Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations

Claudio M. Privitera and Lawrence W. Stark, *Fellow*, *IEEE*

Abstract—Many machine vision applications, such as compression, pictorial database querying, and image understanding, often need to analyze in detail only a representative subset of the image, which may be arranged into sequences of loci called regions-of-interest, ROIs. We have investigated and developed a methodology that serves to automatically identify such a subset of aROIs (*a*lgorithmically detected ROIs) using different Image Processing Algorithms, IPAs, and appropriate clustering procedures. In human perception, an internal representation directs top-down, context-dependent sequences of eye movements to fixate on similar sequences of hROIs (*h*uman identified ROIs). In this paper, we introduce our methodology and we compare aROIs with hROIs as a criterion for evaluating and selecting bottom-up, context-free algorithms. An application is finally discussed.

Index Terms—Eye movements, scanpath theory, regions of interest identification and comparison.

1 Introduction

Eye movements are an essential part of human vision because they must carry the fovea and, consequently, the visual attention to each part of an image to be fixated upon and processed with high resolution. An average of three eye fixations per second generally occurs during active looking; these eye fixations are intercalated by rapid eye jumps, called saccades, during which vision is suppressed. Only a small set of eye fixations, hROIs, *h*uman detected Regions-of-Interest, are usually required by the brain to recognize a complex visual input (Fig. 1, upper panels). We have been studying and defining a computational model of this complex cognitive mechanism based on intelligent processing of digital images.

Image processing algorithms, IPAs, are usually intended to detect and localize specific features in a digital image analyzing, for example, spatial frequency, texture conformation, or other informative values of loci of the visual stimulus. Applying an IPA to an image means transforming that image into a new range of pixel values defining the corresponding algorithm feature. Local maxima in the transformed image represent loci wherein that particular feature is particularly accentuated and they can, consequently, be used as a basis for identifying aROIs, *a*lgorithmically detected Regions-of-Interest. Many local maxima may be generated by an image transformation: Therefore, a clustering procedure is required to reduce the initial large set of local maxima into a final small subset of aROIs (Fig. 1, lower panels).

aROIs and hROIs can be compared to each other through analysis of their spatial locations or structural binding and also analysis of the temporal order or sequential binding. The result of these comparisons measures the capability of an IPA, together with its clustering procedure, to predict hROIs. Thus, our aim is explicit and our measures quantitative. The overriding question is whether IPAs can treat an image in a fashion similar to human sequential glimpses.

In Section 2, the experimental protocol to acquire eye movement data is discussed in detail. Section 3 is devoted to defining a list of IPAs. The clustering and sequencing issue is considered and explained in Section 4. The computational and statistical platform used to compare hROIs and aROIs is introduced in Section 5. Top down vision and human scanpath are discussed in Section 6. In Section 7, the results of the comparisons are discussed and, finally, an application is presented in Section 8.

2 Stimulus Presentation and Eye Movement Measurement

Computer controlled experiments presented pictures and carefully measured eye movements using video cameras [21]. An infrared source light was projected toward the eye of the subject, generating a bright Purkinje reflection on the cornea, reflection that was easy to track by a video camera and the eye-tracking server. The subject was instructed to watch the visual stimuli (for a duration of four seconds, plus a calibration period before and after data acquisition) on a computer screen which was socket-connected to the eye-tracking server. The subject was seated in front of the screen with his head secured onto an optometric chin-rest structure. The viewing distance was approximately 40 cm from the computer screen; stimulus size was an average of 15 cm x 20 cm, yielding a subtended visual angle of approximately 21 × 29 degrees, and the resulting accuracy

- *The authors are with the Neurology and Telerobotics Units, 486 Minor Hall, University of California, Berkeley, Berkeley, CA 94720-2020. E-mail: {claudio@rov, stark@pupil}.berkeley.edu.*

Manuscript received 7 Apr. 1999; revised 28 Oct. 1999; accepted 7 June 2000. Recommended for acceptance by D. Jacobs.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 109560.

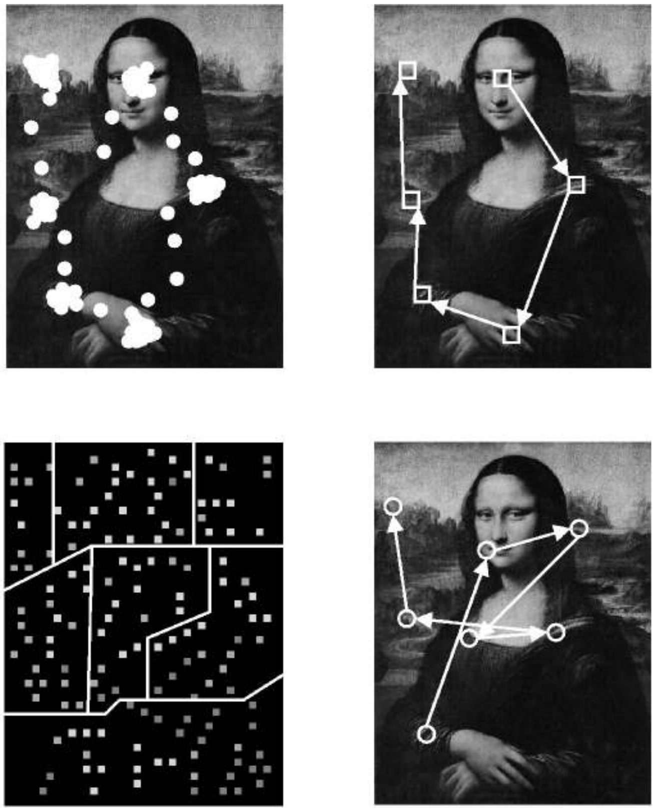

Fig. 1. Computer and human processing—comparing *h*uman identified Regions-of-Interest, hROIs, (upper right) with *a*lgorithmic identified Regions-of-Interest, aROIs, (lower right). Note eye movement sampling, (upper left) and local maxima in the processed image (lower left).

of the eye-position recording system was of the order of one-half to one degree of visual angle.

A fixation analysis algorithm was then applied to the eye movement data to distinguish rapid saccade jumps (Fig. 1, upper right panel, arrows) from location of eye fixations (Fig. 1, upper right panel, squares, note the eye movement sampling, upper left panel).

Seven subjects were used during eye movement experiments. Fifteen different images were utilized, including terrain photographs, landscapes, and paintings. We also used image modifications of some of these stimuli, such as embossed effect or binary thresholding. No specific instructions were given. All subjects had previously seen each picture at least once. Unfamiliarity with the viewed images may affect eye movement patterns [26] and it may correspondingly bias the results for some subjects. Since all observers had some degree of familiarity with the pictures and since no specific tasks were provided, we believe each observer looked at those pictures using intuitive and natural internal cognitive models (Section 6).

Each subject was asked to repeat the experiments within a few days for a total of four viewing sessions over approximately two weeks. By comparing different viewing sessions, we could study consistency in the way each subject looked at specific visual stimuli and we compared the consistent results with algorithmic performance. During each experimental run, the complete sequence of images, each time in different order, was displayed to the subjects.

3 Image Processing Algorithms (IPAs) Used for Identifying aROIs

The information content of a generic image can be abstracted by different image parameters that, in turn, can be identified by relevant IPAs. In this sense, applying algorithms to an image means mapping that image into different domains, where, for each domain, a specific set of parameters is extracted. Those parameters may be related to important attentional features of human vision. In our study, only the loci of the local maxima from each domain were retained in the processed image; these maxima were then clustered in order to yield a limited number of aROIs.

3.1 List of Algorithms

1. $\mathcal{X}$, an x-like mask of 7×7 pixels, positive along the two diagonals and negative elsewhere, was convoluted with the image. We also used different high-curvature mask convolutions, for example, the "$<$"-like mask whose definition is intuitive (see, for example, [14]). A scale of interest of 7×7 pixels corresponded, in our experiments, to a visual angle of 0.3×0.3 degrees × degrees (as a function of the viewer distance from the visual stimulus). This scale was empirically chosen on the basis of a preliminary study and several other factors, such as computational convenience.
2. $\mathcal{S}$, symmetry, a structural approach, appears to be a very prominent spatial relation (see, for example, [9]). For each pixel x, y of the image, we defined a local symmetry magnitude $\mathcal{S}(x, y)$ as follows:

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

where $\Gamma(x, y)$ is the neighborhood of radius 7 of point x, y defined along the horizontal and vertical axis

$$\begin{aligned}&(\Gamma(x,y) = \\ &\quad (x-r,y),\ldots,(x,y),\ldots(x+r,y), \\ &\quad (x,y-r),\ldots,(x,y+r))\end{aligned}$$

and $s((i_1, j_1), (i_2, j_2))$ is defined by the following equation:

$$\begin{aligned}&s((i_1,j_1),(i_2,j_2)) = \\ &\qquad G_\sigma\,(d((i_1,j_1),(i_2,j_2)))\,|cos(\theta_1-\theta_2)|.\end{aligned} \qquad (2)$$

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. Consequently, our definition of symmetry was based on the orientation correspondences of gradients around the centered point [18].

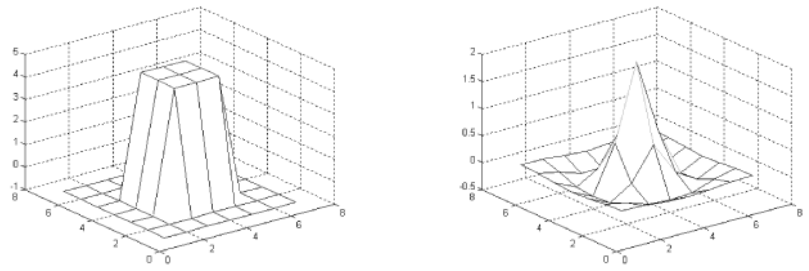

Fig. 2. Kernels $\mathcal{F}$ and $\mathcal{L}$—Algorithm $\mathcal{F}$, the quasi-receptive field mask (left) and algorithm $\mathcal{L}$, the Laplacian of the Gaussian (right).

Alternatively, a normalization of the axial quadratic moment could be used to compute the symmetry transform [6].

3. $\mathcal{W}$, a discrete wavelet transform, is based on a pyramidal algorithm which splits the image spectrum into four spatial frequency bands containing horizontal lows/vertical lows (lh), horizontal lows/vertical highs (lh), horizontal highs/vertical lows (hl), and horizontal highs/vertical highs (hh). This is achieved using a pair of conjugate quadrature filters, CQFs [24], which act as a smoothing filter (i.e., a moving average) and a detailing filter, respectively, (see, for example, [20]). The two filters are separately applied on each row and column of the input image. Each filtering is followed by a down-sampling by a factor of two which finally yields the four octave subbands. The procedure is repeatedly applied to each resulting low-frequency band resulting in a multiresolution decomposition into octave bands.

   We used different orders from the Daubechies $\mathcal{W}_{db}$ and Symlet $\mathcal{W}_{sy}$ family bases [4] to define CQF filters. For each resolution i, only the wavelet coefficients of the highs/highs hh_i matrix were retained (representing the details at each different resolution) and these were relocated into a final matrix HH (with the same dimension as the original image) by the following combination:

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

   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 an up-sampled copy of the input matrix hh: basically, the inverse of the down-sampling operation applied during the filtering process.

4. $\mathcal{F}$, a center-surround 7×7 quasi-receptive field mask, positive in the center and negative in the periphery, was convoluted with the image.

5. $\mathcal{O}$, difference in the gray-level orientation, is possibly also analyzed in early visual cortices (see also, [11]). Center-surround orientation difference was determined first by convoluting the image with four Gabor masks with angles $0°$, $45°$, $90°$, and $135°$, respectively. For each pixel, x, y, the scalars result of the four convolutions was then associated with four unit vectors corresponding to the four different orientations. The orientation vector $\bar{o}(x,y)$ was represented by the vectorial sum of these four weighted unit vectors. We defined the center-surround orientation difference transform as follows:

$$\mathcal{O}(x,y) = (1 - \bar{o}(x,y) \cdot \bar{m}(x,y)) \parallel \bar{o}(x,y) \parallel\parallel \bar{m}(x,y) \parallel, \tag{4}$$

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

6. $\mathcal{E}$, edges per unit area, was determined by detecting edges in an image, using the Canny extension of the sobel operator [3] and then congregating the edges detected with a Gaussian of $\sigma = 3$ pixels (see [19] for perceptive and psychology notions on edges).

7. $\mathcal{N}$, entropy was locally calculated as $\sum_{i \in G} f_i \, log \, f_i$, where f_i is the frequency of the ith gray level within the 7×7 surrounding region of the center pixel and G is the local set of gray levels. Local maxima defined by this factor emphasized texture variance.

8. $\mathcal{C}$, Michaelson contrast, is most useful in identifying high contrast elements, generally considered to be an important choice feature for human vision (see also, [10]). Michaelson contrast was calculated as $\|(\mathcal{L}_m - L_M)/(\mathcal{L}_m + L_M)\|$, where $\mathcal{L}_m$ is the mean luminance within a 7×7 surrounding of the center pixel and L_M is the overall mean luminance of the image. $\mathcal{L}_m$ was also used in our study.

9. $\mathcal{H}$, the discrete cosine transform, DCT, is used in several coding standards as, for example, in the Jpeg-DCT compression algorithm (see Section 8). The image was first subdivided into square blocks (i.e., 8×8 pixels); each block was then transformed into a new set of coefficients using the DCT; finally, only the high frequency coefficients were retained to quantify the corresponding block.

10. $\mathcal{L}$, the Laplacian of the Gaussian, was convoluted with the image (see Fig. 2 for comparisons with algorithm 4, $\mathcal{F}$).

3.2 Biological Principles of the Algorithms

We tried to gather a wide collection of algorithms of all sorts (10 algorithms are studied in this paper). Some of them do fit within the intuitive or partly empirical notions of human vision and visual neurophysiology, but gathered far and wide. We then let our experiments select which of this wide collection of algorithms, when faced with the task of finding aROIs, adhered to a similar identification of hROIs that our human subjects found as indicated by their eye movement fixations (see next section).

Can we look at the *successful* predictable algorithms and decide important truths about how the human brain controls our vision? In a sense, we can. As we said, it appeared that several of the algorithms acted as we might have intuited—looking at center-surround structure with high local contrast, finding symmetrical features, or areas with high edges density. This applies to ordinary pictures and scenes. Of course, the human visual brain is very flexible and, for a particular set of pictures, and for a particular task, and indeed with trained inspectors, quite nonobvious kernels might be utilized to great benefit. We suspect that the brain, with its enormous top-down approach, has the ability to synthesize image processing algorithms well beyond those that have been incorporated (by evolution) into the bottom-up mechanism of the retina and early vision cortical processing (which, alas, have been the ones most studied by vision neurophysiologists).

In other words, even those algorithms that do not have an intuitive and straightforward biological plausibility may be successful in predicting eye fixations. This is why we say that we don't want to select our algorithm a priori: Only a posteriori, e.g., after the comparisons with human data, we can finally identify—select—the best matching algorithms.

Our approach, in general, allows us to study these artificial IPAs and then provides us with a further opportunity to form new assumptions regarding saliency of a specific successful IPA based upon human experimental results.

4 Clustering and Sequencing

In general, in a three-second eye movement experiment, there were about seven to 11 fixations. As asserted in Section 2, a fixation analysis algorithm was applied to the eye movement data to distinguish rapid saccade jumps from location of eye fixations. Consecutive and very near fixations were usually merged into a single (and longer) fixation by the analysis algorithm: Saccades between those fixations are usually referred to as micro saccades. Seven aROIs represented the final average number of fixations from our experiments.

The IPAs resulted in defining many local maxima widely over the image; a clustering procedure was then applied to reduce this large set of local maxima to the final small subset of aROIs ($n \approx 7$). Thus, the resulting string of aROIs were similar in number to human eye movement fixation glances looking at similar images.

The initial set of local maxima was clustered by connecting local maxima and gradually increasing the acceptance radius for joining them. During each step of the clustering process, all local maxima less than a specific radius apart were clustered together (Fig. 3). Each cluster inherited the maximum value of its component points (local maxima): The locus of this highest valued maximum for each cluster then also determined the locus of that cluster. Only that maximum point was retained; all the others composing local maxima were deleted. The procedure was repeated while increasing the acceptance radius at each step. The decision to end the clustering process was reached only when a predefined number n of clusters remained. The values of the remaining clusters, ordered from highest to lowest, permitted us to relate the sequence of clusters, aROIs, to sequences of human fixations.

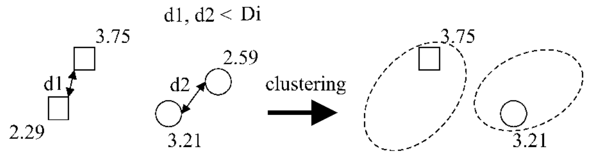

Fig. 3. Clustering Procedure: A Single Step—During each step of the clustering process, all local maxima less than a specific radius D_i apart, are clustered together. Then, the highest valued maximum of each cluster determines the locus for that cluster and all the remaining maxima are removed. The process continues as the acceptance radius D_i at each step increases.

Algorithm $\mathcal{N}$ was applied, for example, to a Chilean desert photo (Fig. 4, upper left panel) and the initial set of local maxima (Fig. 4, upper right panel) was then clustered (Fig. 4, lower left panel; partway through the clustering process). The final ordering is indicated by the arrows connecting the cluster loci and superimposed on the original image (Fig. 4, lower right panel). Note the maximum valued locus for each cluster.

Other clustering procedures have been investigated. However, no significant disparities in the overall performance of our system have been noted when different clustering procedures were compared to each other. This may be quite intuitive; it is the nature of the processed image (i.e., the IPA used), more than the clustering procedure that most influenced the final distribution of aROIs (Privitera et al. [15]).

If we had used only IPAs and not the clustering procedure, we could have selected, for example, the seven highest local maxima directly and defined them to be the aROIs. Those selected aROIs, however, might be much more closely spaced. Thus, the clustering procedure was actually an eccentricity-weighting procedure, where even lower local maxima that were eccentrically located could finally be selected to form an aROI. This is important for comparison to human fixations because subjects often focused their attention on significant eccentric loci on the image.

All the algorithms and the clustering procedure were implemented in Matlab and executed on a Pentium II PC. The execution time for the algorithms ranged from a few minutes to several minutes depending on the algorithm. The clustering procedure typically took about 10 minutes.

5 Comparing and Sorting Procedures

The aROI loci selected by our different IPAs and those loci defined by human eye movement fixations, hROIs, can be compared. In this section, we describe the statistical and computational platform we have been using for these comparisons (see also [16]).

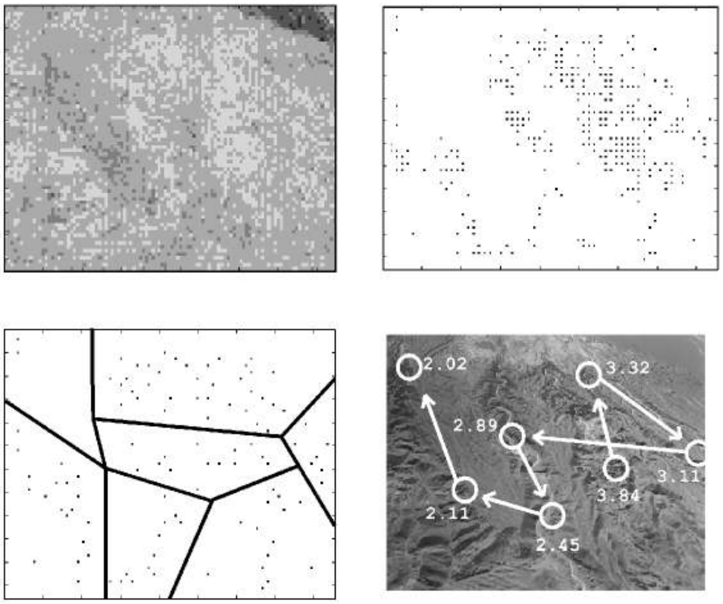

Fig. 4. Clustering Procedure: To completion—Algorithm $\mathcal{N}$, entropy, is applied to a Chilean desert photo (upper left). The initial set of local maxima (upper right) is then clustered using the defined iterative procedure (lower left: partway through the process). The final ordering, superimposed on the original image, is shown in the lower right panel; the maximum-valued locus for each cluster is inserted in the figure.

5.1 Comparison of Two Sets of ROIs

Comparison of final clusters of ROIs began with taking two sets of ROIs (Fig. 5, middle, upper, and lower panels) and clustering these two sets using a distance measure derived from a k-means preevaluation. This evaluation determined a region for calling any ROIs that were closer than this distance coincident and any ROIs that were further apart than this distance noncoincident; the distance was about two degrees and similar in size to human foveal spans for moderate visual acuity. All the coincident ROIs (named *joined-ROIs*) were labeled with the same alphabetic character (Fig. 5, right panel) and they then enabled a similarity metric, S_p, to determine how many ROIs two algorithms (as in the example shown in Fig. 5, note the processed image in the left panels), or two humans, or an algorithm and a human had in common; the final value was normalized based upon string length. The individual sources of the elements, that is, the original ROIs, used in these final interactive steps are preserved as circles and squares (Fig. 5, right panel) to illustrate the procedure.

As mentioned above, ROIs were ordered by the value assigned by the IPA or by the temporal ordering of human eye fixations in a scanpath. Then, the joined-ROIs could finally be ordered into strings of ordered points. Here (Fig. 5), we have, for example: $string_{\mathcal{E}} = abcfeffgdc$ and $string_{\mathcal{S}} = afbffdcdf$. The string editing similarity index S_s was defined by an optimization algorithm [21] with unit cost assigned to the three different operations: $deletion$, $insertion$, and $substitution$.

In summary, our comparisons yielded two different indices of similarity which told how closely two sets of ROIs resembled each other in locus, S_p, and in sequence, S_s. For the example, illustrated in Fig. 5, all the labels of the second string ($afbffdcdf$) are included in the first string ($abcfeffgdc$), yielding an S_p similarity value of one. Sequentially editing the first string to match the second string, however, yields a much lower similarity index S_s. Substituting the third b with e generates $afeffdcdf$, cost 1; inserting b and c after the first a generates $abcfeffdcdf$, cost 2; deleting the last d and f generates, $abcfeffdc$, cost 2; finally, inserting g generates $abcfeffgdc$ which is equal to the second string, cost 1. The total cost is six and, since the original string length was nine, S_s, the sequential similarity between the two strings is $(1 - 6/9) = 0.34$ (see [2] for more details on string editing).

5.2 Y-Matrices and Parsing Diagrams

Similarity coefficients can be sorted and represented for the two measures S_p and S_s and explicitly displayed in a table, named the Y-matrix, having as many rows and columns as the number of the different sequences ROIs to be considered (Fig. 6, upper panels). A complete Y-matrix would be, of course, too large to display and, thus, to read. Parsing diagrams (Fig. 6, lower panels), with averages of similarity coefficients collected from the arrays of the Y-matrices, are a compact and intuitive alternative way to look at the data: R for Repetitive scanpaths, same subject looking at the same picture at different times; L for Local, different subjects, same picture; I for Idiosyncratic, same subject, different pictures; G for Global, different subjects and different pictures.

For our human experiments, the truncated Y-matrix (Fig. 6, upper panels), representing only a small part of the entire set of comparisons, refers to only two images and two subjects. This Y-matrix, however, is sufficient to illustrate how Y-matrices are translated into a parsing diagram. For example, the Y-matrix diagonal represents the S_p autosimilarity coefficients (labeled R) of each subject looking at the same picture over different times; these coefficients then generate a unique averaged coefficient reported in the Repetitive box of the parsing diagrams (Fig. 6, lower panels). The same ordered collection of the coefficients of

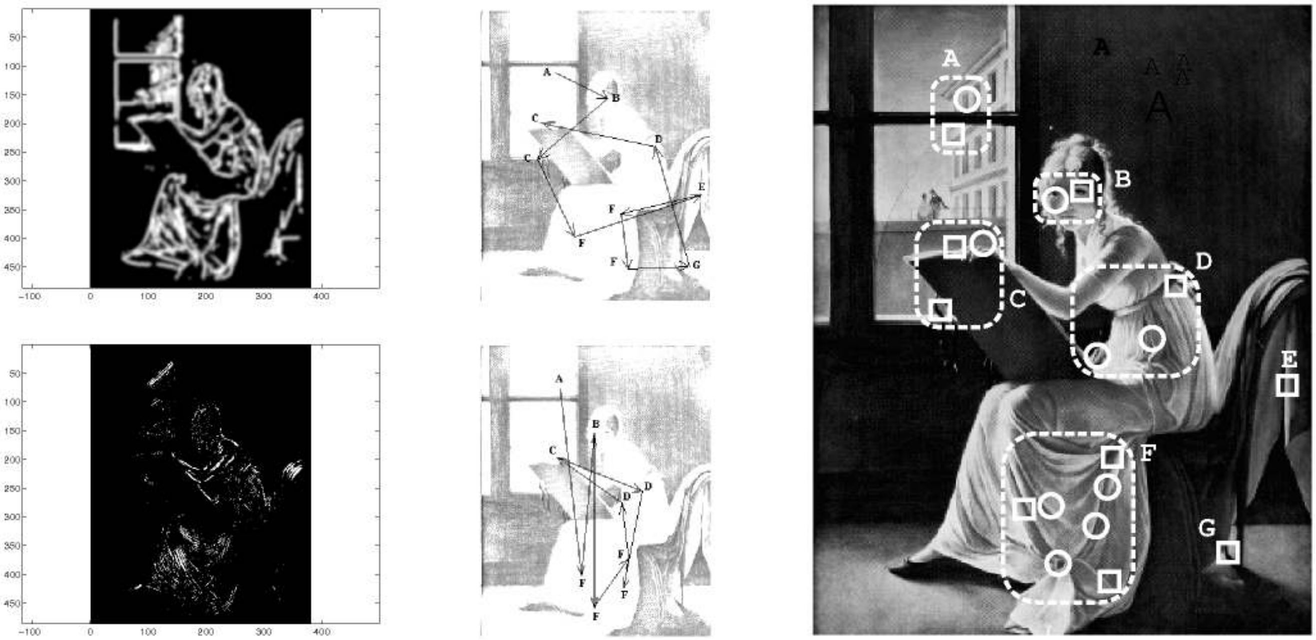

Fig. 5. ROIs Comparisons Procedure—Action of each algorithm yields a transformed image (left column) for two examples, $\mathcal{E}$ (edges per unit area, upper), and $\mathcal{S}$ (symmetry transform, lower). Final aROIs in each image are ordered by value and connected with arrows in analogy to eye movement sequences of fixations (central column, the original image is display at very low contrast to help the viewing of the arrows). The two sets of aROIs are finally combined (right panel) into a number of $joined$-ROIs (labeled with capital letters), further used to measure the distance between the two sets.

the Y-matrix arrays is applied for the other types of comparisons: Local, Idiosyncratic, and Global. The parsing diagram (lower panels, Fig. 6) refers to all images and subjects; in general, we report Y-matrices only when a restricted subset of comparisons needed to be considered.

The most important distinction is that between Repetitive similarity, R (upper left box, Fig. 6, lower panels), and Global similarity, G (lower right box). The R value for human, with the S_p measure, equaled 0.64. This means that the strings for repetitive viewing of the same stimulus for the same subject have loci that were 64 percent within fixational or foveal range—this represented continuing support for the scanpath theory (see the following Section 6). For Global, all different subjects looking at all different stimuli had an S_p value of only 0.28. This number was somewhat different from the expected S_p value of 0.21 based upon consideration of a random model, Ra (bottom box).

6 Top-Down Vision and the Scanpath Theory

The scanpath was defined on the basis of experimental findings. It consists of sequences of alternating saccades and fixations that repeat themselves when a subject is viewing a

Y-MATRICES (partial: 2 subjects)

Sp	Subject1		Subject2	
	Pict1	Pict2	Pict1	Pict2
S1p1	$0.65_{\mathbf{R}}$	$0.38_{\mathbf{I}}$	$0.54_{\mathbf{L}}$	$0.18_{\mathbf{G}}$
S1p2		$0.60_{\mathbf{R}}$	$0.31_{\mathbf{G}}$	$0.47_{\mathbf{L}}$
S2p1			$0.69_{\mathbf{R}}$	$0.33_{\mathbf{I}}$
S2p2				$0.58_{\mathbf{R}}$

Ss	Subject1		Subject2	
	Pict1	Pict2	Pict1	Pict2
S1p1	$0.40_{\mathbf{R}}$	$0.24_{\mathbf{I}}$	$0.31_{\mathbf{L}}$	$0.08_{\mathbf{G}}$
S1p2		$0.39_{\mathbf{R}}$	$0.13_{\mathbf{G}}$	$0.19_{\mathbf{L}}$
S2p1			$0.43_{\mathbf{R}}$	$0.21_{\mathbf{I}}$
S2p2				$0.24_{\mathbf{R}}$

PARSING DIAGRAMS (complete: 7 subjects)

Same Subject	Different Subjects		Same Subject	Different Subjects
Repetitive 0.64	**L**ocal 0.54	Same Picture	**R**epetitive 0.42	**L**ocal 0.28
Idiosyncratic 0.34	**G**lobal 0.28	Different Pictures	**I**diosyncratic 0.21	**G**lobal 0.16
Sp	**R**andom 0.21		**Ss**	**R**andom 0.04

Fig. 6. Y-matrices and Parsing Diagrams—S_p and S_s similarity indices for different subjects (or different algorithms) and for different pictures can be arranged in a Y-matrix (upper panels, only two subjects and two pictures are reported for clarity) with each value being the average of several repetitions. Parsing diagrams (lower panels, for all the subjects who participated in the eye movements experiments and all images) represent the averages of these similarity indices (see bolded letters in the Y-matrix) in a more collected and intuitive fashion.

picture. Only 10 percent of the duration of the scanpath is taken up by the collective duration of the saccadic eye movements, which thus provide an efficient mechanism for traveling over the scene or regions of interest. Thus, the intervening fixations or foveations onto hROIs have at hand 90 percent of the total viewing period (see [1] and also Section 2). The glimpses or fixations place the high resolution foveal center of the retina onto hROIs; of course, low resolution peripheral vision plays an additional important role.

Scanpath sequences appear spontaneously without special instructions to subjects and were discovered to be repetitive; note the high R indices in the parsing diagrams (Fig. 6: 0.64 for S_p and 0.42 for S_s). This repetitiveness suggested to Noton and Stark [12] that a top-down internal cognitive model controls perception and active looking of eye movements in a repetitive sequential set of saccades and fixations, or glances, over features of a scene so as to check out and confirm the model [21]. Other evidence comes from studies of eye movements during visual imagery experiments [2] and ambiguous figures [23].

When the same observer was asked to look repetitively at different modifications of the same pictures over different viewing sessions (embossed and binary thresholding modifications of Madame and After the Shower), we found high R-similarities: 0.45 for S_p and 0.38 for S_s. A possible interpretation of this result is that the same internal model controlled eye movements for different modest modifications of the same pictures, further validating the scanpath theory.

Of course, the objective or task can affect the active looking of eye movements [25], [26]. Nevertheless, without any specific task instruction, for general viewing conditions, when different subjects look at the same picture, they are fairly consistent in identifying regions of interest as indicated in this study by the high L values (Fig. 6: 0.54 for S_p and 0.28 for S_s).

The strong scanpath consistency reported in human experiments when no specific objective is given to the subjects means that only a specific restricted set of representative regions in the internal cognitive model of the picture is essential for the brain to perceive and eventually recognize that picture. This representative set is quite similar, (Fig. 6, $S_p = 0.54$) for different subjects and this brings us to the main scientific objective of our work: Whether it is possible to identify automatically this set by using IPAs.

Comparing aROIs with hROIs is the standard utilized to study and select which IPAs are more successful in this objective; if a specific task is given, different hROIs may result and, consequently, different algorithms may be selected from our collection [15].

The global similarity value, G, represents any invariant components of eye movements—the use of some global eye movement strategy control, as an unlikely example, the tendency to start at the center of the image and then scan circularly around the periphery. Indeed, reading eye movements has a high G value since all English readers start at the upper left and proceed horizontally across each line and descend vertically line by line [21]. Global similarity is actually the antithesis of our basic theory because it would prove that a general and invariant motor program controls eye movements rather than an idiosyncratic internal motor model based on image-specific modeled regions of interest, hROIs. However, our results always showed that this component is very low, but usually somewhat higher than Random. Consequently, G's similarity is considered as a bottom anchor for our scale of comparisons.

Random similarity value, Ra, is more intuitive than global similarity and it is usually considered as a second important bottom anchor for our comparisons: It represents how similar randomly generated scanpaths are to each other. This value is 0.21 and it is equivalent to the similarity value between randomly generated scanpaths and hROIs.

Finally, we would like to guide the reader's attention again to the Local value of the S_p-parsing diagram. This value was 0.54 and it basically says that, when different people view the same image, only an average of 54 percent of their hROIs cohere. This was an important result for this study, in fact, algorithms cannot be expected to predict human fixations better than the coherence among fixations of different persons. Consequently, this Local value must be considered as our main (top-anchor) criterion for evaluating the performance of our algorithms when compared with human data (Section 7).

7 hROIs vs. aROIs Comparisons

We are, of course, most interested in the use of our methodology to test our data—to analyze the capability of IPAs and clustering procedure to predict eye fixations. But, we are also interested in the interrelationships among algorithms. In this section, we present and discuss the different aspects derived from the comparison results.

7.1 Parsing Diagrams

We gathered the crucial comparisons between algorithms and eye fixations together into parsing diagrams (Fig. 7). The ability of all the algorithms (labeled A in Fig. 7) to predict eye fixations was demonstrated by the number in the upper right box, L, of the left panel, S_p. The average for all the algorithms applied to all the pictures was 0.33.

On the basis of having this large a number of measures between algorithms, images, and subjects, we selected a subgroup of algorithms ($A* = \mathcal{W}_{db}$, $\mathcal{L}$, $\mathcal{O}$, and $\mathcal{S}$). For this selection, the S_p similarity rose to 0.36 and the Anova test showed a considerable significance (27.0 related to the F-Fisher critical value of 7.5, see, Appendix).

Fig. 8 shows two qualitative examples of our algorithms for two different images (a painting and a photo of the Chilean desert). aROIs and hROIs are superimposed upon the image and they cohere well: aROIs on the left panels are represented with a sequence of circles and they correspond to algorithms $\mathcal{L}_m$, upper, and $\mathcal{W}_{sy}$, lower; hROIs on the right are represented with a sequence of squares. The S_s parsing diagram showed little coherence, even among all the algorithms, providing support of an earlier preliminary study, [22], that the IPAs and the clustering procedures we used cannot predict sequencing of human eye movements.

The parsing diagrams reported in Fig. 7 represent the result for all scenes. Algorithms can be selected a posteriori

	Repetitive	Local	**Sp**
Same picture	**1**	Alg. = **0.33** (0.04 **18.7**) Alg.* = **0.36** (0.01 **27.0**)	
	Idiosyncratic	**G**lobal	
Different pictures	Alg. = **0.20** (0.03 **1.62**) Alg.* = **0.17** (0.05 **0.03**)	Alg. = **0.26** (0.05 **6.16**) Alg.* = **0.24** (0.05 **4.16**)	

Random = 0.21

Algs vs. Algs — Algs vs. Eye fixations

	Repetitive	Local	**Ss**
Same picture	**1**	Alg. = **0.05** (0.01 **3.31**) Alg.* = **0.05** (0.01 **2.21**)	
	Idiosyncratic	**G**lobal	
Different pictures	Alg. = **0.03** (0.01 **0.12**) Alg.* = **0.02** (0.01 **0.01**)	Alg. = **0.04** (0.01 **0.44**) Alg.* = **0.04** (0.01 **0.32**)	

Random = 0.04

Fig. 7. Parsing Diagrams for Comparing aROIs and hROIs—Crucial comparisons between all algorithms, eye fixations, and images are gathered in the parsing diagrams for S_p, upper panel, and S_s, lower panel. Mean values are in bold outside the parentheses. Tests of significance are represented by the Anova test value (in bold within the parentheses to the right of the standard deviation, see the Appendix). Left column, boxes **R**, Repetitive, and **I**, Idiosyncratic, show the autocorrelation when the same algorithm is applied to the same, **R**, or different, **I**, pictures. Of course, **R** is 1. Right column, boxes **L**, Local, and **G**, Global, show the cross-correlation between algorithms and eye fixations. Note box **L** where the similarities between algorithms and eye fixations refer to the same picture (and then averaged for all pictures).

based on the S_p comparison results between aROIs and hROIs. Four algorithms, for example, $\mathcal{X}$, $\mathcal{S}$, $\mathcal{W}_{db}$, and $\mathcal{F}$, seemed to cohere very well with human data, subjects A, C, H, and T, for a specific class of images (paintings). The corresponding similarity values for those algorithms and human scanpaths were extracted from the original wide Y-matrix (which includes all the comparison values) and explicitly reported in Fig. 9 (average coherence between aROIs and hROIs was 0.56). The remaining algorithms are not shown in the same figure for simplicity and it is implicit that their coherence values with the same human data (and the same class of images) is lower: an average of 0.37. A third coherence (not shown) was achieved for the set of Mars terrain images alone and algorithms $\mathcal{C}$, $\mathcal{W}_{sy}$, and $\mathcal{E}$ (average coherence was 0.43).

Fig. 9 is intended to show how, for a particular class of images, it is possible to identify a few algorithms whose similarity coefficients approach top-anchor criteria. This is despite the overall moderate similarity obtained when the average of the entire collection of algorithms is considered (parsing diagram in Fig. 7).

Several subjects from our lab, extraneous to this project, were asked to qualitatively judge the distribution of aROIs over the pictures for all the algorithms and the pictures that have been used in this study. The subjects that participated in the test were asked to judge, based on their own intuition, the position of aROIs: whether or not the set of aROIs for a specific algorithm was located on significant regions of interest and whether or not the regions of interest of the image were all represented by that set of aROIs. Three different grades were suggested: good, medium, and bad.

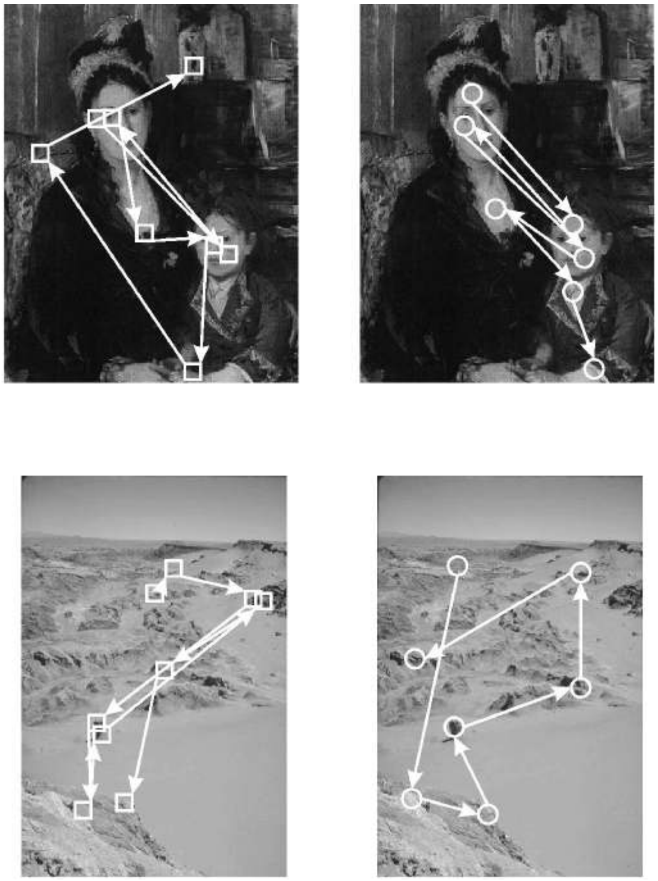

Fig. 8. Qualitative Comparisons of hROIs and aROIs—Two examples of good S_p-similarity between aROIs (right column: $\mathcal{L}_m$, luminance, upper and $\mathcal{W}_{sy}$, wavelet, lower) and hROIs (left column). $S_p = 0.62$ and $S_s = 0.13$, upper panel; $S_p = 0.87$ and $S_s = 0.13$, lower panel. Note low values of S_s indicate that string sequences could not be identified by the algorithms.

First, a total of 64 percent of aROIs were considered acceptable or good. For each image, we could finally rank each of the algorithms based on the average grade it obtained. Then, ordering these results for each algorithm and for each image, we computed the correlation with the ordering generated by our S_p metric. The average correlation was quite high, around 0.7, and it confirmed the relationship between these human qualitative evaluations of aROIs and our metric S_p.

Sp	subject A	subject C	subject H	subject T	*alg. Xmask*	*alg. Symm*	*alg. Wavelt*	*alg. Fmask*
subject A	1	0.23	0.54	0.64	0.60	0.67	0.72	0.64
subject C		1	0.69	0.86	0.78	0.78	0.73	0.40
subject H			1	0.42	0.52	0.60	0.40	0.51
subject T				1	0.42	0.42	0.47	0.28
alg. Xmask					1	0.83	0.87	0.66
alg. Symm						1	0.78	0.85
alg. Wavelt							1	0.51
alg. Fmask								1

Fig. 9. Coherence and Independence among Selected Algorithms—Y-matrices for a selected set of images (paintings) and a selected group of algorithms $\mathcal{X}$ the x-like mask, $\mathcal{S}$ symmetry, $\mathcal{W}_{db}$ wavelet, $\mathcal{F}$ the quasi-receptive field mask, and subjects, A, C, H, and T. Algorithms are selected a posteriori based on cross-similarity with human eye fixations (see average values in the parsing diagrams).

Sp	*alg.*	*alg.*	*alg.*	*alg.*	*alg.*	*alg.*
	Wavlet-db	*Wavlet-sy*	*Laplacian*	*N-entropy*	*F-mask*	*Contrast*
alg. Wavlet-db	-	0.57	0.37	0.21	0.28	0.15
alg. Wavlet-sy		-	0.39	0.18	0.24	0.22
alg. Laplacian			-	0.19	0.25	0.16
alg. N-entropy				-	0.69	0.61
alg. F-mask					-	0.63
alg. Contrast						-

Ss	*alg.*	*alg.*	*alg.*	*alg.*	*alg.*	*alg.*
	Wavlet-db	*Wavlet-sy*	*Laplacian*	*N-entropy*	*F-mask*	*Contrast*
alg. Wavlet-db	-	0.31	0.05	0.02	0.05	0.02
alg. Wavlet-sy		-	0.09	0.00	0.01	0.02
alg. Laplacian			-	0.01	0.01	0.05
alg. N-entropy				-	0.23	0.22
alg. F-mask					-	0.33
alg. Contrast						-

Fig. 10. Coherence and Independence among Algorithms—Cross-comparison values of six algorithms for two indices, S_p and S_s. Enclosed within the dashed boxes are two different groups of algorithms: Each group is internally characterized by high S_p similarity, but cross-similarity in S_p between groups is very low. This means that algorithms are widely chosen and that they may have independent or similar affects on images. Note that S_s values are very low.

7.2 Criterion for Evaluating the Algorithms

A main characteristic of our approach is the definition of two quantitative metrics, S_p and S_s, utilized to validate the algorithms and to analyze eye movement patterns. S_p is the coherence in the location of two sets of regions of interest, ROIs; S_s is the coherence in the ordering of the two sets.

When S_p was used to evaluate the coherence of different human subjects viewing the same image, the resulting coefficients, averaged for all images and subjects, was 0.54 (see the L, local value, in the parsing diagram, Fig. 6). This signifies that, when different subjects view the same image, an average of 54 percent of their hROIs did cohere. Algorithms cannot be expected to cohere with human fixations better than the level of coherence amongst fixations of different humans: This is why the L local value of 0.54 is considered to be the main criterion and the results from our algorithms discussed in the manuscript should be considered in the light of this criterion.

As reported above, when we selected some of the algorithms on the basis of specific types of images (for example, paintings or Mars pictures), the S_p coherence between hROIs and aROIs ranged from 0.43 to 0.56. We consider this level of coherence as a very positive and important result. Thus, our large collection of algorithms can provide different selection policies both for different images and for different tasks [15]. In general, averaging for all images and algorithms, S_p ranged from 0.33 to 0.36 (as reported in Fig. 7): still significantly higher than the chance level.

The qualitative evaluation of the algorithms in the form of a questionnaire was very helpful.

Combining different IPAs may improve this predictability and the versatility of the system for a larger class of images and our algorithms can be used as building blocks (see, for example, [8], [5], [13]). Other models in the literature are often evaluated uniquely from a qualitative point of view, without rigorous metrics and, especially, without really taking into account experimental human data (the L local similarity for example). Contrariwise, we evaluated our algorithms using the same statistical procedures and metrics used to study eye movements and we did document eye movement results: We believe that this provides the best term of comparison.

7.3 Relationships Among IPAs

We wished to obtain as wide a variety of image processing algorithms as possible and to keep the coherence between pairs of image processing algorithms small. Thus, our wide variety of image processing algorithms would have independent actions on the images and they could serve to identify aROIs for a variety of picture types, and for a variety of visual identification tasks [15].

Cross-similarity coefficients are shown (Fig. 10, see also, [16] for preliminary results), for example, for a group of six algorithms, $\mathcal{W}_{db}$, $\mathcal{W}_{sy}$, $\mathcal{L}$, $\mathcal{N}$, $\mathcal{F}$, and $\mathcal{C}$; the average similarity for S_p, left panel, was 0.44 (0.57, 0.39, and 0.37) between the first three algorithms and 0.64 (0.61, 0.63, and 0.69) for the second three. Cross-similarity between elements of the two subgroups was 0.22 (ranging from 0.28 to 0.15), almost the same self-consistency found for different instances of the random algorithm. The values for the first and the second group were significantly higher than the random self-consistency: For the second group, the values really approached the same high significant repetitive self-consistency found in human experiments. The coefficients in Fig. 10 show how algorithms can be widely chosen and that they may have independent or similar effects (bolded rectangulars in Fig. 10) on images. Note that the coefficients for S_s were much lower than the coefficients for S_p.

The other algorithms have not been inserted for simplicity in the table of Fig. 10 because they are not together, a third independent group: When compared with the rest of the algorithms, the resulting S_p cross-similarity values range between the extreme values reported for the two groups in Fig. 10.

7.4 Overall Performance of the Algorithms

Our metrics, Sp and Ss, served to evaluate coherence between algorithms and human subjects. We were not only interested in proving (or counterproving) the predictability

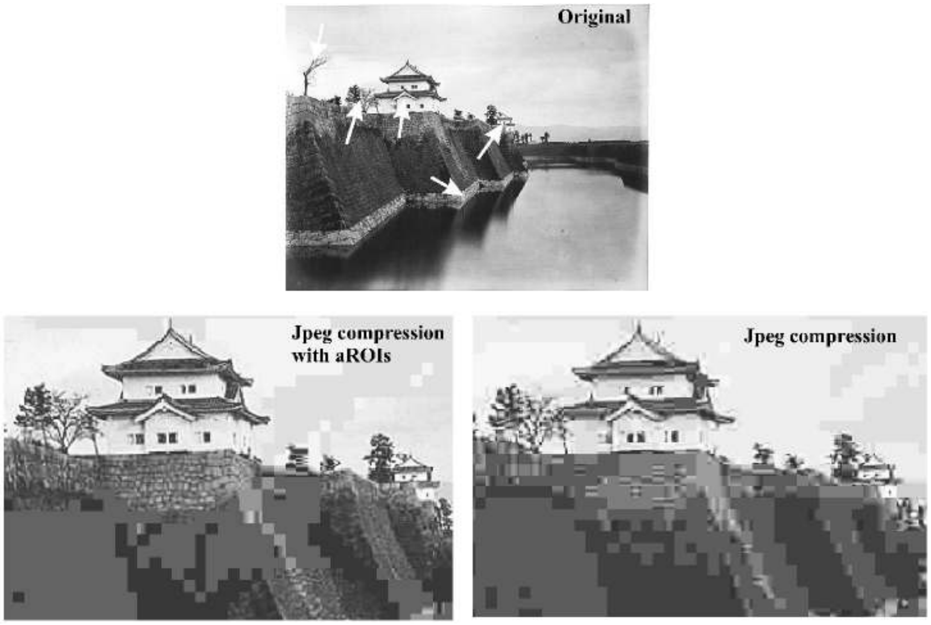

Fig. 11. *Focused JPEG* compression based on aROIs—Five aROIs (arrows) were identified using $\mathcal{O}$, orientation; aROIs were maintained at higher resolution by the *Focused JPEG* compression. The *Focused JPEG* compressed image is shown in the lower left panel and a standard JPEG compression in the lower panel. Total compression was the same in both examples, but the visual fidelity is much higher with *Focused JPEG* compression.

of eye fixations but also in studying the coherence between algorithms in the attempt to obtain as wide a variety of algorithms as possible and eventually to formulate new kernels. Of course, some pairs of algorithms cohered. Nevertheless, our results identified independent groups of algorithms (see example in Fig. 10); thus, our collection of algorithms could be sorted for similarities or for differences in generating aROIs. This sorting of algorithms has three objectives. First, we aim at achieving efficiency by not using redundant algorithms. Second, this will help us in our search for an even wider spectrum of algorithms, some of which might play an important role in particular tasks. Third, we might be able to understand some aspects of human vision.

In general, Wavelets, $\mathcal{W}$, seemed to be very efficient for several classes of images—maybe due to their implicit multiresolution analysis, which allows extracting local maxima corresponding to different feature scales. Symmetry, $\mathcal{S}$, seemed to be very efficient for general images (paintings, for example, as reported previously), confirming some earlier psychophysics results. Contrast, $\mathcal{C}$, cohered very well for terrains: We used rocky Mars terrain images where rocks (likely to be identified by the viewers) were easily discerned from the soil by their darker material strongly contrasted with the uniform brighter color of the soil. Contrast is generally considered to be an important feature for human vision.

Based on the subjective questionnaire results, Edges per unit area, $\mathcal{E}$, and Orientation, $\mathcal{O}$, also well-satisfied the subjective qualitative parameters of the viewers for general images (not including terrains and the aerial photo): They also have some structural relations with features experimentally found in visual receptive fields. The DCT algorithm $\mathcal{H}$ seemed poorly cohered with human data for all the images used in the study.

The goal of this paper is to propose an engineering approach to substitute for human visual attention and eye movements. The set of algorithms we picked up represents, of course, only a small portion of the many different kernels and procedures that could be utilized. However, our selected set of algorithms within our experimental conditions could indeed predict eye fixations; this may have several cognitive and neurological interpretations that are beyond the scope of the present manuscript and that we would like to leave open for future research and the readers' intuition.

8 Application of aROIs

Certain computer vision applications might benefit from an apparatus that automatically identifies regions of visual interest in a digital image. For example, we have defined a JPEG encoder, named *Focused JPEG,* based upon on aROIs (see also [17]). The difference with the standard JPEG baseline is that in our *Focused JPEG* baseline the magnitude of the quantizer factors for each 8×8 pixels JPEG block are adaptively related to the distance from the set of aROIs by means of the following rule:

$$Q(x,y)_i = Q_i S(d_{min}(x,y)), \tag{5}$$

where $d_{min}(x,y)$ is the minimum distance between the block x, y in the image and the set of aROIs. $S(\cdot)$ is a stepwise monotonic function equal to the unity for distance $d_{min}(x,y)$ that is appropriately small (center of ROI) and then increasing with the distance; Q_i is the original standard quantizer coefficient.

Algorithm $\mathcal{O}$ was applied, for example, to a countryside photo and five aROIs have been identified (Fig. 11, upper panel). The *Focused JPEG* compression was then applied to the image based on those identified aROIs where

$S(d(x,y)) = 2$ for $0^\circ \leq d(x,y) < 1^\circ$, degree of the visual angle (note that aROIs are also slightly compressed); $S(d(x,y)) = 250$ for $1^\circ \leq d(x,y)$. The *Focused JPEG* compressed image (Fig. 11, lower left panel) can be compared to the standard JPEG compression (Fig. 11, lower right panel) with the same amount of compression (100 : 4), see, also [27].

9 Discussion

Our method provides a precise task for the IPAs 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 prediction accuracy.

In this paper, we have validated that a constellation of IPAs used in conjunction with a clustering procedure can predict, for S_p, the loci of human fixations. Our results indicate, however, that the algorithms cannot predict the sequential ordering, S_s, of the subfeatures used by a person. Human ordering is idiosyncratic to the subject and to the specific image, as reported in the famous book by Yarbus book [25] and deeply investigated in our lab over the past decades of eye movement experiments. Thus, it is unreasonable to expect an algorithm to be able to predict the temporal order of eye fixations and the poor S_s predictability of the algorithms serves as an important counterexample to our technology.

The wide selection of algorithms gives us an opportunity to study the differences and similarities in terms of the precise task we consider. These algorithm characteristics are of great interest to us as indicators of the general nature of a picture and how either algorithms or humans process it. We might need to provide weighting coefficients for the different algorithms in order to optimize the prediction capabilities of the ensemble.

In addition to finding positive image processing algorithms by comparing their similarity with human data, we can also consider random scanpaths, raROIs, which turn out to have a much smaller average S_p-similarity with human scanpaths for each of the images used in our study, hROIs (see Fig. 7 for example). Let us call the random values the bottom anchors of our working range of similarity. Of course, with any large set of random data, one example might cohere very closely to human data. However, our multiple sets of images and multiple trials make the average values for random data not only low but also robustly so.

The top of the scale is anchored in human studies by the repetitive value, R. More interesting than the Repetitive anchor for our main question is the Local coefficient: how similarly two different subjects look at the same picture. The L coefficient enables us to determine if image-processing algorithms can predict hROIs as well as one subject's eye movement pattern can predict the eye movement pattern of another subject for the same image. Now, we can position an S_p value as lying somewhere in this working range: The highest values are close to the high similarity in loci or S_p of two humans looking at the same image. The lowest values are the random values.

The clustering procedures we used require a good deal of thought and preliminary studies have been reported in Section 4. As indicated, the clustering procedure distributes strings of aROIs in more eccentric locations than they would be in without the clustering procedure. This eccentricity asserted a positive effect on the similarity between aROIs and hROIs.

In summary, the methodology defined in this paper has been tested on a varied set of digital images that ranges from portraits to landscapes and terrain images. A number of subjects were used for the eye movement experiments. Finally, independent subjective evaluations by naive subjects further validated the results. The overall results are very encouraging and we have started to define and implement different applications such as image compression which has been presented in the last section.

Appendix

Study of the ANOVA: ANalysis-Of-VAriance

Regions of interest, ROIs, are generated either in human experiments, hROIs, or using image processing algorithms, aROIs, or randomly, raROIs; each ROI is a two-dimensional vector representing the sequence of x, y coordinates of the regions of interest. A specific ROI is identified by the image to which it corresponds, say I, and the agent that generated it, a for algorithms, h for humans, and ra for random algorithms: The index, I, varies for all the images, h, for all the subjects participating in the experiments and, finally, a for all the algorithms studied in the paper. Agents h and ra are repetitively applied to the same image; the latter is the random algorithm and several different raROIs are generated for the same image even if there is only one random algorithm.

We then considered one factor at a time, for example, different agents applied to the same image and using the metrics S_p and S_s, we generated three different sets of similarity values or treatments: treatment $H = \{S_p(h_i ROIs, h_j ROIs)_I : \forall I, i \neq j\}$, the similarities among different subjects looking at the same picture; treatment $A = \{S_p(aROIs, hROIs)_I : \forall I\}$, the similarities of algorithms and eye movements for the same image; treatment $Ra = \{S_p(RaROIs, hROIs)_I : \forall I\}$, the similarities of the random algorithm and the eye movements for the same image. Put in statistical terms, we have the same *condition*, ROIs defined within the image coordinate space, and three *treatments* that correspond to three different ways to generate and combine these ROIs: Each treatment is quantifiable by S_p (or S_s), which is usually referred to as the *response variable*.

The number in the box L of the parsing diagram, Fig. 7, shows the mean and the variance of second treatment A, which is greater than the mean of Ra, bottom box. This is actually our main criterion (mentioned several times through the paper): It demonstrates that IPAs cohere with humans better than the random algorithm does.

The Anova (usually capitalized, originally put forward by Fisher [7]) is finally applied to further validate whether or not the different experimental treatment means $\mu(H)$,

$\mu(A)$, $\mu(Ra)$, are different enough (compared to the variability within the individual treatments) for us to conclude that they correspond to three different populations. In other words, can we conclude that, based on those means, the same statistical differences generated in our experiments hold for the hypothetical infinite population of all images and viewers? Of course, the Anova test can be applied to any number of treatments; in the case of only two treatments, the Anova corresponds to the Student t-test.

The Anova value is usually compared to a critical value F of the Fisher distribution for a certain level of significance. If the Anova test value is less than the F-Fisher critical value for an α level of significance (0.01 in this paper), then it is possible to infer that the means are not different enough to come from different populations.

Our quantitative conclusions, presented in the results section, and our claims of statistical significance were strongly sustained by the relationship between significant Anova test values and F-Fisher critical value. For each factor, we always applied the Anova to the three treatments, H, A, and Ra, to verify any statistical difference. If so, we then applied the Anova test in pairwise fashion to verify that this difference is due to a difference between treatment Ra and treatment A and not to one between treatment H and treatment A. For example, in Fig. 7, S_p, Local box, Ra, and A have an Anova value equal to 18.7, which is greater than the F-Fisher critical value of 7.5: This means that A and Ra are significantly different. The same procedure can be repeated for different factors by narrowing and/or changing the three treatments: For example, only a specific group of algorithms can be taken into account in H, A (see $A*$ in Fig. 7 and Section 7.1).

Acknowledgments

The authors would like to thank the following sponsors for their support: NASA-Ames Research Center (Stephen Ellis), Fujita Research (Ken Kawamura), and Neuroptics Corporations (Kamran Siminou), our colleagues in the laboratory, Michela Azzariti for her indispensable statistical analysis; Ted Blackmon, Yeuk Fai Ho, Veit Hagenmeyer, and Yong Yu. Others have been generous in their advice—Irwin Sobel at Hewlett-Packard, and, at UCB, Jerry Feldman, computer science, and David Brillinger, statistics; also the anonymous referees for their incisive, constructive, and motivating comments.

References

[1] T.A. Bahill and L.W. Stark, "Trajectories of Saccadic Eye Movements," *Scientific Am.*, vol. 240, pp. 84-93, 1979.

[2] S. Brandt and L.W. Stark, "Spontaneous Eye Movements during Visual Imagery Reflect the Content of the Visual Scene," *J. Cognitive Neuroscience,* vol. 9, no. 1, pp. 27-38, 1997.

[3] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, no. 6, pp. 679-698, 1986.

[4] I. Daubechies, *Ten Lectures on Wavelets.* Philadelphia: SIAM, 1992.

[5] C. De Vleeschouwer, X. Marichal, T. Delmot, and B. Macq, "A Fuzzy Logic System Able to Detect Interesting Areas of a Video Sequence," *Proc. SPIE,* vol. 3,016, pp. 234-245, 1997.

[6] V. Di Gesú and C. Valenti, "The Discrete Symmetry Transform in Computer Vision," Technical Report 011-95, Laboratory for Computer Science (DMA), Univ. of Palermo, 1995.

[7] L. Fisher, *Fixed Effects Analysis of Variance.* New York: Academic Press, 1978.

[8] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1,254-1,259, Nov. 1998.

[9] P.J. Locher and C.F. Nodine, "Symmetry Catches the Eye," *Eye Movements: From Physiology to Cognition,* pp. 353-361, 1987.

[10] S.K. Mannan, K.H. Ruddock, and D.S. Wooding, "The Relationship between the Locations of Spatial Features and Those of Fixations Made during Visual Examination of Briefly Presented Images," *Spatial Vision,* vol. 10, no. 3, pp. 165-188, 1996.

[11] E. Niebur and C. Koch, "Control of Selective Visual Attention: Modeling the 'Where' Pathway," *Advances in Neural Information Processing Systems,* D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, eds., vol. 8, pp. 802-808, MIT Press, 1996.

[12] D. Noton and L.W. Stark, "Scanpaths in Eye Movements During Pattern Perception," *Science,* vol. 17, no. 1, pp. 308-311, 1971.

[13] W. Osberger and A.J. Maeder, "Automatic Identification of Perceptually Important Regions in an Image," *Proc. Int'l Conf. Pattern Recognition,* pp. 17-20, 1998.

[14] R. Plamondon and C.M. Privitera, "The Segmentation of Cursive Handwriting: An Approach Based on Off-Line Recovery of the Motor-Temporal Information," *IEEE Trans. Image Processing,* vol. 8, no. 1, pp. 80-91, 1999.

[15] C.M. Privitera, M. Azzariti, and L.W Stark, "Locating Regions-of-Interest for the Mars Rover," *Int'l J. Remote Sensing,* 2000. to appear.

[16] C.M. Privitera and L.W. Stark, "Evaluating Image Processing Algorithms That Predict Regions of Interest," *Pattern Recognition Letters,* vol. 19, no. 1, pp. 1,037-1,043, 1998.

[17] C.M. Privitera and L.W. Stark, "Focused JPEG Encoding Based upon Automatic Preidentified Regions of Interest," *Proc. SPIE,* vol. 3,644, pp. 552-558, 1999.

[18] D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Context-Free Attentional Operators: The Generalized Symmetry Transform," *Int'l J. Computer Vision,* vol. 14, pp. 119-130, 1995.

[19] W. Richards and L. Kaufman, "Centre-of-Gravity Tendencies for Fixations and Flow Patterns," *Perception and Psychology,* vol. 5, pp. 81-84, 1969.

[20] O. Rioul and P. Duhamel, "Fast Algorithms for Discrete and Continuous Wavelet Transforms," *IEEE Trans. Information Theory,* vol. 38, no. 2, pp. 569-586, 1992.

[21] L.W. Stark and Y. Choi, "Experimental Metaphysics: The Scanpath as an Epistemological Mechanism," *Visual Attention and Cognition,* W.H. Zangemeister, H.S. Stiehl, and C. Freksa, eds., pp. 3-69, 1996.

[22] L.W. Stark and C.M. Privitera, "Top-Down and Bottom-Up Image Processing," *Proc. of IEEE Int'l Conf. Neural Networks,* vol. 4, pp. 2,294-2,299, June 1997.

[23] L.W. Stark, C.M. Privitera, H. Yang, M. Azzariti, Y.F. Ho, A. Chan, C. Krischer, and A. Weinberger, "Scanpath Memory Binding: Multiple Read-Out Experiments," *Proc. SPIE,* vol. 3,644, pp. 495-510, 1999.

[24] P.P. Vaidyanathan, *Multirate System and Filter Banks.* Englewood Cliffs, N.J.: Prentice Hall, 1993.

[25] A.L. Yarbus, *Eye Movements and Vision.* New York: Plenum Press, 1967.

[26] W.H. Zangemeister, K. Sherman, and L.W. Stark, "Evidence for Global Scanpath Strategy in Viewing Abstract Compared with Realistic Images," *Neuropsychologia,* vol. 33, no. 8, pp. 1,009-1,025, 1995.

[27] J. Zhao, Y. Shimazu, K. Ohta, R. Hayasaka, and Y. Matsushita, "A JPEG Codec Adaptive to the Relative Importance of Regions in an Image," *IEEE Trans. Information Processing Soc. Japan,* vol. 38, no. 8, pp. 1,531-1,542, 1997.

Claudio M. Privitera received the Laurea degree in computer science from the University of Pisa, Italy, in 1991. From 1992 to 1995, he held a doctoral fellowship within the National Research Program on Bioelectronic Engineering working at DIST, Department of Informatics, Systems, and Telecommunications of the University of Genoa, Italy. In 1994, he visited Laboratorie *Scribens*, École Polytechnique, Université de Montréal, Canada. From 1995 to 1996, he was a postdoctoral research fellow at ICSI, the International Computer Science Institute of Berkeley working in the Artificial Intelligence Group. In 1996, he joined the Neurology and Telerobotics Units of the University of California at Berkeley, and he has been teaching there in the Mechanical Engineering Department. His research interests cover several aspects of biological and computational vision, image processing, and pattern recognition. His research interests are also in the area of neural computation in motor control, robotics, and artificial intelligence.

Lawrence W. Stark received the AB degree from Columbia University in 1945, the MD degree from Albany in 1948, the ScD h.c. from the State University of New York in 1988, and the PhD, h.c. from Tokushima, 1992. He has been a professor at the University of California, Berkeley since 1968, where he divides his teaching efforts among the Electrical Engineering and Computer Science and Mechanical Engineering Departments in engineering and the Physiological Optics and Neurology Units in biology and medicine. His research interests are centered in bioengineering, with emphasis on human brain control of movement and vision, and symbiotic interactions of this knowledge with the rapidly developing fields of robotic vision and control. He pioneered the application of control and information theory to neurological systems and the use of online digital computers for real-time acquisition of data, for self-organizing pattern recognition, and for physiological modeling. He has published several books and numerous research papers during his career that also included faculty positions at Yale (1954-1960), MIT (1960-1965), and Illinois (1965-1968). His many former graduate students and postdoctoral fellows form a world-wide school of bioengineering and biocybernetics. He is a fellow of the IEEE.