A method for objective edge detection evaluation and detector parameter selection
***IEEE: Transactions on Pattern Analysis & Machine Intelligence* (2003)**

Yitzhak Yitzhaky and Eli Peli

Additional information and results

Contents

Multi-scale effects on correspondence

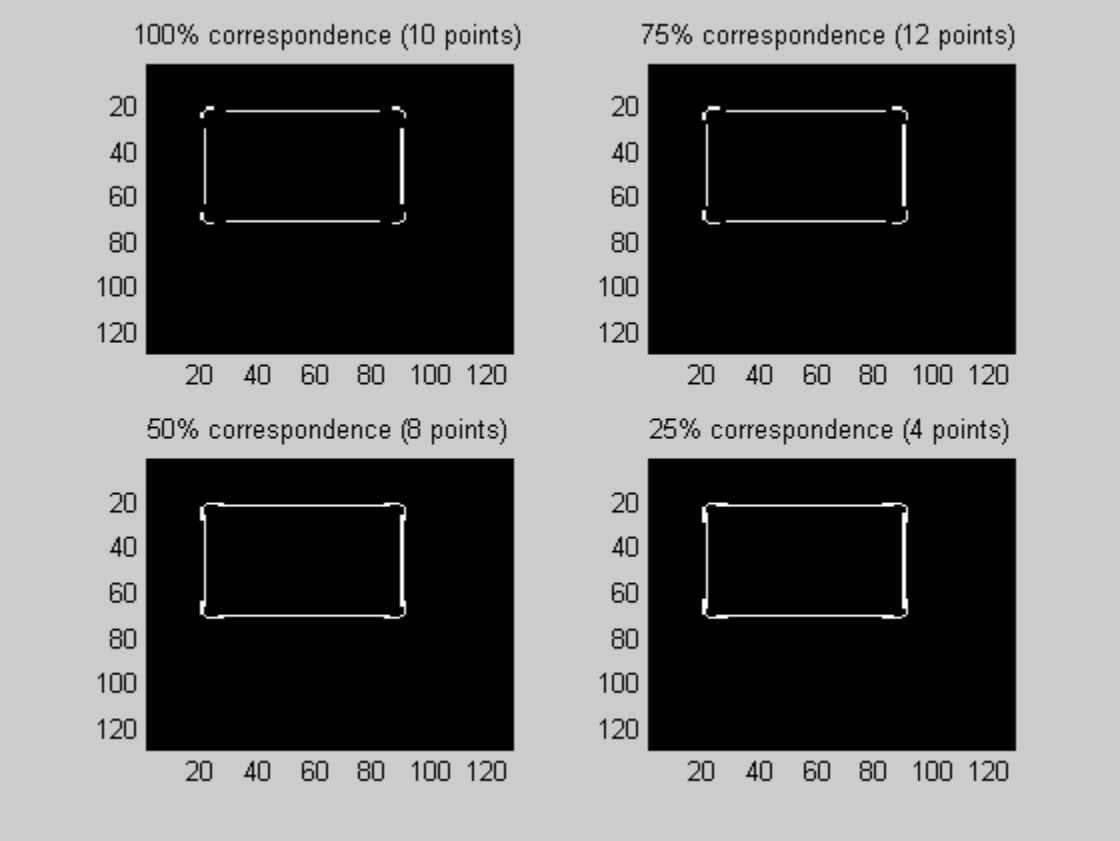


Fig. A1. Multi-scale effects on correspondence (LOG detector) with the same parameters used in the paper: Sharp (synthetic) corners result in miss-localization due to the effects of the smoothing operator. This is a deficiency of the detection method.

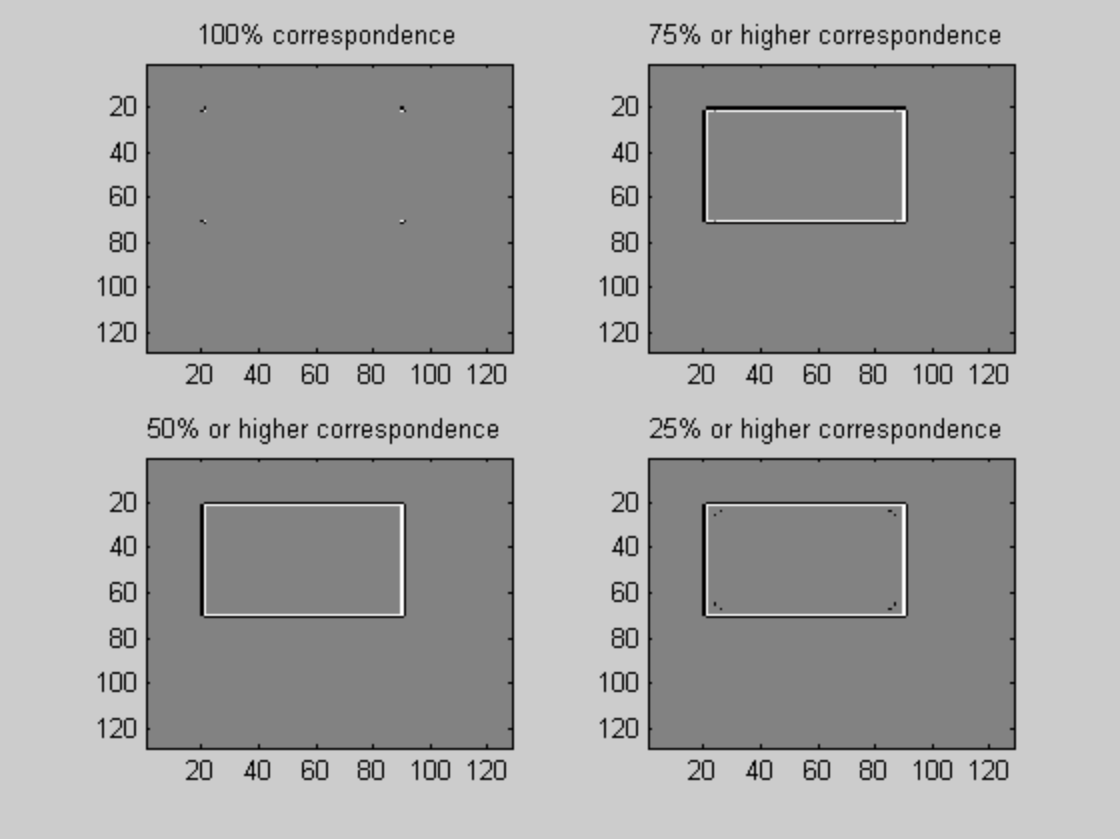


Fig. A2. Same as A1 but for Peli's detector. In this detector several scales are used at each point and therefore the multi-scale miss-localizations are less significant.

([11] E. Peli, "Feature detection algorithm based on a visual system model", *Proceedings of the IEEE*, Vol. 90, pp. 78-93, 2002.)

Examples for different correspondences for different images, showing that usually points with higher correspondence are related to more important features that represent contours of main objects in the image

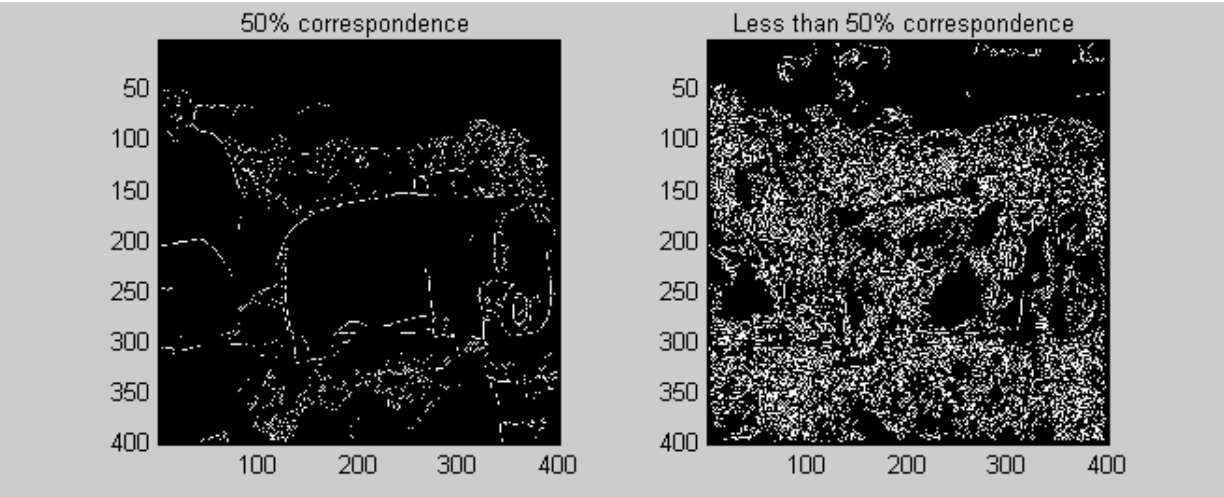


Fig. A3. Canny detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

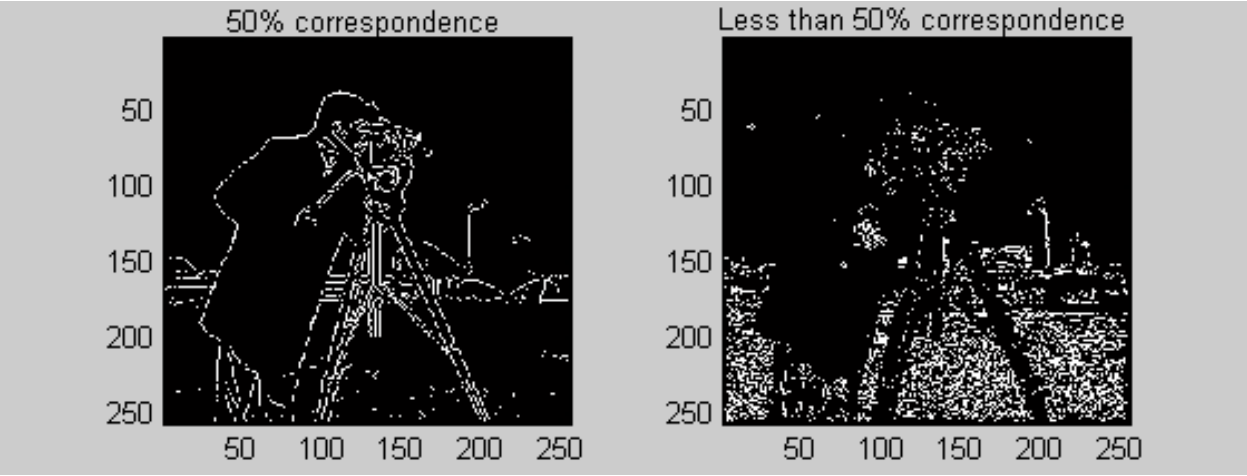


Fig. A4. Canny detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

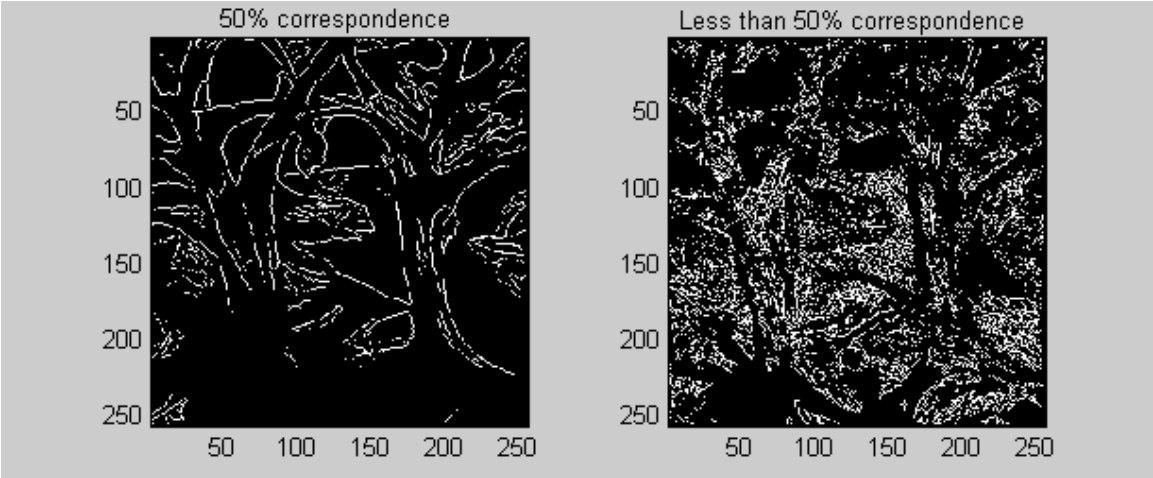


Fig. A5. Canny detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

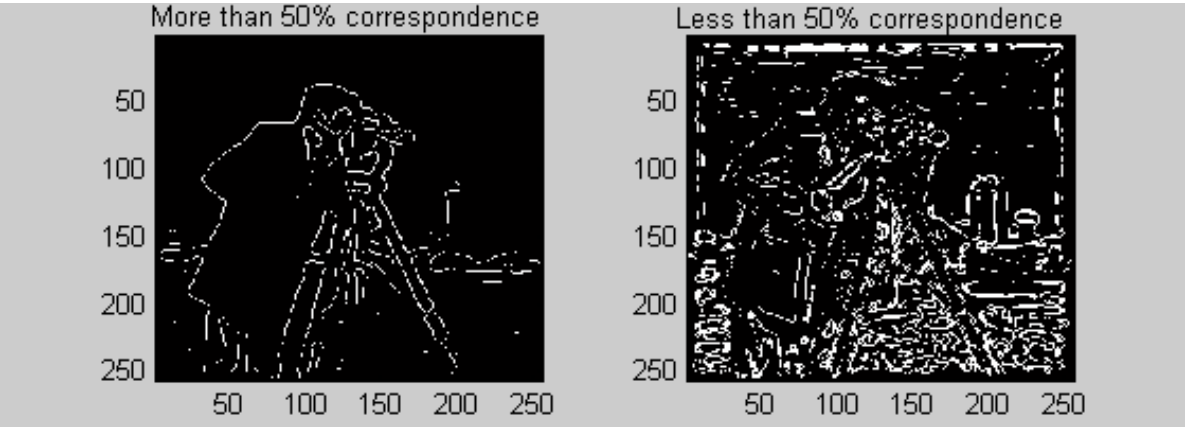


Fig. A6. LOG detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

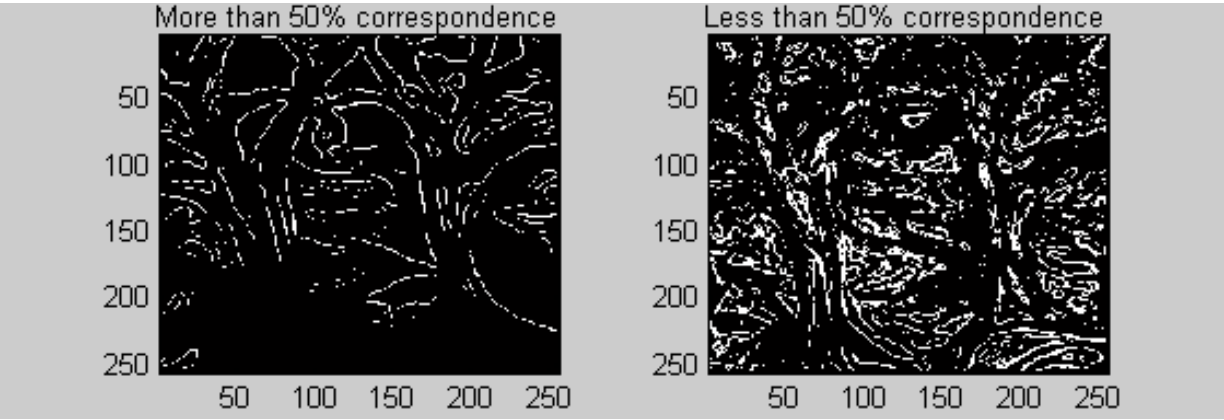


Fig. A7. LOG detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

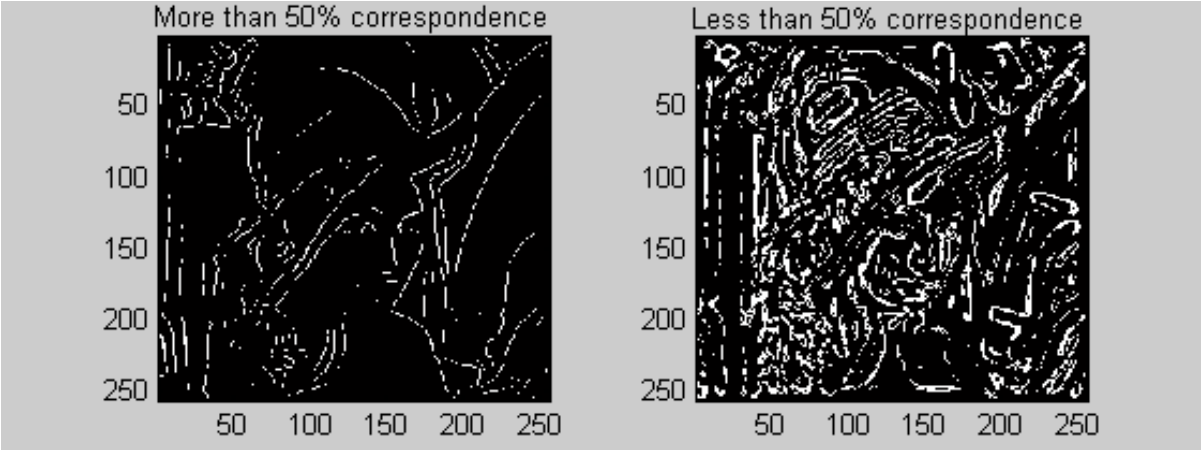


Fig. A8. LOG detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

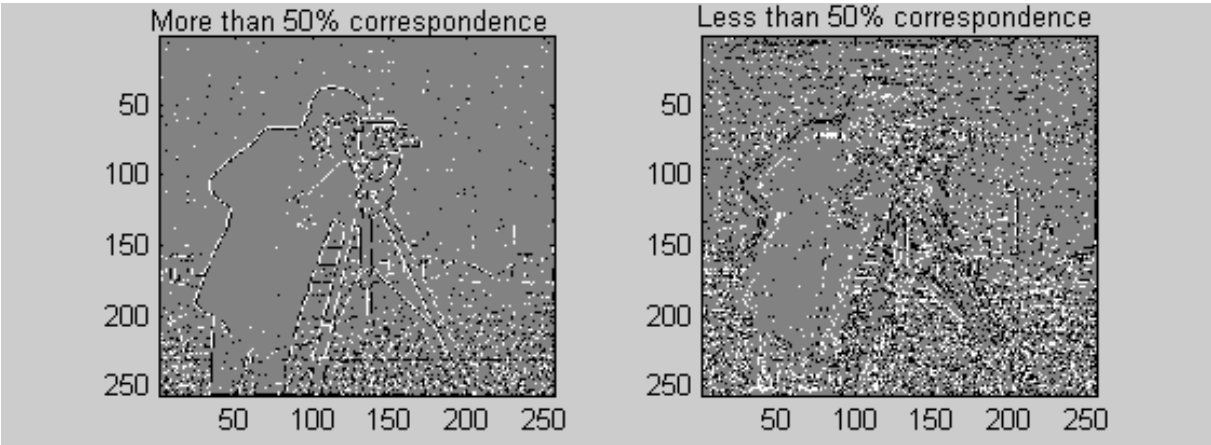


Fig. A9. Peli's detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

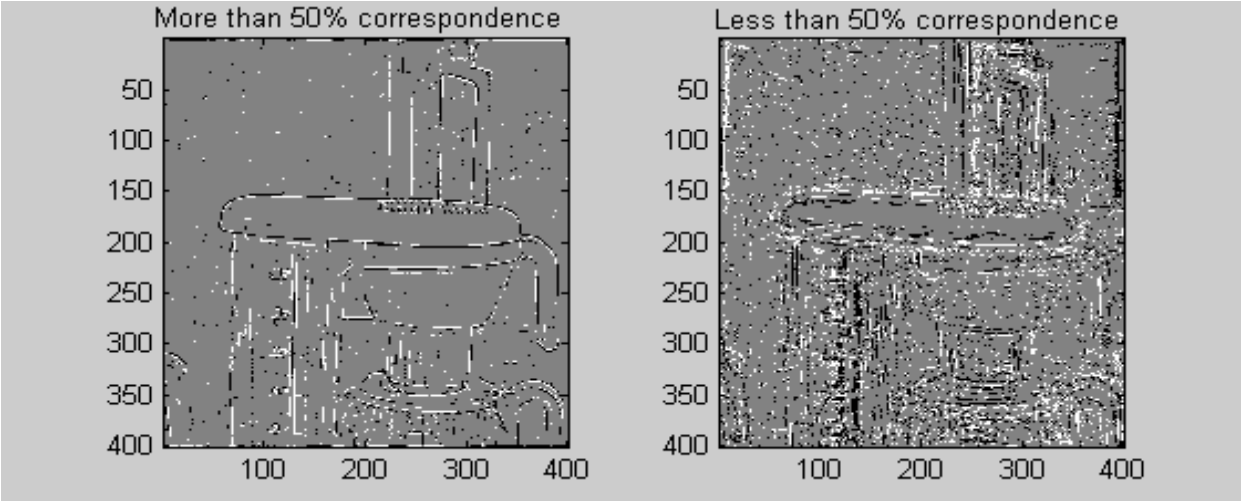


Fig. A10. Peli's detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

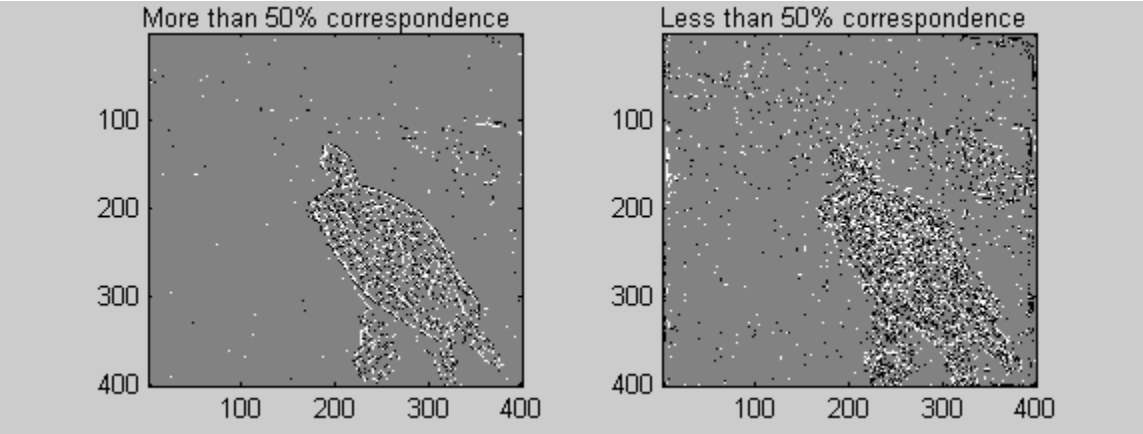


Fig. A11. Peli's detector: (a) Locations with more than 50% correspondence (between 9 and 16); (b) Locations with less than 50% correspondence (between 1 and 8). More important edges appear in (a) where higher correspondence is used.

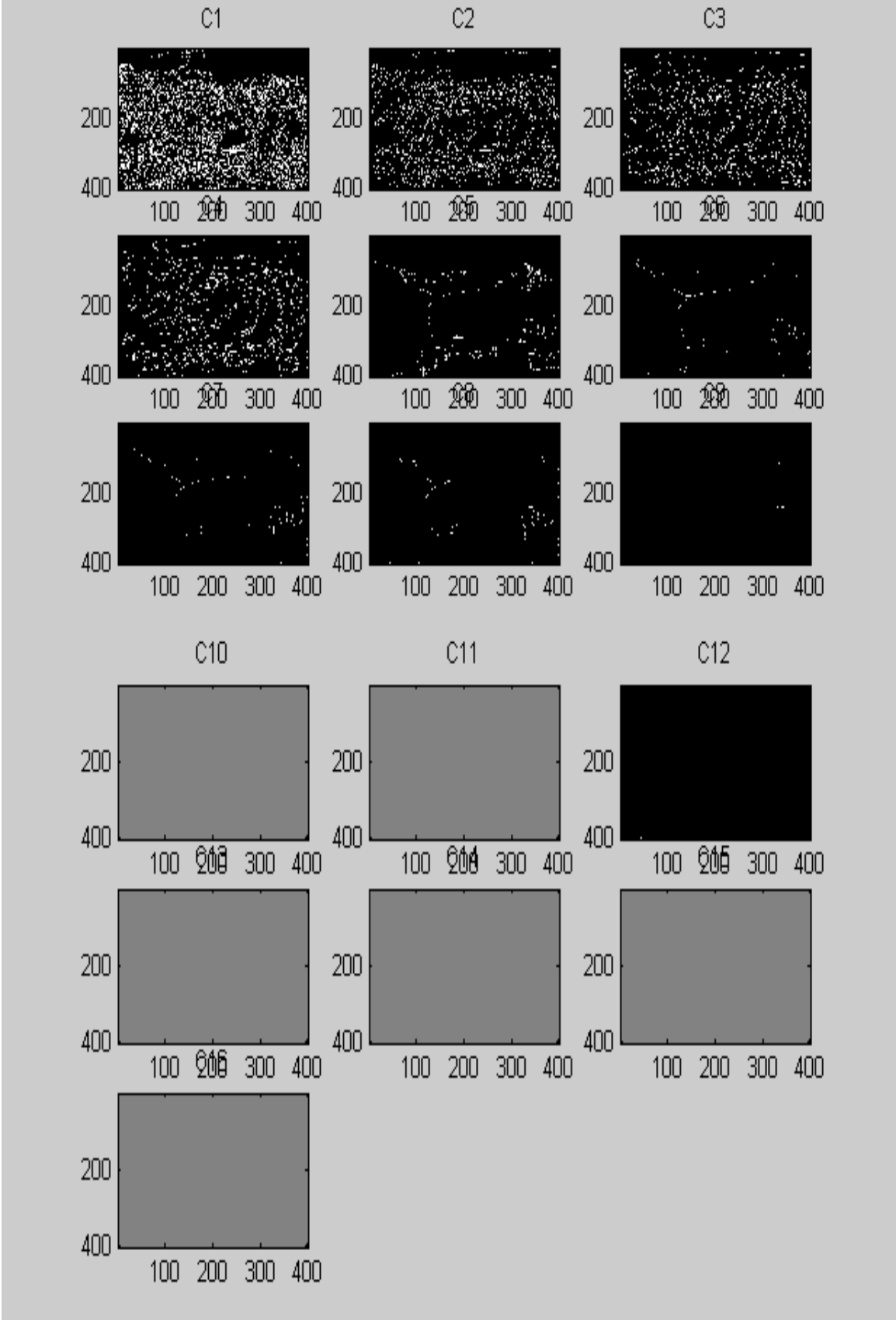

C1
C2
C3
C10
C11
C12
200
400
100
200
300
400

Fig. A12. The 16 detection results for parameter range between: “sigma” = [0.1 - 4.6] and high threshold = [0.05 – 0.95]. In about half of these results the parameters are ‘out of range’. (Detection results are not clear because the subfigures are small.)

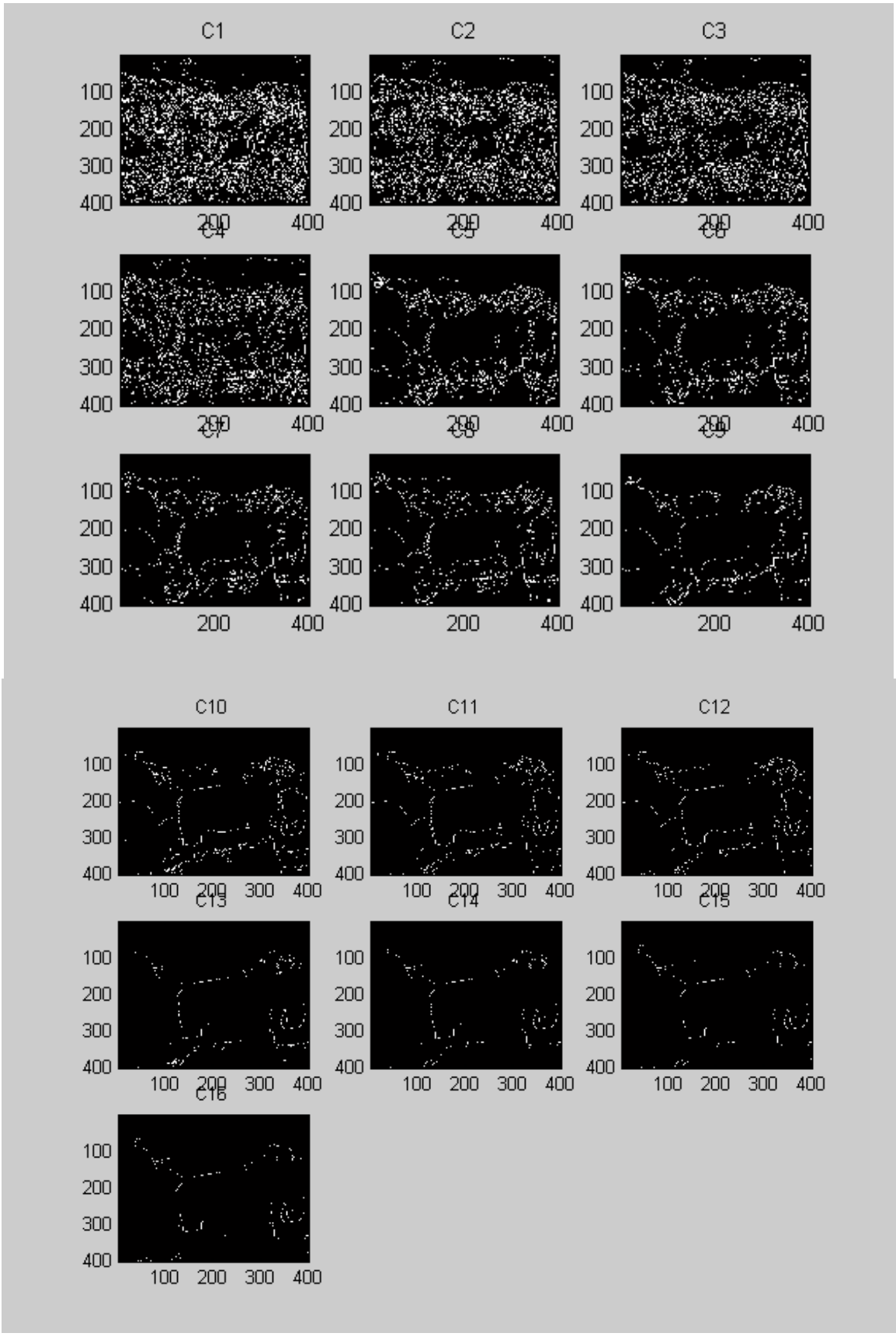

C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
100
200
300
400

Fig. A13. The 16 detection results using the parameter range presented in the paper. It can be seen that the detections include very noisy and very sparse ones. (Detection results are not clear because the subfigures are small.)

Results using 49 parameter sets combinations instead of 16

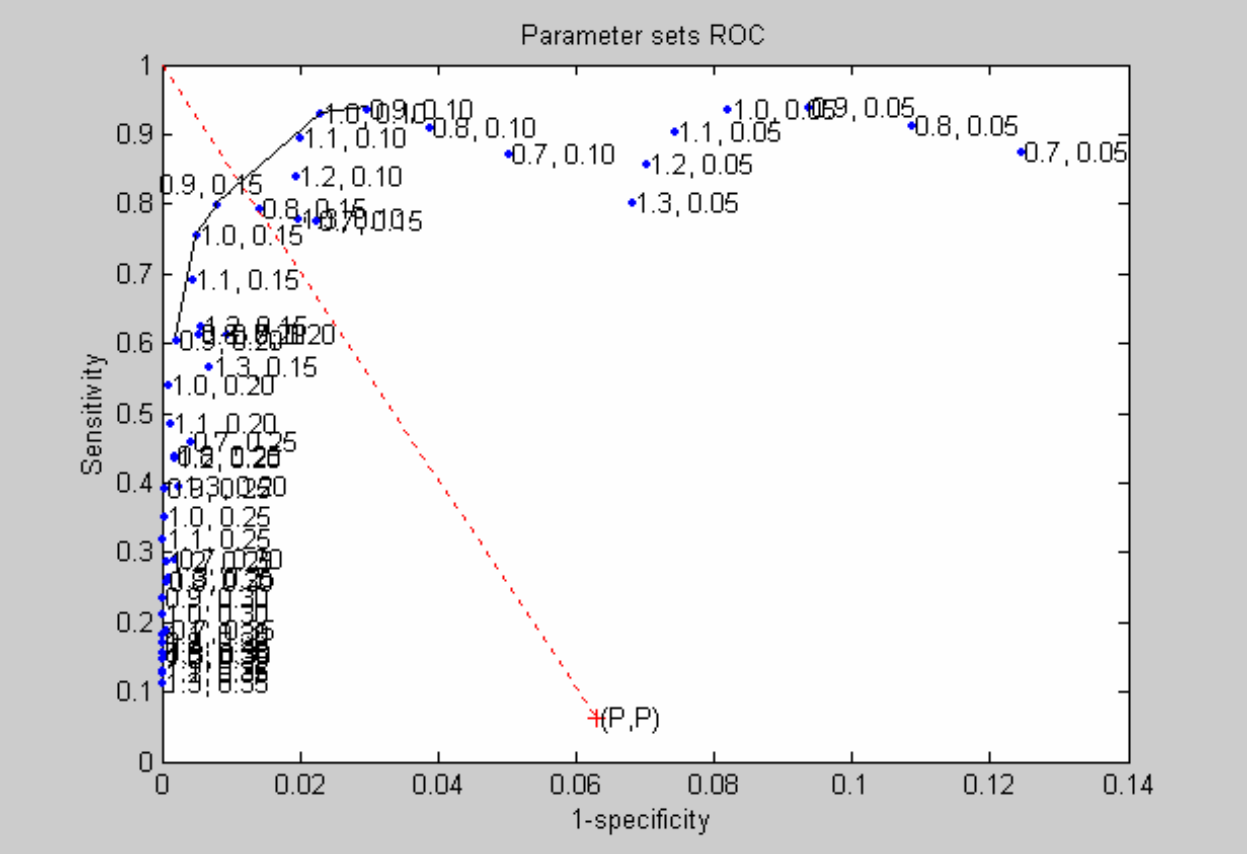


Fig. A14. Implementation of Canny detector with a more refined parameter range (**49 sets instead of 16**) [Elephant image]. This can be **compared to Fig. 8(a1)** in the paper. In both cases the best parameter set was found to be (0.9, 0.15).

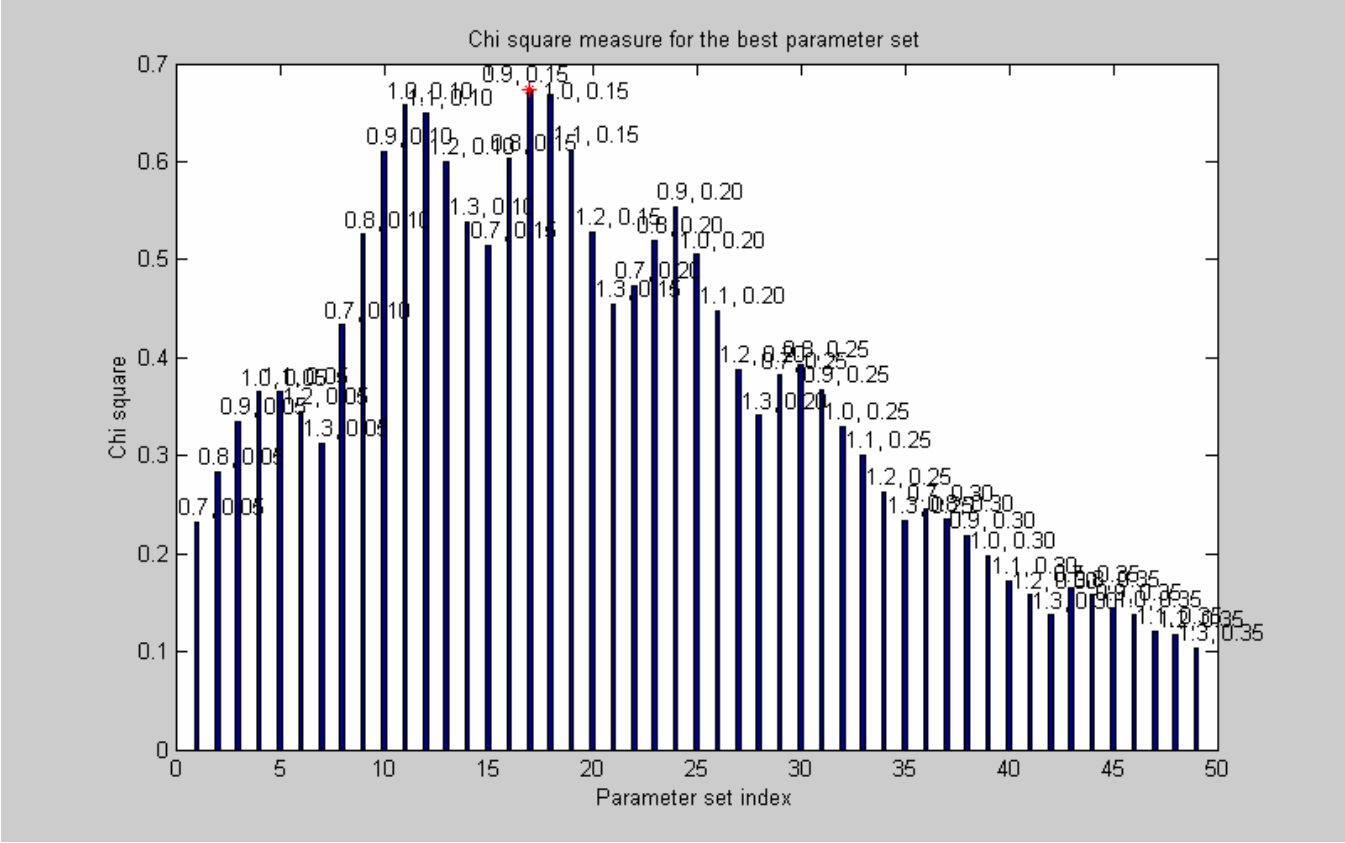


Fig. A15. Implementation of Canny detector with a more refined parameter range (**49 sets instead of 16**) [Elephant image]. This can be **compared to Fig. 8(a2)** in the paper. Again, in both cases the best parameter set was found to be (0.9, 0.15).

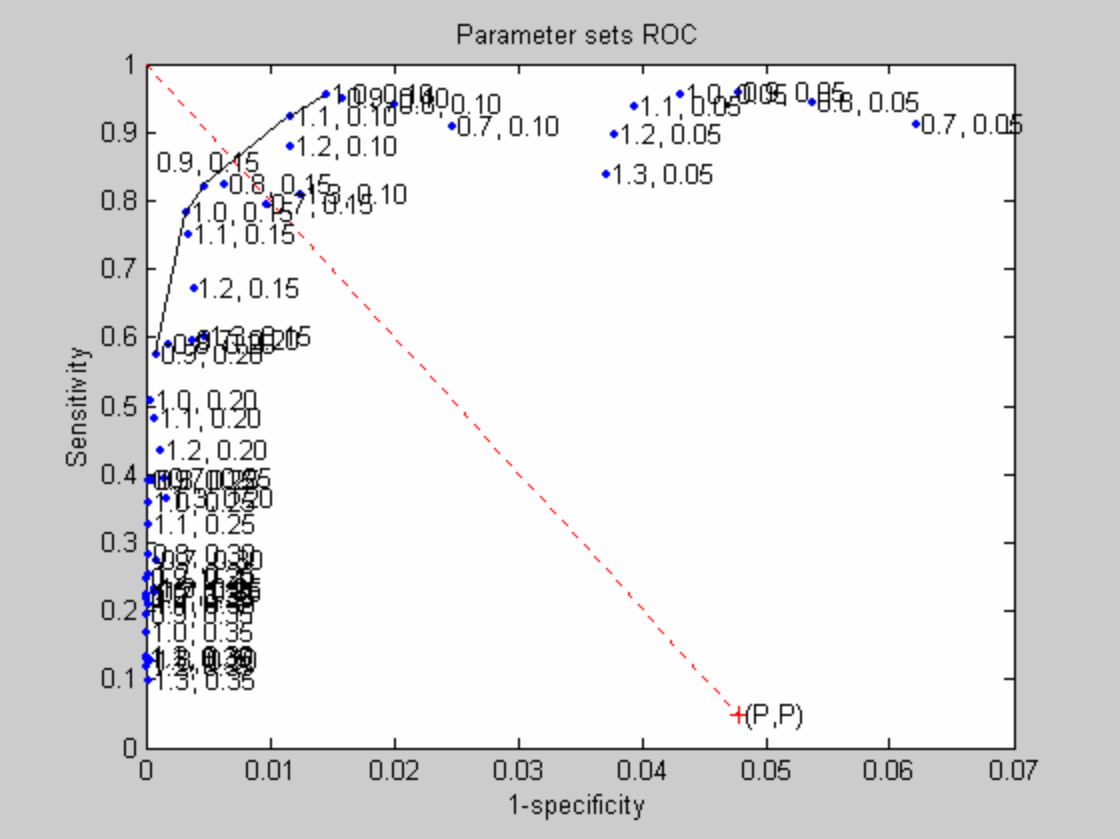


Fig. A16. Implementation of Canny detector with a more refined parameter range (**49 sets instead of 16**) [Grater image]. This can be **compared to Fig. 8(b1)** in the paper. In both cases the best parameter set was found to be (0.9, 0.15).

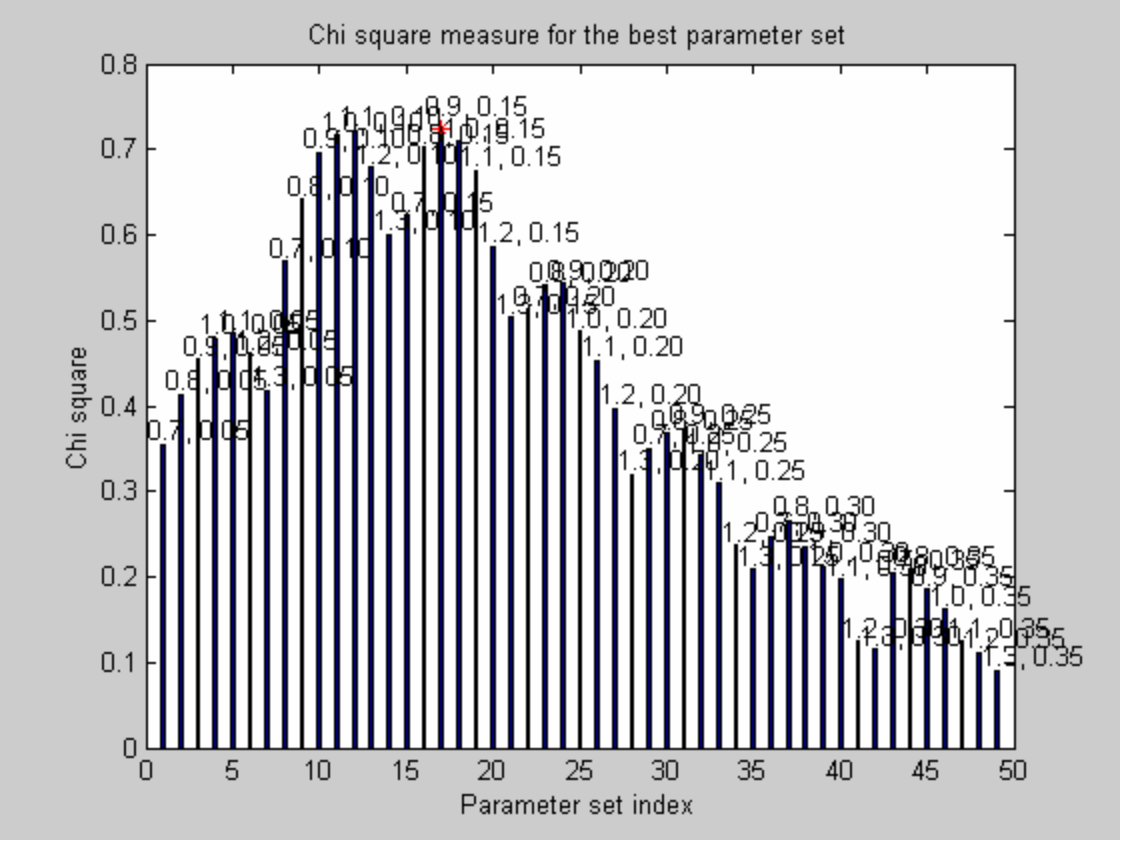


Fig. A17. Implementation of Canny detector with a more refined parameter range (**49 sets instead of 16**) [Grater image]. This can be **compared to Fig. 8(b2)** in the paper. Again, in both cases the best parameter set was found to be (0.9, 0.15).

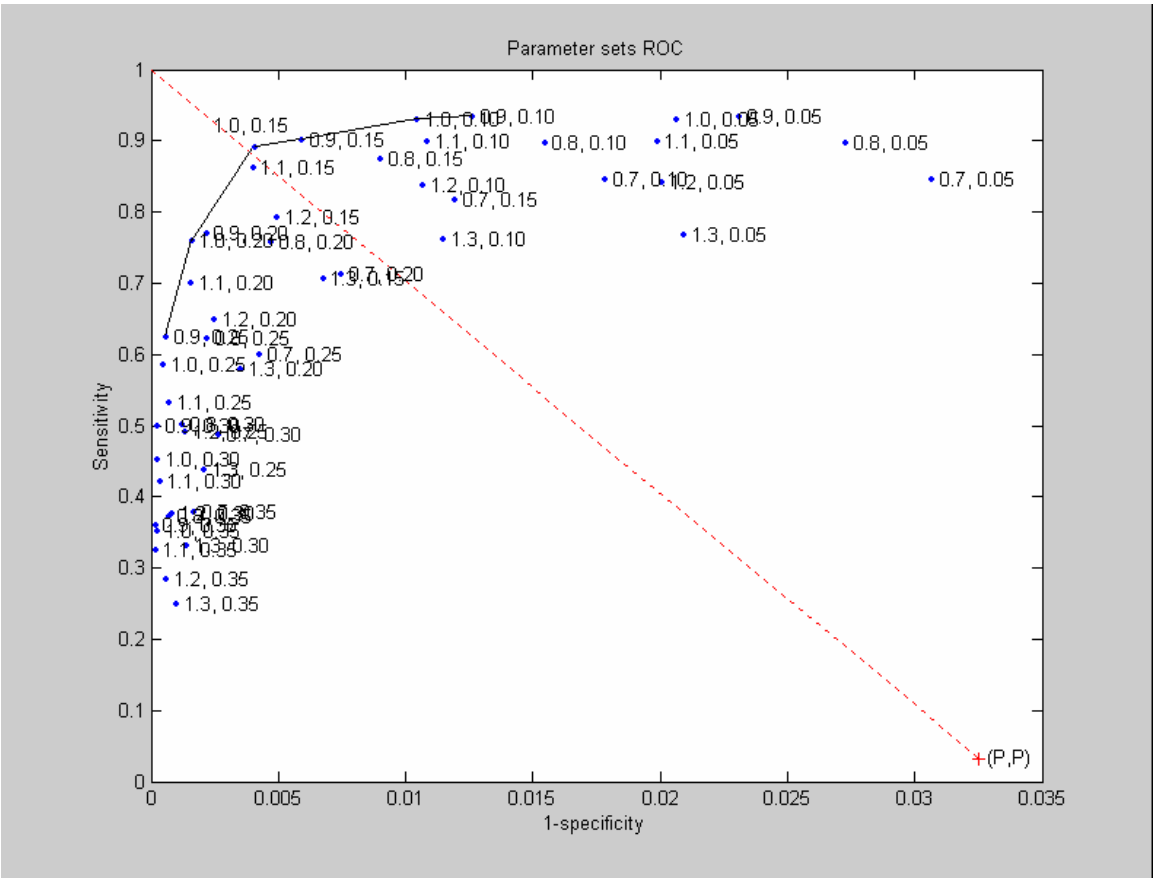


Fig. A18. Implementation of Canny detector with a refined parameter range (**49 sets instead of 16**) [Turtle image]. This can be **compared to Fig. 8(c1)** in the paper. The selected set in the paper was (1.1, 0.15), and here (1.0, 0.15).

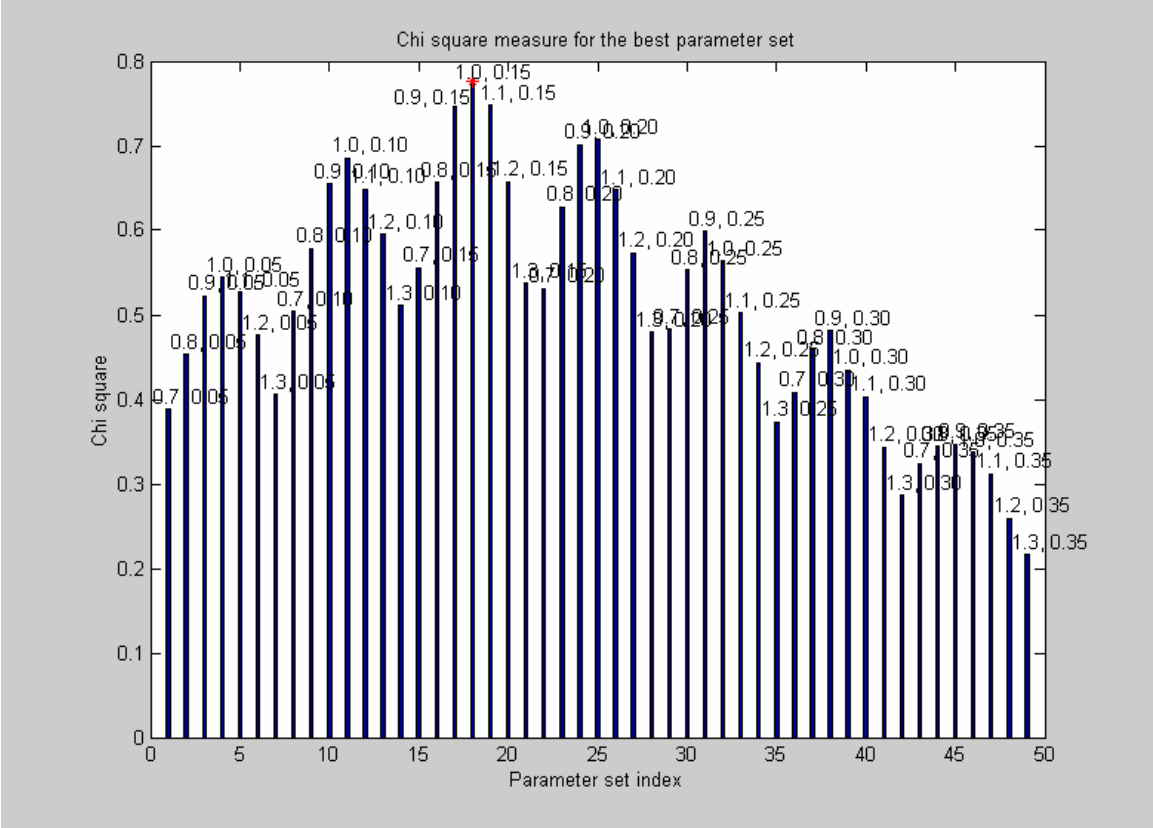


Fig. A19. Implementation of Canny detector with a refined parameter range (**49 sets instead of 16**) [Turtle image]. This can be **compared to Fig. 8(c2)** in the paper. The selected set in the paper was (1.1, 0.15), and here (1.0, 0.15).

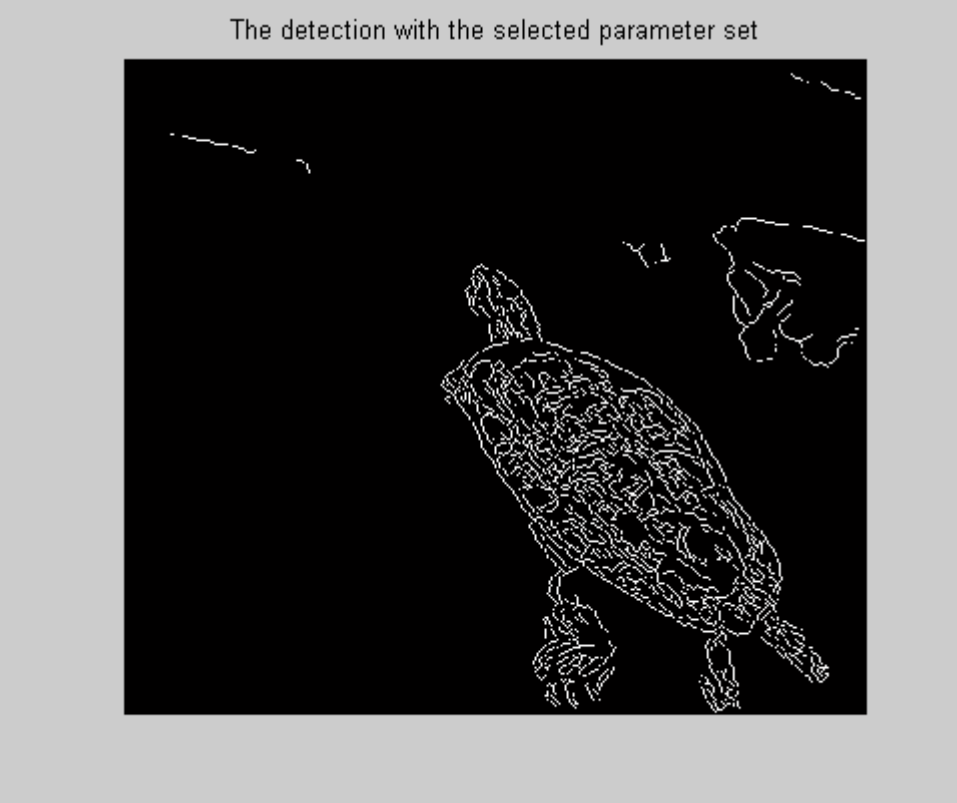


Fig. A20. The detection result with the refined range of parameter set selection (49 sets) is very similar to that with the rougher range (16 sets - Fig. 9(c)).

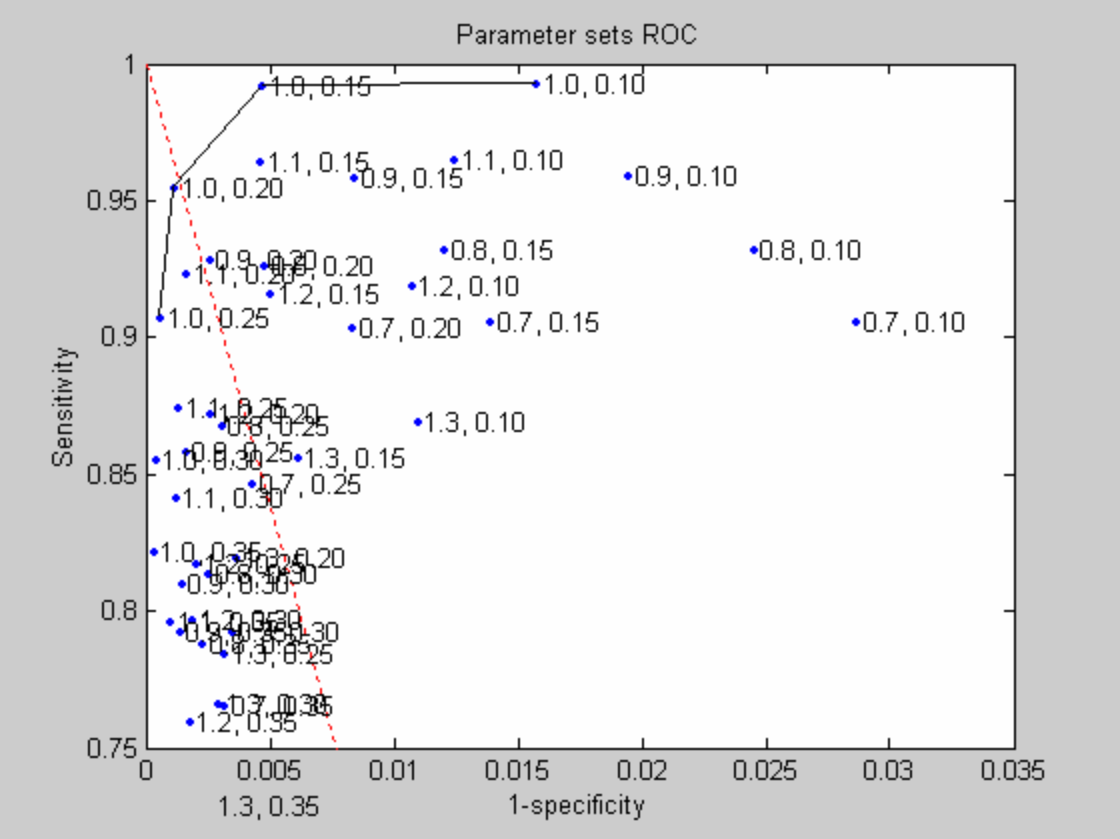


Fig. A21. Implementation of Canny detector with a refined parameter range (**49 sets instead of 16**) [Airplane image] <zoomed figure>. This can be **compared to Fig. 8(d1)** in the paper. The selected set in the paper was (1.1, 0.15), and here (1.0, 0.2).

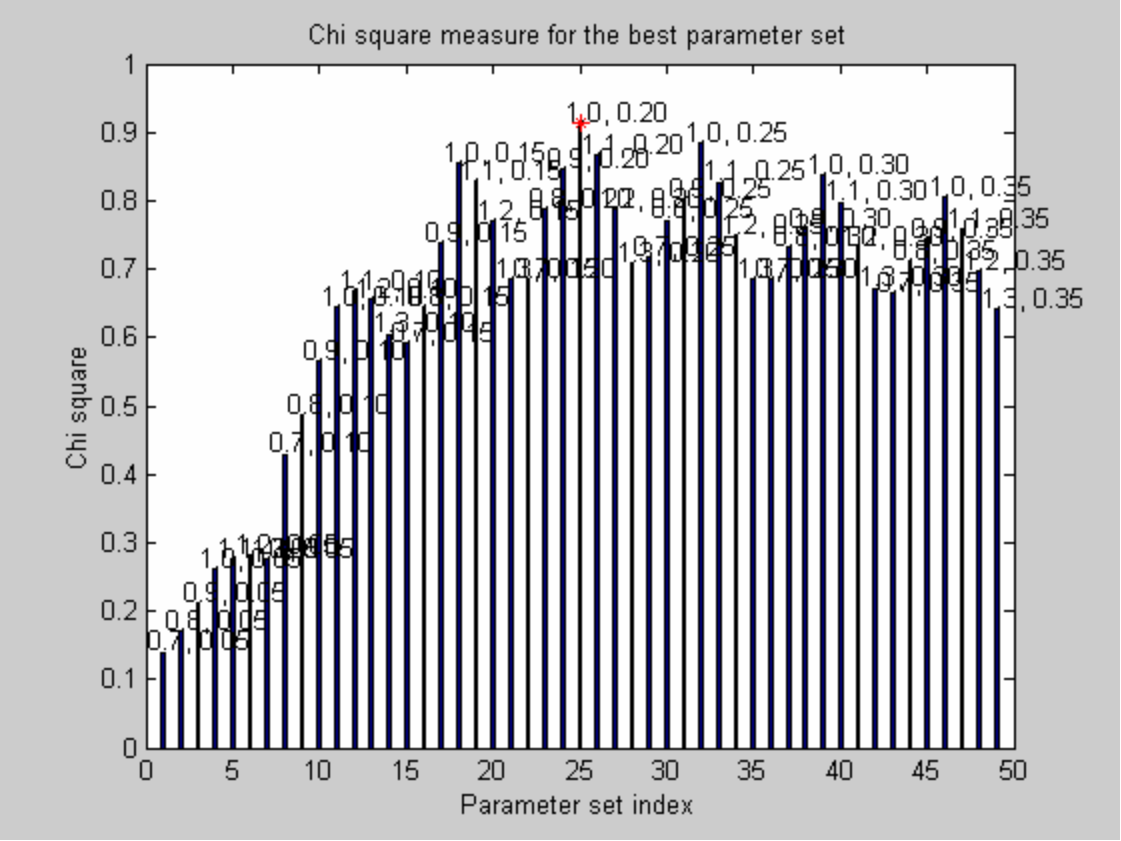


Fig. A22. Implementation of Canny detector with a refined parameter range (49 sets instead of 16) [Airplane image]. This can be compared to Fig. 8(d2) in the paper. The selected set in the paper was (1.1, 0.15), and here (1.0, 0.2).

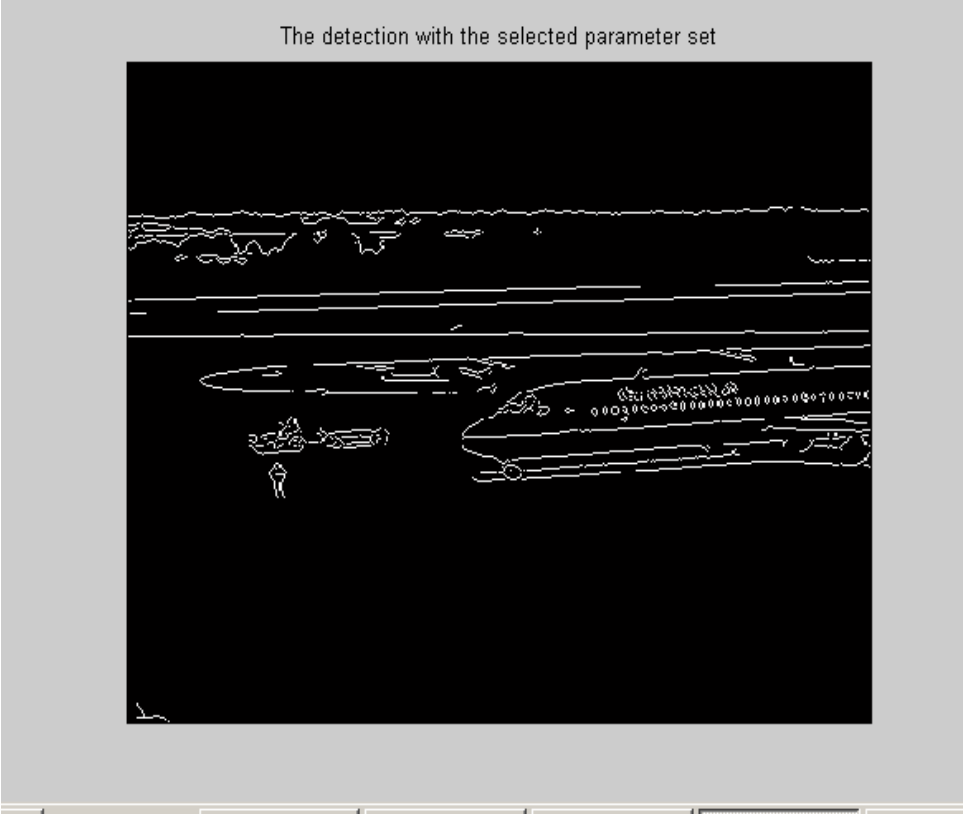


Fig. A23. The detection result with the refined range of parameter set selection (49 sets) is very similar to that with the rougher range (16 sets - Fig. 9(d)).

Additional results – consistency of parameter set selection results across images with spatial similarity

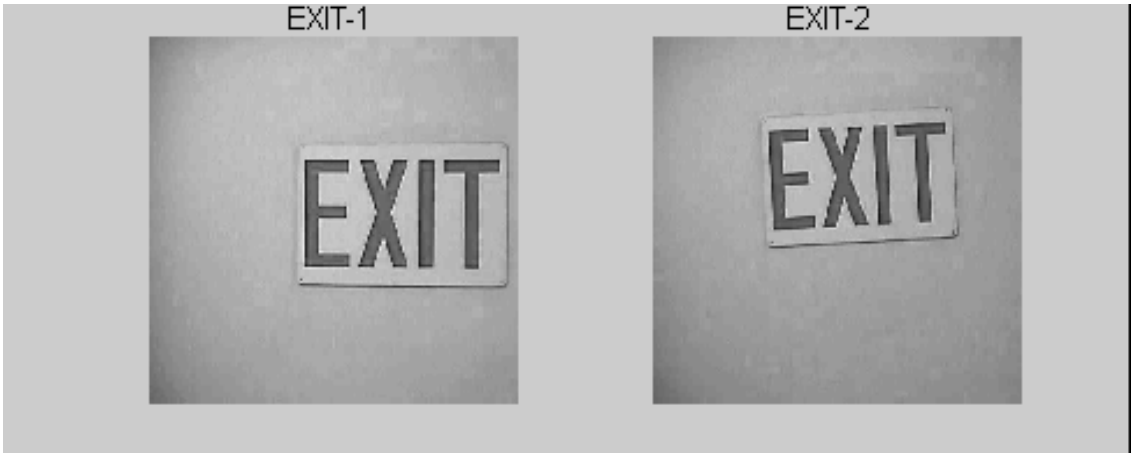


Fig. A24. Two images “EXIT-1” and “EXIT-2” with spatial similarity.

Results for “EXIT-1” image

Results for “EXIT-2” image

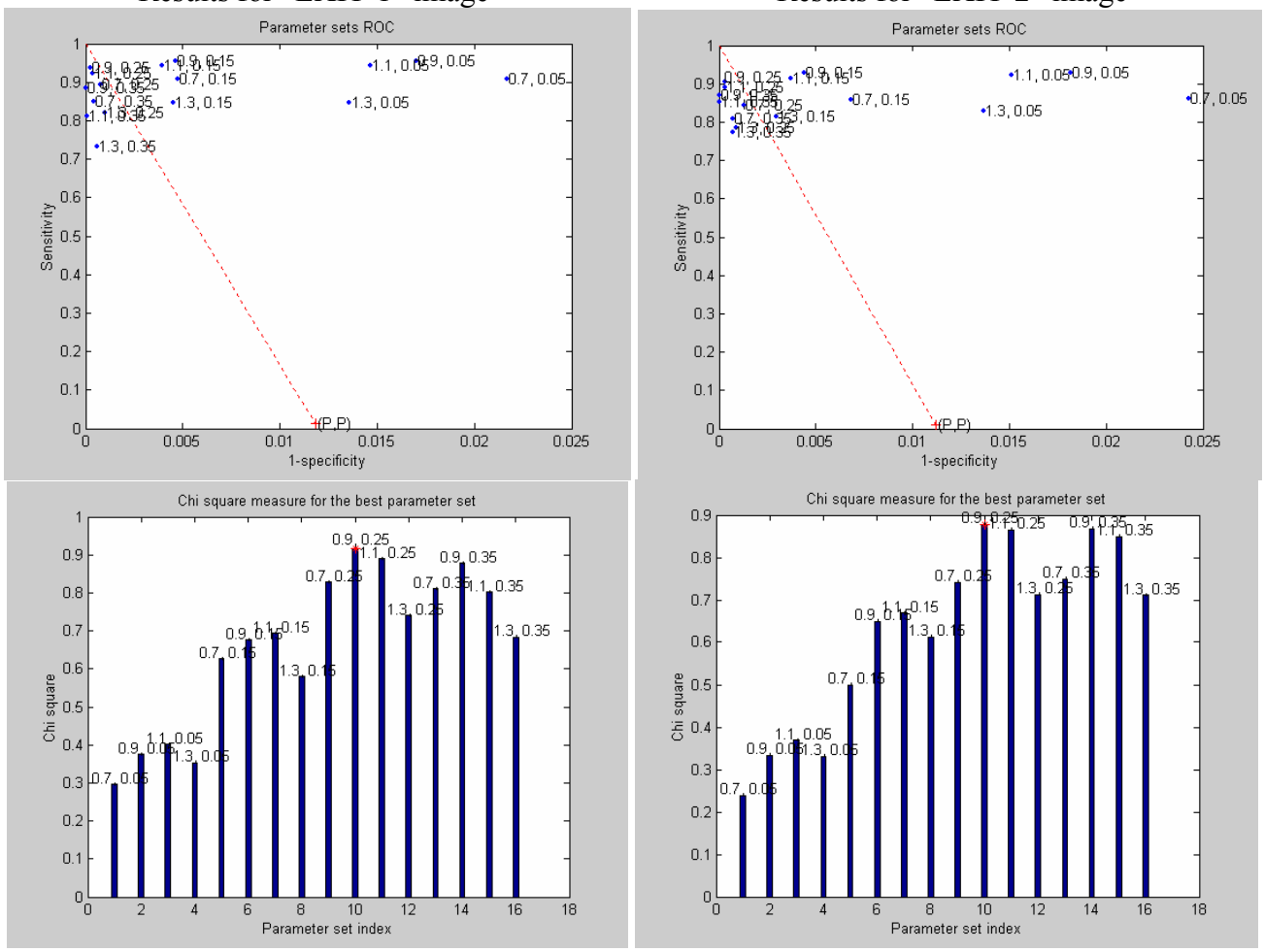


Fig. A25. Canny detector: Best parameter set selection results (ROC criterion – above, Chi-square criterion – below). The similar images produce similar selection results: The distribution of the results that evaluates the parameter sets are similar for both images, and the best sets selected are identical.

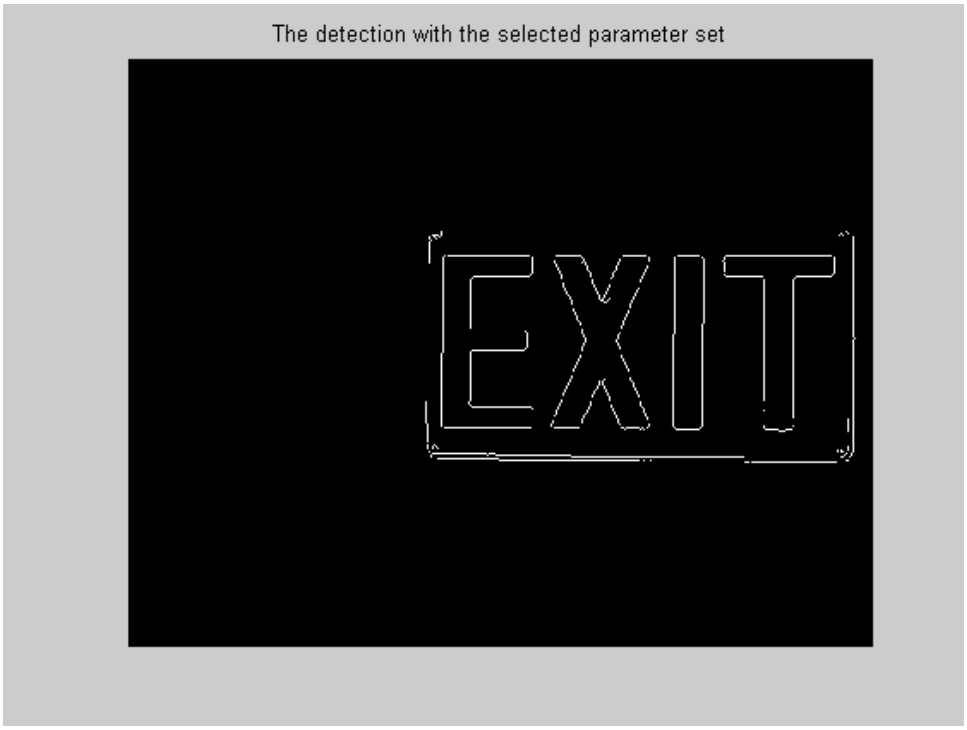


Fig. A.26. Edge detection for “EXIT-1” using the selected parameter set.

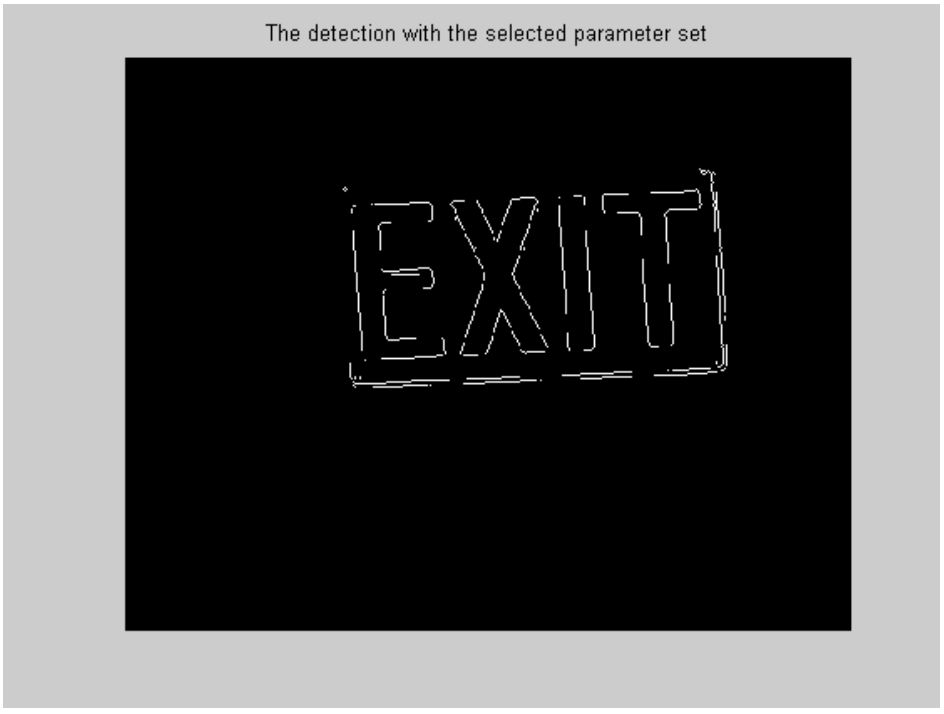


Fig. A.27. Edge detection for “EXIT-2” using the selected parameter set.

Results for “EXIT-1” image

Results for “EXIT-2” image

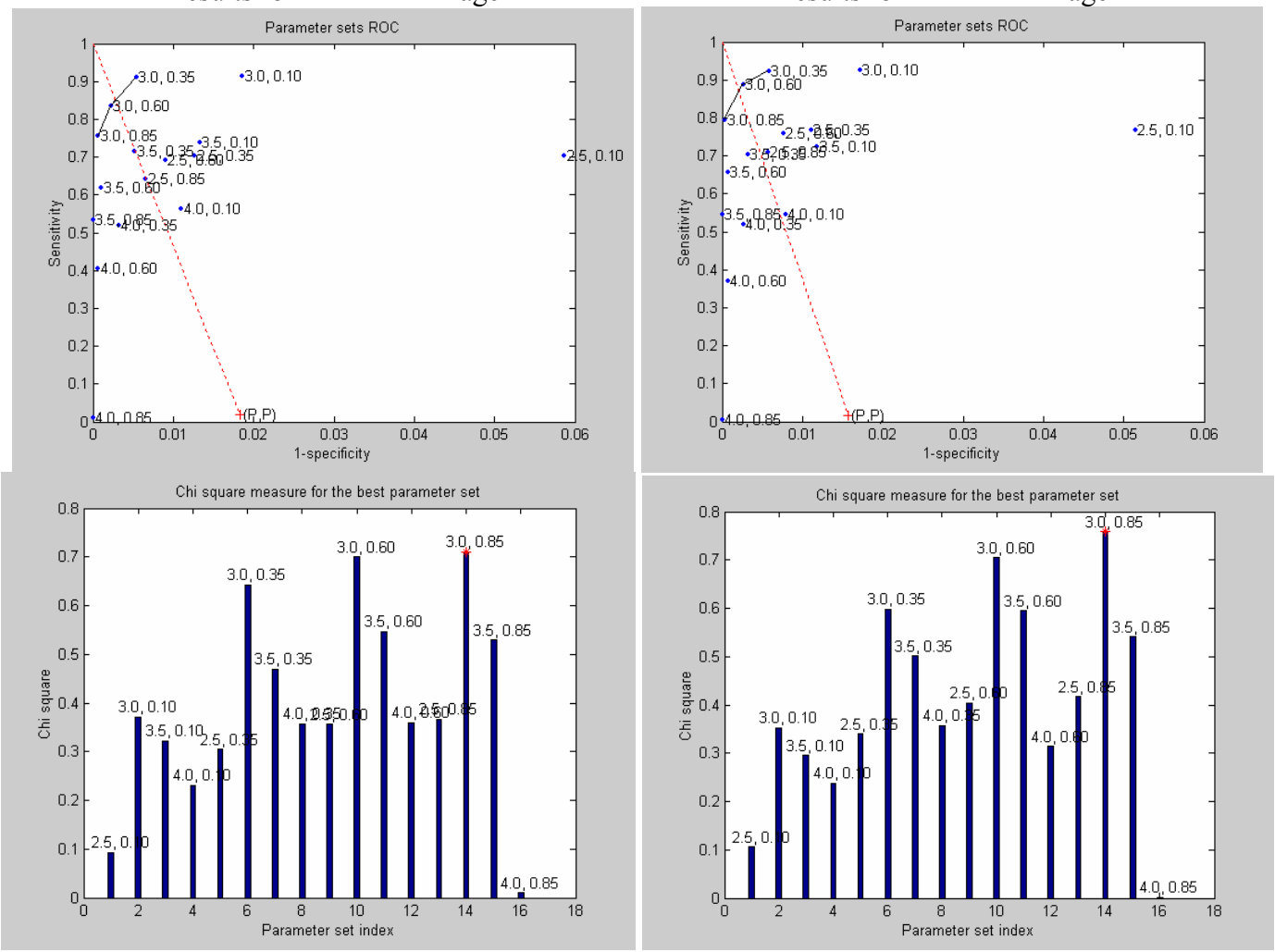


Fig. A28. LOG detector: Best parameter set selection results (ROC criterion – above, Chi-square criterion – below). The similar images produce similar selection results: The distribution of the results that evaluates the parameter sets are similar for both images, and the best sets selected are identical.

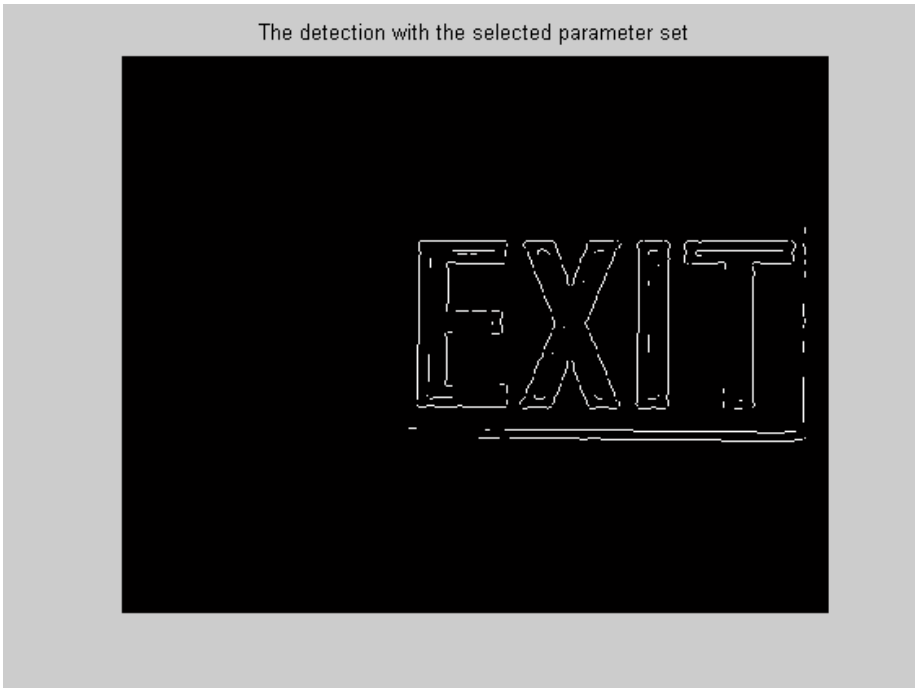


Fig. A.29. Edge detection for “EXIT-1” using the selected parameter set in the LOG detector (according to the chi-square criterion).

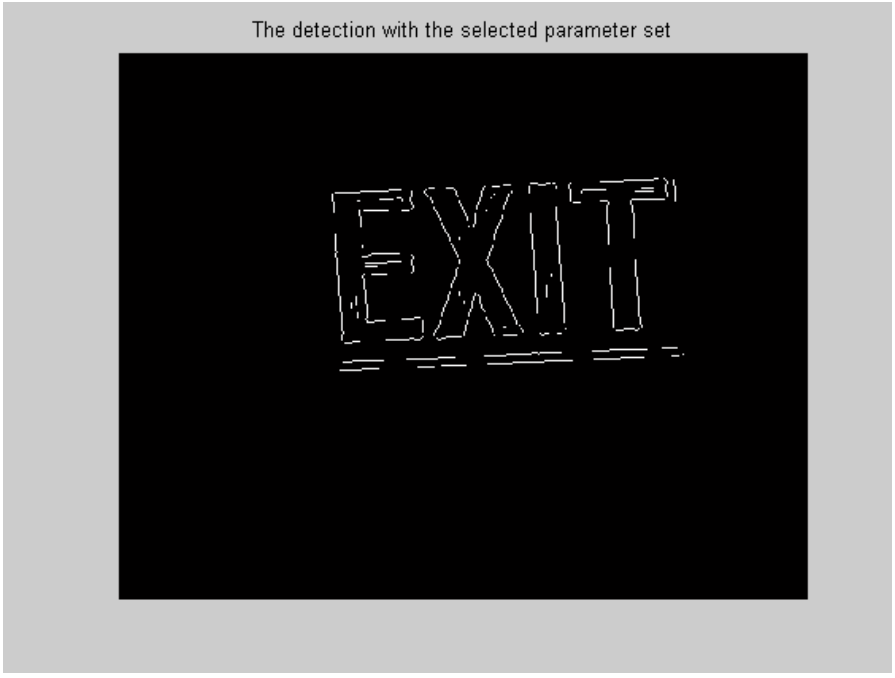


Fig. A.30. Edge detection for “EXIT-2” using the selected parameter set in the LOG detector (according to the chi-square criterion).

Results for “EXIT-1” image

Results for “EXIT-2” image

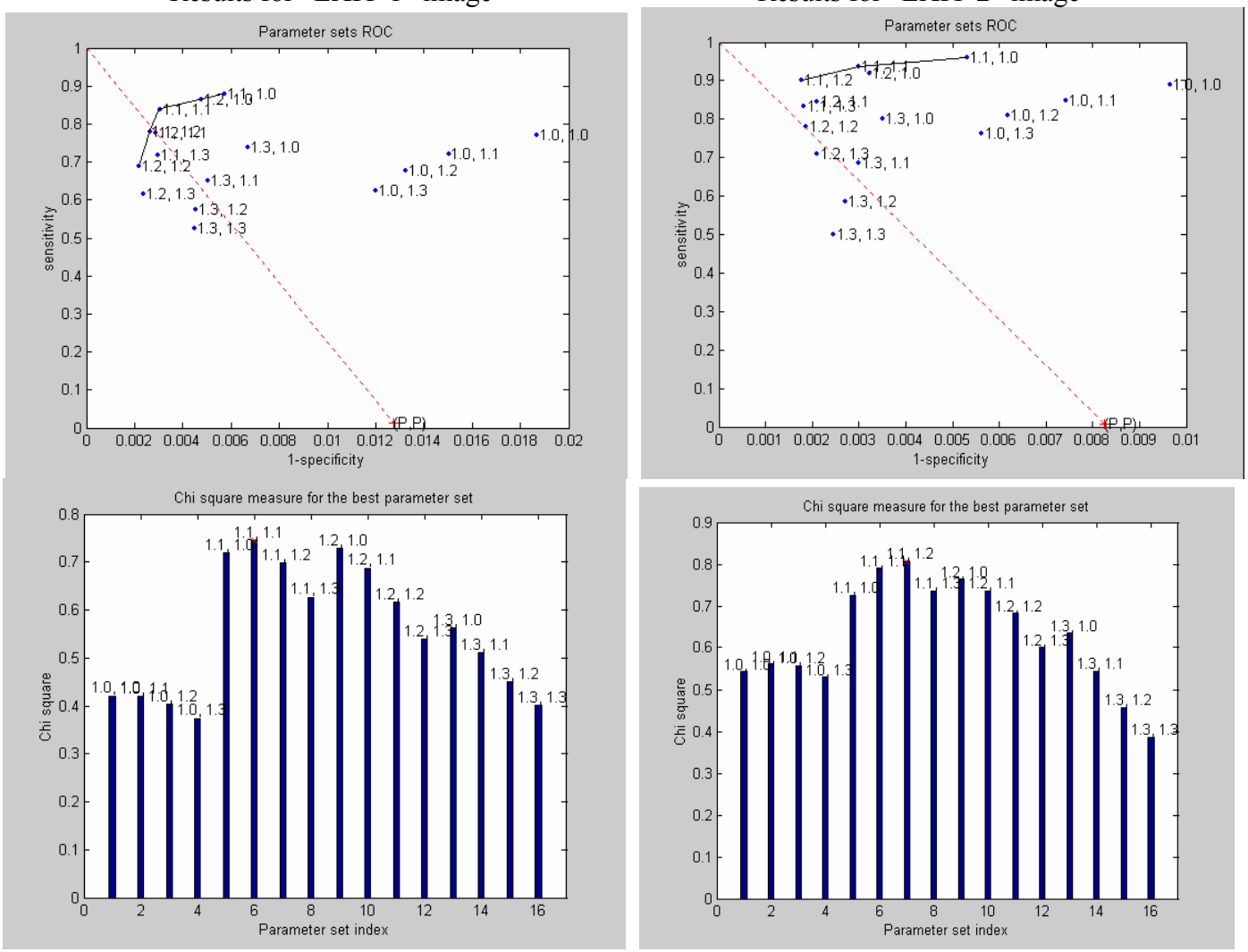


Fig. A31. Peli’s detector: Best parameter set selection results (ROC criterion – above, Chi-square criterion – below). The similar images produce close selection results: The distribution of the results that evaluates the parameter sets are similar for both images, and the best sets selected are very close.

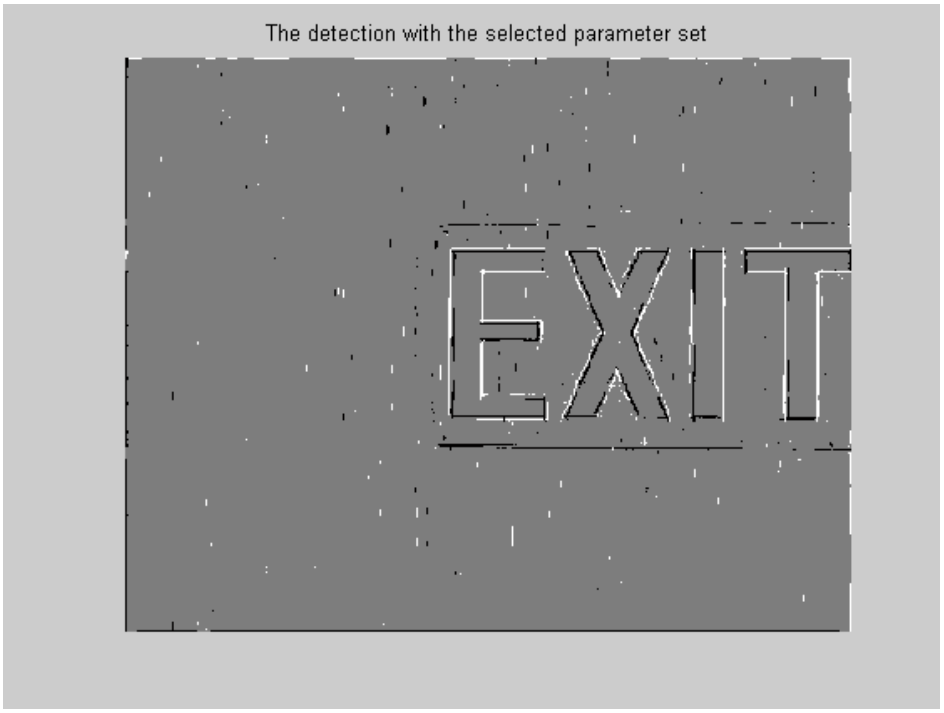


Fig. A.32. Edge detection for “EXIT-1” using the selected parameter set in the Peli’s detector (according to the Chi-square criterion).

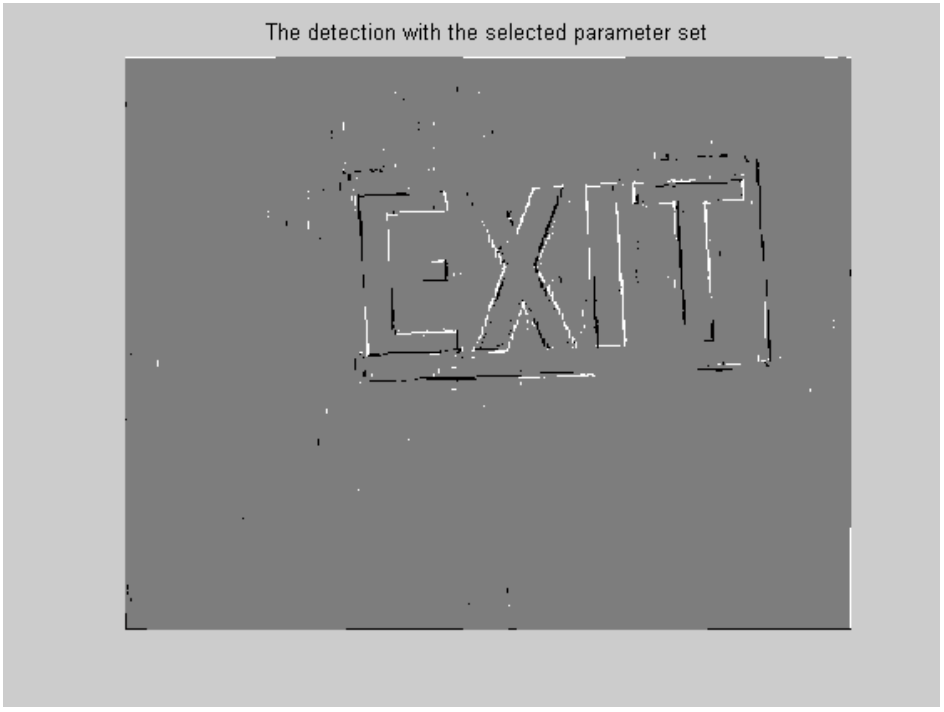


Fig. A.33. Edge detection for “EXIT-2” using the selected parameter set in the Peli’s detector (according to the Chi-square criterion).

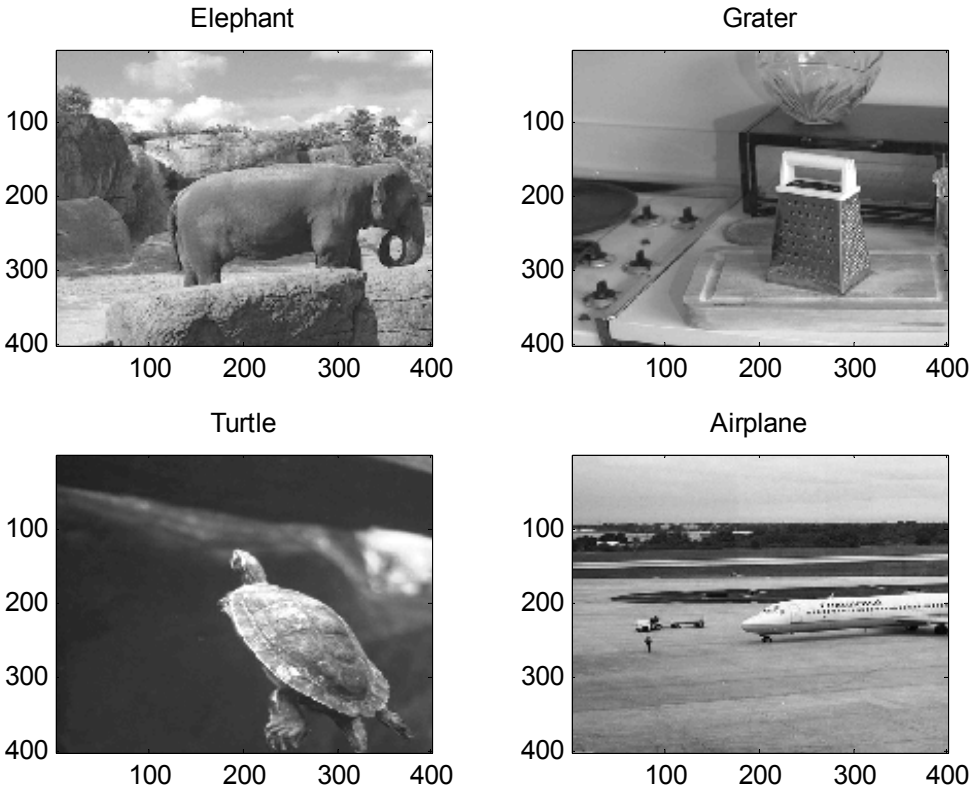
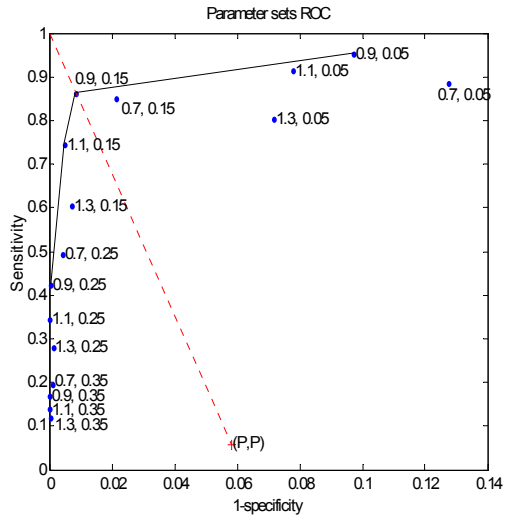
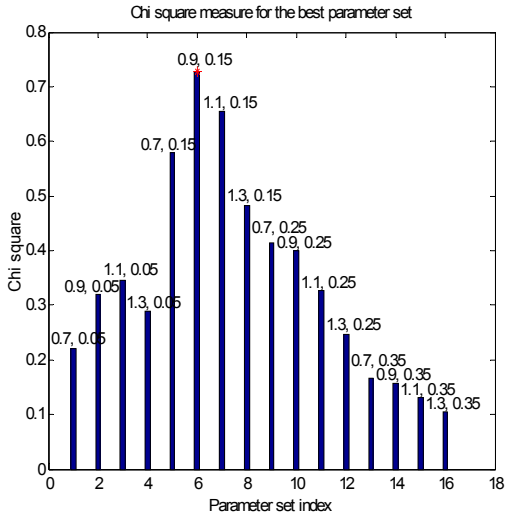


Fig. A34. Images used for demonstration and analysis in the paper (selected for comparison with results in [1], and taken from their web page [19]). Parameter set selection results for the Canny detector are shown in Fig. A35, and edge detection results using the selected parameter sets are shown in Fig. 5 in the paper.

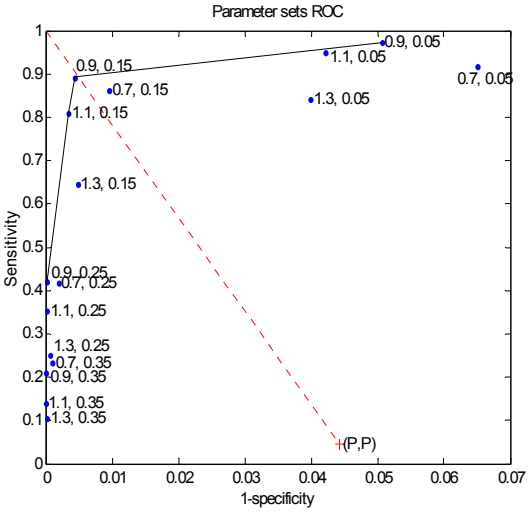
(a1)



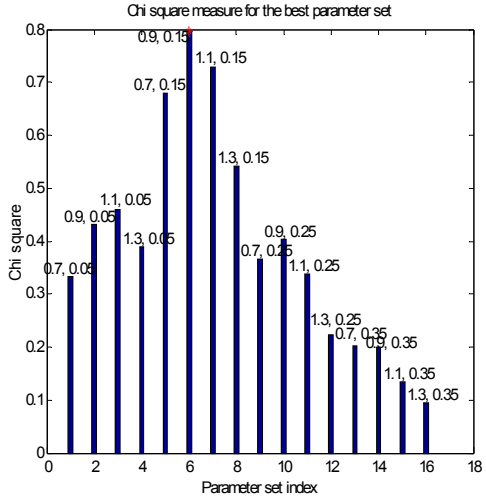
(a2)



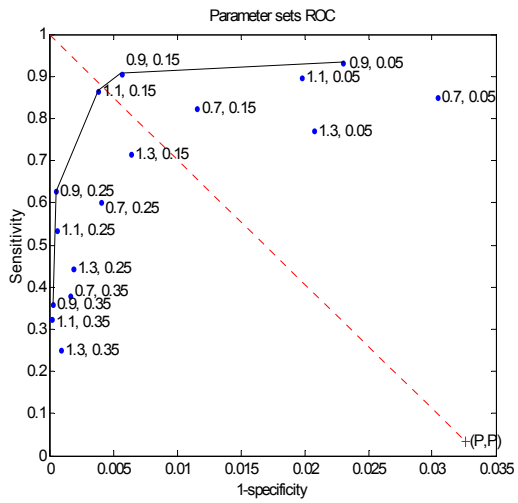
(b1)



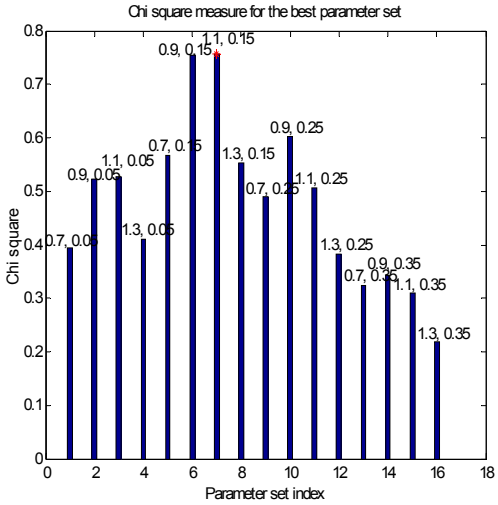
(b2)



(c1)



(c2)



(d1) (d2)

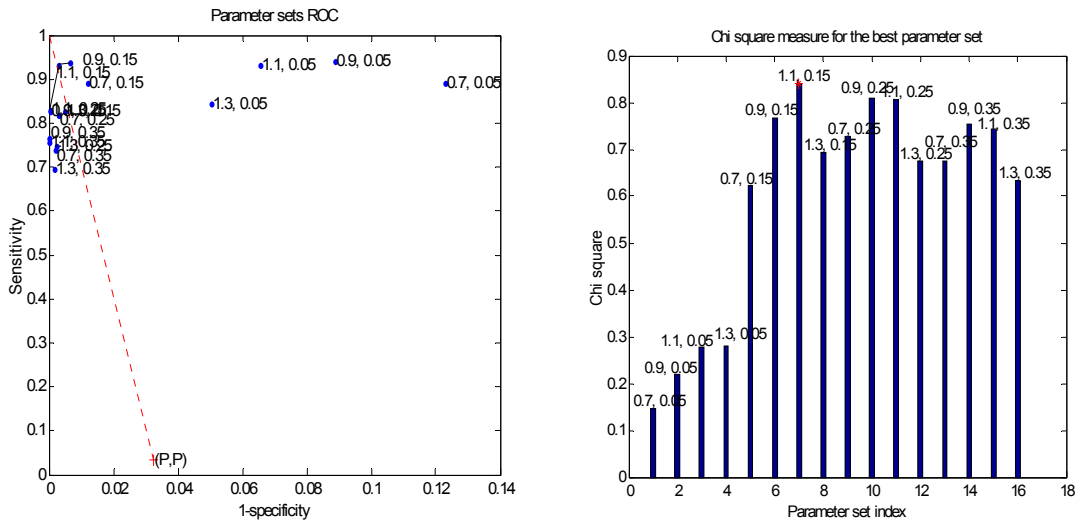


Fig. A35. Results of the Canny detector's parameter set selection using the proposed automatic method (see section 3 in the paper). Graphs (a 1,2), (b 1,2), (c 1,2), and (d 1,2) are the results using the images "Elephant", "Grater", "Turtle", and "Airplane", respectively; The results of the PSROC with the *diagnosis line* are shown in graphs (a1), (b1), (c1), and (d1), where the numbers near the points are the standard deviation and the high threshold parameters of the Canny detector. The results of the Chi-square measure are shown in graphs (a2), (b2), (c2), and (d2). **Results:** For the Elephant image: (a1) the best set is (0.9,0.15); (a2) the best set is the same. For the Grater image (graphs (b1), (b2)), the best set is the same as for the Elephant image. For the Turtle image (graphs (c1), (c2)), the best set is (1.1,0.15) according to both measures. For the Airplane image (graphs (d1), (d2)), the best set is the same as for the Turtle image.

A remark regarding the properties of the CTROC results

As can be seen from the CTROC graphs (Figs. 4(a) and 4(c)) the abscissas have values of about an order of magnitude smaller than the ordinates. The reason is that the number of non-edge points is larger by about an order of magnitude than the number of the edge points. Therefore, the values of TN are relatively high (thus the values of the other probabilities are relatively low) causing low values of FPRs (1 – *specificity*) and high values of TPRs. If the CTROC plane was presented in the full possible range (0,0; 1,1), the CTROC curve would appear very close to the upper left side as if an excellent curve is produced, however, this is misleading since it results merely from the large proportion of TN.

From the CTROCs we can see that in all cases (1 – *specificity*) = 0 at CT level *N* since at that level the PGT has 100% correspondence and therefore no non-edges at the various detections coincide with 'edges' in this PGT, thus, $\overline{\mathrm{FP}}_{\mathrm{PGT}_N} = 0$. The *sensitivity* has always a value of 1 at CT level 1 since at that level all the detected edge points in all the detections are at this PGT's 'edge' points and therefore no detected edge points coincide with 'non-edge' points in this PGT, thus, $\overline{\mathrm{FN}}_{\mathrm{PGT}_1} = 0$. Between these extreme points, the properties of the PGTs can be evaluated according to the relative locations of their points in the CTROC plane. A change of location from the lower left to the upper right of the CTROC plane indicates a smaller CT, while a change of location from the lower right to the upper left of this plane indicates a higher quality (better resemblance to the overall detection results).

Formulation for the extraction of the best parameter set

Defining an edge pixel as "1" and a non-edge pixel as "0", the probabilities in this case, according to table 1(b) are:

$$\mathrm{TP}_{\mathrm{D}_j\mathrm{EGT}} = \frac{1}{\mathrm{K}\cdot\mathrm{L}}\sum_{k=1}^{\mathrm{K}}\sum_{l=1}^{\mathrm{L}}\mathrm{D}_{j_1}\bigcap\mathrm{EGT}_1 \tag{1}$$

$$\mathrm{FP}_{\mathrm{D}_j\mathrm{EGT}} = \frac{1}{\mathrm{K}\cdot\mathrm{L}}\sum_{k=1}^{\mathrm{K}}\sum_{l=1}^{\mathrm{L}}\mathrm{D}_{j_1}\bigcap\mathrm{EGT}_0 \tag{2}$$

$$\mathrm{TN}_{\mathrm{D}_j\mathrm{EGT}} = \frac{1}{\mathrm{K}\cdot\mathrm{L}}\sum_{k=1}^{\mathrm{K}}\sum_{l=1}^{\mathrm{L}}\mathrm{D}_{j_0}\bigcap\mathrm{EGT}_0 \tag{3}$$

$$\mathrm{FN}_{\mathrm{D}_j\mathrm{EGT}} = \frac{1}{\mathrm{K}\cdot\mathrm{L}}\sum_{k=1}^{\mathrm{K}}\sum_{l=1}^{\mathrm{L}}\mathrm{D}_{j_0}\bigcap\mathrm{EGT}_1 \tag{4}$$

where EGT_1 and EGT_0 are 'edge' and 'non-edge' points in the EGT respectively. In order to present the resulting match between the detections using the different sets of parameters and the EGT, on the parameter

set ROC (PSROC) plane, the conditional probabilities will be described here with respect to their definitions in section 2.1.1. The TPR in this case will be:

$$\mathrm{TPR}_{\mathrm{D}_j\mathrm{EGT}} = \frac{\mathrm{TP}_{\mathrm{D}_j\mathrm{EGT}}}{\mathrm{P}}, \tag{5}$$

where the P is the *prevalence* of the number of the positive detections (edges) in the EGT, $P(E_{EGT})$ and does not depend on the detection *j* that only influences the proportion between the components of P: TP and FN. Thus,

$$\mathrm{P} = \mathrm{TP}_{\mathrm{D}_j\mathrm{EGT}} + \mathrm{FN}_{\mathrm{D}_j\mathrm{EGT}} \quad \forall j, \tag{6}$$

The FPR will be:

$$\mathrm{FPR}_{\mathrm{D}_j\mathrm{EGT}} = \frac{\mathrm{FP}_{\mathrm{D}_j\mathrm{EGT}}}{1 - \mathrm{P}}, \tag{7}$$

where

$$1 - \mathrm{P} = \mathrm{FP}_{\mathrm{D}_j\mathrm{EGT}} + \mathrm{TN}_{\mathrm{D}_j\mathrm{EGT}}. \tag{8}$$

The points in the PSROC plane will then be:

$$(\textit{sensitivity, } 1\text{-}\textit{specificity}) = (\mathrm{TPR}_{\mathrm{D}_j\mathrm{EGT}}, \mathrm{FPR}_{\mathrm{D}_j\mathrm{EGT}}). \tag{9}$$

Unlike the case of the GT estimation, where every point has either better *sensitivity* or better *specificity* than each of the others, points in the PSROC plane have no such property. In the CTROC case, as the CT *i* decreases, more 'edge' (positive) pixels are added to the PGT, but no pixels are taken out, and therefore, both $\mathrm{TP}_{\mathrm{PGT}_i,\mathrm{D}_j}$ and $\mathrm{FP}_{\mathrm{PGT}_i,\mathrm{D}_j}$ are not decreasing for each detection, forming a monotonic CTROC curve. However, in the PSROC case, whereby the different detections are matched with the EGT, there is no rule regarding the changes of $\mathrm{TP}_{\mathrm{D}_j\mathrm{EGT}}$ and $\mathrm{FP}_{\mathrm{D}_j\mathrm{EGT}}$ for different detections.

The Chi-square measure in this case will be:

$$\chi^2_{\mathrm{D}_j\mathrm{EGT}} = \frac{(\textit{sensitivity} - \mathrm{Q}_{\mathrm{D}_j\mathrm{EGT}})}{(1 - \mathrm{Q}_{\mathrm{D}_j\mathrm{EGT}})} \cdot \frac{(\textit{speceficity} - (1 - \mathrm{Q}_{\mathrm{D}_j\mathrm{EGT}}))}{\mathrm{Q}_{\mathrm{D}_j\mathrm{EGT}}}, \tag{10}$$

and

$$\mathrm{Q}_{\mathrm{D}_j\mathrm{EGT}_i} = \mathrm{TP}_{\mathrm{D}_j\mathrm{EGT}} + \mathrm{FP}_{\mathrm{D}_j\mathrm{EGT}} \tag{11}$$

Evaluation of the detection results from the PSROC

The properties of the detection results using the different parameter sets can be evaluated according to the relative locations of their points in the PSROC plane using the following 4 possibilities. (1) When a point in the PSROC has higher values of *sensitivity* and (1 – *specificity*), both TP_{D_jEGT} and FP_{D_jEGT} are higher (while P is constant), and therefore, the detection result represented by this point has more 'edge' points (noisier); (2) When a point has lower values of *sensitivity* and (1 – *specificity*), both TP_{D_jEGT} and FP_{D_jEGT} are lower, and therefore, the detection result represented by this point has fewer 'edge' points; (3) When a point has higher *sensitivity* and lower (1 – *specificity*), TP_{D_jEGT} is higher and FP_{D_jEGT} is lower, and therefore, the detection result represented by this point is more similar to the EGT (higher quality of the detection); And (4) when a point has lower *sensitivity* and higher (1 – *specificity*), TP_{D_jEGT} is lower and FP_{D_jEGT} is higher, and therefore, the detection result represented by this point is less similar to the EGT (lower quality of the detection). This analysis of the implication of the point locations in the PSROC plane can help us understand the relative properties of the different detection results just from observing the PSROC plane.

Furthermore, the effect of each detector parameter can be studied from the PSROC plane by observing the change in the location of the resulting point in the PSROC plane according to a change of any of the detector parameters. It can be seen that patterns are formed on the PSROC plane, in terms of quality (similarity to the EGT) and noisiness. An example of that appears in Fig. 7, which is a zoomed view of Fig. 3(c) in the paper. Points with the same center frequencies of the enhancing filters are connected with bold lines and points with the same frequency bandwidths of the filters are connected with thin lines. It can be seen that smaller frequency bandwidths cause greater noisiness in the detection while changes in the center frequencies affects mainly the similarity to the EGT. Similar properties exist also in the PSROC plane of the LOG detector in Fig. 3(a), and the effects of changing each of the parameters can be appreciated regarding noisiness and similarity to the EGT.

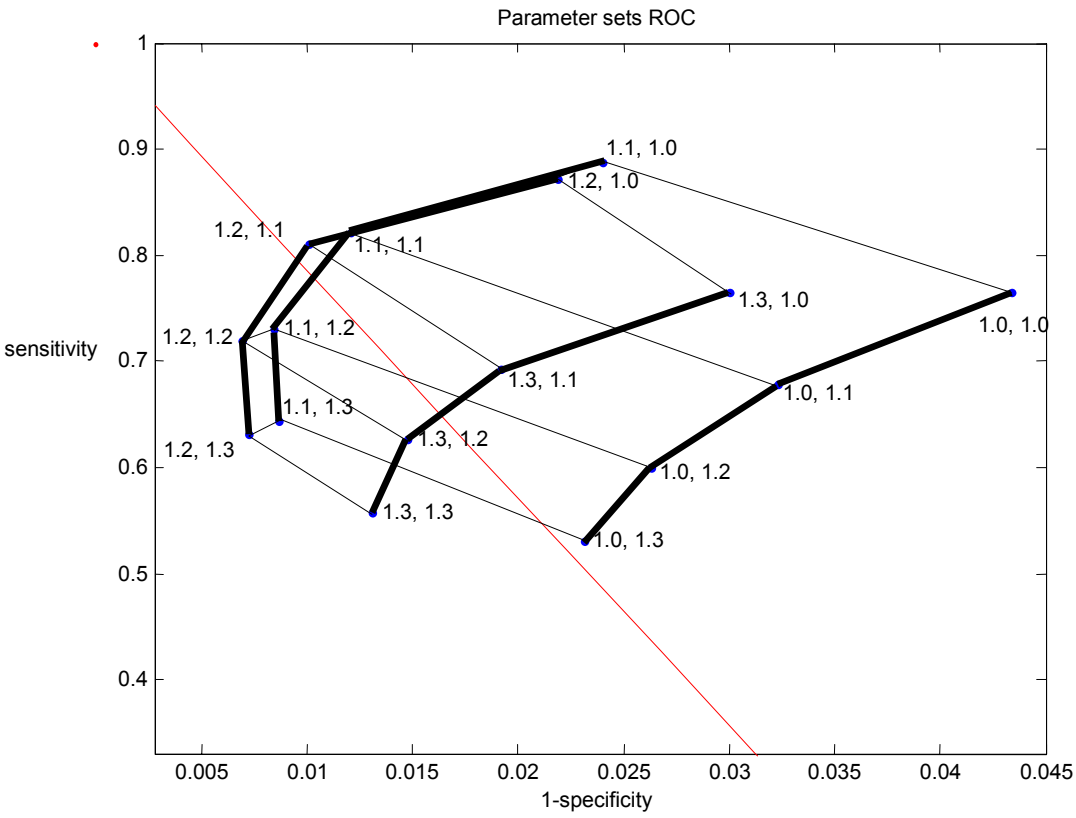


Fig. A36. A zoomed view of Fig. 3(c) in the paper, showing patterns resulting from varying one parameter and keeping constant another parameter in the PSROC plane. Points with the same center frequencies of the enhancing filters are connected with bold lines and points with the same frequency bandwidths of the filters are connected with thin lines.