposing this constraint leads to a very complicated system, so we impose a weaker constraint, namely that each point in the first frame should find a unique match in second. This means that for any i, there is only one j which has V_{ij} as one. This leads to

$$Z = \sum_{F} [\prod_{i} (\sum_{j} e^{|FB_j - A_i|^2/2K^2})].$$

We can now rewrite the partition function as $Z = \sum_F e^{-(1/\alpha K)\mathcal{E}}$, where $\mathcal{E}$ is the energy function we are minimising. Finding the global minimum for $\mathcal{E}$ with respect to F as $K \to 0$ will lead to the value of F which will also be the *maximum a posteriori* estimation.

A similar result can be obtained by following the arguments presented in [9]. We have seen that the energy function $\mathcal{E}$ can be related to a Gibbs like distribution by exponentiation. Let us define the distribution as

$$W(F,K) = \frac{1}{(2\pi K^2)^n n^n} e^{\frac{-\mathcal{E}}{\alpha K}}, \tag{9}$$

and n is the number of points in the frame. This can be rewritten as

$$\prod_{i} \frac{1}{n} \{\sum_{j} \frac{1}{2\pi K^2} e^{\frac{-a_{ij}^2}{2K^2}}\} \tag{10}$$

with d_{ij} as $|FB_j - A_i|$. This can be looked upon as a product of n independent probability distributions, $P_i(F)$

$$P_i(F) = \frac{1}{n}\{\sum_j \frac{1}{2\pi K^2} e^{\frac{-d_{ij}^2}{2K^2}}\} \tag{11}$$

$P_i(F)$ measures the probability that the parameter F moves some feature point of the second frame to within a certain distance of the i^{th} point in the first frame. The distance itself is controlled by the parameter K. We are treating the $P_i's$ as independent here, which may at first sight appear counterintuitive. Since correspondence is unique, the probability of a point being matched to a particular point in the second frame is dependent on whether that point has already been matched. However, we have seen that in our method a weaker version of the uniqueness of match constraint is imposed. Namely that each point in the first frame be matched to only one point in the second. This allows points in the second frame to be matched to multiple points in the first frame, thus making $P_i's$ independent. Such states as would allow multiple matches are however clearly energetically unfavourable.

Thus (10) in essence is the probability of the motion parameter being **F** given the two feature point sets A and B. Note that the value of F which minimises $\mathcal{E}$ maximises P. So equation(10) asserts that in the process of minimising $\mathcal{E}$, we are obtaining the most probable estimate to G such that each point in the first frame finds a match in the second. As with the TSP algorithm, we very slowly reduce the value of K as we iterate.

The minimum of $\mathcal{E}$ can be obtained by doing a gradient descent with respect to the parameter **G**, which is composed of **R** and **D**. Observe that in equation (7), d_{ij} is essentially a function of the parameter **G**. Let us assume that each element g_{ij} of **G** is independent of each other. This allows (7) to be differentiated in a straightforward manner to obtain $\partial\mathcal{E}/\partial g_{ij}$. Each element is updated as $g_{ij}^{l+1} = g_{ij}^{l} - K\partial\mathcal{E}/\partial g_{ij}^{l}$, where l is the iteration number.

As the theoretical descriptions above show, in the limit $K \rightarrow 0$, each point in the first frame would find an exact match in the second frame. Yet this involves a very large number of iterations. To hasten up the process substantially, we modify the scheme somewhat.

1. Carry out some prespecified number of iterations of the gradient descent.

2. Using the G obtained from the above, obtain correspondence by matching each token in A to its nearest neighbour in $\mathbf{G}B$

6 Results of the Simulation

In order to verify the proposed technique, and to examine if our simplifications hold well under experimental conditions, several simulations were performed. In line with [8], data were generated in as points in a cube, each of whose sides were two hundred units, with the origin as one of the corner points. The x, y and z coordinates of the points were chosen as independent random numbers. The data from the second set was obtained by applying a deformation, rotation and translation to the points of the first frame.

The algorithm outlined in the previous section was applied to a large number of data set, and included different values for the motion parameters. As an initial approximation, the matrix G was set to be the identity matrix. It was observed that this gross initial approximation sufficed, in all but a few cases, to obtain correspondence. Also, it was observed that less than a hundred iterations were needed for a sufficiently close approximation to G be obtained, as would allow correspondence to be established. In Table 1, we present the actual F matrix that was used to obtain the second frame from the first, as well as the approximation to G obtained by our method after a hundred iterations. We also show

the product $\mathbf{FG} = \mathbf{FF^{-1}}$ as a measure of the accuracy to which the approximation was obtained. Ideally, this product should be the identity matrix. For case 'a', the translation vector was $[10, 15, 15]^T$ and while for case 'b', the translation vector was $[100, 150, 150]^T$. These examples had ten points in each frame, and all were correctly matched by our method. Evidently, there are instances where $\mathbf{I}$ is not adequate as an initial approximation. In such cases, we can use better approximations if available. We can also reduce the constant involved in the gradient descent equation and continue to iterate longer. Many simulations were performed, with different data sets and deformation tensor values. Some of them are summarised in Table 2. Sets a & b had one $\mathbf{D}$ matrix, sets c & d another. For each run in the sets, a different $\mathbf{R}$ was generated by randomly generating values of tilt, slant and angle of rotation. For sets a & d, these values were between 0 and 40 degrees, for sets b & c they were between 0 and 90 degrees. In the table, we present the mean, median and standard deviation for the number of points correctly matched and the average distance between matched points. We also computed the matrix $\mathbf{E} = \mathbf{F}\ \mathbf{G}$ - $\mathbf{I}$, which should ideally be the zero matrix. To give some idea of how close our approximation was, we computed an error measure for each row of E, defined as $\sum_j |E_{ij}|$. The table presents the statistics of mean, median and standard deviation for each row of $\mathbf{E}$ as well.

In real world situations, the data are often noisy. Noises can be generated in the process of sensing, data acquisition and such. So the algorithm was next tested with noisy data to evaluate it's ability to operate in a real environment. The possible errors in data were modelled as zero mean additive noises. Once again many simulations, with different motion parameters, as well as added noises of different variances, were performed. The results are in line with those we report for the case of rigid motion in [17]. It was observed that

$$\mathbf{F} = \begin{bmatrix} 1.263663 & -0.738239 & 0.193711 \\ 1.613742 & 1.495957 & 0.437550 \\ 0.221344 & 0.988492 & 1.667641 \end{bmatrix}$$

$$\mathbf{G}_{cmp} = \begin{bmatrix} 0.426814 & 0.296231 & -0.126158 \\ -0.541969 & 0.432317 & -0.051356 \\ 0.264561 & -0.296791 & 0.646713 \end{bmatrix}$$

$$\mathbf{FG} = \begin{bmatrix} 0.989465 & 0.003352 & 0.001908 \\ 0.001414 & 0.996066 & -0.001468 \\ -0.001483 & -0.000025 & 0.999872 \end{bmatrix}$$

case(a)

$$\mathbf{F} = \begin{bmatrix} 2.806663 & 1.173833 & 0.442341 \\ 1.632646 & 2.149435 & 0.112575 \\ 0.712118 & 2.522310 & 3.924495 \end{bmatrix}$$

$$\mathbf{G}_{cmp} = \begin{bmatrix} 0.401450 & -0.181767 & -0.034499 \\ -0.332936 & -0.544164 & 0.024661 \\ 0.114616 & -0.303334 & 0.202392 \end{bmatrix}$$

$$\mathbf{FG} = \begin{bmatrix} 0.805406 & -0.006478 & 0.021726 \\ -0.028451 & 0.841037 & 0.010772 \\ -0.029422 & -0.00691 & 0.810839 \end{bmatrix}$$

case(b)

Table 1:

	Error in Rows			No. of Points matched	Average Distance
Mean	0.0310	0.1024	0.0972	10	3.8952
Median	0.0277	0.1024	0.0952	10	3.8584
Std. Dev.	0.0048	0.0026	0.0073	0	0.0854

Run a

	Error in Rows			No. of Points matched	Average Distance
Mean	0.4862	0.3984	0.6113	7.8812	12.4241
Median	0.0411	0.1051	0.1117	10	4.1933
Std. Dev.	0.7793	0.5589	0.8683	3.5589	13.5508

Run b

	Error in Rows			No. of Points matched	Average Distance
Mean	0.4166	0.3450	0.5937	8.27	12.8034
Median	0.0814	0.1273	0.1439	10	5.2291
Std. Dev.	0.6560	0.4978	0.7952	3.2188	12.6602

Run c

	Error in Rows			No. of Points matched	Average Distance
Mean	0.1590	0.0571	0.1893	10	5.8165
Median	0.1620	0.0558	0.1848	10	5.8027
Std. Dev.	0.0219	0.0041	0.0234	0	0.4562

Run d

Table 2: Summary of Results from Synthesised data

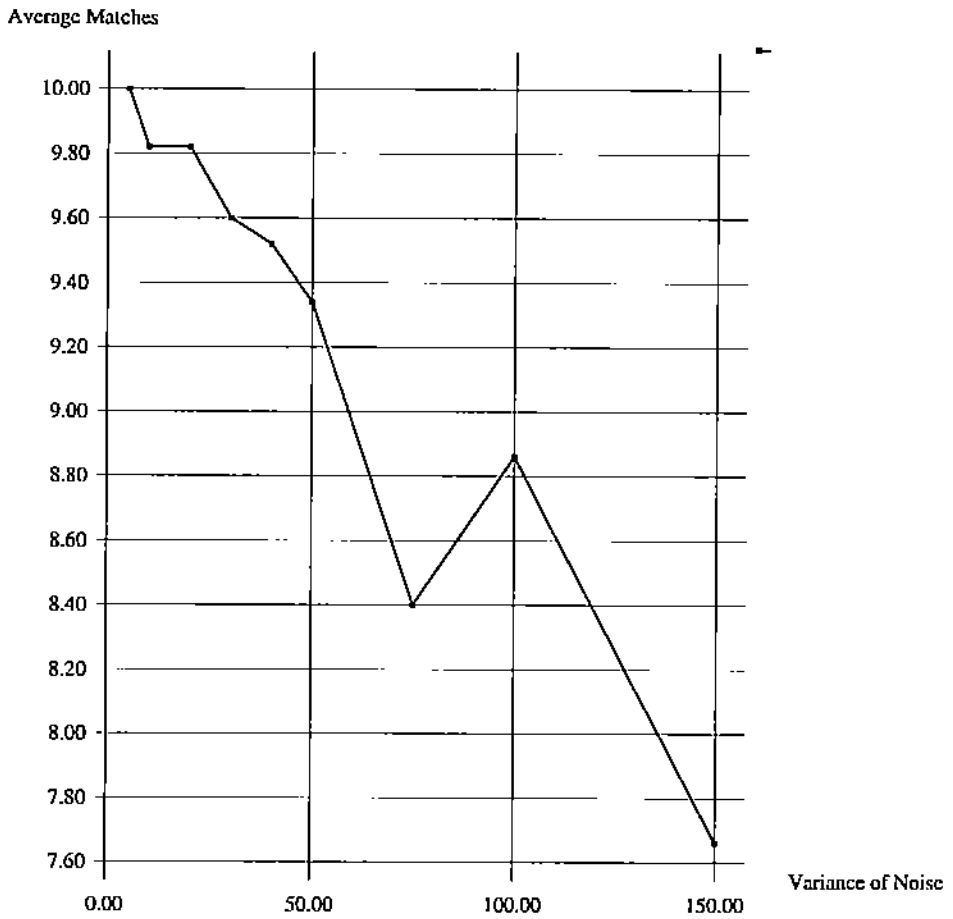

Figure 1: Graph of Average Match versus variance of noise added

the method was able to handle noises with variances of about 50 units with negligible degradation of performance. This variance corresponds to about 25% of the edge length of the cube in which data points were chosen to lie. However, higher values of variances saw a degradation of performance, and the method becomes unreliable after the variation of the added noise increases beyond 100 units. In Figure 1 we plot a graph between the variance of the noise added and the average(over 50 runs) of the number of points correctly matched, which illustrates the above.

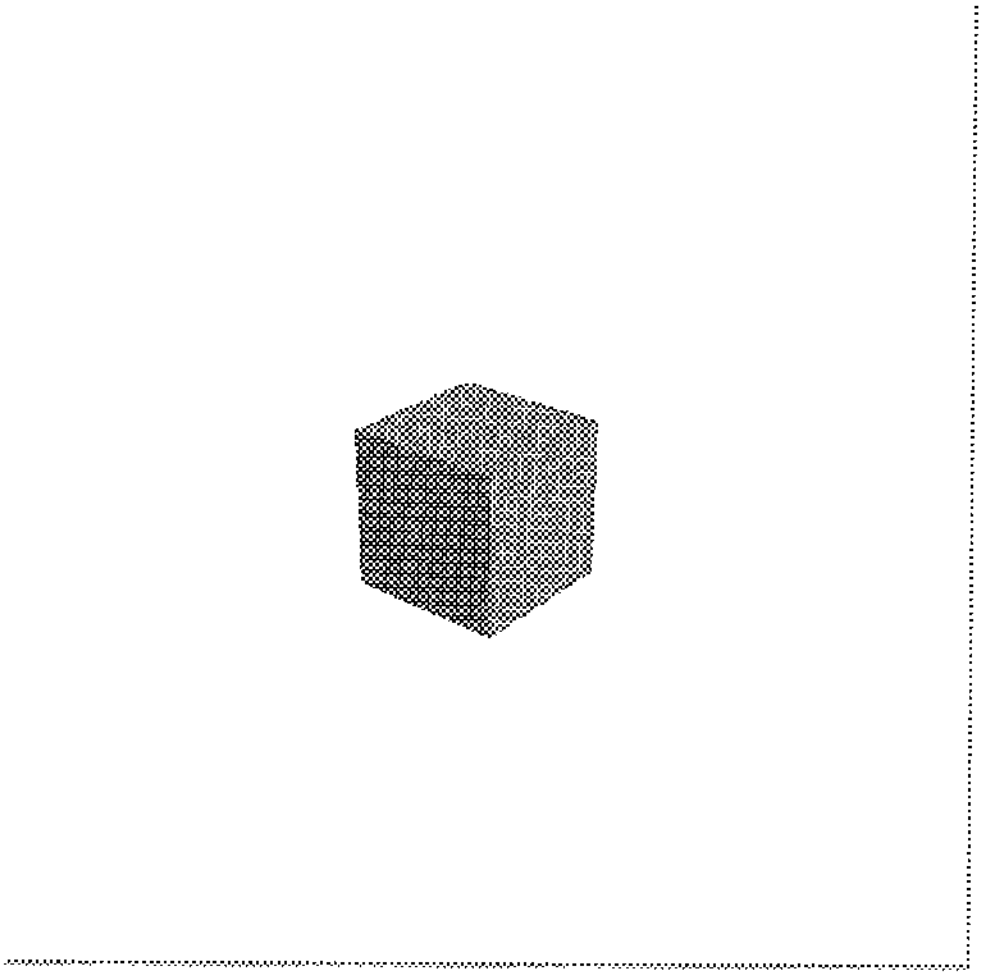

Figure 2: A Cube Shaped Object – Frame 1

In Figures 2, 3, 4 and 5, we show two frames of real objects on which we also did experiments. The first figure is a cube, of which we had data on 8 vertices, and the second is the skeleton of a wrench, which had 20 points.[1] The deformation matrices for these two cases are given in Table 3. In both these cases, our algorithm was able to successfully recover the correspondence of all the points.

[1]We wish to acknowledge the help of Vinod Anupam and Dr. Bajaj of Project Shastra[1] here at the Purdue CS department for their help in obtaining and rendering these range data.

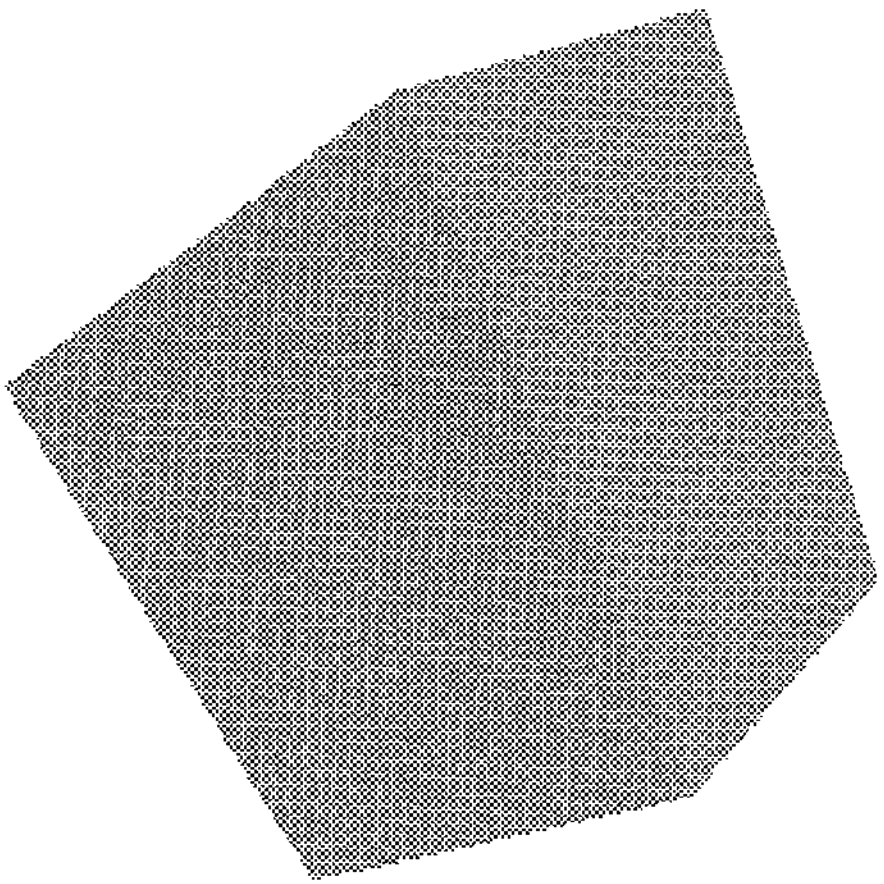

Figure 3: A Cube Shaped Object – Frame 2

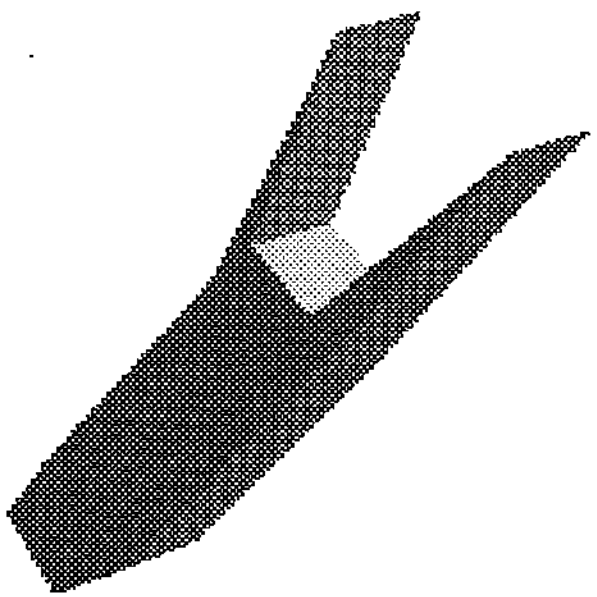

Figure 4: A Wrench Shaped Object – Frame 1

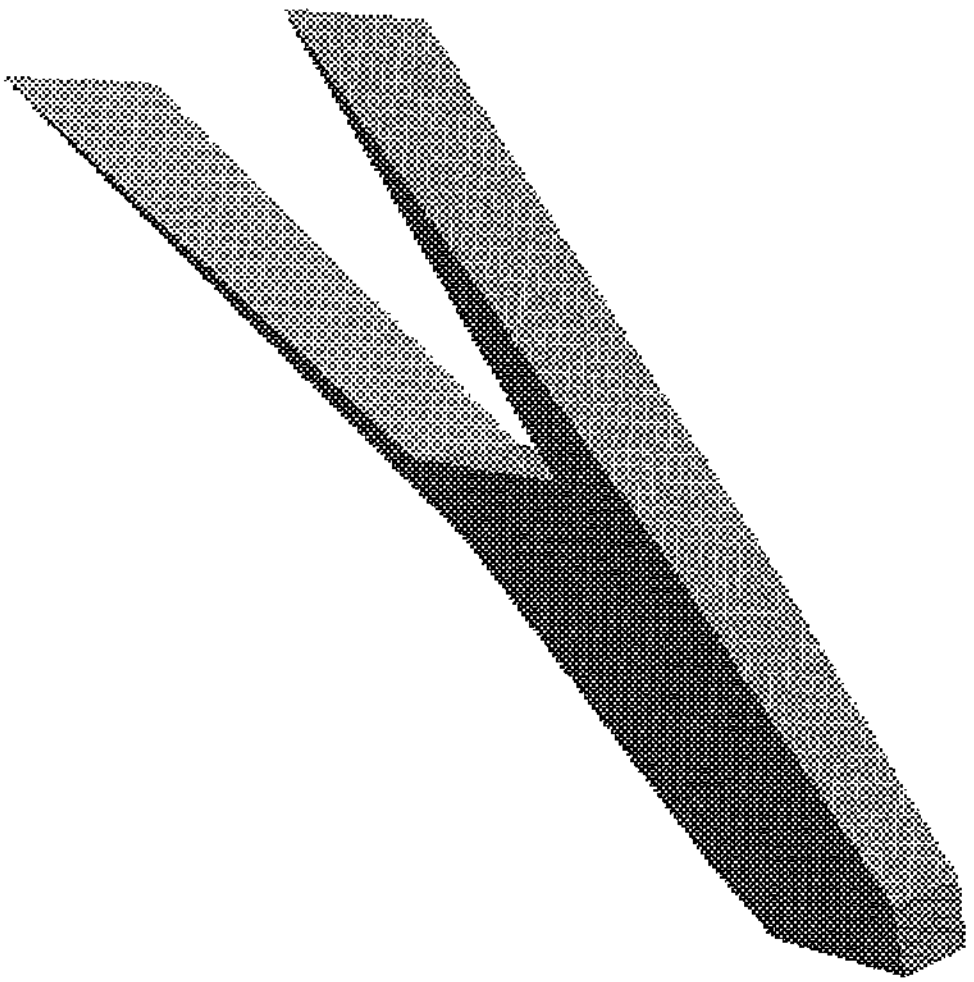

Figure 5: A Wrench Shaped Object – Frame 2

7 Conclusion

While there has been some work done in the area of establishing correspondence under rigid motion assumptions, the problem has been all but ignored for the non rigid case. The work that exists in the area of recovering non rigid motion assumes that the correspondences are already established [28, 6, 7]. Some works try to obtain the nonrigid motion as a series of deformations of an elastic 3D object[35]. Similar techniques have also been used to establish stereo correspondence [4, 5].

We present in this effort a technique that utilises concepts behind elastic nets to obtain correspondence in range data when the underlying motion is not rigid. Since Elastic Nets are inspired by mechanisms to establish ordered neural projections between structures of similar geometry, our method inherits some amount of biological plausibility. It can be implemented on a neural substrate, and is amenable to improved speeds using techniques that exploit the parallelism inherent in Elastic Nets. The method uses approximations to the motion and global deformation parameters to do this. Results of simulations carried out to establish the correctness of the technique have also been presented.

Let us reiterate again that while we obtain approximations to the motion parameters as an intermediate step, we do not strive to compute them very accurately. Once correspondence is established, other approaches like those presented in [6] can be used to obtain the motion parameters. Moreover our method works only for homogeneous deformations. We have also largely ignored in the preceding discussion the questions of choosing parameters when doing the gradient descent. While it is possible to obtain some mathematical insight into this issue [9], we chose the parameters empirically. We have also ignored the question

of obtaining the feature points at which range data is obtained. This subject is well dealt with in other literature.

References

[1] V. Anupam, C.L. Bajaj, and A.V. Royappa, *The shastra distributed and collaborative geometric design environment*, Tech. Report CSD-TR-91-075, Department of Computer Science, Purdue University, 1992.

[2] H. Baker, *Depth from edge and intensity based stereo*, Ph.D. thesis, Computer Science Department, Stanford University, 1981.

[3] E.W. Billington and A. Tate, *The Physics of Deformation and Flow*, McGraw-Hill International, USA, 1981.

[4] D.J. Burr, *A dynamic model for image registration*, Computer Graphics and Image Processing **15** (1981), 102–112.

[5] ———, *Elastic matching of line drawings*, IEEE Transactions on Pattern Analysis and Machine Intelligence **3** (1981), no. 6, 708–713.

[6] S. Chaudhuri and S. Chatterjee, *Estimation of motion parameters for a deformable object from range data*, Proceedings International Conference on Computer Vision, 1989.

[7] ———, *Motion analysis of a homogeneously deformable object using subset correspondences*, Pattern Recognition **24** (1991), no. 8, 739–745.

[8] Homer.H. Chen and T.S. Huang, *Maximal matching of two 3 D point sets*, Proceedings International Conference on Pattern Recognition, 1986, pp. 1048–1050.

[9] R. Durbin, R. Szeliski, and A.L. Yuille, *An analysis of the elastic net approach to the travelling salesman problem*, Neural Computation **1** (1989), 348–358.

[10] R. Durbin and D. Willshaw, *An analogue approach to the travelling salesman problem using an elastic net method*, Nature **326** (1987), 689–691.

[11] O.D. Faugeras and M. Herbert, *A 3D recognition and positioning algorithm using geometrical matching between primitive surfaces*, IJCAI, 1983, pp. 996–1002.

[12] D. Geiger and F. Girosi, *Parallel and deterministic algorithms from MRF's: Surface reconstruction*, IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991), no. 5, 401–412.

[13] S. Geman and D. Geman, *Stochastic relaxation, gibbs distributions and the bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence **6** (1984), no. 6, 721–741.

[14] W.E.L. Grimson and T. Lozano-Perez, *Model based recognition and localisation from sparse range or tactile data*, The International Robotics Research Journal **3** (1984), no. 3, 3–35.

[15] R. Jain and I.K. Sethi, *Finding trajectories of feature points in a monocular image sequence*, IEEE Transactions on Pattern Analysis and Machine Intelligence **9** (1987), 56–73.

[16] A. Joshi and C.H. Lee, *Elastic nets and stereo correspondence*, Proceedings IJCNN '92, Beijing, 1992.

[17] ———, *On the problem of correspondence in range data and some inelastic uses for elastic nets*, Tech. Report CSD-TR-92-058, Department of Computer Science, Purdue University, 1992, submitted to IEEE Trans. Neural Networks.

[18] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: Active contour models*, Proceedings International Conference on Computer Vision, 1987.

[19] Y.G. Leclerc and S.W. Zucker, *The local structure of image discontinuities in one dimension*, IEEE Transactions on Pattern Analysis and Machine Intelligence **9** (1987), 341–355.

[20] C.H. Lee and A. Joshi, *Correspondence problem in image sequence analysis*, Pattern Recognition **26** (1993), 47–61.

[21] Z.C. Lin and T.S. Huang *et al*, *Motion estimation from 3D point sets with and without correspondence*, Proceedings Conference on Computer Vision and Pattern Recognition, 1986, pp. 194–201.

[22] Z.C. Lin, H. Lee, and T.S. Huang, *Finding 3 D point correspondences in motion estimation*, Proceedings International Conference on Pattern Recognition, 1986, pp. 303–305.

[23] E. Maeda, *Generalised lyered neural networks for stereo disparity detection*, Proceedings International Joint Conference on Neural Networks, vol. 1, 1990, pp. 487–490.

[24] M.J. Magee, B.A. Boyter, C.H. Chien, and J.K. Aggarwal, *Experiments in intensity guided range sensing recognnition of three dimensional objects*, IEEE Transactions on Pattern Analysis and Machine Intelligence **7** (1985), 629–636.

[25] A.K. Mal and S.J. Singh, *Deformation of Elastic Solids*, Prentice Hall, Englewood Cliffs, NJ, USA, 1991.

[26] M.S. Moussavi and R.J. Schalkoff, *A neural network approach for stereo vision*, Proceedings of IEEE SouthEastCon90, 1990.

[27] N.M. Nasrabadi and C.Y. Choo, *Hopfield network for stereo correspondence*, IEEE Transactions on Neural Networks **3** (1992), no. 1, 5–13.

[28] A. Pentland and B. Horowitz, *Recovery of nonrigid motion and structure*, IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991), no. 7, 730–742.

[29] A. Pentland and S. Sclaroff, *Closed form solutions for physically based shape modelling and recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991), no. 7, 715–729.

[30] K. Rangarajan and M. Shah, *Estimating motion correspondence*, Proceedings Conference on Computer Vision and Pattern Recognition, 1991.

[31] V. Salari and I.K. Sethi, *Feature point correspondence in presence of occlusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990), 87–91.

[32] M.D. Shuster, *Approximate algorithms for fast optimal attitude computation*, AIAA Guidance and Control Specialists Conference, 1978, pp. 88–95.

[33] D. Terzopoulos and D. Metaxas, *Dynamic 3D models with local and global deformations: Deformable superquadrics*, IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991), no. 7, 703–714.

[34] D. Terzopoulos, A. Witkin, and M. Kass, *Symmetry seeking models for 3D object recognition*, Proceedings International Conference on Computer Vision, 1987.

[35] D. Terzopoulos, A. Witkin, and M. Kass, *Constraints on deformable models: Recovering 3D shape and nonrigid motion*, Artificial Intelligence **36** (1988), no. 1, 91–123.

[36] S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, MA, USA, 1979.

[37] T.D. Williams, *Depth from camera motion in a real world scene*, IEEE Transactions on Pattern Analysis and Machine Intelligence **2** (1980), 511–516.

[38] G.V. Wilson and G.S. Pawley, *On the stability of the travelling salesman problem algorithm of hopfield and tank*, Biological Cybernetics **58** (1988), 63–70.

[39] A.L. Yuille, *Generalised deformable models, statistical physics, and matching problems*, Neural Computation **2** (1990), 1–24.

[40] D. Zhang, *Perspective invariant description of a planar point set and its application to matching*, Proceedings International Conference on Pattern Recognition, 1986.

[41] Y-T. Zhou and R. Chellappa, *Neural network algorithms for motion stereo*, Proceedings International Joint Conference on Neural Networks, vol. 2, 1989, pp. 251–258.

$$\mathbf{F} = \begin{bmatrix} 2.722846 & -0.268833 & 0.440495 \\ 1.394635 & 2.586481 & 0.437888 \\ 0.226062 & 1.348274 & 1.680541 \end{bmatrix}$$

For the cube shaped object

$$\mathbf{F} = \begin{bmatrix} 1.948826 & -1.290351 & 0.201673 \\ 2.354724 & 2.272880 & 0.577809 \\ 0.259152 & 1.322501 & 1.683884 \end{bmatrix}$$

For the wrench shaped object

Table 3: Deformation Matrices