

Copyright © 1999, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**COMPARING HUMAN AND COMPUTER
TOP-DOWN VISION IN A RECOGNITION
TASK: EFFECTS OF BLUR AND NOISE**

by

Masashi Okubo and Lawrence W. Stark

Memorandum No. UCB/ERL M99/20

29 March 1999

COVER

**COMPARING HUMAN AND COMPUTER
TOP-DOWN VISION IN A RECOGNITION
TASK: EFFECTS OF BLUR AND NOISE**

by

Masashi Okubo and Lawrence W. Stark

Memorandum No. UCB/ERL M99/20

29 March 1999

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

COMPARING HUMAN and COMPUTER TOP-DOWN VISION in a RECOGNITION TASK: EFFECTS of BLUR and NOISE

Masashi Okubo and Lawrence W. Stark

Okayama Prefecture University, 719-1197, Japan
Neurology and Telerobotics Units,
University of California, Berkeley, California 94720-2020
Email: okubo@cse.oka-pu.ac.jp, stark@pupil.berkeley.edu

Keywords: top-down computer vision, human vision, matched filters, blur, noise

ABSTRACT

We compared human and computer means of recognizing images. For human vision we presented one target icon and two decoy icons, in random order in a forced choice paradigm. Superimposed were different levels of noise and blur, two of the major degradation processes in images. For computer vision we used matched filters with appropriate shapes for the target and decoy icons as the means of recognition of the same pictures with noise and blur. We found similar behavior in human and computer vision. As expected, clear noiseless pictures resulted in perfect performance, while very degraded pictures reduced performance to chance levels. There were also interesting differences. Humans were better than computers for very blurred pictures with little noise; evidently, humans have a good deal of spatial differentiation capability. Computers were better in regions with moderate blur and a good deal of noise; likely, the noise destroys the shape consistency necessary for human recognition. The scanpath theory of top-down active human perception suggests that human 'matched filters' produced by internal computation or analogic reasoning have similar roles to the masks in the computer vision scheme that were pre-constructed by the human experimenters.

1. INTRODUCTION

Pattern recognition is an important area of current research that impinges on signal processing, information theory, and computer vision. On the other hand, human vision carries out its processes at several different conceptual and anatomical-physiological levels. Lower level vision has to do with early processing of the retina image and thus with physical parameters of the image, such as contrast, spatial frequency, and color. In computer vision, much of the careful work going on at this lower level may be considered as pre-processing of the image for later stages. Middle level vision and binocularity have to do with the construction or interpretation of 3D representation of 2D information arriving onto the two retinas. Higher level vision is largely a top-down process and involves both symbolic non-iconic representation and operational top-down spatial and sequential instructions to eye movements in order to match representations in an iconic fashion to the subfeatures (Stark and Choi, 1996). Explorations are now ongoing with the use of bottom-up image processing and image classification algorithms to substitute for some aspects of top-down recognition (Privitera and Stark, 1998).

Matched filters, MFs, are an efficient way for extracting signals from noise (Turin, 1960). In the early 60's an MIT group developed the use of multiple adaptive filters for classifying sequences of time series events (Stark et al, 1962). Historically, the above series of papers started with simple digital Fourier analyses of biological noise. Here a series of sinusoidal and cosinusoidal waves were cross-correlated, frequency by frequency and time shift by time shift, against the noise time series for which a spectrum was desired. However, by using template filters or MFs constructed as inverse time functions of event shapes, episodic temporal events in the time series were better identified. It then appeared as if a sinusoid could also be considered as a MF for performing frequency

analysis. Recent work from our laboratory includes the application of top-down MFs in computer visual search (Stark et al, 1992) and in 2D pattern recognition (Sun and Stark, 1994, personal communication).

The enemies of easy image recognition are noise, blur and clutter. In the experiments and simulations to be presented, we have used a number of levels and types of noise and a number of levels of blur. We then studied their effect on the ease of image recognition using human subjects and also computer algorithms. Clutter is a more complex matter. Although noise was originally defined as someone else's telephone conversation interfering with one's own; it has now come to be defined in terms of stochastic processes. Thus, interfering signals with patterns similar to the desired signals have had to be renamed; in the image world, the term clutter appears to have been assigned this role.

The aim of present paper is to study similarities and differences in top-down human and computer vision.

2. METHODS

Target and Decoy Images Binary test images for the experiment, 'car', 'van' and 'truck' (Figure 1), were 64 x 64 pixels or 110 x 110 mm in size. These three kinds of vehicles with added noise and blur were used as the targets. Targets free of noise or blur were 60 x 30 mm for 'large' type (left column) and 30 x 15 mm for 'small' type (right column).

Figure 1. Target and Decoy Icons

Three large (left column) and small (small) icons. Coordinates in pixels. Note binary luminance and simple vector line drawings.

Blur and Noise Added. We added three types of noise: 'plus noise' (a cross five pixels in length); 'salt noise' (one pixel set to 255); and 'salt and pepper noise' (one pixel with polarity reversed). Control experiments indicated that the effects of different noise types were similar. Noise ratios were set from 0 to 60 %; the fraction of changed pixels times 100 was the per cent. For each displayed image condition, the type of vehicle and noise type and level was chosen randomly. Stimulus pictures with target and added noise were then blurred to one of seven levels of sigma (see below), also chosen randomly (Figure 2). Clearly, both noise and blur added difficulties for both human and computer recognition processes attempting to identify the target icon type displayed.

Figure 2. Added Noise and Blur

Car with cross-noise (upper row) and blurred with sigma equal to 2 (right); truck with salt noise, 30% (middle row) and blurred with sigma equal to 6 (right); van with salt and pepper noise, 60% (lower row) and blurred with sigma equal to 12 (right).

A Fourier transform method was used to blur the target icons (Figure 3). By multiplying the Fourier transform (upper right) of the original picture (upper left) by the Fourier transform (middle right) of the blurring aperture (middle left) the product (lower right) also in Fourier space was obtained. Then by inverse Fourier transform, a blurred picture back in the spatial domain (lower left) was obtained that could be presented to either human or computer vision processes. Direct convolution of the picture and the aperture is more computationally demanding. The extent of blur is defined by the sigma, the standard deviation of the Gaussian distribution (middle left). The sigma varied from zero to twelve pixels in extent, in increments of two pixels.

Figure 3. Blurring Methodology

Target plus cross noise (upper right) and its Fourier transform (right); aperture (middle row) and its Fourier transform (right). Fourier product of target and aperture (lower row, right) with inverse transform back to spatial domain (left).

Experimental Arrangements. The distance between the subject's eyes and the computer display was set to be 0.5 m. The 64 x 64 pixel stimulus display window on the computer screen was 110 x 110 mm and thus 12.5 degrees of visual angle, considering that one radian equals the 0.5 m distance (Figure 4). The thickness of the

vehicle contour lines were only 1 pixel or 6.8 arc minutes; to place this in visual units, recall that 20/20 vision requires resolution of one arc minute.

Figure 4. Experimental Arrangement

Note distance of subject from display and sizes of images and icons that together define visual angles (see text).

Experimental Protocol The number of target conditions was 266; this was determined by the three types of noise, seven noise levels, seven blur levels, and two target sizes (Figure 5); note that zero noise was one condition for all three types of noise. For each target condition, ten instances of it were randomly presented throughout the experiment to the subject for one second. It then disappeared, leaving only a cross mark on the screen to aid in fixation. The subject made a 'forced choice' signaled by pressing one of three keys, so as to identify which one of the three targets had been presented, even if he felt that he could not make the discrimination. After the subject responded, the next image appeared. The experiments on the large and small targets were separately conducted. While a number of control experiments were done on several subjects, all of the data reported here were collected from one subject in about ten two-hour long sessions. Much of the time was employed in the computer recalculating target conditions because of computer memory limitations.

Figure 5. Multiplicity of Targets, Noise and Blur

Note two sizes of targets (left column), three types of noise (next column) seven levels of noise (next column), and seven levels of blur sigma (right column) yielding 266 target images in all.

Computer Vision Experiment For each condition, the cross-correlation between the possibly noisy and blurred image and each of the six templates was calculated and the vehicle type with the highest correlation was chosen as the recognition response. It should be noted that the MFs for recognition and the target icons were all centered with respect to the 64 x 64 pixel display image area; we did not try to solve the location approximation problem here (Sun and Stark, 1994, personal communication). The fraction of correct responses was adopted as probability of recognition, PR. For comparison with human vision, this MF method was expanded; several somewhat different MFs were generated all employing the template matching methodology. Their comparisons with human vision and with each other proved to be most interesting in terms of defining a likely form for the human matched filter.

3. RESULTS

Human Forced Choice The surface (Figure 6, lower left) represents the recognition behavior of a human subject deciding as to which of the presentations is the target car and correctly rejecting the decoy icons, truck and van. The amount of added noise, in percent of pixels converted to noise signal is indicated on the y-axis and the amount of blur, in size of the spread function sigma in pixels is indicated on the x-axis. The performance, as percent of correct choices, is shown on the vertical, z-axis; 33% is chance level, 100% is perfect choice. A level of 67% represents behavior halfway between these extremes. This level of the surface is indicated by the 'performance function' (thick line) as is its projection on the noise-blur plan. Note very successful performance with low noise and no blurring and only random levels of choice for large amounts of noise and blurring; thus we achieved a full range of behaviors with our domains of blur and noise. Also, there is a slope to the 2-D performance function indicating a trade-off so that fairly successful behavior is possible with much noise and little blurring or with little noise and much blurring. Human vision is quite sensitive to noise; possibly the noise is very effective in destroying shape consistency for recognition. We noted in earlier control experiments the type of noise, when normalized by number of pixels modified, does not seem to be particularly important. Human vision is fairly robust to blur. Remember from information theory that blur does not destroy information but just smears or reallocates it over the image. Blurred information can be reconstructed using lenses for real images or by spatial differentiation. Especially, in regions of small amounts of noise, the human appears to be able to reconstruct the image even in spite of the largest amount of blur with sigma equal to twelve. Size of the pattern to

be recognized, also explored in earlier control experiments, did not seem to interact with the effect of the blur or the ability of the human to reconstruct the image for recognition.

Figure 6. 3D Performance Surfaces

Coordinate axes: Noise in per cent pixels occupied by any of the noise types, and Blur, with sigma in pixels. Vertical Performance scale ranged from perfect recognition success, 100%, to chance levels, 33%. Note "Performance Function" line (heavy line) representing the intersection of the performance surface with the plane at 67% success, halfway between extremes; also the projection of this performance line onto 2D Noise-Blur Plane. Human performance (lower left), Computer performance (lower right), and Joined performance function (upper).

Computer Vision Again the 3D surface (see Figure 6, lower right) can represent the MF choice of the target icon from the decoy icons. Note successful behavior is prominent with low noise and little blurring and choice failure to rise above the chance level is exhibited at high noise levels with much blurring. Trade-off between levels of these two destructive processes can also be seen. In general, noise and blur, two basic characteristics of image degradation, act in similar fashion for both human and computer vision (Figure 6, upper; joining the results of human and computer vision).

Comparison between human and computer vision The 'performance function', the projection of the 67% success line onto the 2D noise-blur plane represents a cut of the 3D recognition surface with the horizontal plane at that level of recognition. This function clarifies the similarities and differences between human (dotted lines) and computer vision (solid lines). Both human and computer functions show trade-off between the two image degradation procedures as expected (Figure 7, upper, human; and lower, computer). The region labeled "b" represents a portion of the plane wherein computer MF processes are more successful in recognition than the human. The region labeled "a" represents a portion of the plane wherein human recognition processes are more successful than the computer MF algorithm. Region "a" defines a low-noise region of the noise-blur plane with a great deal of blur; evidently the human has the capacity to carry out spatial differentiation to overcome some of the blur degradation from the image.

Figure 7. Performance Functions: Matched Filter, MF1, Entire Image Area

Human (upper and dotted lines), computer MF1 (lower and solid lines) and both (middle) performance functions indicating similarities and differences. See text for discussion of regions "a" (vertical hatching) and "b" (oblique hatching). Note the blur is the blur of the target images, not of MF1 (see text).

4.- DISCUSSION

Varying the matched filters. The standard algorithm, MF1, "Entire Image Area", that has been the source of our results above, demonstrated similarities and differences from human vision. To recall, MF1 used the entire image area, either about three times larger in area than the large targets and icons or six times larger in area than the small targets and icons. Examples of these MF1s (Figure 1) yielded the performance and noise-blur plane (Figure 7, lower).

In order to extend and understand those findings, we constructed three additional MF algorithms: MF2, "(within) Contoured Area Only"; MF3, "Vector Lines (of icons) Only"; and MF4, "Blurred Lines (of icons)". MF2 was restricted to the area within the contours (Figure 8, middle row) for the cross-correlation comparisons shown in the noise-blur plane (Figure 8, lower). MF3 employed only the vector graphic lines of the target icons (Figure 9, middle row). MF4 was also restricted to the vector graphic lines of the icons, but these were blurred to the full extent of the image space (Figure 10, middle row). These different properties provided interesting results.

Figure 8. Performance Functions: Matched Filter, MF2, Contoured Icon Area Only

Target image (upper being matched by three MF2 templates equal to only the contoured area (middle row); this was set to be the rectangular area shown in black. Comparison of MF2 (solid line) with human (dotted line (lower). Also see text.

Figure 9. Performance Functions: Matched Filter, MF3, Vector Lines Only

Target image (upper being matched by three MF3 templates equal to only the vector lines (middle row). Comparison of MF3 (solid line) with human (dotted line (lower). Also see text.

Figure 10. Performance Functions: Matched Filter, MF4, Blurred Lines

Target image (upper being matched by three MF4 templates equal to the lines blurred (middle row). Comparison of MF4 (solid line) with human (dotted line (lower). Note close fit! Also see text.

Performance in the noise-blur plane. The result of these additional MFs when cross-correlated with the target images with the additional blur and noise (Figure 8, 9 and 10, upper-panels) are shown in the noise-blur planes (Figure 8, 9 and 10, lower-panels, solid lines). Also shown in these several lower panels are the result of human vision experiments (dotted lines) for comparison; these functions have also been brought together in one figure (Figure 11).

Figure 11. Comparison of Performance Functions of the Four MFs

Comparison of MFs (solid lines) with human (dotted lines). Recall that MF1 used the entire image area. MF2 was restricted to within the contoured icon area only; and MF3 employed only the vector lines of the icons. MF4 used the lines of the icons, yet these were blurred to the full extent of the image space. The properties provided by these different definitions yield the comparative results of this figure. Graphs taken from Figures 7, 8, 9, and 10 that also illustrate filter patterns. Regions "a", "a-prime", "b", "b-prime" and "c" provide insight into the human MFs postulated to exist in top-down human vision and are described in text.

Human vision robustness to blur. When we compare human vision with computer or MF vision, we see that the human performs better in recognition tasks in region "a" of very high blur and low noise in the lower right portion of the noise-blur plane (Figure 11, upper left, MF1). This is consistent for all four MF types. Recall the closer the performance line is to upper right 'failure' corner of the noise-blur plane, the more resistant the recognition processes are to noise and blur. The phenomenon of restricted computer vision MF sensitivity to blur is especially prominent for MF2 and MF3; note regions "a-prime". These two MFs, using only the area within the contour, MF2, or only the vector graphic lines, MF3, become more sensitive to blur even in the regions of low noise.

Computer vision robustness to noise. The original MF1 performed better than human vision in the region of high noise and low blur (Figure 11, upper left, region "b"). This performance advantage shrunk to very tiny remnant for the other three MFs that operated in only limited areas and over vector graphic lines, region "b-prime". There was even reversal shown where MF2 and MF4 of performed not as well as human vision in another tiny region "c". Evidently, using the entire image space made MF1 robust to noise as compared to human and thus it would be the selected filter for operation in the high noise range.

Matching human vision with a matched filter. Finally note the close approximation of MF4 to human vision (Figure 11, lower right); not only are the two performance curves almost identical, but they show the same trade-off slope between noise and blur. Recall that MF4 consists of blurred iconic lines; this suggests that likely the internal human recognition icons, over-learned for this repetitive recognition task, have something of the nature of widely blurred iconic lines.

Human Vision Matched filters are top-down 'a priori' models, that is, previously known shapes of objects to be compared with bottom-up information --- in our experiment the noisy and blurred target and decoy icons. Human vision also has top-down and bottom-up processes. First there is the scanpath, a top-down procedure for visiting important subfeatures of a picture and secondly, something akin to iconic MFs, for comparing with visual information traveling to the brain from the eye (Stark and Choi, 1996). Scenes in pictures are often complex with many simpler subfeatures geometrically and meaningfully arranged in space. Human vision has adapted to these types of informational signal presentations by arranging to scan from subfeature to subfeature. Onto each subfeature, an eye movement fixation superimposes the high-resolution fovea of the eye. Our 'a priori' expectations as to the locations of the subfeatures give rise to repetitive sequences of refixation saccades termed

the "scanpath". Our 'a priori' expectations for an individual subfeature can be thought of as having the appropriate MF available for cross-correlation with the bottom-up image information. Perhaps this takes place via a feedback interaction between layers 1,2 and 3 (the top-down locus) with layers 4 and 5 (the bottom-up locus) of the visual cortex (Stark et al, 1999). Consideration of the likelihood of the serial versus parallel nature of the scanpath in top-down human vision (Noton and Stark, 1971) could be applied to computer vision. Here, parallel processes could more easily, or perhaps with some additional computer complexity, visit the seven, plus or minus two, subfeatures of the picture or object in a synchronous manner.

5. CONCLUSION

Experiments are presented that provide insight into the efficacy of computer vision algorithms, such as matched filters to overcome noise and blur in images in a characteristic, but simplified recognition task. Human vision for the same task has also been studied. Close comparison of these sets of results in the 'Noise-Blur' plane demonstrates similarities and differences between computer vision and human vision. Both computer and human use top-down vision for recognition, 're-knowing' a pattern, both have similar sensitivities to noise and to blur, and both demonstrate trade-off between these two stressors. By varying the nature of the template matched filters, we believe we have come close to understanding the approximate forms that the human 'internal spatial-cognitive models' for this task may take within the iconic comparison mechanism of the visual cortex.

ACKNOWLEDGEMENT

We wish to thank Technical Monitors, Dr Stephen Ellis, NASA-Ames Research Center and Dr. David Dixon, TRAC, White Sands for advice and partial support and our laboratory colleagues and especially, Dr. Claudio Privitera for incisive comments.

REFERENCES

- Noton, D. and Stark, L.W., *Scanpaths in Eye Movements during Pattern Perception*,
Science 171: 308-311 (1971) [105]
- Claudio M. Privitera, C.M. and Stark, L.W., *Evaluating Image Processing Algorithms that Predict Regions of Interest*,
Pattern Recognition Letters 19: 1037-1043 (1998) [393]
- Stark, L.W. and Choi, Y.S., *Experimental Metaphysics: The Scanpath as an Epistemological Mechanism*,
In: *Visual Attention and Cognition*, editors, Zangemeister et al, Elsevier Press, Oxford. Pp. 3-69, (1996) [364]
- Stark, L.W., Okajima, M. and Whipple, G.H., *Computer Pattern Recognition Techniques: Electrocardiographic Diagnosis*,
Comm. Assoc. Computing Machinery, 5: 527-532, 1962 [18]
- Stark, L.W., Privitera, C.M., Yang, H.Y., Ho, Y.F., Azzariti, M., Chan, A., Krischer, C., and Weinberger, A.,
Scanpath Memory Binding: Multiple Read-out Experiments, *Proceedings of SPIE* 6: 291-304 (1999) [396]
- Stark, L.W., Yamashita, I.H., Tharp, G., Ngo, H.X., *Searchpatterns and Searchpaths in Human Visual Search*,
in: *Visual Search: II*, editors, Brogan, D. and Carr, K., Taylor & Francis, London. Pp. 37-58 (1992) [351]
- Turin, G.L., *An Introduction to Matched Filters*, IRE PGIT-6, 311 (1960)
-

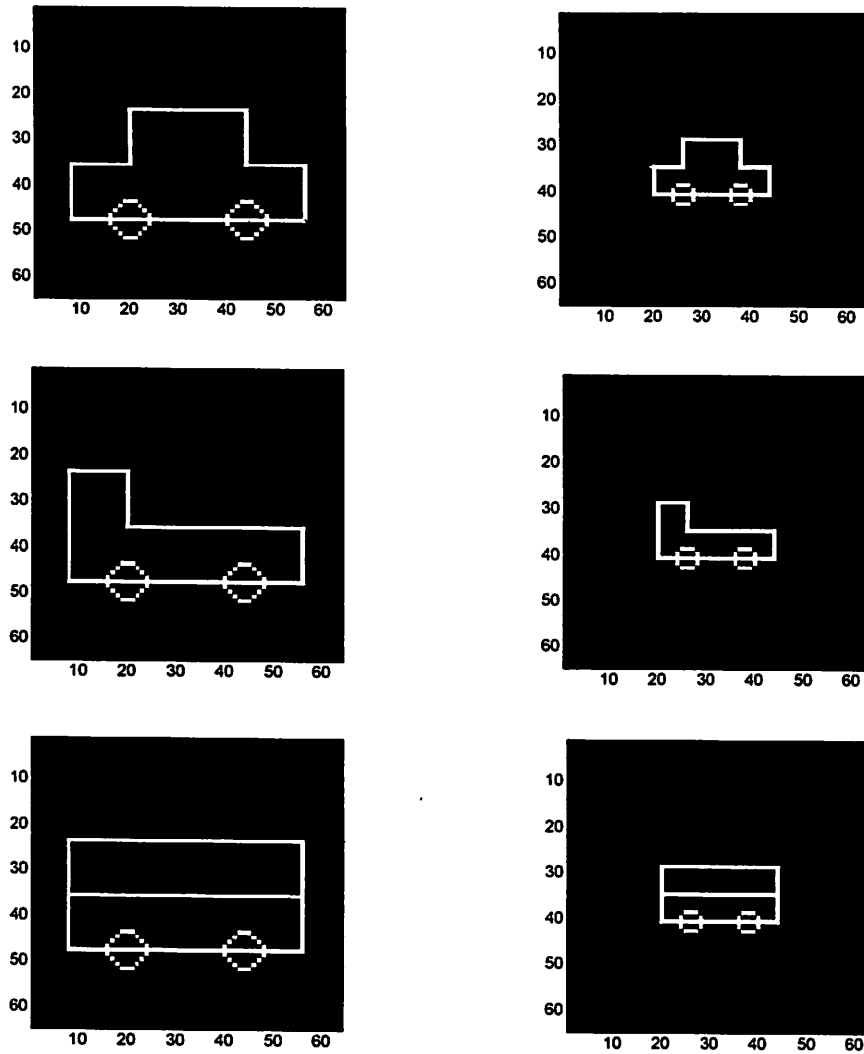


Figure 1. Target and Decoy Icons

Three large (left column) and small (small) icons. Coordinates in pixels. Note binary luminance and simple vector line drawings.

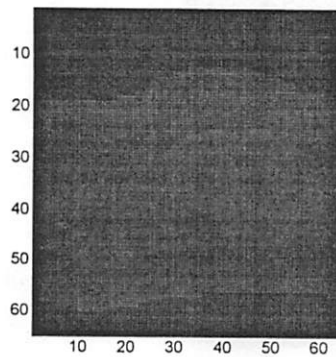
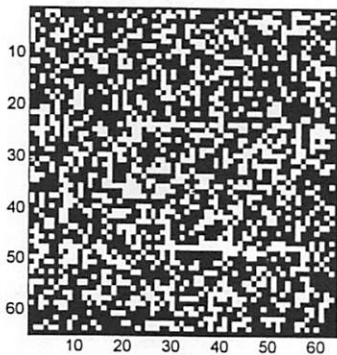
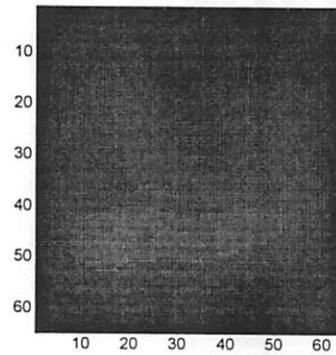
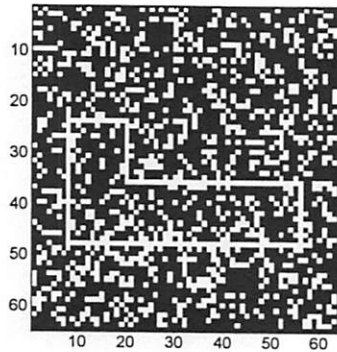
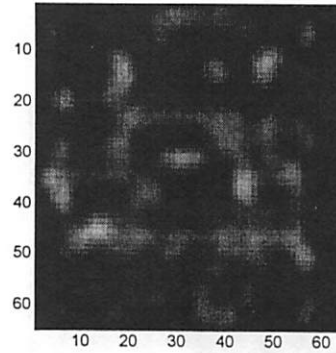
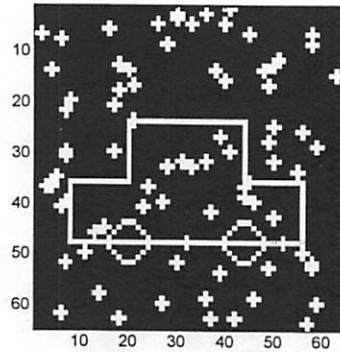


Figure 2. Added Noise and Blur

Car with cross-noise (upper row) and blurred with sigma equal to 2 (right); truck with salt noise, 30% (middle row) and blurred with sigma equal to 6 (right); van with salt and pepper noise, 60% (lower row) and blurred with sigma equal to 12 (right).

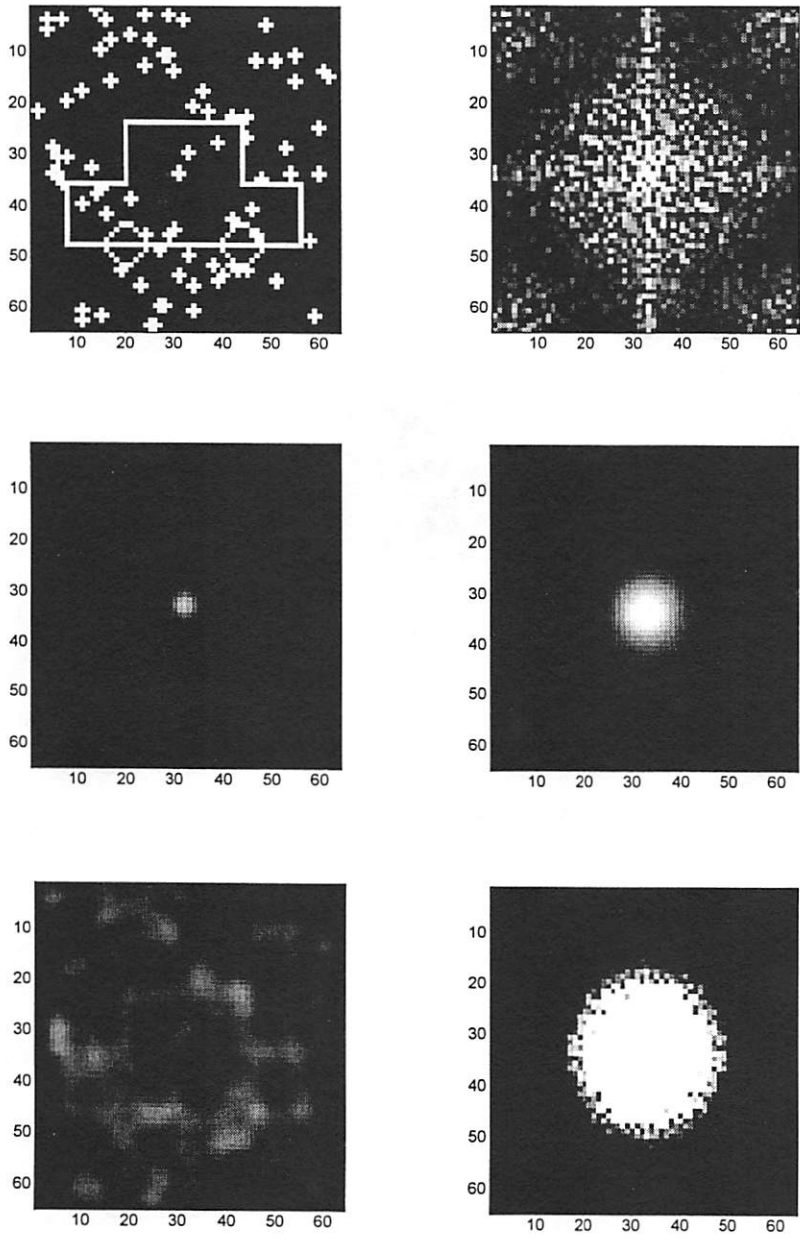


Figure 3. Blurring Methodology

Target plus cross noise (upper right) and its Fourier transform (right); aperture (middle row) and its Fourier transform (right). Fourier product of target and aperture (lower row, right) with inverse transform back to spatial domain (left).

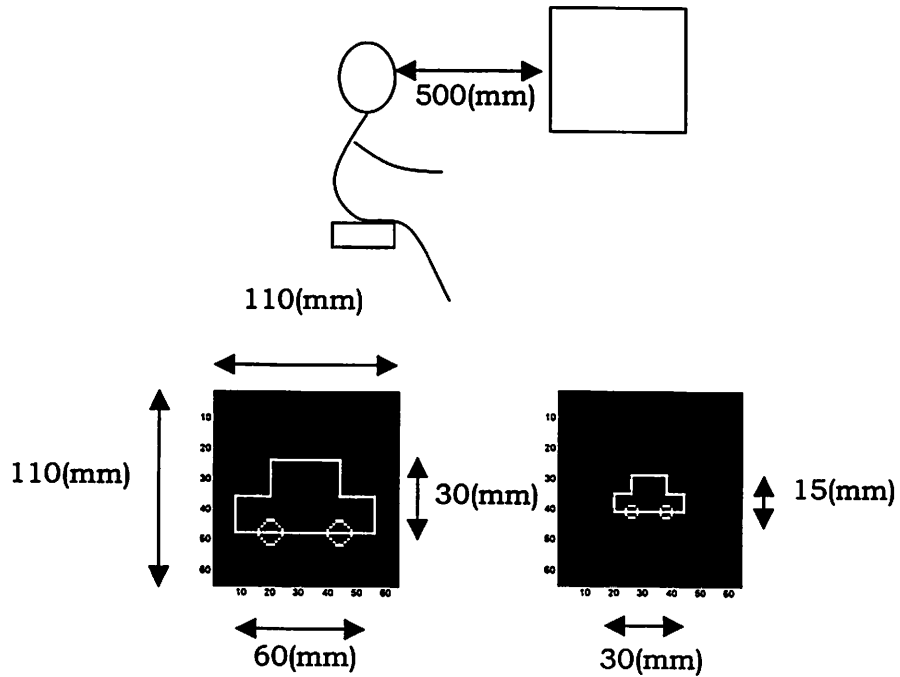


Figure 4. Experimental Arrangement
 Note distance of subject from display and sizes of images and icons that together define visual angles.

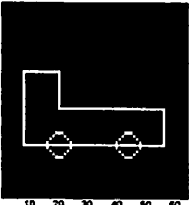
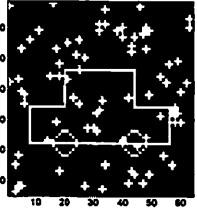
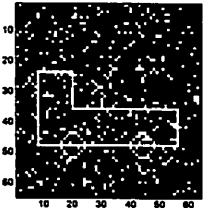
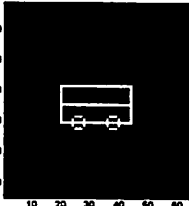
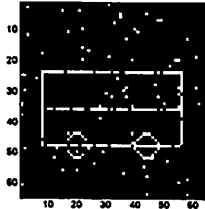
Vehicle Size	Noise Type	Noise Ratio	Blur Sigma
 <p>Large</p>	 <p>Plus</p>	10%	0
	 <p>Salt</p>	20%	2
		30%	4
		40%	6
		50%	8
		60%	10
		0%	12
 <p>Small</p>	 <p>Slat & Pepper</p>		
	W/O Noise	0%	

Figure 5. Multiplicity of Targets, Noise and Blur

Note two sizes of targets (left column), three types of noise (next column) seven levels of noise (next column), and seven levels of blur sigma (right column) yielding 266 target images in all.

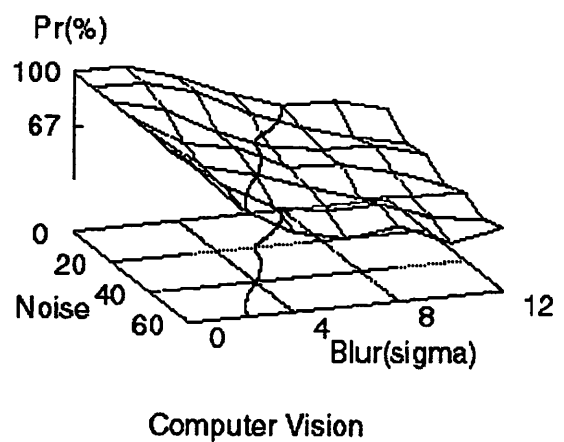
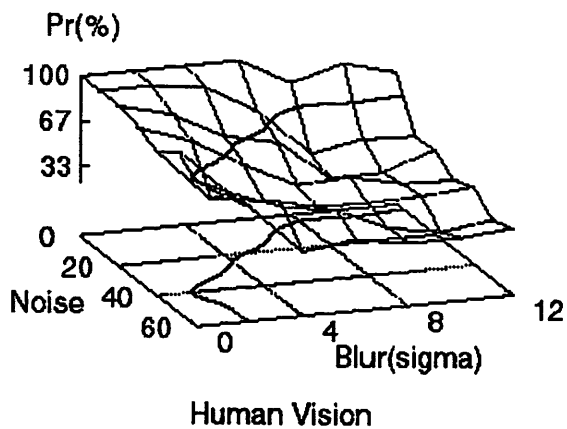
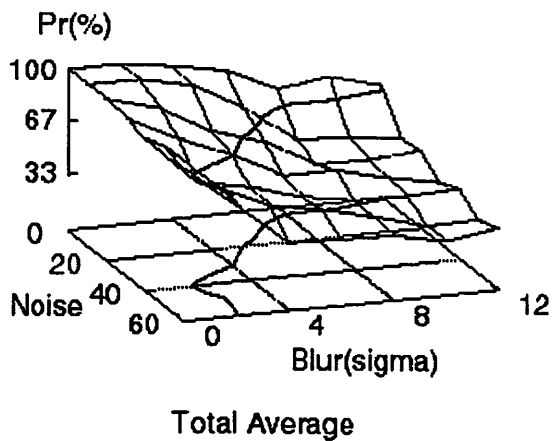


Figure 6. 3D Performance Surfaces

Coordinate axes: Noise in per cent pixel occupied by any of the noise types, and Blur, with sigma in pixels. Vertical Performance scale ranged from perfect recognition success, 100%, to chance levels, 33%. Note "Performance Function" line (heavy line) representing the intersection of the performance surface with the plane at 67% success, halfway between extremes; also the projection of this performance line onto 2D Noise-Blur Plane. Human performance (lower left), Computer performance (lower right) and joined performance function (upper).

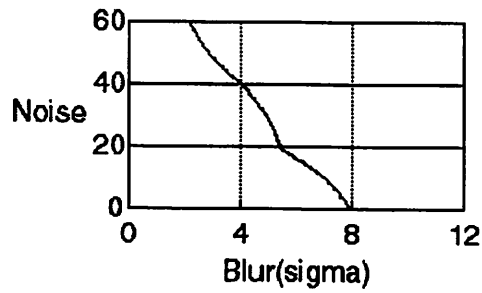
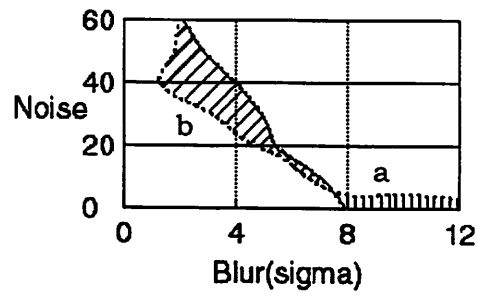
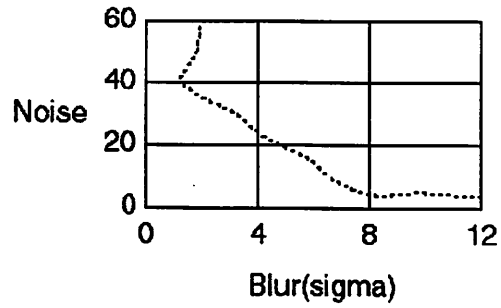


Figure 7. Performance Functions: Matched Filter, MF1, Entire Image Area

Human (upper and dotted lines), computer MF1 (lower and solid lines) and both (middle) performance functions indicating similarities and differences. See text for discussion of regions "a" (vertical hatching) and "b" (oblique hatching). Note the blur is the blur of the target images, not MF1 (see text).

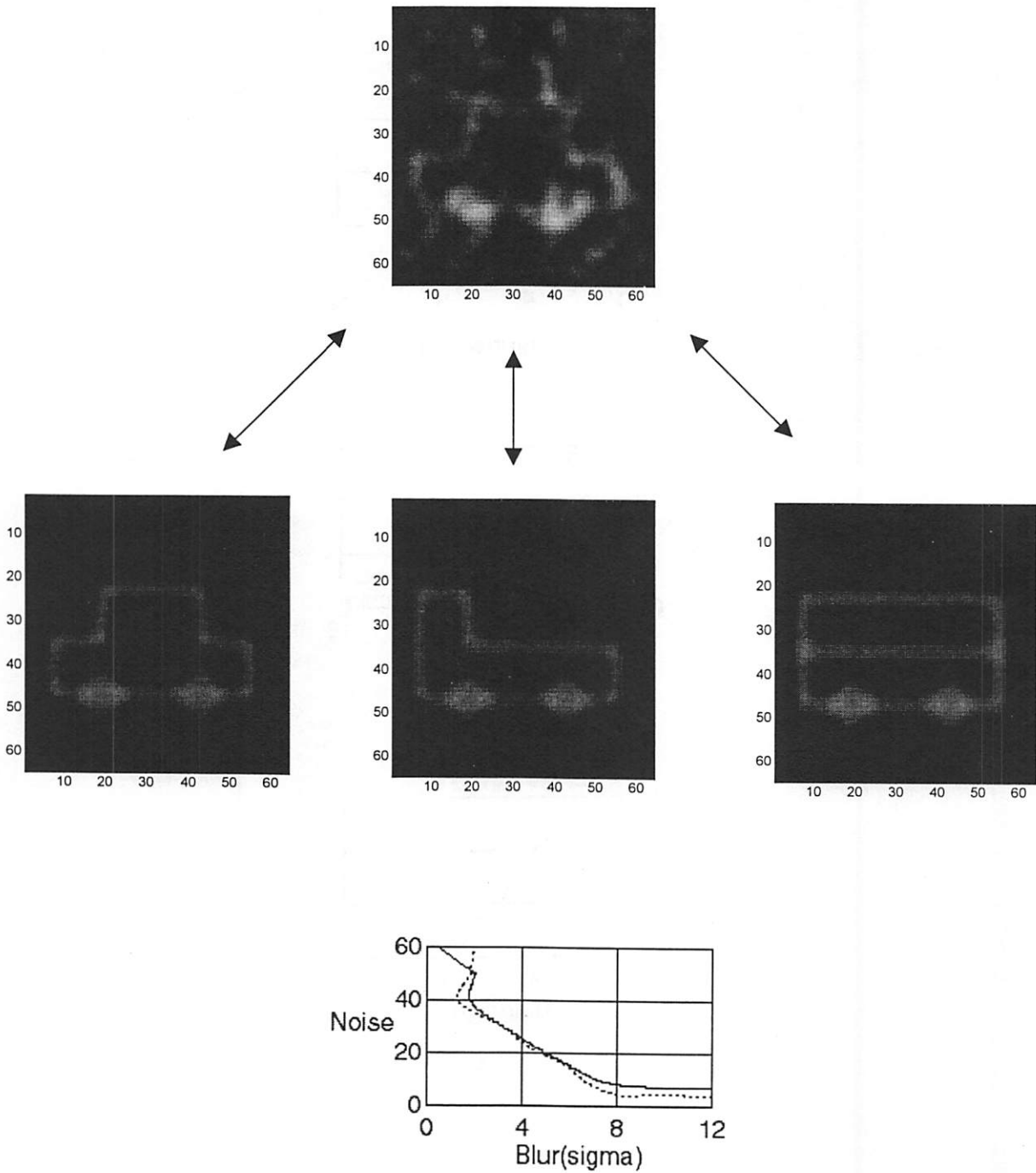


Figure 10. Performance Functions: Matched Filter, MF4, Blurred Lines

Target image (upper being matched by three MF4 templates equal to the lines blurred (middle row). Comparison of MF4 (solid line) with human (dotted line (lower). Note close fit! Also see text.

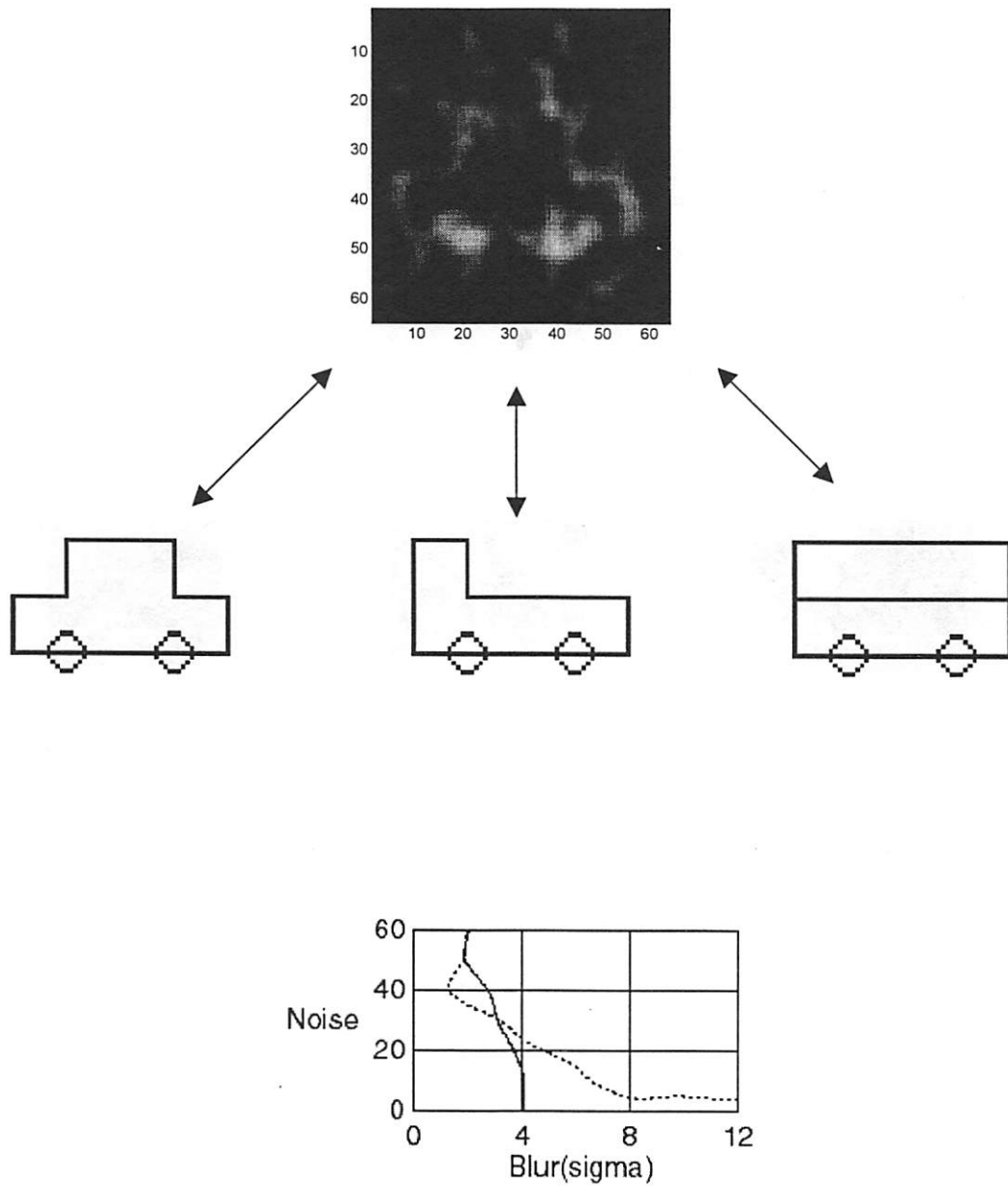


Figure 9. Performance Functions: Matched Filter, MF3, Vector Lines Only

Target image (upper being matched by three MF3 templates equal to only the vector lines (middle row). Comparison of MF3 (solid line) with human (dotted line) (lower). Also see text.

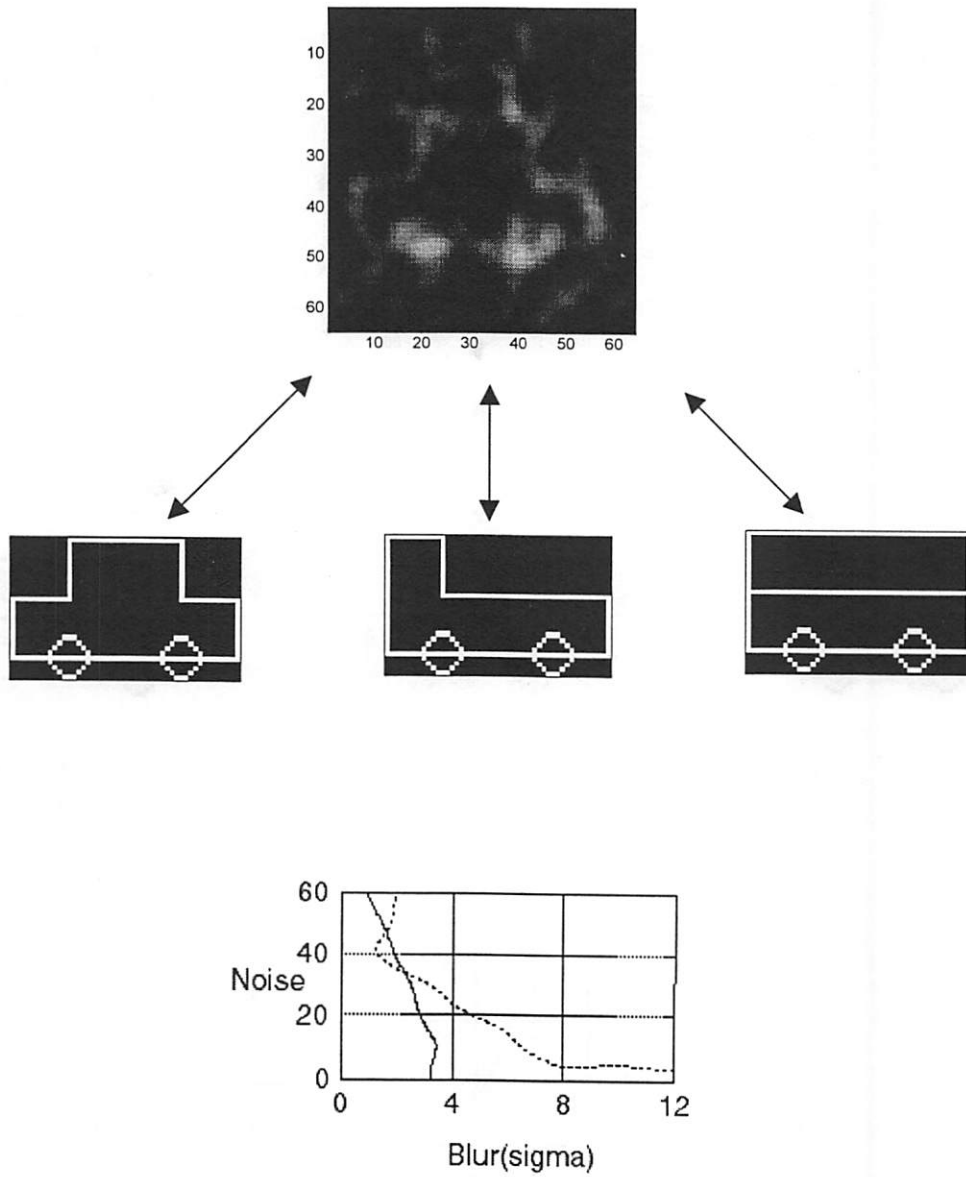


Figure 8. Performance Functions: Matched Filter, MF2, Contoured Area Only

Target image (upper being matched by three MF2 templates equal to only the contoured area (middle row); this was set to be the rectangular area shown in black. Comparison of MF2 (solid line) with human (dotted line (lower). Also see text.

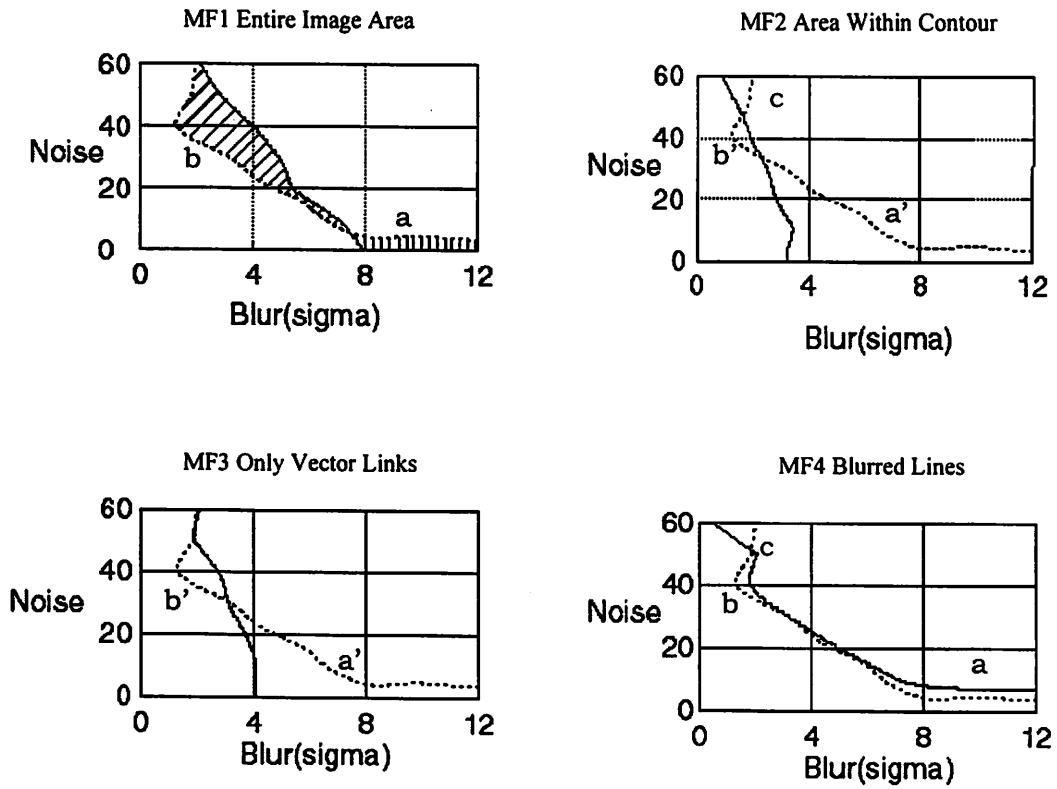


Figure 11. Comparison of Performance Functions of the Four MFs

Comparison of MFs (solid lines) with human (dotted lines) (lower). Graphs taken from Figures 7, 8, 9, and 10 that also illustrate filter patterns. Regions "a", "a-prime", "b", "b-prime" and "c" provide insight into the human MFs postulated to exist in top-down human vision and are described in text.