

Copyright © 1997, 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.

**ALGORITHMS FOR DEFINING VISUAL
REGION-OF-INTEREST:
COMPARISON WITH EYE FIXATIONS**

by

Claudio Privitera and Lawrence W. Stark

Memorandum No. UCB/ERL M97/72

20 September 1997

**ALGORITHMS FOR DEFINING VISUAL
REGION-OF-INTEREST:
COMPARISON WITH EYE FIXATIONS**

Copyright © 1997

by

Claudio Privitera and Lawrence W. Stark

Memorandum No. UCB/ERL M97/72

20 September 1997

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Algorithms for Defining Visual Region-of-Interest:

Comparison with Eye Fixations

Claudio Privitera and Lawrence W. Stark

Neurology and Telerobotics Units

486 Minor Hall, University of California, Berkeley 94720-2020

Abstract

Machine vision often analyzes in detail only a subset of the picture that may be arranged into a sequences of loci called, regions-of-interest, ROIs; preprocessing algorithms for spatial frequency, texture conformation etc. may be used to selected these loci. In human perception an internal representation directs *top-down*, context-dependent, sequences of eye movements to fixate on a similar sequences of ROIs. We compare these two sequences of ROIs as a criteria for evaluating and studying *bottom-up*, context-free, algorithms.

1 Introduction

Image processing, IP, algorithms are usually intended to detect and localize specific features in a digital image in a *bottom-up* fashion, analyzing for example, spatial frequency, texture conformation or other informative values of loci of the visual stimulus. Many algorithms have been proposed in the literature and they might be classified into three principal approaches; for a survey, see Haralick [6] and Reed and Hans Du Buf [9]. Firstly, *structural* approaches based on an assumptions that images have detectable and recognizable primitives distributed according to some placement rules; examples are matched filters. Secondly, *statistical* approaches based on statistical characteristics of the texture of the picture; examples are co-occurrence matrices

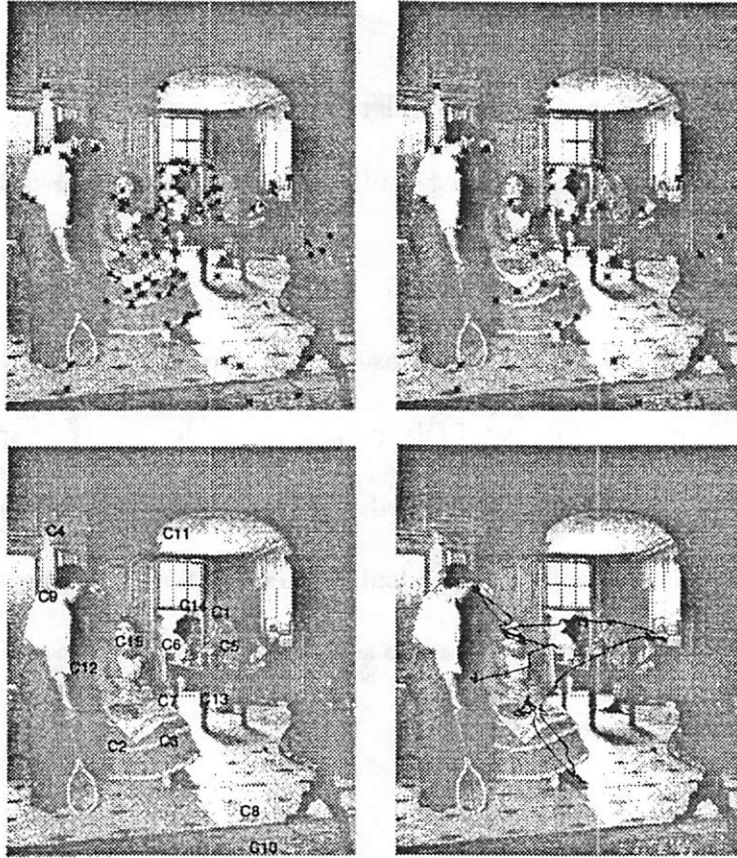


Figure 1: Image Processing and Clustering Algorithms. IP algorithms such as \mathcal{X} here, lead to a large number of candidate high valued loci (upper left). An iterative clustering algorithm then reduced the candidate loci (circa 100, upper left) through intermediate stages (upper right) to a few ROIs (less than 15, lower left) ordered by value. Note similarity to sequential fixations generated by eye movements, EMs, (lower right).

and entropy functions. Thirdly, *model* approaches that hypothesize underlying processes for generation of local regions. These are analyzed on the basis of specific parameters governing these generators: examples are fractal descriptors.

For the purpose of our present study, we have selected elements from this taxonomy in an attempt to simulate certain aspects of human perception. Consequently, our IP algorithms are set a definite task.

From a human perspective, the scanpath theory, Noton and Stark [8], suggests that a *top-down* internal cognitive model of what we *see* controls perception and active looking eye movements, EMs. These EMs are an essential part of vision because they must carry the fovea to each part of an image to be processed with high resolution. Thus, the internal cognitive model drives our EMs in a repetitive sequential set of saccades and fixations, or glances, over features, or regions of interest, ROIs, of a scene or picture so as to check out and confirm the model; supporting evidence is reviewed by Stark and Choi [12].

Thus our aim is explicit and our measures quantitative. The over-riding question is whether IP algorithms can treat a picture in a fashion similar to human sequential glimpses.

2 Stimulus presentation and eye movement measurement

Computer controlled experiments present pictures and carefully measure EMs using video cameras [12, 3]. EMs data are then classified into sequences of alternating rapid jumps or saccades and fixations generally lasting about 0.3 second: this enables the high resolution fovea to abstract information from the fixated ROIs (Figure 1, lower right panel).

Two different pictures, *After the Shower*, (Figure 1) and *Madame* (Figure 2) were used; picture modification yielded four different types (Figure 2) and these eight stimulus, two pictures times four types, were not only presented to the four human subject with several repetitions each, but were also processed by all the IP algorithms described below.

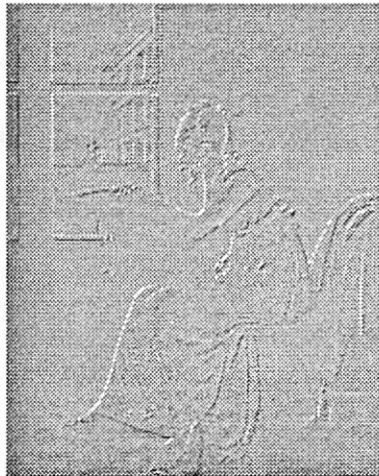


Figure 2: Picture Modification Yielding four Different Types. Original (upper left); line drawing (upper right); embossing modification (lower left); binary modification (lower right).

3 ROI detecting algorithms

The information content of a generic picture can be identified by different image parameters which in turn can be identified by relevant image processing algorithms. In this sense, applying algorithms to a picture means to map that image into different domains, where for each domain, a specific set of parameters is extracted. After the image has been processed, only the loci of the local maxima from each domain are retained; these maxima are then clustered in order to yield a limited number of ROIs. The algorithms we were studied are:

1 – \mathcal{X} , an x -like mask, positive along two diagonal and negative elsewhere, were convolved with the image. We have also used different high-curvature masks convolution as for example the " $<$ "-like mask whose the definition is intuitive: these were rotationally invariant.

2 – \mathcal{S} , symmetry, a structural approach, appears to be a very prominent spatial relation. For each pixel x, y of the image, we define a local symmetry magnitude $\mathcal{S}(x, y)$ as follows:

$$\mathcal{S}(x, y) = \sum_{(i_1, j_1), (i_2, j_2) \in \Gamma(x, y)} s((i_1, j_1), (i_2, j_2)) \quad (1)$$

where $\Gamma(x, y)$ is the neighborhood of radius 7 of point x, y defined along the horizontal and vertical axis ($\Gamma(x, y) = (x - r, y), \dots, (x, y), \dots, (x + r, y), (x, y - r), \dots, (x, y + r)$) and $s((i_1, j_1), (i_2, j_2))$ is defined by the following equation:

$$s((i_1, j_1), (i_2, j_2)) = G_\sigma (d((i_1, j_1), (i_2, j_2))) |\cos(\theta_1 - \theta_2)| \quad (2)$$

The first factor G_σ is a gaussian of fixed variance, $\sigma = 3$ pixels and $d(\cdot)$ represents the distance function. The second factor represents a simplified notion of symmetry: θ_1 and θ_2 correspond to the angles of the gray level intensity gradient of the two pixels (i_1, j_1) and (i_2, j_2) . The factor achieves the maximum value when the gradients of the two points are oriented in the same direction. The gaussian represents a distance weight function which introduces localization in the symmetry evaluation [10] (see also [5]).

3 – \mathcal{W} , discrete wavelet transform is based on a pyramidal algorithm which split the image spectrum into four spatial frequency bands containing horizontal lows/vertical lows (ll), hori-

zontal lows/vertical highs (*lh*), horizontal highs/vertical lows (*ll*) and horizontal highs/vertical highs (*hh*). The procedure is repeatedly applied to each resulting low frequency band resulting in a multiresolution decomposition into octave bands. The process of the image wavelet decomposition is achieved using pair of conjugate quadrature filters (CQFs) [14] which acts as a smoothing filter (i.e. a moving average) and a detailing filter respectively (see for example [11] for details). We have used different orders from the Daubechies family basis, [4], to the define CQF filters. For each resolution i , only the wavelet coefficients of the highs/highs hh_i matrix were retained and finally relocated into an final matrix HH (with the same dimension as the original image) by the following combination:

$$HH = \sum_{i=1}^n \zeta^i(hh_i) \quad (3)$$

where n is the maximum depth of the pyramidal algorithm ($n = 3$ in our case) and where $\zeta(\cdot)$ is a matrix operation which returns a copy of the input matrix hh by inserting alternatively rows and columns of zeros.

4 – \mathcal{F} , a center-surround *on/off* quasi-receptive field mask, positive in the center and negative in the periphery, is convolved with the image.

5 – \mathcal{O} , difference in the gray-level orientation, a statistical-type kernel, is analyzed in early visual cortices. Center-surround difference is determined first convolving the image with four gabor masks of angles $0^\circ, 45^\circ, 90^\circ$ and 135° respectively (see also [7]). For each pixels x, y , the scalar result of the four convolutions are then associated with four unit vectors corresponding to the four different orientations. The orientation vector $\vec{o}(x, y)$ is represented by the vectorial sum of these four weighted unit vectors. We define the center-surround difference transform as follows:

$$\mathcal{O}(x, y) = (1 - \vec{o}(x, y) \cdot \vec{m}(x, y)) \|\vec{o}(x, y)\| \|\vec{m}(x, y)\| \quad (4)$$

where $\vec{m}(x, y)$ is the average orientation vector evaluated within the neighborhood of 7×7 pixels. The first factor of the equation achieves high values for big differences in orientation between the center pixel and the surroundings. The second factor acts as a low-pass filter for

the orientation feature.

6 – \mathcal{E} , concentration of edges per unit area is determined by detecting edges in an image, using the canny operator [2] and then congregating the edges detected with a gaussian of $\sigma = 3$ pixels.

7 – \mathcal{N} , entropy is calculated as $\sum_{i=0}^{255} p_i \log p_i$ where p_i is the probability of the gray level i within the 7×7 surrounding of the center pixel.

8 – \mathcal{C} , Michaelson contrast, is most useful in identifying high contrast elements, generally considered to be an important choice feature for human vision. Michaelson contrast is calculated as $\|(L_m - L_M)/(L_m + L_M)\|$, where L_m is the mean luminance within a 7×7 surrounding of the center pixel and L_M is the overall mean luminance of the image.

9 – \mathcal{H} , discrete cosine transform, DCT, introduced by [1], is used in several coding standards as, for example, in the JPEG-DCT compression algorithm. The image is first subdivided into square blocks (i.e. 8×8); each block is then transformed into a new set of coefficients using the DCT; finally, only the the high frequency coefficients, the ones that are instead discarded in the JPEG algorithm, are retained to quantify the corresponding block.

4 Clustering and sequencing

The algorithmic procedures above result in the defining of local maxima. Usually these algorithms operate at pixel resolution levels, so that it is important to cluster their selected image element to provide regional ROIs — each region consisting of a neighborhood collection of multiple local maxima (Figure 1). The initial set of local maxima is clustered connecting local maxima by gradually increasing the acceptance radius for their joining. Approximately 100 initial local maxima are thus reduced to about nine regions or clusters by setting the termination decision to end the clustering process at this number of domains.

The ROI domains can be assigned values depending upon the value of the highest local maxima incorporated into that domain, or alternatively the number of local maxima included,

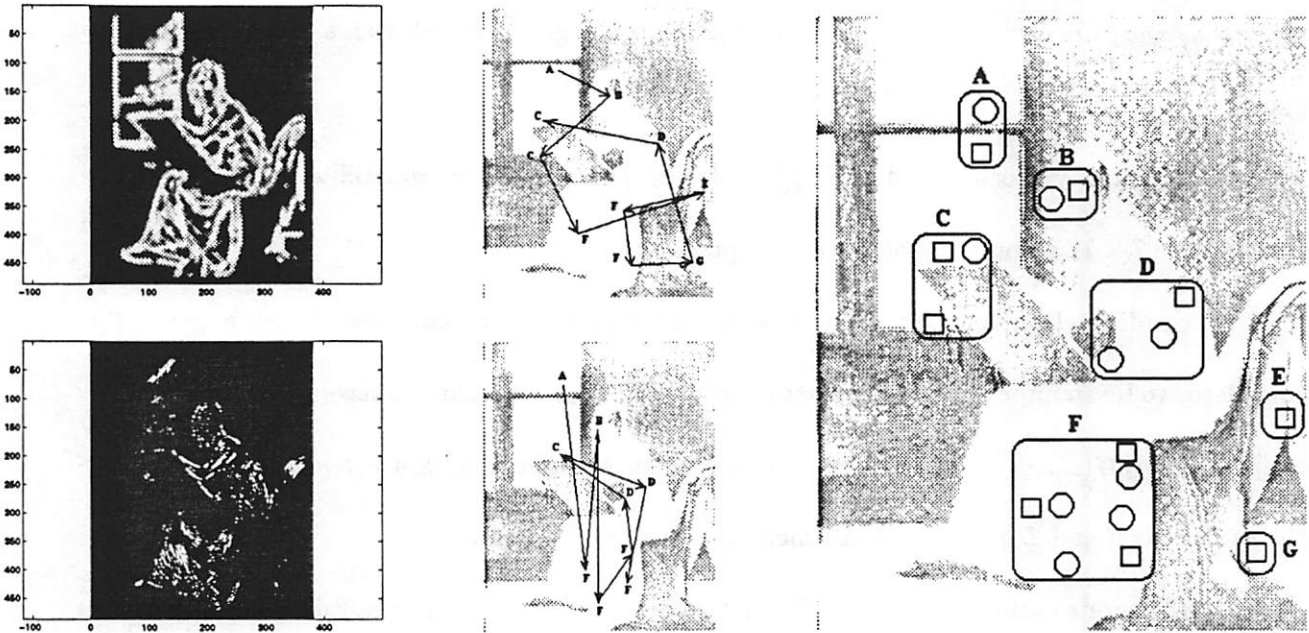


Figure 3: ROIs Sequence Comparisons. Picture transformations are a result of and provide descriptions of actions of each algorithm (left column); two examples, \mathcal{E} (upper), and \mathcal{S} (lower). After transformation, ten final ROIs are ordered by value and the constructed ordered sequences (central column) were connected by vectors (see arrows) in analogy to EM sequences of fixations. Two set of ROIs are finally combined (right panel) into a selected number of *joined-ROIs*, used to define distance measures between the two sets.

\mathcal{S}							\mathcal{F}						
a	b	c	c	e	f	g	a	b	c	d	e	f	g
0	1	0	0	0	0	0	a	0	0	0	0	1	0
0	0	1	0	0	0	0	b	0	0	0	0	1	0
0	0	0	0	0	0	0	c	0	0	0	0	0	0
0	0	1	0	0	0	0	d	0	0	0.5	0	0.5	0
0	0	0	0	0	1	0	e	0	0	0	0	0	0
0	0	0	0	0.33	0.33	0.33	f	0	0.33	0	0.33	0	0.33
0	0	0	0	0	0	0	g	0	0	0	0	0	0

Figure 4: Transition Matrices. Markov matrices derived during the comparisons of the two algorithms shown in Figure 3.

or using other criteria. Each of our IP algorithms, of course, contributes the intensity of its selected parameter to finding the local maxima and the values of resulting clustered ROI domains. The clustering algorithm is actually an eccentricity weighting algorithm, where even lower local maxima that are eccentrically located can be selected to form a domain. Thus, the resulting string of IP ROIs (in Figure 1, left-bottom panel, 15 final ROI domains are shown) were similar in number to human EM fixation glances looking at similar pictures (Figure 1, right-bottom panel).

5 Comparisons and sorting procedure

The ROI loci selected by our different IP algorithms and those defined by human EM fixations can be compared. Comparison of final clusters of ROIs begins with taking two sets of ROIs (Figure 3, middle, upper and lower panels) and clustering these two sets using a distance measure derived from a k-means pre-evaluation. This evaluation determined a region for calling coincident any ROIs that were closer than this distance and non-coincident for ROIs that were further apart than this distance; the distance was about two degrees and similar in size to human foveal spans. The final selection of *joined-ROIs* (Figure 3, right panel) then enabled a similarity metric, S_p , to determine how close were the ROIs identified by two algorithms (as in the example shown in Figure 3), or by two humans, or by an algorithm and a human. The individual sources of the elements, that is the original ROIs, used in these final interactive steps were preserved as circles and squares (Figure 3, right panel) to illustrate the procedure.

As mentioned above, ROIs are ordered by the value assigned by the image processing algorithm or by the temporal ordering of human eye fixations in a scanpath. Then, the joined-ROIs can finally be ordered into strings of ordered points. Here, (Figure 3), we have for example: $strings = abcfeffgdc$ and $string_{\mathcal{F}} = afbffdcdf$. The string editing similarity index S_s was defined by an optimization algorithm [12] with unit cost assigned to the three different operations *deletion*, *insertion* and *substitution*.

Sp				Ss				St										
same subj same alg		diff subjs diff algs		a * h		same subj same alg		diff subjs diff algs		a * h		same subj same alg		diff subjs diff algs		a * h		
same pct same typ	Rep		L		L		Rep		L		L		Rep		L		L	
	h = 0.66	h = 0.56	or em bi dw	0.47 0.58 0.46 0.39	h = 0.59	h = 0.44	or em bi dw	0.20 0.26 0.20 0.17	h = 0.29	h = 0.15	or em bi dw	0.04 0.05 0.04 0.02	h = 0.29	h = 0.15	or em bi dw	0.04 0.05 0.04 0.02	a = 0.47	a = 0.16
same pct diff typs	Rep(t)		L(t)		L(t)		Rep(t)		L(t)		L(t)		Rep(t)		L(t)		L(t)	
	hs = 0.59	hs = 0.51	as = 0.55	as = 0.42	hs = 0.54	hs = 0.42	as = 0.17	as = 0.22	hs = 0.26	hs = 0.14	as = 0.05	as = 0.03	hs = 0.26	hs = 0.14	as = 0.05	as = 0.03		
diff pct diff typs	I		G		G		I		G		G		I		G		G	
	hd = 0.34	hd = 0.23	ad = 0.37	ad = 0.39	hd = 0.35	hd = 0.22	ad = 0.13	ad = 0.13	hd = 0.14	hd = 0.06	ad = 0.03	ad = 0.03	hd = 0.14	hd = 0.06	ad = 0.03	ad = 0.03		
Ra				Ra				Ra				Ra						
0.15				0.04				0.01										

Figure 5: Triple Parsing Diagram. Three panels show comparisons of similarity indices, S_p , S_s , S_t , for scanpaths produced by human EMs (h), by IP algorithms (a), and especially for comparisons between IP algorithms and human scanpaths ($a * h$). Each of the panels is sub-divided into boxes representing different combinations of comparisons between pictures, types of picture and algorithms. Random correlations, Ra , are shown in lower middle box of each panel.

Sp	A	C	H	T	X	S	W	F
A	1	0.23	0.54	0.64	0.60	0.67	0.72	0.64
C		1	0.69	0.86	0.78	0.78	0.73	0.40
H			1	0.42	0.52	0.60	0.40	0.51
T				1	0.42	0.42	0.47	0.28
X					1	0.83	0.87	0.66
S						1	0.78	0.85
W							1	0.51
F								1

Figure 6: Correlation Coefficients in Y-matrix Format. Four human subjects, (A, C, H, T) and four algorithms (\mathcal{X} , \mathcal{S} , \mathcal{W} , \mathcal{F}) have intra-grup correlation coefficients indicated in triangular semi-matrices; for humans (upper left) and for IP algorithms (lower right). Inter-group comparisons between subjects and algorithms in rectangle (upper right). From the data of this kind we have computed average correlation coefficients or similarity indices for the parsing diagrams (Figure 5). The $a * h$ values (rectangle, average for each column) are 0.58, 0.62, 0.60, 0.45, for the four algorithms: an overall average of 0.56.

We can also create, based on the above mentioned joined-ROIs, markov transition matrices, [M1]s (Figure 4, \mathcal{S} left matrix and \mathcal{F} right matrix). The transition similarity index, S_t , is based upon cross-correlation between the coefficients of the [M1]s.

Thus all of our comparisons yielded three different indices of similarity which tells us how closely two set of ROIs resemble each other in locus, S_p , in sequence, S_s , and in transition from one ROI to the next, S_t . For the example illustrated above (Figure 3) we have: $S_p = 1$, $S_s = 0.34$ and $S_t = 0.09$.

5.1 Parsing diagram

The parsing diagram, (Figure 5) represents averages of sorting coefficients for the three measures, S_p , S_s and S_t we have used: (a) for repetitive scanpaths (*Rep*), same subject looking at the same picture and same type at different times; (b) different subjects or different algorithms, same picture and same type (Local = L); (c) same subject or same algorithm, different types and different pictures (Idiosyncratic = I); (d) different subjects and different algorithms, different types and different pictures (Global = G). For different types, but the same picture, we have: (e), same subject or same algorithm ($Rep(t)$) and (f), different subjects or different algorithms, ($L(t)$). In Figures 5 human EM values are labelled h ; IP algorithms values are labelled a and the comparison of IP algorithms and human is labelled $a * h$.

The most important distinction is that between Repetitive similarity, *Rep*, upper left box, and Global similarity, G , lower box: the *Rep* value for human with the S_p measure, (Figure 5 left diagram), equals 0.66. This means that the string for repetitive viewing of the same stimulus for the same subject have loci that were 66% within fixational or foveal range — continuing support for the scanpath theory.

For Global, all different subjects looking all different stimuli had an S_p value of only 0.23. This number was somewhat different from the expected S_p value of 0.15 based up random model, Ra , considerations (bottom box of Figure 5). Comparisons among different algorithms are

generally similar, but perhaps lower than the S_p similarity indices among humans (see also [13] for a previous preliminary result).

We also selected four of the algorithms that seem to cohere with the human subjects after an overall viewing of the data (Figure 5), but further segregated to show each algorithm and each subject separately (Figure 6); this coherence strengthens our overall result by documenting stronger similarity indices among these four selected algorithms; note the high correlations (average equals 0.75) among algorithms (lower triangle). We might further improve EM loci prediction by choosing other sets of structurally different algorithms and combine them and the ones used here in some optimal fashion.

The relationship between human fixation strings and IP algorithms ROI strings for different subjects and different algorithms viewing or processing the same stimulus, that is the same picture and the same type, tests the ability of BU algorithms to predict loci of human fixations (third column, box L , Figure 5, S_p parsing diagram). The diagram shows the value for all the four different types; the average value of 0.47 is very close to the similarity value of human, 0.56, for local, L . Again we found support for this result from an earlier study, [13] and even more support from our study of four selected algorithms (Figure 6, rectangle) where the average values was 0.55.

These quantitative data provide major support for our main result that IP algorithms, even broadly selected, could predict human EM string fixation loci about as well as two different human EM strings could predict each others loci.

An important finding in earlier studies, [12] was that human scanpaths were closely matched for string similarities and for markov transition probabilities. Our studies confirm these earlier results; high repetitive similarities (Rep) were found for the same subject looking at the same stimulus. The repetitive value for S_s and S_t were 0.59 and 0.29, approximately three to four times as great as the Global value, which in turn was approximately four time the random value (Figure 5, S_s and S_t).

5.2 Results with individual algorithms

When we studied the effects of the individual algorithms in generating clusters plus ROIs, they varied greatly as expected, confirming our view that we had assembled a wide set of algorithms.

The x -like mask was a poor predictor for two of the types and good with one; it appears to be opposite in its effect on embossed and cartoon. The " $<$ "-like mask was a good predictor for all four types: it is not symmetrical so that we plan in the future to try different orientations and perhaps different support-resolutions as in a wavelet fashion.

The symmetry transform, S , was a good predictor for the original picture. The wavelet transforms, W , seemed to be very good predictors for all types and without any poor predictions. This is probably due to the rescaling process of their basis function. H was poorly related to the modified types, but worked well for the original picture. F was a very good predictor for three of the types but was poor for the embossed figure, perhaps because it was already spatially differentiated and the second differentiation, due to the Laplacian, introduced some noise.

6 Discussion

In this paper, we have validated that a constellation of BU algorithms, suitably but widely selected can predict the loci of human fixations, for the same picture and type modification; this prediction is termed the $a * h$, local, L , relationship in our S_p parsing diagrams. Further, this prediction is as good as the ability of one human to predict the fixation locations of another subject; the h , local, L , relationship in our S_p parsing diagrams. Our results indicate, however, that the algorithms can not predict the sequential ordering of the subfeatures used by a person, S_s and S_t .

Our method provides a precise task for the image processing algorithms we have studied — to predict human scanpaths, both loci and sequences of eye movement fixations or foveations. The method also provides for quantitative measurements of the accuracy of those predictions.

The wide selection of algorithms gave us an opportunity to study the differences and similarities in terms of the precise task we studied. These algorithm characteristics are of great interest to us as indicating the general nature of an image and how it is processed either by algorithms or by humans. However, we are concerned that we might need to provide weighting coefficients for the different algorithms in order to optimize the predicting capabilities of the ensemble.

Our scale of similarity indices is anchored at the bottom both by the random, *Ra*, values and by the global, *G*, values, that is for all subjects and algorithms and pictures and modifications. The top of the scale is anchored for human studies by the repetitive, *Rep*, value, the closeness of fit of a single subject's scanpaths to her scanpaths with the same picture at another time with the same task instructions. Can we similarly use trivial modifications of the pictures to obtain Repetitive indices for the algorithm studies?

The clustering procedures we used followed a good deal of thought and some preliminary studies. However, as we indicated above, the clustering procedure really acts to choose the string of ROIs to be more eccentrically located than if we did not use the clustering algorithm. Thus it might be termed a non-clustering clustering algorithm ! Our preliminary studies with fuzzy clustering (unpublished work in our lab by Dr. Ulli Oechsner of Hamburg University) has led to a number of conjectures that might be tested in further experiments.

As I. I. Rabi said, each scientific problem gives rise to several new problems, like the heads of the Medusa when cut off. So we have found. How to collect and classify and relate even a wider variety of image processing algorithms? What modifications of a picture leaves the picture information relatively constant and what modifications strongly deforms the picture information? Finally, we intend to apply our methodology to a much richer collection of pictures, scenes and works of art. These might range widely from natural and constructed landscapes and city-scapes to groups of persons and animals and objects to single portraits and still lives.

Acknowledgements

We thank our sponsors for partial support:– NASA-Ames Research Center (Dr. Stephen Ellis), Fujita Research (Dr. Ken Kawamura) and Neuroptics Corporations (Dr. Kamran Siminou). Our colleagues in the laboratory, Michela Azzariti, Ted Blackmon, Yeuk Fai Ho, Veit Hagenmeyer, Yong Yu, and others have been generous in their advice — Irwin Sobel at Hewlett-Packard, and at UCB, Jerry Feldman, CS, and David Brillinger, Statistics.

References

- [1] N. Ahmed, T. Natarajan, and K. Rao. Discrete cosine transform. *IEEE Trans. Computer*, 23:90–93, 1974.
- [2] J. Canny. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [3] Y.S. Choi, A.D. Mosley, and L.W. Stark. String editing analysis of human visual search. *Optometry and Vision Science*, 72(7):439–51, 1995.
- [4] I Daubechies. Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.*, 41:909–96, 1988.
- [5] V. Di Gesù and C Valenti. The discrete symmetry transform in computer vision. Technical Report 011-95, Laboratory for Computer Science (D.M.A.), University of Palermo, 1995.
- [6] R.M. Haralick. Statistical and structural approaches to texture. *Proc. IEEE*, 67:786–804, 1979.

- [7] E. Niebur and C. Koch. Control of selective visual attention: Modeling the “where” pathway. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 802–808. The MIT Press, 1996.
- [8] D. Noton and L. Stark. Eye movements and visual perception. *Scientific American*, 224:34–43, 1971.
- [9] R. T. Reed and J. M. Hans Du Buf. A review of recent texture segmentation and feature extraction techniques. *CVGIP: Image Processing*, 57(3):359–372, 1993.
- [10] D. Reifeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision*, 14:119–130, 1995.
- [11] O. Rioul and P. Duhamel. Fast algorithms for discrete and continuous wavelet transforms. *IEEE Trans. on Information Theory*, 38(2):569–86, 1992.
- [12] L. Stark and Y. Choi. Experimental metaphysics: The scanpath as an epistemological mechanism. In W. H. Zangemeister, H. S. Stiehl, and C. Freksa, editors, *Visual Attention and Cognition*, pages 3–69. Elsevier, Amsterdam, 1996.
- [13] L. Stark and C. Privitera. Top-down and bootm-up image processing. In *Proc. of IEEE International Conference on Neural Networks*, volume 4, pages 2294–2299. Houston, TX, June 9-12 1997.
- [14] P.P. Vaidyanathan. *Multirate System and Filter Banks*. Prentice-Hall, Englewood Cliffs, NJ, 1993.