Abstract—The use of grey-scale contours, and fingerprints derived from this, has recently been used to analyse images for object recognition. The processing of these data can take a number of different forms.

This paper describes a method for using characteristic aspects of the fingerprint, and geometrical relationships between them, to reduce an image to a set of simple geometric features. These features allow for compact storage in a database and very fast scanning to create a short list of candidate matching objects.

Keywords—artificial vision; machine vision; computer vision; image contours; artificial intelligence

I. INTRODUCTION

The problem of object recognition by computers has been active for more than two decades, and as yet has not been solved. To solve this would create many new opportunities for the automation industry, and increase efficiency in many other industries. Many current vision systems can distinguish between objects if the image is clear and without occlusions. The system described in this paper proposes a solution that is designed to identify objects that are partially occluded or viewed from different angles.

This paper forms part of a generalised object recognition system that is described further elsewhere [1,2]. The function of the method described here is to extract, from the image contours, features and relationships that can be represented in a database. A number of database objects can be selected that have a high probability of being contained within a captured image. The method must select all of the object images contained within the captured image (no false negatives) while including only a modicum of the others (false positives).

II. BACKGROUND

Recognition of general objects within a scene is a complex problem with vast potential. People have been trying to develop methods to deal with this for many years, with varying degrees of success. If an artificial vision system is created to a high enough standard, automation of many repetitive tasks will be possible.

The most common problems encountered are occlusions over part of the image, scaling and rotation of the image, and distortion of the image. Systems have been designed for recognition of objects from very specific sets, such as skeletal object recognition, which works very well for objects with complex skeletal structures [3]. Template matching [4] can identify many different object types, but only with small database sizes due to the complex calculations required. Many object recognition systems have been attempted that use contours to characterise each object, with many different ways to store and search through the data. One example [5] is similar to the system described this paper, by splitting a contour and analysing the components; however the system compares all the contour parts, which is slow, and still susceptible to occlusions.

The field lies open for a system efficient enough to be scaled up to larger databases. One such candidate [1,2] is based on the use of image contours and the extraction of the contour fingerprint.

The approach can be broken down into three main sections.

- The image is translated into a set of contours, which follow the outlines and significant grey level edges.
- The contours are reduced into a set of simple geometric features that describe the most significant parts of the original image.
- The features are processed to form a search word, which is compared to a database to find which objects are likely to be in the image.

This approach eliminates data not unique to the general object definition, such as colour and texture.

A. Fingerprints and Contour Generation

The image is translated into a set of contours; these are sets of adjacent points that have the same grey level. They will occur around edges or outlines in the image as can be seen in figures 1 and 2.

Figure 1. Original image of the mug.
The fingerprint for each contour is given by the first derivative of the contour local tangent angle, which is the local rate of curvature. This means sharp corners on the contour show up as spikes on the fingerprint (see figure 3), straight lines are successive points close to the x-axis, and arcs are successive points of equal value away from the x-axis. The use of fingerprints makes the recognition of significant features far easier, and the endpoints are found very accurately.

There will be many contours for each image. Depending on the complexity of the image this can range from 200 to over 1000. Because a human can still easily recognise the object from the contours, it can be assumed that these contours contain all the data necessary to identify the object within the image.

III. FEATURE-BASED IMAGE RECOGNITION

We come now to the extraction of features from the contours of the image. The features must be extracted one by one from the multitude of contours, which is a time-intensive process. Currently there are three different types of feature:

- **Line**: The line is defined by its centre point, length, and tangent angle. It is found by isolating the flat spots that occur at zero amplitude on the fingerprint plot. The line itself is fitted to the two endpoints on the part of the contour associated with the flat spot.

- **Arc**: The arc is defined by its middle point, length, radius and tangent angle. It is found similar to the line, but the flat spots must occur at higher amplitudes. The arc is then fitted to the two endpoints and the middle point along the part of the contour associated with the flat spot.

- **Lobe**: The lobe is defined by its middle point, two endpoints, length, tangent angle, and the angle the contour changes from start to finish. The exact contour cannot be recreated from this definition but the most important parts are covered, and it provides a very good approximation. The lobe is fitted to each spike on the fingerprint plot using the corresponding contour endpoints and tangents.

Each contour’s fingerprint has to be examined to extract the relevant features. The features are stored with the associated contour in the program. The number of total features usually ranges from one thousand to ten thousand.

All the features collected must be combined and reduced to find only the most important ones. This is done by checking for redundancy, then intelligently combining smaller features into larger ones.

- All types of features go through a grouping algorithm which finds matches using the various attributes for each feature type. A set of average feature attributes is created for each group.

- All the groups are checked for sufficient redundancy to ensure the feature occurs on multiple contours, meaning it is part of a significant outline. This means the number of features is now vastly reduced so a more intelligent grouping algorithm can be performed.

- Lines of different length are combined provided they are parallel and in close proximity. Calculations are done to make a new line that encompasses both lines that created it.

- Arcs are combined similarly to lines but the radius, intersection point, and local tangent angles are also taken into account. Again calculations are done to ensure it takes both smaller arcs into account.

- Lobes are combined if they are offset from one another; this is checked by finding the vectors between each of the three sets of points. The new lobe is calculated by moving the points a weighted distance along the vectors.

All of these reduction steps mean an image’s main outlines are described by only a handful of ‘super features’ (see figure 4), making the next set of processing steps far more efficient.
A. High-speed Database Search

The super-features are processed to find geometric relationships in the image, a sample of which are shown in the table. An array of Boolean variables is used to store what relationships are present. Most of the relationships must be programmed individually, and to date only the main ones have been added. The Boolean variables for the geometric relationships operate in a similar way to the V1 cells in our own vision system [7].

- Parallel lines
- Circles
- Squares
- Triangles
- Parallelograms
- Symmetrical lines
- Axis of symmetry common to three feature pairs
- Squares with rounded corners
- Rotationally-symmetric lines
- Right-angle line-lobe-line triples meeting at tangents

Using a single bit for each relationship means that if the object is occluded or some parts are missing, only a few relationships will not be present, giving reliable results for occluded images.

The array of bits can be converted to a search word useable in database software such as MySQL. By comparing the search word with a database of objects and their associated search words the system can then come up with a shortlist of objects that could be in the image. The comparison is done by using a bitwise AND operation between the two words and counting how many bits match. This gives a list ranking the most likely object in the image. By selecting only the objects with a high percentage of hits, the list to search through can then be narrowed down to a handful of objects instead of the whole database, which will contain approximately 300,000 different views[1]. The more intensive search of the image can then be implemented to find which objects are actually present; this routine will compare the super features found in the image with those stored in the database for each object.

IV. EXPERIMENTAL

Tests were undertaken on one hundred images, mainly from an office environment, some of which are shown in 6. Two of these images (figure 5) were chosen as 'gold' images to reside in the database of objects. Many of the images contained one, or both, of the gold objects in various orientations, sizes and degrees of occlusion. All one hundred and two images were then processed, in the manner described in section III, and marked as either containing the object or not.

The number of matching bits required to indicate that the object was present was tuned so that the number of false negatives was minimal while keeping the number of false positives from ballooning out.

A. Results Data

Table 1 below summarises the results of the tests. For both of the gold images the results are recorded separately where the gold object is unobstructed and partially occluded (including where the back of the object is presented to the camera).

The Images row indicates the number of images that include the object and the Passed row indicates the number of those images where its presence was detected (true positives). The Rejected row indicates the number of images that were correctly rejected as not containing an object (true negatives).

The total number of images is one hundred and two including the gold images themselves.

<table>
<thead>
<tr>
<th>Images</th>
<th>Mug (unobstructed)</th>
<th>Mug (occluded)</th>
<th>Tape (unobstructed)</th>
<th>Tape (occluded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mug (unobstructed)</td>
<td>15</td>
<td>35</td>
<td>23</td>
<td>43</td>
</tr>
<tr>
<td>Mug (occluded)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tape (unobstructed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tape (occluded)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passed</td>
<td>15</td>
<td>35</td>
<td>19</td>
<td>34</td>
</tr>
<tr>
<td>Rejected</td>
<td>23</td>
<td></td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>False positives</td>
<td>44</td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>False negatives</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>
VI. DISCUSSION

Interpretation of the results must take into account that this is a screening process where the emphasis must be on few false negatives. False positives will be rejected by subsequent processes.

The method finds all of the images containing mugs, including those partially occluded. It finds a little over one false positive for every true positive, which the subsequent process can be expected to remove.

On the other hand, the ratio of false acceptances to true rejections was almost double, suggesting that there was a low level of discrimination. Certainly, with so few images not containing gold objects, the test was somewhat skewed and it remains to be seen whether this figure becomes more manageable with larger data sets.

Almost 90% of the tape measures in the unobstructed set were correctly accepted, falling to about 70% where partial occlusion was present. This is likely to be because the arc feature becomes inadequate when the tape is turned slightly side on, as the circle turns into an ellipse.

The discrimination of the method was more acceptable for the tape with the ratio of false acceptances to true rejections being about a half. As for the mug, however, this will need to drop significantly for larger data sets.

In summary, the initial results for the method are encouraging and, with continued work, will become an efficient way to create a shortlist of objects that could be found in an image.

VII. CONCLUSIONS

The method has been shown to recognise simple geometrical features from a fingerprint, and can combine these to capture the essentials of the image using only a few main features.

Geometric relationships between the main features can be found to provide a brief description of what is contained in the image.

These descriptions enable a search routine to generate the likelihood that a given object is in the image.

A test image set found 100% of the mugs, and 79% of the tape measures.

B. Future Improvements

The issue of low discrimination needs to be examined with larger, and more diverse, data sets with a view to increasing it.

While the three feature types currently used provide good results for some objects, on others they will be inadequate. The addition of an ellipse or partial ellipse feature will enable the curves to be represented with greater accuracy.

More feature relationships are also required to ensure search data is collected from all significant aspects of the set of super features.

VIII. REFERENCES