Australian Government
Department of Defence
Science and Technology

Implementation of the Scale Invariant Feature Transform Algorithm in MATLAB®

Leonid K Antanovskii

Weapons and Combat Systems Division
Defence Science and Technology Group

DST-Group–TR–3347

ABSTRACT

This report addresses the description and MATLAB implementation of the Scale-Invariant Feature Transform (SIFT) algorithm for the detection of points of interest in a grey-scale image. Some illustrative simulations for code verification are conducted.

RELEASE LIMITATION

Approved for public release

Published by

Weapons and Combat Systems Division
Defence Science and Technology Group
PO Box 1500
Edinburgh, South Australia 5111, Australia

Telephone: *1300 333 362*
Facsimile: *(08) 7389 6567*

February, 2017
AR-016–811

APPROVED FOR PUBLIC RELEASE

Implementation of the Scale Invariant Feature Transform Algorithm in MATLAB®

Executive Summary

The most important problem in *Computer Vision* is to detect an object from its images taken from various positions and at variable illumination. The only way to recognize an object from its images, some of which may play the role of training images, is to associate points of interest to which distinctive features can be assigned and matched between different images. The matching procedure will be successful only if the extracted features are nearly invariant to scale and rotation of the image.

This report addresses the description and MATLAB implementation of the Scale-Invariant Feature Transform (SIFT) algorithm for the detection of points of interest in a grey-scale image. Some illustrative simulations for code verification are conducted.

THIS PAGE IS INTENTIONALLY BLANK

Author

Leonid K Antanovskii
Weapons and Combat Systems Division

Born in Siberia, Leonid Antanovskii holds a Master of Science (with distinction) in Mechanics and Applied Mathematics from the Novosibirsk State University and a PhD in Mechanics of Fluid, Gas and Plasma from the Lavrentyev Institute of Hydrodynamics of the Russian Academy of Science. Since graduation he worked for the Lavrentyev Institute of Hydrodynamics (Russia), Microgravity Advanced Research & Support Center (Italy), the University of the West Indies (Trinidad & Tobago), and in private industry in the USA and Australia.

Leonid Antanovskii joined the Defence Science and Technology Group in February 2007 working in the area of weapon–target interaction.

THIS PAGE IS INTENTIONALLY BLANK

Contents

1 INTRODUCTION 1

2 DESCRIPTION OF THE SIFT ALGORITHM 1

3 CODE EVALUATION 4

4 DISCUSSION 4

5 REFERENCES 5

APPENDIX A: FIGURES 7

List of Figures

A1 Synthetic image 1 generated by VIRSuite in visible band 7
A2 Difference-of-Gaussian image 1 for $\sigma = 3.2$ 7
A3 Keypoint locations and orientations in image 1 8
A4 Histogram of keypoint scales in image 1 8
A5 Synthetic image 2 generated by VIRSuite in infrared band 9
A6 Difference-of-Gaussian image 2 for $\sigma = 3.2$ 9
A7 Keypoint locations and orientations in image 2 10
A8 Histogram of keypoint scales in image 2 10
A9 Original and transformed images 11
A10 Putative matches 11
A11 Inlier matches 12
A12 Original and recovered images 12

THIS PAGE IS INTENTIONALLY BLANK

1 Introduction

The most important problem in *Computer Vision* is to detect an object from its images taken from various positions and at variable illumination. The only way to recognize an object from its images, some of which may play the role of training images, is to associate points of interest to which distinctive features can be assigned and matched between different images. The matching procedure will be successful only if the extracted features are nearly invariant to scale and rotation of the image.

Scale-Invariant Feature Transform (SIFT) algorithm has been designed to solve this problem [Lowe 1999, Lowe 2004*a*]. Up to date, this is the best algorithm publicly available for research purposes. It is worthwhile noting that the commercial application of SIFT to image recognition is protected by the patent [Lowe 2004*b*]. The main idea of the SIFT algorithm is based on progressive smoothing and resizing an image, and taking local extrema of the difference-of-Gaussian functions in the three-dimensional space of pixel coordinates and scales. The points of interest, also called keypoints, are the corrected local extrema to achieve a better accuracy to a sub-pixel level and extra stability by eliminating noise. Then each keypoint is assigned an orientation (or even multiple orientations) defined by the histogram of local gradient of the image intensity. Relative to the orientation, a descriptor is computed from the keypoint neighbourhood, which is invariant to image scale and rotation, but yet highly distinctive.

Synthetic imagery generated by VIRSuite are used for testing the developed code. The VIRSuite software is a real-time scene generator developed in Defence Science and Technology Group (see e.g. [Swierkowski et al. 2014]). It is designed to provide closed-loop dynamic simulations of complex scenarios that include moving objects, composite backgrounds, sources of radiation and precise radiometry. The software is capable of generating high-fidelity multimodal live imagery comprising range data and passive imagery in the visual and infrared bands.

This report addresses the description and MATLAB® implementation of the SIFT algorithm for the detection of points of interest in a grey-scale image. Some illustrative simulations for code verification are conducted. The developed MATLAB code may be released on request. It can be used as a prototype for an advanced and optimized software.

2 Description of the SIFT algorithm

Denote the intensity of the input grey-scale image by $I_0(x, y)$ where x and y are pixel coordinates. We assume that the intensity $I_0(x, y)$ is normalized to the range $0 \le I_0(x, y) \le 1$. We rescale the image by doubling its size (aliasing) and apply initial smoothing (anti-aliasing) with the blur amount of $\sigma = 0.5$. It is claimed in [Lowe 2004*a*] that, on average, this simple pre-processing will increase the number of detected keypoints by a factor of 4.

To blur images we use Gaussian filter with the kernel

$$G(x, y, \sigma) = \frac{1}{2\,\pi\,\sigma^2}\,\exp\left(-\frac{x^2 + y^2}{2\,\sigma^2}\right) \tag{1}$$

where σ is the blur amount or scale. The blurred image intensity $I(x, y, \sigma)$ is given by the expression

$$I(x, y, \sigma) = G(x, y, \sigma) * I_0(x, y) \tag{2}$$

where the asterisk denotes convolution with respect to x and y. Due to the convergence of $G(x, y, \sigma)$ to the delta-function $\delta(x, y)$ as σ goes to zero, we have $I(x, y, 0) = I_0(x, y)$. The points of interest are closely related to the local extrema (maxima or minima) of the Laplacian-of-Gaussian function

$$L(x, y, \sigma) = \sigma^2 \, \Delta I(x, y, \sigma) \tag{3}$$

in the scale-space of (x, y, σ), where Δ denotes the Laplacian operator with respect to x and y. Due to the formula

$$\lim_{k \to 1} \frac{G(x, y, k\,\sigma) - G(x, y, \sigma)}{k - 1} = \sigma^2 \, \Delta G(x, y, \sigma)\,, \tag{4}$$

the Laplacian-of-Gaussian function can be approximated by the difference-of-Gaussian function

$$D(x, y, \sigma) = I(x, y, k\,\sigma) - I(x, y, \sigma) \;\approx\; (k - 1)\, L(x, y, \sigma) \tag{5}$$

where k is some constant close to 1. Since the factor $(k - 1)$ does not affect extrema location of $L(x, y, \sigma)$, the scaled Laplacian-of-Gaussian function is replaced with the difference-of-Gaussian function $D(x, y, \sigma)$ which can be computed more efficiently.

The computation of local extrema of $D(x, y, \sigma)$ is based on generating the so-called *scale-space pyramid*. Progressive blurring is applied to the image for carefully chosen discrete values of scales σ separated by a constant factor k such that $k > 1$ [Lowe 2004*a*]. In order to speed up the algorithm, the blurred image is down-sampled by resizing it by a factor of 0.5 when the value of σ is doubled, and then the blurring procedure is repeated. This operation creates the octaves of progressively blurred images. Doubling of σ implies that $k = 2^{1/N}$ where N is the number of scale intervals. In order to find local extrema, we need to produce $D(x, y, \sigma)$ at $(N + 2)$ levels of σ (one above and one below), and therefore generate $(N + 3)$ blurred images for each octave of the pyramid. It is recommended in [Lowe 2004*a*] to use 4 octaves, 3 scale intervals ($N = 3$), and set the initial scale to $\sigma = 1.6$ (prior blurring).

The discrete values of $D(x, y, \sigma)$ form a three-dimensional array for each octave, so the local extrema can be computed by examining every triplet (x, y, σ) with its 26 immediate neighbours in the hexagonal mesh (8 neighbours at the same level of σ and 9 neighbours at the level above and below). The associated value of σ is called the scale of the keypoint (x, y).

Normally, a lot of extrema will be detected in a typical image, so an appropriate elimination of noise is essential to achieve stability. This is accomplished by the following procedure. First, the gradient and Hessian matrices of $D(x, y, \sigma)$ are estimated from its discrete values at the local extrema. Denoting partial derivatives by the corresponding subscripts, we compute the corrections to x, y and σ by the formula

$$(\mathrm{d}x, \mathrm{d}y, \mathrm{d}\sigma) = -\,(D_x, D_y, D_\sigma) \begin{pmatrix} D_{xx} & D_{xy} & D_{x\sigma} \\ D_{xy} & D_{yy} & D_{y\sigma} \\ D_{x\sigma} & D_{y\sigma} & D_{\sigma\sigma} \end{pmatrix}^{-1} . \tag{6}$$

Only reasonably small corrections satisfying the conditions $|\mathrm{d}x| < 0.5$ and $|\mathrm{d}y| < 0.5$, as well as $\sigma + \mathrm{d}\sigma > 0$, are taken into account, for which we compute the corrected sub-pixel coordinates and scale

$$(\hat{x}, \hat{y}, \hat{\sigma}) = (x, y, \sigma) + (\mathrm{d}x, \mathrm{d}y, \mathrm{d}\sigma) \tag{7}$$

and then find the corrected value of the difference-of-Gaussian function

$$D(\hat{x}, \hat{y}, \hat{\sigma}) = D(x, y, \sigma) + \frac{1}{2}\left[D_x(x, y, \sigma)\,\mathrm{d}x + D_y(x, y, \sigma)\,\mathrm{d}y + D_\sigma(x, y, \sigma)\,\mathrm{d}\sigma\right] . \tag{8}$$

In order to eliminate extrema with low contrast or poorly localized at edges, we leave only those keypoints which satisfy the inequalities

$$|D(\hat{x}, \hat{y}, \hat{\sigma})| > \alpha\,, \quad \frac{D_{xx}\,D_{yy} - D_{xy}^2}{(D_{xx} + D_{yy})^2} > \beta\,, \tag{9}$$

where α and β are some thresholds. Their optimal values, $\alpha = 0.03$ and $\beta \approx 0.08$, obtained from conducted numerical experiments over a variety of images are recommended in [Lowe 2004*a*].

The next step is to assign a consistent orientation to each keypoint $(\hat{x}, \hat{y})$ with scale $\hat{\sigma}$. By construction, the gradient (D_x, D_y) vanishes at the corrected point $(\hat{x}, \hat{y}, \hat{\sigma})$, and therefore its value does not determine an orientation. However, this gradient can be averaged over a neighbourhood of the keypoint with the Gaussian kernel (1) centred at the keypoint. It is recommended in [Lowe 2004*a*] to choose $1.5\,\hat{\sigma}$ as the scale of the Gaussian weight. In principle, the direction of the averaged gradient can be used as the reference angle with respect to which a rotation-invariant keypoint descriptor will be computed. A better option is to assign potentially multiple orientations to a keypoint [Lowe 2004*a*] when several peaks of the gradient magnitude with respect to the gradient angle have comparable values. This is achieved by building a histogram of the weighted gradient magnitudes with respect to the gradient angles. The orientation histogram has 36 bins covering 360 degrees. Peaks of the orientation histogram exceeding 80% of the highest peak are used to create a keypoint with that quantized orientation. So, multiple keypoints at the same location may be created. It is claimed in [Lowe 2004*a*] that about 15% of keypoint locations would have multiple orientations, but these contributed significantly to the stability of matching. Finally, the keypoint orientation is corrected by interpolating the peak position using a quadratic spline.

The orientation of a keypoint is used to extract a descriptor from a neighbourhood of the keypoint location, which is invariant with respect to a similarity transform of the image. The keypoint descriptor is a 128-dimensional vector of unit length. Its construction is described in detail in [Lowe 2004*a*].

The matching of a descriptor to a database of descriptors is quite straightforward. We find a candidate database descriptor whose Euclidean distance d from the given descriptor is minimal. The candidate descriptor will be accepted if d is less than some parameter ρ times the minimum distance over all the rest database descriptors, but rejected otherwise. The parameter ρ called the rejection ratio should be in the range $0 < \rho < 1$. The optimal value $\rho = 0.8$ is recommended in [Lowe 2004*a*].

3 Code evaluation

Three simulations have been conducted for code verification. The first two simulations are used to test keypoint detection, and the third one for keypoint detection and descriptor matching.

Figure A1 shows a synthetic image generated by VIRSuite in the visible spectrum. The image size is 1024-by-1024 pixels. The image is first converted to grey scale, and its intensity is normalized. The SIFT keypoint detection algorithm has been applied to the resulting image with the default control parameters recommended in [Lowe 2004*a*]. The number of detected keypoints is equal to 3,524. An image of the scale-space pyramid is shown in Figure A2. The keypoint locations (red dots) and orientations (green arrows) are shown in Figure A3. It is seen multiple orientations assigned to some locations. Figure A4 shows the histogram of keypoint scales. It is seen from the histogram that the majority of keypoints are detected at relatively small scales.

Figures A5–A8 show similar information for a synthetic image generated by VIRSuite in the infrared spectrum. The infrared image size is also 1024-by-1024 pixels. The number of detected keypoints is equal to 309 which significantly smaller than the number in the previous example.

The last simulation recovers a similarity-transformed image by means of an estimated two-dimensional homography. The grey-scale image of a camera man (`cameraman.tif`) has been loaded. This image is supplied with MATLAB for demonstration purposes, and its size is 256-by-256 pixels. Another image is obtained from this image by a similarity transformation; namely, it is scaled by 90% and rotated anticlockwise by 5 degrees. The original and modified images are shown in Figure A9. Keypoints have been detected in both images and their descriptors matched. The rejection ratio is equal to 80%. The putative matches are shown in Figure A10, which definitely contain outliers. A two-dimensional homography has been estimated from the putative matches using the RANSAC algorithm. This procedure is described in detail in [Hartley & Zisserman 2003]. The number of inlier matches is equal to 92 out of 106 putative matches. The inlier correspondences are shown in Figure A11. Using the estimated homography, the modified image is transformed back. The original and recovered images are shown in Figure A12.

4 Discussion

The SIFT algorithm has been described and implemented in MATLAB. The conducted numerical simulations demonstrated that it works reasonably well. The current version of the code is not optimized yet. Proper optimization can be done in a multi-threaded environment.

5 References

Hartley, R. & Zisserman, A. (2003) *Multiple View Geometry in Computer Vision*, 2nd edn, Cambridge University Press, Cambridge.

Lowe, D. G. (1999) Object recognition from local scale-invariant features, *in Proc. Int. Conf. Computer Vision*, Vol. 2, pp. 1150–1157.

Lowe, D. G. (2004*a*) Distinctive image features from scale-invariant keypoints, *in Int. J. Computer Vision*, Vol. 60, pp. 91–110.

Lowe, D. G. (2004*b*) Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image. US Patent 6,711,293.

Swierkowski, L., Christie, C. L., Antanovskii, L. K. & Gouthas, E. (2014) Real-time scene and signature generation for ladar and imaging sensors, *in Proc. SPIE 9071 Infrared Imaging System: Design, Analysis, Modeling, and Testing XXV*, Vol. 90711E, Baltimore, USA.

THIS PAGE IS INTENTIONALLY BLANK

Appendix A: Figures

Figure A1: Synthetic image 1 generated by VIRSuite in visible band

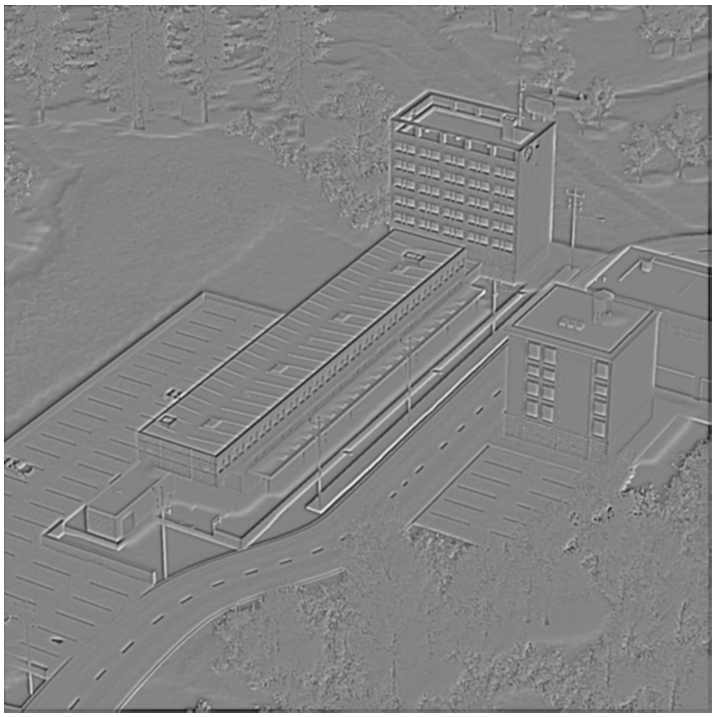

Figure A2: Difference-of-Gaussian image 1 for $\sigma = 3.2$

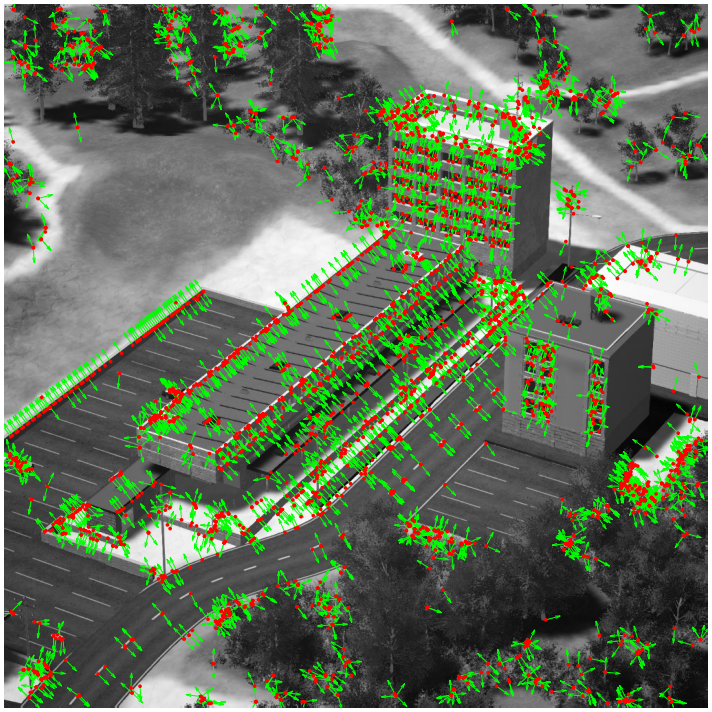

Figure A3: Keypoint locations and orientations in image 1

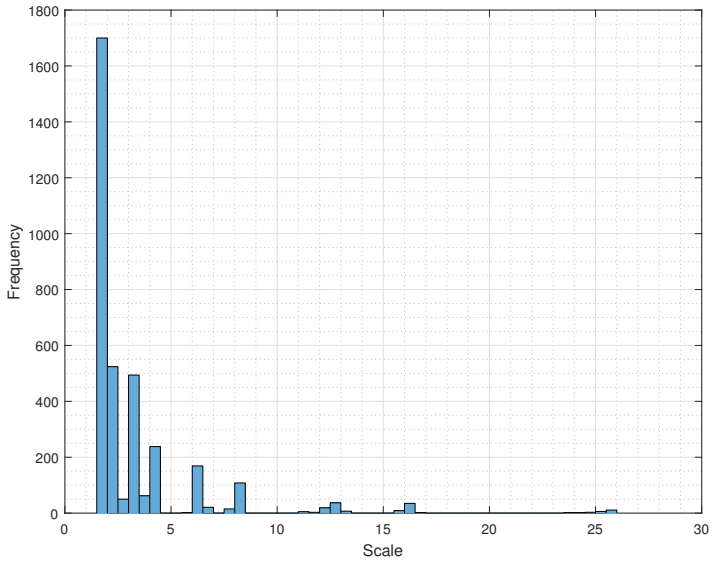

Figure A4: Histogram of keypoint scales in image 1

Figure A5: Synthetic image 2 generated by VIRSuite in infrared band

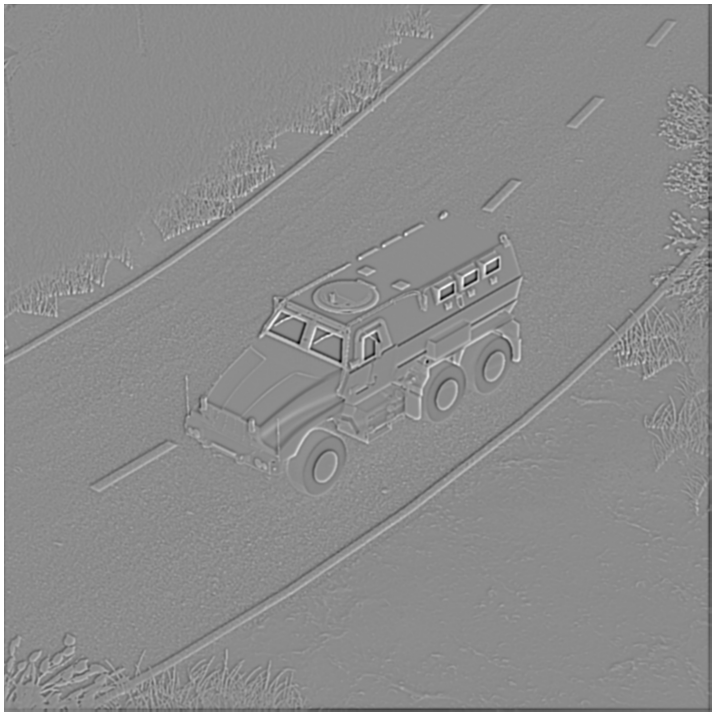

Figure A6: Difference-of-Gaussian image 2 for $\sigma = 3.2$

Figure A7: Keypoint locations and orientations in image 2

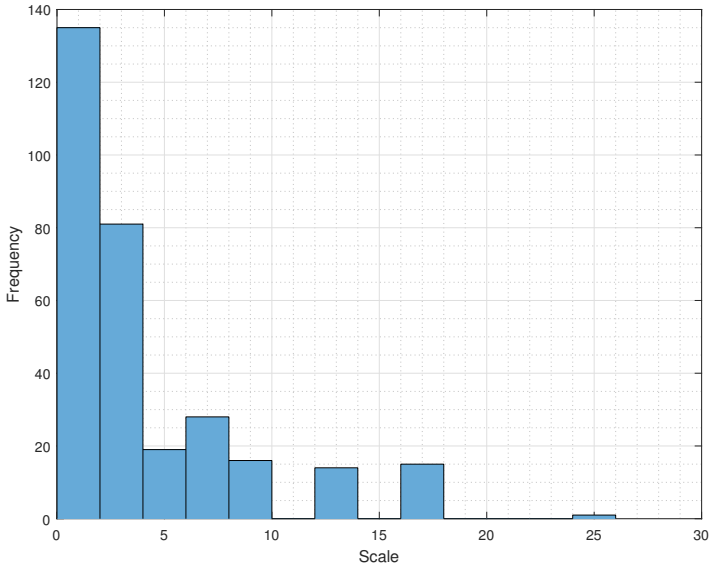

Figure A8: Histogram of keypoint scales in image 2

Figure A9: Original and transformed images

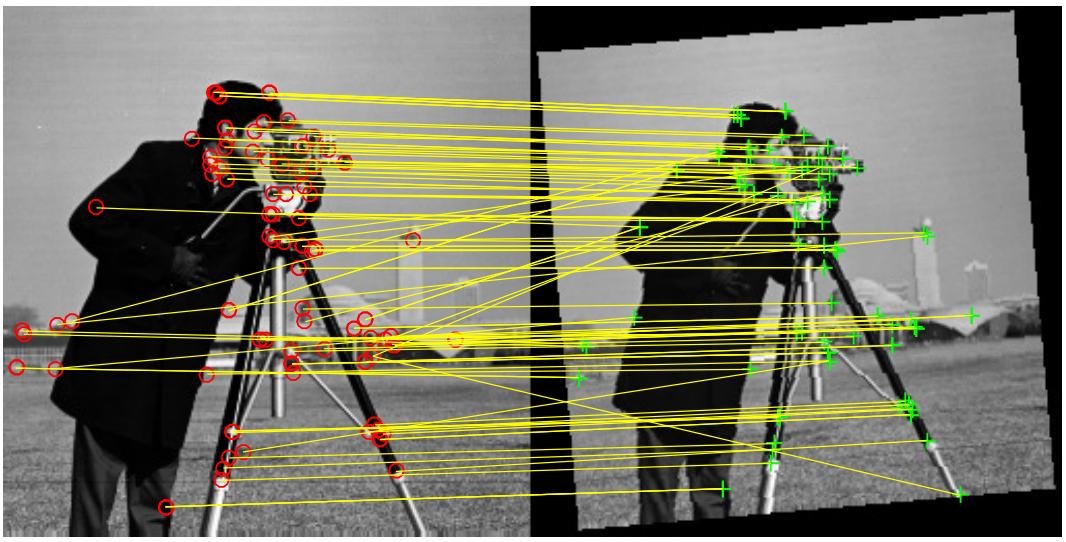

Figure A10: Putative matches

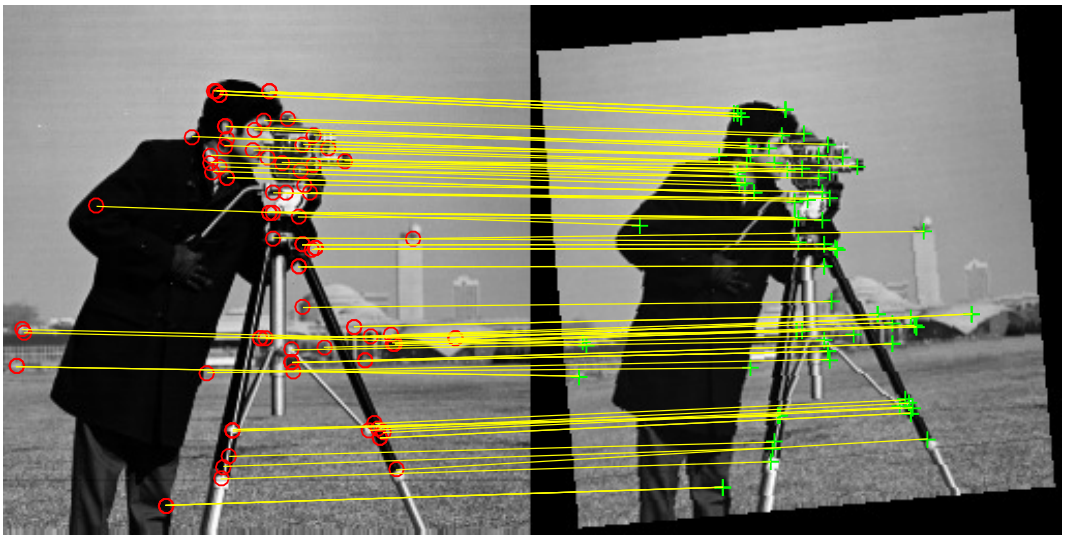

Figure A11: Inlier matches

Figure A12: Original and recovered images

UNCLASSIFIED

DEFENCE SCIENCE AND TECHNOLOGY GROUP DOCUMENT CONTROL DATA			1. DLM/CAVEAT (OF DOCUMENT)
2. TITLE Implementation of the Scale Invariant Feature Transform Algorithm in MATLAB®		3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)	
4. AUTHOR Leonid K Antanovskii		5. CORPORATE AUTHOR Defence Science and Technology Group PO Box 1500 Edinburgh, South Australia 5111, Australia	
6a. DST Group NUMBER DST-Group–TR–3347	6b. AR NUMBER 016–811	6c. TYPE OF REPORT Technical Report	7. DOCUMENT DATE February, 2017
8. Objective ID AV12877489	9. TASK NUMBER AIR07/213	10. TASK SPONSOR RAAF Air Combat Group	
13. DST Group Publications Repository `http://dspace.dsto.defence.gov.au/dspace/`		14. RELEASE AUTHORITY Chief, Weapons and Combat Systems Division	
15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT *Approved for public release* OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SOUTH AUSTRALIA 5111			
16. DELIBERATE ANNOUNCEMENT			
17. CITATION IN OTHER DOCUMENTS No Limitations			
18. RESEARCH LIBRARY THESAURUS Science, Mathematics, Algorithms, Computer Vision, Object Recognition			
19. ABSTRACT This report addresses the description and MATLAB implementation of the Scale-Invariant Feature Transform (SIFT) algorithm for the detection of points of interest in a grey-scale image. Some illustrative simulations for code verification are conducted.			

UNCLASSIFIED