Recognition of Cylindrical Objects Using Occluding Boundaries Obtained from Colour Based Segmentation

Dekun Yang, Josef Kittler and George Matas
Department of Electronics and Electrical Engineering,
University of Surrey, Guildford. GU2 5XH

Abstract

This paper describes a method for model-based recognition of cylindrical objects from occluding boundaries obtained by computationally efficient colour segmentation of a 2D image. The models are invoked by combining geometric and colour features. Occluding boundaries of hypothesized objects are generated using colour segmentation and ground plane constraint. Hypothesis verification is achieved by evaluating the fit between occluding boundary generated by the hypothesised object and the edge data. This method differs from existing methods in that it integrates multiple measurements and prior knowledge to achieve robust object recognition. Experiments with real images have been carried out and the results are promising.

1 Introduction

The ultimate task in computer vision is concerned with interpreting image data in terms of objects so as to provide a symbolic scene description and aid understanding. It involves separating objects of interest from the background, inferring their 3D positions, identifying them and consequently obtaining symbolic description of the scene. In order to achieve this task efficiently, it may be beneficial to integrate different techniques and to make use of some prior knowledge about the scene. This paper describes a method for model-based recognition of 3D objects which combines different sources of information such as geometric primitives, colour features and object location constraints. It complements a plethora of existing object recognition techniques and may be invoked by a vision system in situations when image description in terms of edge segments is either distorted or over segmented. Such situations need not be detected in vision systems such as the VAP (Vision as Process) vision system where multiple recognition knowledge sources are launched in parallel and many hypotheses generated by them are evaluated and acted on by a hypotheses manager.

Model-based object recognition is a process of interpreting image data in terms of object models. This problem arises when we analyse and attempt to interpret a 3D scene. Surveys of object recognition literature for pre-1985 may be found in [1, 6] and recent work may be consulted in [8, 2] and the reference therein. Most of the previous work has been devoted to recognizing polyhedral objects. which is

BMVC 1994 doi:10.5244/C.8.43

very limited since the world contains a wide variety of curved objects. In recent years an important area of research in computer vision has been the recognition of curved objects [5, 13]. Attempts have been made to infer 3D information about curved objects from their occluding contours by employing geometric constraints such as the extremum principle [4]. Most of the existing techniques rely on the edge based detection of object outline and on the detection of curvature extrema to achieve model invocation and hypotheses verification. However, there are certain problems that seem to be inherent in these techniques. The extraction of the outline of object from image data can be unreliable due to the inaccuracy of the edge detection process. It is also difficult to detect curvature extrema accurately, especially when the curves are noisy. The difficulties in obtaining good edge data and estimates of curvature are the main obstacles in most existing approaches.

In this paper we show a way of alleviating the difficulties by integrating information from different sources to achieve model invocation and hypothesis verification. The underlying ideas are the following. First, model invocation can be achieved by using other properties of objects such as colour rather than relying entirely on geometric cues. Under the ground plane constraint, i.e., when objects to be analysed are located on a known plane, the hypothesis of a 3D object identity can be generated based on the 2D region obtained from colour segmentation. Second, hypothesis verification can be robustly achieved by refinement using edge data. Accepting and rejecting the generated hypotheses based on multiple sets of measurements, i.g., colour and edge measurements, has an obvious advantage over a single set because the accumulation of evidence leads to a more reliable decision. To demonstrate the ideas, we consider in this paper the problem of recognising cylindrical objects. The reason for concentrating on cylindrical objects is that many man made objects appearing in our visual sensing scenario, e.g. objects on breakfast table, have this shape.

The method presented in this paper consists of two processes: model invocation and hypothesis verification. For the former, 2D regions corresponding cylindrical objects are first obtained by computationally efficient colour based segmentation. For each 2D region obtained from colour segmentation, the 3D information such as location, width and height of the possible cylindrical object is then inferred by imposing ground plane constraint. Object hypothesis is generated using colour, width and height as the indices to the model library. Occluding boundaries of the hypothesized objects are computed as model features. For the latter, hypothesis is first refined by integrating edge data from Canny edge detector. This is achieved by searching optimal estimate of the object location which minimises the distance between the model features and the edge data obtained from Canny edge detector. Hypotheses verification is then achieved by evaluating the fit of model features to the edge data.

This method differs from existing methods in the following respects. Firstly, it generates object hypothesis from 2D regions obtained from colour segmentation rather than curvature estimates obtained from edge data. Previous work on finding cylindrical objects was based on the detection of ellipse features in the image data. The most common method of ellipse detection [3] performs the fitting of an ellipse to a set of chained points based on the measurement of curvature along the chain. However, several difficulties may arise when curvature estimates are noisy or the

edges of ellipse are not connected well. In contrast with previous work, our method can generate more reliable object hypothesis due to the robust colour segmentation. Secondly, it refines object hypotheses by integrating other information, i.e., edge data, before they are confirmed or rejected. The refinement of hypotheses leads to a more robust and reliable decision regarding object identity.

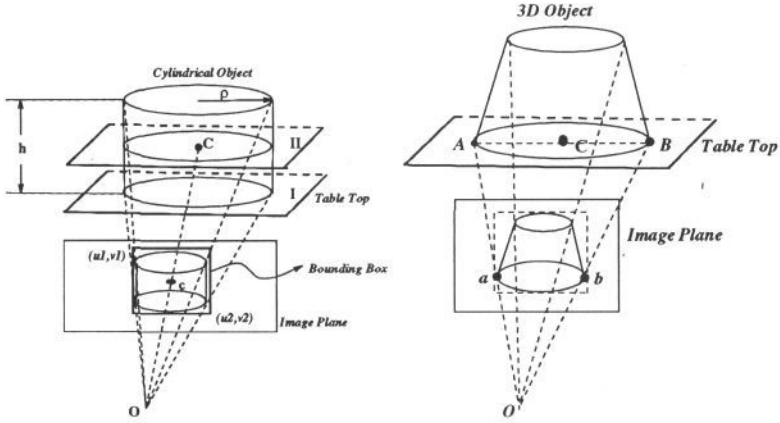
The paper is organised as follows. In the next section the approach adopted to object hypothesis generation is described. The problem of hypothesis verification is addressed in Section . Section 4 details experiments performed to illustrate the method. Finally Section 5 offers some conclusions.

2 Hypothesis Generation Using Colour Based Segmentation

The occluding boundary of an object contains important shape information which can be used for object recognition. As pointed out in the previous section, it may be difficult to obtain reliable occluding boundaries of scene objects if their extraction is based edge detection. In this section we describe a method which generates the occluding boundaries of hypothesized cylindrical objects using colour based segmentation and ground plane constraint.

Colour segmentation process divides the image into homogeneous regions using colour information at each pixel. Colour based methods for extracting object boundary offer an advantage over edge based methods in that the regions belonging to the same object can easily be extracted by grouping. In the literature existing methods fall into two categories depending upon the grouping mechanisms involved: clustering [9] or recursive region splitting [12]. Colour segmentation can also be achieved by exploiting chromatic differences in images. Recently, a method for extracting object contour by integrating colour and motion has been presented by Dubuisson and Jain [7]. In this paper we use an illumination invariant colour recognition method described elsewhere [10]. In order to deal with image sequences efficiently, the capability of the colour recognition method was extended to provide a segmentation method based on chromatic differences in images of a scene before and after a dynamic event. The segmentation method can be applied only when the camera is static. For a moving camera a different hypothesis generation and maintenance mechanisms are employed. The use of chromatic values [14], rather then direct colour values, allows us to discern between shadows cast by the new object and changes at locations where the new object occluded the background. Connected components of pixels with high chromatic difference are found. Therefore, we are able reliably to segment regions of different colour when we either know prior colour model of objects or know that the objects can be moved in the scene. Colour segmentation provides a number of uniform regions that are referred to as mask. The occluding boundary of an object is obtained as the occluding boundary of the mask. The bounding box for each occluding boundary is used as a 2D region hypothesis.

Region hypotheses obtained from the colour feature space are in the 2D domain. In order to generate 3D object hypotheses, it is necessary to integrate them with another knowledge source. Under the assumption that objects to be analysed are



(a) (b)

Figure 1: The 3D-2D geometric relation

located on a reference plane, e.g., table top, we can obtain 3D information about a given 2D region. We begin the detailed analysis with some definitions and notation. We assume an idealised pinhole camera by which the image is formed through a perspective projection. The camera is associated with an XYZ coordinate system with origin O at the optical center and Z axis along the optical axis. The image plane is identical with the plane $Z = f$ where f is referred to as the focal length. The image coordinate system UV is defined such that U and V axes are parallel to the X and Y axes, respectively. Given a 3D point $M = (x, y, z)$, its image $m = (u, v)$ is the intersection of the image plane and the line through M and the optical center, which can be written by the following relation

$$\begin{pmatrix} \omega u \\ \omega v \\ \omega \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} f/s_u & 0 & u_0 \\ 0 & f/s_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{1}$$

where ω is an arbitrary scalar and the elements in the 3×3 matrix are the camera's intrinsic parameters. The quantities s_u and s_v can be interpreted as the size in the camera coordinate unit of the horizontal and vertical pixel, respectively, and (u_0, v_0) is the coordinate of the principal point of the camera, i.e., the intersection between the optical axis and the image plane.

Figure 1 (a) shows the geometric relation between the 3D scene and the 2D image plane. Given a 2D image coordinate $m = (u, v)$, a ray which its corresponding 3D point must lie on, is determined by

$$\boldsymbol{p} = \{\boldsymbol{O} + \mu(\tilde{m} - \boldsymbol{O}), \mu > 0\} \tag{2}$$

where $\boldsymbol{O}$ is the optical center of the camera and $\tilde{m} = A^{-1}(u, v, 1)^t$. If the corresponding 3D point is on a know plane $f(x, y, z) = 0$, then the 3D point can be recovered by intersecting the ray $\boldsymbol{p}$ into the plane $f(x, y, z) = 0$. Usually, we may

define the world coordinate system $X_w Y_w Z_w$ such that its X_w and Y_w axes lie on the plane $f(x, y, z) = 0$. Assuming that the camera coordinate system and the world coordinate system are related by

$$(x, y, z)^t = R(x_w, y_w, z_w)^t + T \tag{3}$$

where superscript t denotes transpose and R and T are the rotation matrix and the translation vector, we obtain the one-to-one mapping between image coordinate $m = (u, v)$ and its 3D point on the plane $z_w = 0$, $M = (x_w, y_w, 0)^t$, as follows.

$$\omega(u, v, 1)^t = G(x_w, y_w, 1)^t \tag{4}$$

where $G = A[r_1\ r_2\ T]$ is a 3×3 matrix of rank of 3 which represents a projective transformation between the plane $z_w = 0$ and the image plane, r_1 and r_2 are the first and second rows of R respectively.

Let us now proceed to the problem of generating occluding boundary of a cylindrical object based on colour. Each cylindrical object is modelled by two parallel circles (top and bottom) and two line segments. Its geometric characteristic is specified by the radius of circle ρ and height h (see figure 1 (a)). Since the symmetry of cylindrical objects is preserved under 3D-2D projection transformation, the center of the occluding boundary is a good approximation of the projection of the center of the 3D cylindrical object. We recover the center of cylindrical object using the hypothesise-verify approach as follows. For each 2D region, we obtain the center of bounding box and assume that it is identical to the center of the occluding boundary. We choose each model in the library as the hypothesized cylindrical object. With known table top and the height of the hypothesized cylindrical object we obtain the plane I that is parallel to the table top and passes through the center of the cylindrical object. We define the world coordinate system such that its X_w and Y_w axes lie on the plane I. Using equation (4), we obtain the 3D point of the center of the hypothesized cylindrical object. We then project the hypothesized cylindrical object onto image plane. If the projected model occluding boundary is close to the detected occluding boundary extracted using colour based segmentation then we choose this object hypothesis as valid hypothesis to be further verified. The criterion of similarity between the projected occluding boundary and the actual occluding boundary detected from image data will be addressed in detailed in the next section.

We can extend the proposed method to deal with conic objects as shown in figure 1 (b). To achieve this, we use the occluding boundary obtained from colour segmentation rather than the bounding box. We then detect two salient local feature points a and b as shown in figure 1 (b). Generally, these two points can be detected by employing curvature detection methods. However, colour segmentation provides connected occluding boundary S_c in image data, which allows us to obtain the two points simply by finding two points which have minimum x component and maximum x component respectively. Once the image coordinates for the two salient points are obtained, we can compute their 3D positions on table top, A and B. The center of the hypothesized conic object can be easily found as the mid-point of A and B.

The next step is to combine the geometric and colour information, i.e., three attributes, radius, height and colour, of cylindrical objects to form indices to access

entries in the object library. Usually a small number of models rather than a unique model are retrieved from the object library due to the inaccuracy of the 2D region based hypothesis generation or simply because of genuine ambiguities in object identities in terms of these attributes. We verify each hypothesis as described in the next section.

3 Hypotheses Refinement and Verification

Object hypotheses generated using colour information and ground-plane constraint are ambiguous, due to the inadequacy of colour segmentation process. In particular, (i) for a given 2D region there may be false positive recognition hypothesis and (ii) there is uncertainty regarding the pose of the hypothesised 3D objects. In order to remove the ambiguities, another independent information source, edges obtained from Canny edge detector, is used for refining and verifying the hypotheses since discontinuities of image intensity convey important information for object boundaries. The task of hypothesis refinement is to estimate the best pose of hypothesised 3D objects while the task of hypothesis verification is to obtain a unique interpretation for a given 2D region. To do this, we project the hypothesised 3D object model into the image and evaluate how well model points fit the image data.

For cylindrical objects, the projections contain conics because conics are always projected onto conics by perspective projection. Given a conic Q represented by

$$(x, y, 1)^t Q(x, y, 1) = 0 \tag{5}$$

its projection on the image plane is given by

$$Q' = k(G^{-1})^t Q G^{-1} \tag{6}$$

where k is an arbitrary nonzero constant and G is the projective transformation between the plane in which the conic Q lies on and the image plane. With calibrated camera and the prior knowledge that the objects to be analysed are located on the table top, we can easily obtain the projection of each hypothesised cylindrical object. We remove the hidden boundaries to obtain the entire visible occluding boundary of the model object. We sample the occluding boundary to obtain a set of model points denoted by $S = \{\boldsymbol{m}_i = (u_i, v_i),\ i = 1, \cdots, N\}$ where N is the number of model points that depends on the sampling rate.

Given a set of edges obtained from Canny edge detector, $S' = \{\boldsymbol{m}'_i = (u'_i, v'_i),\ i = 1, \cdots, N'\}$, we formulate the problem of hypothesis refinement as the problem of estimating the location (C_x, C_y) of the hypothesised cylindrical object on the table top, i.e.,minimising the overall Euclidean distance between the model points and the edge points:

$$\mathcal{F}(C_x, C_y) = \frac{1}{\sum_{i=1}^{N} \lambda_i} \sum_{i=1}^{N} \lambda_i d^2(\boldsymbol{m}_i(C_x, C_y), S') \tag{7}$$

with respect to the parameters C_x and C_y, where

$$d(\boldsymbol{m}_i, \boldsymbol{S}') = \min_{\boldsymbol{m}'_j \in S'} |\boldsymbol{m}_i - \boldsymbol{m}'_j| \tag{8}$$

is the closest distance between point $\boldsymbol{m}$ and the set of points S. The quantity λ_i takes value of "1" if $\boldsymbol{m}_i$ is closest to the set S' and its distance is less than a threshold ε_1, and "0" otherwise.

Since the objective function $\mathcal{F}$ is non-linear, we need an iterative minimization procedure. With an initial estimate of location (C_x, C_y), we perform the minimization using the method described in [15]: (i) We find the closest distance for each point m_i, i.e., $d(\boldsymbol{m}_i, \boldsymbol{S}')$. We then set values 0 or 1 to λ_i depending on whether the closest distance is less than threshold ε_1. (ii) The values λ_i are updated through a statistical analysis of distances between pairs of matched points. We compute the mean μ and the deviation σ of the distances. We set $\lambda_i = 0$ if the distance is greater than $\mu + 3\sigma$. This treatment is justified because the difference between paired points is nearly the same for most pairs due to small changes of the model location. (iii) We then determine the optimal estimate of location $(\hat{C}_x, \hat{C}_y)$ which minimises the cost function $\mathcal{F}$. We apply the downhill simplex method to perform the nonlinear two-dimensional minimization. The method performs a descent through the two-dimensional topography until it reaches a local minimum. It requires 3 initial estimates which can be obtained by perturbing the given initial estimate along x and y directions respectively. (iv) We update the set of model points using the estimate of cylinder location. (v) We repeat the above steps until $|d^{(k+1)} - d^k|$ is less than a threshold ε_2, where k is the index of steps.

It is clear that a higher sampling rate leads to a better estimate of the cylinder location but less efficient performance due to more model points being involved. A way to alleviate the computational difficulty is to employ a multi-resolution scheme for sampling. At the beginning of the iterative minimisation we use a coarse sampling by choosing some salient points of the object outline as the model points. We increase the sampling rate while the iteratively minimisation proceeds.

Because the function $\mathcal{F}$ measures the similarity between the projected model and the image data, it can be also used to measure the adequacy of the hypothesis. The decision whether the hypothesis is true or not is made according to

$$\mathcal{F}(C_x, C_y) \begin{array}{ll} < \delta & \text{accept} \\ > \delta & \text{reject} \end{array} \tag{9}$$

Given the distribution of image points $\boldsymbol{m}'$ reflecting the noise introduced by the process of extracting the outline of the object, we can find the distribution of $\mathcal{F}$ so as to set the threshold δ. However, the derivation of the distribution of $\mathcal{F}$ is tedious in practice. Consequently, the choice of δ in our implementation is heuristic based on the number of object models to be analysed.

4 Experimental Results and Discussion

We ran experiments on a sequence of real images with a resolution of 720 by 576 pixels. The sequence recorded a 'breakfast scene' where there were many different cylindrical objects such as milk-jugs, sugar-bowls and cups. Due to the space limit we show experiments in detail only for an image which is shown in figure 2 (a). Figure 2 (b) shows the edge map extracted using Canny edge detector. Because of noise introduced by the low level vision process, the object boundaries

are recovered, but with many object internal edges and the background noise edges. Therefore curvature based methods for recognising objects from occluding boundaries are not expected to obtain robust results. However, Canny edge map is good enough to be used in the proposed method because imperfect boundaries are not used to generate hypotheses. Instead the proposed method uses colour information [11] to determine the 2D regions associated with the scene objects. Figure 2 (c) shows the occluding boundaries obtained. Figure 2 (d) shows the corresponding bounding boxes for the 2D regions extracted. The bounding boxes are computed from the minima and maxima of the boundaries. As we can see, each object in the image has been well segmented. Small regions and holes were removed.

Compared to the Canny edge map, we can observe that (i) the contours obtained from the colour based segmentation provide good 2D regions which correspond to different objects but do not describe the object boundary accurately; (ii) the contour segments obtained from the Canny edge detector give good information of object boundary since they are detected based on gradient intensities, but it is difficult to group the contours into objects. Therefore, it is desirable to fully exploit information of different sources in order to improve the recognition performance. This is the reason why we use contours from the segmentation process to generate object hypotheses while contours of Canny edges are used for verification.

With known camera calibration and ground-plane constraint, we recovered 3D information, radius, height and position, of 2D regions obtained from colour segmentation. We used colour, radius and height as indices to generate 3D hypotheses of objects. We considered four cylindrical objects, milk-jug, sugar-bowl, cup and saucer. Their models are described in figure 2 (e). Since it is very likely that a cup will be placed on a saucer we consider them as a compound model so as to improve efficiency in recognition. Thus in total we have 5 cylinder-like objects. We obtained 6 2D regions based on colour segmentation. The biggest region and the smallest region shown in figure2 (d) correspond to a hand and a spoon respectively. They are rejected because their 3D sizes recovered are either too big or too small with respect to model data. For each of the remaining 4 2D regions we have 2 or 3 object hypotheses. For each objet hypothesis we project the model object into image plane from the hypothesised location. We first removed the hidden model contours and sampled the contours for comparison with Canny edges. We sampled each model contour so that each set of model contains 150 points.

For each hypothesis we refined the estimate of location by non-linear minimization. Firstly, the closest point is computed for each model point to establish a match between model and Canny edge strings. Secondly, the match is updated by statistical analysis of the distance of matched points. Thirdly, an optimal estimate is computed based on matched points and new model point correspondence are obtained by applying the estimate. On average, it takes less than 15 iterations to converge which takes about 30 seconds (including the time for colour segmentation) on a SUN Sparc 10 workstation. The result is shown in figure 2 (f) where the models superimposed onto the scene objects are highlighted in white.

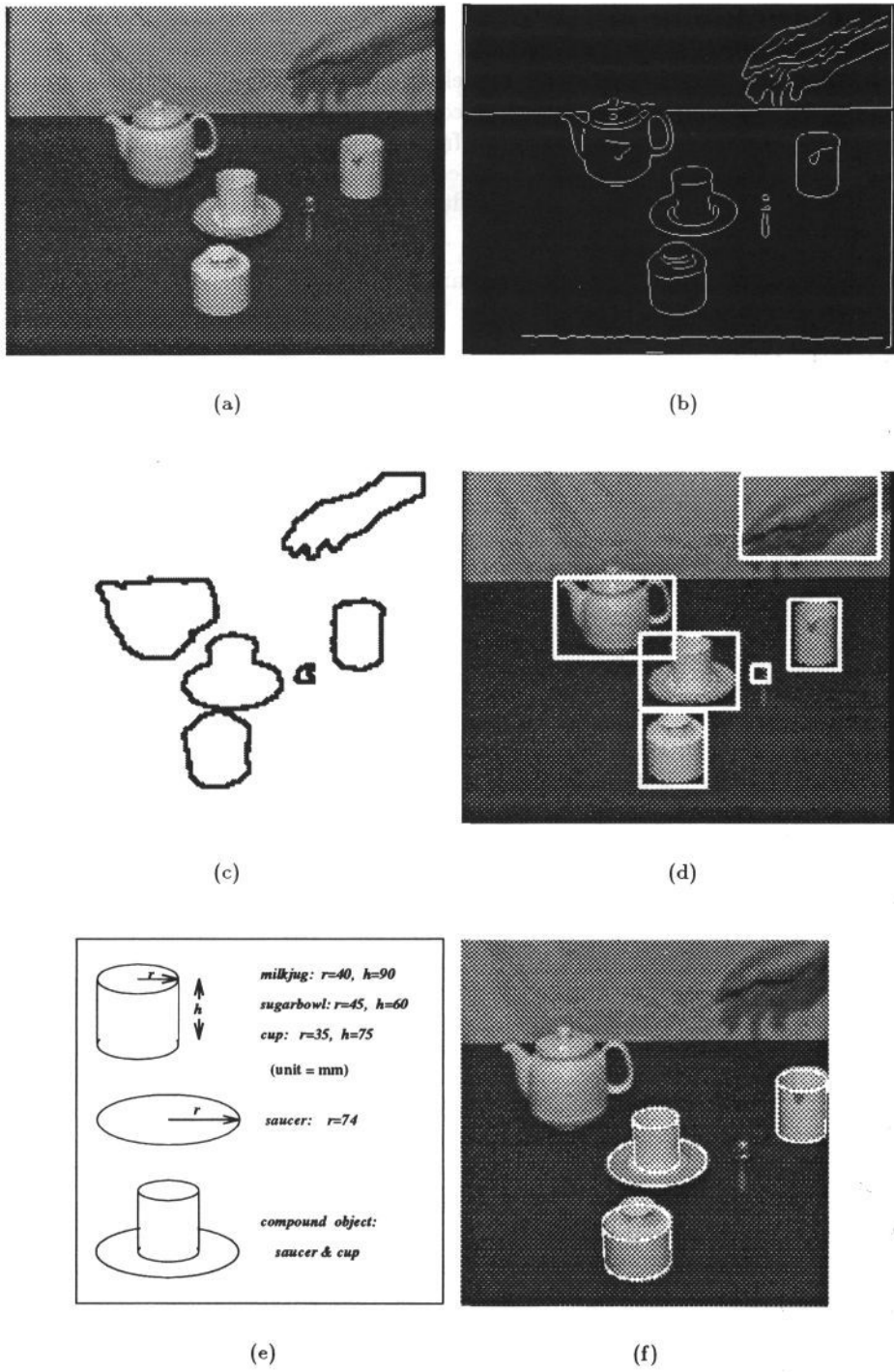


Figure 2: (a) An Image of the scene. (b) Edge data obtained from Canny edge detector. (c) Occluding boundaries obtained by colour based segmentation. (d) Bounding boxes of the boundaries. (e) Models of Cylindrical Objects (f) Recognition result

5 Conclusions

A method has been presented for model-based recognition of cylindrical objects from occluding boundaries obtained by computationally efficient colour segmentation of a 2D image. This method differs from existing methods in that it integrates multiple measurements and prior knowledge to achieve robust object recognition. Experimental results obtained on real data demonstrate its viability and advantages.

Acknowledgement: This work was carried out under the ESPRIT project 7108 "*Vision as Process II*".

References

[1] P. Besl and R. Jain. Three-dimensional object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:75–145, 1985.

[2] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239–256, 1992.

[3] F. Bookstein. Fitting conic sections to scatered data. *Computer Vision, Graphics and Image Processing*, 9:56–71, 1979.

[4] M. Brady and A. Yuille. An extremum priciple for shape from contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:288–301, 1984.

[5] C. H. Chien and J. K. Aggarwal. Shape recognition from single silhouettes. In *First International Conference on Computer Vision*, London, UK, 1987, pages 481–490.

[6] R. Chin and C. Dyer. Model-based recognition in robot vision. *ACM Computing Surreys*, 18:67–108, 1986.

[7] M. P. Dubuisson and A. K. Jain. Object contour extraction using color and motion. In *Fourth International Conference on Computer Vision*, Berlin, Germany, 1993, pages 471–476.

[8] P.J. Flynn and A.K. Jain. CAD-based computer vision: From CAD models to relational graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:1066–1075, 1991.

[9] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data.* Prentice Hall, 1988.

[10] J. Matas, R. Marik, and J. Kittler. Generation, verification and localization of object hypotheses based ob colour. In *British Machine Vision Conference*, Surrey, UK, pages 539–548. BMVA Press, 1993.

[11] J. Matas, R. Marik, and J. Kittler. Illumination invariant colour recognition. In *British Machine Vision Conference*, York, UK, BMVA Press, 1994.

[12] R. Ohlander, K. Price, and R. Reddy. Picture segmentation using a recursive region splitting method. *Comp. Graph. and Image Proc.*, 8:313–333, 1978.

[13] J. Ponce. Invariant properties of straight homogeneous generalised cylinders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:951–965, 1989.

[14] G. Wyszecki and W. S. Stiles. *Color Scence: Concepts and Methods, Quantitative Data and Formulea.* John Wiley, 1982.

[15] Z. Zhang. On local matching of free-form curve. In *British Machine Vision Conference*, Leeds, UK, pages 347–356. Springer-Verlag, 1992.