Space-Variant Optical Character Recognition

Ping-Wen Ong †‡ Richard S. Wallace † Eric L. Schwartz †‡§*

†Vision Applications
611 Broadway #714
New York, NY 10012

‡Courant Institute of
Mathematical Sciences
New York University
New York, NY 10012

§Computational
Neuroscience Lab
NYU Medical Center
New York, NY 10016

Abstract

An active vision system with a multiresolution sensor has the advantages of a wide visual field, high central resolution and low data rate. We use a special space-variant sensor based on complex log mapping to do character recognition in real scenes. Here, we describe experiments with feature designs based on several statistical OCR methods: characteristic loci, partition and projection. We have also examine several rotation-invariant feature designs based on the partition and projection methods. After studying the problem of recognizing strings in real scenes, we derived a rule based approach to solve the attention problem in the context of pattern recognition. Finally, we describe an active vision system which successfully recognizes characters on a license plate in a real scene.

1 Space-variant active vision

The application of optical character recognition (OCR) to printed characters has received a great deal of attention during the past three decades [4,5], and represents an area of increasing importance for application of pattern recognition and image processing. OCR in natural scenes, as opposed to document images, is an area which has a natural requirement for wide-area, high resolution imaging. In a typical post-office application, the location of a package and the address block on the package might vary over a relatively large workspace, while a high resolution digitization of the address characters is necessary. A common approach to this problem is to acquire the entire scene at high resolution with a conventional video sensor, and then process the resulting very large number of pixels. In the present paper, we consider another approach, using recently developed space-variant sensor methods that allow the coverage of a large work space, but which provide a high resolution area in a small central (foveal) region whose resolution is a monotonically decreasing function of distance from the center of the sensor.

Space-variant sensing is partially motivated by the visual architecture of the higher vertebrates, including the human, which make use of strongly space-variant (retinal) architectures [8]. Recently, several research groups have developed space-variant sensors based on CCD , CMOS and firmware (see [2] for citations). Because the particular space-variant geometry used in these papers is motivated by the logarithmic structure of the human visual system, we will use the term "logmap" to describe this architecture.

The use of a "foveating" sensor requires the application of active vision techniques: the sensor must be able to actively move its center of fixation in order to optimally use its architecture.

Space-variant systems have the same field of view and central resolution as uniform sensor systems. But because they have a greatly reduced number of pixels, the overall processing requirements are reduced correspondingly. This reduction has been estimated to be roughly 10,000:1 for the space-variant human visual system [7], and is in the range of one to three orders of magnitude for achievable machine vision system. Recently, we have constructed a very low cost prototype active vision system based on a space-variant sensor [2].

In the present paper, we investigate several algorithmic problems associated with performing OCR using a space-variant active vision system. Our benchmark application is to read license plates on moving cars in near real time (e.g. five frames/second). We have chosen this benchmark because it extends a well understood pattern recognition technique (OCR) to a space-variant context, along with some of the typical motion control, target acquisition, attentional and im-

*This work supported in part by DARPA Contract #N00014-90-C-0049.

0-8186-2915-0/92 $3.00 © 1992 IEEE

age processing problems which must be solved in order to apply space-variant imaging to pattern recognition.

2 Feature Design

We have investigated three OCR methods which have been developed for uniform resolution sensors. This choice of methods for initial investigation was based on the use of statistically defined features, good invariance properties, and the use of a binary image. All of them are translation and scale invariant. These methods are:

The *characteristic loci* [3,9] method counts the number of white-to-black crossings along lines extending from each white background pixel. Several line directions should be considered. We use 0, 45, 90 and 135 degrees lines in our experiments. We oversampled the log-space by a factor of two in order to find the best matching directions for each pixel in the log space.

The *partition* method uses the distribution of the pixels in the bounding box of the character image to form the feature vector. A two dimensional function can be used to assign weights to all the positions around one partition for smoothing out all the discontinuities on the boundary of any partition.

The *projection* method uses the vertical and horizontal projections of a character image as a basis to find useful feature values [1]. One way to derive a set of useful feature values from character projections is by using the Walsh function decomposition [6]. Another projection method is to simply use the normalized raw projection data as the feature vector.

The method of *Principal Axes* allows us to process rotated character images. The principle axes of the character determine a rotated bounding box. By rotating the character, or the derived features into canonical position, the previous non-rotationally invariant methods may then be applied.

3 Experiments

We tested the three methods outlined above on an alpha-numeric character set (0 to 9 and A to Z in Helvetica font). Since the space-variant sensor is not shift-invariant, the stability of the method (i.e. no performance decrement for small translational shifts) is important, and requires a training approach that is significantly different from traditional OCR. Using one image for every character from 0 to 9 and A to Z, we used 49 different fixation points to construct 49 different character images in the log space. Generalization to other fixation points was tested with a set of 2500 different fixation points. The size and rotation parameters were varied over three sets of different size character images for the normal methods and three sets of different orientation character images for the rotation invariant methods. Two different classifier techniques were used: k-nearest neighbor and maximum likelihood.

4 Results

Results for the accuracy and speed of recognition is summarized in following table.

	methods	accuracy (%)	speed (ch./sec.)
vertical characters only	char. loci.	90	2
	partition	98	2 – 14
	projection	98	25
rotation invariant	partition	90	2 – 9
	projection	95	12

If a feature design is stable, the feature vector derived from the same image will remain more or less the same despite different positions of the fixation points. One way to evaluate stability is to find the relationship between the distance of two fixation points and the distance of the feature vectors derived from these fixation points. The rotation-invariant raw-projection method is the stablest and the Walsh method is the least stable. All the rotation invariant methods are quite stable and so are the partition methods. Non-rotational projection methods (raw and the Walsh) are the most unstable ones.

5 Reliability Analysis

Our results show that the recognition rate depends on the size of the input characters, even though we designed the features be be size invariant. Further investigation reveals that there is another factor affecting the recognition rate: the position of the fixation point relative to the character image.

From the layout of a logmap sensor, it is apparent that a character image is most recognizable when it is close to the center of the sensor. If we plot the probability for recognizing a character as a function of the position of the fixation point, we find that there is a flat area (with probability equal to one) around the the center of the character images (Figure 1). These regions, called the *reliable area*, represent the positions where we can point our sensor and recognize this letter correctly. Character images of different sizes have different sized reliable areas. The reliable area depends on both the size of the character image and the geometry of the sensor. In general, smaller character images have smaller reliable areas. Therefore, the reason smaller characters have a lower recognition rate in our experiments is that for all characters, we foveate at the

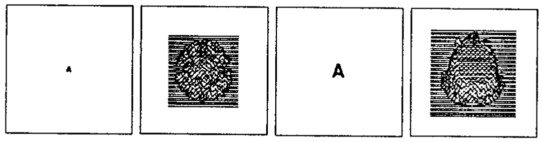

Figure 1: Two images of character 'A' and the corresponding recognition-reliability plots.

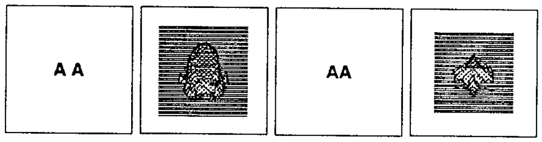

Figure 2: Two images of two 'A' characters and the corresponding recognition-reliability plots. Left one has two well-separated 'A's and right one has two close 'A's.

same number and position of places. For small characters, many of these points fall outside the reliable area yielding a low recognition rate.

Further investigation showed that recognizing different size characters using only the center of the character as a fixation point brings the recognition rate much closer to that of previous results. Therefore, to recognize any individual character images, the center of the character is the best place to point the sensor.

The definition of reliable area can be extended to multiple character strings. For example, a reliable area for an image of two adjacent characters is the region we can point the sensor to and recognize both characters correctly. This composite reliable area depends on the reliable areas and the *separable area* of these two characters. The separable area is the area we can point the sensor to and separate these two characters. This phenomenon makes the separable area an important factor in computing the composite reliable area. When two characters are too close to each other, the composite reliable area falls between these two characters. Therefore, it is impossible to point the sensor to the center of either one of the characters and recognize both of them correctly at the same time. The only good place to point the sensor is near the midpoint between these two characters. Figure 2 shows an example of this situation.

6 Attentional Algorithm to find Strings in static scenes

The use of space-variant sensing can provides orders of magnitude reduction in the number of pixels, for a given ratio of work-space to maximum acuity [7].

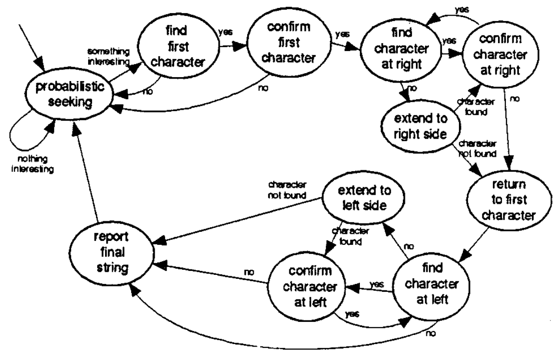

Figure 3: Transition diagram for attentional algorithm.

In order to exploit this advantage, it is crucial to provide an attentional algorithm to allow the system to rapidly locate and fixate "interesting" regions in the scene. We have implemented a finite state machine which incorporates a number of heuristic attentional algorithms. These are based on the following attentional qualities:

Maximum likelihood: The best place to "look next" is where a character is likely to be detected. For the current scene, a segmentation based on binarization and connected components [10] is performed, and all scene locations are scored for character likelihood. The sensor is then move to the center of the maximum likelihood neighborhood that is closest to the current fixation point.

Separation: If we detect that there are characters that are close to each other and will merge boundaries if we move to the center of any one of them, we move the fixation point to the place where the sensor can best separate the characters.

Probabilistic seeking: If there are no suitable character candidate visible, we use sub-features of the character to attempt to locate characters. For example, since characters often have high contrast with the background, edges can be used as an indication of the possible existence of a character.

Line searching: If a high likelihood candidate is found, search along the line computed by using the orientations of the bounding box of this character.

Using these rules, we have derived a very simple attentional algorithm for finding a string in a scene. Figure 3 depicts this algorithm in terms of a transition diagram.

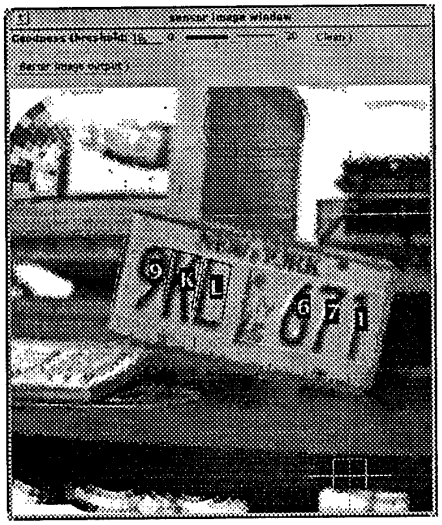

Figure 4: Snapshot of the program's output. This program blends multiple logmap images to form a composite image and shows the recognition results for a scene with rotated characters.

7 A System to Read Strings in Natural Scenes

We have built an active vision system to search for and read strings in natural scenes. The system setup includes a SUN SPARCstation, a CCD camera, a frame grabber and a pan-tilt actuator. Two programs run concurrently. One program controls the motor, image input and transforms the video image into a logmap image. The second program uses Unix sockets to communicate with the sensor program and get the logmap image for the recognition task. We have tested this system with real license plates placed in front the camera and let the program control the motor to find the characters on the license plates. Figure 4 shows an example of the results of this string reading system.

8 Conclusions

The work described in this paper suggests the following four major conclusions:

1. Space-variant image processing provides a means of building low-cost hardware for OCR in contexts which have relatively unconstrained work-space size. We have described the hardware architecture of a prototype space-variant active vision system in an accompanying report [2], while the present report describes the details of pattern recognition in this task.

2. Some existing OCR methods can be adapted to space-variant sensing with good performance.

3. We have constructed a finite-state machine which incorporates a small set of visual attentional heuristics which provides good performance for OCR.

4. The adaptation of uniform resolution OCR to space-variant sensing is a good test-bed for considering more general pattern recognition applications using space-variant active vision.

References

[1] Adnan Amin and Grald Nasini. Machine recognition of multi font printed Arabic texts. In *8th ICPR*, pages 392–395, 1986.

[2] Benjamin B. Bederson, Richard S. Wallace, and Eric L. Schwartz. A miniaturized active vision system. In *11th International Conference on Pattern Recognition*, 1992.

[3] H.A. Glucksman. Classification of mixed-font alphabetics by characteristic loci. In *Digest of 1st Ann. IEEE Computer Conf*, pages 138–141, 1967.

[4] J. Mantas. An overview of character recognition methodologies. *Pattern Recognition*, 19(6):425–430, 1986.

[5] George Nagy. Optical character recognition: theory and practice. In *Handbook of statistics*, volume 2, pages 621–649. Elsevier North-Holland, 1982.

[6] A. Rajavelu, M.T. Musavi, and M.V. Shirvaikar. A neural network approach to character recognition. *Neural Networks*, 2:387–393, 1989.

[7] A. S. Rojer and E. L. Schwartz. Design considerations for a space-variant visual sensor with complex-logarithmic geometry. *10th ICPR, Vol. 2*, pages 278–285, 1990.

[8] E. L. Schwartz. Spatial mapping in primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.

[9] Ching Y. Suen. The role of multi-directional loci and clustering in reliable recognition of characters. In *6th ICPR*, pages 1020–1022, 1982.

[10] Richard S. Wallace, Ping-Wen Ong, Benjamin B. Bederson, and Eric L. Schwartz. Space-variant image processing. Technical Report 256, New York University, Computer Science, Robotics Research, 1991.