FULLY SCALE AND IN-PLANE INVARIANT SYNTHETIC DISCRIMINANT FUNCTION BANDPASS DIFFERENCE OF GAUSSIAN COMPOSITE FILTER FOR OBJECT RECOGNITION AND DETECTION IN STILL IMAGES

Saad Rehman, Rupert Young, Phil Birch, Chris Chatwin, Ioannis Kypraios
Lasers and Optics Research Group
University of Sussex, Brighton,UK
Email : S.Rehman@sussex.ac.uk

ABSTRACT

A difficult pattern recognition problem is to recognize objects despite distortions in position, orientation and scale in cluttered backgrounds. A system capable of detecting target objects despite any kind of geometrical distortion has many potential applications since it will be able to detect target objects when the orientation and position of the target or camera is unknown. In this work we report the use of fully invariant correlation filters for object detection in still images despite any kind of geometrical distortion of the target object. A mapping technique is combined with a bandpass difference of gaussian composite correlation filter, capable of creating invariance to various types of distortion of the target object.

Keywords: *Object Recognition, Random geometrical distortion, fully invariant correlation filters, mapping technique, gaussian composite correlation filter*

1. INTRODUCTION

Synthetic Discrimination Function (SDF) based techniques provide a solution to the problem of invariant correlation filter design, expected distortions being included in the filter design. A log r-theta mapping can be applied to the input image to give invariance to in-plane rotation and scale by transforming rotation and scale variations of the target object into vertical and horizontal shifts[7,8]. The SDF filter is then trained using the log-mapped image. A Difference of Gaussian band pass filter is added in the design of the filter to provide edge enhancement of the input images and so obtain sharper correlation peaks. Areas producing a strong correlation response can then used to determine the position, in-plane rotation and scale of the target objects in the scene.

Log-polar mapping as a pre-processing operation for correlation filters may offer the capability to extend the range of distortions over which correlation filters can detect and recognise a target object, possibly also contained in a highly cluttered environment. In this paper we summarise the log-polar pre-processing operation and discuss methods of multiplexing correlation filters to accommodate orientation changes of the target object. We show that a log-polar pre-processed appropriately

designed multiplexed correlation filter is able to detect and recognise a target object from a highly cluttered environment, although the correlation response is not ideal due to limitations, in the current implementation, of the clutter model employed. A set of training images of a Jaguar model car were constructed, in which each image differed by 5 degrees in orientation angle. Thus a total of 72 images were constructed to cover the whole range to 360 degrees.

2. LOGMAP PRE-PROCESSING

To detect and recognize[1-3] target objects in a scene despite differences in scale or in-plane rotation to the target reference images, a log r-θ mapping, or logmap, can be employed [4-9]. The structure of the sensor is based on a Weiman polar exponential grid [10-13] and consists of concentric circles of pixels which are exponentially spaced and increase in size from the centre to the edge. Each sensor pixel on the circular region of the x-y Cartesian space is mapped into a rectangular region in polar image space r-θ. The sensor's geometry maps concentric circles in the Cartesian space into vertical lines in the polar space and radial lines in the Cartesian space into horizontal lines in the polar space. The log-polar mapping is not shift

invariant, so the properties described above hold only if they are with respect to the origin of the Cartesian image space. The complex logarithmic mapping can be described as [10]:

$$w = \log z$$
(1)

By applying the complex form notation:

$$z = x + i\,y$$
(2)

where $r = \sqrt{x^2 + y^2}$ and $\theta = \arctan\left(\dfrac{y}{x}\right)$, we have

$$w = \log r + i\theta = u + iv$$
(3)

where $u = \log r$ and $v = \theta$. Hence, an image in the z-plane with co-ordinates x and y is mapped to the w-plane with co-ordinates u and v. The mapped image from the Cartesian space (z-plane) in to the Polar space (w-plane) is referred to as the log-polar mapping or the logmap. This process can be reversed to produce the inverse mapping from Polar space (w-plane) to Cartesian space (z-plane). Complex log-polar mapping is shown in figure 1.

2.1 In-Plane Invariance

If an image is rotated by an angle ϕ about the origin, then [10]:

$$z = r\,e^{i(\theta + \phi)}$$
(4)
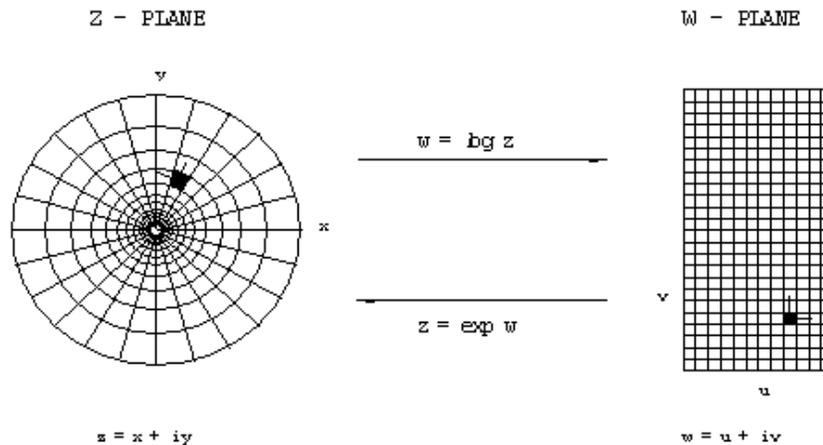
Figure 1 Complex log-polar mapping

Thus,

$$w = \log r + i\theta + i\phi = u + iv + i\phi$$
(5)

In effect, rotating the image by the angle ϕ has resulted in a vertical shift in the mapped image by the rotation angle.

2.2 Scale Invariance

If the image in the z-plane is scaled by a factor α, then [10]:

$$z = r\,\alpha\,e^{i\theta}$$
(6)

Thus,

$$w = \log r + \log\alpha + i\theta = u + \log\alpha + iv$$
(7)

In effect, scaling the image by the factor α has resulted in a horizontal shift in the mapped image by the scaling factor.

3. INVERSE MAPPING

The inverse mapping can be shown in the following way. Since we know that $z = e^w$ and $w = u + iv$, combining the two equations we get [10-13]:

$$z = e^{u+iv}$$
(8)

Expanding the equation (8):

$$z = e^u[\cos(v) + i\sin(v)]$$
(9)

As we know from the Eq (2) that $z = x + iy$, the equation (9) thus becomes:

$$e^u[\cos(v) + i\sin(v)] = x + iy$$
(10)

Equation (10) implies that:

$$x = e^u \cos(v)$$
(11)

$$y = e^u \sin(v)$$
(12)

We also know that:

$$r = \sqrt{[x^2 + y^2]}$$
(13)

Putting the values of equation (11) and (12) in equation (13) we get

$$r = \sqrt{[e^u \cos(v)]^2 + [e^u \sin(v)]^2}$$
(14)

Equation (14) implies that $u = \log r$ and similarly:

$$\theta = \arctan\left(\frac{e^u \sin(v)}{e^u \cos(v)}\right) = v$$
(15)

4. DIFFERENCE OF GAUSSIAN BAND PASS

The Difference of Gaussian (DOG) function is a readily computed circular symmetric wavelet which approximates the Maxican hat wavelet. DOG bandpass filters an image and performs the edge enhancement operation on the objects present in the image[19]. The DOG function is defined as the difference of two differently scaled Gaussian functions, $g_i(x, y)$, where $i = 1,2$.

$$g_i(x, y) = \frac{1}{2\pi\sigma_i^2}.\exp(-\frac{x^2 + y^2}{2.\pi.\sigma_i^2})$$
(16)

Therefore the DOG function is given by

$$g(x, y) = g_1(x, y) - g_2(x, y)$$
(17)

that is:

$$g(x,y) = \frac{1}{2\pi\sigma_1^2}\exp[-\frac{x^2+y^2}{2\pi\sigma_1^2}] - \frac{1}{2\pi\sigma_1^2}\exp[-\frac{x^2+y^2}{2\pi\sigma_2^2}]$$
(18)

where $(\sigma_1\sigma_2)$ represents the standard deviations of the two Gaussian functions. The frequency domain representation of the DOG function is given below as:

$$G(u,v) = \exp[-2.\pi^2.\sigma_1^2.(u^2+v^2)] - \exp[-2.\pi^2.\sigma_2^2.(u^2+v^2)]$$
(19)
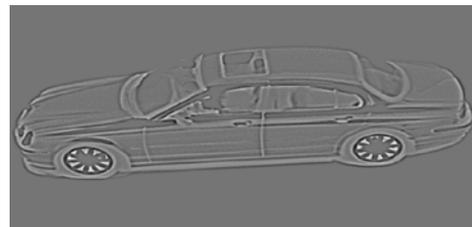
Figure 2 Filtered image with standard deviation of ratio 1.6 and Bandpass set at 0.35 (S_f)

The scaling of the band pass is critical. The value of the band pass maximum frequency response should be chosen to give the best compromise between intra-class distortion tolerance and inter-class discrimination of the resulting filter.. The low frequencies must be reduced to enhance the discrimination ability of the filter. The higher

frequencies must be reduced enough to give adequate target distortion tolerance. The precise choice for the band pass location, will thus depend on the trade-off desired between these two conflicting requirements. The best performance of the DOG filter occurs at its closest approximation to the Mexican hat wavelet when the ratio of σ_1 to σ_2 is 1.6 [19]. Figure 2 shows a DOG filtered jaguar car image at 1.6 ratio of standard deviation values and the bandpass set at 0.35 of the maximum sampling frequency (S_f). Gray background in figure 2 is at zero intensity in order to show the negative values in the image as darker regions.

5. SYNTHETIC DISCRIMINATION FUNCTION

In the SDF [14-16,18] design method, the expected object distortions are included in the correlation filter by multiplexing the weighted versions of the target object into a composite image. The resulting correlation outputs at the origin of these cross-correlations are constrained to be the same and are equal to a pre-specified constant. Let $h(x,y)$ denote the composite image and $t_i(x,y)$ denote the training image set where $i = 1,2,...,N$ and N is the number of the training images used in the synthesis of the SDF. Then the value at the origin of the correlation plane between the composite image and each of the training images is assumed to be equal to a constant c.

Thus:

$$h(x, y).t_i(x, y) = \int \int t_i(x, y)h^*(x, y)dxdy$$

(20)

The composite image is assumed to be a linear combination of the N training images:

$$h(x, y) = a_1t_1(x, y) + a_2t_2(x, y) + ... + a_Nt_N(x, y)$$

$$= \sum_{i=1}^{N} a_i t_i(x, y)$$

(21)

where the coefficients $a_i (i = 1,2,..., N)$ are determined to satisfy the constraint c. By substituting equation (21) into (20) we have:

$$\sum_{i=1}^{N} a_i^* R_{ij} = c$$

(22)

Where

$$R_{ij} = \int \int t_i(x, y)t_j^*(x, y)dxdy$$

(23)
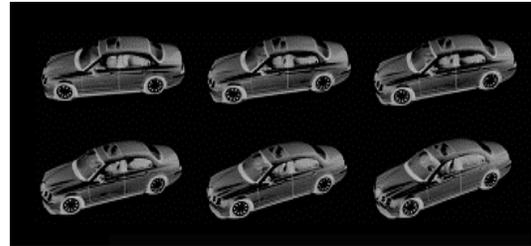
Figure 3 Examples of training images

Figure 3 shows some of the training images used in the developed system. The composite image is constructed by using the images of the Jaguar car rotated at 0,10,15,20,25 and 30 degrees. The resulting composite image is shown in Figure 4.
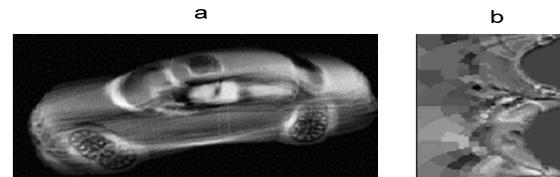
Figure 4 (a) Composite Image (b) Logmap of Composite

6. PERFORMANCE METRICS

To measure the performance of the correlation filters some basic measures have to be calculated. The basic performance measure is correlation output peak intensity (COPI). It signifies the maximum intensity value of the correlation output plane. It is defined as [20]:

$$COPI = \max_{x,y}\{| C(x, y)|^2\}$$

(24)

where C(x,y) is the output correlation amplitude value at (x,y). A filter with high COPI shows good performance and a high detection ability. Another important performance measure is Peak-to-correlation energy measure (PCE). The basis of the PCE is that the COPI should be as high as possible while at the same time the over all correlation

plane energy should be as low as possible. It is defined as [20]:

$$PCE = \frac{COPI}{Energy_c}$$

(25)

Where $Energy_c$ is the total correlation plane energy and is defined as:

$$Energy_c = \sum |C(x,y)|^2$$

(26)

7. DEPENDENCE OF FREQUENCIES ON STANDARD DEVIATION

The filtering operation performed by the DOG filter can be controlled by the standard deviations. The pass band of frequencies can be altered by changing the standard deviation in the DOG construction whilst keeping ratio of 1.6. This in tern translates to different peak widths in the correlation plane. In the results that follow, composite image remains the same, target image is the car image out-of-plane rotated at 30 degrees.
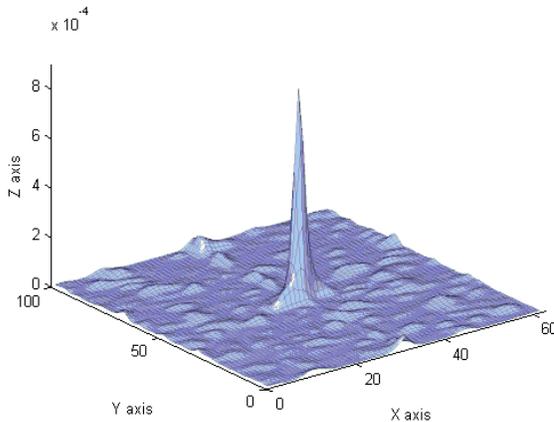
*Figure 5 Correlation plane for standard deviation ratio of 1.6 , bandpass set at $0.35S_f$ and COPI = 1.12*10-5, PCE = 0.22*
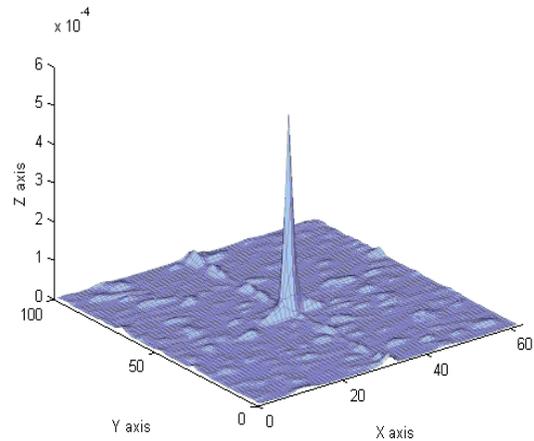
Figure 6 Correlation plane for standard deviation ratio of 1.6, bandpass set at $0.46S_f$ and COPI = $6.5*10^{-6}$, PCE = 0.30
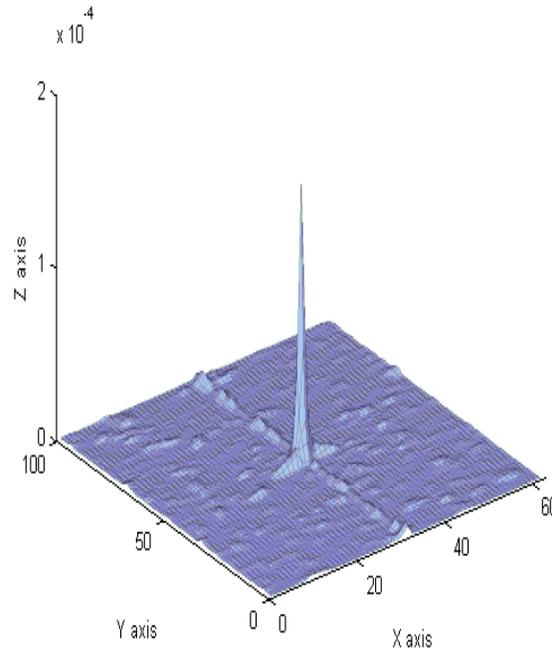
Figure 7 Correlation plane for standard deviation ratio of 1.6, bandpass set at $0.58S_f$ and COPI = $1.64*10^{-6}$, PCE = 0.36

8. IN-PLANE ROTATIONAL INVARIANCE OF THE BANDPASS SDF

In this section we discuss the results of the in-plane rotational invariance of the SDF filter. Composite image remains the same. The in-plane rotated images employed are shown in Figure 8. All the simulation results are obtained using the standard deviation ratio of 1.6 with bandpass set at $0.35S_f$.
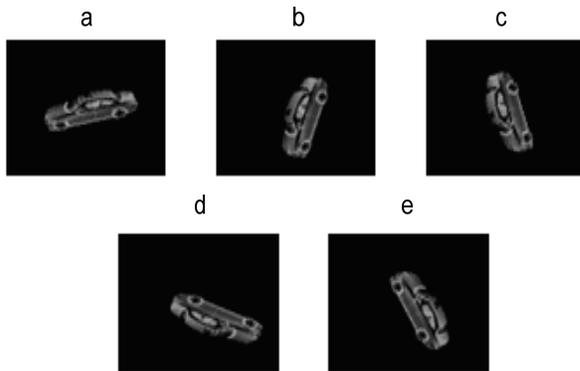
Figure 8 (a) 10 degrees in-plane rotation (b) 70 degrees in-plane rotation (c) 100 degrees in-plane rotation (d) 150 degrees in-plane rotation (e) -70 degrees in-plane rotation

The correlation planes generated by the rotated images are shown in figure 9 to figure 13.

Figure 10 Correlation plane for 70 degrees in-plane rotation, COPI = 4.9*10^{-6}, PCE = 0.27

Figure 9 Correlation plane for 10 degrees in-plane rotation, COPI = 1.2*10^{-5}, PCE = 0.24
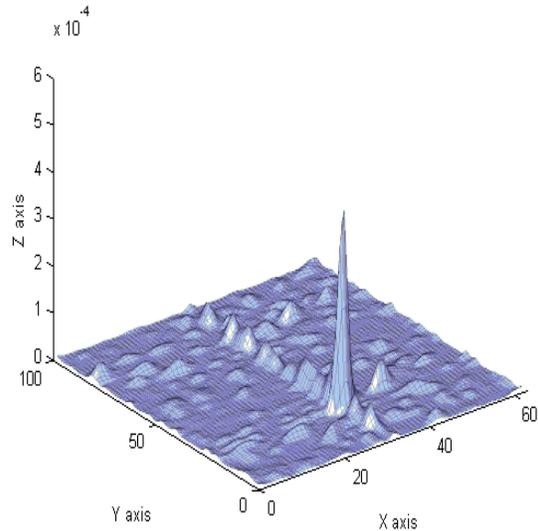
Figure 11 Correlation plane for 100 degrees in-plane rotation, COPI = 6.48*10^{-6}, PCE = 0.23
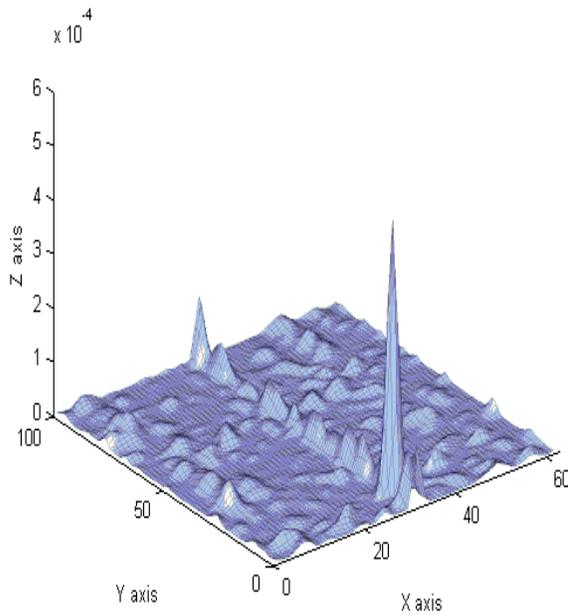
Figure 12 Correlation plane for 150 degrees in-plane rotation, COPI = 5.34*10⁻⁵, PCE =0.28
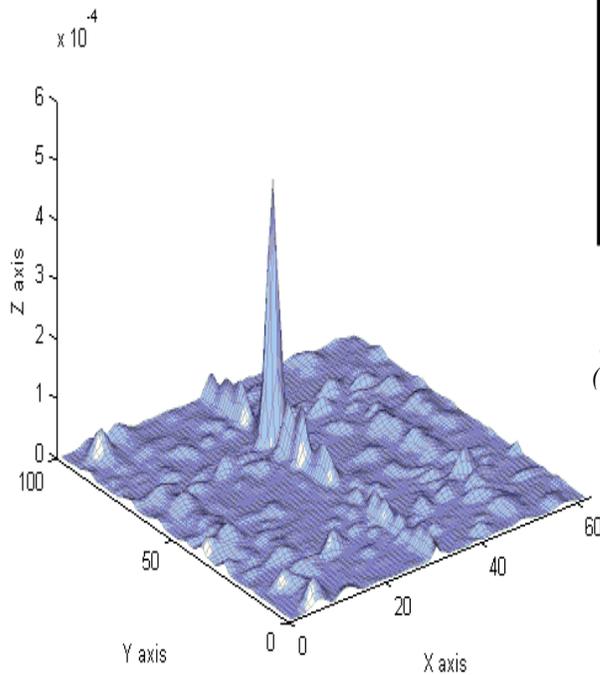
Figure 13 Correlation plane for -70 degrees in-plane rotation, COPI = 1.039*10⁻⁵, PCE = 0.223

In the results shown in figures 9-13 it is apparent that the correlation peaks move across the correlation plane as the rotational angle is changed, in linear proportion. Thus by combining the in-

plane rotation invariance of the logmap and the distortion invariance of a band pass SDF filter, a rotation and distortion invariant SDF filter can be realised for object recognition.

9. OUT-OF-PLANE INVARIANCE OF THE BANDPASS SDF

In this section the out-of-plane invariance of the SDF filter is tested. The composite image remains the same. We introduce a number of test images without background noise which are rotated and scaled for the detection of the car. The test images are shown in figure 14. The correlation planes for the above test images are shown below in figures 15-17. All the simulation results are obtained using the standard deviation ratio of 1.6 and bandpass set at 0.35 S_f.
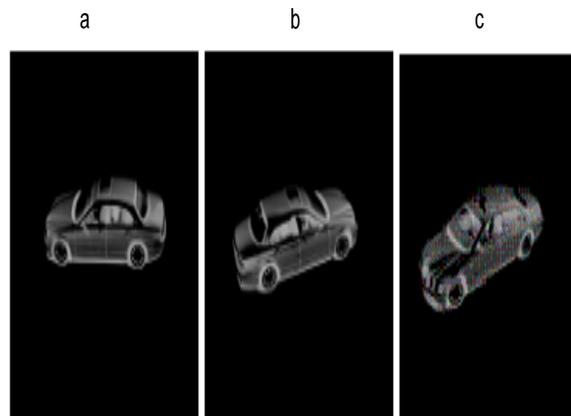
Figure 14 (a) Test Image car rotated at 355 degrees (b) Test Image car rotated at 195 degrees (c) Test image car rotated 30 degrees and scaled at 90%
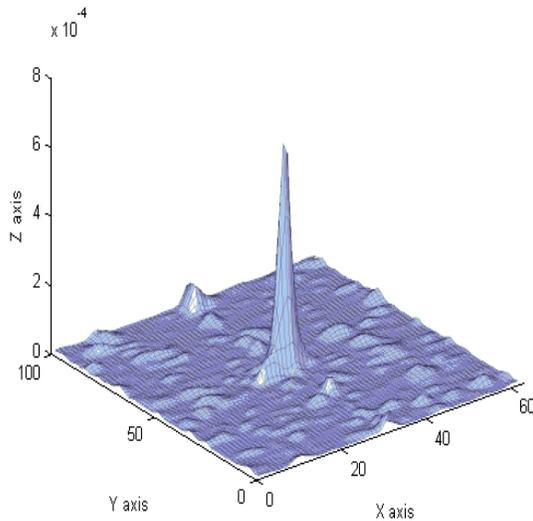
Figure 15 Correlation plane for 30 degrees out-of-plane rotated and 90% scaled car, COPI = 1.3822*10-5, PCE = 0.21
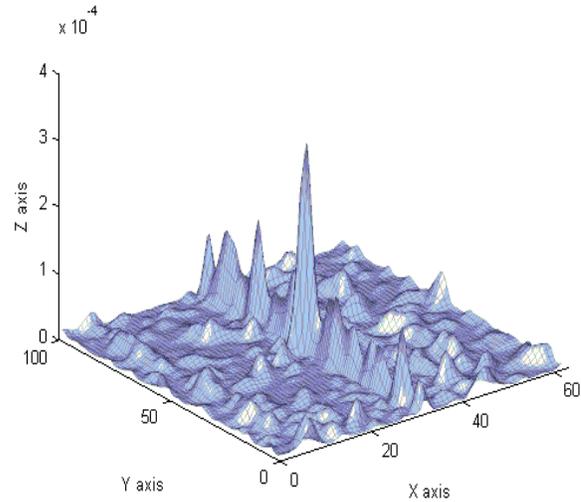
Figure 17 Correlation plane for 355 degrees out-of-plane rotated car, COPI = 1.853*10^{-5}, PCE = 0.26

It can be seen from figures 15 to 17 that as the out-of-plane rotational angle increases, the peak height decreases. Since there is no background noise the correlation output plane shows minimal disruption. The correlation peaks are localised due to the inclusion of the difference of Gaussian band pass in the filter design. The composite reference image accomodates the out-of-plane rotation of the car allowing maintenance of the correlation peak height over the swathe of angles covered within the composite image.

The correlation plane corresponding to figure 16 is shown in figure 17.

Let us combine all the three distortions and test the full invariance of the bandpass SDF filter. Suppose that the image of a car with an out-of-plane rotation of 30 degrees , resized at 85% of the original image and finally in-plane rotated at 30 degrees. The correlation plane for the above simulation is shown in figure 18.
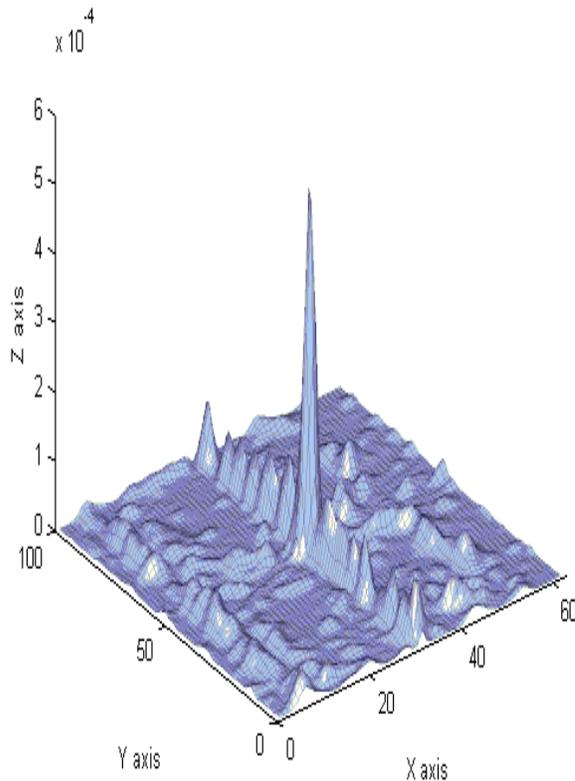
Figure 16 Correlation plane for 195 degrees out-of-plane rotated car, COPI = 1.19*10^{-5}, PCE =0.22
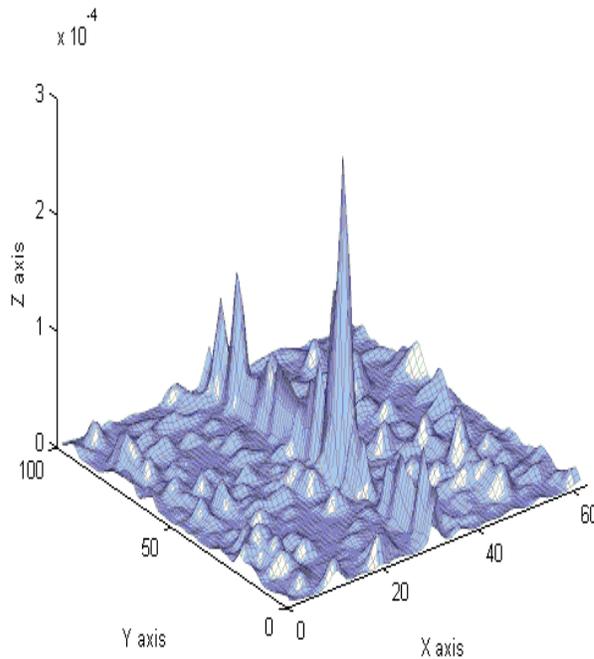
*Figure 18 Correlation plane for car image with 30 degrees out-of-plane rotation, 85% seized and 30 degrees in-plane rotated, bandpass set at 0.35 Sf and COPI = 1.113*10-5, PCE = 0.225*

It can be seen that the performance of the filter decreased as all the distortions were combined. However, the bandpass SDF filter was still able to detect the car.

10. BACKGROUND CLUTTER TOLERANCE

In this section the background clutter tolerance of the DOG-SDF filter is assessed. Figure 19 shows a scene where a 35 degrees out-of-plane rotated car is super imposed. This car is resized to 85% of the original size and 5 degrees in-plane rotated.

Figure 19 Target car 35 degrees out-of-plane rotated, 5 degrees in-plane and 85% resized superimposed on an image scene
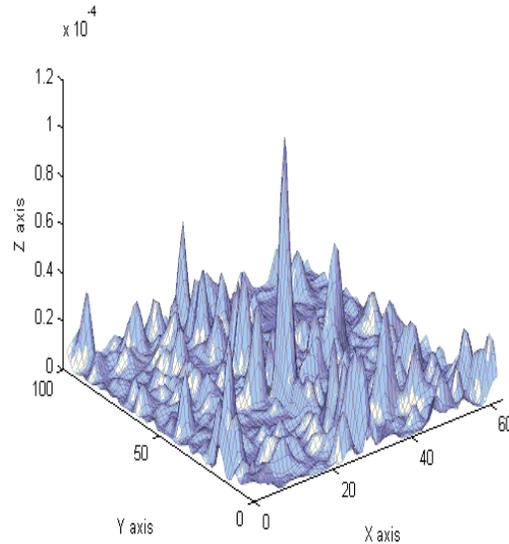
*Figure 20 Correlation plane for car image with 35 degrees out-of-plane rotation, 5degrees in-plane and 85% resized , bandpass set at 0.35 S_f and COPI = 2.013*10^{-5}, PCE = 0.2*

Although the car was detected in the presence of the clutter, significant degradation of the correlation plane background is clear.

11. CONCLUSION

The results presented indicate that the DOG band pass SDF filter was able to detect the rotated and scaled car, when used in conjunction with a logmap pre-processing operation. The disruption present in the log-mapped images increased significantly when scaling and in-plane rotating the reference image due to finite sampling. In order to reduce the disruption, interpolation was adopted for smooth rendering of the images which reduced the disruption but did not eliminate it. Methods to further reduce the effects of finite sampling will be investigated in future work in order to reduce its impact.

REFERENCES

[1]   Richard O.Duda, Peter E. Hart, David G. Stork , "Pattern Classification", *John Wiley & Sons*, 2nd Edition.

[2] D. Casasent and D. Psaltis, "Position, rotation, and   scale invariant optical correlation",

Applied Optics Vol. 15, No. 7, 1795-1799 (1976)

[3]   K. Mersereau and G. Morris, "Scale, rotation, and shift invariant image recognition", *Applied Optics* Vol. 25, No. 14, 2338-2342 (1986)

[4]   Weiman, C.F.R., and Chaikin, G., "Logarithmic spiral grids for image processing and display," *Computer Graphics and Image Processing* Vol. 11, 197-226 (1979)

[5]   Sandini, G., and Dario, P., "Active vision based on space-variant sensing," in: *5th International symposium on Robotics Research,* 75-83 (1989)

[6]   Schwartz, E.L., Greve D., and Bonmasser, G., "Space-variant active vision: definition, overview and examples," *Neural Networks*, Vol. 8, No. 7/8, 1297-1308 (1995)

[7]   P.Bone, R. C. D. Young, C. Chatwin, "Position-, rotation-, scale- and orientation-invariant multiple object recognition from cluttered scenes", *Optical Engineering*, Vol. 45, No. **7**, 077203-1 to 8, (July 2006)

[8]   P.Bone, I.Kypraios, R.C.D.Young. C.R. Chatwin, "Fully invariant object recognition in cluttered scenes", Invited paper, *Proc. SPIE, Information Technologies*, Editors: A. Andriesh, V. Perju, Chisinau, Moldova, May 2004.

[9]   Shilpi Goyal, Naveen K. Nishchal, Vinod K. Beri, and Arun K. Gupta, "Wavelet-modified maximum average correlation height filter for rotation invariance that uses chirp encoding in a hybrid digital–optical correlator", *Applied Optics*, Vol. 45, No. 20, 10 July 2006.

[10]   Weiman, C.F.R., "3-D sensing with polar exponential sensor arrays," in: *Digital and Optical Shape Representation and Pattern Recognition*, Proc. SPIE Conf. on Pattern Recognition and Signal Processing, Vol. 938, 78-87 (1988)

[11]   Weiman, C.F.R., "Exponential sensor array geometry and simulation," in: *Digital and Optical Shape Representation and Pattern Recognition*, Proc. SPIE Conf. on Pattern

Recognition and Signal Processing, Vol. 938, 129-137 (1988)

[12]   C-G. Ho, R.C.D. Young, C.R. Chatwin, "Sensor Geometry and Sampling Methods for Space-Variant Image Processing", *Journal of Pattern Analysis and Applications*, Vol. 5, 369-384, (2002)

[13]   Weiman, C.F.R., and Chaikin, G., "Logarithmic spiral grids for image processing and display," *Computer Graphics and Image Processing* Vol. 11, 197-226 (1979)

[14]   H. J. Caulfield and W. Maloney, "Improved discrimination in optical character recognition", *Applied Optics*, *Vol.* 8, *No.* 11, 2354-2356 (1969)

[15]   C. F. Hester and D. Casasent, "Multivariant technique for multiclass pattern recognition", *Applied Optics*, Vol. 19, 1758-1761 (1980)

[16]   Z. Bahri and B. V. K. Kumar, "Generalized synthetic discriminant functions", *Journal of Optical Society of America*, Vol. 5, *No.* 4, 562-571 (1988)

[17]   I. Kypraios, R. Young and C. R. Chatwin, "An investigation of the non-linear properties of correlation filter synthesis and neural network design", *Asian Journal of Physics*, Vol. 11, No. 3, (2002)

[18]   Saad Rehman, Peter Bone, Rupert Young, Chris Chatwin, "Object Detection and Recognition in Cluttered Scenes Using Fully Scale and In-Plane Invariant Synthetic Discriminant Function Filters", *Journal of Computer and Information Sciences*, Vol 1, No 1, 15-18, 2007.

[19]   D. Marr, E. Hildreth, "Theory of edge detection", *Proc. R. Soc. Lon. B 20*, pp.187-217, (1980).

[20]   B.V.K. Kumar and L. Hassebrook, "Performance measures for correlation filters", *Applied Optics*, Vol. 29, No. 20, pp. 2997-3006, (1990).