Evolving object detectors for infrared imagery: a comparison of textural analysis against simple statistics.

Daniel Howard and Simon C. Roberts

Software Evolution Centre, dhoward@dera.gov.uk, Tel:(44)(1684)894480,
Building U50, Systems & Software Engineering Centre,
Defence Evaluation and Research Agency,
Malvern, WORCS WR14 3PS, UK.

Abstract. A multi-stage Genetic Programming method evolves a detector of vehicles for the automated inspection of the infrared line scan imagery obtained by low flying aircraft. Two sets of GP terminals, i.e. both are irrotational statistics of pixel data, are compared on ten images. The discrete Fourier transform set is found to be marginally superior to the simpler statistics set that includes an edge detector. It is also shown that an anlysis of detector formulae can reveal detection principles.

1 The problem and its practical importance

In airborne reconnaissance by low flying aircraft, vast amounts of imagery taken by the aircraft must be analysed at a workstation. Analysts cannot visually inspect all of the imagery in allowable time, and would benefit from a fast automatic way to bring promising areas to their attention.

Genetic Programming is used to evolve a vehicle detector for infrared line scan (IRLS) images using examples of 'object' and 'non-object'. The detector is an algebraic function of pixel information that is applied pixel by pixel to an entire image.

2 The staged Genetic Programming method

A multi-stage GP method was presented in [1–3] to evolve fast and accurate detectors in short evolution times. Given a rectangular image of dimensions N x $M = O(10^6)$ pixels containing objects that can be bounded in sub-windows of size n x $n = O(10^2)$ pixels, the task is to construct a function of pixel data with support n x n. An object is detected at a pixel when the function evaluation is positive.

A number of images is selected with a 'truth' of known object locations prepared for each image. A true positive (TP) results when the detector correctly detects an object from the truth; a false positive (FP) occurs when the detector incorrectly detects an object - also known as a 'false alarm'. Similarly a true negative (TN) is a correct non-object detection and a false negative (FN) is an object missed.

Evolution of the detector to discriminate the object pixels from all of the non-object pixels in the image would involve evaluations at all of the N x M pixels and would be prohibitively expensive.

Instead, a first stage of GP takes a random selection of non-object pixels and all the object pixels from the truth as test points. The fittest detector from this evolution is applied to the N x M pixels of each image. And this produces a set of FP. A second stage of GP, uses the discovered FP and all of the object pixels from the truth as test points to evolve a second detector. It has a tougher job than the first stage because it must discriminate like from like.

The fittest detector from both stages are combined, i.e. an object is detected only when both return a positive value. The first detector is applied to all N x M pixels of an image and the second detector only to the positive returns of the first - a small number compared to N x M.

In order to detect objects having large variability this two-stage GP process was modified [4] by decomposing the second stage into a number of sub-stages, e.g. $2i$, $2ii$, $2iii$,... Sub-stage $2i$ tries to eliminate all the FP while hitting as many targets as possible. Sub-stage $2ii$ takes the objects missed by $2i$ and tries to detect as many of them while eliminating all the FP. The sub-stage process stops when there are enough sub-detectors to detect all objects. This is illustrated in Figure 4.

The overall detector becomes:

`if value of 1st AND (2i OR 2ii OR ...) detectors` $>$ `0 THEN pixel is an object.`

Detectors evolved this way can challenge the Grossberg stability-plasticity dilemma [5] because the overall detector can learn new objects without forgetting what it has learned in the past.

3 GP specifications

Ten IRLS images were used to evolve the detectors with GP terminals and functions as in Table 1. GP 'statistic' terminals were pre-computed at the beginning of the run and stored for use by GP as in [6]. These corresponded to two sets of experiments. The first used simple statistics, while the second used textural statistics, i.e. 1D discrete Fourier transform (DFT). In this steady-state GP with cross-over by tournament selection, a constant terminal could mutate to another of any constant type. The raw image consisted of 8 bit grey-scale pixels. Each image was normalised to have a mean of 64.0 and standard deviation of 16.0 to guarantee certain uniformity.

Table 1. GP parameters.

Parameter	Description	Parameter	Description
Terminals	*all stages*:255 allowed integers (-127:128);	α	was varied between 1 and 6
	+ve reals (0.005:1.0) in 0.005 steps; and	β	was varied between 1 and 7
	-ve reals (-1.0:-0.005) in 0.005 steps.	Population	5000 (all stages).
	1st and 2nd stages: mean, s.d. and	Max nodes	1000 (all stages).
	edge stats - section 3.1	Mate radius	5000 (all stages).
	2nd stages: F_{Di} - section 3.2	Kill tourn.	size 2: for steady-state GP
Functions	+, -, *, protected %, min, max.	Breed tourn.	size 4: for steady-state GP.
Genetic	95% cross-over; 5% mutation of constants.	Max gens	10 (1^{st} stage), 20 (2^{nd} stages).

3.1 Simple statistics

Rotationally invariant statistics of pixel information were defined on concentric circles of single pixel thickness as in Figure 1. These were drawn with a diameter in pixels, e.g 11, with statistics defined on the pixel values. The circumference pixel positions were computed from $r^2 = x^2 + y^2$ but using integer arithmetic such that every pixel position only had two neighbouring pixel positions, i.e. corners cannot occur. Four types of irrotational statistics were defined on each circle: a circumference average, P_D; the number of edges found on the circumference, N_D; a circumference standard deviation, S_D; and an edge distribution measure, E_D, where the subscript D stands for 'diameter'. For example, for a diameter of 33 pixels the statistics are: P_{33}, S_{33}, N_{33} and E_{33} respectively. The circles were scaled according to the altitude of the aircraft when the image was taken. Hence, the circle diameters: 11, 23, 33, and 45 apply to images taken at the reference height of 300ft.

The number of edges, $N_D - 1$, was computed by taking the difference between consecutive pixels on the ring, and comparing against a threshold. An edge was identified when the difference between consecutive pixels on a given ring exceeded half the standard deviation of the entire image. The edge algorithm took account of inflexion as illustrated in Figure 1. Always N_d equaled the number of detected edges plus one to avoid $N_D = 0$ if no edges were detected.

The edge distribution measure E_D was defined as:

$$E_D = \sum_{i=1}^{N_D-1} (\frac{s_i}{L})^2 \tag{1}$$

where s_i is the arc length between the ith and the (i-1)th edges, L is the total arc length or circumference; if no edges were detected E_D was set to 1 - both for consistency and to prevent $E_D = 0$.

This choice of statistics provided the evolutionary process with: (a) an irrotational character; (b) robustness to high variability in pixel data; and (c) both angular and radial information.

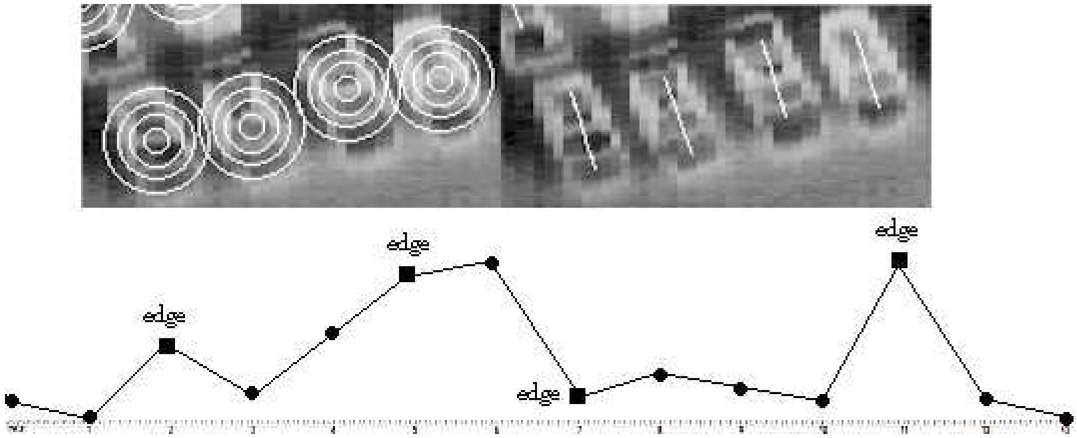

Fig. 1. Graphical illustration of edge detector statistic along the circumference of the circle. At each pixel a set of rules is applied to decide whether it is an edge. These rules take into account the magnitude of the jump and whether or not there is an inflexion. In this illustration which covers a good number of cases, the algorithm detects four edges, resulting in $N_D = 5$. Top: circles on vehicles and vehicles defined with lines of pixels.

3.2 Textural statistics

Second stage detectors were evolved using two types of statistics: simple, and DFT based on the 1D discrete Fourier transform, e.g. pg. 40-55 in [7] - over pixels lying on the circumference of a circle, thus giving a one dimensional transform.

GP is required to detect a vehicle regardless of the corresponding altitude and thus similar vehicles at different altitudes should produce similar terminal values. As images corresponding to lower altitudes contain more detail than those corresponding to higher altitudes, they have a stronger high frequency content. For this reason, only low frequency information was taken from the DFTs.

Consider a circle with a circumference of N pixels giving a DFT with N spectral coefficients. As pixels are real valued, i.e. they have no imaginary component, the DFT gives symmetrical spectral coefficients such that the coefficient for frequency (N/2)+i is equal to the coefficient for frequency (N/2)-i. Therefore, the number of unique coefficients is (N/2)+1. The smallest circumference gave 7 unique coefficients and so the 7 lowest frequency coefficients were stored for each DFT calculation.

The actual value output by a terminal node F_D, where the subscript D is the circle diameter, was governed by an associated variable which was set to a random value between 0 to 6 inclusively when the node was created. The statistic F_{Di} was defined as the ith spectral coefficient of a DFT of circle diameter D. The value 0 gave the lowest frequency coefficient (i.e. the steady-state or background level) and 6 gave the highest frequency coefficient of the 7 stored coefficients. As with the simple statistics the four reference rings were defined at D values of: 11, 23, 33 and 45 (at 300ft) and scaled with altitude.

3.3 Fitness measure

The fitness measure f_M and a figure of merit FOM were computed with:

$$f_M = \frac{\alpha TP}{n_s + \beta FP} \qquad\qquad FOM = \frac{TP}{n_t + FP}$$

where n_s is the number of vehicles of the truth for a GP stage, while $n_t = 93$ is the total number of vehicles in the truth (see Figure 3 and Table 7); the parameters α and β control the balance between the miss rate and the false alarm rate. The optimum parameter settings were unknown before the evolution runs and hence each detector stage was evolved with multiple settings. Figure 2 shows the miss rate versus the false alarm rate for multiple runs, with different randomiser seeds, at given α and β settings. A higher α:β ratio typically gave a lower miss rate at the expense of a higher false alarm rate.

The α and β settings depended on the required task for a given detector. For example, the 1st detector was required to hit all vehicles hence α:β was high. However, the 2i detector should minimise the false alarm rate hence α:β was low, whereas later 2nd detector stages were required to hit all remaining

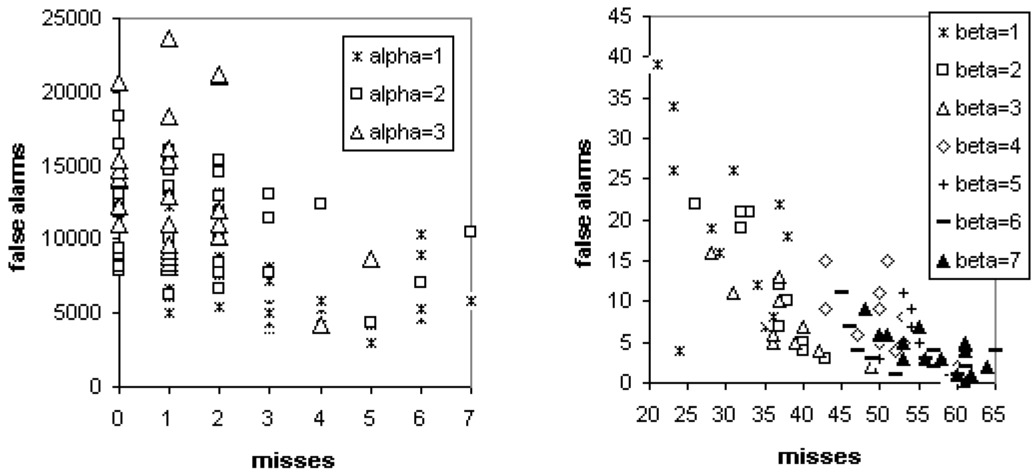

Fig. 2. Evolution behaviour of (FN,FP) with different α and β; (left) 1st GP stage, (right) 2i stage for simple statistics. It is very difficult to control the exact (FN,FP) result with an α and β choice.

missed vehicles hence α:β was high. Furthermore, the settings reflected the balance between vehicle and non-vehicle training data. For example, each 2nd detector processes *all* the false alarms returned by the 1st detector, but 2nd detectors process successively fewer vehicles.

4 Interpretation of Evolved Detectors

Using a *Mathematica* back-end [8] we simplified the fittest detector formulae and traced the paths taken by detectors, i.e. down the min and max branches in GP trees, to extract the expressions responsible for the object detection.

The complexity of the best detectors was substantial. However, as shown in Table 2, it was always possible to find a detector with similar performance as the best one but smaller in size. The luck in the random seed is the main reason for the large size difference. A detector formula of such small size can be understood, mainly by grouping terms and observing the value of differences between such terms.

4.1 Example

The simplified version of the substitute for the 1st phase in Table 2 is:

$$\begin{aligned} \texttt{IF} \quad min(N_{11}, g_1) \geq 0 \quad \texttt{THEN VEHICLE}. \\ g_1 = -N_{11} - N_{23} - P_{45} + max(E_{45}, P_{33}) + min(g_2, g_3) \\ g_2 = -E_{11} - N_{45}/N_{23} + S_{33} - S_{45} \\ g_3 = max(-0.155, -17.155 + N_{45} - S_{33}) \end{aligned} \tag{2}$$

Table 2. Best detector and its substitute.

		BEST					SUBSTITUTE				
statistic	stage	TP	FP	FOM	after gens	size	TP	FP	FOM	after gens	size
simple	1st	93	9431	0.009765	10	333	93	11937	0.007731	10	31
simple	2i	69	4	0.7113	20	601	56	12	0.5333	20	93
	2ii	84	10	0.8155	20	239	79	6	0.7980	20	63
	2iii	90	10	0.8738	20	125	89	12	0.8476	20	73
DFT	2i	73	12	0.6952	20	185	66	14	0.6168	20	57
	2ii	87	14	0.8131	20	437	81	21	0.7105	20	73
	2iii	93	15	0.8611	20	103	-	-	-	-	-

This detector has two min operations and two max operations that offer six possible paths, however, only three paths are observed when the detector is applied over the ten images:

$$N_{11} \tag{3}$$

$$-N_{11} - N_{23} - P_{45} + P_{33} - 0.155 \tag{4}$$

$$-N_{11} - N_{23} - P_{45} + P_{33} - E_{11} - N_{45}/N_{23} + S_{33} - S_{45} \tag{5}$$

And the count for the paths by return type is given in Table 3. As expected equation (3) is only to be found for positives since the N_D statistics is by definition positive, hence it does not discriminate. Table 3 reveals that equation (5) is mostly responsible for discrimination between target and non-target. This equation is not simply a 'spot' detector of bright light, as are most infrared preliminary filters normally used for this type of work, but instead it represents a subtle balance between a differences of means, i.e. $P_{33} - P_{45}$, a difference of standard deviations, and a positive term whose magnitude is a sum of edge counts. The term E_{11} is positive and is usually equal to one. When looking at the data one sees mostly a competition between the difference in P_D and the sum of N_D (which acts as a variable positive threshold) though at times the difference in S_D is larger and can settle the outcome.

Table 3. Pixel count for detector paths: 1st detector.

All			TP			FN			FP			TN		
eqn.	pixels	percent	eqn.	pixels	percent	eqn.	pixels	percent	eqn.	pixels	percent	eqn.	pixels	percent
(5)	7817064	96.6%	(3)	6839	56.5%	(5)	112336	85.8%	(3)	5011	42%	(5)	7696344	97%
(4)	264380	3.3%	(5)	3621	29.9%	(4)	18605	14.2%	(5)	4763	39.9%	(4)	241975	3%
(3)	11850	0.1%	(4)	1637	13.5%				(4)	2163	18.1%			

5 Detector performance

Table 4 gives the average and standard deviation of TP, FP, FOM and hit coverage (i.e. hits per vehicle) for evolving 1st detectors with various α and β settings and randomiser seeds. It can be seen that TP was consistently high whereas FP was more variable with randomiser seed and tended to increase with α as can also be seen from Figure 2. As shown in Figure 4 the 1st detector was required to give a high hit coverage per vehicle to provide 2nd detectors with more opportunities for discriminating vehicles from the false alarms. Table 4 shows that higher α gave higher hit coverage as expected.

Table 5 and gives the results for evolving 2nd detectors with various α and β settings and randomiser seeds. It can be seen that the results were more variable for 2i detectors as these hit the majority of vehicles. Note that all results presented in this section were obtained by running a given configuration with at least 10 different randomiser seeds.

Table 4. 1st detectors obtained after 10 generations, and which trained on 12246 pixels, effect of α and β.

α	β	TP		FP		FOM		coverage	
		avg.	s.d.	avg.	s.d.	avg.	s.d.	avg.	s.d.
1	1	90.18	1.73	7257.82	2524.86	0.013782	0.004748	85.50	14.15
2	1	91.38	1.66	11335.33	3136.70	0.008744	0.002940	112.07	14.15
3	1	91.82	1.17	13260.96	4480.28	0.007787	0.003172	115.29	26.13

6 Processing simple statistics versus DFTs

The overall performance of detectors processing simple statistics was similar to that for detectors processing DFTs. This suggests that these different types of terminals identify similar qualities from the images.

Table 5. Superiority of DFT statistics in multi-stage GP scheme for IRLS vehicle detection, effect of α and β.

discrete Fourier transform									simple statistics								
stage	α	β	TP		FP		FOM		stage	α	β	TP		FP		FOM	
			avg.	s.d.	avg.	s.d.	avg.	s.d.				avg.	s.d.	avg.	s.d.	avg.	s.d.
2i	1	1	66.62	2.62	17.15	7.46	0.6062	0.0264	2i	1	1	62.92	5.65	19.23	9.92	0.5615	0.0539
	1	2	61.23	6.94	10.69	8.82	0.5898	0.0347		1	2	57.20	4.77	12.40	7.29	0.5424	0.0169
	1	3	57.67	4.73	9.83	7.06	0.5609	0.0274		1	3	55.50	5.46	7.90	4.20	0.5492	0.0387
2ii	1	1	82.00	1.81	16.82	2.75	0.7469	0.0179		1	4	42.50	4.50	8.08	3.90	0.4201	0.0379
	2	1	84.65	1.96	25.75	6.79	0.7143	0.0278		1	5	37.89	3.57	4.89	3.35	0.3868	0.0315
	3	1	84.89	1.79	30.28	9.30	0.6915	0.0395		1	6	37.92	6.33	3.62	2.68	0.3918	0.0592
	4	1	86.89	2.56	38.44	11.98	0.6649	0.0430		1	7	36.36	4.88	3.93	2.46	0.3743	0.0433
2iii	1	1	87.60	1.56	14.40	3.55	0.8187	0.0390	2ii	1	1	80.59	2.00	9.76	4.92	0.7851	0.0217
	2	1	89.78	0.79	13.44	2.73	0.8438	0.0187		2	1	84.26	1.65	22.37	7.64	0.7330	0.0412
	3	1	90.17	1.04	15.47	3.23	0.8321	0.0277	2iii	2	1	87.50	0.95	11.25	2.58	0.8396	0.0137
	4	1	90.63	0.77	17.18	5.89	0.8245	0.0357		3	1	87.71	0.95	11.57	2.16	0.8390	0.0133
	5	1	91.27	0.75	22.18	8.32	0.7962	0.0526		4	1	87.75	0.97	14.75	6.35	0.8165	0.0369
										5	1	87.75	0.97	16.50	8.65	0.8054	0.0517
										6	1	88.30	0.78	25.20	12.25	0.7540	0.0678

Table 6. Data sheet for Table 5.

	discrete Fourier transform		simple statistics	
stage	pixels trained on	for generations	pixels trained on	for generations
2i	19368	20	19368	20
2ii	10307	20	10641	20
2iii	9672	10	9675	10

Indeed, the simple statistics include pixel average which is equal to the first frequency component in a DFT. However, the edge statistics explicitly capture shape information, whereas this information may only be captured implicitly by low frequency coefficients in a DFT. It is thought that the calculation of terminal values from the aforementioned concentric circles allows detectors to extract similar brightness and shape information regardless of whether they process the simple statistics or DFTs.

Regarding the 2i detector, Tables 5 shows that the DFT approach achieved marginally higher $FOMs$. However, it also shows that the best 2i detector produced significantly more false alarms using the DFT approach, which carries through to the 2ii results as TP and FP correspond to fusing 2ii detectors with the best associated 2i detector. Hence, the table shows that the FOM at the 2ii stage was typically lower using the DFT approach.

The DFT approach tended to segregate the vehicles better, such that the vehicles processed by 2iii detectors appeared to have more in common, i.e. they all corresponded to high altitude images. Conversely, the 2iii detectors for the simple statistic approach had to process a diverse set of vehicles, varying greatly in brightness and shape. Consequently, only the DFT approach succeeded to hit all vehicles, as shown in Table 5. The DFT approach may be better at vehicle segregation by corresponding altitude due to the difference in textural quality at different altitudes, as low altitude images have stronger higher frequency component resulting from the sharper clarity.

It was required that the image processing speed was minimised. Consequently, the DFT technique was unsuitable for evolving 1st detectors as it was far more computationally intensive than the simple statistics approach, being up to 12 times slower for a single image. Furthermore, the DFT technique generally produced a higher false alarm rate and lower hit coverage (i.e. hits per vehicle). The 2nd detectors only process the hits returned by the 1st detector which was typically 1% of an image, thus computation speed was less important for 2nd detectors.

Table 7 compares the performance of the best detectors for the simple statistics approach against those for the DFT approach. Note that the 1st detector was the same in each case and that it hit all vehicles, thus the number of vehicles in each image was equal to TP for the 1st detector. The boolean

outputs of the detectors (hit or miss) were fused together as follows:
`fused = 1st AND (2i OR 2ii OR 2iii)`

Therefore, the fused detector accumulated the second detector hits that were in common with first detector hits. Note that multiple second detectors could hit the same vehicle thus the second detectors' TP do not necessarily sum to the fused TP. The 2i detectors were influenced most strongly by images

Table 7. Progress through the stages; comparison between both statistics for each image.

image	detector	simple		DFT		image	detector	simple		DFT		image	detector	simple		DFT	
		TP	FP	TP	FP			TP	FP	TP	FP			TP	FP	TP	FP
no. 2	1	29	345	29	345	no. 7	1	1	940	1	940	no. 10	1	9	711	9	711
	2i	20	0	23	1		2i	1	1	1	0		2i	5	0	2	1
	2ii	6	1	9	0		2ii	0	0	0	0		2ii	3	1	3	0
	2iii	7	0	2	0		2iii	0	0	0	0		2iii	1	0	5	1
	fuse	28	1	29	1		fuse	1	1	1	0		fuse	9	1	9	2
no. 4	1	43	509	43	509	no. 8	1	1	1040	1	1040	no. 11	1	3	452	3	452
	2i	39	0	42	0		2i	0	0	0	2		2i	2	0	1	0
	2ii	17	0	2	0		2ii	1	0	0	1		2ii	1	0	1	0
	2iii	3	0	0	0		2iii	0	0	1	0		2iii	0	0	1	0
	fuse	43	0	43	0		fuse	1	0	1	3		fuse	3	0	3	0
no. 6	1	1	1299	1	1299	no. 9	1	3	1394	3	1394	no. 13	1	2	1832	2	1832
	2i	0	1	0	1		2i	0	1	1	3		2i	1	0	2	1
	2ii	1	0	1	0		2ii	1	1	2	0		2ii	0	1	0	1
	2iii	0	0	0	0		2iii	1	0	0	0		2iii	0	0	0	0
	fuse	1	1	1	1		fuse	2	2	3	3		fuse	1	1	2	2
no. 15	1	1	909	1	909												
	2i	1	1	1	3												
	2ii	1	2	0	0												
	2iii	0	0	0	0												
	fuse	1	3	1	3												

2 and 4 as these contained the majority of vehicles, as seen in Figure 3. It appears that the 2i detector evolved using the DFT approach focused on vehicles with stronger high frequency components, i.e. more contrast, often hitting on edges of vehicle roofs and windows. Conversely, the 2ii detector evolved using the DFT approach focused on relatively uniform vehicles, albeit bright or dark. Consequently, the 2iii detectors focused on vehicles in images corresponding to high altitudes (i.e. with weak high frequency components) without relatively large uniform areas.

The 2i detector evolved using simple statistics also appears to segregate high-contrast vehicles, hitting many of the same vehicles in similar locations as for the DFT approach. This demonstrates how the multi-stage approach allows the detectors to *discover* the characteristic features of vehicles. However, the 2i detector evolved using simple statistics hit more vehicles in images corresponding to high altitudes (e.g. images 10 and 11), resulting in the aforementioned diverse vehicle set for the 2iii detector.

7 Conclusions

This study compared two types of GP terminal on a set of ten IRLS images in order to determine which set of statistics was more useful to this problem - automated vehicle detection.

The DFT approach is marginally superior to the simple statistic approach achieving a higher hit rate. At each second stage it achieved a more consistent object segregation, allowing all objects to be detected. Conversely, for the simple statistics, the detector from the last second evolution stage failed to generalise across a more diverse set of objects.

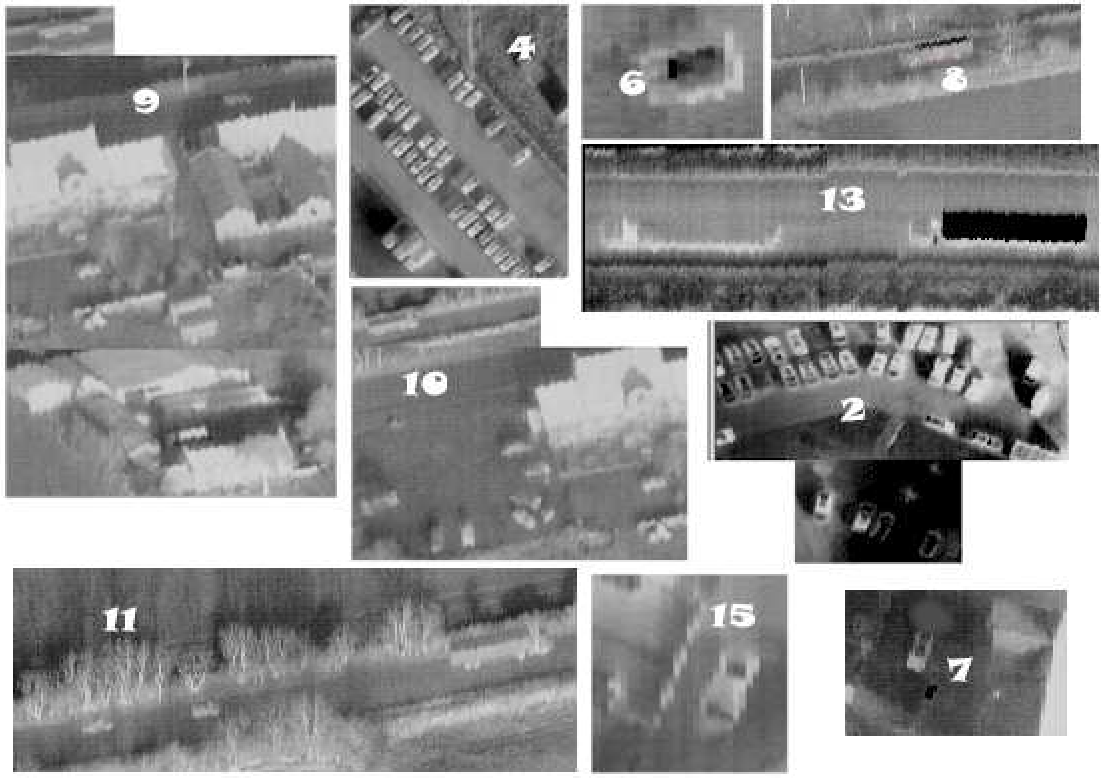

Fig. 3. Selected portions of the ten IRLS images containing vehicles; image numbers as in Table 7. Note that the image portions are not shown to scale and the relative size of their objects does not indicate altitude.

References

1. Howard D. and Roberts S.C. (1999) *Evolution of Ship Detectors for Satellite SAR Imagery*, Proc. Second European Workshop in Genetic Programming, Springer LNCS, Göteborg, Sweden, .
2. D. Howard, S. C. Roberts, R. Brankin (1998) *Target detection in SAR imagery by genetic programming*, Genetic Programming 1998: Late Breaking Papers, pp. 67-75.
3. Howard D., Roberts S.C., and Brankin R. (1999) Advances in Engineering Software, 30 (303-311).
4. Roberts S.C., and Howard D. (1999) *Evolution of Vehicle Detectors for Infrared Line Scan Imagery*, EvoIASP'99 and EuroEcTel'99 Joint Proceedings, pp. 111-125, Göteborg, Sweden, Springer LNCS.
5. Grossberg, S. (1982) *Studies of Mind and Brain*, Reidel, Boston.
6. Poli, R. (1996) *Genetic Programming for Image Analysis*. In Koza, John R., Goldberg, David E., Fogel, David B., and Riolo, Rick L. (editors). Genetic Programming 1996: Proceedings of the First Annual Conference, Cambridge, MA: The MIT Press.
7. Jähne, B.(1997) *Digital Image Processing*, Springer-Verlag, 4th Ed., ISBN 3-540-62724-3.
8. Mathematica v.3.0 software: ©1996 Wolfram Research, Inc.

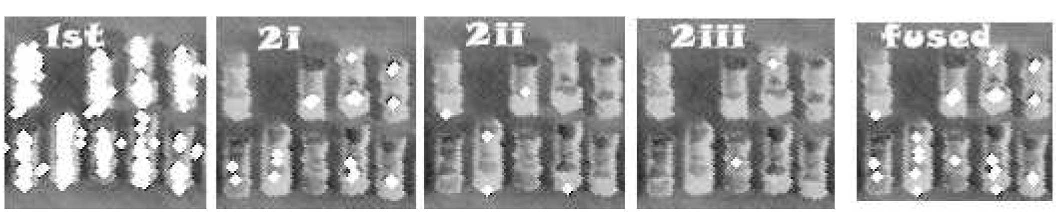

Fig. 4. Example of multi-stage detection (white areas).