Pose estimation for objects with planar surfaces using eigenimage and range data analysis

Ovidiu Ghita[1], Paul F. Whelan[1], David Vernon[2] and John Mallon[1]

[1]Vision Systems Laboratory

School of Electronic Engineering

Dublin City University, Dublin 9, Ireland

(ghitao@eeng.dcu.ie)

[2] Etisalat University College

Department of Computer Engineering

Sharjah, United Arab Emirates

Abstract

In this paper we present a novel method for estimating the object pose for 3D objects with well-defined planar surfaces. Specifically, we investigate the feasibility of estimating the object pose using an approach that combines the standard eigenspace analysis technique with range data analysis. In this sense, eigenspace analysis was employed to constrain one object rotation and reject surfaces that are not compatible with a model object. The remaining two object rotations are estimated by computing the normal to the surface from the range data. The proposed pose estimation scheme has been successfully applied to scenes defined by polyhedral objects and experimental results are reported.

Keywords: Image segmentation, eigenimage analysis, range data, surface orientation, 3D pose estimation.

1. Introduction

During the past few years a large number of strategies to determine the spatial orientation for 3D objects have been developed. Earlier approaches attempted to determine the spatial orientation (or attitude) of the object by computing the spatial transformation between the coordinates of a limited number of points on the objects in the scene and their corresponding location in the model object (Aguado et al, 2002, Faugeras and Hebert 1986; Fischler and Bolles 1981; Horaud et al. 1989; Phong et al. 1995). These approaches are in general robust but several limitations are worth mentioning. The first is the difficulty in selecting relevant points on the objects' visible surfaces. In practice these points may not be easily detected due to occlusions and self-occlusions and in this situation the object pose cannot be estimated. In order to alleviate the problems introduced by occlusion, one possible solution is to infer the object pose by analysing the spatial transformation of less ambiguous image primitives such as lines (Ansar and Daniilidis 2003), conics (Forsyth et al. 1991; Tarel and Cooper 2000), curves (Brox et al. 2005, Rosenhahn et al. 2005) and surfaces (Blane et al. 2000; Campbell and Flynn 2001) that can be determined even if they are partially occluded (Bolles and Horaud 1986).

The main problem with the aforementioned techniques is the difficulty to robustly match the set of features extracted from scene data with the set of features associated with a model object in cases where the scene is defined by multiple objects with random orientations.

Visual learning methods based on eigenimage analysis (Black and Jepson 1998; Edwards 1996; Krumm 1996, Liu et al. 2002; Murase and Nayar 1995) have been also proposed to estimate the object pose. As opposed to pose estimation methods based on a structural description of the objects (Arun et al. 1987; Bahnu 1987; Fischler and Bolles 1981; Forsyth et al. 1991; Phong et al. 1995), eigenimage analysis estimates the object's pose by matching its appearance (Black and Jepson 1998; Edwards 1996; Krumm 1996, Mittrapiyanuruk et al. 2004; Turk and Pentland 1991). In this regard, Murase and Nayar (1995) proposed an appearance-based approach to learn and recognise a set of complex objects. Although their method addresses pose estimation along with object recognition, the pose space is constrained by only one *degree of freedom* (DOF) since the image set is generated by rotating the object about a single axis. This problem was specifically addressed in the paper by Edwards (1996) where an active pre-normalisation scheme was applied to reduce the object space from 6 DOF to 3 DOF. His approach can handle the pose estimation for one-object scene with the object's tilt angle limited to 30 degrees with respect to the camera position. Black and Jepson (1998) proposed an alternative solution to match/track gestures of a moving hand. To accommodate the affine transformations between the eigenspace and the input image they employed an optical flow-based technique to estimate the warp transformation. Although interesting, this approach can be applied only when a long sequence of images is available under the assumption that the pixels brightness remain constant and only their location may change from image to image within the sequence. This translates to a requirement to have constant illumination conditions during the image acquisition (and database generation) process and this implementation is more suitable for tracking rather than pose estimation.

In practice, real scenes are defined by multiple objects and the task to infer the pose is significantly more difficult as the scene objects may be partially occluded. Johnson and

Hebert (1999) developed an object recognition scheme that is able to identify multiple 3D objects in scenes affected by clutter and occlusion. In this regard, they applied eigenimage analysis to match surface points using the spin image representation. The main attraction of this approach resides in the use of spin images which are local surface descriptors, hence, they can be easily identified in real scenes that contain clutter. The reported results are impressive but several problems are worth mentioning. The first is the fact that this approach is better suited to objects that have a complex 3D appearance with accurate range data and most importantly, the pose cannot be easily determined as the spin images are local descriptors and are not unique for polyhedral objects unless the pose estimation process is augmented with a geometrically driven model-to-scene verification procedure. Moreover their paper is focused on the object recognition and no results regarding pose estimation are reported.

In this paper we address the problem of full pose estimation for rigid objects with planar faces using a combination of geometrical and visual learning strategies. This task comprises two main components. The first component of the developed system performs region segmentation in order to extract the meaningful surfaces associated with the scene objects while the second component deals with inferring the object pose. The pose estimation scheme removes the main limitation of the standard eigenimage analysis, namely the requirement to sample the object pose in full 6 DOF pose space. Thus, our pose estimation scheme constrains 2 DOF by computing the normal vector for each detected region. As this information is sufficient to determine two rotation angles, the last rotational DOF, namely the angle about the camera axis, is sampled by matching the appearance of the segmented regions with those contained in the model database. This paper is organised as follows. Section 2 presents an overview of the developed algorithm. Section 3 describes the image segmentation algorithm. Section 4 details the pose estimation problem while a number of experimental results are presented and discussed in Section 5. Section 6 includes some concluding remarks.

2. Overview of the proposed approach

The complete approach comprises two distinct components. The first component deals with the database generation and consists of the following off-line operations.

- Principal component analysis (PCA) training with the object surfaces obtained from the segmentation process. The eigenvector representation is generated using a coarsely-sampled set of object poses by varying only one rotational DOF, i.e rotation about *z* axis relative to the camera and the range sensor. For this implementation each object surface has its own eigenvector representation.
- Refinement of the approximate estimation of the rotation about *z* axis by interpolation in the eigenspace.

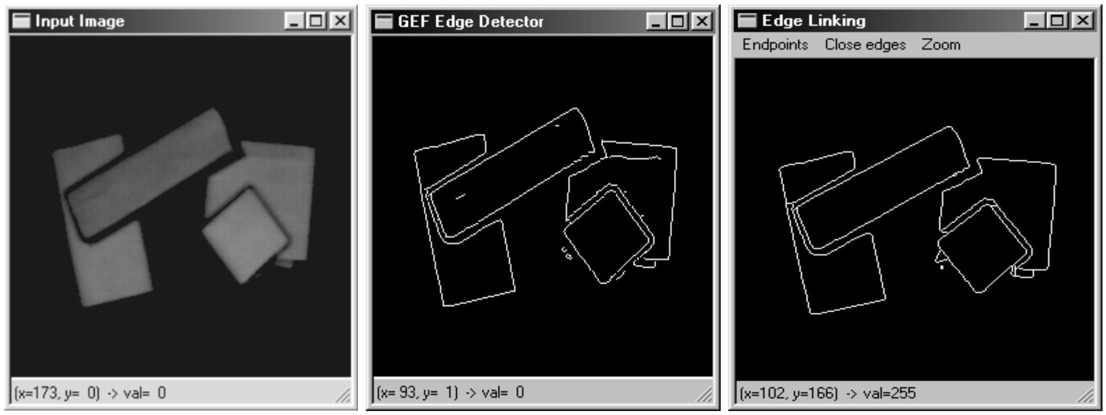
The second component is on-line and deals with the identification and estimation of the pose of the scene objects. The operations required by the second component are summarised below.

- Edge-based segmentation of the input image describing a cluttered scene into distinct planar regions.

- Normal calculation for each segmented region using 3D data and estimation of 2 rotational DOFs (i.e. rotation about *x* and *y* axes).
- Projection of the segmented region on a plane perpendicular on the *z* axis.
- Estimate of the rotation about *z* axis using eigenimage analysis.

3. Image Segmentation

The image segmentation framework employed in this implementation uses edge information to decompose the input image into disjoint regions. When explicit depth information is available, the segmentation process is typically applied to range images since edges are associated with abrupt changes in the depth structure. However this approach is appropriate only when the scene under analysis exhibits significant depth discontinuities and the range sensor has high accuracy. The range sensor employed in this application is based on active depth from defocus (Ghita and Whelan 2001) and offers a 7-bit resolution for a depth range between 0 and 9 cm. Since our implementation deals with a set of small textureless objects, more accurate results were obtained when the segmentation process was applied to intensity images (this approach is also motivated by the fact that for DFD sensors the depth is typically calculated from two defocused intensity images and as a result the range data and intensity data are registered). We also tried to augment the segmentation process with range data but the poor correlation of the edges from the intensity data with the depth discontinuities in the range data motivated us to develop an edge linking strategy in order to improve the edge structure returned from the intensity image.



(a) (b) (c)

Figure 1. Edge linking and noise removal algorithm. (a) The input image. (b) Edge information. (c) Edge linking results (note the removal of the unconnected edge segments).

The quality of the segmentation process is highly dependent on the precision of the edge operator involved. Edge extraction is generally based on analysing the information associated with the first and second derivatives (Marr and Hildreth 1980). However, the recovered edge map either contains false edge points that are generated by image noise or exhibits gaps in edge structure due to a low variation in the pixel intensity distribution. For this application we employed the Gradient Exponential Filter (GEF) edge operator that has been originally developed by Shen and Castan (1992). The performance of this edge operator closely match that offered by the more ubiquitous Canny edge detector (Canny 1986) but it is worth noting that the computational overhead for GEF operator is significantly lower than that associated with the Canny edge operator. In order to refine the initial edge information we applied a method based on thresholding with hysteresis (Ghita and Whelan 2002).

As mentioned earlier the edge map is affected by errors such as false responses that are generated by image noise. But more importantly due to a small change in the image intensity distribution, gaps in the edge structure that may be associated with physically meaningful object features are present. These false edge points and the gaps in the edge structure generate segmentation errors and in order to alleviate these problems we have employed a morphological-based strategy for edge linking (Ghita and Whelan 2002). For this implementation the edge gaps are bridged by analysing the optimal linking path based on minimising a cost function (for more details the reader can refer to Ghita and Whelan (2002)). As we are interested in closed edge structures the unconnected edge structures are removed from the final edge map. Results of the edge linking algorithm are depicted in Figure 1.

4. Pose estimation

Our approach to pose estimation describes the objects in terms of their visible surfaces. In this regard, for each object of interest its appearance is sampled over a range of viewing directions. The resulting images define an image set which encodes the attitude of the object in question. The attitude of an object contained in the scene can be determined by matching an image contained in the image set. To be accurate, this approach requires very large image sets and as a consequence the matching process will be computationally intensive. Fortunately, as the images that form the image set are highly correlated, the computational burden associated with the matching process can be significantly alleviated if an image compression technique is applied.

Principal component analysis (PCA) or eigenimage analysis (Sirovich and Kirby 1987; Moghaddam and Pentland 1997) is a well-known technique for computing a low-dimensional representation (eigenspace) that describes the entire image set. In this formulation, the eigenspace is generated by computing the eigenvectors of the covariance matrix of the image set. Then, by projecting the image set on the eigenspace, the result is a collection of low dimensional vectors which are the compressed representation of the image set (Turk and Pentland 1991).

4.1 PCA Technique. Mathematical background

Let P be the number of images contained in the image set of a given object. To organise the image set as a matrix it is necessary to convert each image into a row vector I_i of size D (image dimension 256×256). To increase the variance between the images contained in the image set, it is necessary to subtract the average image of the image set from each image.

$$I_l^{'}=I_l-I_{av},\ \ l=1\ldots P,\ \ S=[I_1^{'},I_2^{'},\ldots,I_P^{'}]^T \tag{1}$$

where I_{av} is the average image of the image set, S is the image set matrix and T denotes a transpose operation. For each image in the training set the background is discarded and the object surfaces are centered within the image.

The next operation consists of computing the covariance matrix of the image set $C = S^T S$. The eigenvector decomposition of the covariance matrix C results in D orthonormal components that can be determined by solving the eigenvector equation (Press et al. 1992):

$$Cu_i = v_i u_i \tag{2}$$

where u_i is the i^{th} eigenvector and v_i is the corresponding eigenvalue. It is worth noting that the dimension of the covariance matrix C is $D \times D$, a fact that makes the calculation of its eigenvectors impractical. If the number of images P is smaller than D, the reduced covariance matrix $R = SS^T$ can be used instead of the covariance matrix C, but the dimension of the space is limited to P. This dimension can be further decreased since the eigenvectors derived from small eigenvalues have a negligible discriminative power. The eigenvalues are sorted in descending order and the eigenspace dimension can be selected in conjunction with a small threshold value ε as follows:

$$\frac{\sum_{i=1}^{M} v_i}{\sum_{i=1}^{P} v_i} \geq \varepsilon \tag{3}$$

where $M << D$ (for this implementation we set M = 24). The eigenspace is obtained by multiplying the matrix of eigenvectors $U = [u_1, \ldots, u_M]$ with the image set matrix S.

The next operation involves the projection of the image set on the eigenspace and the result is a collection of vectors α_i which defines the compressed version of the images contained in the image set. Since these vectors are *M* dimensional, the amount of compression is *M/D*.

4.2 The sampling problem

The eigenspace representation described in Section 4.1 has several limitations such as sensitivity to image conditions (background noise, image shift and illumination changes) (Fortuna et al. 2002). Since the scene is segmented into disjoint regions, the problems derived from different levels of illumination do not have a significant impact on this implementation. To compensate for the remaining problems, for each image the background is discarded (Murase and Nayar 1995), and the objects are centred within the image.

In line with the image set normalisation procedure described above, the problem of sampling the object's appearance is a critical issue. To sample the full 6 DOF object pose it is necessary to generate an image set that captures all possible orientations of the object under analysis. There is no doubt that such an approach is quite impractical since even at a coarse rate of object pose sampling it would require an extensive number of images. For example to sample the object pose at a rate of 10 samples/DOF requires 10^6 images (Edwards 1996). Consequently, the 6 DOF object pose has to be reformulated in order to reduce the size of the image set. In this sense the translation components constrain 3 DOF and can be easily determined by analysing the coordinates of the centroid of the object's surface. Thus, in this paper we focus on the estimation of the rotation parameters.

In this paper we reformulated the problem of pose estimation as follows: 2 rotational DOF (i.e. rotation about *x* and *y* axes) are determined by statistical calculation of the normal vector to the detected scene regions. Then, the scene regions are projected on a plane perpendicular to the *z* axis, and the last rotational DOF (rotation about *z* axis) is determined using an eigenimage representation. This procedure will be detailed in the following sections.

4.3 Normal vector calculation

The normal vector to a planar surface can be easily computed if we know the coordinates of at least 3 non-collinear 3D points. Unfortunately, computing the normal vector using only a small number of 3D points is not robust as this procedure is extremely sensitive to errors in depth estimation. As we know that the 3D points associated with the segmented region lie on a planar surface, the normal vector can be locally computed using the assumption that the elevation (or the *z* coordinate) is functionally dependent on the *x* and *y* coordinates. Thus, given the set of *n* points $Q = [x, y, z]^T = [x_1 ... x_n, y_1 ... y_n, z_1 ... z_n]^T$ from the range data that belong to the surface in question, the normal vector can be statistically computed by a planar fitting of the 3D points (Lancaster and Salkauskas 1986). As the equation for a planar surface is $z = a_1x+a_2y+a_3$, the best fit can be determined in the least square sense (Nash 1990) by minimizing the errors between the z_i and the plane's values $a_1x_i+a_2y_i+a_3$ as follows:

$$Err(\hat{a}) = \sum_{i=1}^{n} (\hat{a}_1 x_i + \hat{a}_2 y_i + \hat{a}_3 - z_i)^2 \tag{4}$$

where $\hat{a} = [\hat{a}_1, \hat{a}_2, \hat{a}_3]^T$ are the estimated values (since the least square planar fitting minimises the errors in the functional $z = a_1x+a_2y+a_3$ in our analysis we have adopted the homogenous form for the normal vector: $[\hat{a}_1, \hat{a}_2, -1]^T$). Equation 4 generates a simultaneous system where the unknown values are $\hat{a}$ (see Equation 5).

$$\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} y_i^2 & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i & n \end{bmatrix} [\hat{a}] = \begin{bmatrix} \sum_{i=1}^{n} x_i z_i \\ \sum_{i=1}^{n} y_i z_i \\ \sum_{i=1}^{n} z_i \end{bmatrix} \tag{5}$$

The normal vector associated with the surface $Q = [x, y, z]^T$, is represented in homogenous form as $N = [n_x, n_y, n_z, 1]^T = [\hat{a}_1, \hat{a}_2, -1, 1]^T$. Referring to Figure 2, the aim is to transform a plane so that the normal vector lies along the z direction of the reference frame. Within the orthographic projection assumption, the image of the transformed plane can be simply formed by ignoring the z component of the transformed points (see Figure 3). For this image of the transformed plane, the rotation about the z axis is estimated using PCA as will be detailed in the next section.

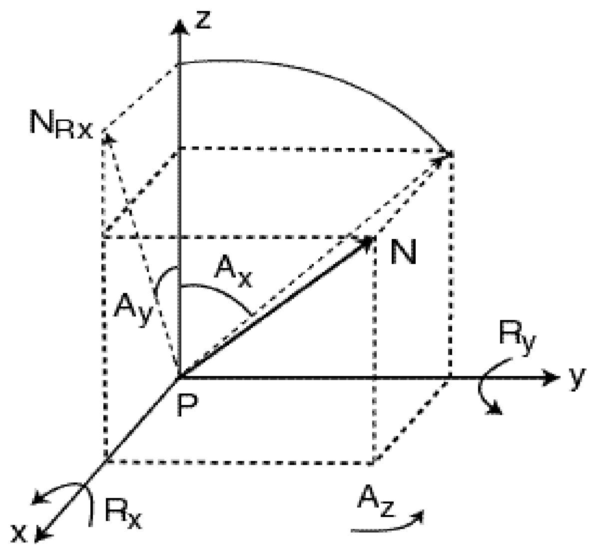
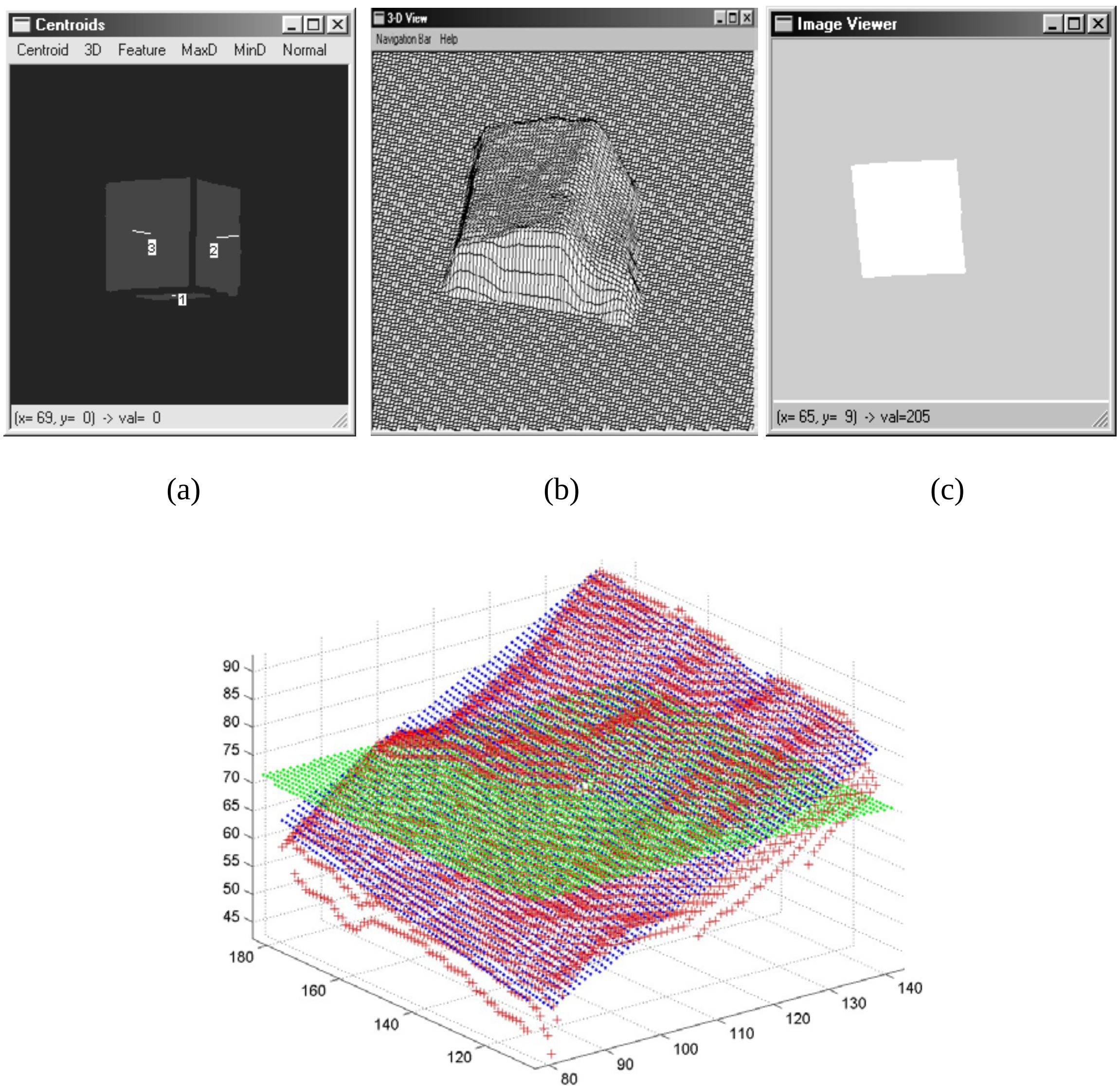


Figure 2. The rotations constrained by the normal vector *N* to the object surface. The angle A_z which describes the rotation about *z* axis is computed from the PCA analysis.

The desired transformation is formulated as $H = T_0^{-1} R_y R_x T_0$, where T_0 is a transformation that centres the points *Q* about the origin, and R_x and R_y are rotations about the *x* and *y* axis respectively, as shown in Figure 2.



(a) (b) (c)

(d)

Figure 3. Orthographic projection. (a) Surface segmentation image of a cubic object (note the orientation of the normal vector for each of the visible surfaces of the cubic object relative to the camera/sensor view). (b) Depth estimation. (c) Output image illustrating the transformed planar of the surface marked with label 3. (d) 3D view illustrating the orthographic projection (red-3D surface data-points, blue-least square planar fitting of the 3D points, green-transformed plane perpendicular to the *z* axis).

T_0 has the form $\begin{bmatrix} I_3 & -M \\ 0_3^T & 1 \end{bmatrix}$, where $M = [m_x, m_y, m_z]^T$, is the mean vector $M = \frac{1}{n}\sum_{i=1}^{n} P_i$ and I_3 the 3×3 identity matrix. Rotations R_x, R_y have the following forms:

$$R_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos A_x & -\sin A_x & 0 \\ 0 & \sin A_x & \cos A_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad R_y = \begin{bmatrix} \cos A_y & 0 & \sin A_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin A_y & 0 & \cos A_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

where $A_x = \tan 2^{-1}(n_y, n_z)$. The rotation angle about y is computed using the transform $N_{Rx} = R_x N = [n_{rx}, n_{ry}, n_{rz}, 1]^T$, as $A_y = -\tan 2^{-1}(n_{rx}, n_{rz})$, where $tan2^{-1}$ is the four quadrant inverse tangent.

4.4 3 DOF object pose estimation

The method described in the previous section constrains two rotational DOF, namely, the rotation about x and y axes and all segmented surfaces are projected on a planar surface perpendicular on the z axis. This allows us to employ eigenimage analysis to constrain the rotation about z axis as the dimensionality of the pose space is reduced to 1 DOF. Therefore, every region is projected on the eigenspace and its projection is compared with those contained in the database.

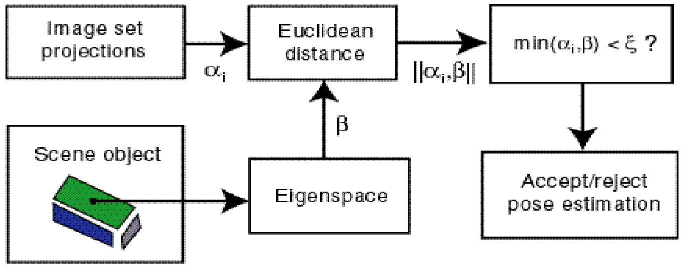


Figure 4. The surface matching process.

The input image approximates an image contained in the database if the minimum distance between its projection on the eigenspace β and the projections derived from the image set α_i is smaller than a threshold value ζ.

$$d_i = \|\beta - \alpha_i\| \leq \varsigma \tag{7}$$

The value of this threshold was set experimentally and defines the maximum allowable distance for a positive estimation stage. The scene surfaces are ranked and the pose is estimated for best positioned surface that is approximated with the smallest error. The surface matching process is illustrated in Figure 4.

5. Experiments and results

The initial tests were conducted on synthetic data defined by a planar surface parallel to the *xy* plane in order to evaluate the correctness of our pose estimation algorithm. Our aim is to identify the errors returned by our algorithm in estimating the rotation angles about *x* and *y* axes of the synthetic data that has been rotated about *x* and *y* axes using the transformation R_xR_y (angles A_x and A_y defined by the user). The experimental data (see Table 1) indicates that our algorithm is able to identify the angles specified in the transformation R_xR_y within the computer error generated by the calculation of the trigonometric functions.

Simulated planar orientation	Rotation *x*: 0.0^0 Rotation *y*: 0.0^0	Estimated planar orientation	Rotation *x*: 0.0^0 Rotation *y*: 0.0^0
Simulated planar orientation	Rotation *x*: 28.5^0 Rotation *y*: -35.0^0	Estimated planar orientation	Rotation *x*: 28.500002^0 Rotation *y*: -35.00004^0
Simulated planar orientation	Rotation *x*: 77.751^0 Rotation *y*: 19.312^0	Estimated planar orientation	Rotation *x*: 77.750999^0 Rotation *y*: 19.311989^0
Simulated planar orientation	Rotation *x*: 1.709^0 Rotation *y*: -5.315^0	Estimated planar orientation	Rotation *x*: 1.708969^0 Rotation *y*: -5.315021^0
Simulated planar orientation	Rotation *x*: -33.49^0 Rotation *y*: -16.27^0	Estimated planar orientation	Rotation *x*: -33.490002^0 Rotation *y*: -16.269997^0
Simulated planar orientation	Rotation *x*: -17.27^0 Rotation *y*: -66.45^0	Estimated planar orientation	Rotation *x*: -17.269991^0 Rotation *y*: -66.449997^0
Simulated planar orientation	Rotation *x*: 12.572^0 Rotation *y*: 45.001^0	Estimated planar orientation	Rotation *x*: 12.572009^0 Rotation *y*: 45.000999^0

Table 1. Estimation of rotation angles about *x* and *y* axes from synthetic data.

To evaluate the performance of the proposed pose estimation scheme when applied to real 3D data obtained from the range sensor, we selected 5 different polyhedral objects that are used to create various scenes.

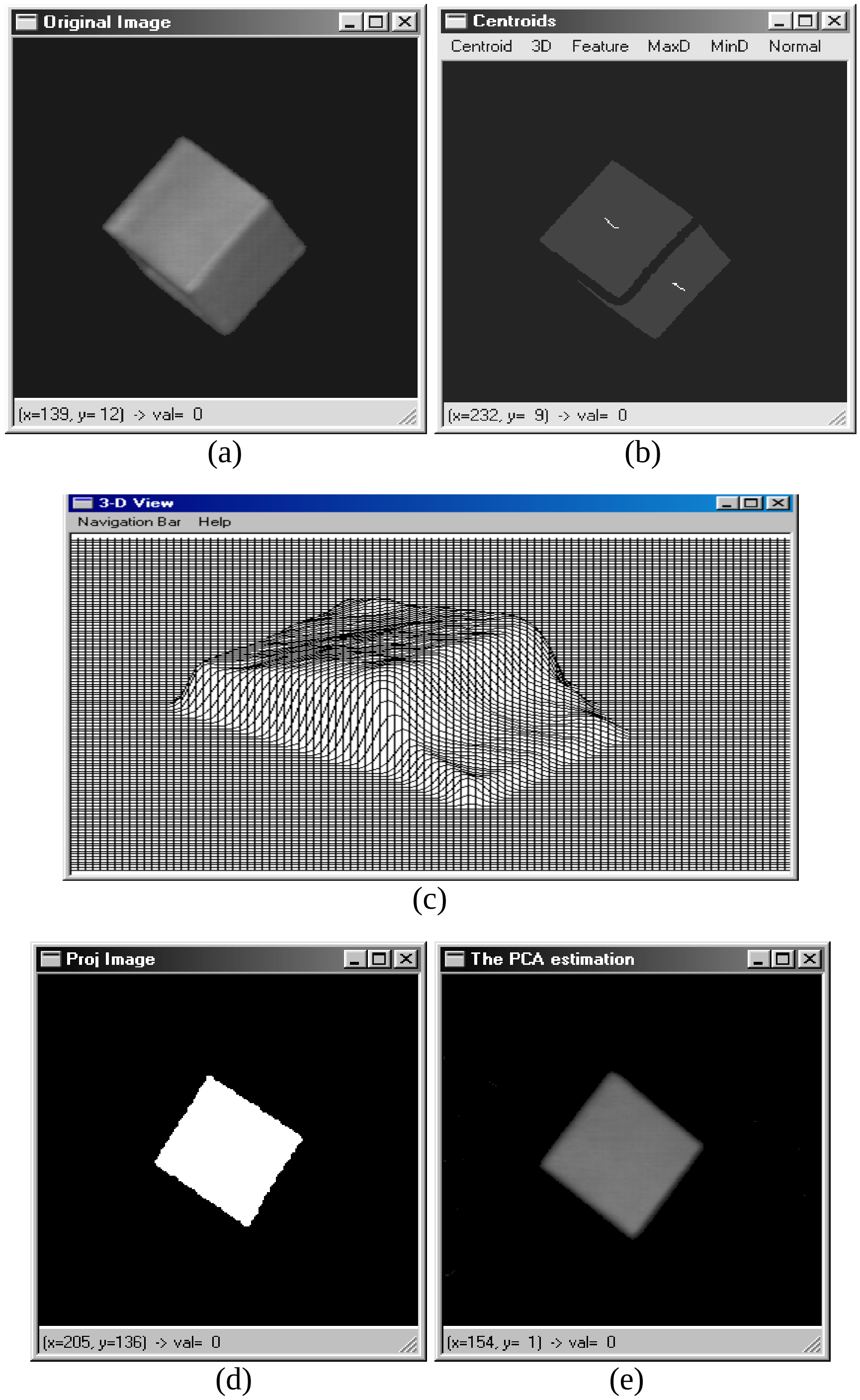


Figure 5. One object scene. (a) Input image. (b) Surface segmentation (normal vectors relative to the camera/sensor position). (c) Depth estimation. (d) Orthographic projection for best estimated surface (A_x = -24.80^0, A_y= -25.08^0). (e) PCA estimation.

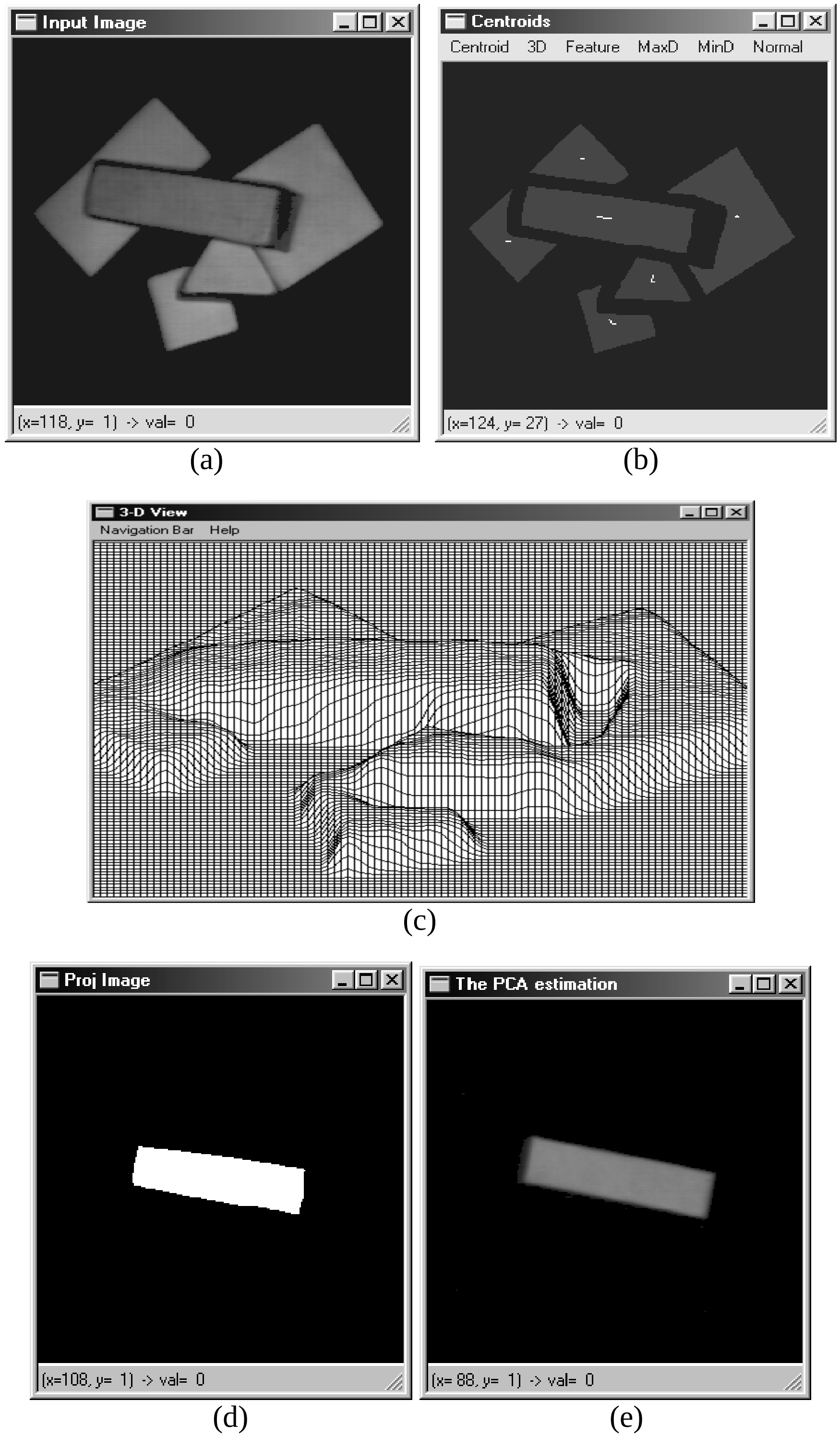


Figure 6. Multiple object scene. (a) Input image. (b) Surface segmentation (normal vectors relative to the camera/sensor position). (c) Depth estimation. (d) Orthographic projection for best estimated surface (A_x = -20.94^0, A_y= 4.21^0). (e) PCA estimation.

While the pose estimation process entails two distinct stages, we analysed the pose estimation error for each stage separately. Initially, errors in two DOF namely the rotation about *x* and *y* axes are evaluated and as expected the pose error is in direct relation to the quality of the depth estimation. In Table 2 we compared the estimation achieved by our algorithm detailed in Section 4.3 with the results obtained when the rotation about *x* axis is estimated by choosing manually relevant non-colinear points from range data (the error in estimation the rotation angle about *y* axis is similar). It can be observed the good correlation between the results returned by our algorithm and the estimation of the rotation angle calculated using the 3D data points selected manually from range data. Our set-up includes a range sensor based on active depth from defocus and its accuracy is 3.4% of the overall ranging distance from the sensor (Ghita and Whelan 2001). The depth error tends to be higher around depth discontinuities and to alleviate this problem the planar surfaces resulting after the application of the segmentation process were approximated by employing a least square planar fitting. This solution also alleviates other depth errors such as those caused by specular characteristics of the object surfaces. In our experiments we have investigated the pose error on scenes defined by a single object (see Figure 5) to estimate the feasibility of the proposed implementation and on scene containing clutter to asses the validity of the proposed pose estimation scheme (see Figure 6).

As discussed in Section 4, the rotation about *z* axis has been analysed by applying eigenimage analysis. In this way, for each object, we acquired 24 images where the rotation angle is sampled uniformly with the object lying flat on a worktable. In our experiments the object's eigenspace is 24 dimensional and the manifold has been re-sampled to 720 points by using linear interpolation. (The PCA space is generated by 24 images that are able to sample uniformly the rotation about *z* axis with a resolution of 15^0. The resolution of the PCA manifold has been increased by calculating new PCA projections using linear interpolation that will generate 30 interpolated projections between any two adjacent projections produced by the 24 images (training set). This would result in a PCA manifold that has 720 projections and is able to sample linearly the rotation about *z* axis with a resolution of 0.5^0. For more details about this procedure refer to Murase and Nayar 1995).

Actual orientation. Rotation only about *x* axis		0^0	5^0	10^0	20^0	30^0	40^0	45^0
Estimated orientation. Our algorithm		-0.54^0	6.67^0	11.83^0	22.05^0	27.68^0	35.07^0	41.81^0
Estimated orientation. Manual selection of 3D points	Set 1	-0.49^0	7.15^0	11.57^0	21.81^0	27.39^0	34.38^0	41.35^0
	Set 2	-0.38^0	7.05^0	11.42^0	22.85^0	27.17^0	34.73^0	43.80^0

Table 2. Estimation of rotation angles about *x* axis using the range data generated by our depth from defocus sensor.

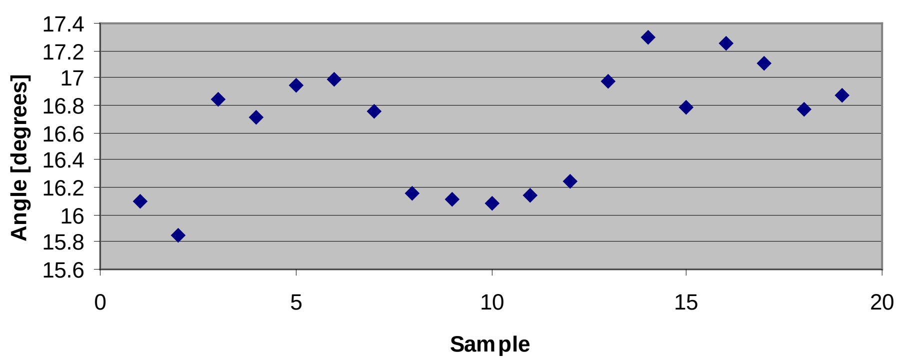


Figure 7. Repeatability test for 20 successive measurements. Object rotated about *x* axis with an angle of 16 degree. Resulting mean value 16.64^0, standard deviation 0.44.

Since the region of interest that is projected on the eigenspace has been aligned to be perpendicular on the *z* axis using the range information, the absolute error is influenced by the errors in the estimation of the plane associated with the region in question. The experiments indicate that the error in estimating the rotation about *z* axis is in direct relation with the object's tilt angles (rotation angles about *x* and *y* axes).

Actual object orientation	Rotation x: 0^0 Rotation y: 0^0 Rotation z: 0^0	Estimated object orientation	Rotation x: 1.63^0 Rotation y: -1.82^0 Rotation z: 0.5^0
Actual object orientation	Rotation x: 0^0 Rotation y: 0^0 Rotation z: 25^0	Estimated object orientation	Rotation x: 1.42^0 Rotation y: -0.98^0 Rotation z: 22.5^0
Actual object orientation	Rotation x: 0^0 Rotation y: 0^0 Rotation z: 45^0	Estimated object orientation	Rotation x: 0.95^0 Rotation y: -0.84^0 Rotation z: 43.5^0
Actual object orientation	Rotation x: 15^0 Rotation y: 0^0 Rotation z: 0^0	Estimated object orientation	Rotation x: 17.58^0 Rotation y: 3.72^0 Rotation z: 2.5^0
Actual object orientation	Rotation x: 15^0 Rotation y: 0^0 Rotation z: 25^0	Estimated object orientation	Rotation x: 17.65^0 Rotation y: 3.32^0 Rotation z: 24.0^0
Actual object orientation	Rotation x: 30^0 Rotation y: 0^0 Rotation z: 0^0	Estimated object orientation	Rotation x: 26.40^0 Rotation y: 2.88^0 Rotation z: 3.5^0
Actual object orientation	Rotation x: 30^0 Rotation y: 0^0 Rotation z: 25^0	Estimated object orientation	Rotation x: 26.75^0 Rotation y: 3.35^0 Rotation z: 21.0^0
Actual object orientation	Rotation x: 45^0 Rotation y: 0^0 Rotation z: 0^0	Estimated object orientation	Rotation x: 42.24^0 Rotation y: 4.28^0 Rotation z: 3.5^0
Actual object orientation	Rotation x: 45^0 Rotation y: 0^0 Rotation z: 25^0	Estimated object orientation	Rotation x: 42.13^0 Rotation y: 4.45^0 Rotation z: 21.5^0
Actual object orientation	Rotation x: 45^0 Rotation y: 0^0 Rotation z: 45^0	Estimated object orientation	Rotation x: 41.35^0 Rotation y: 4.93^0 Rotation z: 39.5^0

Table 3. Pose estimation accuracy. Estimation of the rotation about *z* axis for various object rotations about *x* and *y* axes.

The performance of the developed pose estimation algorithm is illustrated in Table 3 and it can be observed that the errors in the estimation of the rotation angle about z axis increase for large rotations about x and y axes that are generated by the low resolution depth estimation. Figure 7 illustrates the repeatability test for 20 successive measurements when the cubic object illustrated in Figure 5(a) was rotated 16 degrees about x axis.

6. Conclusions

This paper described the development of a two-stage pose estimation algorithm. In the standard form the eigenspace analysis technique has several limitations such as sensitivity to illumination changes, background conditions and partial occlusion. To address these issues we have applied an edge-based segmentation in order to decompose the input image into disjoint regions that describe the scene objects. However, to determine the 6 DOF object pose using the standard eigenspace technique is not a practical approach since a prohibitive number of images are required to sample the object's appearance in all possible orientations. To overcome this issue we employed range data to constrain 2 object rotations while the estimation of the remaining object rotation is determined by applying eigenimage analysis. It is worth noting that this pose estimation scheme has the advantage that no spatial relationships between adjacent scene surfaces are necessary to determine the pose of the scene object. The experimental results indicate that reasonable accurate pose estimation is obtainable from this approach and we believe that this pose estimation technique is particularly useful when dealing with polyhedral objects or objects with well-defined surfaces.

References

[Aguado et al. 2002] A.S. Aguado, E. Montiel and M.S. Nixon "Invariant characterisation of the Hough transform for pose estimation of arbitrary shapes", *Pattern Recognition*, vol. 35, no. 5, pp. 1083-1109, 2002.

[Arun et al. 1987] K.S Arun, T.S. Huang, S.D. Blostein, "Least-square fitting of two 3-D point sets", *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 9, no. 5, pp. 698-700, 1987.

[Ansar and Daniilidis 2003] A. Ansar, K. Daniilidis, "Linear pose estimation from points or lines", *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 25, no. 5, pp. 578-589, 2003.

[Bahnu 1987] B. Bhanu, "Representation and shape matching of 3-D points", *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 6, no. 3, pp. 340-351, 1987.

[Black and Jepson 1998] M.J. Black, A.D. Jepson, "EigenTracking: Robust matching and tracking of articulated objects using a view-based representation", *Int. J. Comp. Vision*, vol. 26, no.1, pp. 63-84, 1998.

[Blane et al. 2000] M. Blane, Z. Lei, H. Civi, D.B. Cooper, "The 3L algorithm for fitting implicit polynomial curves and surfaces to data", *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 22, no.3, pp. 298-313, 2000.

[Bolles and Horaud 1986] R.C. Bolles, P. Horaud, "3DPO: A three dimensional part orientation system", *Int. J. Robotics Res.*, vol. 5, no. 3, pp. 3-26, 1986.

[Brox et al. 2005] T. Brox, B. Rosenhahn, J. Weickert, "Three-dimensional shape knowledge for joint image segmentation and pose estimation", DAGM-Symposium 2005, Vienna, Austria, pp. 109-116, 2005.

[Campbell and Flynn 2001] R. Campbell, P. Flynn, "A survey of free-form object representation and recognition techniques", *Computer Vision and Image Understanding*, vol. 81, no. 2, pp. 166-210, 2001.

[Canny 1986] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 698-700, 1986.

[Edwards 1996] J. Edwards, "An active, appearance-based approach to the pose estimation of complex objects", *Proc. of the IEEE Intelligent Robots and Systems Conference*, Osaka, Japan, pp. 1458-1465, 1996.

[Faugeras and Hebert 1986] O.D. Faugeras, M. Hebert, "The representation, recognition and locating of 3-D objects", *Int. J. Robotics Res.*, vol. 5, no. 3, pp. 27-52, 1986.

[Fischler and Bolles 1981] M.A. Fischler, R.C. Bolles, "Random shape consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Comm. of ACM*, vol. 24, no. 6, pp. 351-395, 1981.

[Forsyth et al. 1991] D. Forsyth, J.L. Mundy, A. Zisserman, C. Coelho, A. Heller, C. Rothwell, "Invariant descriptors for 3-D object recognition and pose", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 10, pp. 971-991, 1991.

[Fortuna et al. 2002] J. Fortuna, D. Schuurman, D. Capson, "A Comparison of PCA and ICA for object recognition under varying illumination", *Int. Conference on Pattern Recognition*, vol. 3, pp. 11-16, Quebec City, Canada, 2002.

[Ghita and Whelan 2001] O. Ghita, P.F. Whelan, “A video-rate range sensor based on depth from defocus”, *Optics & Laser Technology*, vol. 33, no.3, pp 167-176, 2001.

[Ghita and Whelan 2002] O. Ghita, P.F. Whelan, “Computational approach for edge linking”, *Journal of Electronic Imaging*, vol. 11, no. 4, pp 479-485, 2002.

[Horaud et al. 1989] R. Horaud, B. Conio, O. Leboulleux, B. Lacolle, “An analytic solution for the perspective 4-point problem in computer vision”, *Graphics, and Image Processing*, vol. 47, no.1, pp. 33-44, 1989.

[Johnson and Hebert 1999] A. Johnson, M. Hebert, “Using spin images for efficient object recognition in cluttered 3D scenes”, *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 21, no. 5, pp. 433 – 449, 1999.

[Krumm 1996] J. Krumm, “Eigenfeatures for planar pose measurement of partially occluded objects”, *Proc. of the Computer Vision and Pattern Recognition*, San Francisco, USA, pp. 55-60, 1996.

[Lancaster and Salkauskas 1986] P. Lancaster, K. Salkauskas, *Curve and Surface Fitting: An Introduction*, London: Academic Press, 1986.

[Liu et al. 2002] Q. Liu, R. Huang, H. Lu, S. Ma, “Face recognition using kernel based Fisher discriminant analysis”, *IEEE Int. Conference on Automatic Face and Gesture Recognition*, Washington DC, USA, 2002.

[Marr and Hildreth 1980] D. Marr, E. Hildreth, “Theory of edge detection”, *Proc. of Royal Society*, B 207, London, pp. 187-217, 1980.

[Mittrapiyanuruk et al. 2004] P. Mittrapiyanuruk, G.N. DeSouza, A.C. Kak, “Calculating the 3D pose of rigid objects using active appearance models”, *Int. Conference in Robotics and Automation*, New Orleans, USA, 2004.

[Moghaddam and Pentland 1997] B. Moghaddam, A. Pentland, "Probabilistic visual learning for object representation", *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 19, no. 7, pp. 971-991, 1997.

[Murase and Nayar 1995] H. Murase, S.K. Nayar, "Visual learning and recognition of 3-D objects from appearance", *Int. J. Comp. Vision*, vol. 14, pp. 5-24, 1995.

[Nash 1990] J.C. Nash, *Compact numerical methods for computers: Linear algebra and function minimisation*, 2nd edition, Bristol, England, Adam Hilger, 1990.

[Phong et al. 1995] Q.T. Phong, R. Horaud, A. Yassine, D.T. Pham, "Object pose from 2-D to 3-D point and line correspondencies", *Int. J. Comp. Vision*, vol.15, no. 3, pp. 225-243, 1995.

[Press et al 1992] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical recipes in C*, Cambridge University Press, 1992.

[Rosenhahn et al 2005] B. Rosenhahn, C. Perwass, G. Sommer, "Pose estimation of free-form contours", *Int. J. Comp. Vision (IJCV)*, vol. 62, no. 3, pp. 267-289, 2005.

[Shen and Castan 1992] J. Shen, S. Castan, "An optimal operator for step edge detection", *CVGIP: Graphical Models and Image Processing*, vol. 54, no.2, pp. 112-133, 1992.

[Sirovich and Kirby 1987] L. Sirovich, M. Kirby, "A low-dimensional procedure for characterization of human faces", *J. Opt. Soc. Amer*,. vol. 4, no.3, pp. 112-133, 1987.

[Tarel and Cooper 2000] J.P. Tarel, D.B. Cooper, "The complex representation of algebraic curves and its simple exploitation for pose estimation and invariant recognition", *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 22, no. 7, pp. 663-674, 2000.

[Turk and Pentland 1991] M. Turk, A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, vol.3, pp. 71-86, 1991.

Ovidiu Ghita received his B.E. and M.E. degrees in Electrical Engineering from Transilvania University, Brasov, Romania. From 1994 through 1996 he was an Assistant Lecturer in the Department of Electrical Engineering at Transilvania University. Since then, he has been a member of the Vision Systems Group at Dublin City University (DCU), during which time he received his Ph.D. for work in the area of robotic vision. Currently, he holds a position of postdoctoral Research Assistant in the Vision Systems Group at DCU. His current research interests are in the area of range acquisition, machine vision and medical imaging.

Paul F Whelan received his B.Eng.(Hons) degree from the National Institute for Higher Education Dublin, a M.Eng. degree from the University of Limerick, and his Ph.D. (Computer Science) from the University of Wales, Cardiff. During the period 1985-1990 he was employed by Industrial and Scientific Imaging Ltd and later Westinghouse (WESL), where he was involved in the research and development of industrial vision systems. He was appointed to the School of Electronic Engineering, Dublin City University (DCU) in 1990 and is currently a full Professor and holds a Personal Chair. Prof. Whelan set-up the Vision Systems Laboratory and its associated Vision Systems Group in 1990 and currently serves as its director. As well as a wide range of scientific publications, Prof. Whelan co-edited *Selected Papers on Industrial Machine Vision Systems* (1994), and was the co-author of *Intelligent Vision Systems for Industry* (1997) and *Machine Vision Algorithms in Java* (2000). His research interests include applied morphology, texture analysis, machine vision and medical imaging. He is a Senior Member of the IEEE, a Chartered Engineer and a member of the IEE and IAPR. He served on the IEE Irish centre committee (1999-2002) and is member of the governing board of the *International Association for Pattern Recognition (IAPR)* and the President of the *Irish Pattern Recognition and Classification Society.*

David Vernon is a Professor of Computer Engineering at Etisalat University College, UAE, and a visiting researcher at the University of Genoa. He is the coordinator of the European Network for the Advancement of Artificial Cognitive Systems (www.eucognition.org) and a member of the RobotCub project team working on the creation of the iCub, an open-source cognitive humanoid robot (www.robotcub.org). Over the past 27 years, he has held positions at Westinghouse Electric Corp., Trinity College Dublin, the European Commission, the National University of Ireland Maynooth, and Science Foundation Ireland. He has authored two and edited three books on computer vision and has published over eighty papers in the fields of Computer Vision, Robotics, and Cognitive Systems. His research interests include Fourier-based computer vision and enactive approaches to cognition.

John Mallon received the B.Eng. (H1) degree in Mechatronic Engineering in 2001 and PhD degree in 2006 from Dublin City University, Ireland. He is currently a post-doctoral researcher in the Vision Systems Group in DCU. His research interests include multiple view geometry and 3D sensing, optical aberrations and modelling and robotics.