A Probabilistic Interpretation of the Saliency Network

Michael Lindenbaum * and Alexander Berengolts **

Computer Science Department, Technion,
Haifa 32000, ISRAEL

Abstract. The calculation of salient structures is one of the early and basic ideas of perceptual organization in Computer Vision. Saliency algorithms aim to find image curves, maximizing some deterministic quality measure which grows with the length of the curve, its smoothness, and its continuity. This note proposes a modified saliency estimation mechanism, which is based on probabilistically specified grouping cues and on length estimation. In the context of the proposed method, the well-known saliency mechanism, proposed by Shaashua and Ullman [SU88], may be interpreted as a process trying to detect the curve with maximal expected length.
The new characterization of saliency using probabilistic cues is conceptually built on considering the curve starting at a feature point, and estimating the distribution of the length of this curve, iteratively. Different saliencies, like the expected length, may be specified as different functions of this distribution. There is no need however to actually propagate the distributions during the iterative process.
The proposed saliency characterization is associated with several advantages: First, unlike previous approaches, the search for the "best group" is based on a probabilistic characterization, which may be derived and verified from typical images, rather than on pre-conceived opinion about the nature of figure subsets. Therefore, it is expected also to be more reliable. Second, the probabilistic saliency is more abstract and thus more generic than the common geometric formulations. Therefore, it lends itself to different realizations of saliencies based on different cues, in a systematic rigorous way. To demonstrate that, we created, as instances of the general approach, a saliency process which is based on grey level similarity but still preserve a similar meaning. Finally, the proposed approach gives another interpretation for the measure than makes one curve a winner, which may often be more intuitive to grasp, especially as the saliency levels has a clear meaning of say, expected curve length.

1 Introduction

The human visual system (HVS) is capable of filtering images and finding the important visual events so that its limited computational resources may be focused on them and used efficiently. This discrimination between the important

* email address:mic@cs.technion.ac.il
** email address:aer@cs.technion.ac.il

parts of the image, denoted "figure", and the less important parts, denoted "background", is done before the objects in the image are identified, and using general rules (or cues) indicating what is likely to be important [Wer50].

Presented, for example, with a binary image containing points and/or curves (such as those resulting from edge detection), it turns out that this perceptual process prefers to choose for figure, a subset of points lying on some long, smooth and dense curve.

To account for this phenomenon with a computational theory [Mar82], Shaashua and Ullman suggested a particular measure, denoted saliency, that is a particular quantification of the desirable smoothness and length properties. They have shown that indeed, the image subsets, associated with high saliency are those considered as more important by common human subjective judgment [SU88]. One important advantage of this computational theory is that this global optimization may be formulated as a dynamic programming task and consequently may be carried out as an iterative process running on a network of simple processors getting only local information. This makes the theory attractive because the proposed process is consistent with common neural mechanisms.

The saliency measure of [SU88] was re-analyzed recently as well, revealing some deficiencies. A generalization, stating that every saliency measure which satisfy some conditions set in [SU88], can be optimized in the same way, was suggested in [AM98]. Other measures of saliency, based on non-iterative local support [GM93], eigenvectors of an affinity matrix [SB98] and stochastic models for particle motion [WJ96] were suggested as well. A survey on different saliency methods is described in [WT98]. The aim of the work on saliency remains to explain perceptual phenomena such as Figure from Ground abilities and illusory contours perception, but also to provide a computer vision tool for intermediate level sorting and filtering of the image data. Work on Figure from Ground discrimination such as [HH93,HvdH93,AL98], do not address explicitly the saliency issue but, implicitly, calculate a (binary) saliency as well.

Having its origin in an attempt to explain a perceptual phenomena, most of the work on saliency does not emphasize the justification for the HVS preference of long smooth curves. It just tries to find a computational mechanism that produces such preference. Note that the particular saliency measure proposed in [SU88] is one particular quantification of the intuitively phrased desired properties. It may be (slightly) modified (by say, replacing the curvature value by twice its value), yielding a measure which is as plausible and computationally efficient, but leading to a different choice of the most salient curve.

For perceptual modeling, the "best" saliency measure may be decided by psychophysical experimentation. For computer vision applications, however, optimizing the saliency measure requires first to agree on a quantitative criterion. The initial motivation of this work is to provide an interpretation and another justification of the original saliency concept.

We show here how saliency like measures may be derived within a more general framework, namely the quantification of grouping reliability using probabilities. The method is conceptually built on considering the curve starting at a feature point, and estimating the distribution of the length of this curve, iteratively. Different saliencies, like the expected length, may be specified as different functions of this distribution. Although central to the explanation of the method, there is no need, in practice, to propagate the actual distribution during the iterative process, which indeed would have required a substantial computational effort.

The proposed view and corresponding algorithm is different than that considered in [SU88] (for example, with regard to the treatment of virtual (non-feature) points), but it shares the iterative dynamic programming like algorithm. When phrased in terms of our algorithm, the original saliency of [SU88] corresponds to a curvature/distance based grouping cue. Maximizing it at a point corresponds to maximizing the expected length of the curve on which this point lies. This way, the traditional saliency measure gets a different interpretation, of looking for objects associated with maximal expected perimeter.

The new characterization of saliency using probabilistic cues is associated with the following advantages:

1. reliability — Basing the search for the “best group” on the probabilistic characterization, which may be derived from typical images, (using ground truth), rather than on pre-conceived opinion about the nature of figure subsets is expected to give better choices of significant groups.
2. generality — the probabilistic saliency is more abstract and thus more generic than the original geometric formulation. Therefore, it lends itself to different realizations of saliencies based on different cues. To demonstrate that, we run the same saliency method with two different cues: low curvature and grey level similarity.
3. another perspective — consider the SU saliency not only by its original curvature based interpretation, but also by its probabilistic interpretation gives another interpretation for the measure than makes one curve a winner, and may often be more intuitive to grasp, especially as the saliency levels has a clear meaning of say, expected curve length.

The paper continues as follows. First, in section 2, we present the length distribution concept, and show how different saliency measures may be built upon it. The proposed saliency process is described in section 3, where we consider the iterative calculation, some shortcuts allowing not to calculate or to keep the actual distribution, convergence issues, and the formulation of the SU saliency as an instance of the new saliency. Some experiments, demonstrating the different types of saliencies resulting from the proposed saliency algorithm, are described in section 4.

2 Probabilistic Saliency

2.1 Length distributions

Let x_i be a directional feature point in the image (e.g. an edgel). Such a point may or may not belong to some curve which extends l_+ length units to one side and l_- length units to the other side. Here we consider these lengths as random variables associated with the feature x_i, and characterize them by the distributions $\mathbf{D}^i_+(l)$ and $\mathbf{D}^i_-(l)$, respectively. The direction, used to keep the order in the curve, is specified relative to say, the direction of the gradient at this feature point, and may take one of the two $\{+, -\}$ values. The parts of the curve lying in the positive and negative directions are denoted positive and negative extensions, respectively. Our basic intuition is that points with long extensions, correspond to larger objects and deliver more significant information about the content of the image. Therefore we shall try to find those feature points associated with the $\mathbf{D}^i_+(l)$ and $\mathbf{D}^i_-(l)$ distributions, which put more weight on longer l values.

When no connectivity information is available, all features are not known to belong to any curve. Then, all distributions are concentrated on very short lengths, corresponding to the length of the corresponding feature themselves. For simplicity, we assume that all these initial distributions are identical and denote this initial distribution by $\mathbf{D}^*(l)$.

2.2 Length distribution update rules

Consider two features, x_i and x_j, which belong to some curve, such that x_j lies in the positive extension of x_i. Suppose that $\mathbf{D}^j_+(l)$ is known. Then, $\mathbf{D}^i_+(l)$ can be written as

$$\mathbf{D}^i_+(l) = \mathbf{D}^{j \to i}_+(l) = \mathbf{D}^j_+(l - l_{ij}) \tag{1}$$

where l_{ij} is the distance from x_i to x_j (on the curve). This follows by observing that a positive extension of length l associated with x_j implies that the length of the positive extension of x_i is $l + l_{ij}$. The notation $\mathbf{D}^{j \to i}_+(l)$ explicitly emphasizes that this is an inference of the length distribution associated with the $i-$th feature from the known distribution associated with the $j-$th feature.

In the common situation in image analysis, we can never be sure that two features lie on the same curve. In a non-model-based context, we can only estimate the probability for this event based on local information such as perceptual organization cues [Low85]. Let $c(x_j)$ denote the curve on which x_j lies and let P_{ij} be the probability $Prob\{x_i \in c(x_j)\}$. This probability, denoted as "the grouping cue" is expected to be inferred from perceptual information. Specifying the affinity value between the two feature points x_i and x_j, in this probabilistic abstract way, allows to calculate a saliency like measure, based on different grouping cues and not only on the co-circularity cue used in [SU88]. As we shall see, this probabilistic formulation provides a common meaning for the different saliencies associated with the different cues, independently of the different types of information they employ.

Consider now an algorithm trying to find a path between the feature points, and some hypothesis about a particular path, in which the feature x_j lies on the positive extension of the feature x_i. If the length distribution $\mathbf{D}^j_+(l)$ is known then the expected value of the length distribution $\mathbf{D}^i_+(l)$, is

$$\hat{\mathbf{D}}^{j \to i}_+(l) = P_{ij}\mathbf{D}^j_+(l - l_{ij}) + (1 - P_{ij})\mathbf{D}^*(l). \tag{2}$$

Note that this is an estimate of the length distribution of the positive extension of x_i, under a particular hypothesis regarding the path. The possibility that x_i belongs to some other path (or curve) which does not contain x_j is not taken into account. Therefore the only options for x_i are either to be connected to this curve or to be disconnected from anything (in the positive direction). An alternative formulation, where all curves to which x_i may belongs are taken into account, leads to a Bayesian estimate of $\mathbf{D}^i_+(l)$. See section 5 for a discussion of this alternative and its relation to the saliency like approach developed in [WJ96].

Suppose now that a path $\Gamma = \{x_1, x_2, \ldots, x_N\}$ starts at the feature point x_1, such that x_{i+1} is on the positive extension of x_i, $i = 1, \ldots, N-1$. Then, the length distribution associated with x_1 may be recursively calculated: $\mathbf{D}^N_+(l) = \mathbf{D}^*(l)$, $\hat{\mathbf{D}}^{N-1}_+(l) = \hat{\mathbf{D}}^{N \to N-1}_+(l)$, $\ldots$, until $\hat{\mathbf{D}}^1_+(l)$ is finally estimated. A distribution estimated this way, from a path of length N, is denoted (when we we want to make it explicit), $\hat{\mathbf{D}}^i_{+N}(l)$.

2.3 Probabilistic Saliency

Let $Q[\mathbf{D}^i_+(l)]$ be a (scalar) quality measure computable from the length distribution, and quantifying, in some way, the desired property of a long curve. Typical measures may be the average length or other moments. This measure serves as a one-sided-saliency, and we shall look for features points maximizing it and for curves containing such points. Note that every feature point is associated with two one-sided saliencies, corresponding to the two directions. Some possible choices for the saliency are

Maximum one-sided expected length — A straightforward saliency measure is the expected value of the extension length random variable, which is denoted expected length and is easily calculated from its distribution.

Maximum two-sided expected length — Unless the curve is close and very tightly connected. maximizing the expected length in the two directions is done independently for the two sides. Then, the sum of these one-sided saliencies at a point is just the expected length of the curve on which the point lies.

Maximum confidence one-sided curve — Some common object recognition process, which rely on curve invariants, need some continuous curve from the object. In such scenario, some reasonably long curve associated with high reliability is preferred over a longer curve with lower reliability. Here, the preferred curve is characterized by a distribution concentrating around one value, in contrast to an uncertain estimate, characterized by a closer to uniform distribution.

In the rest of this paper, the one-sided expected length saliency is usually used. although one example, demonstrating the advantages of the confidence emphasizing approaches is considered in the experiments. The expected length is the measure corresponding to SU saliency and its interpretation is simple and clear. We shall also see that it has algorithmic advantages.

For feature points on closed curves, the meaning of the saliency as expected length is distorted, because the length of points of the curve is counted twice or more (after a sufficient number of iterations). The increase of the saliency of close curves is often considered desirable because closer curves have usually higher significance over their open counterparts with the same length [SU88]. Calculating the expected length for closed curves can be done using the technique described in [AB98] and shall not be repeated here.

3 The probabilistic saliency optimization process

The aim of the optimization process is, for every feature point, to find a path, starting at this point and maximizing the saliency of that point (calculated relative to this path).

(We should mention here that the proposed method is similar, in principle, to that proposed by Shaashua and Ullman (see [SU88,AB98]), and is brought here only because some details differ (due to the use of distributions) and for completeness. We tried to use similar notations when possible. The calculation of saliency in the sense of [SU88], for a sparse set of feature points (i.e. without virtual feature points) was considered also in [AM98].)

Calculating this optimum is easy for short paths (e.g. $N = 1, 2$) but is generally exponential in N. Fortunately, it may be calculated by a simple iterative process using dynamic programming if the quality criterion (or saliency) is extensible [SU88]. That is, if the saliency associated with the best (length N) path starting from x_i satisfies

$$Q[\mathbf{D}^i_{+N}(l)] = \max_j F\left(q_{ij}, \mathbf{D}^j_{+(N-1)}(l)\right)$$

where q_{ij} is a quantity calculated from the feature points x_i and x_j, $\mathbf{D}^j_{+(N-1)}(l)$ is the distribution associated with the best (length $N-1$) path, associated with the highest saliency, starting from x_j, and the maximization is done over all neighbors x_j of x_i. Note that this condition is a bit more general that that suggested in [SU88], as the new saliency calculation may use the distribution and not only a function of it. In fact, all information about the best path may be used as well, as the more general condition is that the optimal solution contains within in optimal solutions to subproblem instances [CLR90]. Note that the expected length is an extensible quality criterion.

3.1 The iterative process

Preprocessing:
A neighborhood is specified for every feature point.

At the k-th iteration ($k = 1, 2, 3, \ldots$)
For every feature point x_i

1. For all neighbors x_j $j = 1, 2, 3 \ldots$ of x_i
   (a) calculate the grouping cue P_{ij}.
   (b) update the length distribution $\mathbf{D}_{+}^{j \to i}(l)$ using eq. (2), and calculate $Q[\mathbf{D}_{+}^{j \to i}(l)]$.
2. Choose the neighbor x_j maximizing the quality measure and update the length distribution to $\mathbf{D}_{+}^{j \to i}(l)$.

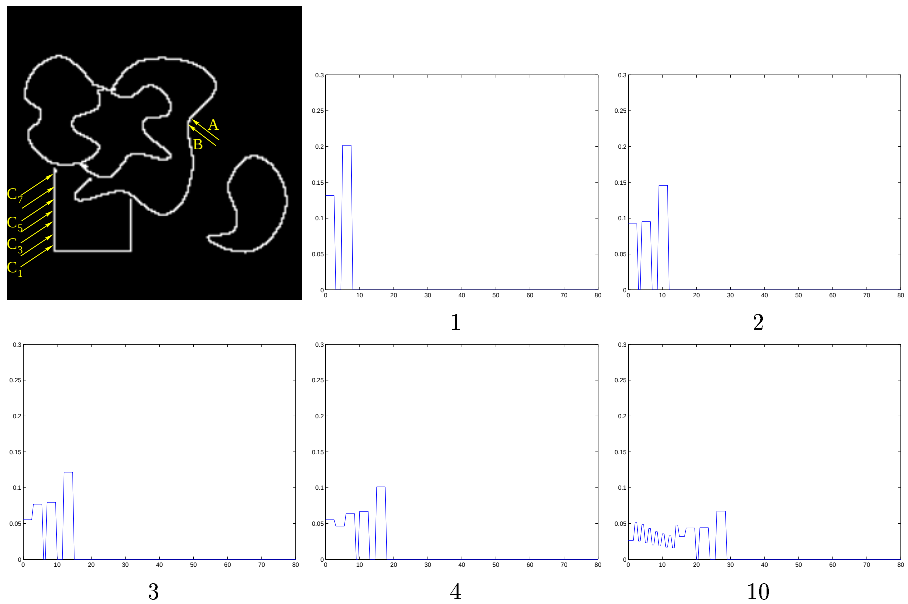
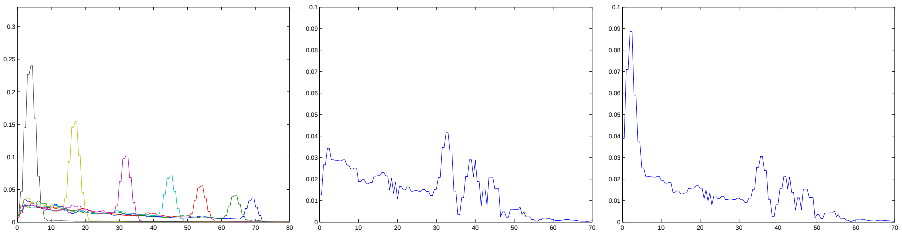


Fig. 1. The one sided length distribution at the point C_3 (in the top left illustration), plotted for 1,2,3,4 and 10 iterations. Note that for such a smooth curve (a straight line segment), the distribution quickly develops a significant weight for the large values.

The procedure starts when all feature points are associated with the basic distribution $\mathbf{D}_{+}^{*}(l)$. For saliencies prefering long curves, the process behaves as follows: At the first stage, every feature point x_i chooses the best perceptually connected neighbor x_j, so that P_{ij} is maximal for it, "improves" its distribution, and increase its saliency. At the next iterations, the prefered neighbor is chosen not only by its perceptual affinity but also by its own saliency, as generated in the previous iterations. See Figure 1, illustrating the development of the length distribution associated with a particular point and Figure 2 which describe some (roughly) stable distribution obtained after many iterations.

Apart from building the length distributions, the process also specifies, for every feature point, the next feature point on its extension. Thus, starting from salient points, the iterative process finds also the long, well connected, curves which contributed and supported this high saliency.



(a) (b) (c)

Fig. 2. The left graph (a) describes some distributions corresponding to the different points $C_1, \ldots, C_7$ (in the previous Figure) after 80 iterations. Note that points which are close to the end (C_1 is the closest) cannot develop large value, and correspond to the distributions with peaks on small l values. The point A is gets support from a smooth curve and is associated with a distribution having significant weight in the high values (b). The point B is weakly connected to A, and therefore, its distribution is an average of the initial distribution, focusing on low values and that of A, which makes its roughly bimodal (c).

3.2 Shortcuts

Apparently, one deficiency of the proposed saliency is the need to update a length distribution for every feature point, which is costly in time and space. To alleviate this problem we suggest to store and update only the statistics required to calculate the prefered saliency. For example, For calculating the expected length quality measure, let $E^*[l]$ be the expected length associated with the distribution $\mathbf{D}^*_+(l)$. Then, the distribution update rule is changed to the following expected length update rule

$$E[l]^i_+ = E[l]^{j \to i}_+ = P_{ij}(l_{ij} + E[l]^j_+) + (1 - P_{ij})E^*[l]$$

Other statistics (e.g. variance) may be propagated similarly, and there is usually no need to propagate the entire

3.3 Optimality and Convergence

By the same arguments made in standard dynamic programming and in [SU88], after N iterations, the length distribution of the i-th feature is associated with the maximal saliency. The maximum is over all possible curves of length N starting in the i-th feature point. This optimization happens for all feature points simultaneously. One (or more) of them will also achieve the global saliency measure. Therefore, the process finds also the maximal quality curve, as measured by the saliency of its endpoint.

After N iterations, all the open paths of length N or less, which start at x_i make their maximal contribution. If N is set as the number of feature points in the image, then the process should converge after N iterations. The exception is of course closed curves, which are equivalent to infinite chains. We show now that even for closed curves the length distribution converges. The proof takes follows some principles from [AB98].

Consider, for example, a feature point on a closed path of length N_c. Let this point be the i-th point and let the direction be such that this i-th point updates its distribution based on the $(i+1)$-th point. Until the N_c-th iteration, the closure does not effect the distribution associated with the feature point. At the N_c-th iteration, the saliency of the i-th point may be written as

$$\begin{aligned}\mathbf{D}^i_{+N_c}(l) &= (1-P_{i,i+1})\mathbf{D}^*(l) + P_{i,i+1}\mathbf{D}^{i+1}_{+(N_c-1)}(l-l_{i,i+1}) \\ &= (1-P_{i,i+1})\mathbf{D}^*(l) + P_{i,i+1}(1-P_{i+1,i+2})\mathbf{D}^*(l-l_{i,i+1}) \\ &\qquad +P_{i,i+1}P_{i+1,i+2}\mathbf{D}^{i+2}_{+(N_c-2)}(l-l_{i,i+1}-l_{i+1,i+2}) \\ &= \dots \\ &= (1-\Pi_{j=0}^{N_c-1}P_{i+j,i+j+1})\bar{\mathbf{D}}^*(l) + \Pi_{j=0}^{N_c-1}P_{i+j,i+j+1}\mathbf{D}^i_{+0}(l-\textstyle\sum_{j=0}^{N_c-1} l_{i+j,i+j+1}) \\ &= (1-\alpha)\bar{\mathbf{D}}^*(l) + \alpha\mathbf{D}^i_{+0}(l-L) \\ &= (1-\alpha)\bar{\mathbf{D}}^*(l) + \alpha\mathbf{D}^*_{+}(l-L)\end{aligned} \tag{3}$$

$\bar{\mathbf{D}}^*(l)$ is an average distribution of $\mathbf{D}^*(l), \mathbf{D}^*(l-l_{i,i+1}), \mathbf{D}^*(l-l_{i,i+1}-l_{i+1,i+2}), \dots$ (with non-equal coefficients), $L = \sum_{j=0}^{N_c-1} l_{i+j,i+j+1}$, and $\alpha = \Pi_{j=0}^{N_c-1}P_{i+j,i+j+1}$. Note that while the distribution of the i-th feature point is no longer the initial distribution, this update is not reflected yet in the way it supports itself through the closed curve. From the next iterations however, the change of the i-th feature point histogram will be reflected in this support, and after N_c additional iterations the histogram will change to

$$\begin{aligned}\mathbf{D}^i_{+2N_c}(l) &= (1-\alpha)\bar{\mathbf{D}}^*(l) + \alpha\mathbf{D}^i_{+N_c}(l-L) \\ &= (1-\alpha)\bar{\mathbf{D}}^*(l) + \alpha[(1-\alpha)\bar{\mathbf{D}}^*(l-L) + \alpha\mathbf{D}^*_{+}(l-2L)]\end{aligned} \tag{4}$$

After $K \cdot N_c$ iterations,

$$\mathbf{D}^i_{+KN_c}(l) = (1-\alpha)\sum_{k=0}^{K}\alpha^k\bar{\mathbf{D}}^*(l-kL) + \alpha^K\mathbf{D}^*_{+}(l-KL) \tag{5}$$

Consider now any finite moment or order m associated with the length distribution. Note that $\mathbf{D}^*_{+}(l-kL)$ (and $\bar{\mathbf{D}}^*(l-kL)$) has zero weight on any length l higher than kL. Therefore, after the KN_c-the iteration, this moment, denoted $M^i_{+KN_c}$, is bounded:

$$\begin{aligned}M^i_{+KN_c} &\leq (1-\alpha)\textstyle\sum_{k=0}^{K}\alpha^k(kL)^m + \alpha^K(KL)^m \\ &= (1-\alpha)L^m\textstyle\sum_{k=0}^{K}\alpha^k k^m + L^m\alpha^K K^m\end{aligned} \tag{6}$$

For any reasonable cue, $\alpha = \Pi_{j=0}^{N_c-1}P_{i+j,i+j+1}$ is strictly smaller than one and the bounds on the moments converges. The moments themselves are increasing with K, and therefore converge, and hence the distribution converge.

3.4 Relation to the original SU saliency

The original saliency measures, proposed in [SU88], meant to mimic the human visual system (HVS) behavior and to model the priority it gives to long smooth curves, even when they are fragmented. Our approach, on the other hand, is based on a statistical characterization of grouping cues, which is believed to be available. It is well known that the HVS is very successful in grouping tasks, therefore, the statistics of grouping cues must have been learned and incorporated into its grouping mechanisms. Thus, it is expected that our method will also give results, which are compatible with the HVS preferences. For cues based on co-circularity, which is the principle used in [SU88], the results of both methods are expected to be similar.

We shall show now that in the context of curvature/distance based cue, the SU algorithm corresponds to an instance of our algorithm: The saliency of the $i-$th feature, specified in [SU88] is updated by the local rule

$$E_i^{(n+1)} = \sigma_i + \rho_i \max(E_j^{(n)} f_{ij})$$

where the maximum is taken over all the features in the neighborhood of the $i-$th feature, and

- $E_i^{(n)}$ is the saliency of the i-th feature after the $n-$th iteration,
- σ_i is a "local saliency" which is set as a positive value (e.g. 1) for every real feature,
- ρ_i is a penalty for gaps which is set to one in features (no gap) and to a lower value when the feature is virtual. Finally,
- f_{ij} is a "coupling constant" which decreases with the local curvature.

In the framework of [SU88] features could be "real" (where we have, say, an edge point), or "virtual" where there is no local image based evidence for an edge. This choice allows to hypothesize an image independent and1g1g parallel local architecture which is a plausible model for a perceptual process. A virtual feature does not add to saliency and therefore is associated with null σ_i. It should also attenuate the currently existing saliency and is therefore associated with lower than one ρ_i parameter.

In our framework, all features are real. For them, the co-circularity may be interpreted as a measure for the grouping probability: by the general assumption that smooth curves are likely, a low curvature implies that connection is more probable then high curvature. Thus, for real feature points, the SU update formulae may be interpreted as

$$E_i^{(n+1)} = 1 + \max(E_j^{(n)} P_{ij}). \tag{7}$$

The ability to continue the curve over gaps is interpreted as follows: Suppose that the three feature points x_i, x_j, x_k are consecutive along the curve, are chosen as such by the SU algorithm, and let x_j be a virtual feature point. Then, the SU saliency of x_i is (roughly) $E_i^{(n+1)} = 1 + f_{ij} E_j^{(n)} = 1 + f_{ij} f_{jk} \rho_k E_k^{(n-1)} = 1 + P_{ik} E_k^{(n-1)}$. Thus, the effect of a missing point may be replaced by a lower probability $P_{i,k} = f_{ij} f_{jk} \rho_k$. The probability of the feature point x_i to be part of the curve $c(x_k)$ on which x_k lies, is indeed lower when there is a gap between x_i and x_k. Moreover, the process of calculating a cue between two distant points may be considered as an explicit search for a path between them, which minimizes a cost function.

Recall now (from section 3.2) that the expected length propagates as

$$\begin{aligned} E[l]^i_+ &= (1 - P_{ij})E^*[l] + P_{ij}(l_{ij} + E[l]^j_+) \\ &= E^*[l] + P_{ij}(l_{ij} - E^*[l] + E[l]^j_+) \end{aligned} \tag{8}$$

which, for inter-pixel distance of l_{ij} equal to the expected length of one edgel $E^*[l]$, and both equal to 1, yields

$$E[l]^i_+ = 1 + P_{ij}E[l]^j_+ \tag{9}$$

Therefore we conclude that the co-circularity and the gap attenuation, used in SU saliency, may be interpreted as measures of the grouping probability used here, and that the overall saliency maximized there, is, according to this interpretation, the expected length.

There are also other differences, but they are technical, and result from our use of directional feature points, implying that we can work with the actual features, and not with the arcs between the features as done in [SU88].

4 Implementation and Experiments

In contrast with [SU88,AB98], we considered only real (non-virtual) feature points. They were oriented using the gradients direction. The positive (negative) extension neighbors of every feature point were all (real) neighboring feature points, s.t. the vector $x_i x_j$ is making an angle in $[\pi/6, 5\pi/6]$ ($[-5\pi/6, -\pi/6]$) with the gradient. The neighborhood was usually a disk of radius 10 pixels. The initial length distribution was set to have equal weights on the values $0, 1$ and 2.

Traditional co-circularity cue. First we considered the classical cue, using curvature (or weighted angle differences). Following [SU88], we set the cue as

$$P_{ij} = \exp\{\|x_{ij}\|^2/50\} \cdot \exp\{-\tan(GradAngleDiff/2)\},$$

where $GradAngleDiff$ is just the difference between the two gradient angles in the two points. Note that as the distances between a feature points and its neighbors is no longer constant, we added a preference to short distances. Interestingly this dependency needs to reduce the cue faster than $\exp\{\|x_{ij}\|\}$ because otherwise the process always prefers the far neighbors. (Going to that neighbor through another, closer neighbor, gives a lower expected length, which follows directly from the update rules.)

We view this experimental work as an intermediate stage, because the actual probabilities are not those determined by this parametric form. Our current work focuses on measuring these cues empirically.

Here (Figures 3,4 are two examples of the implementation. They include the original image (synthetic and real), the edge points detected with standard

DRF (Khoros) operator, the two one-sided saliencies and their sum, and the thresholded saliency. Note that the saliency image has a concrete meaning: it is the expected length on which the point lies. For the one sided case for example, if one starts from a point associated with saliency of 38 (a typical value for the strong curves, on say, the lizard back), he can expect to find about 38 neighbors on the curve in one of the directions.

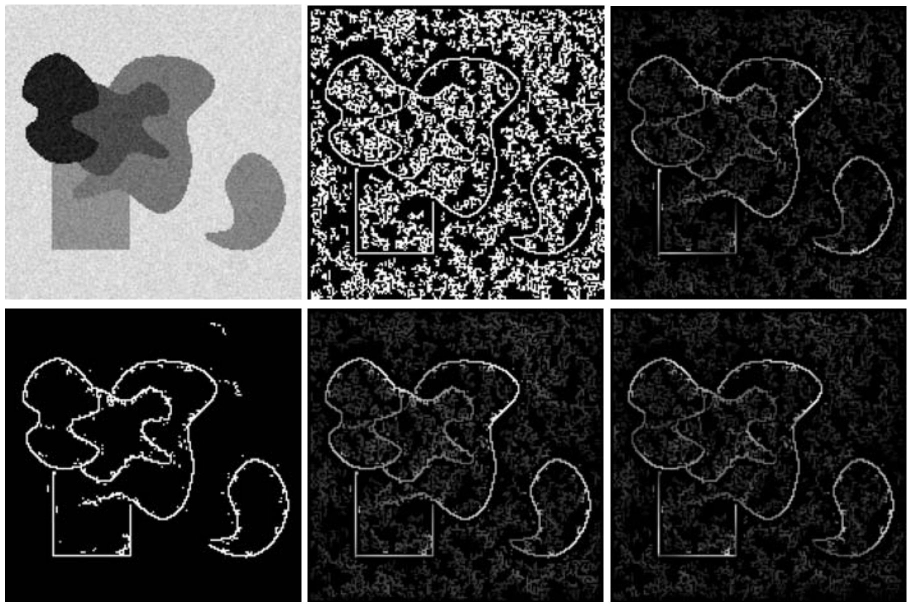


Fig. 3. A typical saliecy calculation with an **angle** cue done on a heavily corrupted noise: (starting from upper left, clockwise) The original image, edges, positive and negative saliencies, sum of saliency, thresholded saliency.

Saliency with a Grey Level Cue. Next we took the same saliency process and just changed the cue, which now, measure the similarity in grey levels and not the smoothness of the curve. Specifically, we set

$$P_{ij} = \exp\{\|x_{ij}\|^2/50\} \frac{1}{1 + \frac{30\, GreyLevelDiff^2}{GradSize(i) GradSize(j)}}.$$

The $GreyLevelDiff$ is the difference in grey levels between the two feature points, and $GradSize(i)$ is the gradient size at x_i. See Figure 5. Note that most unwanted additions to the thresholded saliency image are in inner points where the grey level is similar and random high gradients exist. Note also that the saliency value has the same meaning: expected length of the curve (either to one side or to both). Actually, the results were better than we expected and in a sense outperform the use of the angle-based cue. We intend to investigate this issue farther and with real images as well. To conclude, this experiment demonstrates that a saliency process which is similar, in principle, to that proposed in [SU88], can work also with other sources of information.

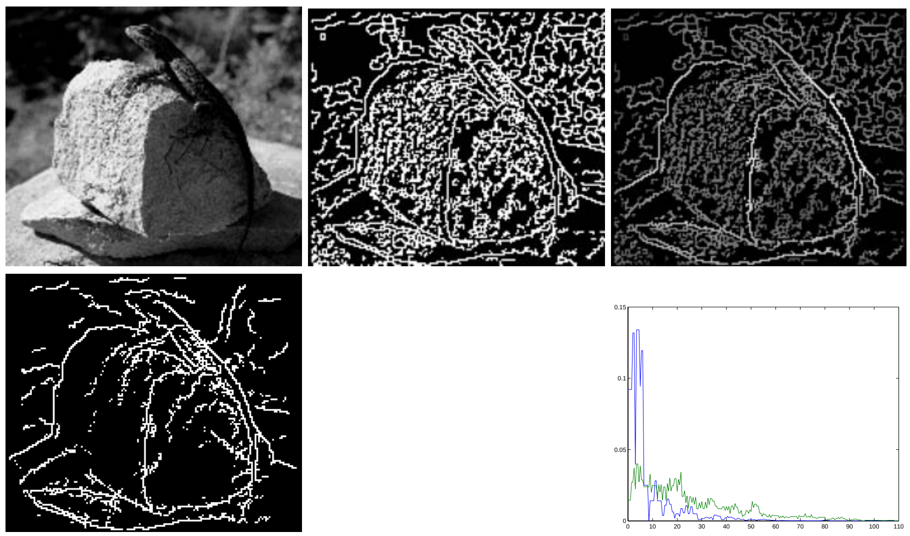


Fig. 4. Saliency calculation for a real, complex, image: (From upper left, clockwise) original image, edges, sum of positive and negative saliencies, two length distributions associated with a stone point (dark) and the lizard back (lighter), and thresholded saliency.

A Saliency measure emphasizing confidence. It may happen that a relatively weakly connected sequence of edgles will yield a substantial expected length (or SU) saliency. Indeed this was the case, for example, in the real "lizard" image, where many points on the texture were associated with large saliency. Thus, a quality measure emphasizing the connectedness over the long length may be preferred. One such measures is the "expected square root length", specified as $\sum_l \mathbf{D}^i_+(l)\sqrt{l}$, which prefers shorter curves associated with higher confidence: consider, for example, two length distributions, one giving a full weight to the length $l = 10$ and another sharing the weight between $l = 0$ and $l = 20$. While the expected length associated with the distributions is identical, the expected square root length clearly prefers the more "concentrated" distribution where full confidence is given to the $l = 10$ value, and give it a saliency value of $\sqrt{10}$ which is larger by a factor of $\sqrt{2}$ than the saliency associated with the other distribution. Indeed we found that such saliency may have advantages when working on real images (see Figure 6).

Note however that such saliency has one severe theoretical deficiency: it is not extensible, and thus global maximization is not guaranteed.

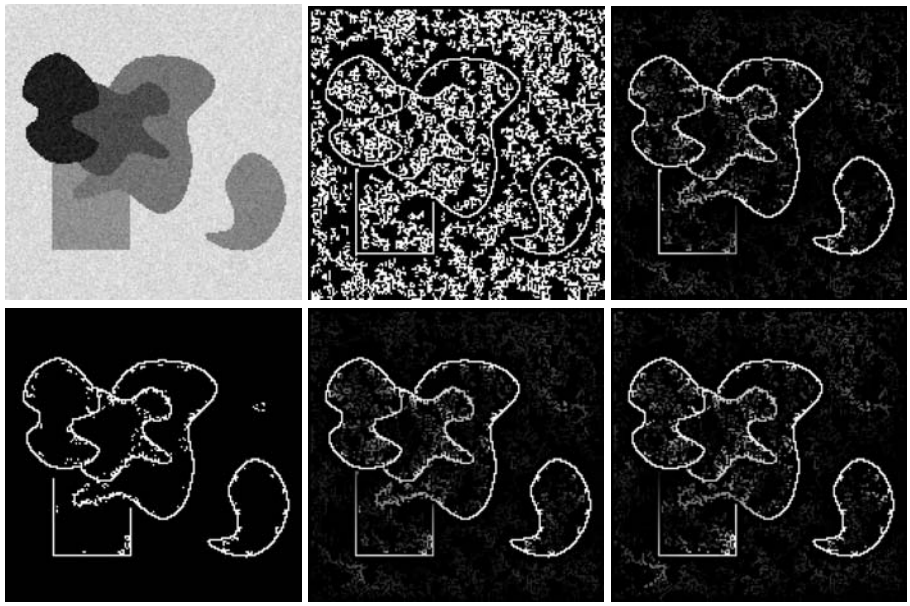


Fig. 5. A typical saliency calculation with an **Grey Level** cue done on a heavily corrupted noise: (starting from upper left, clockwise) The original image, edges, positive and negative saliencies, sum of saliency, thresholded saliency.

5 Discussion

This note presented a framework and an algorithm for calculating a well defined saliency measure which is based on estimating the length distribution and the expected length of curves. The work was motivated by the SU saliency [SU88], which, in our opinion, was build on good principles but lacked in interpretation, at least for computer vision practitioners. One result of the proposed work is that, when interpreting high curvature and gaps as factors, which decrease the probability to connect, then the SU saliency calculates the expected length of the curve on which every pointy lies. This is of course in agreement with [AB98] where the saliency of a straight line of length l and no gaps is found to be l.

The work is now in progress and we are exploring many interesting issues related to the proposed saliency mechanism. One interesting question is whether we can make the saliency invariant to scale (at least in the sense that the ratio between saliencies of two different curves stay the same over scale), and thus solve one of the problems raised in [AB98]. This is possible in principle because we are no longer limited to the curvature cue but can design other cues as well.

An even more interesting question is whether there is extensible useful saliency quality function of the distribution, which is different than the expected value (or weighted expected value). The variance, for example, is not such a function, because it is not necessary that the path of length N associated with, say, the least variance, contains a path of length $N-1$ associated with the lowest variance as well.

Fig. 6. Comparison between the expected value saliency (left) and the square root expected saliency (right). Both saliencies were thresholded so that only the points associated with saliency in the top 10 % are marked. The square root measure preserves more un-fragmented figure and contains less background texture (although the differences are not that large). This was the case also for other thresholds.

The claimed added advantage of higher reliability is not fully proved yet in this paper. Our current goal is to develop methods for characterizing the probability P_{ij} empirically and for constructing cues which are associated with a higher reliability than simply measuring the curvature. We expect to gain in the overall reliability when such cues are constructed.

The interpretation of cues as probabilities was considered in [WJ96], where the stochastic motion of a particle was used to model completion fields and elicits a saliency process as well (as observed in [WT98]). The saliency induced by this process is different than that suggested in [SU88] mainly because it is not associated with a single "best" curve but with some average of all curves in the image. Interestingly, a modified form of the proposed saliency form may be created by updating the length distribution not according to the best curve but according to the average of all curves with weights, which are just the corresponding probabilities. This way we get an alternative estimate of the length distribution (and the expected length). Which one is better? As we see it, the saliency method that we proposed here, (and that of [SU88]) is a maximum likelihood approach to saliency and length estimation, because it calculates the saliency relative to the best parameter. (This "parameter" is a path in this case). The second approach is essentially Bayesian and allows to get contributions from many alternatives. Note that both methods can be used to calculate the expected length estimate. We actually expect the second, Bayesian, method to give more visually pleasing saliency plots. Observe however, that it does not provide an estimate of the best path with it.

6 Acknowledgments

This research was supported by the fund for the promotion of research at the Technion.

References

[AB98] T.D. Alter and R. Basri. Extracting salient curves from images: An analysis of the saliency network. *IJCV*, 27(1):51–69, March 1998.

[AL98] A. Amir and M. Lindenbaum. Ground from figure discrimination. In *CVPR98*, pages 521–527, 1998.

[AM98] Laurent Alquier and Philippe Montesinos. Representation of linear structures using perceptual grouping. In *Presented in The 1st workshop on Perceptual Organization in Computer Vision*, 1998.

[CLR90] T.H. Cormen, C.E. Leiserson, and R.L. Rivest. *Introduction to Algorithms.* MIT Press, 1990.

[GM93] G. Guy and G.G. Medioni. Inferring global perceptual contours from local features. In *CVPR93*, pages 787–787, 1993.

[HH93] Laurent Herault and Radu Horaud. Figure-ground discrimination: A combinatorial optimization approach. *PAMI*, 15(9):899–914, Sep 1993.

[HvdH93] Friedrich Heitger and Rudiger von der Heydt. A computational model of neural contour processing: Figure-ground segregation and illusory contours. In *ICCV-93, Berlin*, pages 32–40, 1993.

[Low85] David G. Lowe. *Perceptual Organization and Visual Recognition.* Kluwer Academic Publishers, 1985.

[Mar82] D. Marr. Vision: A computational investigation into the human representation and processing of visual information. In *W.H. Freeman*, 1982.

[SB98] S. Sarkar and K.L. Boyer. Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. *CVIU*, 71(1):110–136, July 1998.

[SU88] Amnon Sha'ashua and Shimon Ullman. Structural saliency: The detection of globally salient structures using locally connected network. In *ICCV-88*, pages 321–327, 1988.

[Wer50] Max Wertheimer. Laws of organization in perceptual forms. In Willis D. Ellis, editor, *A Source Book of Gestalt Psychology*, pages 71–88, 1950.

[WJ96] L.R. Williams and D.W. Jacobs. Local parallel computation of stochastic completion fields. In *CVPR96*, pages 161–168, 1996.

[WT98] L. Williams and K. Thornber. A comparison of measures for detecting natural shapes in cluttered backgrounds. In *ECCV98*, 1998.