

Improved maximum likelihood estimation of object pose from 3D point clouds using curves as features

Harshana G. Dantanarayana*, Jonathan M. Huntley
Wolfson School of Mechanical, Electrical and Manufacturing Engineering,
Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK

ABSTRACT

Object recognition and pose estimation is a fundamental problem in automated quality control and assembly in the manufacturing industry. Real world objects present in a manufacturing engineering setting tend to contain more smooth surfaces and edges than unique key points, making state-of-the-art algorithms that are mainly based on key-point detection, and key-point description with RANSAC and Hough based correspondence aggregators, unsuitable. An alternative approach using maximum likelihood has recently been proposed in which surface patches are regarded as the features of interest¹. In the current study, the results of extending this algorithm to include curved features are presented. The proposed algorithm that combines both surfaces and curves improved the pose estimation by a factor up to 3 \times , compared to surfaces alone, and reduced the overall misalignment error down to 0.61 mm.

Keywords: pose estimation, maximum likelihood, curve features, edge features, surface features, point clouds

1. INTRODUCTION

Pose estimation of known objects is a standard problem in the manufacturing industry for automated quality control and automated assembly. Even though modern 3D point cloud measurement systems reduce the problem in both scale and affine transformations relative to 2D, the problem is still complex due to occlusions and clutter².

Most state-of-the-art pose estimation techniques start by detecting key points of a 3D model and a scene³, and describing the neighborhoods around those key points using key point descriptors⁴. The descriptors are used to create a correspondences map⁵, with aggregation via a Hough transform⁶ or RANSAC algorithm⁷ to find a rough estimate of the pose. This is then refined with a (time consuming) iterative closest point (ICP) algorithm⁸. The main shortcoming of the above technique in a manufacturing engineering setting is that many real world objects in assembly plants do not contain many unique key points that can be reliably matched, but instead have smoothly varying surfaces and edges.

In a previous paper¹ a probabilistic algorithm based on maximum likelihood (ML) estimation was presented, in which surface features were used to estimate the pose of a modelled object in a cluttered scene as an alternative approach that works well on smoothly varying surfaces. In the current paper this algorithm is extended to include curve features, as well as the original surface features, in both the model and the measured scene. Preliminary results are presented to demonstrate that this provides significant improvements in the accuracy of the estimated pose.

The underlying theory behind the method of maximum likelihood, and its extension to include curve features, are introduced in Section 2. In Section 3, the methods used to create models and match scenes combining both surfaces and edges are described. Section 4 summarises and discusses the improved results from the extended method, with conclusions drawn in the final section.

2. THEORY

In a previous paper by the same authors¹, the localisation problem was defined as how to estimate the pose $\Theta = (x, y, z, \omega, \kappa, \varphi)^T$ of an object, given a set of features $\mathbf{s} = \{\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_N\}$ that is detected in a scene. The likelihood L of the pose estimator, where the object is known to contain a set of pre-determined ‘model’ features $\mathbf{m} = \{\mathbf{m}_0, \mathbf{m}_1, \dots, \mathbf{m}_M\}$, is directly proportional to the probability of observing features \mathbf{s} , given pose Θ of the object. When all features are independent, the probability of observing all features is given by

*H.G.Dantanarayana@lboro.ac.uk; phone +44(0) 1509227565

$$L(\Theta | \mathbf{s}) \propto P(\mathbf{s} | \Theta) = \prod_{i=1}^N f_i(\mathbf{s}_i | \Theta), \quad (1)$$

where $f_i(\mathbf{s}_i | \Theta)$ is the probability density function for the i^{th} scene feature following rotation and translation of the model by Θ . Due to the uncertainty in matching scene features to model features, in general a scene feature \mathbf{s}_i should be compared with all M model features. This is represented mathematically as

$$f_i(\mathbf{s}_i | \Theta) = \sum_{j=1}^M g_{ij}(\mathbf{m}_j | \Theta) + g_0, \quad (2)$$

where g_{ij} is the probability density for the i^{th} scene feature due to the j^{th} model feature following a rotation and translation by Θ . g_0 is the background probability in order to allow for a scene feature not matching any model feature¹.

2.1 Surface and curve feature detection

As in the previous study¹, surfaces are detected by clustering neighboring pixels with a small deviation of the angle of the surface normal as well as the distance. The maximum allowed deviations were set at 0.045 rad and 10 mm for the angle difference of the surface normal and the distance between pixels in 3D. Similarly, the curve features are detected by clustering neighboring pixels with a small deviations of angle of the tangent as well as distance. Note that the surface derivatives required to calculate surface normals and curve tangents, are calculated from the neighbourhoods around the point of interest (x,y,z) in the point cloud.

Given the surface derivatives of the z -coordinate with respect to the x -coordinate $z_x = \partial z / \partial x$ and surface derivative with respect to the y -coordinate $z_y = \partial z / \partial y$, the surface normal is defined as⁹

$$\mathbf{n} = \frac{\langle -z_x, -z_y, 1 \rangle}{| \langle -z_x, -z_y, 1 \rangle |}. \quad (3)$$

Given the second derivatives z_{xx}, z_{xy}, z_{yy} the first fundamental form of the surface is⁹

$$\mathbf{F}_I = \begin{bmatrix} 1 + z_x^2 & z_x \cdot z_y \\ z_x \cdot z_y & 1 + z_y^2 \end{bmatrix} \quad (4)$$

and the second fundamental form of the surface is⁹

$$\mathbf{F}_{II} = \frac{1}{| \langle -z_x, -z_y, 1 \rangle |} \begin{bmatrix} z_{xx} & z_{xy} \\ z_{xy} & z_{yy} \end{bmatrix} \quad (5)$$

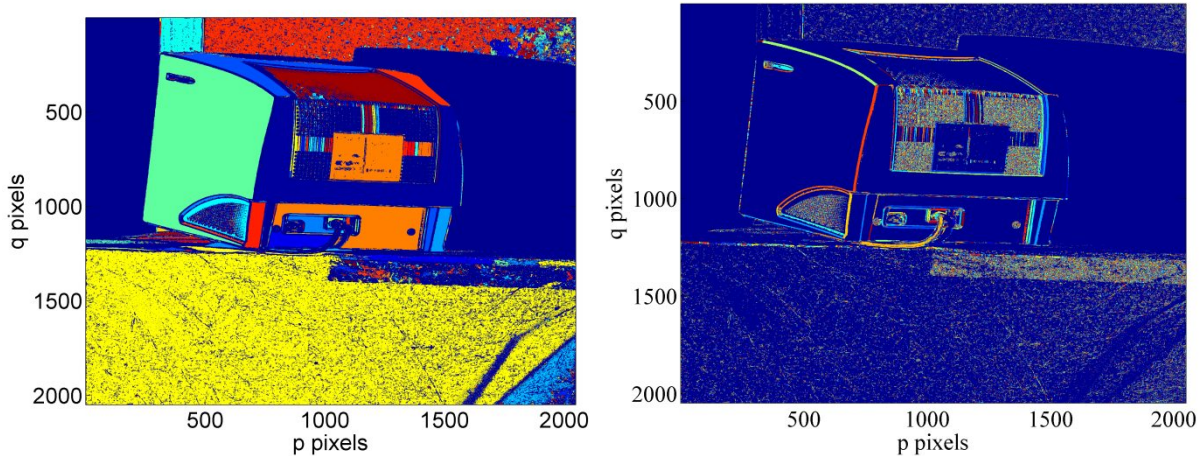


Figure 1 (a) Surface segmentation and (b) curve segmentation of shape data from a pose of a CRT monitor. Different segments are shown in different colours.

The first general eigenvalue of F_I and F_{II} gives the largest curvature and its corresponding eigenvector gives the direction about which the surface curves the most. The second eigenvector is orthogonal to the first and hence is the tangent of the curve if the first eigenvalue is larger than a threshold value.

Figure 1(a) shows the segmented surfaces of the CRT monitor and Figure 1(b) shows the segmented curves of the same object in the same pose. The curves act as separators between surfaces in Figure 1(a), and corners act as the separators between curves in Figure 1(b).

2.2 Surface and curve feature representation

The detected surface segments are cut into smaller surface patches as shown in figure 2(a), and the detected curves are cut into smaller curve segments as shown in Figure 2(b). In both cases, $u - v - w$ are the principal axes of the surface and curve respectively, found using principal component analysis (PCA) on the surface points and curve points for each surface and curve.

For each smaller surface patch and smaller curve segment, a second PCA is performed on the points within the smaller set. A surface patch is then represented by the 6 coefficients $(\alpha_1, \dots, \alpha_6)$ of the quadric surface $Q(u, v) = \alpha_1 u^2 + \alpha_2 v^2 + \alpha_3 uv + \alpha_4 u + \alpha_5 v + \alpha_6$. A curve segment is represented by the 6 coefficients $(\alpha_{1u}, \alpha_{2u}, \alpha_{3u}, \alpha_{1v}, \alpha_{2v}, \alpha_{3v})$ of the parametric curve defined in Figure 2(b). For surface patches the centre \mathbf{c} and surface normal \mathbf{n} are kept as additional matching parameters, whereas for curve segments the centre \mathbf{c} and the curve tangent \mathbf{t} are kept as additional matching parameters.

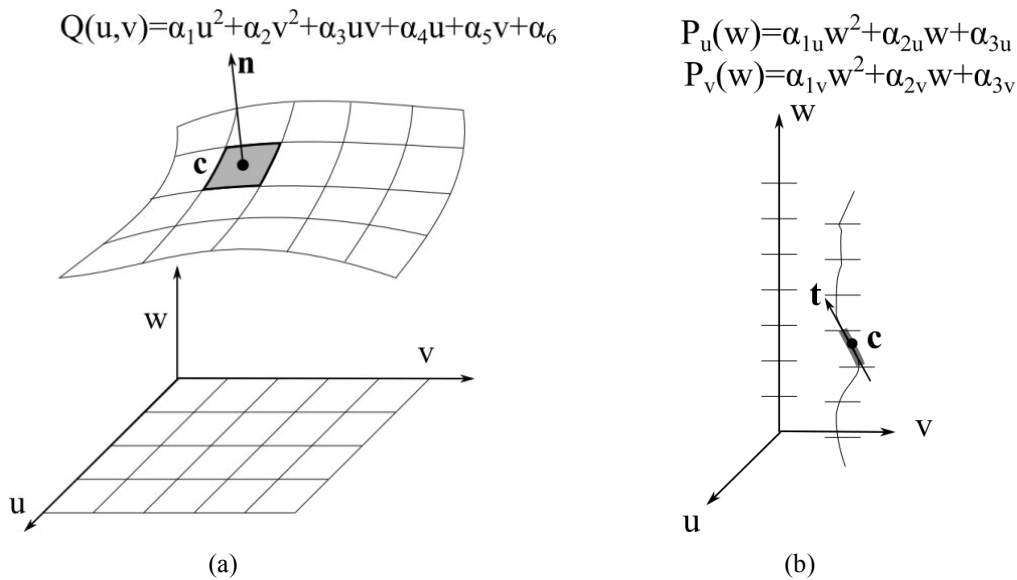


Figure 2. Representation of (a) a surface as a collection of surface patches and properties of one of those patches, (b) a curve as a collection of curve segments and properties of those segments.

2.3 Probability density function of a surface patch and a curve segment

When probabilistically matching surface or curve features, the features are first transformed into the $u-v-w$ frame of the j th model feature¹. The centre \mathbf{c}_{si} and the surface normal \mathbf{n}_{si} of the i th scene patch is then translated along the patch to a location directly above or below the model centre \mathbf{c}_{mj} using the quadric function $Q_{sj}(u, v)$. For curve segments, the centre \mathbf{c}_{si} and curve tangent \mathbf{t}_{si} are similarly translated along the curve to a location on the same constant w plane as the model centre \mathbf{c}_{mj} using the parametric functions $P_{usi}(w)$, $P_{vsi}(w)$.

The mutual matching function¹ g_{ij} can be expressed as a combination of matching parameters of the i th scene surface patch and the j th model surface patch, or the i th scene curve segment and j th model curve segment, as follows:

$$g_{ij}(\mathbf{m}_j|\boldsymbol{\theta}) = \frac{1}{\sigma_{ijw}\sigma_{ijr}\sigma_{ij\phi}(2\pi)^{3/2}} \exp\left(-\frac{w_{ij}^2}{2\sigma_{ijw}^2}\right) \exp\left(-\frac{r_{ij}^2}{2\sigma_{ijr}^2}\right) \exp\left(-\frac{\phi_{ij}^2}{2\sigma_{ij\phi}^2}\right) \quad (6)$$

where r_{ij} is the distance within the u - v plane between the two centre points, w_{ij} is the corresponding distance along the w axis, ϕ_{ij} is the angle between the patch normal (surfaces) or line segment tangents (curves); and σ_{ijk} is the standard deviation of the distribution in the vertical ($k = w$), radial ($k = r$) and angle ($k = \phi$) directions.

The standard deviations, σ_{ijr} , σ_{ijw} and $\sigma_{ij\phi}$ are estimated by the variation of the points relative to centre \mathbf{c}_{mj} and variations of surface normal relative to the patch surface normal \mathbf{n}_{mj} for each model patch. For each curve segment, the standard deviations, σ_{ijr} , σ_{ijw} and $\sigma_{ij\phi}$ are estimated by the variation of the points relative to centre \mathbf{c}_{mj} and variations of curve tangent relative to the curve segment tangent \mathbf{t}_{mj} .

2.4 Combining surfaces and curves for localisation

The log likelihood objective function used previously¹ is modified in this study by combining the log likelihood functions of surfaces ($\log L_{surf}(\boldsymbol{\theta}|\mathbf{s}_{surf})$) and curves ($\log L_{curv}(\boldsymbol{\theta}|\mathbf{s}_{curv})$) into a single function, by normalising and weighting their relative contributions as follows:

$$\log L(\boldsymbol{\theta}|\mathbf{s}) = \frac{\gamma}{N_{surf}} \log L_{surf}(\boldsymbol{\theta}|\mathbf{s}_{surf}) + \frac{(1-\gamma)}{N_{curv}} \log L_{curv}(\boldsymbol{\theta}|\mathbf{s}_{curv}). \quad (7)$$

In the above, γ (where $0 < \gamma < 1$) is a weighting factor determining the relative contributions of the surface features and curve features in the matching process. N_{surf} and N_{curv} are, respectively, the numbers of surfaces and curves detected in the scene.

3. METHOD

3.1 The ‘gold standard’ model

A model was created from scans of multiple poses of the model object. The coordinate transform between the poses was obtained by placing white circular markers either on the object or on the platform the object is rotated upon. By matching the centres of these circular markers, one can obtain the required rigid body transformations¹. Surface and curve segmentation is performed on the scan data from each pose, and the point clouds of each surface and curve identified in different poses of the model object are then merged. The surface and curve ‘gold standard’ models thus obtained are shown in Figures 3(a) and 3(b) respectively.



Figure 3. (a) Surface feature model and (b) curve feature model of a CRT monitor model.

3.2 Scenes

For each scene containing the CRT model, a two-step process is used to localise the object within the scene as discussed in the previous study¹. In the coarse localisation stage the log-likelihood function was broadened by scaling the standard deviations σ_{ijw} , σ_{ijr} and $\sigma_{ij\phi}$ until the algorithm is able to move away from the starting position. In addition, the contribution of surfaces (γ) was set to a large value ($\gamma = 0.9$) which, through maximising the log-likelihood function, provided a coarse estimate of the location and orientation of the model object within the scene.

Once the correct coarse location was found, pose refinement was implemented by using the smaller values for σ_{ijw} , σ_{ijuv} and $\sigma_{ij\phi}$ and reducing the surface contribution down to ($\gamma = 0.1$). This gave more prominence to the curve contribution to the log likelihood function when finding the refined pose.

4. RESULTS

Multiple scenes both with the CRT monitor in isolation and within clutter were analysed in order to observe the improvement in accuracy when employing curve details in addition to the surface information. The closest points were calculated from each scene point cloud to the rotated and translated model point cloud within the scene, from which alignment error maps were generated. Figure 4(a) shows the logarithmic error map obtained for the CRT model in isolation using only surfaces while Figure 4(b) shows that using both surfaces and curves. One can see that the latter is a significant improvement on the former. The mean absolute alignment error over the object within the scene has improved from 1.11 mm down to 0.36 mm.

Figure 5(a) and 5(b) show the error maps for a cluttered scene containing the CRT monitor, when the pose estimation was obtained from the surface-only method, and the combined surfaces and curves method, respectively. Once again, the improvement is visible from the images. The mean absolute alignment error over the object has improved in this case from 1.18 mm down to 0.61 mm. Table 1 shows the mean absolute alignment errors obtained for more scenes containing the CRT monitor as evidence for the improvement in accuracy obtained by employing the proposed technique. As can be observed, for all scenes considered, the mean absolute alignment error reported for the combined method is less than 0.61 mm and for most scenes the mean absolute error is below 0.5 mm. Therefore, an improvement in accuracy of pose estimation between 1.5 \times and 3 \times has been achieved.

Table 1. The mean absolute alignment error over the whole object within the scene calculated for the pose estimates obtained using ML algorithm with only surfaces and ML algorithm with both surfaces and curves.

CRT Pose	Surfaces only	Surfaces and curves
Isolated (1)	0.55 mm	0.39 mm
Isolated (2)	1.11 mm	0.36 mm
Isolated (3)	1.02 mm	0.44 mm
Isolated (4)	0.74 mm	0.50 mm
Clutter (1)	1.18 mm	0.61 mm
Clutter (2)	0.76 mm	0.46 mm

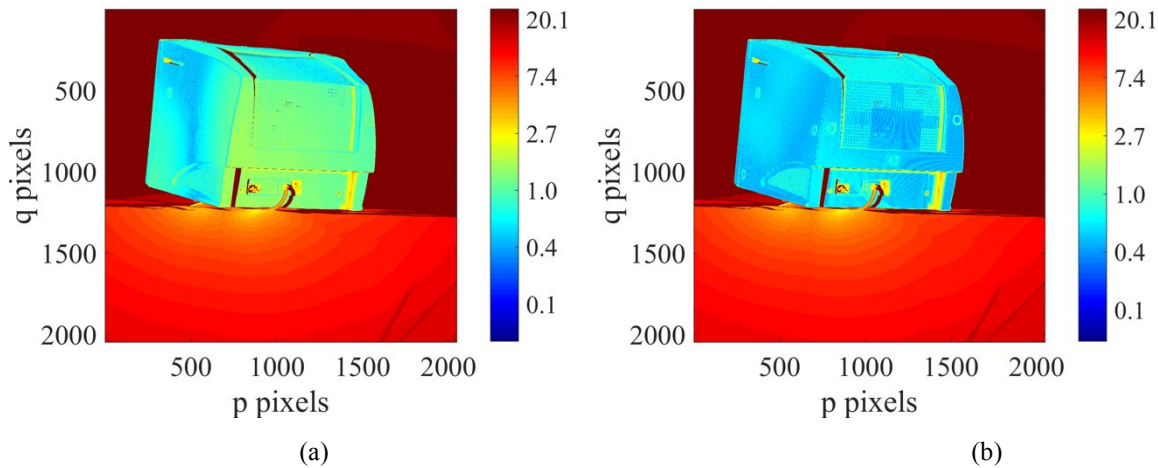


Figure 4. Logarithmic absolute error map (units: mm) between the scene and the model transformed by the pose obtained by (a) the maximum likelihood algorithm with only surfaces, (b) the algorithm with the proposed coarse ($\gamma = 0.9$) and refinement ($\gamma = 0.1$) stages for the CRT monitor in isolation.

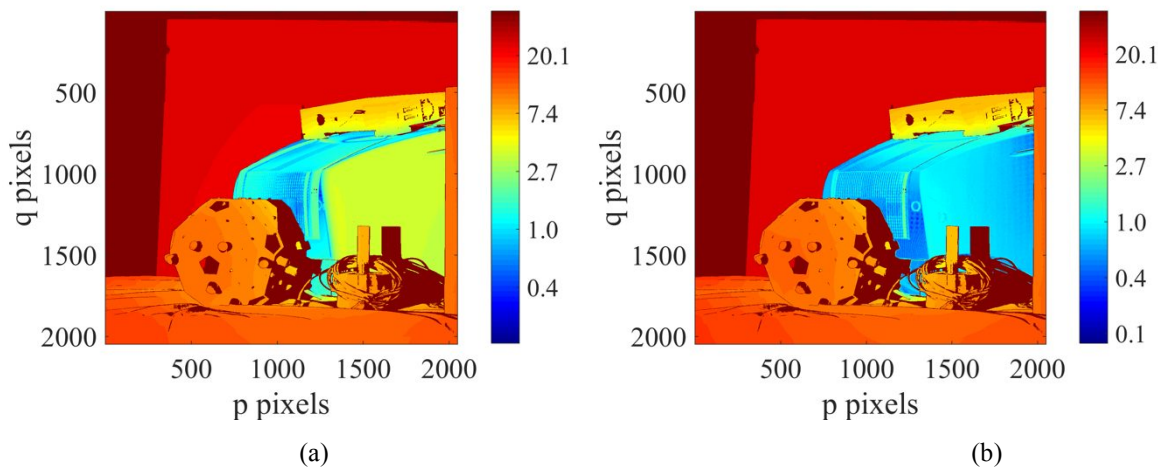


Figure 5. Logarithmic absolute error map (units: mm) between the scene and the model transformed by the pose obtained by (a) the maximum likelihood algorithm with only surfaces, (b) the algorithm with the proposed coarse ($\gamma = 0.9$) and refinement ($\gamma = 0.1$) stages for the CRT monitor in clutter.

5. CONCLUSION

Pose estimation is a fundamental problem in automated verification and assembly in manufacturing industry. A probabilistic object recognition and pose estimation algorithm based on maximum likelihood was presented recently¹. In the current paper, the algorithm has been extended to include curve features in objects. The details of the algorithm, including the detection and characterisation of curves in 3D point clouds, and a probabilistic matching equation that allows the relative weighting of surface and contributions to be controlled with a single coefficient (γ), have been presented.

The two step process of coarse and fine localisation¹ is modified so that the coarse localisation is achieved with more contribution from the surface likelihood function ($\gamma = 0.9$), and the pose refinement is achieved with a larger contribution from the curve likelihood function ($\gamma = 0.1$). From the results obtained for a CRT monitor model for a number of scenes, we conclude that the inclusion of curves in the maximum likelihood algorithm improves the overall accuracy of the pose estimation by a factor of $1.5 \times - 3 \times$. The mean alignment error for all scenes was lower than 0.61 mm.

ACKNOWLEDGMENT

The research was funded by the Engineering and Physical Sciences Research Council under the Light Controlled Factory project EP/K018124/1.

REFERENCES

- [1] Dantanarayana, H.G. and Huntley, J.M., "Object recognition in 3D point clouds with maximum likelihood estimation," *Proc. SPIE*, 95300F–95300 (2015).
- [2] Besl, P.J. and Jain, R.C., "Three-dimensional object recognition," *ACM Comput. Surv.* 17(1), 75–145 (1985).
- [3] Yu, T.H., Woodford, O.J. and Cipolla, R., "A performance evaluation of volumetric 3D interest point detectors," *Int. J. Comput. Vis.* 102(1-3), 180–197 (2013).
- [4] Mikolajczyk, K. and Schmid, C., "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.* 27(10), 1615–1630 (2005).
- [5] Andreopoulos, A. and Tsotsos, J.K., "50 years of object recognition: directions forward," *Comput. Vis. Image Underst.* 117(8), 827–891 (2013).
- [6] Mukhopadhyay, P. and Chaudhuri, B.B., "A survey of Hough Transform," *Pattern Recogn.* 48(3), 993–1010 (2015).
- [7] Fischler, M.A. and Bolles, R.C., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* 24(6), 381–395 (1981).
- [8] Yuan, C., Yu, X. and Luo, Z., "3D point cloud matching based on principal component analysis and iterative closest point algorithm," *IEEE Intl. Conf. Audio, Language and Image Processing*, 404–408 (2016).
- [9] Besl, P.J. and Jain, R.C., "Invariant surface characteristics for 3D object recognition in range images," *Comput. Vis. Graph. Image Process.*, 33(1), 33–80 (1986).