MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1526 February, 1995

Direct Object Recognition Using No Higher Than Second or Third Order Statistics of the Image

Kenji NAGAO **B. K. P. HORN**

This publication can be retrieved by anonymous ftp to publication.ai.mit.edu.

Abstract

Novel algorithms for object recognition are described that directly recover the transformations relating the image to its model. Unlike methods fitting the conventional framework, these new methods do not require exhaustive search for each feature correspondence in order to solve for the transformation. Yet they allow simultaneous object identification and recovery of the transformation. Given hypothesized corresponding regions in the model and data (2D views) — which are from planar surfaces of the 3D objects — these methods allow direct compututation of the parameters of the transformation by which the data may be generated from the model. We propose two algorithms: one based on invariants derived from no higher than second and third order moments of the image, the other via a combination of the affine properties of geometrical and differential attributes of the image. Empirical results on natural images demonstrate the effectiveness of the proposed algorithms. A sensitivity analysis of the algorithm is presented. We demonstrate in particular that the differential method is quite stable against perturbations — although not without some error — when compared with conventional methods. We also demonstrate mathematically that even a single point correspondence suffices, theoretically at least, to recover affine parameters via the differential method.

Copyright © Massachusetts Institute of Technology, 1995

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-91-J-4038. BKPH was also supported by NSF grant Smart Vision Sensors 9117724-MIP.

1 Introduction

Object recognition is one of the central problems in computer vision. The task of model-based object recognition (e.g., [4]) is to find the object model in the stored library that best fits the information from the given image. The most common methods of model based object recognition fall into two categories from the point of view of how the objects are represented and how they are matched:

The first represents objects by a set of local geometrical features — such as vertices that can be fairly stably obtained over different views — and matches the model features against the image features, typically in a exhaustive manner. In general, this type of method *simultaneously* identifies the object and recovers the transformation. Equivalently, it recovers the pose of the object that would yield an image from the object model in which the projected features best matches those found in the given image (e.g., [4, 7, 24, 11]). One such method is based on the 'hypothesize and test' framework. It first hypothesizes the minimum number of correspondences between model and image features that are necessary to compute the transformation e.g.,[7, 24]. Then, for each hypothesized set of corresponding features, the transformation is computed and then used to reproject the model features onto the image features. The hypothesized match is then evaluated based on the number of projected features that are brought into close proximity to corresponding image features, and the pair of the transformation and model with the best match is selected.

While this approach has achieved remarkable success in recognizing objects, particularly in dealing with the problem of occlusions of object surfaces, it still has practical computational problems, due to its exhaustive search framework. For example, even with a popular algorithm [7] for matching model objects with m features with image data with n features, we have to test on the order of m^3n^3 combinations, where m and n are easily on the order of several hundreds in natural pictures.

On the other hand, approaches in the second category represent objects by more global features. One method of this type is the moment invariant method. It combines different moments to represent the object, and matches the object model and image data in moment space[6, 19, 1]. The chosen combinations of moments are designed so that they are invariant to the image transformations of concern, such as translations, dilation, and rotations. Thus, emphasis is mainly placed on the *identification* of the object in terms of the object model represented by the combinations of the moments, rather than on the recovery of the transformation between the model and the image data.

In addition, most authors have not addressed the problem of general affine transformation case (instead only treating translation, dilation and scaling). An exception is the method by Cyganski et. al.[2] based on tensor analysis. They developed a closed form method to identify a planar object in 3D space and to recover the affine transformation which yields the best match between the image data and the transformed model. The basis of their method is the contraction operation of the tensors[12, 9] formed by the products of the contravariant moment tensors of the image with a covariant permutation tensor that produces unit rank tensors. Then, further combining those with zero-order tensors to remove the weight, they derived linear equations for the affine parameters sought after. This method is quite elegant, but, it turns out that it needs at least moments up to fourth order. In general, the second type of method is very efficient when compared with the first type of method, that is, methods based on local features plus exhaustive search. At the same time, methods based on invariants tend to be very sensitive to perturbations in the given image data. For example, Cyganski's algorithm is known to be very efficient computationally, however, since higher order moments are notorious for their sensitivity to noise[18], it is very fragile when it comes to perturbations in the image data, being particularly sensitive to local occlusions of object surfaces.

The algorithm that we propose in this paper can be classified in the second category for the reason given below. It is more efficient than conventional approaches in the first category, yet more stable than conventional methods of the second category: (1) it relies on the presence of potentially corresponding image fragments over different views, that are from planar patches on the surface of the 3D objects, (2) it provides a non-recursive, that is, closed-form, method for object recognition. The method does not require complete image regions to be visible and does not depend on the use of local features such as edges or 'corners.' Our method also recovers the transformation from the object model to the image data, but, unlike Cyganski's method, it does not use moments of order higher than second or third order. Therefore, compared with Cyganski's method, it should be less sensitive to perturbations. In addition, we also present another new approach to robust object recognition using differential properties of the image.

Thus, we propose two different algorithms: one based on an affine invariant unique to the given image, which uses up to second or third order moments of the image, and the other via a combination of second order statistics of geometrical and differential properties of the image. Both algorithms recover the affine parameters relating a given 2D view of the object to a model composed of planar surfaces of a 3D object under the assumption of orthographic projection[20, 10]. We also demonstrate that such methods based on the differential properties of the image are fairly stable against perturbations. Of course, the results are not perfect in the presence of perturbations, but the new method does provide much better results than conventional methods using global features. Although we do not explicitly address the problem of how to extract corresponding regions for planar patches in different views, it is known to be fairly feasible using one of several existing techniques (e.g.,[22, 21, 23, 13]). Once we have recovered the affine transformation for the planar patches, we know that by using the 3D object model we can immediately recover the full 3D information of the object [7]. Therefore, our algorithm is aimed at direct 3D object recognition, by first recognizing planar surfaces on the object, and then recovering full 3D information, although the recovery of 3D information is

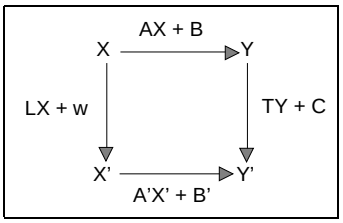

Figure 1: Commutative Diagram of Transformations
Given model feature X and corresponding data feature X', we seek conditions on the transformations A, A' such that this diagram commutes.

not explicitly addressed in this paper. Some experimental results on natural pictures demonstrate the effectiveness of our algorithm. We also give here an analysis of the sensitivity of the algorithm to perturbations in the given image data.

2 Recovering affine parameters via an affine invariant plus rotation invariant using no higher than second/third order moments

In this section, we present a closed form solution for recovering the affine parameters with which a given image can be generated from the model, using an affine invariant theory that we have recently proposed. We first summarize the affine invariant description (up to rotations) of the image of planar surfaces. Then, using this property, we show how the affine parameters are recovered via direct computation in conjunction with the rotation invariant using moments of the image.

2.1 An affine invariant up to rotations: a unique class of linear transformations

In [15, 16, 17], we showed that there exists a class of transformations of the image of a planar surface which generates unique projections of it up to rotations in the image field. It was precisely shown that this class of transformations is the *only* class of linear transformations which provides invariance up to rotations, as long as we are concerned with no higher than second order statistics of the image(see [15]). This property is summarized in the following theorem.

[Theorem]
Let X be a model feature position and X' be the corresponding data feature position in the 2D field. We can relate these by

$$X' = LX + \omega \tag{1}$$

where L is a 2×2 matrix and ω is a 2D vector. Now suppose both features are subjected to similar linear transformations

$$Y = AX + B \tag{2}$$
$$Y' = A'X' + B' \tag{3}$$
$$Y' = TY + C \tag{4}$$

where A, A', T are 2×2 matrices and B, B', C are 2D vectors. Then, if we limit T to an orthogonal matrix, a necessary and sufficient condition for these linear transformations to commute (i.e. to arrive at the same values for Y') for all X, X' (see Figure 1), as long as only up to second order statistics of the features are available, is

$$A = cU\Lambda^{-\frac{1}{2}}\Phi^T \tag{5}$$
$$A' = cU'\Lambda'^{-\frac{1}{2}}\Phi'^T \tag{6}$$

where Φ and Φ' are eigenvector matrices and Λ and Λ' are eigenvalue matrices of the covariance matrices of X and X' respectively, U and U' are arbitrary orthogonal matrices, and c is an arbitrary scalar constant. The terms $[\cdot]^{\frac{1}{2}}$ denote square root matrices[8] and $[\cdot]^T$ means matrix transpose. □

Furthermore, it was shown[15] that when (1) represents the motion of a plane, and both Φ and Φ' represent rotations/reflections simultaneously, and U and U' are set to some rotation matrices, then T in (4) can be constrained to be a rotation matrix. As another aspect of this normalization process, we know that transformations A, A' defined in (5) and (6) transform the respective distributions to have a covariance matrix that is the identity matrix. Arguments were also given on the physical explanations of this property for the rigid object case. In [15, 16, 17], to recover the affine parameters using this property, we used clustering technique to derive three potentially corresponding clusters in the model and data 2D features and used their centroids as matching features in the alignment framework.

In this section, we present other methods to directly recover the affine parameters using this invariant property. Recall that once we have normalized the image using the transformations given in (5), (6), the shapes are unique up to rotations. Thus, if we can compute the rotation matrix T in (4) which relates the normalized data image from the normalized model we can recover the affine transformation L by

$$L = A'^{-1}TA \tag{7}$$

where the translational component has been removed, using the centroid coincidence property[2, 15]. Note however, that, since this normalization process transforms the covariance matrices into identity matrices times a scale factor, the covariances can no longer be used to compute the rotation angle between the normalized model and data features. So, we need to use some other information to determine this rotation angle.

2.2 Computing the rotation angle using second order weighted moments of the image

Although the binary image of the model and the data are normalized by the matrices A, A' so that they have identity covariance matrices, the *weighted* moments of the image function – for instance brightness of the image – are *not* normalized in that sense. Therefore, we can compute the rotation angle between the normalized binary images of model and data by first using the orientation of the major axes of the image computed in terms of the *weighted* moments with respect to fixed coordinates. We then take the difference between the absolute

orientations of the model and the data computed in this fashion to give the relative rotation angle.

$$\tan 2\theta = \frac{2M_{1,1}}{M_{2,0} - M_{0,2}} \tag{8}$$

where $M_{i,j}$'s are second order weighted moments of the normalized image given in the following:

$$M_{2,0} = \sum_x \sum_y (x^2) f(x,y) \tag{9}$$

$$M_{1,1} = \sum_x \sum_y (xy) f(x,y) \tag{10}$$

$$M_{0,2} = \sum_x \sum_y (y^2) f(x,y) \tag{11}$$

where the origins of the normalized coordinate have been centered at the centroid of each normalized region, θ is the orientation of the normalized image, and $f(x,y)$ is an image function — such as brightness — defined on the normalized coordinate. For the 'image' function, however, brightness may not necessarily be the best choice. A desirable property of the 'image' function here is stability under varying ambient light conditions and the relative orientation of the object surface with respect to the camera and the light source in 3D space. From the shape of the formula of (8) with (9)—(11) it is clear that the rotation angle thus recovered is never affected by scale change of the image function between the model and data views. Therefore, the property we need here from the image function is not a perfect constancy, but merely a constancy within a scale factor under different illumination conditions. This is not a hard requirement in practice because we are now focusing on the properties of planar surfaces. For example, if the sensor channels are narrow band it is known that the outputs are invariant up to a consistent scale factor over the entire surface(see e.g.[14]). By equation(8), we get two different candidate angles (by taking the direction of the eigenvector with the larger eigenvalue). To select the correct one, we can align the given image data with the reconstructed image from the model using the recovered affine parameters based on (7), and pick the one that gives the best match.

2.3 Rotation angle via Hu's moment invariants: using 3rd order moments

If the image does not have enough texture, or if it is a binary image, we can not use weighted moments of the image to compute the rotation angle between the normalized image data and the model. In this case, however, we can use the third order moments of the binary image. The use of higher order moments for invariance to rotation was extensively discussed in pattern recognition(e.g.[6, 19, 1]). As a by-product of the study of invariance, in [6] a method for computing the rotation angle using higher order moments was also presented, which we rewrite here:

$$\begin{aligned} I'_{30} &\equiv (N'_{3,0} - 3N'_{1,2}) - i(3N'_{2,1} - N'_{0,3}) \\ &= e^{i3\theta}[(N_{3,0} - 3N_{1,2}) - i(N_{2,1} - 3N_{0,3})] \\ &\equiv e^{i3\theta} I_{30} \end{aligned} \tag{12}$$

$$\begin{aligned} I'_{21} &\equiv (N'_{3,0} + N'_{1,2}) - i(N'_{2,1} + N'_{0,3}) \\ &= e^{i\theta}[(N_{3,0} + N_{1,2}) - i(N_{2,1} + N_{0,3})] \\ &\equiv e^{i\theta} I_{21} \end{aligned} \tag{13}$$

where N_{pq} and N'_{rs} are respective third order moments of the normalized binary image for model and data given in the following (shown only for the normalized model view) and I_{pq}, I'_{pq} are the complex moment invariants proposed in [6].

$$N_{3,0} = \sum_x \sum_y (x^3) \tag{14}$$

$$N_{2,1} = \sum_x \sum_y (x^2 y) \tag{15}$$

$$N_{1,2} = \sum_x \sum_y (xy^2) \tag{16}$$

$$N_{0,3} = \sum_x \sum_y (y^3) \tag{17}$$

where the sums are taken over the entire region of the binary image in the normalized coordinate in which the coordinate origin has been centered at the centroid of the region.
Thus, we have:

$$\tan 3\theta = \nu_3/\delta_3 \tag{18}$$

$$\tan \theta = \nu_1/\delta_1 \tag{19}$$

where,

$$\begin{aligned} \nu_3 &= (N'_{3,0} - 3N'_{1,2})(3N_{2,1} - N_{0,3}) \\ &\quad -(3N'_{2,1} - N'_{0,3})(N_{3,0} - 3N_{1,2}) \end{aligned} \tag{20}$$

$$\begin{aligned} \delta_3 &= (N'_{3,0} - 3N'_{1,2})(N_{3,0} - 3N_{1,2}) \\ &\quad +(3N'_{2,1} - N'_{0,3})(3N_{2,1} - N_{0,3}) \end{aligned} \tag{21}$$

$$\begin{aligned} \nu_1 &= (N'_{3,0} + N'_{1,2})(N_{2,1} + N_{0,3}) \\ &\quad -(N'_{2,1} + N'_{0,3})(N_{3,0} + N_{1,2}) \end{aligned} \tag{22}$$

$$\begin{aligned} \delta_1 &= (N'_{3,0} + N'_{1,2})(N_{3,0} + N_{1,2}) \\ &\quad +(N'_{2,1} + N'_{0,3})(N_{2,1} + N_{0,3}) \end{aligned} \tag{23}$$

An aspect to which we must pay careful attention in using Hu's moment invariants is that of n-fold rotational symmetry. As argued in [6], moment combinations I_{pq}'s with the factor $e^{iw\theta}$, where w/n is not an integer, are identically zero if the shape has n-fold rotational symmetry, so that we can not use those moments for recovering the rotation angle (see [6] for detail). For example, for the 4-fold symmetry case such as a square, both of the formula given in (18) and (19) are useless, and we need higher than third order moments. This happens if the original surface shape is a rectangle when viewed from particular direction in 3D space (of course including the frontal direction). This is because if some image of the surface can be a rectangle, then, no matter from whatever direction it is viewed, its normalized binary image becomes a square, and hence has 4-fold symmetry. This is the consequence of the normalization process we are using(see [15] for detail). We will see this case in the experiment soon.

2.4 Results in using invariants on natural pictures

We now show experimental results obtained using the proposed algorithm for recovering affine parameters based on affine invariants of natural pictures. All the pictures shown here were taken under natural light conditions. The image regions, which are from planar patches on the object surfaces, were extracted manually with some care, but some perturbations may be introduced by this step of the procedure. Figure 2 shows the results on images of a Cocoa-Box. The upper row of pictures show the two gray level pictures, of the same Cocoa-Box taken from different view points: the left view was used for the model, while the right was used for the data.The left and right figures in the middle row show the respective normalized images up to a rotation. Indeed, we see that the two figures coincide if we rotate the left figure by 180 degrees around its centroid. The left and right figures in the lower row are the respective reconstructed image data from the model view (shown in the upper left) by the recovered affine transformation using, lower left: affine invariant plus second order weighted moments of the gray level, lower right: third order moments of the binary image for computing the rotation angle. If the method works correctly, then those reconstructed images should coincide with the corresponding image portion found in the upper right figure. Indeed, we see that both of the methods worked very well for recovering the transformation parameters.

In Figure 3 the results are shown for pictures of a Baby-Wipe container. The upper row of pictures shows the source gray level pictures of a Baby-Wipe container of which the front part was used for the experiment: the left view was used for the model, the right view was used for the data. The left and right figures in the middle row show the respective normalized images. The lower figure is the reconstructed image data from the model view using the affine transformation recovered by means of affine invariant plus second order weighted moments for computing the rotation angle. We would expect that the reconstructed image coincides well with the image in the upper right. From the figure, we see that this method, i.e., affine invariant plus second order weighted moments worked very well for recovering the parameters. As observed in the figures, the normalized images are almost 4-fold rotationally symmetric, so that — as described previously — we can not use the third order moments of the normalized binary image to recover the rotation angle.

Figure 4 shows the results on some Tea-Box pictures. The upper row shows the pictures of a Tea-Box: the left view was used for the model, while the right view was used for the data. The left and right figures in the middle row are the respective normalized images up to a rotation. The left and right figures in the lower row show the respective reconstructed image data from the model view using the recovered affine transformation based on affine invariant plus second order weighted moments of the gray level (left) and third order moments of the binary image (right) for recovering the rotation angle. From the figure, we see that both of the reconstructed images coincide well with the original data shown in the upper right. Though both the methods worked fairly well, the method using second order weighted moments performed slightly better. Considering that both of the reconstructed images are tilted a little bit in a similar manner, perhaps some errors were introduced in the manual region extraction.

3 A sensitivity analysis in the use of affine plus rotation invariant

In this section we analyze the sensitivity of the proposed algorithm for recovering affine transformations using affine invariant plus second order weighted moments of the image function to perturbations in the image data. Perturbations are caused, for example, by errors in region extractings, by lack of planarity of the object surface, or by occlusions. From (7), we know that the sensitivity of the recovered affine parameters against perturbations solely depends on the stability of A', the matrix normalizing the given binary image, and T, the rotation matrix relating the normalized model and the data views, as we assume that the model, so that A, does not include any perturbations. As described in (2.1), the transformation A' can be computed solely using eigenvalues and eigenvectors of the covariance matrix of the original binary image, i.e., the set of (x, y) coordinates contained in the image region. Therefore, if the given image contains perturbations, these have effects on the matrix A', but only through the covariances. In other words, the errors in A' can be completely described by the perturbations expressed in terms of covariances. On the other hand, the effect of the perturbations on the recovered rotation matrix differs according to which algorithm we take for computing rotation, namely, the weighted moments of the image attributes, or the third order moments of the binary image of the objects. In this section, we only show the case for second order weighted moments of the image attributes. The perturbation analysis of the algorithm based on third order moment may be presented in a subsequent paper.

3.1 Analytical formula for sensitivity

In the following, we derive the sensitivity formulas for the affine parameters to be recovered, given perturbations in the image data with respect to the model. Let the ideal description(without any errors) for the normalization process be presented as:

$$\tilde{A}'\tilde{L}A^{-1} = \tilde{T} \tag{24}$$

and the affine parameters are recovered by(c.f.(7)):

$$\tilde{L} = \tilde{A}'^{-1}\tilde{T}A \tag{25}$$

Throughout the subsequent parts of the paper, we consistently use the notation $[\tilde{\cdot}]$(tilde) for ideal parameter values and one without tilde for actually observed values, unless otherwise stated. Then, the perturbations ΔL happening on L is given as follows:

$$\begin{aligned} -\Delta L &= (A'^{-1}T - \tilde{A}'^{-1}\tilde{T})A \\ &= \{(\tilde{A}' - \Delta A')^{-1}(\tilde{T} - \Delta T) - \tilde{A}'^{-1}\tilde{T}\}A \end{aligned}$$

$$
\begin{aligned}
&= \{\tilde{A'}^{-1}(I-\Delta A'\tilde{A'}^{-1})^{-1}(\tilde{T}-\Delta T)-\tilde{A'}^{-1}\tilde{T}\}A \\
&= [\tilde{A'}^{-1}\{\sum_{k=0}^{\infty}(\Delta A'\tilde{A'}^{-1})^k\}(\tilde{T}-\Delta T)-\tilde{A'}^{-1}\tilde{T}]A \\
&= \tilde{A'}^{-1}(\Delta A'\tilde{A'}^{-1}\tilde{T}-\Delta T)A+O(\Delta^2) \qquad (26)
\end{aligned}
$$

where $-\Delta A'$ and $-\Delta T$ are respective perturbations of A' and T such that $-\Delta A' = A' - \tilde{A'}$, $-\Delta T = T - \tilde{T}$. The minus signs for the perturbations are for consistency with the perturbation of the covariances which will appear soon. Thus, ignoring the higher than first order terms, we now know that our job is to derive formulas for ΔT and $\Delta A'$ in terms of perturbations contained in the image data.

[Perturbations in A']

As observed in (6), A' is a function of eigenvalues λ_r's and eigenvectors Φ_r's of the covariance matrix Σ' such that $A'_{ij}(\lambda,\Phi) = \lambda_i^{-\frac{1}{2}}\Phi_{ji}$ where λ_r is the rth eigenvalue and Φ_{sr} is the sth component of the corresponding rth eigenvector Φ_r. Let $\tilde{A'}_{ij}$ be the ideal value for A'_{ij}, the ij component of the matrix A'. Then, we get a formula for the perturbations $\Delta A'_{ij}$ from the Taylor expansion of A' in terms of λ and Φ as follows:

$$
\begin{aligned}
-\Delta A'_{ij} &= A'_{ij} - \tilde{A'}_{ij} \\
&= A'_{ij}(\tilde{\lambda_i}-\Delta\lambda_i, \tilde{\Phi_{ji}}-\Delta\Phi_{ji}) - A'_{ij}(\tilde{\lambda_i}, \tilde{\Phi_{ji}}) \\
&= -\frac{\partial A'_{ij}}{\partial \lambda_i}\Delta\lambda_i - \frac{\partial A'_{ij}}{\Phi_{ji}}\Delta\Phi_{ji} + O(\Delta^2) \\
&= \lambda^{-\frac{3}{2}}(\frac{1}{2}\Delta\lambda_i\tilde{\Phi_{ji}} - \tilde{\lambda_i}\Delta\Phi_{ji}) + O(\Delta^2) \qquad (27)
\end{aligned}
$$

where perturbations of the eigen properties are defined as $-\Delta\lambda_i = \lambda_i - \tilde{\lambda_i}$,
$-\Delta\Phi_i = \Phi_i - \tilde{\Phi_i}$.

Here, from *perturbation theory*[3], we have:

$$
\begin{aligned}
-\Delta\Phi_k &= -\frac{\tilde{\Phi_l}^T\Delta\Sigma'\tilde{\Phi_k}}{\tilde{\lambda_k}-\tilde{\lambda_l}}\tilde{\Phi_l} \qquad (28) \\
-\Delta\lambda_k &= -\tilde{\Phi_k}^T\Delta\Sigma'\tilde{\Phi_k} \qquad (29)
\end{aligned}
$$

where $(k,l) \in \{(1,2),(2,1)\}$ and $-\Delta\Sigma'$ is the perturbation of the given covariances such that $-\Delta\Sigma' = \Sigma' - \tilde{\Sigma'}$. The minus sign of the perturbation of covariances accounts for the occlusions (being occluded by some other surface) occurring in the given image data. Substituting (28) into (27), we obtain:

$$
\begin{aligned}
-\Delta A'_{11} &= -\tilde{\lambda_1}^{-\frac{1}{2}}\frac{\tilde{\Phi_2}^T\Delta\Sigma'\tilde{\Phi_1}}{\tilde{\lambda_1}-\tilde{\lambda_2}}\tilde{\Phi_{21}} \\
&\quad +\frac{1}{2}\tilde{\lambda_1}^{-\frac{3}{2}}(\tilde{\Phi_1}^T\Delta\Sigma'\tilde{\Phi_1})\tilde{\Phi_{11}} \qquad (30) \\
-\Delta A'_{12} &= -\tilde{\lambda_1}^{-\frac{1}{2}}\frac{\tilde{\Phi_2}^T\Delta\Sigma'\tilde{\Phi_1}}{\tilde{\lambda_1}-\tilde{\lambda_2}}\tilde{\Phi_{22}} \\
&\quad +\frac{1}{2}\tilde{\lambda_1}^{-\frac{3}{2}}(\tilde{\Phi_1}^T\Delta\Sigma'\tilde{\Phi_1})\tilde{\Phi_{21}} \qquad (31) \\
-\Delta A'_{21} &= -\tilde{\lambda_2}^{-\frac{1}{2}}\frac{\tilde{\Phi_1}^T\Delta\Sigma'\tilde{\Phi_2}}{\tilde{\lambda_2}-\tilde{\lambda_1}}\tilde{\Phi_{11}} \\
&\quad +\frac{1}{2}\tilde{\lambda_2}^{-\frac{3}{2}}(\tilde{\Phi_2}^T\Delta\Sigma'\tilde{\Phi_2})\tilde{\Phi_{12}} \qquad (32) \\
-\Delta A'_{22} &= -\tilde{\lambda_2}^{-\frac{1}{2}}\frac{\tilde{\Phi_1}^T\Delta\Sigma'\tilde{\Phi_2}}{\tilde{\lambda_2}-\tilde{\lambda_1}}\tilde{\Phi_{12}} \\
&\quad +\frac{1}{2}\tilde{\lambda_2}^{-\frac{3}{2}}(\tilde{\Phi_2}^T\Delta\Sigma'\tilde{\Phi_2})\tilde{\Phi_{22}} \qquad (33)
\end{aligned}
$$

The equations (30)—(33) give the first order approximation of the perturbation $\Delta A'_{ij}$ for A'_{ij}, that is a linear combination of the perturbation $\Delta\Sigma'_{pq}$ such that $\Delta A'_{ij} = \sum_{pq}\tau^{ij}_{pq}\Delta\Sigma'_{pq}$ where τ_{pq} are coefficients that are composed of the eigen properties of the covariances matrix of the *ideal* image data, that are uniquely determined by (30)—(33) and are independent of the perturbations.

[Perturbations in T: the rotation matrix]

In deriving an analytical formula for the perturbation ΔT, we rely on the formula given in (8)—(11), relating the rotation angle to the second order weighted moments of the image (as we have fixed the orientation of the model, orientation of the given image can be seen to be equivalent to the rotation angle). Further, we have the following relation between the weighted moments of the original and the normalized images.

$$
\begin{aligned}
[\tilde{M'}] &= \tilde{A'}[\tilde{m'}]\tilde{A'}^T \qquad (34) \\
[M'] &= A'[m']A'^T \qquad (35) \\
&= (\tilde{A'}-\Delta A')([\tilde{m'}]-[\Delta m'])(\tilde{A'}-\Delta A')^T \qquad (36)
\end{aligned}
$$

where $[m'], [M']$ are respective symmetric matrices of original and transformed weighted moments defined in the following and the term $[\Delta m']$ is the matrix of the perturbation contained in the original image data in terms of the weighted moments:

$$
\begin{aligned}
[m'] &= \begin{pmatrix} m'_{20} & m'_{11} \\ m'_{11} & m'_{02} \end{pmatrix} \qquad (37) \\
[M'] &= \begin{pmatrix} M'_{20} & M'_{11} \\ M'_{11} & M'_{02} \end{pmatrix} \qquad (38) \\
[-\Delta m'] &= [m'] - [\tilde{m'}] \qquad (39)
\end{aligned}
$$

Let $\theta, \tilde{\theta}$ be the recovered and ideal rotation angle and $-\Delta\theta$ be the corresponding perturbation, where we assume that $\Delta\theta$ is small enough such that:

$$
\begin{aligned}
-\Delta T &= \begin{pmatrix} \cos(\tilde{\theta}-\Delta\theta) & -\sin(\tilde{\theta}-\Delta\theta) \\ \sin(\tilde{\theta}-\Delta\theta) & \cos(\tilde{\theta}-\Delta\theta) \end{pmatrix} \\
&\quad -\begin{pmatrix} \cos(\tilde{\theta}) & -\sin(\tilde{\theta}) \\ \sin(\tilde{\theta}) & \cos(\tilde{\theta}) \end{pmatrix} \\
&= \begin{pmatrix} \Delta\theta\sin(\tilde{\theta}) & \Delta\theta\cos(\tilde{\theta}) \\ -\Delta\theta\cos(\tilde{\theta}) & \Delta\theta\sin(\tilde{\theta}) \end{pmatrix} + O(\Delta^2) \qquad (40)
\end{aligned}
$$

In the following, we derive a formula for perturbation $\Delta\theta$ in using second weighted moments of the image. As we assumed that $-\Delta\theta = \theta - \tilde{\theta}$ is small enough, we can approximate it as:

$$
-\Delta\theta = \frac{1}{2}(2\theta - 2\tilde{\theta})
$$

$$\approx \frac{1}{2}\tan(2\theta - 2\tilde{\theta})$$
$$= \frac{1}{2}\cdot\frac{\tan(2\theta)-\tan(2\tilde{\theta})}{1+\tan(2\theta)\tan(2\tilde{\theta})}$$
$$\approx \frac{1}{2}\cdot\frac{\tan(2\theta)-\tan(2\tilde{\theta})}{1+\{\tan(2\tilde{\theta})\}^2} \quad (41)$$

Substituting the relation presented in (8), we get:

$$-\Delta\theta \approx \frac{1}{1+\{\tan(2\tilde{\theta})\}^2} \times\left(\frac{M'_{1,1}}{M'_{2,0}-M'_{0,2}} - \frac{\tilde{M'_{1,1}}}{\tilde{M'_{2,0}}-\tilde{M'_{0,2}}}\right) \quad (42)$$
$$\approx \frac{1}{1+\{\tan(2\tilde{\theta})\}^2}\cdot\frac{1}{(\tilde{M'_{2,0}}-\tilde{M'_{0,2}})^2}\cdot J \quad (43)$$
$$= \frac{1}{(\tilde{M'_{2,0}}-\tilde{M'_{0,2}})^2+(2\tilde{M'_{11}})^2}\cdot J \quad (44)$$

where

$$J \equiv M'_{11}(\tilde{M'_{20}}-\tilde{M'_{02}}) - \tilde{M'_{11}}(M'_{20}-M'_{02}) \quad (45)$$

Substituting (37)—(39) into (45) we get:

$$J = e_{11}\Delta A'_{11} + e_{12}\Delta A'_{12} + e_{21}\Delta A'_{21} + e_{22}\Delta A'_{22} + f_{20}\Delta m'_{20} + f_{11}\Delta m'_{11} + f_{02}\Delta m'_{02} + O(\Delta^2) \quad (46)$$

where e_{ij}'s, f_{pq}'s are respective coefficients of $\Delta A'_{ij}$ and $\Delta m'_{pq}$ that are composed of the components $\tilde{A'_{ij}}$ and $\tilde{m'}_{rs}$ that are independent of the perturbations involved in the given image data.
Then, combining (40), (44), and (46), we get ΔT.

[Perturbation in L]
Finally, combining the formulas for ΔT thus derived with (30)—(33) and substituting it into (26), we obtain the perturbation ΔL:

$$-\Delta L_{ij} \approx \sum_{r,s}(\xi^{ij}_{rs}\Delta\Sigma'_{rs} + \kappa^{ij}_{rs}\Delta m'_{rs}) \quad (47)$$

where ξ, κ are coefficients that are exclusively composed of the components $\tilde{A'_{ij}}$, $\tilde{m'}_{rs}$ and A_{ij}, that are independent of the perturbations, and $(r,s) \in \{(2,0),(1,1),(0,2)\}$. By this linear combination of $\Delta\Sigma'_{rs}$ and $\Delta m'_{rs}$, we have obtained the first order approximation of ΔL_{ij}, the perturbation of L_{ij}, given the perturbations in the original image data in terms of the second order moments of the binary image ($\Delta\Sigma'_{rs}$), and the second order weighted moments of the image attributes ($\Delta m'_{rs}$).

3.2 Experiments

Now we show the sensitivity of the proposed algorithm for recovering affine parameters based on affine invariant plus second order weighted moments to perturbations of the given image region. From (47), we know that perturbation of each recovered component L_{ij} is the linear combination of perturbations of moments of the given image. Here, for simplicity, we try to capture the *overall* trend of the sensitivity of L to perturbations in the given data by examining the following formulas:

$$\delta \equiv \sqrt{\frac{\sum_{i,j}\Delta L^2_{ij}}{\sum_{i,j}\tilde{L_{ij}}^2}} \quad (48)$$

against:

$$\sigma^2 \equiv \sqrt{\frac{\sum_{i,j}\{\bar{f}^2(\Delta\Sigma'_{ij})^2+(\Delta m'_{ij})^2\}}{\sum_{i,j}\{\bar{f}^2(\tilde{\Sigma'_{ij}})^2+(\tilde{m'_{ij}})^2\}}} \quad (49)$$

where $\bar{f}$ is a balancing parameter, and in the following experiments we set it to 255/2. The terms δ, σ^2 express respectively the normalized errors of the recovered affine parameters and the normalized perturbations in terms of the moments of the image. We expect that those two formulas show monotonic relations when perturbations in the moments are small. Of course, we know from the above arguments that there will be some complicated interactions between the two, but we hope some insight may be obtained by observing those two formulas. We use the same picture of a Cocoa-Box used in the earlier experiments. To study the effects of occlusion, perturbations in the image data were produced by dropping particular connected regions from the (almost) perfect image data, as given in Figures 5. The upper pictures show examples of the perturbed image data for which some percentage of the image region was dropped: left 5%, middle 15%, right 25%. The lower pictures show the respective reconstructed image data. Figure 6 shows δ (vertical axis) versus σ^2 (horizontal axis), in which the perturbations were taken from 2.5% to 25% by 2.5% step. From the figure, we see that δ, accuracy in recovering affine parameters, is almost proportional to σ^2, the perturbations, when it is small, but the slope increases a lot as σ^2 increases.

4 Using differential properties of the image: without invariants

In this section, we derive another constraint equation on affine parameters based on the differential properties of the image, and combine it with the canonical geometrical constraint given in (1) to recover the affine transformation. We rewrite here the geometric constraint on the motion of planar surfaces for convenience:

$$X' = LX \quad (50)$$

where the translational component has been eliminated (based on the invariance of the region centroids). Deriving the covariance matrices on both sides of (50) gives:

$$\Sigma_{X'} = L\Sigma_X L^T. \quad (51)$$

where indices of the covariances $\Sigma_{X'}, \Sigma_X$ show the corresponding distributions. Due to the symmetry of covariance matrix, we have only three independent equations in (51) for four unknowns that are the components of L. Therefore, we apparently need another constraints to solve for L. (comments: The constraint of ratio of the image area $det[L] = \text{AREA}(X')/\text{AREA}(X)$ is redundant here when one employs (51).) From this point

of view, what we have done in the preceding sections can be seen as imposing constraints of the rotations between the normalized (up to rotations) images either in terms of weighted moments of some image attribute or using third order moments of the binary image. Here, we will seek another constraint which does not use invariants, based on the differential property of the image which is related to the underlying geometry of the image.

4.1 Deriving another constraint based on differential properties of the image

To derive another constraint on affine parameters, suppose that we have an image attribute $E(X)$ — some scalar function of position X in the image field — that is related to $E'(X')$ of the corresponding point X' in another view by:

$$E(X) = \frac{1}{\rho}E'(X') \tag{52}$$

where X and X' are related by (50) and ρ is a scalar constant. This represents a constraint that the changes of the function E between the different views are only within a scale factor that is consistent over the specified region. Again, we can claim, as in the previous discussion of 2.2, that this constraint is a fairly reasonable one. Taking the gradient of both sides of (52),

$$(E_x, E_y)^T = \frac{1}{\rho}J^T(E'_x, E'_y)^T \tag{53}$$

where E_s's denote partial derivatives of E in terms of the variable s, and J is the Jacobian of X' in terms of X such that

$$J = \begin{pmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{pmatrix} \tag{54}$$

$$= L \tag{55}$$

we get a similar constraint to that on the geometry given in (50), in the differential image, that includes the same affine parameters L:

$$U = \frac{1}{\rho}L^T U' \tag{56}$$

where $U = (E_x, E_y)^T$ and $U' = (E'_x, E'_y)^T$. Taking the covariances brings another constraint on affine parameters in terms of the second order statistics of the differential image as follows:

$$\Sigma_U = \frac{1}{\rho^2}L^T \Sigma_{U'} L \tag{57}$$

Thus, we have obtained two constraint equations in (51),(57) on affine parameters which are composed of up to second order statistics of the geometry and the differential properties of the image.

4.2 Solving for the matrix L

We show how we can solve for the affine transformation L, combining the constraints of the geometry and the differential properties of the image. We anticipate that in practice, due to the limited dynamic range of the sensor device as well as its spatial resolution, the geometrical constraint would probably be more reliable than the differential constraints. Therefore, we incorporate all the three geometrical equations given in (51) with one of the three differential constraints given in (57) to get a solution for L. But, for the purpose of stability, we will try all the possible combinations of the set of the three from (51) with every one of (57), and choose the best-fit match in terms of the alignment of the model with the image data, just as in the case of using the affine invariant.

Combining (51) and (57) we immediately get:

$$\rho = \sqrt{\frac{det[\Sigma_{X'}]det[\Sigma_{U'}]}{det[\Sigma_X]det[\Sigma_U]}} \tag{58}$$

Since covariance matrices are positive definite and symmetric, it is not hard to see from equation (51) that L can be written as:

$$L = \Sigma_{X'}^{\frac{1}{2}} Q \Sigma_X^{\frac{1}{2}} \tag{59}$$

where $\Sigma_{X'}^{\frac{1}{2}}, \Sigma_{X'}^{\frac{1}{2}}$ are respective positive definite symmetric square root matrices of $\Sigma_{X'}, \Sigma_X$, that are unique[8], and Q is an orthogonal matrix, accounting for the remaining one degree of freedom. Considering the fact that $0 < det[L] = det[\Sigma_{X'}^{\frac{1}{2}}]det[Q]det[\Sigma_X^{\frac{1}{2}}]$ we know that Q must be a rotation matrix, so that Q may be written as:

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{60}$$

thus we have:

$$L = \begin{pmatrix} e_{11}\cos\theta + f_{11}\sin\theta & e_{12}\cos\theta + f_{12}\sin\theta \\ e_{21}\cos\theta + f_{21}\sin\theta & e_{22}\cos\theta + f_{22}\sin\theta \end{pmatrix} \tag{61}$$

where the coefficients e_{ij}, f_{ij} are composed of the elements of $\Sigma_{X'}^{\frac{1}{2}}$ and $\Sigma_X^{\frac{1}{2}}$ and those are uniquely determined by (59). Substituting (61) into each of the two equations (we have already used one for solving for ρ) in (57) yields:

$$k_{ij}(\cos\theta)^2 + 2l_{ij}(\cos\theta)(\sin\theta) + m_{ij}(\sin\theta)^2 = \rho^2 p_{ij} \tag{62}$$

where k_{ij}, l_{ij}, m_{ij} are the respective coefficient of $(\cos\theta)^2$, $(\cos\theta)(\sin\theta)$, and $(cos\theta)^2$ in the ij components of the resulting matrices in the left hand side, that are composed of coefficient e_{pq}'s,f_{rs}'s, and elements of $\Sigma_{U'}$, and p_{ij} in the right hand side. Solving for equation (62) we get:

$$\cos\theta = \pm\sqrt{\frac{\nu_c}{\delta_c}} \tag{63}$$

$$\sin\theta = \pm\sqrt{\frac{\nu_s}{\delta_s}} \tag{64}$$

where,

$$\nu_c = 2l^2 + (m - \rho^2 p)(m - k) \pm\sqrt{4l^2(l^2 - (\rho^2 p - m)(\rho^2 p - k))} \tag{65}$$

$$\delta_c = (m - k)^2 + 4l^2 \tag{66}$$

$$\nu_s = 2l^2 + (k - \rho^2 p)(k - m) \pm\sqrt{4l^2(l^2 - (\rho^2 p - k)(\rho^2 p - m))} \tag{67}$$

$$\delta_s = (k - m)^2 + 4l^2 \tag{68}$$

where indices have been suppressed for simplicity. By substituting this back into (61), we finally obtain the four possible candidate of L. To select the best one out of this candidate set, we will try out all the candidates using the alignment approach and pick the one that fits best.

The advantage of using gradient distributions of the image functions, compared with using only geometrical properties, is that their covariances may not be as strongly disturbed by local missing regions or occlusions. Actually, we show below a demonstration of this using experiments on natural images. In this section we described a method that combines differential and geometrical properties of the image, but we might be able to derive a different method for recovering the affine parameters if we had more than one reliable image attributes. By combining those two image constraints, instead of incorporating geometry, we may be able to evelop a method that would be less affected by missing regions.

Since the major disadvantages of the use of global features such as moments is the apparent senisitivity to local disturbances, this approach — that is, the use of differential properties — could be a key issue for improving the stability of the algorithms. In the Appendix we also show — at least mathematically — that even a single point correspondence between the model and data 2D views suffice to recover affine parameters, if some invariant image function is available under the change of orientation of the surface.

[Summary]
In this section so far, we have mathematically derived a constraint equation on affine parameters based on the differential properties of the image in terms of its second order statistics. Then, combining this constraint with the canonical geometric constraint — again in terms of second order statistics — we shown how we can solve for the affine parameters by a direct computation.

4.3 Results using differential properties on natural pictures

Results using the algorithm via combination of the geometrical and differential properties of the image are shown on the same natural pictures used in the earlier experiments for the method based on affine invariants. We used the gradient of the gray level (brightness) image function for the differential data. Note that even though the picture given in the following shows only the data for the manually extracted region used for recognition, we actually use the original image when calculating the gradient at each point. As a result, the artificially introduced edges of the extracted region do *not* have any effect on the derivation of the gradient distribution. Note that this is very important in demonstrating the effectiveness of our method, because otherwise larger contributions on the covariances of gradient distributions would be made by the artificially constructed edges.

Figure 7 shows the results on the Cocoa-Box pictures. The left and right figures in the upper row show the respective gradient distribution — the horizontal axis is f_x and the vertical axis is f_y — for the model and the data views. The lower figure shows the reconstructed image data from the model view by the affine transformation that was recovered. We expect this figure to coincide with the corresponding portion of the upper right picture in Figure 2. From the figure, we see that the algorithm performed almost perfectly.

In Figure 8 the results on the Baby-Wipe container pictures are given. The left and right figures in the upper row show the respective gradient distribution for the model and the data view. The lower figure is the reconstructed image data. We expect this to coincide with the corresponding portion of the upper right picture of the Figure 3. The accuracy is again fairly good, although not as good as that obtained by affine invariant plus second order weighted moments. Likewise, Figure 9 shows the results on the Tea-Box pictures. The result is almost as good as that obtained using affine invariant.

In Figure 10, we show the reconstructed image data given the perturbation in the original image. We used the same data as that used in the sensitivity tests for the affine invariant method. The figures show the respective results for the fraction of missing region 5%(left), 15%, 25%. In Figure 11, the values of δ (vertical axis), accuracy in recovering affine parameters, are plotted against the percentage of the missing region (horizontal axis) in the given image data. We compared this results with the one obtained by the affine invariant method presented previously. Apparently, the results by differential method (plotted as blocks) are less sensitive to perturbations than those by obtained by the affine invariant method (plotted as stars). Probably, this is due to the use of differential distribution as described previously.

5 Conclusion

In this paper, we proposed new algorithms for 3D object recognition that provide closed-form solutions for recovering the transformations relating the model to the image. We proposed two different algorithms: The first one is based on the affine plus rotation invariants using no higher than second or third order moments of the image. Some results on natural pictures demonstrated the effectiveness of the proposed algorithm. An error analysis was also given to study the sensitivity of the algorithm to perturbations. The second algorithm used differential properties of the image attribute. Results demonstrated that the use of differential properties of image attributes allows a recovery of the parameters that is insensitive to missing regions in the given image. This suggested a new direction of object recognition in the sense that it may provide a robust technique using global features for recovering transformations relating the model to the image. Differential properties have been extensively used in motion analysis(e.g.,[5]), but limited to infinitesimal motions of the object. In contrast to the case of motion analysis, our case is not limited to infinitesimal motion. The new method can deal with any motion of the planar surface, as long as the change of the image attribute is constrained within a scale factor at each position on the object. Though all the demonstrations were only on planar patches, as we described, it can connect with the full

3D model of the object to recover the full 3D information via direct computation.

Acknowledgments

Kenji Nagao is grateful to Eric Grimson for the discussions on his research as well as for his various supports in doing this research. He also thanks Greg Klanderman, Aparna Lakshmi Ratan, and Tao Alter for the discussions on his research.

References

[1] Y. S. Abu-Mostafa, D. Psaltis, "Recognitive Aspects of Moment Invariants", *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. PAMI-6, No.6, November 1984.

[2] D. Cyganski, J. A. Orr, "Applications of Tensor Theory to Object Recognition and Orientation Determination", *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. PAMI-7, No.6, November 1985.

[3] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press 1972.

[4] W. E. L. Grimson, *Object Recognition by Computer*, MIT Press, 1991.

[5] B. K. P. Horn, *Robot Vision*, Cambridge MA:MIT Press, 1986.

[6] M. K. Hu, "Visual Pattern Recognition by Moments Invariants", *IRE Trans. on Infor. Theory*, pp. 179–187, 1962.

[7] D. P. Huttenlocher, S. Ullman, "Recognizing Solid Objects by Alignment with an Image", *Int. J. Comp. Vision*, 5:2, pp.195–212, 1990.

[8] M. Iri, T. Kan, *Linear Algebra*, Kyouiku-Syuppan, pp.120–147, 1985 (in Japanese).

[9] M. Iri, T. Kan, *Introduction to Tensor Analysis*, Kyouiku-Syuppan, pp.207–261, 1973 (in Japanese).

[10] J. J. Koenderink, A. J. Van Doorn, "Affine structure form motion", *J. Opt. Soc. Am.*, 8:377–385, 1991

[11] Y. Lamdan, J. T. Schwartz, H. J. Wolfson, "Affine Invariant Model Based Object Recognition", *IEEE Trans. Robotics and Automation* vol. 6, pp. 238–249, 1988.

[12] D. Lovelock, H. Rund, *Tensors, Differential Forms, and Variational Principles*, Dover, pp.1–53, 1975.

[13] K. Nagao, M. Sohma, K. Kawakami and S. Ando, "Detecting Contours in Image Sequences", Transactions of the Institute of Electronics, Information and Communication Engineers in Japan on Information and Systems, vol. E76-D, No.10, pp. 1162–1173, 1993 (in English)

[14] K. Nagao, W. E. L. Grimson, "Recognizing 3D Objects using Photometric Invariant", A.I. Memo 1523, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, February 1995.

[15] K. Nagao, W. E. L. Grimson, "Object Recognition by Alignment using Invariant Projections of Planar Surfaces", A.I. Memo No. 1463, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, February 1994.

[16] K. Nagao, W. E. L. Grimson, "Object Recognition by Alignment using Invariant Projections of Planar Surfaces", In Proc. 12th ICPR, 1994.

[17] K. Nagao, W. E. L. Grimson, "Object Recognition by Alignment using Invariant Projections of Planar Surfaces of 3D Objects", In Proc. DARPA Image Understanding Workshop, November, 1994.

[18] W. H. Press et. al., *Numerical Recipes in C*, Cambridge University Press, pp.610–614, 1985.

[19] S. S. Reddi, "Radial and Angular Moment Invariants for Image Identification", *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. PAMI-3, No.2, March 1981.

[20] A. Shashua, "Correspondence and Affine Shape from two Orthographic Views: Motion and Recognition", A.I. Memo No. 1327, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, December 1991.

[21] M. J. Swain, Color Indexing, PhD Thesis, Chapter 3, University of Rochester Technical Report No. 360, November 1990.

[22] T. F. Syeda-Mahmood, "Data and Model-driven Selection using Color Regions", *In Proc. ECCV Conf.*, pp.321–327, 1992.

[23] W. B. Thompson, K. M. Mutch and V. A. Berzins, "Dynamic Occlusion Analysis in Optical Flow Fields", *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-7, pp.374–383, 1985.

[24] S. Ullman and R. Basri, "Recognition by Linear Combinations of Models", *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-13, pp.992–1006, 1991.

Appendix: Recovering affine parameters via single point correspondence

In this appendix we give theoretical arguments showing that even a single point correspondence between two different views suffices to recover the affine parameters by using differential properties of the image. To do this we assume that we have a nice image attribute (function) I which has the perfect invariant property between different views such that: $I(X) = I'(X')$ and $I \in C^2$ where $X' = LX$.
(Comments: This complete invariance assumption may seem to be unrealistic in practice. But, again, as argued in [14] when the ambient light is not changed it is known that the ratios of the sensor outputs of different channels are invariant if the sensors are narrow band.)

Taking the gradient of I we have:

$$I_x = L_{11} I'_{x'} + L_{21} I'_{y'} I_y = L_{12} I'_{x'} + L_{22} I'_{y'} \qquad (69)$$

Deriving the second order derivatives, we have:

$$I_{xx} \quad = \quad L_{11}^2 I'_{x'x'} + 2 L_{11} L_{21} I'_{x'y'} + L_{21}^2 I'_{y'y'} \qquad (70)$$

$$
\begin{aligned}
I_{xy} &= L_{11}L_{12}I'_{x'x'} + (L_{11}L_{22} + L_{12}L_{21})I'_{x'y'} \\
&\quad + L_{21}L_{22}I'_{y'y'} \qquad (71) \\
I_{yy} &= L_{12}^2 I'_{x'x'} + 2L_{12}L_{22}I'_{x'y'} + L_{22}^2 I'_{y'y'} \qquad (72)
\end{aligned}
$$

From (69) we get $L_{21} = (I_x - L_{11}I'_{x'})/I_y$, and substituting this to (70) and rearranging we have

$$
(I'_{x'x'}{I'_{y'}}^2 - 2I'_{x'}I'_{y'}I'_{xy} + I'_{y'y'}{I'_{x'}}^2)L_{11}^2 - 2(I'_{y'}I'_{x'y'} + I'_{x'}I'_{y'y'})I_xL_{11} + I_x^2 I'_{y'y'} - {I'_{y'}}^2 I_{xx} = 0 \quad (73)
$$

Likewise, we have a similar equation for L_{22}(and L_{12}). Then, solving for these quadratic equations we obtain:

$$
\begin{aligned}
L_{11} &= \frac{(-I'_{y'}I'_{x'y'} + I'_{x'}I'_{y'y'})I_x \pm |I'_{y'}|\Omega_1}{I'_{x'x'}{I'_{y'}}^2 - 2I'_{x'}I'_{y'}I'_{xy} + I'_{y'y'}{I'_{x'}}^2} \qquad (74) \\
L_{21} &= \frac{(-I'_{x'}I'_{x'y'} + I'_{y'}I'_{x'x'})I_x \mp (I'_{x'}/I'_{y'})|I'_{y'}|\Omega_1}{I'_{x'x'}{I'_{y'}}^2 - 2I'_{x'}I'_{y'}I'_{xy} + I'_{y'y'}{I'_{x'}}^2} \quad (75) \\
L_{21} &= \frac{(-I'_{y'}I'_{x'y'} + I'_{x'}I'_{y'y'})I_y \mp (I'_{y'}/I'_{x'})|I'_{x'}|\Omega_2}{I'_{y'y'}{I'_{x'}}^2 - 2I'_{y'}I'_{x'}I'_{xy} + I'_{x'x'}{I'_{y'}}^2} \quad (76) \\
L_{11} &= \frac{(-I'_{x'}I'_{x'y'} + I'_{y'}I'_{x'x'})I_y \pm |I'_{x'}|\Omega_2}{I'_{y'y'}{I'_{x'}}^2 - 2I'_{y'}I'_{x'}I'_{xy} + I'_{x'x'}{I'_{y'}}^2} \qquad (77)
\end{aligned}
$$

where

$$
\begin{aligned}
\Omega_1 &= \sqrt{I_x^2({I'_{x'y'}}^2 - I'_{x'x'}I'_{y'y'}) + I_{xx}(I'_{x'x'}{I'_{y'}}^2 - 2I'_{x'}I'_{y'}I'_{x'y'} + I'_{y'y'}{I'_{x'}}^2)} \\
\Omega_2 &= \sqrt{I_y^2({I'_{x'y'}}^2 - I'_{y'y'}I'_{x'x'}) + I_{yy}(I'_{y'y'}{I'_{x'}}^2 - 2I'_{y'}I'_{x'}I'_{x'y'} + I'_{x'x'}{I'_{y'}}^2)}
\end{aligned}
$$

Figure 2: Results by affine invariant method on the Cocoa-Box pictures.
The upper row of pictures shows two gray level pictures of the same Cocoa-Box taken from two different view points: the left view was used for the model, the right view for the data. The left and right figures in the middle row show the corresponding normalized images. Indeed, we see that the two figures in this row coincide if we rotate the left one by 180 degrees around its centroid. The left and right figures in the lower row are the respective reconstructed image data from the model view (shown in the upper left) by the recovered affine transformation using, lower left: affine invariant plus second order weighted moments of the gray level, lower right: third order moments of the binary image for computing the rotation angle. If the method works correctly, then those reconstructed images should coincide with the corresponding image portion found in the upper right figure. Indeed, we see that both of the methods worked very well for recovering the transformation parameters.

Figure 3: Results by affine invariant method on the Baby-Wipe pictures.
The upper row of pictures shows two gray level pictures of a Baby-Wipe container of which the front part was used for the experiment: the left view was used for the model, while the right view was used for the data. The left and right figures in the middle row show the respective normalized images. Indeed, we see that the two figures coincide if we rotate the left figure by about 180 degrees around its centroid. The bottom figure is the reconstructed image data from the model view by the recovered affine transformation using affine invariant plus second order weighted moments for computing the rotation angle. We expect that the reconstructed image coincides well with the image in the upper right.

Figure 4: Results by affine invariant method on the Tea-Box pictures.
The upper row shows the pictures of a Tea-Box: the left view used for the model, while the right voew was used for the data. The left and right figures in the middle row are the respective normalized images up to a rotation. The left and right figures in the lower row show the respective reconstructed image data from the model view using the recovered affine transformation based on affine invariant plus second order weighted moments of the gray level (left) and third order moments of the binary image (right) for recovering the rotation angle. From the figure, we see that both of the reconstructed images coincide well with the original data shown in the upper right. Though both the methods worked fairly well, the method using second order weighted moments performed slightly better. Considering that both of the reconstructed images are tilted a little bit in a similar manner, perhaps some errors were introduced in the manual region extraction.

Figure 5: Sensitivity analysis against perturbations in the given image.
The upper pictures show examples of the perturbed image data for which some percentage of the image region was dropped: left 5%, middle 15%, right 25%. The lower pictures show the respective reconstructed image data. The perturbations in the image data were produced by dropping particular connected regions from the (almost) perfect image data.

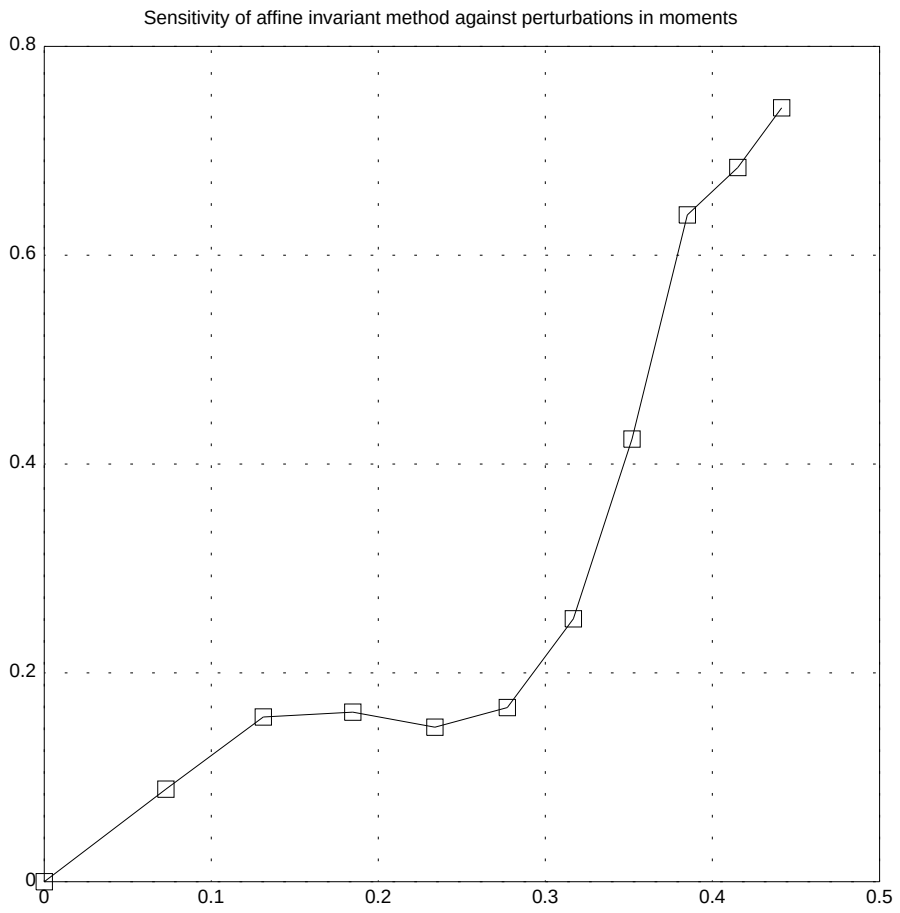

Figure 6: Sensitivity of the recovered parameters by affine plus rotation invariants against perturbations. The horizontal axis is σ^2 while the vertical axis is δ. The values of δ, accuracy in recovering affine parameters, is almost proportional to σ^2, the perturbations, when it is small, but the slope increases rapidly as σ^2 elevates.

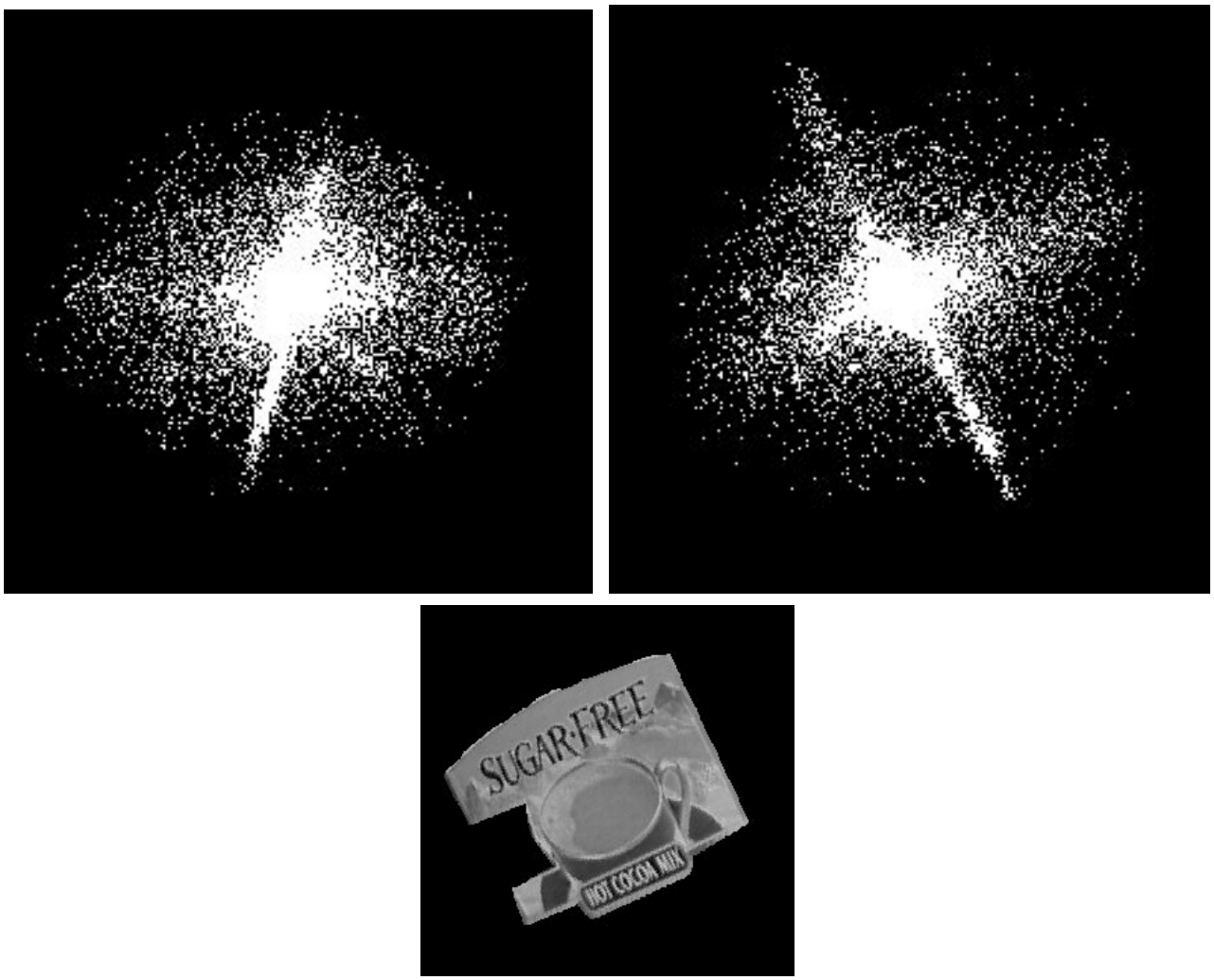

Figure 7: Results by combination of geometric and differential properties on the Cocoa-Box pictures. The left and right figures in the upper row show the respective gradient distribution — the horizontal axis is f_x and the vertical axis is f_y — for the model and the data views. The lower figure shows the reconstructed image data from the model view by the affine transformation that was recovered. We expect this figure to coincide with the corresponding portion of the upper right picture in Figure 2. From the figure, we see that the algorithm performed almost perfectly.

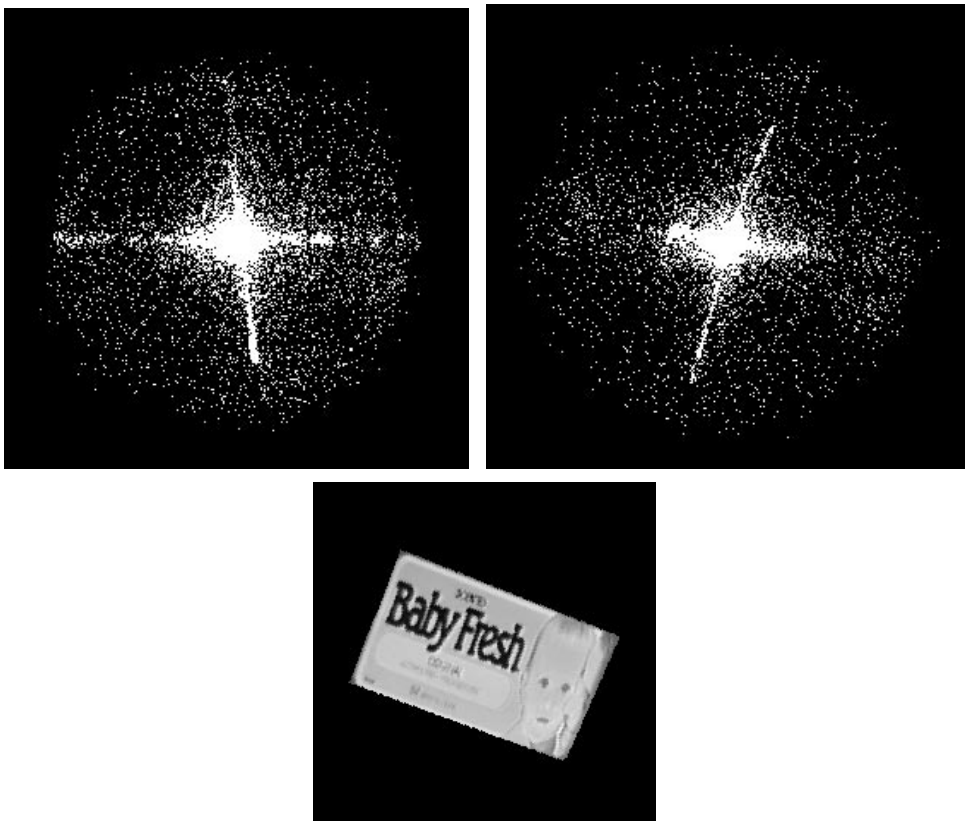

Figure 8: Results by combination of geometric and differential properties on the Baby-Wipe pictures. The left and right figures in the upper row show the respective gradient distribution for the model and the data view. The lower figure is the reconstructed image data, that we expect to coincide with the corresponding portion of the upper right picture of the Figure 3. The accuracy is again fairly good, though not as good as that obtained by affine invariant plus second order weighted moments.

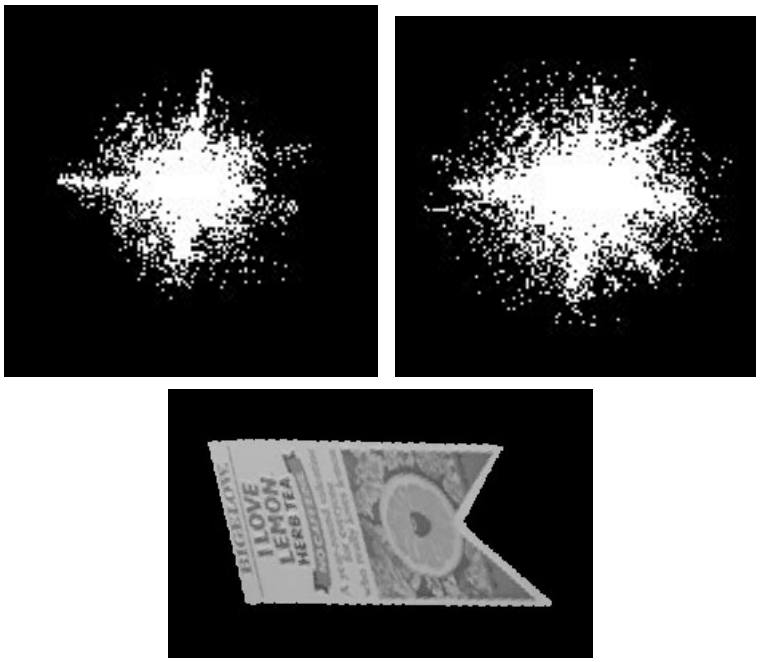

Figure 9: Results by combination of geometric and differential properties on the Tea-Box pictures.
The left and right figures in the upper row show the respective gradient distribution for the model and the data view. The lower figure is the reconstructed image data, that we expect to coincide with corresponding portion of the upper right picture of Figure 4. The result is almost as good as the one by using affine invariant.

Figure 10: Sensitivity of differential method against perturbations.
The figure shows the reconstructed image data for the same perturbed images as those used in the sensitivity tests for affine invariant method. The pictures show respective results for the perturbation percentage in the given image: left 5%, middle 15%, right 25%.

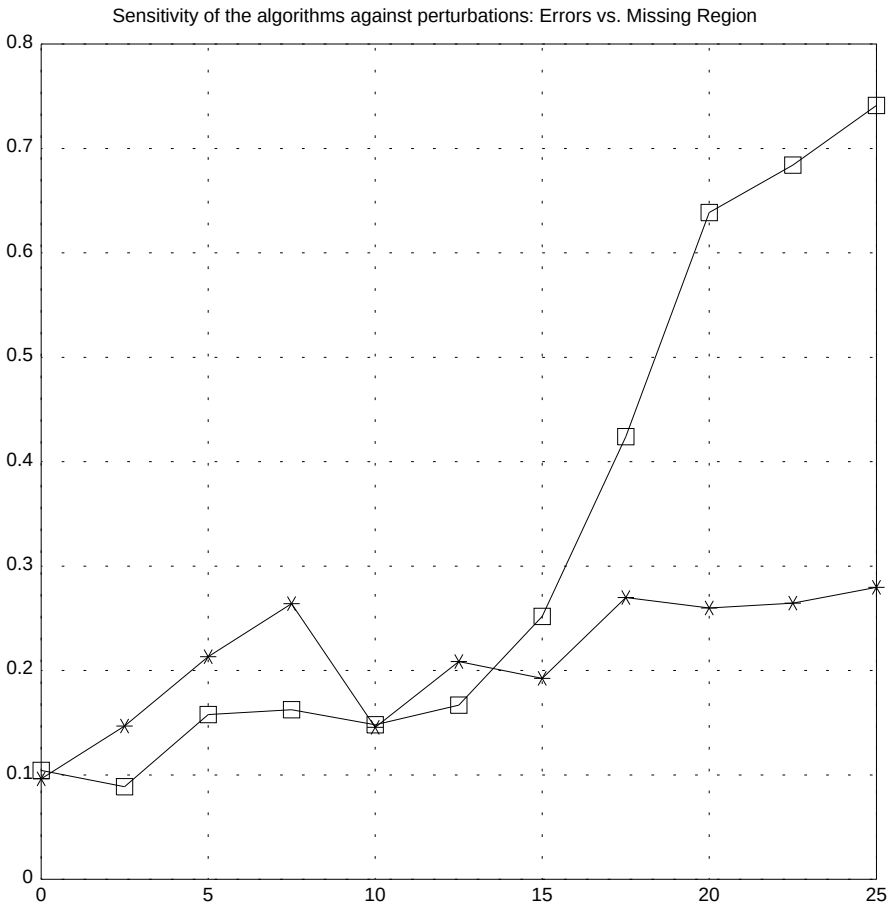

Figure 11: Sensitivity of the recovered parameters by differential method against perturbations. The values of δ (vertical axis), accuracy in recovering affine parameters, are plotted against the percentage of the perturbation in the given image data (horizontal axis). The results by affine invariant are plotted using blocks, while those by differential method are plotted using stars. Apparently, the results by differential method are less sensitive to perturbations than those by affine invariant method.