Sensing Objects for Artificial Intelligence

RORY C. FLEMMER AND HUUB H. C. BAKKER
Institute of Engineering and Technology, Massey University,
Private Bag 11 222, Palmerston North, New Zealand.
email: R.C.Flemmer@Massey.ac.nz

Abstract

In order to produce a workable system of vision for artificial intelligence, a method is proposed which is based on contours of constant brightness within a monochromatic image. These contours are expressed as their second derivative with respect to arc length. This makes them rotation invariant and allows objects to be learned and then recognized in a cluttered or occluded image using a data base such as MYSQL.

Keywords: Image analysis, machine vision, object recognition, contour, database, fingerprint.

1 Introduction

As we seek to create an artificial intelligence, we look to biological intelligences as a model. The overwhelming majority of terrestrial creatures use vision as their primary source of knowledge about the world. It can be argued that much of human intelligence evolved before speech [1] and it would be hard to argue that speech is a dominant mode of information acquisition for humans.

It follows that, for the development of an autonomous artificial animal, a robust technology for image processing must be developed. This is the impetus for the developments disclosed in this paper.

Image analysis is about half a century old and might be regarded as a mature science. Its literature is very large and some of it is very clever. It has tended to base itself on the notion that the ineluctable precursor to image analysis is the detection of edges. From a knowledge of the edges in an image, coupled with lighting, shadow and texture, an attempt can be made to construct the geometry which is defined by these features. This is a very hard task and has not been overwhelmingly successful.

Experiment has shown that unless kittens (raised in the dark) can investigate objects actively by perceptual hypothesis followed by confirmation and further hypotheses, they remain blind [2]. This suggests that the visual process in animals does not proceed upwards from the retina in ever-refined analysis. Further, the eye's saccadic movements [3] indicate clearly that the "picture" is built up on the basis of successive approximations.

In summary, using edges to build up the "picture" is extremely difficult and has not been successful except, industrially/medically where the visual situation is well defined and we know what we expect to see – in which case, the technology has far surpassed the eye in speed and acuity.

Edges as a Special Case of Contours

Consider a monochrome image, derived from standard hardware. This is typically 640 x 480, with 256 grey levels. Imagine the "picture" to lie flat and the grey levels to be plotted on a vertical axis. This results, conceptually, in a landscape where the brightest points (up to 255) are the peaks and the dark points lie in the valleys. Then construct contours of constant grey level (to sub-pixel accuracy) at intervals of say 10 grey levels over the whole height of the landscape.

In those regions where several different contours are almost collinear, we view this "cliff" as an edge. In the nature of the visual landscape, the number of contours so juxtaposed varies along the line of the edge and, as, in edge-following, we are constrained to follow the greatest height of the "cliff", our trajectory often veers off along a line which we do not consider to limn the object. Further, the edge-finding literature tends to assume that this topographical surface is smooth, extended, and differentiable. We know that this is not the case and that some features are defined by a small number of pixels.

If we consider the general case of a family of contours rather than its subset of those collinear contours called edges, we find that we have considerably more information in our data. It is observed, generally, that one or several contours do indeed limn each part of the object - as seems appropriate to our eye.

2 Literature

Because the image analysis literature is so vast, it is not profitable to summarize it here. Suffice it to say that we have not found any work which relates closely to the current scheme.

3 Implementation

An object was photographed, Figure (1) and Microsoft Visual Basic 6.0 was used to acquire the contours to sub-pixel accuracy.

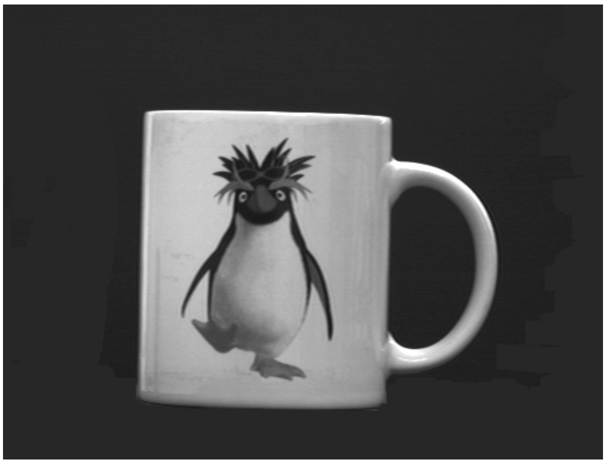

Figure 1: Initial Image

Contours were acquired at grey levels from 20 to 220 in increments of 10. This resulted in 211 contours defined by data doublets at each point. (X(i), Y(i)). The distance between contour points was approximately one pixel. In order to avoid ambiguity, the rubric “bright on the right” was used to determine which was the head and which the tail of each contour. As you proceed from the tail to the head, the pixels on your right are brighter than those on your left. Figure 2 shows a greyed-out image of the object with the 211 contours superimposed

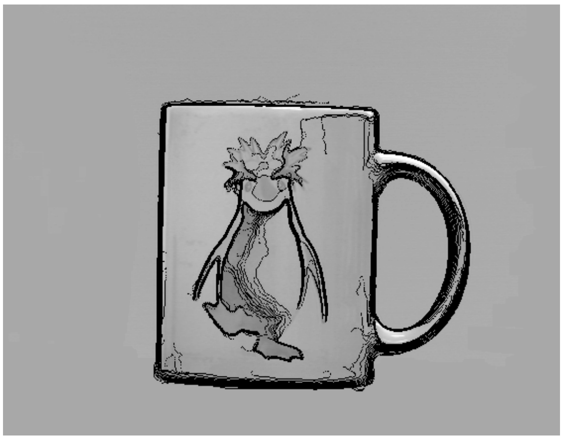

Figure 2: All Contours Superimposed

The contours were then winnowed to remove those which were shorter than and collinear with other contours. This produced a total of 39 contours, shown in fig. 3

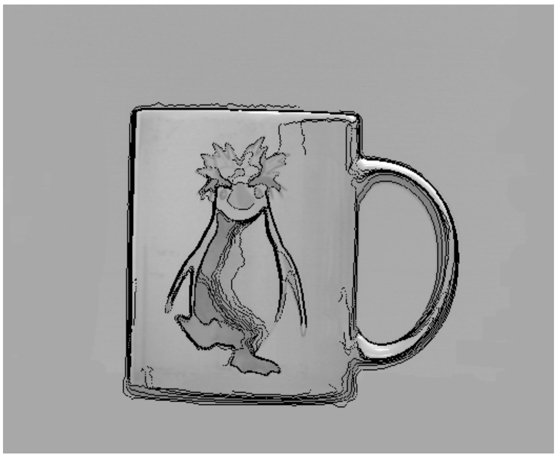

Figure 3: Selected Contours

The contours were then expressed as their second derivative with respect to arc length. I.e. they were plotted as the change in angle per pixel along the contour. This stratagem has the advantage that it makes the function invariant under rotation. It has the disadvantage that any error in the original contour is magnified. For the contour which circumscribes the cup, the raw data is plotted as the upper trace in figure 4. The points had an average positional error in X and Y of about 0.03 pixels as shown by the noise in the upper trace of fig 4 – about 1 degree.

This data was then transformed into frequency space using Haar wavelets [4], the difference coefficients of level 2 and 3 were set to zero and the data was then transformed back to yield the smoothed lower trace of figure 4. This trace represents that particular contour which circumscribes the mug.

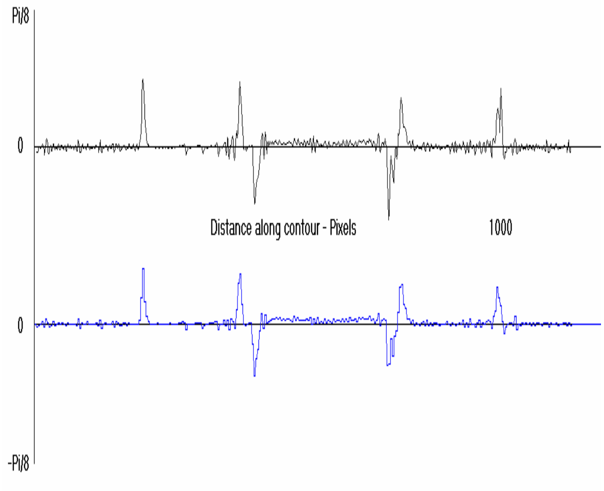

Figure 4: Change of Angle per pixel vs Contour Length

In interpreting this trace, note that the contour line starts just above the lower left corner of the mug and proceeds with bright on the right (clockwise). It proceeds upwards to the first corner, which is represented by the first positive spike in the curve. The second spike represents the bend down towards the handle. The third spike, which is negative, represents the contour turning outwards to follow the top of the handle. The slightly raised portion represents curving around the handle. The trace was caused to wrap around and recapitulate the start of the trace. This was done in case the contour started midway through a feature.

Software was written to demarcate these features into lobes. This trace has seven lobes. The area contained under a lobe corresponds to the angle through which the contour turns as it passes through the lobe. The first lobe has an area of 1.5 radians which represents the angle by which the contour rotates as it navigates from the almost vertical left side of the mug to the almost horizontal top. The fourth, long, lobe has an area of about 3.1 radians, which is also appropriate as it describes a rotation of about 180 degrees around the handle.

The fact that lobe area represents the angle of rotation of the contour makes data mining easier because, if we identify the lower trace in figure 4, merely by the area of the lobes, this is independent of image size. We call this smoothed trace the fingerprint of the contour.

As the orientation of the object changes, the fingerprint changes correspondingly. We observe that a change of 20 degrees about an axis lying within the paper causes a perceptible change in the fingerprint but it remains clearly recognizable through it transformation. This means that, provided that a specimen object is within 20 degrees of a known image, we can recognise it.

Learning a New Object

New objects are learned by placing them in a carefully lit, shadow-less, environment, taking a picture, changing the lighting, taking another picture and then extracting those contours which are multiply redundant, to end up with an intrinsic, light- independent data set. Each image must be photographed in each of 20 orientations. These should be defined by the directions of a ray from the centre of the object to the centre of each face of a regular icosahedron. To rotate from any such image to its neighbour requires a rotation of 36 degrees. Therefore, if all 20 sets of data are known, a sample image will be within 18 degrees of a known image. Therefore, with 20 sets of data, we can recognize the object and infer its orientation fairly precisely.

Methodology for Object Recognition

A large corpus of objects must be learned. Rough calculations show that 15 000 objects would cover almost all objects known by an educated person [5]. These objects would each be stored as 20 views and each view would have up to 100 fingerprints. These would be stored as lobe information as well as the compressed wavelet data. This data would not place unreasonable demands on a PC.

This form of storage allows the mature technology of databases to be invoked. It is our intention to use the capabilities of MYSQL.

When an unknown scene is presented to the system, it will reduce the image to a large number of fingerprints – perhaps 100. It will then run a database search to find those candidate fingerprints which have about the right number and size of lobes. It will then check on the order of lobes and the proportional distance between them. The short list can be transformed from the wavelet frequency domain back into the actual fingerprint and be directly compared with the sample fingerprint.

4 Discussion

We have proposed what seems a novel approach to image analysis. It has the drawback of being computationally quite intensive during the acquisition of the contours. It has the further drawback that it requires considerable data mining to pull out the matching gold image of the object view. But it has the significant advantage that it seems to work in obscure and occluded fields. Further, it gives a high degree of certainty in the match.

We view this as a preliminary excursion into the field and recognize that our technique will be considerably improved.

What is more exciting is that several other fields open immediately;

Firstly, we must deal with the question of universals: “Is this a cup or a mug or a jug?” We believe that our paradigm will lend itself naturally to homomorphic distortions as we go from cup to mug to jug. We might generalize and say that, in handling objects, we have thus far handled nouns. The question of universals introduces adjectives. From a small mug (cup) to a mug to a tall, big mug (jug)

Secondly, there is the question of articulation. Is a straight leg the same as a bent leg? This might be viewed as a generalization of the discussion of nouns in this paper to a discussion of prepositions as we consider conjoining objects.

Given that this calculus is aimed at an autonomous robot who must wake up, learn the objects of the world and then form opinions as to what happens, we will almost immediately get into verbs and adverbs.

References

[1] “The Brain and Language”, *Cambridge Encyclopeadia of Human Evolution, Cambridge University Press, 1996*

[2] Hein A, Held R and Gower EC “Development and segmentation of visually controlled movement by selective exposure during rearing”. *Journal of Comparative Physiological Psychology (2): 181& 1970*

[3] Luria, A.R. “The working Brain: An introduction to Neuropsychology.” *Penguin, Harmondsworth, England, 1983.*

[4] Mallat,S. “A theory of multiresolution signal decomposition: the wavelet representation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 11(7):674 – 693,1989.

[5]. Bryson, B. "The Mother Tongue: English and how it got that way" *William Morrow and Co., New York, 1990.*