AUTONOMOUS IMAGE PROCESSING ALGORITHMS LOCATE REGION-OF-INTERESTS: THE MARS ROVER APPLICATION

by

Claudio Privitera, Michela Azzariti, Lawrence W. Stark

Memorandum No. UCB/ERL M98/8

10 March 1998

AUTONOMOUS IMAGE PROCESSING ALGORITHMS LOCATE REGION-OF-INTERESTS: THE MARS ROVER APPLICATION

by

Claudio Privitera, Michela Azzariti, Lawrence W. Stark

Memorandum No. UCB/ERL M98/8

10 March 1998

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

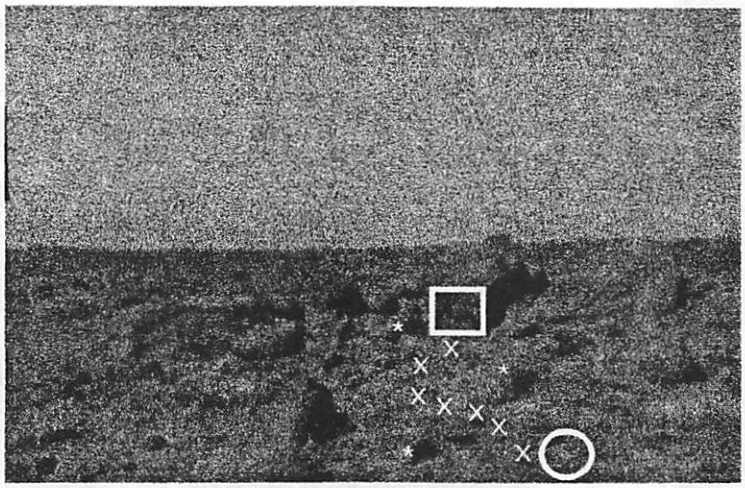


Figure 1: Daily Plan for geological exploration and detailed planned navigational route for autonomous control by rover..

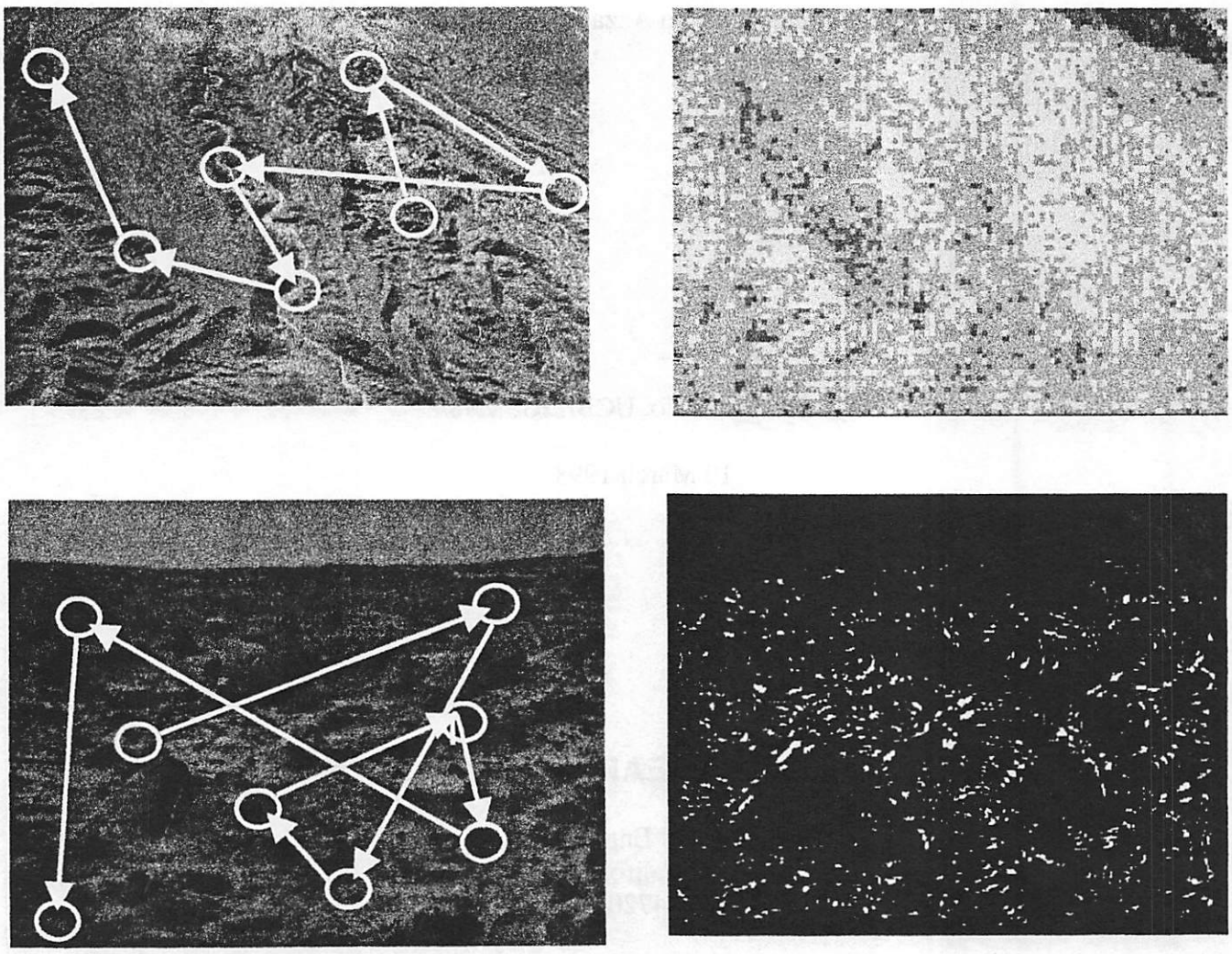


Figure 2: Terrain Pictures Processed by IPAs to Obtain Ordered ROIs (entropy above and symmetry below).

Autonomous Image Processing Algorithms locate Region-of-Interests: the Mars Rover application.

Claudio Privitera, Michela Azzariti, Lawrence W. Stark

Neurology and Telerobotics Units
486 Minor Hall, University of California, Berkeley 94720-2020

Abstract:

In this report, we demonstrate that bottom-up IPA's, image-processing algorithms, can perform a new visual task ---- to select and locate ROIs, regions-of-interests. This task has been defined on the basis of a theory of top-down human vision, the scanpath theory. Further, using measures, --- Sp and Ss, the similarity of location and ordering, respectively, ---- developed over the years in studying human perception and the active looking role of eye movements, we could quantify the efficient and efficacious manner that IPAs can imitate human vision in located ROIs.

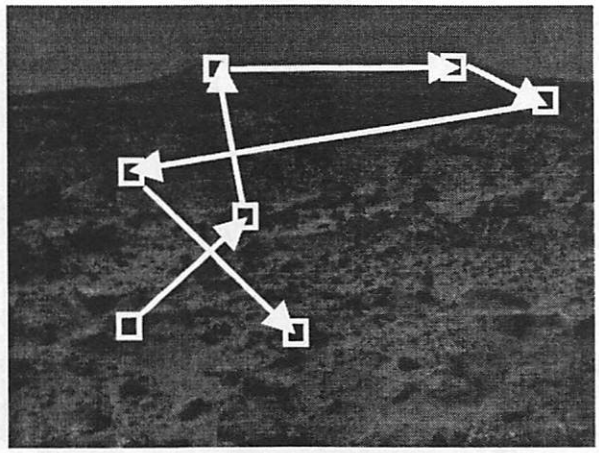
The means to quantitatively evaluate IPA performance has been an important part of our study. In fact, these measures were essential in choosing from the initial wide variety of IPAs, that particular one that best serves for a type of picture and for a required task. It should be emphasized that the selection of efficient IPAs has depended upon their correlation with actual human chosen ROIs for the same type of picture and for the same required task accomplishment.

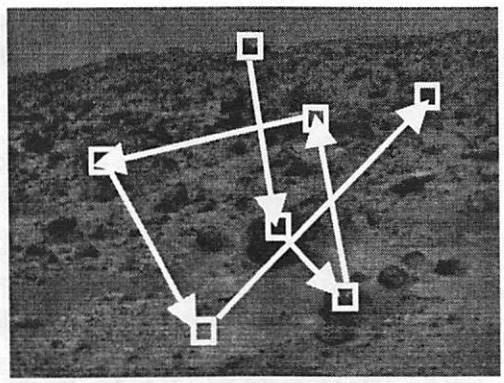
1.- INTRODUCTION

The tasks. There are three problems that we are trying to address: --- i) to assist the once-a-day supervisory control procedures for the Mars Rover; ii) to enable robust autonomous control of the Mars Rover during absence of earth control; iii) to assist in reducing the down-loaded Mars picture data interpretation load when this becomes excessive, and thus aid in selecting amongst a large data set, a smaller set wherein human navigational or geological study planning might be most productive.

An example of a daily plan for the navigational route was superimposed onto a picture of Mars terrain obtained July 1997 (Figure 1); note, circle representing an IC, initial condition, position of the Rover and square representing an intended goal of possible geological interest. This was the situation facing the engineering supervisory controller for Rover navigation each day at the Earth-station; a navigational plan next takes shape considering factors such gradient, possible hidden obstacles, energy supplies, and above all, integrity of the Rover. This detailed route has to be planned (see *x*'s along path); but the real time trajectory must be accomplished by on-board control, allowing for local modifications so as to minimize gradients and to provide for avoidance of obstacles. Also navigational benchmarks may be placed (large asterisks) as future route markers to supplement existing landmarks and to indicate features of possible geological interests.

Autonomous algorithms. Our overall goal has been to describe and document the methodology we have employed in approximating top-down human vision with bottom-up autonomous IPAs, image processing algorithms (Figure 2). The early experiments herein reported provide encouraging results and suggest a number of future research tasks.





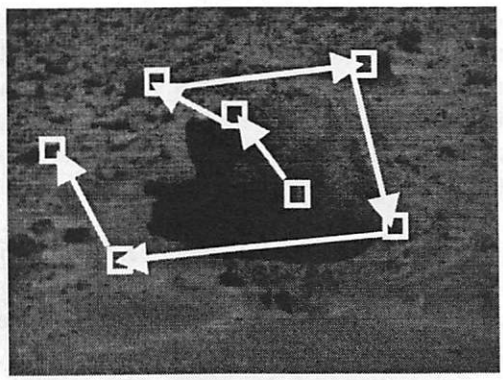


Figure 3: Terrain Pictures and ROIs chosen by Human Viewers.

Scanpath Theory. Human vision relies on active looking eye movements organized into scanpaths in order to check on the internal cognitive model of what a person was actually viewing. Our human subjects were asked to choose important features or objects in terrain pictures presented to them; particular geological and navigational tasks were specified since we found that different tasks called for different features to be selected (Figure 3).

Selection of algorithms. A specific aim of this report is to describe how human vision experiments can provide criteria for selection of appropriate IPAs. A wide variety of IPAs were applied to the same terrain pictures and those IPAs that cohered most closely to human ROIs, region of interests, were selected. The scheme for measuring similarities between humanly chosen, ROIs, noted as chROIs, and IPA-identified ROIs, noted as aROIs, is carefully explained below.

2.- METHODS

2.1 Scanpath Theory and Human identification of ROIs

The scanpath theory suggests that a top-down internal cognitive model of what we see controls perception and active looking eye movemnts, 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 region of interests, ROIs, of a scene or picture so as to check out and confirm the model (see Noton and Stark, 1971). The present study differs from the standard scanpath EM experiments: in fact, ROIs were identified by ordered sequences of cursor positioning over a picture. This is "semi-classical" scanpath of top-down human vision depends upon conscious human choice and the sequencing of the chROIs. Questions can be raised about the relationship between scanpaths provided by sequences of human eye movement fixations and the ordered sequences of chROIs chosen by cursor positioning.

The corpus of pictures we have used in this study consists of four Mars terrain pictures, 1m,2m,3m,4m, (Figures 2 and 3) taken by the July 4th, 1997 Mars mission. We also used two pictures of Earth terrain obtained from a preliminary NASA expedition to the Chilean desert, 5c,6c, (Figure 2, upper). The definition of "terrain" has been widened by planetary geologists to include surfaces of other planets. These were made available to us courtesy of Dr. Virginia Gulick, Planetary Geologist, NASA-Ames Research Center. Superimposed on these pictures were seven chROIs, selected by human observers, as containing subfeatures of special interest in the overall picture; arrows indicate the sequence in which the chROIs were chosen (Figure 3).

The viewers were asked to consider themselves as being faced with a number of different tasks, such as inspecting the pictures for general landforms or inspecting the pictures for possible navigable pathways for a small Rover vehicle. The subjects were recruited from our laboratory personnel and properly informed according to the rules of the Committee for the Protection of Human Subject of the University of California, Berkeley, were all geologically naïve. Each was instructed to act as if he was a planetary geologist and approached the pictures with a particular scientific or controller task in mind. They were to answer in their own minds questions about the following tasks.

The first four tasks were geological :--

1) general landform features?
2) characteristic of the rocks in the terrain structure. What is their structure? Copious in amount? Having sharp or rounded edges?
3) evidence of sand or water action?

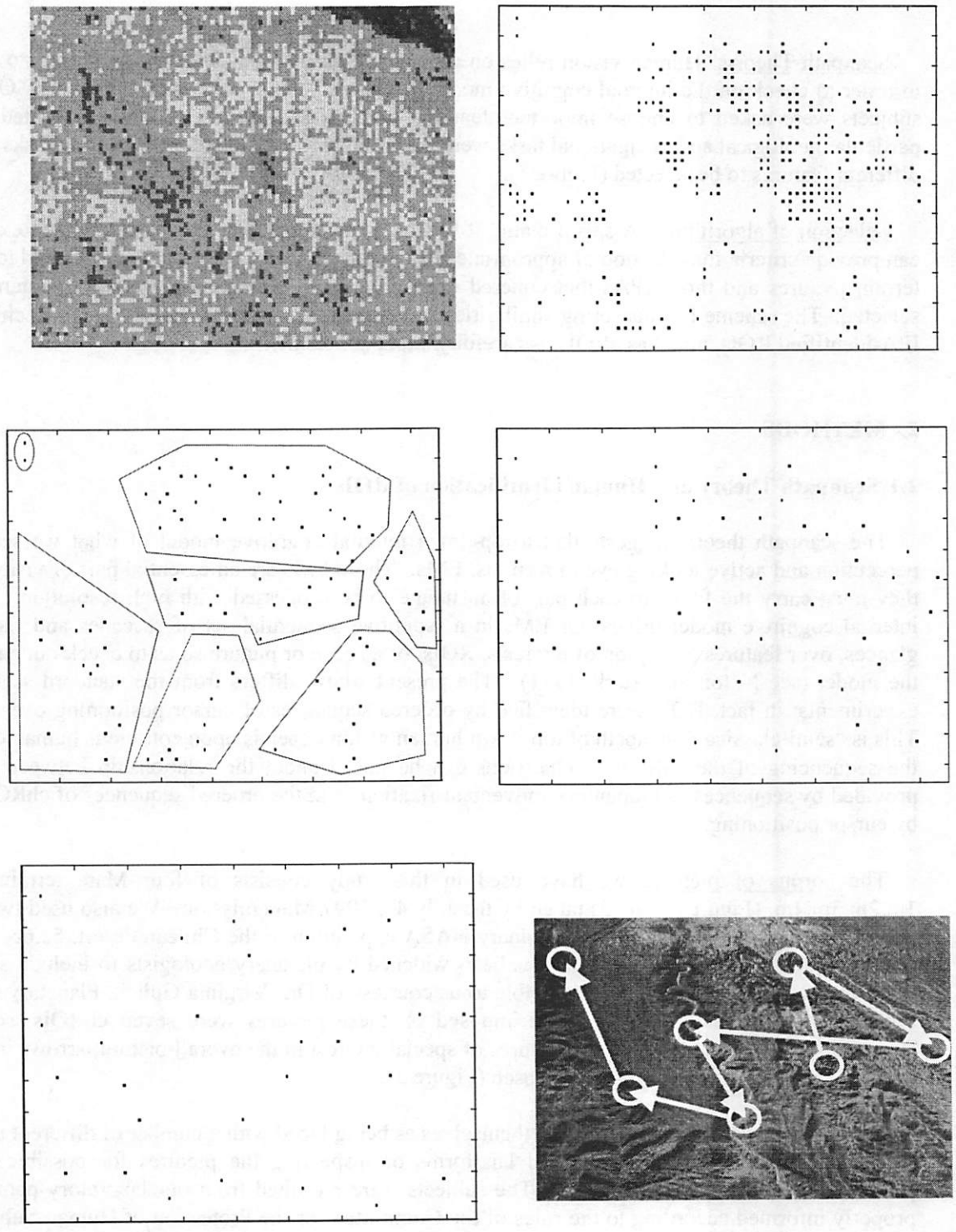


Figure 4: Clustering Method.

4) strata or layered rocks or cliffs present?

For the last two tasks they were to act as if they were a supervisory controller planning a path for the Mars Sojourner Rover moving on the surface of Mars:--

5) plan a path for the Rover vehicle on Mars terrain.

6) locate obstacles to such a path, especially obstacles that could completely block the Rover or place it in jeopardy.

2.2 Terrain Pictures Processed by IPAs to obtain ordered aROIs

According to the Scanpath Theory, an internal spatial-cognitive model of a scene or object controls active looking and human perception in a top-down procedure. Our question was, " How closely can we imitate this process using bottom-up IPAs?".

IPAs 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 1979, and Reed and Han Du Buf, 1993). 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 and entropy functions. Thirdly, model approaches that hypothesize underlying processes for generation of local regions, and 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. A list of the algorithms used in this follows:

A- Statistical kernels:

entrp -- entropy is calculated within a surrounding of the center pixel;

michaelson -- michaelson contrast generally considered to be an important choice feature for human vision;

ortn – difference in the gray-level orientation, a statistical-type kernel, is analyzed in early visual cortex;

lummax (lummin) – simply maxima (minima) in the luminance are detected.

B- Structural kernels:

filtermask -- an x-like mask, positive along the two diagonals and negative elsewhere, was convoluted with the image;

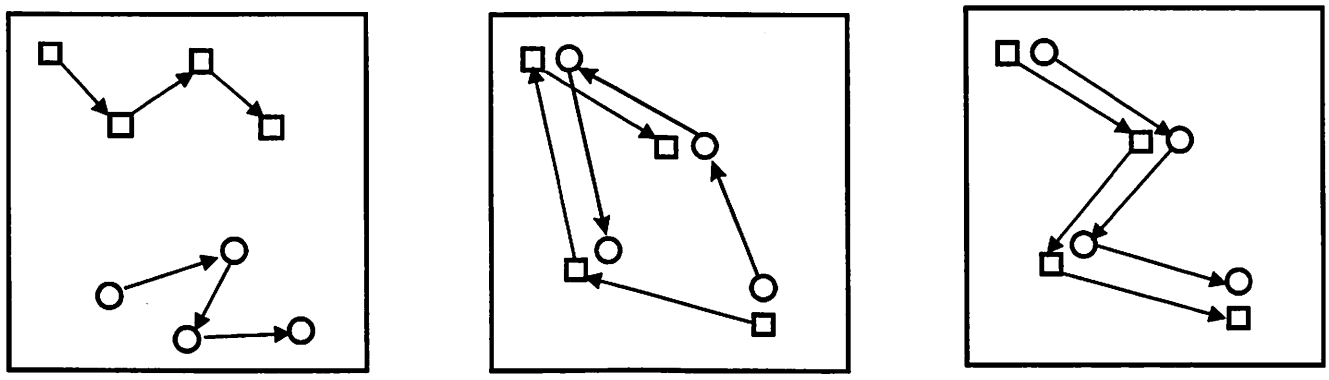
symm -- symmetry transform, appears to be a very prominent spatial relation;

edge -- concentration of edges per unit area;

lapl -- the laplacian of the gaussian is convoluted with the image.

C- Wavelet kernels:

discrete wavelet transform: different wavelets (each with different numbers of vanishing points) are taken into consideration (daubechies, biorthogonal, symlet): only three levels of coefficients were retained.



Sp = 0
Ss = 0

Sp = 1
Ss = 0

Sp = 1
Ss = 1

Figure 5: Illustration of Sp and Ss Similarities.

Two sets of ordered ROIs (left) whose loci are widely separated: low similarity. Two sets of ROIs (middle) closely located but whose ordered sequences are different: Sp high, Ss low. Two sets of ROIs (right) whose loci and ordered sequence are similar: Sp high, Ss high.

Sp	mich	ortn	lmax	lmin
entrp	0.26	0.33	0.25	0.25
mich		0.11	0.11	1.00
ortn			0.11	0.11
lmax				0.13

Ss	mich	ortn	lmax	lmin
entrp	0.00	0.11	0.07	0.00
mich		0.00	0.02	1.00
ortn			0.00	0.00
lmax				0.02

Figure 6: Y-matrix of Sp (left): locus similarities of pairs of IPAs. Y-matrix of Ss (right): string similarities of pairs of IPAs.

2.3 Clustering

First, the pictures were processed by each algorithm (Figure 2, left to right). Secondly, the IPAs choose approximately 1000 local maxima pixels that cohere with the particular algorithm kernel most closely. Third, a clustering algorithm was utilized (Figure 4) to reduce these thousand points down to an ordered set of seven to eight clusters superimposed upon the original terrain scene. The fit of the aROIs chosen by the algorithms (Figure 2, entropy, upper and symmetry, lower) onto important subfeatures of these terrain pictures was depicted .

The initial set of local maxima was clustered connecting local maxima by gradually increasing the acceptance radius for their joining (Figure 4). Partway through the clustering process the clusters have been encircled in their wide extent (middle left), although further reduction, using the clustering algorithm, occurs (middle right and lower left). Finally, each cluster inherits maximal value of its component points. This provides an ordering as indicated by the arrows connecting the cluster loci (right lower). The highest valued point of each cluster actually determines the final locus of the cluster.

2.4 Definition and use of Sp and Ss

The aROI loci selected by our different IP algorithms and the chROIs defined by humans can be compared. A similarity measure for comparing these two sets of loci is Sp = 1 – 'd', where 'd' is a distance measure, summed over a set of 'di's, where each 'di' is a distance between an algorithm ROI and an human identified ROI or between two algorithm ROIs or between two human identified ROIs. Each 'di' is first calculated based upon a threshold distance, 'di' = *gamma*. A 'di' is set equal to zero if below the threshold and 1 if above. The threshold *gamma* was ascertained using k-mean evaluation of EM fixation distances. The final value for 'di' is normalized based upon the value of the index i , which is equal to the string length. Finally, string editing similarities were defined by an optimization algorithm and yield Ss similarity indices.

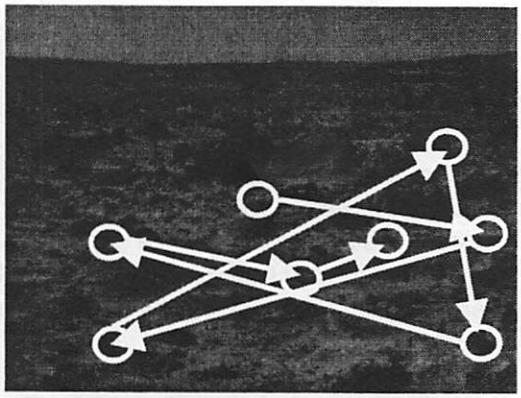
Various idealized Sp and Ss similarities for two sets of ordered ROIs (Figure 5) help to understand these measures: --- low similarity for Sp and Ss (left);); high value for Sp but not for Ss (middle); high similarity for Sp and Ss (right).

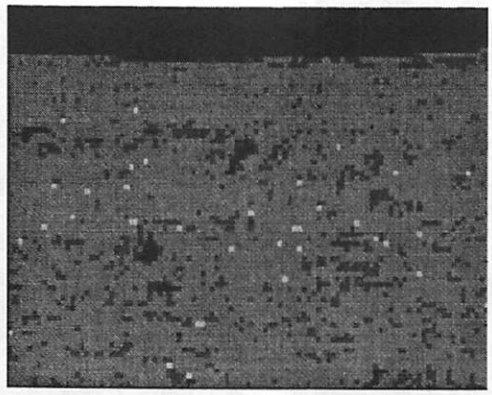
The Sp measure was used for similarities between the aROIs chosen by each pair of the IPAs, and these values, varying between 1 and 0 were arranged in a matrix, the Y-matrix (Figure 6). The upper right triangle of the matrix and the lower left triangle (omitted) were symmetrical; the diagonal elements (also omitted) would all have the value, 1. Each matrix coefficient thus measures the coherence between a pair of IPAs. In developing and selecting these IPAs we, of course, wished not to have IPAs that cohere as these would be redundant, as for example in the case where the value of 1.00 shows identical function for lummax, maximal luminance, and mich, michaelson (Figure 6, fourth column, second row, for both Sp and Ss matrices). Note also that in these Y-matrix, the values of Ss were much lower than the values of Sp.

Average values for Sp for human identification of chROIs were usually assembled in a two by two "parsing diagram". This display is perhaps best described in detail in connection with our Results.

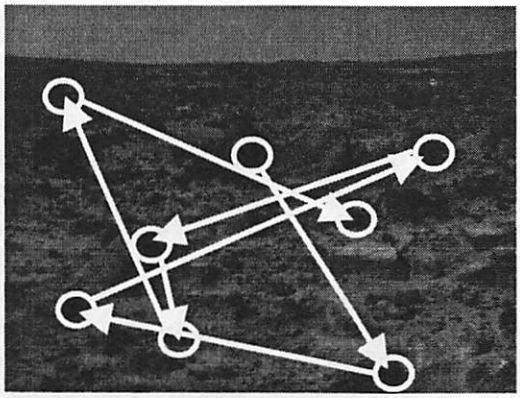
2.5 ANOVA

In order to better evaluate and interpret the final results, we also used an Anova, analysis-of-variance. The issue with Anova is whether or not the means of the observed data are different enough from the random mean to conclude that the means of the distribution corresponding to the observed data and the random data are different. The Anova value is compared with a critical value 'F' of a

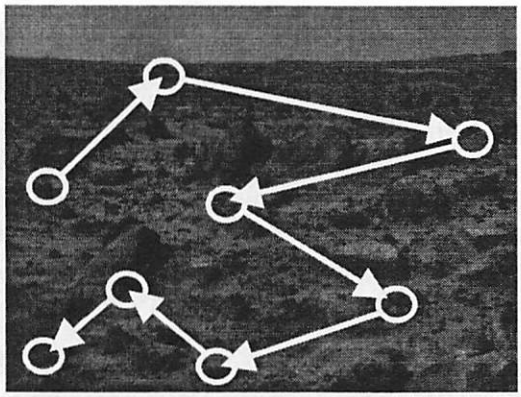


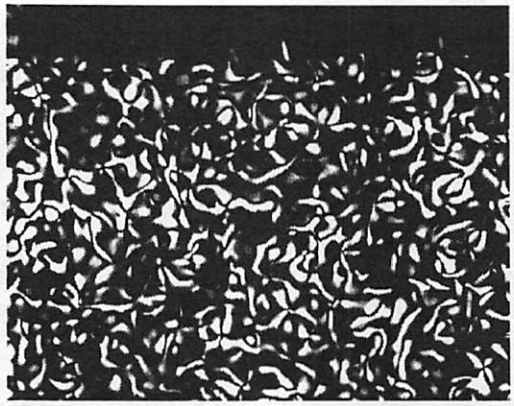


Entr



Mich





Ornt

Figure 7: Action of Algorithms on Mars terrain.

Fisher distribution with k-1 degree of freedom in the numerator (where k is the number of distributions that we are comparing) and n-k in the denominator (where n is the total number of observations in the k distributions). If the Anova test value is less than the F-Fisher critical value for a *alpha* level of significance (for example, in this paper, *alpha* was set equals to 0.01), then it is possible to infer that the two means are not different enough to come from different distributions; on the other hand, if the Anova test value is greater than the F-Fisher critical value this signifies that the means likely come from different distributions.

Our standard format for presenting our data in the parsing diagram (for example, we selected a data set from Figure 9, upper left box) was as follows: 0.65 (0.05, 355.51) with 0.65 equal to the mean value, 0.05 equal to +/- standard deviation, and 355.51 equal to the Anova test value. Our quantitative conclusions presented in the result section below were strongly sustained by the relationship between significant Anova test values and F-Fisher critical value (for *alpha* = 0.01) of 10.04.

3.- RESULTS

3.1 Action of Algorithms on Terrain Pictures

As described above, an IPA operateed on a picture and a transformed picture was obtained that illustrates the action of the algorithm (Figure 7, right and Figure 2). This visual presentation was a way of obtaining intuition as to the mathematical process of convoluting (or other procedure) the picture pixels with the IPA process. The 1000 points with the maximal values were then used in the clustering procedure, explained above, to obtain an ordered set of eight ROIs selected by that algorithm (shown superimposed on the terrain picture, Figure 7, left column).

3.2 Relationships among the IPAs

We were, of course, interested in several aspects of the IPAs. The first was, we wishded to obtain as wide a variety of IPAs as possible. Thus, we wanted the coherence between pairs of IPAs to be as small as possible. In this way, our wide variety of IPAs would have independent actions on the pictures. Thus, they could serve to identify ROIs for a variety of picture types, and for a variety of visual identification tasks.

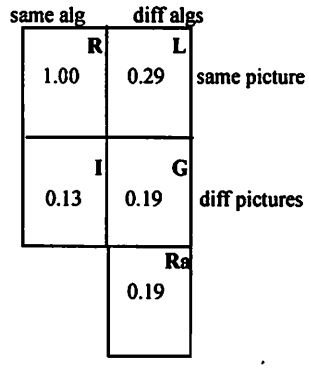
Y-matrices. The coefficients of the Y-matrix (Figure 8) indicated the coherence between each pair of algorithms as explained above. Whereas, the coefficients value of 1 (fourth column, second row) demonstrated complete coherence between those two IPAs, the value of 0.26 for the coefficient between the michaelson IPA and the entropy IPA (first column, first row) demonstrated moderate independence. A string-editing similarity coefficient of zero (first column, first row) represents complete independence of two compared sequences. Note again, that the coefficients for Ss (string-editing similarity) were much lower than the coefficients for Sp. Note the horizontal and vertical lines separate coherence values for different types of IPAs --- statistical, structural and wavelets.

Parsing diagram. The means of these coefficients could be put into a "parsing diagram" to help us to understand the average coherences relating the same and different IPAs acting on the same and different pictures (Figure 8 right).

For Sp, the repetitive, R, value of 1.00 (upper left box) defined the identical similarity for the same IPA viewing the same picture. When the same IPA views different pictures, the similarity equals 0.13 (lower left box, I, idiosyncratic). Since there was no special reason why an algorithm should select the same set of aROIs for different pictures, this low value was expected; for comparison as a bottom-anchor for the scale, the random, Ra, value equaled 0.19 (right, lowest box). Different algorithms

mich	ortn	lmax	lmin	symm	filtm	edge	lapl	db2	db5	db10	db20	biorl	bior3	bior6	symlet	Sp
.26	.33	.25	.25	.31	.32	.38	.36	.26	.27	.29	.22	.27	.38	.18	.26	entrp
	.11	.11	1.	.27	.22	.29	.22	.32	.18	.24	.26	.22	.25	.27	.32	mich
		.11	.11	.15	.28	.40	.30	.29	.26	.15	.26	.29	.30	.37	.29	ortn
			.13	.16	.27	.23	.14	.27	.20	.18	.24	.22	.27	.18	.27	lmax
				.27	.22	.29	.22	.32	.18	.24	.26	.22	.25	.27	.32	lmin
					.34	.30	.17	.23	.34	.24	.17	.21	.32	.23	.23	symm
						.38	.48	.32	.34	.28	.17	.47	.28	.39	.32	filtm
							.33	.42	.44	.33	.33	.46	.26	.35	.42	edge
								.29	.39	.28	.20	.36	.27	.38	.29	lapl
									.40	.23	.19	.37	.27	.33	1.	db2
										.29	.29	.36	.33	.31	.43	db5
											.26	.30	.10	.31	.24	db10
												.25	.22	.22	.19	db20
													.29	.46	.40	biorl
														.27	.27	bior3
															.35	bior6

mich	ortn	lmax	lmin	symm	filtm	edge	lapl	db2	db5	db10	db20	biorl	bior3	bior6	symlet	Ss
.00	.11	.07	.00	.07	.09	.09	.10	.13	.10	.05	.00	.07	.05	.05	.13	entrp
	.00	.02	1.	.05	.02	.05	.07	.05	.00	.00	.05	.05	.02	.02	.05	mich
		.00	.00	.02	.04	.07	.11	.07	.04	.05	.09	.09	.07	.06	.07	ortn
			.02	.02	.05	.05	.05	.05	.00	.00	.02	.07	.00	.02	.05	lmax
				.05	.02	.05	.07	.05	.00	.00	.05	.05	.02	.02	.05	lmin
					.07	.04	.08	.07	.12	.00	.02	.05	.05	.11	.07	symm
						.08	.05	.06	.18	.07	.04	.07	.05	.09	.06	filtm
							.07	.02	.14	.02	.04	.10	.07	.05	.02	edge
								.06	.15	.05	.05	.13	.10	.11	.06	lapl
									.07	.05	.07	.09	.02	.09	1.	db2
										.05	.09	.02	.09	.12	.07	db5
											.07	.04	.02	.00	.05	db10
												.04	.00	.00	.07	db20
													.14	.16	.09	biorl
														.11	.02	bior3
															.09	bior6



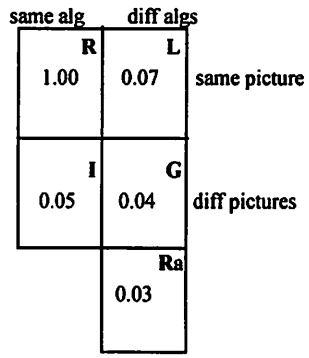


Figure 8: Y-matrices (left table) showing interrelationships among algorithms (Sp, upper and Ss, lower). Parsing diagram for IPAs coherence (right).

The coefficients indicate the coherence between each pair of algorithms. For example, the coefficient of 1.00 (4th column, second row) indicates that the Michaelson IPA and the Lumin IPA are identical. On the other hand, the value of 0.26 for the coefficient in the first column and first row shows moderate coherence between the Michaelson IPA and the entropy IPA. Note that the coefficients for Ss (string editing distance) are much lower than the coefficients for Sp (see text).

SAME TASK

Sp

Same person	Different persons
R 0.65 (0.05, 355.51)	L 0.48 (0.08, 41.53)
I 0.32 (0.17, 7.48)	G 0.32 (0.2, 7.26)
	Ra 0.23 (0.04)

Ss

	Same person	Different persons
same picture	R 0.23 (0.10, 54.83)	L 0.13 (0.02, 83.22)
different pictures	I 0.07 (0.09, 2.89)	G 0.09 (0.09, 16.78)
		Ra 0.04 (0.02)

F-Fisher critical value F(0.01) = 10.04

DIFFERENT TASKS

Sp

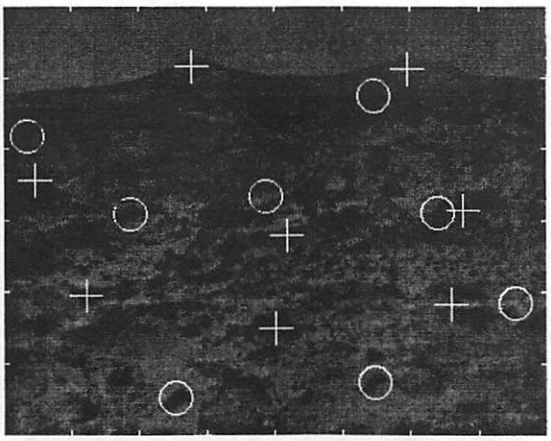
Same person	Different persons
R 0.38 (0.06, 66.55)	L 0.36 (0.05, 30.96)
I 0.27 (0.12, 3.49)	G 0.27 (0.09, 3.78)
	Ra 0.23 (0.04)

Ss

	Same person	Different persons
same picture	R 0.10 (0.04, 24.52)	L 0.09 (0.04, 22.16)
different pictures	I 0.04 (0.06, 0.03)	G 0.04 (0.09, 0.47)
		Ra 0.04 (0.02)

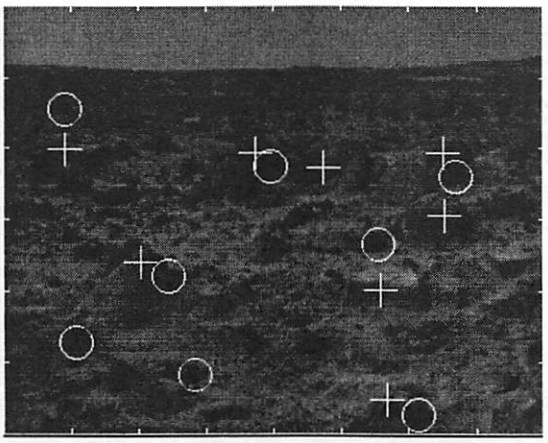
F-Fisher critical value F(0.01) = 10.04

Figure 9: Parsing diagrams for human choice coherence and for the same (upper) and different (lower) tasks; mean (standard deviation and F).



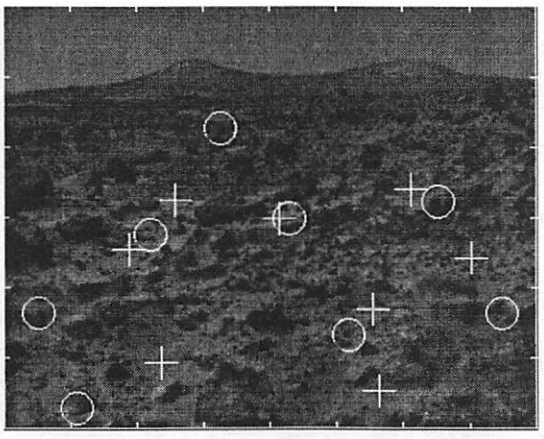
Task1, symm

Sp=0.5



Task2, mich

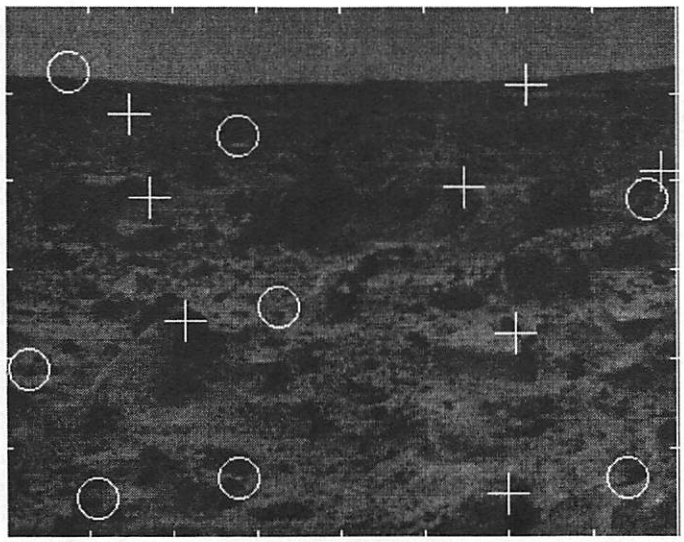
Sp = 0.88



Task3, symlet

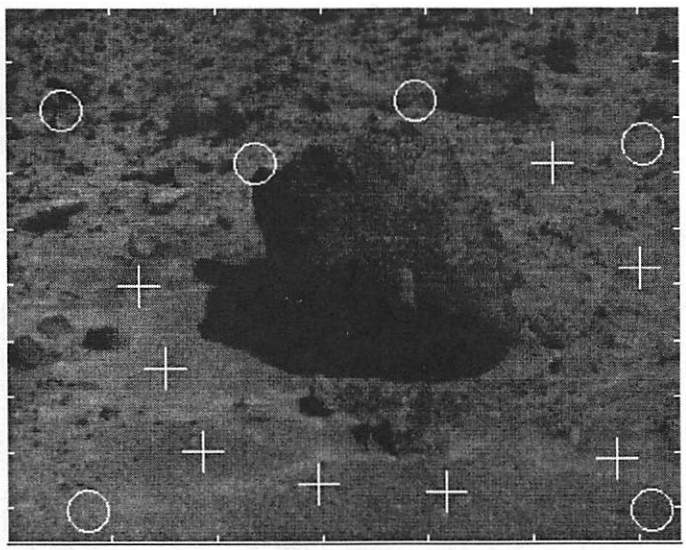
Sp = 0.63

Figure 10: Examples of Excellent Prediction of Human ROIs Locations ('+') by IPAs (circles).



Task1, bior1

Sp = 0.13



Task3, db10

Sp = 0

Figure 11: Examples of Poor Prediction of Human Locations by IPAs.

This poor correlation indicates the necessity of selection of algorithms on the basis of picture type and task requirements, as documented in this report.

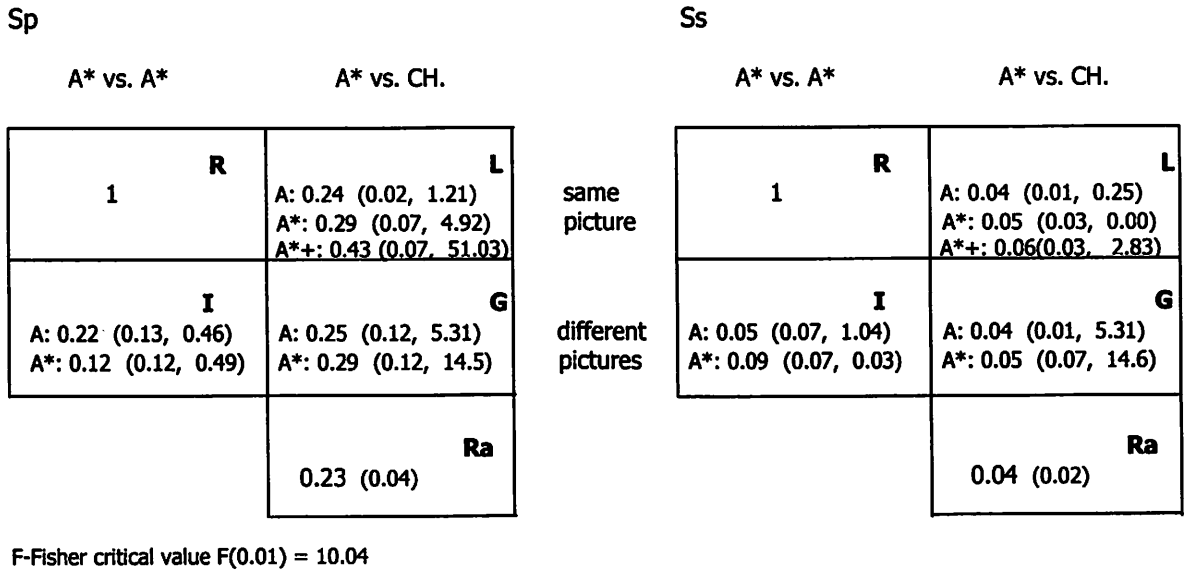
Task/fig Sp

	Fig.1	Fig.2	Fig.3	Fig.4	Fig.5	Fig.6	Mean
Task 1	0.19	0.26	0.25	0.37	0.05	0.60	0.29
Task 2	0.36	0.26	0.30	0.47	0.06	0.54	0.34
Task 3	0.04	0.40	0.12	0.33	0.01	0.43	0.22
Task 4	0.30	0.19	0.19	0.31	0.06	0.61	0.28
Task 5	0.18	0.41	0.33	0.34	0.09	0.40	0.30
Task 6	0.23	0.35	0.40	0.33	0.04	0.47	0.31
Mean	0.21	0.31	0.26	0.36	0.05	0.50	(0.29, 0.15, 23.53) tasks (0.29, 0.04, 3.86) figures

Task/fig Ss

	Fig.1	Fig.2	Fig.3	Fig.4	Fig.5	Fig.6	Mean
Task 1	0.07	0.06	0.08	0.01	0.	0.05	0.05
Task 2	0.08	0.13	0.04	0.11	0.04	0.05	0.08
Task 3	0.	0.08	0.02	0.05	0.	0.02	0.03
Task 4	0.	0.01	0.01	0.08	0.	0.11	0.04
Task 5	0.	0.11	0.01	0.05	0.04	0.01	0.04
Task 6	0.02	0.05	0.05	0.08	0.01	0.04	0.05
Mean	0.03	0.07	0.04	0.06	0.01	0.05	(0.04, 0.02, 0.01) tasks (0.04, 0.01, 0.0) figures

Figure 12: Confrontation tables of Sp and Ss similarities between IPA*s and choices.



F-Fisher critical value F(0.01) = 10.04

Figure 13: Parsing diagram for Sp (left) and Ss (right) similarities between IPAs and human choices. Note the values for IPA* (a combination of three different algorithms, see text) and IPA*+ (IPA* applied to Figures 4 and 6 only).

looking at different pictures cohered even less with a mean value of 0.19 for another bottom-anchor, the G, global value (middle right box). However when different IPAs act on the same picture, their aROIs were related with a local, L, mean similarity value of 0.29 (upper right box). Local was a critical measure for our study.

When considering string-editing similarities, Ss, we find (Figure 8, lower right), except for the trivial R value of 1.00 for the same IPA acting on the same picture, all values approximate the bottom anchors; both Ra and G values. Although, there may have been coherences between different (redundant) IPAs acting on a particular picture, this result suggested that the string sequences were not coherent with the bottom-up algorithmic procedures.

3.3 Human choice and multiple tasks

As described above, the human subjects, all geologically naive laboratory personnel who volunteered as subjects, were given a set of six different task instructions. They were requested to select eight regions of interest, pertaining to each terrain picture and to each of the six instructed tasks; these data were summarized in the parsing diagrams (Figure 9).

For the same task (upper panel), and for the same person looking at the same picture the Sp R-similarity 0.65 and Ss R-similarity was 0.23. This was evidence for the "scanpath theory" now found with the new paradigm of "human choice". The L-similarities were also significant indicating that different persons looked at the same picture with somewhat similar scanpaths. The decrease of the local, L-coefficients 0.48 and 0.13 from the R-values 0.65 and 0.23 indicate that the scanpaths were somewhat different from person to person. Clearly, again the Ss similarities were less that the Sp similarities. All of these conclusions were strongly sustained by the relationship between the Anova test values and F-Fisher critical value (for *alpha* = 0.01) of 10.04.

When the task results were not segregated, but rather results for all tasks were merged the similarities become much reduced (lower panel). This strongly supports our use of "task definition" as an important protocol condition following the famous paper of Yarbus (Figure 109, page 174, Yarbus 1967) in Eye Movements and Vision, Alfred L. Yarbus, Plenum press, NY, 1967) The R-similarity and the L-similarity were still significant with respect with the F-Fisher critical value. The Sp value remained higher than the Ss value.

3.4 Algorithms vs. choices

Recall that we explored loci and sequences of aROIs with a wide variety of IPAs; certain of these IPAs were excellent predictor of human chROIs (Figure 10). Three different algorithms (Figure 10, upper, middle and lower) were related each to a particular subject and to a particular task. Their Sp similarity measures were 0.5, 0.88, 0.63, very good coherence indeed. Other algorithms were poor predictors of human chROIs (Figure 11); two different algorithms were each related to a particular subject and a to particular task and had low Sp similarity measures: -- Sp equal to 0.13 and 0.0 respectively (Figure 11, upper and lower). These figures served to illustrate the wide range of individual results we obtained.

On the basis of these large number of measures between algorithms, figures, tasks, and subjects we selected a combination of three different algorithms (michaelson, edge, symlet) that were combined to give a single new composite IPA, IPA*.

Next we averaged the similarity measure for Sp (Figure 12, upper) and Ss (lower) in confrontation tables that averaged over subjects but kept the similarity measures segregated for the six figures and the six tasks. Again note the uniformly low values for the Ss confrontation table; thus IPA* did not predict sequences for human choices.

However, the values for a particular task and for a particualr figure ranged up to a similarity value of 0.61 with an average value of 0.29. Note that the means for particular task averaged over all figures were fairly consisted in this was supported by a significant F-value. Contrariwise, the means for particular figures, but for different tasks were not significant correlated by F-test.

Again, we gathered the crucial comparisons between algorithms and human choices together into a parsing diagram (Figure 13). The ability of the algorithms to predict human choices was demonstrated by the numbers in the upper right box, L, of the left panel, Sp, for each picture segregated, but for all tasks combined. The average of all the twenty algorithms was only 0.24; for comparisons, human choices for different persons, again segregated for pictures but averaged for all tasks was 0.36 (Figure 9). For the combination of the three selected algorithms the value rose to 0.29. Finally, for this combination applied only to the data relating to Figures 4 and 6, the values rose to 0.43 and the Anova test showed very considerable significance (51 related to the critical value of 0.01).

This 0.43 value was averaged over the two figures and thus it related more closely, not to values segregated by picture, L, local, but to values averaged over all pictures, G, global. Note that the A and A* value were very similar for L and G, so this question was moot. Since we did not selected the algorithms on the basis of the particular task, this average of all tasks can be considered as a lower bound of a more careful selection procedure.

4.- DISCUSSION

4.1 Major Accomplishment

Results so far obtained and described are encouraging. We have accomplished our main goal -- to demonstrate that bottom-up autonomous algorithms can select and locate ROIs, regions-of-interest, that are related spatially to those chosen by human viewers. Indeed, in quite a few cases the algorithms, either particular ones in our original set, or the combination IPA* (of three algorithms) tentative selected by us as embodying important processes, actually performed very well for individual pictures. The level of prediction and identification by IPAs (Sp = 0.29) of important chROIs was almost as close to human as two humans might be expected to be (Sp = 0.36).

Underlying methods. The top-down structural binding of the spatial cognitive models, internal to the higher level human vision system, provides for the selection of the loci of the chROIs, that is where the eyes fixated or where the subjects choose important sub-features. With a bottom-up IPA, we are determining the aROIs in a bottom-up fashion from the picture information in the scene. Our research document our ability to do this. On what does this ability rest?

First, we developed an initial wide variety of IPAs and demonstrated very little redundancy amongst them. Second we developed a "selection" procedure based upon a matrix scheme originally developed to study human visual perception and EMs. The matrix used was Sp and Ss which calculated the similarity (1-distance) between pairs of ROI sets in terms of their location and sequences respectively.

4.2 Assisting the Mars Rover Exploration

Further we have established this in the context of pictures of Mars terrain and of geological exploration being carried our by the semi-autonomous Rover vehicle that landed on Mars in July 1997. Thus, these algorithmically-selected aROIs are now appropriate to the Martian terrain pictures being viewed and to the tasks of geological interpreters and of supervisory controllers of the remote planetary vehicle. In the future we plan to develop extensions for more precise task descriptions. We may recruit experts capable of this refinement of the tasks so that our new list selected algorithms will focus

on a particular task with perceptual and visual search dimensions and parameters that are explicitly more clearly specified.

4.3 Pictures

We are in the process of developing a corpus of geological pictures and of navigational terrain pictures so that specified tasks can be clearly attributed to certain regions for testing the algorithms. A much wider set of pictures, perhaps geological pictures from text books, and a wide inventory collectible over the Web, should be available for our future studies. Another interesting aspect that we have begun to look at is the influence of zooming in or out on the structure of navigational or geological scenes. How much intuitive mental zooming does the human perceptual brain utilize? We also want to study possible differences between pictures that have a considerable internal structure or organization, as for example in paintings of landscapes. These are in contrast to the more random, unstructured, or disorganized appearance of the planetary pictures we have used in the present report. This may lead us into interesting hypotheses regarding how the perceptual brain handles pictures with and without clear structure and organization.

4.4 Human Studies

We are especially interested in obtaining selection of chROIs from trained and experienced subjects. They should belong to two categories --- professional geologists and professional remote vehicle navigators. Our group plans to present pictures onto monitor screens at NASA-Ames Research Center and to collect the data directly via the internet as these professional subjects carry out their selections.

Choice vs Eye Movement Fixations. Another aspect of human studies, wherein we have already carried out some further studies (Privitera, Azzariti, Stark in preparation), attempts to answer an important question --- "How different or similar are the chROIs selected by 'conscious' human choice as compared with 'unconsciously' commanded sets of eye movement, emROIs, and their intervening fixations or glimpses that naturally are used to scan a picture.

4.5 Further development of IPAs

A planned study, which has been initiated is to modulate the scale and shift multiplier for a number of our already selected IPAs. These geometrical mapping functions may in some sense dominate the action of the IPAs and may provide important enrichment to the already wide variety of IPAs we have assembled. As described in the body of the report, the clustering algorithm plays an important role in our procedure, in reducing the selected points from 1000 to seven or eight, and in providing an ordering sequence. We have ongoing some preliminary studies on other varieties of clustering algorithms, (Krishnan, Privitera, Stark, in preparation) and are testing them to see how they enhance or degrade selective functions of the IPAs with respect to localization of the aROIs. Finally, we are interested in trivial modulations of the pictures so as to obtain R, repetitive parsing numbers; this top-anchor may be as useful as the bottom-anchors, G and Ra, in estimating the range of the similarity indices and in evaluating significance of the similarity numbers.

4.6 'A posteriori' Construction of IPA Kernels

In the 1950's and 1960's a MAMF procedure was intensely studied (Stark et al. 1962; Okajima et al. 1965) that used an adaptive routine to essentially capture kernel information from events being monitored and identified with bottom-up processes. Although these kernels were then one dimensional for events in time, they have been recently studied by Sun and Stark (personal communication) for two-dimensional icons in a random visual search field. The coefficients and the procedures resemble

ANNs, artificial neural networks with some important differences; the network is segmented according to the dimensions of the pre-processed data and the development of the coefficients, under autonomous (learning without a teacher) processes can be easily visualized and judged.

The convergence therefore is much more rapid and transparent than with more classical neural nets (and these latter most often learn by means of a teacher administering rewards and punishments. The adaptive resonance ANN, developed many years later by a researcher in Boston has many elements in common with the MAMF scheme. We have a plan to try the established MAMF method to develop alternative sets of structural and statistical kernels for our current well-defined and quantitatively measurable task.

5.- SUMMARY

We have been able to demonstrate that bottom-up IPAs, image processing algorithms can perform a new visual task --- to locate aROIs, regions-of-interest. These ROI loci are defined on the basis of a theory of top-down human vision, the scanpath theory. Further, using measures, --- Sp and Ss, the similarity of location and ordering, respectively, ---- developed over the years in studying human perception and the active looking role of eye movements, we can quantify the efficient and efficacious manner in which IPAs imitate human vision in located ROIs.

The structural binding of internal spatial-cognitive models is accompanied by a sequential binding that may be further or additionally activated by virtue of the sequential nature of human eye movements and of attentional shifts. This sequential binding has no parallel in IPAs and our results document that IPAs do not order the ROIs in a similar sequence to human scanpath: Ss similarity values are closed to random. Fortunately, one need not imitate this aspect of perception in order to develop useful application.

In any case, the means to quantify and evaluate IPA performance has been an important part of our study. These quantitative similarty measures are essential in selecting from the initial wide variety of IPAs, those particular ones that best serve for a type of picture and for a required task. Note that the selection depends upon the correlation with actual human chosen chROIs for the same type of picture and for the same required task to be accomplished.

Acknowledgment:

We are pleased to thank Drs. Theodore Blackmon, Michael Sims, Virginia Gulick, Jeff Cuzzi, Ted Roush all from Nasa-Ames for partial support and for suggestions, and, as well our laboratory colleagues, Drs. Huiyang Yang and Yeuk Fai Ho, for contributions to this study.

References:

Haralick, R.M. *Statistical and structural approaches to texture.* Proc. IEEE, 786-804 – 67, 1979.

Noton, D. and Stark, L. *Eye movements and visual perception.* Sci. American, 224:34 – 3, 1971.

Okajima, M., Stark, L., Whipple, G.H. and Yasui, S. *Computer Pattern Recognition Techniques: Some Results with Real Electrocardiographic Data.* IEEE Trans.Biomedical Electronics, 106-114 – (BME10), 1963

Privitera, C. and Stark, L.W. *Algorithms for defining visual Region-of-Interests: comparison with eye fixations.* Memorandum No. UCB/ERL M97/72, University of California, Berkeley.

Reed, R.T. and Hans Du Buf, J.M. *A review of recent texture segmentation and feature extraction techniques*. CVGIP: Image Processing, 379-72 – 57(3), 1993.

Stark, L. and Privitera, C. *Top-down and bottom-up image processing.* Proc. of IEEE ICNN'97, Houston, TX, June 9-12 1997, (4) 2294-99.

Stark, L., Okajima, M. and Whipple, G.H. *Computer Pattern Recognition Techniques: Electrocardiographic Diagnosis* Comm. Assoc. Computing Machinery, 527-532 – 5, 1962.

Yarbus, A.L. *Eye movements and Vision*, Plenum press, NY, 1967.