Shape extraction in a hexagonal-image processing framework

Article · January 2000

3 authors:

Lee Middleton
University of Southampton
65 PUBLICATIONS 713 CITATIONS
SEE PROFILE

Jayanthi Sivaswamy
International Institute of Information Technology, Hyderabad
120 PUBLICATIONS 1,380 CITATIONS
SEE PROFILE

G. Coghill
University of Auckland
63 PUBLICATIONS 482 CITATIONS
SEE PROFILE

Shape Extraction in a Hexagonal-Image Processing Framework.

Lee Middleton, Jayanthi Sivaswamy, and George Coghill.
Department of Electrical and Electronic Engineering,
Univeristy of Auckland, Private Bag 92019,
Auckland, New Zealand.
{l.middleton,j.sivaswamy,g.coghill}@auckland.ac.nz

Abstract

For databases of images new sorts of query will have to be developed to make access to content efficient and convenient. Generally these queries will be based upon a number of broad classes which can be extracted from the image. This work looks at one such class, shape, and how to extract it from an image. The approach is based upon the use of an attention window to extract local information about the image. This information is integrated via a biologically inspired search mechanism to generate a description of the object's shape. Fundamental to the system's operation is a novel hexagonal representation scheme. The preliminary results of the system are promising showing proficiency at the shape extraction task with minimal error. As well, the representation generated has application to more general recognition problems.

1 Introduction

With the current rapid growth in computing technology and the Internet, more and more applications are arising which make use of pictorial information. Consequently, management and querying of this information has become a fundamental issue. It is thus of paramount importance that systems are developed that can facilitate this process to make it relatively painless to the user. Thus, there has been a large amount of research in the area of content-based image retrieval in the last few years [3, 8, 1]. A typical content-based image retrieval system performs a query via a number of generic classes which have been extracted from the image. These include, but are not limited to texture, colour, and shape.

The work presented here focuses on the issue of shape. The shape of an object can be thought to be defined by the its dominant edges. The approach taken is based on the human visual system. When the visual system explores a visual scene it does so by a succession of large jumps which are known as saccades [10]. These saccades serve to isolate points of interest in the image [17]. The next step is contour following [7] which involves following the dominant edges from these saccadic fixation points. By linking these together a simple object-centred description of the shape involved is generated. These steps inspired the proposed shape extraction scheme. The algorithm is built around a hexagonal lattice with successive saccades extracting a local hexagonally shaped attention window. It is envisaged that this work will eventually be applied to an image database of trademarks. In this work, only monochrome images are examined.

This paper will first present some general information about the proposed model. The various aspects of it will then be discussed in turn. Finally, the proposed model will be evaluated on some simple example images.

2 Shape extraction system

The proposed model is illustrated in figure 1. Initially, the data is preprocessed to find the critical points in the image. Once this has been performed, the first critical point is chosen as a starting point for the rest of the algorithm. At this point a hexagonal attention window is used to extract information about a region of the image. All subsequent processing is performed upon this. The next significant block is the feature extraction one which extracts two sorts of features (weight and orientation) from the attention window data. Next, the integration process takes the feature information and uses it to decide where to move the attention window next. As well, the current information is added to the overall description of the image. Each of these blocks will now be discussed in turn.

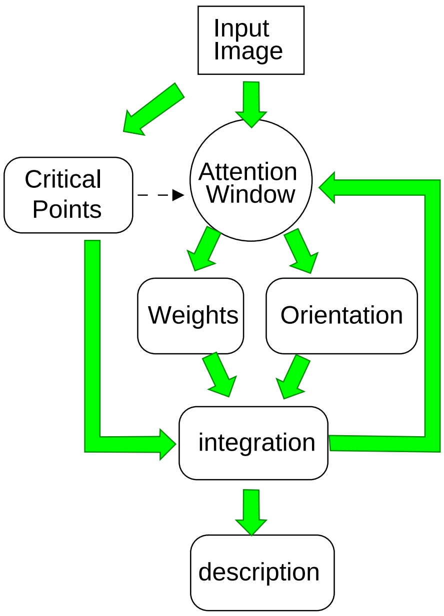


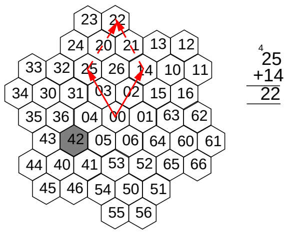
Figure 1: Block diagram of the proposed shape identification system.

2.1 Hexagonal Attention Window

Central to the attention window strategy is the use of hexagonal sampling. The attention window is created by hexagonally sampling about a current point of interest. Hexagonal sampling is not new and has been explored by many researchers [11, 14, 6]. Amongst the advantages cited are the ability to better represent curved structures [13]. For a practical Hexagonal-image processing (HIP) framework, such as has been employed here, [12] there are four key aspects : image conversion from square to hexagonal lattices, image addressing and storage, processing, and display. Each of these will now be addressed in turn as they are crucial to the operation of the attention window and thus to the rest of the proposed system.

2.1.1 Square- to Hexagonal-image conversion

To convert an image from a square lattice to a hexagonal lattice a process known as image re-sampling is employed [16]. Each point in the original image can be modelled as a square tile, with the same intensity as the pixel. The generation of the hexagonal points is achieved then by using an appropriate sampling lattice to draw sampled points from this space. The generators for this lattice are given by the basis vectors : $B = \{(1,0), (\frac{1}{2}, \frac{\sqrt{3}}{2})\}$. Note, that in order that a square space be fully represented, extra points are required. This is a direct result of the sample points being more densely packed than in the



+	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	63	15	2	0	6	64
2	2	15	14	26	3	0	1
3	3	2	26	25	31	4	0
4	4	0	3	31	36	42	5
5	5	6	0	4	42	41	53
6	6	64	1	0	5	53	52

(b)

Figure 2: (a) The hexagonal image structure with indices (b) Balanced ternary addition

square sampled case. Of course, if specialised sampling hardware were available this would not present a problem and the sampled data would more closely represent the data [15].

2.1.2 Addressing and Storage Issues

It is difficult to address all the points defined by the hexagonal lattice defined in section 2.1.1. One approach is to use three coordinates [5] each aligned to one of the axes of symmetry of the hexagon. Another is to use Cartesian coordinates and employ a skewed y (or x) axis [4]. For this work an alternate approach was examined based on the symmetry of the hexagon.

Consider a single hexagon to be a tile at layer 0. This can then be surrounded moving anti-clockwise by a further 6 hexagons. This new structure forms a tile at layer 1, with the new tile being the super-tile of layer 0. Similarly, this process can be extended to any number of layers. An example of layer 2 is illustrated in figure 2. The number of hexagons that are contained in a given layer L super-tile are 7^L. Each hexagon can then be numbered uniquely as a sequence of numbers each one giving its position in a given tile. The highlighted hexagon in figure 2a is in the fourth tile in layer 2 and the second tile at layer 1. Thus, all the hexagons in a layer L super-tile can be addressed uniquely by a L-digit base 7 number. Note, that this digit also encodes the spatial location of the hexagon. Finally, as each location is given by a unique number, only one coordinate is required to address any point. This means that storing an image only requires a vector. This coordinate system is used extensively throughout the rest of the work described here.

2.1.3 Image Processing operations

Many common image processing operations require the neighbourhood of a pixel to be known. The neighbourhood of a hexagonal pixel is defined by walking, at most, a given distance from a central pixel. Thus, the first hexagonal neighbourhood N_1 contains 6 pixels, and the next neighbourhood N_2 contains 19 pixels. There is another permissible neighbourhood that is allowed by the data structure. This is defined by aggregations of tiles at a given level. This neighbourhood is defined as $N_{g(i)}$ and due to the nature of the hexagonal lattice, it is approximately circular in shape.

To find the neighbours of a given point requires the use of balanced ternary arithmetic [9] which is analogous to vector addition (see the example in figure 2). The first neighbourhood, N_1, is simply found by balanced ternary addition of the numbers 1 to 6 to the central pixel. Neighbourhoods N_2 and $N_{g(2)}$ are also found by balanced ternary addition.

A simple operation that uses neighbourhoods is convolution with a mask. To perform the convolution for a hexagonal-image, only two nested loops are required where a standard square-image requires four. Since, all points in a given hexagonal-image have index less than 7^L (L being the number of layers) the boundary of the image is easy to compute. Finally, another task that is simpler to perform is the derivation of an image pyramid. This can be performed by starting at the origin and averaging all the pixels in the $N_{g(i)}$ neighbourhoods. Successive layers of the pyramid may be found by increasing the value of i.

2.1.4 Display of Hexagonal-images

The first issue that needs to be addressed when considering displaying the hexagonal-image is conversion from a unique index to a screen location. This process is straightforward and requires the use of polar coordinates. As the index can be written $d_{n-1} \cdots d_0$, each number ,d_i, causes a change in the coordinates as shown below :

$$r = \begin{cases} 1, & i = 0 \\ (\sqrt{7})^i, & i > 0 \end{cases} , \theta = i.\tan^{-1}(\frac{\sqrt{3}}{2}) + (d_i - 1)\frac{\pi}{3}$$

The actual (x, y) coordinates just require these offsets to be summed for a given point. After a Cartesian coordinate is computed, the image has to be displayed on the screen. This is performed using a suitable aggregation of pixels that, when put together, looks hexagonal in shape. This is known as a hyper-pixel and takes advantage of the oblique effect in human vision [2].

2.2 Critical Point Extraction

In biological vision systems the image is examined via a series of points of interest. Taking this as a starting point the critical point extraction subsystem

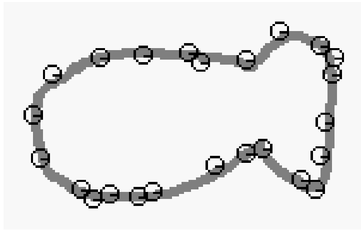


Figure 3: The critical points extracted for a typical image.

was developed. The purpose of the critical point extraction subsystem is twofold. The first is to provide a series of potential points for the system to examine, and the second is to provide alternate points to examine when the algorithm finishes at this current point of interest. To find candidate points every point in the image is examined through an attention window (as described in section 2.1). The critical points are defined as points where the weight (see section 2.3.1) in a given region exceeds a particular threshold. An example of a simple candidate image and the discovered critical points are illustrated in figure 3. The coordinates of the critical points are stored using the scheme described in section 2.1.2.

2.3 Feature Extraction

There are two sorts of feature extraction that are carried out on the image data. These are the discovery of weight and dominant orientation. They are both computed in a hierarchical fashion exploiting the inherently pyramidal fashion of the hexagonal data structure used in the attention window.

2.3.1 Weight Extraction

The weight of the image refers to the sum of the gray levels which comprise it. For the hexagonal structure that is used for the attention window the weight can be computed for several resolutions easily. For instance, for a 2-layer system as shown in figure 2 the weight can be computed by summing the cells with the same first digit. This process can be repeated to produce the weight at the lowest possible resolution (1 pixel). Thus, for an attention window with 7^L pixels there will be $\frac{7^L-1}{6}$ distinct weight values. This is the sum of the number of weights for each layer in the image. For example, the 2-layer system has 8 weights. This involves 1 weight for the first layer (entire image) and 7 for the second layer.

2.3.2 Orientation Extraction

Extraction of the orientation image from the attention window is a two step process. Firstly, the image is filtered using the Sobel edge operators [12]. For each group of seven points the dominant direction is assigned to be the direction at the central pixel (as given using the two distinct Sobel masks). By reducing the order of the image (as described for weight extraction) this process can be repeated again. As for the weight this will reduce an attention window of size 7^L to $\frac{7^L-1}{6}$ distinct values. As the representation described in section 2.1.2 has vector like properties, balanced ternary numbers are stored as the direction. For practical reasons these are scaled so as to point into the nearest group of seven pixels. At the lowest resolution of the pyramid extracted via this algorithm is the dominant direction of the entire attention window.

2.4 Integration

Integration takes the information from the feature extraction stage and uses it to decide where to move next. The weight gives the direction where the most pixels are concentrated. If this is in agreement with the orientation for the current attention window then this is a good candidate for the direction of movement. If not, then a higher resolution is examined. From this top-down process a list of candidates for possible movement are generated. These are ordered from low to high resolution. Once a candidate for the movement is found, it is then examined for three things. Firstly, it should not cause a reverse in the direction (ie travel back to where it came from last iteration) as this causes redundancy. If this is the case then the candidate is removed and the next one is examined. Secondly, if the movement passes sufficiently close to a critical point then this becomes the next point. The movement is adjusted to be the correct value by just subtracting the coordinates. Finally, the movement is examined to see if it closes a loop. A closed loop is considered to mean that the shape of the object has been found and that the algorithm may end.

Whilst the algorithm is progressing, each iteration produces a new point that is added to the list that corresponds to the image's shape. Each successive point is stored as a balanced ternary number relative to the image's centre. This representation of the shape is very compact. Also, as movements are also stored using a balanced ternary number then the new point to focus attention is found by simply adding the movement to the current position.

Shape	n	$\overline{\Delta}$	σ^2
square	35	0.29	0.46
fish	34	0.29	0.35

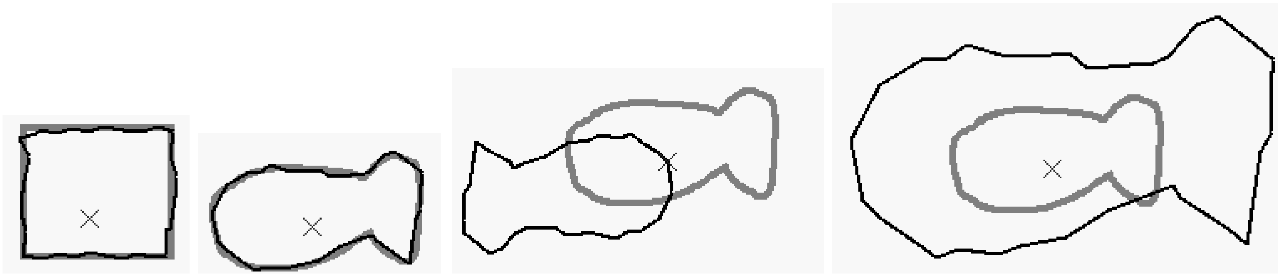
Table 1: Comparison of the number of incorrect points.

3 Results

The algorithm outlined in section 2 was evaluated using simple shapes. The input image size was 256×256 pixels in size and contained binary data only. The attention window employed was a 3 level hexagonal structure (343 hexagonal pixels). Some representative results are illustrated in figures 4a and 4b. In each image the original image is shown (in grey) along with the shape that is extracted (in black). Note that the original image is artificially lightened to aid in illustrating the algorithms performance. The results illustrate performance with two distinctly different shapes. The square is made predominantly of straight lines and the fish is made of a mixture of straight and curved lines.

Examination of both of the images and the resulting derived shape show that the algorithm performs well. In most cases the derived shape lies within the original shape. For the square the top left corner is poorly represented but the result is still good enough to distinguish the shape. The fish shape is much better represented with most significant aspects of its shape being identified. In both shapes there were a few points which were outside the original image. An analysis of these points is given in table 1. In both cases the number of points, n, is practically the same, as is the mean difference, $\overline{\Delta}$. The mean squared error, σ^2, is different between the two shapes. This is due to the errors in the square case being larger. However, in both cases the difference is acceptable.

For a generalised recognition scheme a representation that is view (and hence size and rotation) invariant must be employed. This can be achieved via a representation that is relative to the objects centre and can be easily manipulated. A successful database shape search strategy has thus to include shape manipulation. These are achieved in the proposed scheme via the positional nature of balanced ternary arithmetic. In figures 4c and 4d), transformations on the shape are illustrated. In figure 4c, the data has a constant offset added to it and then is multiplied by 4 (rotation by 180^o). Figure 4d, illustrates how the data can be easily scaled by a simple multiplication. In this case the data was multiplied by 63 (a scale factor of 2).



(a) (b) (c) (d)

Figure 4: Results of the shape determination algorithm (a)a square (b)a fish (c)the fish rotated (d) the fish scaled

4 Concluding Remarks

The work presented here is an attention driven exploration strategy that has been applied to the problem of shape determination of monochrome images. This task was accomplished using the HIP framework and the pyramidal structure contained within it. The system exhibits novelty based upon the hexagonal coordinate system. As a result of this coordinate system simple geometric transformations can be performed upon the extracted shape. In the future work will concentrate upon one of two areas. The first is in cleaning up the generated data to provide a more minimal dataset to represent the shape. This can be achieved by examination of the data and examining the change between successive pairs of points. Only large changes will be kept. The second area to examine is to simplify the rules needed for contour following. This could be achieved by the use of a neural network architecture or a fuzzy logic system. An additional improvement upon the work here is in automatically recognising the lines width. This can be achieved by examining the weights at each of the points that make up the shape. Larger weights will indicate thicker lines. Generally, it is considered that the work here is a promising beginning. The strength of the approach lies in the use of the novel object-centred representation scheme for the shape. This representation scheme could be applied to a wider class of recognition problems. The manipulation aspect gives it a significant advantage over representations based on primitives (such as geons or generalised cones) which can be computationally expensive.

References

[1] J. P. Eakins and M. E. Graham. Similarity Retrieval of Trademark Images. *IEEE Multimedia*, 52:53–63, April-June 1998.

[2] R. M. Gray, P. C. Cosman, and K. L. Oehler. *Digital Images and Human Vision*, chapter 4, pages 35–52. MIT Press, 1993.

[3] V. Gudivada and R. V.V. Content-based image retrieval systems. *IEEE Computer*, 28(9):18–22, 1995.

[4] N. P. Hartman and S. L. Tanimoto. A Hexagonal Pyramid data structure for Image Processing. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-14(2):247–256, Mar/Apr 1984.

[5] I. Her. Geometric Transforms on the Hexagonal Grid. *IEEE Transactions on Image Processing*, 4(9):1213–1222, September 1995.

[6] I. Her and C.-T. Yuan. Resampling on a Pseudo-hexagonal Grid. *CVGIP : Graphical Models and Image Processing*, 56(4):336–347, July 1994.

[7] H. S. Hoffman. *Vision and the Art of Drawing.* Prentice Hall (Englewood Cliffs, N.J.), 1989.

[8] Y. S. Kim and W. Y. Kim. Content-based trademark retrieval system using a visually salient feature. *Image and Vision Computing*, 16:931–939, 1998.

[9] D. E. Knuth. *The Art of Computer Programming : Seminumerical Algorithms*, volume 2. Addison Wesley, 1969.

[10] S. M. Kosslyn and O. Koenig. *Wet Mind : The New Cognitive Neuroscience.* The Free Press (New York), 1995.

[11] R. M. Mersereau. The processing of Hexagonally Sampled Two-Dimensional Signals. *Proceedings of the IEEE*, 67(6):930–949, June 1979.

[12] L. Middleton and J. Sivaswamy. Edge Detection in a Hexagonal-image Processing Framework. In D. Pairman and H. North, editors, *Image and Vision Computing New Zealand 1999*, pages 217–222. Landcare Research, 1999.

[13] I. Overington. *Computer Vision : a unified, biologically-inspired approach*. Elsevier Science Publishing Company, 1992.

[14] R. C. Staunton. Hexagonal image sampling : a practical proposition. *Proc. SPIE*, 1008:23–27, 1989.

[15] D. Whitehouse and M. Phillips. Sampling in a two-dimensional plane. *Journal of Physics A : Mathematical and General*, 18:2465–2477, 1985.

[16] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, 1990.

[17] A. L. Yarbus. *Eye Movements and Vision*. Plenum Press (New York), 1967.