Combining Conspicuity Maps for hROIs Prediction

Claudio M. Privitera[1], Orazio Gallo[2], Giorgio Grimoldi[2],
Toyomi Fujita[1], and Lawrence W. Stark[1]

[1] Neurology and Telerobotics Units, Optometry School,
University of California, Berkeley, 94720, CA
[2] Department of Bioengineering of Politecnico di Milano,
P.zza Leonardo da Vinci, 20133 Milan

Abstract. Bottom-up cortical representations of visual conspicuity interact with top-down internal cognitive models of the external world to control eye movements, EMs, and the closely linked attention-shift mechanisms; to thus achieve visual recognition. Conspicuity operators implemented with image processing algorithms, IPAs, can discriminate human Regions-of-Interest, hROIs, the loci of eye fixations, from the rest of the visual stimulus that is not visited during the EM process. This discrimination generates predictability of the hROIs. Further, a combination of IPA-generated conspicuity maps can be used to achieve improved performance over each of the individual composing maps in terms of hROI predictions.

1 Introduction

Saccadic eye movements and the closely linked mechanisms of visual attention shifts have been attracting scientists for decades. These are very complex functions which utilize many different aspects of the human nervous system from early visual areas and brain stem motor competencies up to high level semantic elaborations. The classical work performed by Yarbus back in the sixties on task- and intention-dependent eye movements, EMs, [1], is still acknowledged and very influential.

Several models have been presented in the neuroscience and computational vision community. In general, these models seem to support theories by which selective attention shifts are strictly related to, or even controlled by a visual multi-feature conspicuity cortical representation; see, for example the pioneering work of Koch and Ullman, [2] and more recently an interesting review by Itti and Koch, [3] and the approach of Parkhurst et. al [4]. Visual attention, which is linked to EMs, is thought to be a process guided by this map (likely located in the frontal eye field, [5], although neurophysiological data support the existence of several conspicuity maps in different cortical areas such as temporal and pre-frontal cortex, [6], [7]). Such a map is also thought to preserve the topography of the external viewed world and measures, in the z-coordinate, levels of conspicuity or saliency; the highest points or local maxima of the map define the loci (here referred to as hROIs, human Regions-of-Interest) of sequential steps or saccades that are governed by a sequential winner-take-all procedure [2], [3].

The nature and number of cortical topographical conspicuity representations vary; in general they must be related to bottom-up early visual conspicuity features such as

L. Paletta et al. (Eds.): WAPCV 2004, LNCS 3368, pp. 71–82, 2005.

center-surrounded operators, high-contrast contours and areas, and color and selective orientations [8], [9], [3]. The work by Reinagel and Zador, [10] for example, shows how subjects look at hROIs that have high spatial contrast and poor pair-wise pixel correlation; a definition of internal intensity non-uniformity. Similarly, higher-order statistics have recently been used by Krieger et al. [11] to characterize hROIs and to propose features like corners, occlusions, and curved lines as salient detectors. Different representations and different visual scales can be processed at different sequential temporal stages during active looking, [12] or in parallel and then weighted and combined into a final *supra-dimensional* or *supra-feature* conspicuity map [13].

Cognition must influence these conspicuity representations in a top-down manner, [14], however, how this influence is exercised and integrated in the overall control of EMs/attentional shifts is still not well understood and open to different interpretations. Interesting models wherein top-down and bottom-up processing are integrated into Bayesian or Markovian frameworks can also be found in the literature [15], [16], [17].

Our standpoint is strongly based on the scanpath theory put forward by Noton and Stark, [18], [19]. The scanpath was defined on the basis of experimental findings. It consists of sequence of alternating saccades and fixations that repeat themselves when a subject is viewing a picture. Only ten percent of the duration of the scanpaths 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 ninety percent of the total viewing period. Scanpath sequences appear spontaneously without special instructions to subjects and were discovered to be repetitive, [18], [19].

The scanpath theory proposes that an internal spatial cognitive model controls both perception and the active looking EMs of the scanpaths sequence, [18], [19], and further evidence for this came from new quantitative methods, experiments with ambiguous figures, [20] and more recently from experiments on visual imagery [21], [22] and from MRI studies on cooperating human subjects [23].

A top-down internal cognitive model of the external world must necessarily control not only our recognition but also the sequence of EM jumps which direct the high-resolution and centrally located fovea into a sequence of human regions-of-interest, hROI loci. We usually refer to these hROI loci as *informative* regions because they carry the information needed (and thus searched) by our cognition to validate or propose new internal cognitive models during active looking EM scanpaths, [24]. Recognition is re-cognition, which signifies to know again, to recall to mind a predefined model; eye movements and cognition must act in synergy.

Conspicuity and informativeness are thus interconnected, and conspicuity can be used to predict informativeness. Our studies have indeed demonstrated that bottom-up image processing algorithms, IPAs, can be successfully used, together with an eccentricity clustering procedure to generate algorithmic Regions-of-Interest, aROIs. These aROIs are able to predict the hROI loci of human observers, obtained during general viewing EM experiments, [25]. The level of the prediction, meaning the proximity of an IPA-generated sequence of aROI loci and an experimental instance of an hROIs sequence, or scanpath, can be measured by a spatial metric, S_p, which has been defined on the basis

of physiological and experimental observations. The metric S_p is important in many applications, [26].

Different IPAs yield different results depending on the class of images and the specific task involved, [27]. Also, as discussed above, different IPAs can be combined together into a *supra-feature* IPA. The resulting combination might be used to achieve improved performance over each of the individual IPAs in terms of hROIs S_p-prediction and in terms of consistency for larger numbers of images. A study specifically designed for the geological/planetary exploration application, [28], has already shown interesting preliminary results.

In the present paper we want to elucidate the important relationship between image conspicuity distribution and experimental human scanpath hROIs, for different image classes and for more general viewing conditions; the S_p metric will be utilized for this purpose. Heterogeneous sets of bottom-up features can be combined with ad-hoc selection and fusion processes. The results will confirm that such a combination can perform better than using only a single feature.

2 Sp Similarity Metric and Saliency Discrimination

2.1 Experimental hROI Database

Standard procedures were used to extract hROIs, the loci of eye fixations, ([22], circles superimposed on the raw eye movement data, Figure 1, upper left panel); an average of seven hROIs was collected per image, which usually corresponds to about three seconds of viewing. Different classes of images for a total of forty images were used; they were photos of *Interior*, i.e., environments like a theater, *Traffic*, i.e., road intersection in a big city, *Natural*, i.e., a canyon or a lake, *Geological*, i.e., rocks or terrain, *Ambient*, i.e., a kitchen or a bedroom, *Urban*, i.e., snapshots of city life.

Data were from new EM experiments and also provided by the Neurology and Telerobotics Units archive. The eye tracker system used in the Units resides in a Pentium3 two-processor computer which controls both visual stimuli presentation and the video camera analysis. The first corneal Purkinje reflection from an infrared light source is tracked in real time using a flying-spot algorithm and then later processed by a parsing algorithm for the identification of hROIs. For each image, two calibration sessions composed by a sequence of 3×3 fixation points are used to map video camera eye positions into the stimulus linear space. The experiment is repeated if calibration drift or similar deleterious experimental artifacts are detected during or between the two calibrations. The subjects were secured in a head-rest optometric apparatus and no specific viewing instructions were provided.

2.2 Algorithmic Identification of Regions-of-Interest, aROIs

Several sources supplied the collection of IPAs; they were in general inspired by early-stage neurophysiology, such as center-surround filters, localized projectors, or concentric detectors; other IPAs were simply based on intuition or different approaches presented in the literature. Sixteen IPAs were extracted from the collection and used in this study. They were all defined on a local support size corresponding to two degrees of the visual

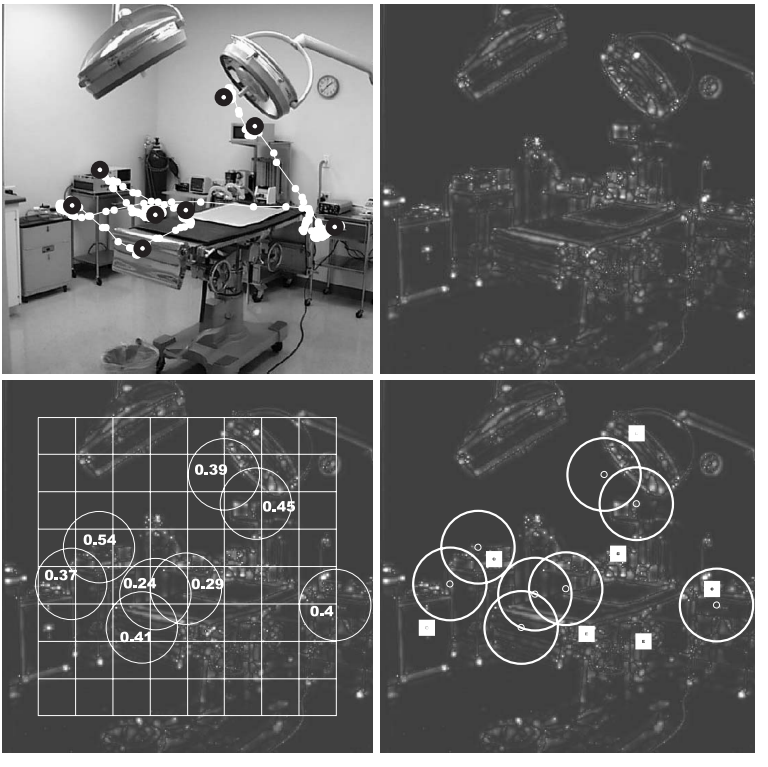

Fig. 1. Loci of eye fixations, hROIs (black small circles, upper-left panel), are extracted from raw eye movement data (white trace, upper-left panel) using standard procedures. An energy map is created using a symmetry operator (see text, upper-right panel). Conspicuity is measured for the hROIs (large circles, bottom-left panel, peak of energy values are also reported) and for the rest of the image (grid, bottom-left panel). The energy map is also used to identify algorithmic Regions-of-Interest, aROIs (squares, bottom-right panel) whose loci can finally be compared with hROIs

field, similar to the subtended angle of the fovea in our experimental setup. Algorithms A_1 to A_4 are four different Gabor filters rotated by 0, 45, 90, 135 degrees, [29]. Algorithm A_5 was a difference in the gray-level orientation operator, [30]. We then have A_6, local entropy, [25], A_7, Laplacian of a Gaussian, [31], A_8, contrast operator, [25], A_9, x-like operator, [25], A_{10}, radial symmetry, [32], A_{11}, quasi receptive, [25], $A_{12} - A_{14}$, frequency analysis, low, medium, high pass filters based on DCT, [25], $A_{15} - A_{16}$, color opponence operators, red-green and yellow-blue, [30].

Each IPA gives emphasis to a different bottom-up visual conspicuity when applied to an image and correspondingly defines its own energy map. A symmetry operator, for example, highlights symmetry, and the resulting output energy map (Figure 1, upper right panel) defines both the level of this feature (the z-dimension is coded in grey-level, where brighter patches represent centers of high symmetricity) and its spatial distribution

over the image. Local maxima in the output energy map define peaks of conspicuity and the corresponding loci are possible candidates to be selected as the algorithmic Regions of Interest, aROIs.

An eccentricity clustering procedure was introduced in [25] to select eccentric-located local maxima as the final aROI loci; the algorithm was based on an incremental merging of neighboring local maxima, and the highest peak locus was finally held for each local neighborhood. There is probably little biological plausibility for the mechanism but important practical convenience; the selection of aROI maxima is a parameter-free procedure independent of scale and image size. The eccentricity is necessary to maximize the extent of the image that is covered with a limited number of hROIs, EM jumps. The winner-take-all algorithm utilized by many authors (see for example [3]) in conjunction with an inhibition-of-return (or *forgetting function*) mechanism, [33] has more neuronal justifications; of course, the spatial size of the inhibition needs to be properly set in this case. We decided on the same size defined above for the IPA local support and corresponding to two degrees of the visual field. In the winner-take-all algorithm absolute local maxima are iteratively selected as aROIs; for each absolute maximum, the forgetting function zeroes all the energy around it and the aROI selection process continues with the next absolute maximum. The top seven aROIs were retained for each image. How these two selection procedures affect the final S_p outcome is still under investigation; preliminary results, however, seem to indicate that the final distribution of aROIs is independent of the selection procedure (see also [34]).

2.3 Sp Similarity Index

Each image and experimental instance is thus represented by a pair of vectors; the vector of hROIs (Figure 1, lower left panel, circles) is the original scanpath and is experimentally defined for each single subject and experimental trial. The vector of aROIs (Figure 1, lower right panel, squares) is the artificial scanpath and is computed from the IPAs as just described. The loci similarity between hROIs and aROIs (but also between two hROIs or two aROIs) is expressed by a spatial similarity index, S_p, [25] which represents the percent of experimental hROIs which are in close proximity to at least one aROI. An S_p value of 1 means that all hROIs are captured within the aROIs; zero means a complete spatial dissociation between the two vectors. The threshold distance defining this proximity is approximately 2 degrees (Figure 1, lower two panels, big circumferences) and is based on experimental intra-subject EM observation (for a full review see discussions on the EM Repetitive similarity in [25]); for example, two aROIs are in two (out of eight) different hROI circumferences (Figure 1, lower right panel), which yields in this case an S_p value equal to $2/8 = 0.25$.

Randomly generated aROIs can fall by chance within an hROI; this explain why the S_p similarity of such randomly generated aROIs with human scanpath, averaged for all images and repetitions, is 0.2, what we considered a bottom anchor or the lowest fiduciary point for S_p. The opposite (top) anchor or highest fiduciary point is the inter-subject S_p similarity whose average is about 0.6; this is the average of the scanpath S_p similarity indices when different subjects look at the same picture (then averaged for all pictures). A value of 0.6 says that an average of 60 percent of their hROIs cohere

which indicates that viewers were fairly consistent in identifying regions of interest. This is an important results for this study, in fact, algorithms cannot be expected to predict human fixations better than the coherence among fixations of different persons.

3 Saliency Discrimination of Eye Fixation Loci

In human scanpaths, only a small portion of the entire image falls on the fovea, the high-resolution area of the retina, to be sequentially attended and analyzed by the brain; the rest of the image is processed only at much lower resolution by peripheral vision and is ignored by the eye fixation sequence. Subjects tend to repeat their scanpaths, and also cohere with other subjects in the loci of fixations, when they look at the same image (see above, [25] and the recent review in [22]).

This visual image spatial discrimination, which is implicit in the EM process, can be also studied in terms of the conspicuity distribution of a given IPA operator; how different is the distribution within the hROIs compared with the rest of the image that was not selected by the experimental scanpath? A measure that can be carried out by the D-prime index from the theory of signal detection:

$$D' = \frac{\mu_{max(hROIs)} - \mu_{max(NhROIs)}}{\sqrt{\sigma^2_{max(hROIs)} + \sigma^2_{max(NhROIs)}}} \tag{1}$$

In the D-prime definition above, $max(hROIs)$ is the peak of the energy in each hROI circumference (Figure 1, lower-left panel). The rest of the image is divided in a regular grid of 2×2 degree NhROI (an acronym for Non-hROI) blocks and for each of these blocks, the corresponding energy peak is also retained and used in the equation ($max(NhROIs)$). The contour of the image is very unlikely to be fixated during human scanpath and is thus not considered in the D-prime definition (Figure 1, lower-left panel). A positive value for D-prime indicates that the IPA energy (or more specifically the peak energies) inside the hROIs is on the average greater than that for the rest of the image – a necessary condition for a given IPA to be chosen as a conspicuity model for that specific hROI set.

Values of D-prime and S_p are indeed well correlated, (Figure 2, a correlation of 0.73); each $< S_p, D-prime >$ data point corresponds to a unique ternary: one hROIs instance, one image and an IPA. High values of D-prime conspicuity discrimination correspond to optimal scanpath S_p predictability; values of S_p that are low or below the random bottom anchor level are associated with poor or even negative discriminability. This indicates a spatial mismatch of the energy distribution with the hROIs portion of the image having less energy than the outside portion. Positive values of D-prime correspond to S_p similarities that are above the random bottom anchor $S_p = 0.2$. Note that most of the $< S_p, D-prime >$ curve (Figure 2) lays on the second part of the plot, above zero, which indicates that a positive discrimination is present in most of the experimental hROIs instances.

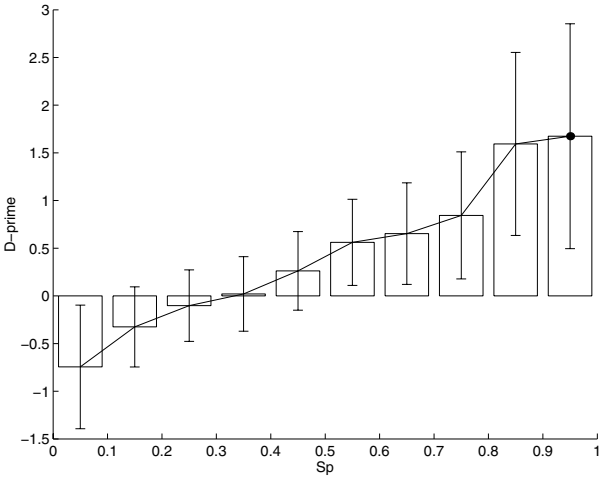

Fig. 2. Correlation between D-prime conspicuity discriminability of hROIs and the corresponding S_p hROI-aROI loci similarity. Positive values of D-prime correspond to S_p similarities that are above the random bottom fiduciary level, $S_p = 0.2$

4 Polynomial Combination of Energy Maps

Different energy maps can be weighted and combined together in a polynomial manner into a *supra-feature* map which is hereafter referred to as A* (Figure 3). We want to derive A* and verify that it can in fact improve the spatial S_p matching of aROIs with the experimental hROIs.

4.1 Optimization for Each Single Experimental hROIs Instance

We initially tried to define the best combination for each single image and hROIs; the objective was to determine the vector of weights, $\overline{\alpha}$, which maximizes the corresponding S_p for A*, a polynomial of the input energy maps $A_1, A_2, ..., A_{16}$ (Figure 3). An optimization problem that we implemented in Matlab with the built-in medium scale algorithm of the constrained minimization routine (based on the Sequential Quadratic Programming method, [35]).

Although S_p is the index to be maximized, D-prime is a more suitable objective function for the optimization; it is a continuous n-dimensional function, appropriate for a canonical gradient-descent searching process in the $\overline{\alpha}$ space. The metric S_p is discontinuous; an infinitesimal variation in the input space $\overline{\alpha}$ can cause a different energy maximum (aROI) to be selected in the A* map (as a consequence of the iterative absolute maximum selection procedure defined in the previous Section), which can result in a substantial change of S_p. Recall that S_p and D-prime are correlated (as shown in Figure 2). The object function is thus based on D-prime and defined as follow:

$$\max_{\overline{\alpha}} \{D'(A^*), \alpha_i > 0, \sum_i \alpha_i = 1\} \tag{2}$$

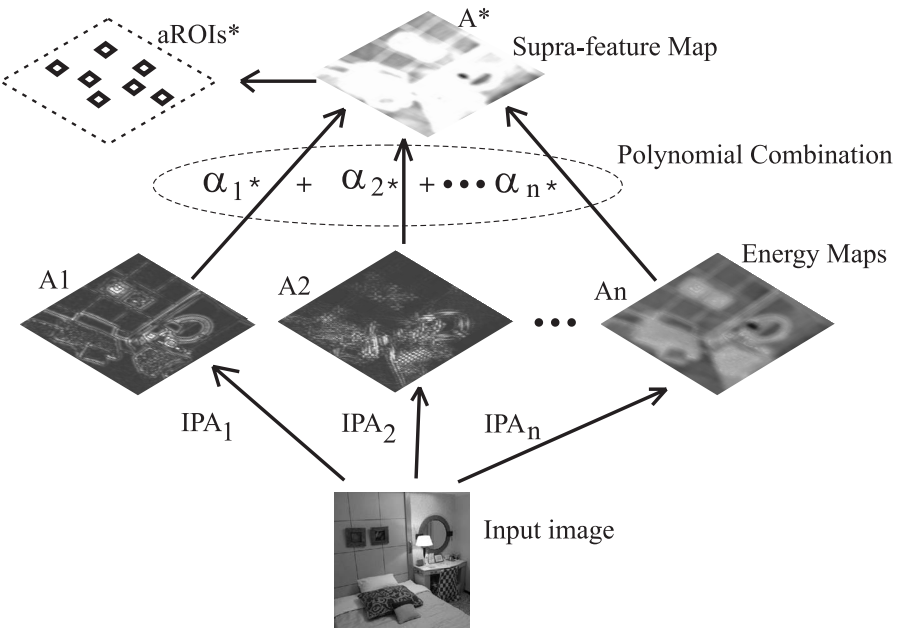

Fig. 3. A polynomial (weighted) combination of different conspicuity maps defines a *supra-dimension* conspicuity map A*

An image from the Ambient image can exemplify the optimization process (Figure 4); all IPAs were initially applied to the image and the best result was achieved with A_{15} (Figure 4, left) which yielded an S_p value of $5/7 = 0.71$. The aROIs* from the A* map (Figure 4, right) defined by the optimal set of weights $\overline{\alpha}$* as resulted at the end of the optimization process, have an S_p value equal to 1. Each experimental hROIs data was used in a similar optimization process and the resulting combinatorial A* used to select aROIs* which could be finally compared with the corresponding hROIs. In general, more than 80% of the optimizations resulted in an A*-S_p that was higher or at least equal to the best of the single IPAs. The residual 20% of unsuccessful optimizations might be related to local maxima (not easily overcome in such a complex multi-dimensional search space) and to the not unitary correlation between the objective function D-prime and S_p.

4.2 Optimization for Image Classes

These single optimizations show that it is in general possible to define an improved polynomial combination of energy maps whose matching with experimental hROIs is superior to each of the composing single maps.

A more fundamental problem is to define a super polynomial combination A** which might be applied more consistently to all the images (and hROIs) of the same class. To define the corresponding vector of weights, $\overline{\alpha}$**, we tried initially to review the definition of the objective function in order to contemplate a multi-dimensional D-prime index

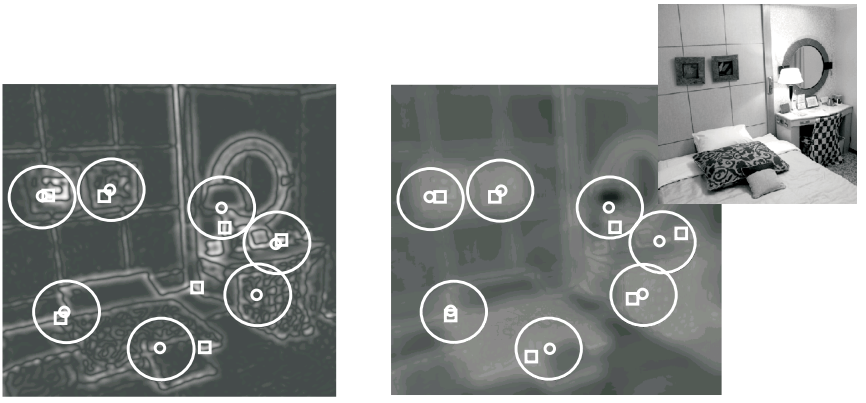

Fig. 4. Example of result for a single image/hROIs instance optimization; the best algorithm (algorithm A_15, left panel) compared with the final polynomial combination A* (right panel). Circles identify hROIs, squares aROIs; the A* S_p index is 7/7 vs. the 5/7 for the best algorithm

Table 1. The A** optimization for each single image class. The average S_p value of A** (right column) is always superior to the mean of all IPAs (middle column) and to the best IPA of each class (left column, the index (1 to 15) of the IPA is also reported in parentheses)

Class	Best IPA	Mean of IPAs	A**
Interior	(5) 0.54 (0.04)	0.45 (0.05)	0.62 (0.04)
Traffic	(9) 0.44 (0.09)	0.36 (0.05)	0.45 (0.04)
Natural	(6) 0.54 (0.07)	0.44 (0.06)	0.58 (0.05)
Geological	(4) 0.55 (0.06)	0.46 (0.06)	0.63 (0.06)
Ambient	(6) 0.6 (0.03)	0.49 (0.04)	0.64 (0.04)
Urban	(9) 0.57 (0.05)	0.39 (0.04)	0.59 (0.02)

for all the images in the data set and then to run the same optimization process; the computational repercussions were however too costly and thus impracticable. We then tried to use the same vectors, $\overline{\alpha}$*, already calculated from all the single optimizations. All classes of images were considered separately: for each class, all the vectors $\overline{\alpha}$* were clustered using a K-means algorithm and the center of the densest was cluster retained as the final $\overline{\alpha}$**. This center in the multi-feature $\overline{\alpha}$ space likely represents the most common characteristics of that class and it is reasonable to select it as the corresponding representative. Other clustering possibilities are under investigation and evaluated with a clustering analysis procedure (such as the Akaike's criterion) that can serve to explicate the significance of each $\overline{\alpha}$**.

The results seem to support the initial hypothesis; the S_p behavior of the polynomial algorithm A**, averaged for all images and hROIs within a class, is consistently superior (see S_p value, table 1, right column) if compared with the mean of all composing IPAs (table 1, middle column) and with the mean of the best IPA for each class (table 1, left

column). Almost all the A** (except for the Traffic class) yield an average S_p (table 1, right column) which is very close to the inter-subject S_p top anchor similarity of 0.6 discussed in the previous Section.

The best algorithm (table 1, number in parenthesis, left column) is different for each image class; also different classes generated different distributions of weights, $\overline{\alpha}$**, (Figure 5, the values in the ordinate indicate the weight given to the corresponding IPA, in that specific class) further supporting the notion that different image types might be characterized by different types of conspicuity.

5 Conclusions

Different conclusions emerge from this study. First, experimental EM fixations, which are distributed over only a specific portion of the image, are discriminable in term of conspicuity and D-prime. Fifteen different IPAs were used to represent different type of conspicuities. More importantly, the level of discriminability correlates with the capability of a particular IPA to be used in an aROIs generation process, to match the human hROIs scanpath as expressed by the similarity index S_p. Secondly, a polynomial combination of IPAs yields to a better S_p prediction of experimental hROIs. This was demonstrated for most of the single experimental instances and for the complete set of images within a class. A procedure was proposed to define for each class, the vector of weights $\overline{\alpha}$** which defines the influence of each composing IPA in the *supra-feature* map A**.

As mentioned above, it has been known for some time that the implicit or explicit task-setting in which the subject is immersed can strongly modify the scanpath. Thus as a subject continually looks at the scene she may change her point of view, think

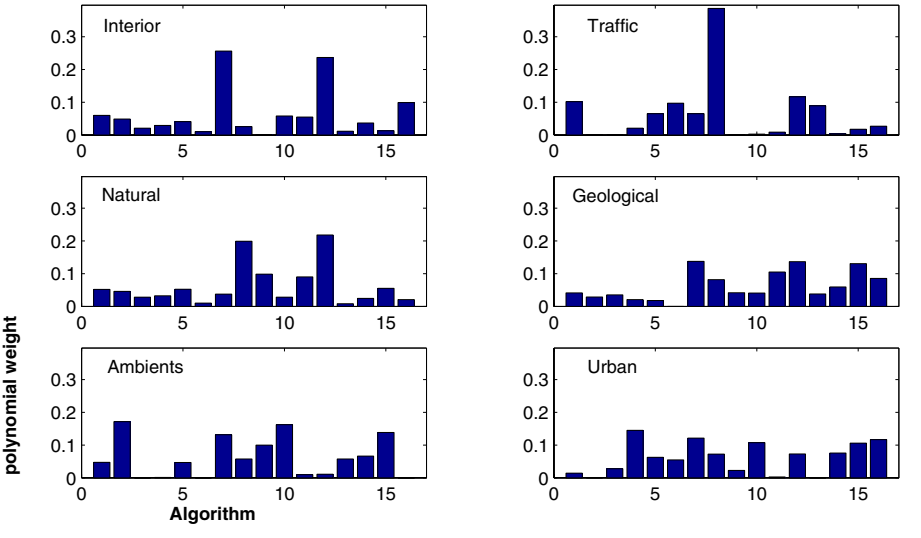

Fig. 5. Six different image classes yielded six different distributions of $\overline{\alpha}$** weights (ordinate) which define A**; algorithm index is in the abscissa

of different task-goals and modify the scanpath [1] (a planetary exploration task was recently investigated by Privitera and Stark, [28]). The vector of weights $\overline{\alpha}^{**}$ can thus be susceptible to the task-setting.

A discrimination of hROIs in term of visual conspicuity is experimentally plausible in human vision and analyzed in the literature from different perspectives; we propose an adaptive and composite nature for conspicuity which can vary for different image classes and probably tasks and applications. Cognition might modulate the combination of different types of conspicuity as a function of the internal TD hypothesis or the visual expectation: different environmental setting, for example Urban vs. Natural, can be associated to different clusters of conspicuity weights, $\overline{\alpha}^{**}$, that can be then utilized in the generation of the scanpath motor sequence.

Acknowledgements

We wish to thank Profs. Giancarlo Ferrigno, Alessandra Pedrocchi and Antonio Pedotti, all from the Department of Bioengineering of Politecnico di Milano, for their support and stimulating discussions. Our colleague in the laboratory, Dr. Yeuk Ho, was indispensable for his management of the laboratory computational resources and the eye-tracker system. A special thanks goes also to Dr. Tina Choi for her editorial assistance.

References

1. Yarbus, A.: Eye movements and vision. Plenum, New York (1967)
2. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuity . Human Neurobiology **4** (1985) 219–227
3. Itti, L., Koch, C.: Computational modelling of visual attention. Nature Neuroscience Reviews **2** (2001) 194–203
4. Parkhurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vision Research **42** (2002) 107–123
5. Bichot, N.P., Schall, J.D.: Saccade target selection in macaque during feature and conjunction visual search. Visual Neuroscience **16** (1999) 81–89
6. Stein, J.E.: The representation of egocentric space in the posterior parietal cortex. Behavioral and Brain Sciences **15** (1992) 691–700
7. Nothdurft, H.: Salience from feature contrast: additivity across dimensions. Vision Researh **40** (2000) 1183–1201
8. Engel, F.L.: Visual conspicuity, visual search and fixations tendencies of the eye. Vision Research **17** (1977) 95–108
9. Findlay, J.M., Walker, R.: A model of saccade generation based on parallel processing and competitive inhibition. Behavioral and Brain Sciences **22** (1999) 661–721
10. Reinagel, P., Zador, A.M.: Natural scene statistic at the center of gaze. Network: Comput. Neural Syst. **10** (1999) 341–350
11. Krieger, G., Rentschler, I., Hauske, G., Schill, K., Zetzsche, C.: Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. Spatial Vision **13** (2000) 201–214
12. Neri, P., Heeger, D.: Spatiotemporal mechanisms for detecting and identifying image features in human vision . Nature neuroscience **5** (2002) 812 – 816

13. Muller, H.J., Heller, D., Ziegler, J.: Visual search for singleton feature targets within and across feature dimensions. Perception and Psychophysics **57** (1995) 1–17
14. Henderson, J.M., Hollingworth, A.: in Eye guidance in reading and scene perception. In: Eye movements during scene viewing: an overview. North-Holland/Elsevier (1998)
15. Chernyak, D.A., Stark, L.W.: Top-down guided eye movements. IEEE Trans. on SMC, part B **31** (2001) 514–522
16. Rimey, R.D., Brown, C.M.: Controlling eye movements with hidden markov models. International Journal of Computer Vision **7** (1991) 47–65
17. Schill, K., Umkehrer, E., Beinlich, S., Krieger, G., Zetzsche, C.: Scene analysis with saccadic eye movements: top-downand bottom-up modeling. Journal of Electronic Imaging **10** (2001) 152–160
18. Noton, D., Stark, L.W.: Eye Movements and visual Perception. Scientific American **224** (1971) 34–43
19. Noton, D., Stark, L.W.: Scanpaths in Eye Movements during Pattern Perception. Science **171** (1971) 308–311
20. Stark, L.W., Ellis, S.R.: in Eye Movement: Cognition and Visual Perception. In: Scanpaths revised: cognitive models direct active looking. Lawrence Erlbaum Associates, Hillside, NJ (1981) 193–226
21. Brandt, S.A., Stark, L.W.: Spontaneous eye movements during visual imagery reflect the content of the visual scene. J. Cognitive Neuroscience **9** (1997) 27–38
22. Stark, L.W., Privitera, C.M., Yang, H., Azzariti, M., Ho, Y.F., Blackmon, T., Chernyak, D.: Representation of human vision in the brain: How does human perception recognize images? Journal of Electronic Imaging **10** (2001) 123–151
23. Kosslyn, S.: Image and Brain: The Resolution of the Imagery Debate. MIT Perss, Cambridge, MA (1994)
24. Stark, L.W., Choi, Y.S.: in Visual Attention and Cognition. In: Experimental Metaphysics: The Scanpath as an Epistemological Mechanism. Elsevier Science B.V. (1996) 3–69
25. Privitera, C.M., Stark, L.W.: Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations. IEEE Trans. PAMI **22** (2000) 970–982
26. Privitera, C.M., Stark, L.W., Ho, Y.F., Weinberger, A., Azzariti, M., Siminou, K.: Vision theory guiding web communication. In: Proc. SPIE - Invited paper. Volume 4311., San Jose, CA (2001) 53–62
27. Privitera, C.M., Azzariti, M., Stark, L.W.: Locating regions-of-interest for the Mars Rover expedition. International Journal of Remote Sensing **21** (2000) 3327–3347
28. Privitera, C.M., Stark, L.W.: Human-vision-based selection of image processing algorithms for planetary exploration . IEEE Trans. Image Processing **12** (2003) 917–923
29. Daugman, J.G.: Two-dimensional spectral analysis of cortical receptive field profiles. Vision Research **20** (1980) 847–856
30. Itti, L., Kock, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. PAMI **20** (1998) 1254–1259
31. Marr, D., Hildreth, E.: Theory of edge detection. Proc. Roy. Soc. London **B207** (198) 187–217
32. Loy, G., Zelinsky, A.: Fast radial symmetry for detecting points of interest. IEEE Trans. PAMI **25** (2003) 959–973
33. Klen, R.M.: Inhibition of return . Trends Cogn. Sci **4** (2000) 138–147
34. Privitera, C.M., Krishnan, N., Stark, L.W.: Clustering algorithms to obtain regions of interest: a comparative study. In: Proc. SPIE. Volume 3959., San Jose, CA (2000) 634–643
35. Gill, P.E., Muray, W., Wright, M.H.: Practical Optimization. Academic Press, London (1981)