

Evaluation of regions-of-interest based attention algorithms using a probabilistic measure

Martin Clauss, Pierre Bayerl and Heiko Neumann

University of Ulm, Dept. of Neural Information Processing, 89081 Ulm, Germany
{mc,pierre,hneumann}@neuro.informatik.uni-ulm.de

Computational mechanisms of attention that select salient locations in images request for quantitative measures for evaluation. We present a new measure for evaluation of algorithms for the detection of regions of interest (ROI). In contrast to existing measures, the present approach handles situations of order uncertainties, where the order for some ROIs is crucial, while for others it is not. We compare the results of several measures in various test scenarios. We further demonstrate how our measure can be used to evaluate algorithms for ROI detection, particularly the model of Itti and Koch for bottom-up data-driven attention.

Motivation

Biological vision is based on the dynamic selection of regions of interest (ROI) through the guidance of the gaze towards selected scenic regions. Several ROIs are scanned in a serial manner by fixating the high-resolution fovea of the eye on these suspicious locations. Stark and Choi [Stark1996] showed that the sequence of these fixations for one stimulus differs significantly between observers and is not even unique for one single observer. Several computational models exist that can detect and extract such ROIs from a given stimulus. In this study, we employed the model of Itti and Koch [Itti1998] for computation of ROIs. In this model a global saliency map is calculated from multiple feature maps to select ROIs as target locations.

The evaluation of attention algorithms is complicated since no formal underlying theory exists that can be utilized to guide this process. Therefore, only a few approaches for evaluation of such models have been proposed. For example, Itti and Koch evaluated their model with respect to noise exposure or the analysis of different feature combination strategies [Itti1999]. Current measures have difficulties accounting for variations in the order of ROIs. That is, if there are two or more ROIs that change order frequently because they are equally attractive, current measures cannot handle this correctly. We present a new measure that is able to handle such variations. The proposed measure will be compared to two other measures namely string edit distance and a simple measure which is obtained by calculating the percentage of the ROIs that were correctly detected in comparison to the ground truth.

The attention model of Itti and Koch

Itti, Koch and Niebur [Itti1998] presented a popular model for computing regions of interest using a saliency map which is obtained from a pyramidal representation of the input data. The computation is purely data driven, as it does not incorporate any feedback or knowledge-based mechanisms (compare with, e.g. [Tsotsos1995] for an extended bottom-up and top-down driven approach).

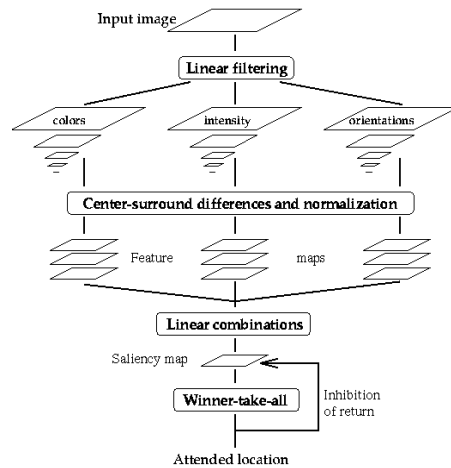


Fig. 1.:The model of Itti and Koch [Itti1998]

First a gaussian pyramid is built from the input stimulus. This pyramid is then split into several channels selective to different features like colour, intensity or orientation. Following this step, a centre-surround operation is performed on each of these multi-scale representations. All these maps are then combined into a single saliency map. Different strategies of combining these maps have been discussed and analysed in [Itti1999]. The simplest method is just summing up all the maps. In order to achieve this, the maps have to be scaled to the same spatial size prior to summation. More advanced methods we

applied perform iterative centre-surround inhibition to sharpen the data and extract local maxima [Itti1999].

A winner-take-all algorithm is applied to the resulting saliency map which determines the most salient location to be attended next. The sequential targeting of salient locations is realized by a mechanism that releases the currently selected position after some time. In order to avoid that this location is also attended in the next steps, the currently attended location is inhibited in the saliency map for a certain number of iterations (inhibition of return).

Methods of evaluation

In this section we will present two simple measures for evaluating models that select ROIs and their respective saliency that leads to a ranked order of target locations. We will further explain how our new measure is calculated. In order to calculate any measure a reference is needed first. This reference (“ground truth”) contains designated ROIs determined from detected salient locations. The diameter of the ROIs may be defined either manually or by the algorithm employed. For each test run the chosen ROIs have to be matched against the ROIs of the ground truth. In our tests

those ROIs are considered to be the same when the distance between the ROIs is less than their radius. Only then we assume the same ROI was detected again.

Order-independent measure. The measure presented in this section is a simple method to evaluate a model for ROI selection. Assume that we have a ground truth consisting of N ROIs each having a number assigned that denotes its saliency. Further, it is assumed that the model has identified M ROIs in the test data which are ordered sequentially according to the computed saliency measure. The result is then evaluated by calculating the ROIs from the ground truth that were found again within the set of M ROIs with highest rank order detected by the model. For example, if the ground truth has 5 ROIs and 3 of them are also found during the test run, the result is simply 0.6 or 60%. This measure can be implemented easily and also be calculated very quickly. Its major limitation is that it does not at all take care of the order in which the ROIs were chosen.

Order-dependent measure (string edit distance). Another measure to evaluate models that deal with sequentially ordered events is the "string edit distance" [Corman1997]. This measure counts the minimum number of operations needed to transform one string or sequence of ROIs into a given second string or sequence, with allowed operations being insertion, deletion and reassignment. The minimum amount of operations required for this transformation is usually computed using dynamic programming. For example, assume there are three ROIs A, B and C in the ground truth with corresponding saliency as such that they are detected exactly in this order. If the model detects them in the order A, C, B, the string edit distance is 2 (2 reassignments or 1 insertion and 1 deletion).

One limitation arising with this measure is that it is not possible to define two regions of interest of equal importance. Instead one ROI has always to be preferred over another when setting up the ground truth and its labelling order.

Proposed new measure. In order to circumvent the limitations of the two measures presented above, we develop a new measure which considers the order of the ROIs and that also accounts for systematic variations. We achieve this by approximating the probability distribution of relative order of ROIs through statistical means. These probability distributions are stored in matrices which are compared using normalized cross correlation.

The main steps for calculating these matrices are summarized as follows:

- first, the ROIs need to be determined. This may be done by any kind of source: Human/animal observers, by a person or by a computational model. The ROIs need to be assigned numbers denoting the corresponding column and row inside the matrix;
- the relative order of the ROIs is then voted in the matrix for further processing as shown in Fig. 2;
- in order to reliably estimate the probability distribution, the previous steps need to be performed multiple times;
- finally, the resulting matrix is normalized by dividing it through the number of iterations of the previous steps.

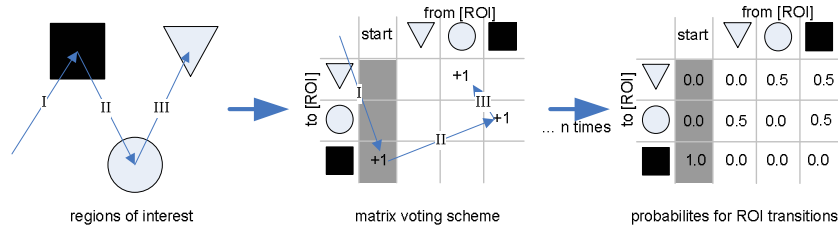


Fig. 2.: The matrix created from the relative order of ROIs. The abscissa denotes the label of the ROI “A” that preceded the current one. The ordinate of the matrix denotes the label of the current ROI “B”. As the very first ROI detected does not have any predecessor, an additional column is needed. Each value in this “pre-occurrence matrix” represents how often ROI “A” preceded ROI “B”. Through normalization the values are transformed into probabilities. If there are several ROIs that should have the same importance the probabilities are simply spread over several possible predecessors. In order to account for regions of interest that were not present in the ground truth, an extra row and column exist, which sum up those additional regions of interest. For the ground truth those cells are all zero.

The obtained matrix encodes the probabilities for all pairs (a,b), i.e. denoting that ROI a precedes ROI b.

First, the calculation presented above has to be done once for a ground truth run resulting in a matrix A. This matrix describes the relative order of ROIs for the ground truth. After that, the calculation is performed for one or more test runs and a corresponding matrix B is returned. The two matrices are then compared by calculating the normalized cross correlation of the two matrices. The measure presented in this section is capable of handling ROI sequences of both strict as well as loose order.

Results

We now compare our statistical measure against string edit distance and the simple measure. In the model of Itti and Koch we employed the local iterative inhibition algorithm as feature combination strategy ([Itti 1999]).

We first select a scenario which all of the measures are expected to be able to handle. The corresponding input stimulus is shown in Fig. 3. It consists of six radially aligned circles of constant diameter and decreasing intensity on a constant background. As the brightness of the circles decreases, so does their contrast against the background. The corresponding saliency of the discs is also ordered in accordance to their contrast. Therefore, there is a strict order of preference of the ROIs. When the noise level is increased, the number of discs that can be detected decreases monotonically. We therefore expect that all three measures employed for evaluation decrease roughly linear when the stimulus is corrupted by a certain amount of noise. Fig. 4 demonstrates that this is approximately the case. All measures return comparable results.

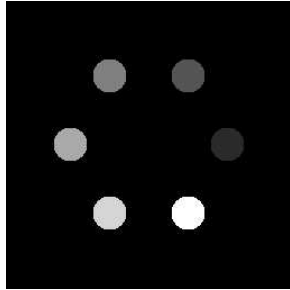


Fig. 3.: Input stimulus consisting of six discs with monotonically increasing luminance on constant background

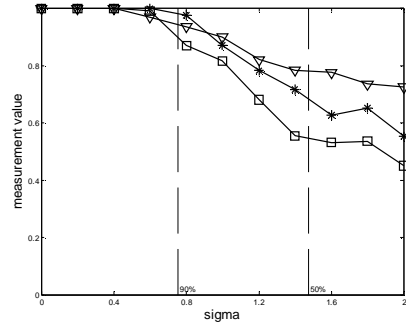


Fig. 4.: Results for Fig. 3. String edit distance (squares), our proposed measure (stars) and simple measure (triangles) all behave similar and start decreasing monotonically at 90% correct sequences (vertical dashed mark).

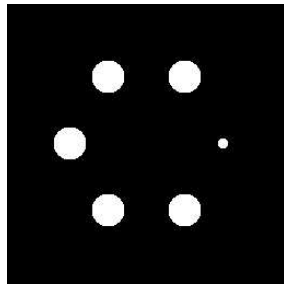


Fig. 5.: Input stimulus. Six discs with constant intensity, one small and five large discs on constant background

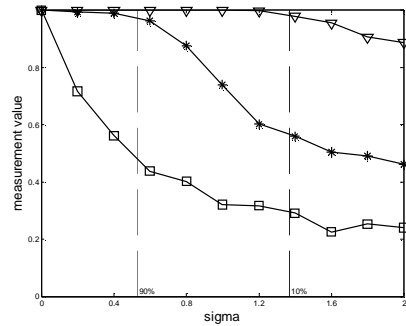


Fig. 6.: Results for Fig. 5. The two vertical dashed bars denote where 90% and 10% of the detected ROI sequences were correct. String edit distance (squares) decreases mostly before the 90% mark. The simple measure (triangles) remains almost up to the 10% mark. Our proposed measure (stars) decreases especially between the two marks.

Next we calculate all three measures for the input stimulus shown in Fig. 5. This stimulus again consists of circularly arranged discs. Here, all discs but one have the same diameter, the odd one is smaller in size. The discs all have the same brightness; the background is constant. The saliency here is defined by the size contrast of the discs. The results are shown in Fig. 6.

Once again we perform local iterative inhibition prior to feature map combination. For noise corrupted versions of the stimulus the five equal discs are likely to vary in their order of saliency. Consequently, their rank order changes and we expect that the string edit distance measure is not able to interpret the output of the algorithm correctly. In contrast, the simple measure is supposed to degrade too slowly, as the

five large discs can be easily detected even in noisy situations. Our measure is able to handle this situation: It is sensitive with respect to order of the discs that have different sizes whereas the order of equal sized discs is ignored.

Conclusion

Our results demonstrate that in addition to the two extreme cases (strictly linear order and no order at all) our proposed measure is able to also handle situations where the order of ROIs is partially ambiguous. This is an improvement in comparison to the other two measures analyzed in this work. This flexibility is necessary in order to compare computational mechanisms with human or animal test subjects.

Acknowledgements

This work is funded by the state of Baden-Württemberg as a part of the project “Systemarchitekturen zur Gewährleistung sicherer und Ressourcen schonender Mobilität im Straßenverkehr”

References

- [Cormann1997] Corman, T. H., Leiserson, C. E., Rivest, R. L.: *Introduction to Algorithms*, New York: McGraw-Hill, 1997
- [Itti1998] Itti, L., Koch, C., Niebur, E.: *A model of saliency-based visual attention for rapid scene analysis.*, IEEE Transactions on PAMI, **20**, 1254-1259, 1998
- [Itti1999] Itti, L., and Koch, C.: *A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems*, SPIE human vision and electronic imaging IV(HVEI99), San Jose, CA, pp. 473-482
- [Stark1996] Stark, L. W., Choi, Y. S.: *Experimental Metaphysics: The scanpath as an epistemological mechanism.*, W. H. Zangemeister, H. S. Stiehl, & C. Freska (Eds.), Visual attention and cognition, pp. 3-69. Amsterdam: Elsevier Science B.V
- [Tsotsos1995] Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., Nuflo, F.: *Modeling visual attention via selective tuning*, Artificial Intelligence, **78**, 507-545, 1995