Department of Electrical and Computer Systems Engineering

Technical Report MECSE-9-2003

A Scale Invariant Object Detector: An Implementation for License Plate Detection

M. Gubara and D. Suter

MONASH UNIVERSITY

A scale invariant Object Detector: An Implementation for License Plate Detection

Mohamed Gubara and David Suter
Department of Electrical and Computer Systems Engineering
Monash University, Clayton Vic. 3800, Australia
{osman.gobara; d.suter }@eng.monash.edu.au

Abstract

In this paper we present a new method for 2D object detection. This method uses the Hough transform as a robust platform for edge detection and extends it in the scale space. This extension enables the Hough transform to detect the scale changes within an image. The method starts with a linear scale-space representation of the image. At each level of this space edges are extracted through a normalized derivative function and accumulated in the Hough space. Higher features are mapped from the Hough space into the spatial space and then tagged with a normalized descriptor. The feature with the maximum descriptor over scale is then selected as the object of interest

The new approach is applied to the classical problem of the license plate detection. This problem is characterized by a unique solution that can be easily verified by the human observation.

1. Introduction

Object detection is an old problem that remains essentially unsolved. In this paper we present a general approach for texture-based object detection. The algorithm starts by detecting interest points at different levels of the scale space.

An interest point is any point in the image that is characterized by distinctive neighboring features. This includes L-corners, T-junctions, Y-junctions and textured areas. In [1] the interest point detectors are categorized into three main groups: Contour-based methods, Intensity-based methods and Parametric-model based methods. The Contour-based methods define interest points either at the intersections of grouped line segments or at the maximum curvature of approximated contours. Intensity-based methods define interest points through the illumination distribution of the neighborhood. In most cases these algorithms are based on the second moment matrix, which is a mathematical measure for the neighboring gradient distribution. Parametric-based methods define interest points at regions that fit a predefined analytical intensity model.

Interest point detectors are extremely sensitive to image scale variations: they have different spatial locations at different image scales. This problem was addressed in the recent work of [2, 3, 4 and 5]. Schmid and Mohr [2] use the Harris detector to extract interest points in the form of circular patches around corners. These patches are defined to be rotationally invariant. The work of [3 and 4] extends this idea to true scale invariance by defining a normalized descriptor for each interest point over the scale-space. Lowe [3] defines the descriptor as the maximum output of a difference of Gaussian filters. Lindeberg [4] defines the descriptor as the maxima of a scale-normalized Laplacian as well as the magnitude of a normalized gradient. Mikolajczyk and Schmid [5] use the descriptor of Lindeberg to define the so-called Harris-Laplacian detector. In this detector a scale-space representation is built for the Harris function of eq.1 by the second moment matrix $C(z,s)$, where z is the spatial coordinates of the interest point, L_x and L_y the x and y gradients of z respectively, G the Guassian kernel and s the scale of G. At each level s of the scale-space, candidate points with the local spatial maximum of the Harris function are selected. Interest points are then defined as the candidates with the local-scale maximum of the Laplacian function.

$$\text{Harris function} = \det(C) - \alpha \text{trace}^2(C) \qquad (1)$$

$$\textit{where} \quad C(z,s) = s^2 G(z,s) * \begin{bmatrix} L_x^2(z,s) & L_x L_y(z,s) \\ L_x L_y(z,s) & L_y^2(z,s) \end{bmatrix}$$

Tuytelaars and Van Gool [7] define scale-invariant regions in the form of parallelograms that start from a corner and its nearby edges. Each edge is extended to a point having a maximum photometric quantity. The corner together with the two extended edges form a

parallelogram. The extracted parallelograms are invariant to scale and geometric transformations but they are extremely dependent on the initially extracted corners. To solve this problem Tuytelaars and Van Gool [8] propose a complementary approach. In this approach a non-maximum suppression algorithm extracts interest points with local maximum in the image intensity. These interest points are characterized by a non-planar neighborhood and any slight change in their spatial location does not affect the constructed regions. Next the image intensities along a finite set of rays originating from each interest point are detected and the point with the maximum illumination difference along each ray is then selected. The set of selected points for each interest point are then linked to enclose an affinally invariant region. Finally an ellipse with similar shape, up to the second moment, replaces the detected region. In all the above techniques the size of detected regions are small, which makes them robust against occlusion and both background and viewpoint changes. Therefore these techniques are suitable for applications that include matching such as indexing and stereo correspondences.

In this paper we present an approach that detects textured objects at different scales. The success of this approach is based on [1] a robust implementation of the Hough transform and [2] a normalized derivative descriptor that characterizes textured regions over both spatial and scale spaces. In this approach the image is initially represented by a linear scale-space. At each level of the scale-space; (a) interest points are detected by a normalized derivative function (i.e. value under Σ of eq.6), accumulated in the Hough-space and then mapped to the spatial space as line segments, (b) intersections between line segments are built up into rectangular regions and each region is identified by a normalized descriptor. Finally the region with the maximum descriptor over both scale and spatial spaces is selected as the target object.

Overview. This paper is organized as follows. In section 2 we discuss different implementations for the scale-space. The Hough-space is summarized in section 3. The proposed algorithm for license plate detection is described in section 4 and finally experimental results are presented in section 5.

2. Scale-space

An image represented at the finest scale usually contains a substantial amount of spurious information. To identify the coarser details of the image it needs to be smoothed with a neighborhood operator or kernel where the increased amount of smoothing suppresses structures with characteristic length less than the kernel size. The Gaussian distribution as described in the literature of Lindeberg and Romeny [11] is considered the most appropriate neighborhood operator. The stack of images as a function of increasing inner scale is known as the linear scale-space.

2.1 Space Representation

A linear scale-space is defined by the solution of the following diffusion equation;

$$\frac{\partial L(z,s)}{\partial s} = \frac{1}{2}\nabla^2 L(z,s) = \frac{\partial_{xx} L(z,s) + \partial_{yy} L(z,s)}{2} \qquad (2)$$

with the initial condition that $L(z, 0)$ is equal to the original image $I(z)$, ∇^2 is the laplacian kernel and z is the spatial coordinates of the interest point. Equivalently a linear scale-space can be defined by the convolution with the Guassian kernel of eq.3.

$$G(z,s) = \frac{1}{2\pi\sqrt{s}} e^{-z^2/2s} \qquad (3)$$

To reduce the amount of smoothing around edges Perona and Malik [6] propose the use of anisotropic diffusion as a generalization of the linear scale space representation.

$$\frac{\partial L(z,s)}{\partial s} = \frac{1}{2}\nabla^2\left(h(z,s)\ \nabla L(z,s)\right) \qquad (4)$$

where $h(z, s)$ is defined to be dependant on the image gradient. A number of possible solutions of $h(z, s)$ are presented by eq.5 where k defines the range of gradients in an image and thus controls the amount of smoothing at point z.

$$h(z,s) = e^{-\frac{|\nabla L(z,s)|}{K}},\ \ or\ \ h(z,s) = \frac{1}{1+\frac{|\nabla L(z,s)|}{k}} \qquad (5)$$

2.2 Space Dimensionality

The dimension of the scale-space is represented by an initial scale σ_0, final scale σ_f and a number of levels n. It can also be represented by σ_o, a factor k of scale change between successive levels and the number n. In our application we have adopted the exponential scheme of [9] for gradual incremental changes in the successive levels of the scale. This scheme is defined by $s_m = k^m s_0$ where k is the scale change factor, m the number of levels, s_0 the initial scale and s_m the scale at level m.

3. Hough-space

The Hough space is usually used to map the interest points to higher featured shapes such as line segments, circles and ellipses. In our application we are interested in detecting straight lines and therefore the two dimensions of the Hough space are the perpendicular distance r from the origin and the orientation θ of r with respect to the x-axis. As we intend to impose a new scale-dimension to this space, the resolution of r and θ becomes a critical constraint on the feasibility of the new approach.

The resolution of r depends on the size of the image and the area, around the origin, within which we intend to detect the interest points. In our application the license plate can be located anywhere within the image and therefore the image size remains an upper bound on the range of the r-dimension. The shape of the license plate can be modeled as an upright rectangle. This assumption requires few bins in the θ-dimension and therefore it becomes quite feasible to add a new dimension to the Hough space.

A major disadvantage in the Hough transform is that it cannot identify the right scale of a detected interest point. This leads in most cases to the accumulation of interest points of different scales in a single (θ, r) location within the Hough space, which in turn results in the detection of spurious features that affects the localization accuracy of the plate detector. For the Hough transform to perform reliably interest points of different scales must be accumulated at different locations. This suggests we add a new scale-dimension to the Hough transform. Taking a linear scale-space representation of the image, at each level of this space interest points with values under Σ in eq.6 greater than a predefined threshold G are accumulated in a separate Hough array.

For an accurate localization of the license plate the value of G must be defined to be proportional to the image illumination. In order to detect the value of G automatically a fourth dimension was added to the Hough transform. The new dimension is represented by a range of n thresholds where at each threshold step new interest points are detected and accumulated in the Hough array. To prove the validity of the new approach different tests were carried out with a wide range of images being captured at different illumination levels.

At this stage it is essential to define a characteristic descriptor for the detected regions. The descriptor must first identify the contents of the region and then sustain a maximum value for license plates over spatial, scale and gradient dimensions.

4. License Plate Detection

A good survey of the recently developed approaches for license plate detection is presented in [10]. Among these approaches are a number of algorithms that rely on the fact that license plates are characterized by a unique and homogenous color (gray-level). A different approach focuses on the high contrasts between the plate numbers and the background. A third approach detects plates through the well-known method of colored-texture analysis. In our approach we define the license plate as a highly textured region with a significant amount of variations in brightness. This definition suits the scale-space representation because contrasts are defined by derivative functions and again normalized derivative functions have the ability to localize characteristic regions within the scale-space.

4.1 License Plate Classifiers

Two different classifiers C_1 and C_2 have been defined for the license plate. C_1 and C_2 define the brightness and contrast levels respectively of a detected region R and we will refer to them as the histogram and texture measures. To calculate C_1 we first build a histogram for R and then identify the region within the histogram that contains the brightest p% of the pixels of R. The upper and lower bounds of this region are the rightmost column of the histogram (i.e. brightest color) and column C_1 respectively. The higher the value of C_1 the more probable R is considered a license plate. If C_1 is less than a predefined threshold d_{th} then R is rejected. The problem with the histogram measure is that it is totally dependent on the illumination and it usually fails with darker images.

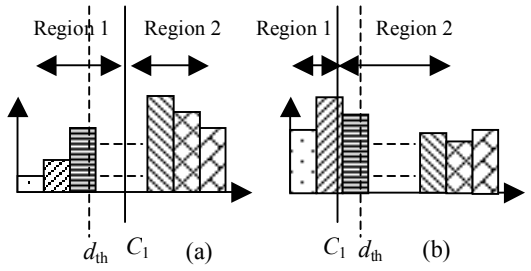


Figure.1 histogram of region R with threshold d_{th} (a) probably R is a license plate i.e. $C_1 > d_{th}$ (b) R is not a license plate i.e. $C_1 < d_{th}$.

The texture measure is defined by eq.6 as the average magnitude of the normalized Sobel gradients of region R, where l_x and l_y are the gradients in the x and y

directions respectively, s is the scale and r is the size in pixels of region R.

$$c_2 = \frac{1}{r}\sum_{n=1}^{r}\left(s\sqrt{l_x^2 + l_y^2}\right) \qquad (6)$$

4.2 Region Detection

For a point (x, y) in the spatial space we calculate a normalized Sobel gradient at each (s, G_{th}) level with scale s and gradient threshold G_{th}. If the normalized gradient is greater than G_{th} then point (x, y) is defined as an edge. Edges of each (s, G_{th}) level are accumulated in a single Hough array within a range of θ around the horizontal and vertical lines. Points of each Hough array are mapped into the spatial space as line segments. Intersections between lines segments are marked and rectangular regions are identified. Finally each group of overlapping regions is merged into one common boundary.

4.3 Plate Detection

First the input image is converted to the gray scale. The size of the radius-dimension of the Hough transform is set to be proportional to the image size. A finite set of scale levels are calculated from the exponential scheme of section 2.2. At each scale level s a discrete Gaussian kernel of size $3.7s$-0.5 is applied to the image through convolution. The smoothed image is then differentiated with a normalized Sobel gradient. Edges are detected over the spatial coordinates in incremental threshold steps. Rectangular regions are extracted from each step and characterized with a normalized classifier as defined in section 4.2. Finally the region with the maximum classifier over both the gradient and scale dimensions is selected as the most fitting region for a license plate.

The above algorithm provides an automatic selection of the gradient thresholds of the plate boundaries. Its performance is independent from the size, orientation, structure and color of the plate. It uses a texture measure that is designed to be somewhat independent from the illumination and scale variations. Moreover its implementation is feasible. In the following section we will conduct several experiments with a large dataset of images to verify the reliability of this approach.

5. Experimental Results

A number of experiments were carried out to (a) tune the parameters of the Hough transform for a robust performance, (b) choose an appropriate classifier for the license plate, (c) define a feasible linear scale-space and (d) verify the effectiveness of the new approach.

5.1. Data Set deduced

The results were based on a large dataset that contains 116 images for different cars captured at different scales, illumination levels and viewing conditions. The license plates are of different sizes, colors, shapes and structure of the contents. Figure.2 shows some of the extreme cases that we have dealt with. The accuracy of the test results (i.e. extracted plates) is defined by the four fuzzy states, *very good*, *good*, *bad* and *failure*. *Very good* means that the plate is extracted up to its boundaries, *good* means that the extracted region does not fit exactly the plate boundaries but it contains the whole plate information, *bad* means only part of the plate is extracted and *failure* means the plate was not extracted at all. These fuzzy states are shown in figure.3.

(a) (b) (c)

(d) (e) (f)

Figure.2 extreme cases (a) bent plate (b) one plate on top of another (c) plate foreground and background are of same color (d) plate boundaries are not clear because of dirt (e) plate without a boundary and (f) numbers are not clear.

5.2 Experiments

It was experimentally determined that license plates have a characteristic height/width ratio that is less than or equal to 0.6 and greater than 0.29 and accordingly a detected region must satisfy this constraint to be considered a license plate. It was also determined that the maximum and minimum lengths of the plate boundaries are 120 and 17 pixels respectively and accordingly lines outside this range are omitted from the Hough array. The thresholds for the gradient were represented by an optimal range of 70 levels starting with G_0=10 and ending with G_f=80. The radius

dimension of the Hough transform ranged from 0 to the maximum diagonal length of the image while the θ-dimension ranged from 80 to 110 degrees for horizontal lines and 0 to 10 degrees for vertical lines.

(a) (b) (c)

Figure.3 extracted plates at 3 different fuzzy states (a) very good (b) good and (c) bad.

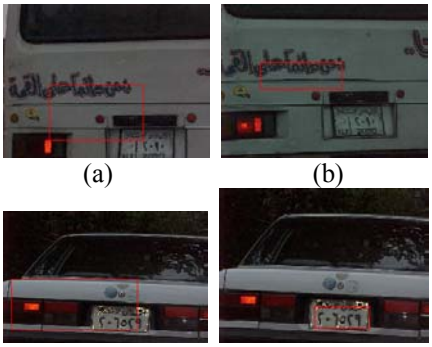
A number of experiments were conducted without the scale-space to test the reliability of the current implementation of the Hough transform. These tests included the histogram measure with k=50 and d_{th}=90 and a simple contrast-stretching scheme. Plates were extracted on an average speed of 3.08 seconds on a 1.6 GHz celeron processor. The results are as shown in table.1. In test 1 no contrast stretching was applied while in 2 the image illumination was increased by a constant value. In test 3 the brightening scheme of 2 was applied to images with a histogram measure less than 50. In test 4 again no brightening was applied and detected regions with height/width ratio less than 0.29 were rejected. Test 5 was a combination of both 4 and 3. The texture measure was used in test 6 together with the constraint of 4.

	Test number					
Accuracy	1	2	3	4	5	6
V. good	59.5%	57.8	58.6	65.5	58.6	66.4
Good	10.3%	13.8	10.3	5.1	13.8	8.6
Bad	9.5%	5.2	4.3	6.9	6.03	10.3
Failure	20.7%	23.2	26.7	22.5	21.6	14.7

Table.1

The texture measure was able classify the *failure* cases much better than the histogram measure. The average speed for plate detection was still 3.08 seconds. The performance of the algorithm at this stage was unstable to small variations in scale and viewing conditions. Some of these results are shown in figure.4 where in the first image pair, (a) and (b), the algorithm failed to locate the plates but nearly located the same regions in the two images. In the second image pair the plate was detected more properly in (d) than (c).

To improve the results of figure.4 we added a linear scale-space to the system. The exponential scheme of section 2.2 was used to define a finite set of scale levels with k=1.5 and s_0=0.75. The number of scale levels was



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

Figure.4 (a) and (b) differ in their viewing conditions while (c) and (d) differ by a simple stretching.

increased in unit steps from 1 to 4 and at each step the results were recorded in table.2. Plates were detected at an average speed of 5.25 seconds for the first 2 scales and increased to an average of 9 seconds for the third scale and finally reached an average of 14 seconds for the 4th scale. Adding scale 2 the algorithm performed much better with the *good* cases leading to a substantial increase in the number of *very good* cases while the percentage of *bad* cases remained the same. At scale 3 the algorithm improved the results of the *bad* cases and again increased the number of *very good* cases. The number of *good* cases remained unchanged for both scales 2 and 3.

	Number of scales			
Accuracy	1	2	3	4
V. good	58.6%	61.2	73.3	61.2
Good	12.9%	8.6	8.6	14.7
Bad	14.7%	15.5	6	9.5
Failure	13.4%	12.9	12.1	14.7

Table.2

At scale 4 the performance of the algorithm was setback and the improvements of scale 3 were misinterpreted at scale 4. On the other hand the percentage of *failure* cases proved to be dependent only on the type of classifier used. This is clear from the results of the 2 tables where the percentage remained steady for the first 5 tests of table.1 and did not decrease until a new classifier was applied in test 6. Also the percentage remained unchanged along test 6 of table 1 and the different tests of table 2 because their results were based on the same classifier. To summarize the above discussion we can say that the *good* and *bad* cases result from detections at

inappropriate scales while the *failure* cases result from misleading classifications. On the other hand scale 2 and 3 are more suitable for the *good* and *bad* cases respectively.

The problem that not all the *good* cases were detected probably at scale 2 and similarly for the *bad* cases at scale 3 is due to the quantization errors of the scale space and in order to improve these results more scale levels should be involved. Moreover, most of the *very good* cases that were misjudged at scale 4 were those images classified as *very good* cases at scales 1 through 3. This means that any feature can be detected correctly over a limited range of scales after which it decays and become unrecognizable.

Our previous discussions can be further verified by the results of figure.5. Referring to figure.4 both images (a) and (b) are of the same size but of different viewing angles and looking back to the sequence (a) of figure.5 we find that both plates were detected at scale 2 and up to scale 4. On the other hand the image pair (c) and (d) of figure.4 are of different sizes but with the same viewing angle. Again looking back to sequence (b) of figure.5 the plate of the smaller car was detected at scale 1 and up to scale 3 while the plate of the larger image was detected later at scale 2 and up to scale 4.

Figure.5 plates of the 2 images of figure.4 detected at 4 consecutive scales.

Similar to the results of Lindeberg [4] figure.5 shows that detected regions with normalized derivative features remain dominant over scale-space. Moreover we add that the normalized derivative feature of an object remains dominant only for a limited interval of scales after which it diminishes and becomes insignificant. The power of expression of a feature over scale space differs from one object to another. In figure.6 we represent a group of license plates that has been detected correctly in the 4 consecutive scales.

Figure.6 detected plates with different sizes, colors, orientation and structures.

6. Conclusion and Future Work

In this work we have defined a new scale-invariant object detector. This detector is characterized by an automatic selection of the gradient threshold of edges. The performance was verified through license plate detection of a large data set of car images. In most cases plates were detected in a good shape. The proposed algorithm was robust to illumination and viewpoint changes. In order to improve the performance of our approach we need to define new feature descriptors that are in general invariant to affine transformations.

References

[1] C. Schmid, R. Mohr and C. Bauckhage. Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2): 151–172, 2000.

[2] C.Schmid and R. Mohr. Local gray value invariants for image retrieval. *PAMI*, 19(5): 530–535, 1997.

[3] D. Lowe. Object recognition from local scale-invariant features. *ICCV99*, pages 1150–1157, 1999.

[4] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2): 79-116, 1998.

[5] K. Mikolajczyk and C. Shmid. Indexing based on scale invariant interest points. *Proceedings of the 8th International Conference on Computer Vision*, Vancouver, Canada, pages 525-531, 2001.

[6] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. IEEE Trans, *PAMI*, 12(7): 629-639, 1990.

[7] T. Tuytelaars and L.Van Gool. Content-based Image Retrieval based on Local Affinely Invariant Regions. *Int'1 Conference on Visual Information Systems*, pages 493-500, 1999.

[8] T. Tuytelaars and L.Van Gool. Wide baseline stereo matching based on local affinely invariant regions. *11th British Machine Vision Conference*, University of Bristol, UK, pages 412-425, 2000.

[9] K. Mikolajczyk and C. Shmid. Indexing based on scale invariant interest points. *Proceedings of the 8th International Conference on Computer Vision*, *Vancouver*, *Canada*, pages 525-531, 2001.

[10] K. Kim, k. Jung and J. Kim. Color Texture-Based Object Detection: An Application to License Plate Localization, *ECCV*, pages 293-309, 2002.

[11] T. Lindeberg and B.M. ter Haar Romeny. Linear scale-space II: early visual operations in Geometry-Driven Diffusion in Computer Vision, ed. B.M. ter Haar Romeny, Kluwer Academic Publishers, Dordrecht, Netherlands, pages 39-71, 1994.