Bottom-up saliency maps – a review

(Przegląd algorytmów map wyróżnień typu bottom-up)

mgr inż. MONIKA BISKUPSKA

Instytut Maszyn Matematycznych, Warszawa

In 1985 the authors Koch and Ullman introduced a concept of a saliency map [1]. It was used to model the human attention and the shifting focus connected with sight and visual stimuli. The saliency map for a given image represents how distinctive the image regions are and in what order the eye and the nervous system process them. In their paper, Koch and Ullman explain that saliency is a measure of difference of the image regions from their surroundings in terms of elementary features such as colour, orientation, movement or distance from the eye. From later works of Koch with Itti and Niebur comes the precise definition:

The purpose of the saliency map is to represent the conspicuity – or "saliency" – at every location in the visual field by a scalar quantity and to guide the selection of attended locations, based on the spatial distribution of saliency [2].

Form this concept many algorithms arose working within Koch and Ullman framework. Various methods of selecting and analysing feature maps and then detecting local irregularities of the visual input [3] were designed to produce an accurate and computationally-efficient saliency map, where each pixel is represented by single value indicating its importance.

The purpose of this paper is to collect and present the best-known methods of calculating saliency maps, with an emphasis on bottom-up methods that work without any previous assumptions of the content of processed images. The top-down methods that are designed to work with only a narrow class of images presenting known objects are outside the scope of this review.

A model of saliency-based visual attention for rapid scene analysis (Itti/Koch algorithm) [2]

Based on the earlier publication of one of the authors [1], it was assumed that it is possible to design an algorithm reflecting the behaviour of the human eye and nervous system while looking at the image. The list of distinguishing features from the previous paper was refined and in its final form it includes colour, intensity and orientation. Itti and Koch's saliency model is one the earliest and the most widely used in later works for comparison purposes [4].

The authors present a draft of an algorithm examining the differences of colours, saturation and orientation on a given image. The saliency value is dependent on detected local spatial discontinuities of those features. Moreover, the algorithm measures relative positions of isolated fragments and attaches less saliency to elements lying close to each other and more to those far apart.

In the first phase the image is converted to nine different scales (from one to eight reduction factor). An operator denoting "across-scale difference" is introduced for the "centre-surround" operation to detect locations which stand out from their surroundings.

In the second step, the image is decomposed into a series of 42 feature maps.

The first set of feature maps consists of intensity contrast and colours double opponency maps, generated from colour channels of the image. First, the intensity *I* is calculated as an average of colour values:

$$\mathrm{I} = \frac{\mathrm{r+g+b}}{3}$$

Then, the four colour channels for red, green, blue and yellow are calculated, normalised by intensity and additionally reduced to zero in places where the intensity doesn't reach the threshold of 1/10 of the maximum value for the image. Also, the colours are modified in relation to each other according to pattern:

$$\mathrm{R} = \frac{\mathrm{r(g+b)}}{2} \quad \text{and} \quad \mathrm{Y} = \frac{\mathrm{r+g}}{2} - \frac{|\mathrm{r-g}|}{2},$$

with negative values set to zero.

With five Gaussian pyramids for those five channels for each scale of the image, six feature maps are generated using the introduced centre-surround difference operator on scaled intensity maps, denoted **I**(s,c), where $c \in \{2,3,4\}$ and $s = c + \delta$, $\delta \in \{3,4\}$ are scale reduction factors. This set of maps is made to represent the intensity contrast, which, in mammals, is detected by neurons sensitive either to dark centres on bright surrounds or to bright centres on dark surrounds [5].

Next, there are twelve chromatic double opponency maps calculated for the centre-surround differences between red-green ($\mathcal{RG}(\mathrm{c,s})$) and blue-yellow ($\mathcal{BY}(\mathrm{c,s})$) pairs of opposing channels. They are paired this way because in the centre of their receptive fields, neurons are excited by one colour (e.g., red) and inhibited by another (e.g., green), while the opposite is true in the surround. Such spatial and chromatic antinomy exists for the red/green, green/red, blue/yellow, and yellow/blue colour pairs in human primary visual cortex [6]. One double opponency map is responsible for two such pairs: ($\mathcal{RG}(\mathrm{c,s})$) for red/green and green/red and ($\mathcal{BY}(\mathrm{c,s})$) for blue/yellow and yellow/blue.

The final map set consists of the local orientation maps generated via oriented Gabor pyramids from intensity maps. There are twenty-four such maps ($\mathcal{O}(\mathrm{c,s,\theta})$, where $\theta \in \{0°, 45°, 90°, 135°\}$ is the preferred orientation) encoding orientation differences.

All of those forty-two feature maps of various scales are collapsed into three conspicuity maps for intensity $\overline{\mathcal{I}}$,, colour $\overline{\mathcal{C}}$ and orientation $\overline{\mathcal{O}}$. In the next step they are normalised using a normalising operator $\mathcal{N}$ designed to globally promote maps where a small number of strong peaks of activity is present, while globally suppressing maps which contain numerous comparable peak responses. The output saliency map is then in consequence

$$\mathrm{S} = \frac{\mathcal{N}(\overline{\mathcal{I}}) + \mathcal{N}(\overline{\mathcal{C}}) + \mathcal{N}(\overline{\mathcal{O}})}{3}.$$

The algorithm identifies the maximum value of the saliency map (SM) as a centre of attention and then uses the winner-takes-all neural network to identify secondary maximums to generate an approximation of switching the visual attention of the human eye.

The so called Itti/Koch algorithm is a very popular method of looking for salient elements on the image, mostly because of the iLab Neuromorphic Vision C++ Toolkit (http://ilab.usc.edu/toolkit/), the freely available implementation in C++ licensed under GPL and developed by Itti and his team.

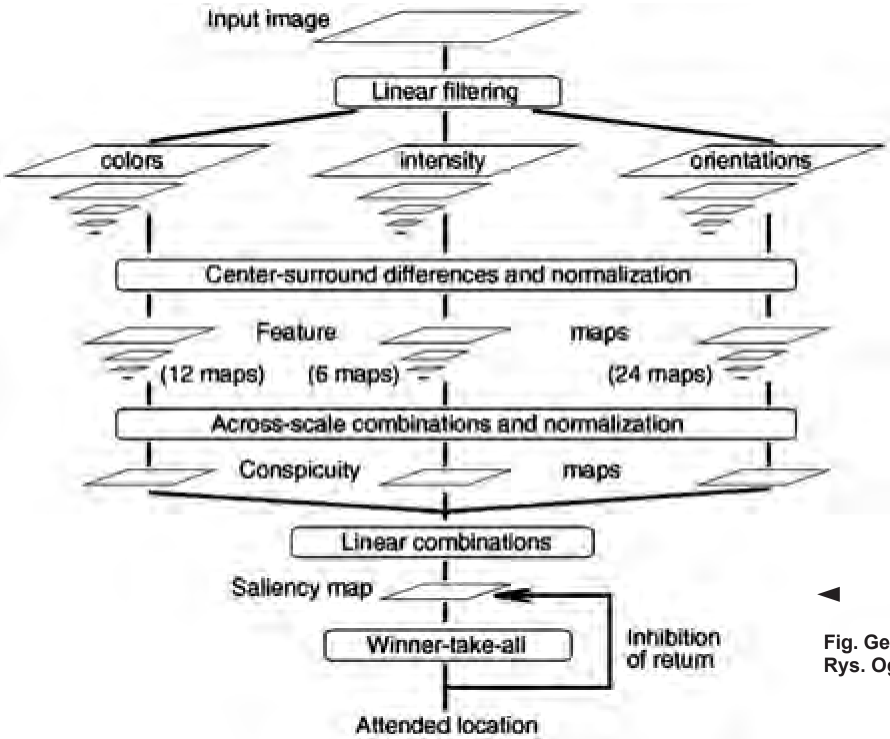

Fig. General architecture of Itti/Koch model. Source: [2]
Rys. Ogólna architektura modelu Ittiego/Kocha

Region enhanced scale-invariant saliency detection ("scale-free") [7]

The authors of this method target the fixed scale problem of the feature maps in Itti/Koch algorithm.

The approach described in this paper is a hybrid method combining pixel-based saliency detection with region information from the segmentation algorithm. The output saliency map is built by averaging the saliency value of all pixels in the resulting segments. The saliency itself is calculated using multiple scaled weighted contrast maps, as the contrast is the only feature taken into account in this algorithm.

The contrast maps are built in four steps. First the image is transformed into a perceptually uniform colour space. Then the Gaussian pyramid is built on the scaled images, with its height (n_1) corresponding to the width and height of the original image:

$$\mathrm{n_l} = \log_2\left(\frac{\min(\mathrm{w,h})}{10}\right).$$

In the next step the contrast pyramids are constructed for each of the scaled images, with contrast values defined as the weighted sum of the differences between the particular pixel and its neighbourhood:

$$\mathrm{C_{i,j,l}} = \sum\nolimits_{\mathrm{q}\in\Theta} \mathrm{w_{i,j,l}}\,\mathrm{d}\left(\mathrm{p_{i,j,l}}, \mathrm{p_q}\right),$$

where the weight is defined as $\mathrm{w_{i,j,l}} = 1 - \frac{\mathrm{r_{i,j,l}}}{\mathrm{r_{l,max}}}$, Θ is the neighbourhood of the pixel (*i*, *j*) at the scale *l*, *p* is pixel's colour, with *q* denoting the pixel belonging to the Θ, *d* is the distance between the colours in L^2 norm, and $r_{i,j,l}$ is the spatial distance between the pixel (*i, j*) at the scale *l* and the image centre, with $r_{l,max}$ denoting the maximum distance. The *r* factors are introduced to make the centre of the image in general more salient than its edges. As the last step, the saliency values are generated by summing up the contrast maps across all the scales. This method ensures that small details are salient in small-scale maps and large scale features are salient only at coarse scale. The resulting saliency values consist of both fine and coarse features of the image.

In the second phase of this algorithm, the region information is extracted from the image via image segmentation method of choice and is used to modify the raw saliency values. The saliency of each region is set to the average saliency of all pixels belonging to this region. Those averaged values of all image regions make the final saliency map.

This hybrid approach reduces high contrast differences that are not salient, like horizon lines separating the grass from the sky, and enhances the separated salient regions, especially those in the centre of the image.

Saliency, Attention and Visual Search: Saliency, attention, and visual search: An information theoretic approach (AIM) [8]

In this paper the authors present a theoretic formulation of a human visual attention model and a visual saliency algorithm basing on those formulae, named Attention based on Information Maximization (AIM). The paper itself is focused on explaining the AIM as a model of human saliency determination, with the algorithm itself presented only in a draft form.

The core of the algorithm is the feature extraction phase. It is conducted using a large number of empirical matrixes called "learned filters". Random image patches (square pixel grids with three colour channels) have been produced from pre-processed natural images (outdoor natural scenes). They are used in the independent feature extraction (ICA) phase of the algorithm as basis functions for getting basis coefficients of each patch of the image.

The algorithm itself is computationally complex to the point where the authors decided to run efficiency tests on the smaller filter patches of 21×21 pixels instead of their usual basis functions of 31×31. However, it behaves amiably in the parallel environment similar to that present in the human brain.

According to the authors' research, the ROC scores of this algorithm are significantly greater than Itti/Koch's. The Matlab code of the algorithm is available on one of the authors' websites (http://www-sop.inria.fr/members/Neil.Bruce/), but it is optimized only for small-scale images.

Dynamic visual attention: searching for coding length increments (DVA) [3]

The authors of this work propose a dynamic visual attention model based on the rarity of features. The described method is different from other presented algorithms, as it allows for continuous sampling of features. It is designed to approximate the "ebbs and flows of visual attention," in particular it tries to emulate human eye movements (saccade), and works as a temporally continuous model. The authors decide that their solution solves "critical issues when applying the Koch and Ullman framework to video or other forms of continuous observations".

The saliency here is calculated using energy function. The energy is distributed using the Incremental Coding Length (ICL), corresponding to rarity of a particular feature. Saliency is explained as "unexpected features", and those by definition of the ICL have high energy value. Image parts consisting of rare features in the process of calculating ICL receive the highest energy and become salient.

The process of calculating saliency is iterative and can be easily applied to video frames and other data that changes over time.

SUN: A Bayesian Framework for Saliency Using Natural Statistics (SUN) [4]

The authors of this paper propose a probabilistic formula for saliency. The algorithm can be based on both difference of Gaussians (DoG) feature selection known from Itti/Koch and similar methods (called by authors "biologically plausible linear filters"), or the ICA algorithm known from AIM method.

The DoG version of this method begins with the separation of three colour components of an image (red, green and blue). Next, the structures of colour channels and intensity are calculated:

$$I = r + g + b, RG = r - g, BY = b - \frac{r+g}{2} - \frac{\min(r-g)}{2}.$$

The difference of Gaussians forming the DoG filters for the (x, y) pixel is defined as:

$$DoG(x,y) = \frac{1}{\sigma^2} * \exp\left(-\frac{x^2+y^2}{\sigma^2}\right) - \frac{1}{(1.6\sigma)^2} * \exp\left(-\frac{x^2+y^2}{(1.6\sigma)^2}\right).$$

Next, the DoG filter is convolved with three colour channels (I, RG and BY) at four σ-scales (4, 8, 16 and 32 pixels) to create 12 feature response maps.

In this algorithm it is necessary to measure the probability distribution over the features. The authors decided to use natural image statistics from a collection of natural images, so the next step is to use the twelve feature response maps in conjunction with 138 images of natural scenes to generate the estimation of the probability distribution over the observed values of each of the 12 features. Using the exponential power distribution, the estimated distribution is then parameterised:

$$p(f;\sigma,\theta) = \frac{\theta}{2\sigma\Gamma\left(\frac{1}{\theta}\right)} \exp\left(-\left|\frac{f}{\sigma}\right|^{\theta}\right),$$

where Γ is the gamma function, θ is the shape parameter, σ is the scale parameter and f is the response filter output.

The feature maps can also be obtained via ICA algorithm applied to a training set of natural images. In such a case 362 feature maps were generated (from 11×11 pixel patches in 3 colour channels).

The output saliency is calculated from probability logarithm generated from either *DoG* filters or ICA-acquired patches, with the only difference in the span of the summation:

$$\log s = -\log p(F=f) = \sum_i \left|\frac{f}{\sigma}\right|^{\theta} + const,$$

where i spans from 1 to 12 in the *DoG* generated filters or from 1 to 362 in the ICA features case.

According to the authors' research, this saliency map performs on par with other known algorithms.

Graph-Based Visual Saliency (GBVS) [9]

The authors present a method of finding saliency map radically different form the classic Itti/Koch algorithm. The process is divided into three stages: extraction of features to feature vectors, generating activation maps from feature vectors and normalisation and combination of activation maps into a single saliency map. The extraction part is omitted in this work and the algorithm itself assumes there are pre-existing feature maps. Both activation and normalisation phases use Markov chain interpretation of the image.

The first step of the algorithm is to decompose the image into series of feature maps and perform the preliminary analysis by other methods. The selection of those features is also delegated to outside solutions. The purpose of this step is to reduce the image resolution by switching from "pixels" to "regions" to simplify the calculations. The remainder of the algorithm deals with a single feature map M.

The dissimilarity function is introduced to measure differences between regions (pixels) of the feature map and defined as

$$\mathrm{d}\big((\mathrm{i},\mathrm{j})||(\mathrm{p},\mathrm{q})\big) = \left|\log\frac{\mathrm{M(i,j)}}{\mathrm{M(p,q)}}\right|.$$

The map is now converted to a fully-connected directed graph $\mathbf{G_A}$ by treating each region as a node and drawing a fully-connected graph on those nodes. The weight of the directed edge is defined as directly proportional to dissimilarity of its two ends and their relative position on the map:

$$\mathrm{w_1}\big((\mathrm{i},\mathrm{j}),(\mathrm{p},\mathrm{q})\big) = \mathrm{d}\big((\mathrm{i},\mathrm{j})||(\mathrm{p},\mathrm{q})\big) * \mathrm{F}(\mathrm{i}-\mathrm{p},\mathrm{j}-\mathrm{q}),$$

where $\mathrm{F(a,b)} = \exp\left(-\frac{\mathrm{a^2+b^2}}{2\sigma^2}\right)$ and σ is a free parameter.

The next step is to define a Markov chain from the $\mathbf{G_A}$ graph by treating the nodes as states and the edge weights as transition probabilities. The weights of outbound edges of each node are normalised to 1 beforehand. In the equilibrium distribution of this chain the mass accumulates in the states with high dissimilarity with their surrounding nodes. The chain itself is ergodic because the graph that it is based on is strongly connected, so the equilibrium distribution exists and is unique. This equilibrium distribution is the basis to create an activation map A.

The authors refer to this solution as "organic", because as neurons work independently but are influencing each other, here each state/node works independently and the result is the equilibrium state achieved with the input from all of states/nodes.

The activation map has to be normalised in the process similar to the former. The map is once again converted to a graph $\mathbf{G_N}$ with nodes representing regions and edges with weight proportional to the activation map value and relative distance between nodes:

$$\mathrm{w_2}\big((\mathrm{i},\mathrm{j}),(\mathrm{p},\mathrm{q})\big) = \mathrm{A(p,q)} * \mathrm{F}(\mathrm{i}-\mathrm{p},\mathrm{j}-\mathrm{q})$$

with F defined as before. With the outbound edges weights normalised to 1, the Markov chain is defined as before, with nodes as states and weights as transition probabilities. In the equilibrium distribution the mass flows to the nodes with high activation value. This equilibrium distribution is the basis to create the output saliency map.

The authors' experimental research shows that this algorithm predicts human fixations on the salient regions more reliably than other tested algorithms (including Itti/Koch algorithm). Also, the salient regions found using this method are more cohesive than with other methods while maintaining high accuracy.

There is an official implementation of this algorithm by Harel (one of the authors) for Matlab, freely available for education purposes, along with his implementation of Itti/Koch and Image Signature algorithms (http://www.klab.caltech.edu/~harel/share/gbvs.php).

Image Signature: Highlighting Sparse Salient Regions [10]

In this work the authors present a simple yet powerful algorithm to approximate the decomposition of the image into foreground and background parts using Discrete Cosine Transform (DCT). This approximation is in turn used to generate a saliency map. Additionally, the authors present the experimental results of correlating this saliency map with reaction time of subjects in the *change blindness* experiment.

This algorithm assumes that the input image can be represented as

$$x = f + b;\ x, f, b \in \mathbb{R}^N,$$

where *f* is the foreground of the image – a feature signal, sparsely supported in the standard spatial basis – and *b* is the background of the image, sparsely supported in the basis of Discreet Cosine Transform, which means that both *f* and *b* have only small number of nonzero components. The Discreet Cosine Transform is employed to separate *f* and *b*, given only their sparseness.

The eponymous operator is defined as

$$ImageSignature(x) = sign(DCT(x)).$$

The foreground of the image is then calculated by computing the reconstructed image from the *ImageSignature* via Inverse Discreet Cosine Transform (IDCT):

$$\overline{x} = IDCT[ImageSignature(x)].$$

This is the basis on which the saliency map is generated, as the foreground part of the image is usually the most prominent. The saliency map is defined as

$$m = g * (\overline{x} \circ \overline{x}),$$

where $\overline{x}$ is the reconstructed image, ° is the Hadamard (entry-wise) product operator, * is the convolution operator and *g* is the Gaussian kernel. The authors deemed the Gaussian smoothing necessary because the support of a salient object is not only spatially sparse, but also localized in contiguous region.

The first step of the algorithm is to resize the input image to a coarse pixel grid of 64x48 and split into the three colour channels x^i. The authors tested this solution for both RGB and CIELAB colour channels and from their research the CIELAB decomposition produces better results.

On the three colour channels the *ImageSignature* is calculated, the Inverse Discreet Cosine Transform is run to reconstruct the image and finally the saliency map is generated as

$$m = g * \sum_i (\overline{x^i} \circ \overline{x^i}).$$

The standard deviation of the Gaussian blurring component is a parameter that is later used to fine-tune the results.

The authors conduct a series of experiments comparing their algorithm to other important saliency algorithms: Itti/Koch, DVA, GBVS, AIM and SUN. According to their data, their algorithm based on the CIELAB decomposition (Lab-Signature) predicts human fixation slightly better than other methods, while running much faster. The calculation speed is caused by small number of channels and calculations, as Lab-Signature runs on three channels and a single spatial scale, when the other algorithms are much more complicated: Itti/Koch and GBVS run on several feature channels and multiple spatial scales, in DVA there are 192 filters and equal number of dimensions, AIM uses 25 filters of 1323 dimensions and SUN runs with 362 filters of 363 dimensions.

Finally, the authors execute a unique experiment on the *change blindness* effect. The effect is encountered in experiment where two very similar images are presented to the observer. The second image is a slightly altered (but still visually plausible) version of the first. When the images alternate continuously, the perceptual motion or flicker is clearly noticeable. However, when there is interval between the switches where none of the images is visible (a black mask), the observer has to rely on his visual memory to spot the difference. This proves to be more difficult and the time needed to identify the change is much longer.

The authors conducted such an experiment and then tried to correlate the distance on the saliency maps of the two images (a Hamming distance $D(x^1, x^2) = \left\|sign\left(\widehat{x^1}\right) - sign\left(\widehat{x^2}\right)\right\|_0$ for Lab-Signature and GIST [11] distance for other algorithms) and the response time of experiment's subjects. According to their research, the Lab-Signature algorithm predicts the perceptual distance between such image pairs more accurately than other algorithms.

Saliency detection: A Spectral Residual Approach (SR) [12]

The authors of this method chose to move from image proper to its spectral domain. The algorithm uses log spectrum of the image derived from averaged Fourier spectrum.

The core of the method is the notion that similarity is redundancy, and redundancy should be removed to provide efficient signal consisting of only new (innovative) information. This innovation of the image is called here spectral residual and calculated as a difference between logarithmised transformation and its generalised shape.

With Inverse Fourier Transform the spectral residual is converted to spatial domain, where it is used to construct saliency map. Saliency values are then squared and Gaussian smoothed to produce final saliency map.

The final experiments conducted by authors proved the method to be fast and produce reliable results.

Salient Region Detection and Segmentation (AC) [13]

This method focuses on segmenting salient object from high-resolution maps basing on local contrast processed in different scales.

Saliency map is here obtained as a normalized, pixel-wise sum of maps in different scales. The different scaled maps are built using two regions for each pixel – inner, always same size, and outer, which size is halved for each new scale.

The comparison between this algorithm and up-scaled saliency maps from Itti/Koch solution, according to the authors, shows similar accuracy, but with much greater speed and more cohesive salient regions achieved with presented algorithm.

Frequency-tuned Salient Region Detection (IG) [14]

The authors of this paper address the problem of low resolution and general scarcity of details of saliency maps with approach that produces full resolution maps with uniformly salient object with well-defined boundaries. The solution is frequency-based. Saliency is here defined as

$$S(x,y) = \left\| \boldsymbol{I}_{\mu} - \boldsymbol{I}_{\omega_{hc}}(x,y) \right\|,$$

where $\boldsymbol{I}_{\boldsymbol{\mu}}$ is the (mean) source value of a feature and $\boldsymbol{I}_{\omega_{hc}}$ is the Gaussian blurred value of the original image. The Difference of Gaussian blurring is executed on the image to dispose of imperfections of the image, texture patterns and other artefacts.

The authors conducted a series of experiments on four others saliency methods (including previous incarnation of the saliency map algorithm of one of the authors, AC). The experiments were focusing on spatial frequency qualities of each of the saliency maps. Based on the results, the authors concluded that their method is better than others in terms of speed, ease of implementation and suitability for further segmentation and object isolation.

Saliency-based Image Retargeting in the Compressed Domain [15]

This paper presents a different approach to saliency mapping. Instead of direct operations on the source image, the authors propose taking advantage of the usual means of image delivery and storage and derive saliency information from the compressed stream of JPEG data.

The saliency value is measured on 8x8 blocks of the image, the same blocks that JPEG compression algorithm works with. From the DCT coefficients of each block the feature info is derived. DC coefficient (average energy) is used to determine the intensity and colour of the block, and AC coefficient (orientation) is used for calculating orientation of the block. Colour features are built with attention to YCrCb to RGB conversion, and double opponency channels are generated. From Y channel of AC coefficient, the orientation feature is drawn.

Differences between colour, intensity and orientation features are calculated and the model of human contract sensitivity is applied to the final result to promote salient frequencies. Final saliency map is obtained by integrating composed and weighted feature maps.

According to authors' research, the proposed algorithm is better both in accuracy and in speed than others tested (Itti/Koch, SR and IG).

Global Contrast based Salient Region Detection [16]

This method parses image histogram to measure saliency.
The saliency of a specific pixel is based on its global colour contrast:

$$S(I_k) = \sum_{\forall I_i \in I} D(I_k, I_i) = S(c_l) = \sum_{j=1}^{n} f_j D(c_l, c_j),$$

where $D(x,y)$ is the distance of the colour values x and y, n is the number of different colours used in the image and f_j is the frequency of colour c_j. The authors propose a method to reduce the number of considered colours by quantization and elimination. In effort to reduce quantization artefacts, the colour space smoothing is introduced to additionally refine saliency value.

There is additional method of measuring saliency presented in this paper, the region contrast. The region contrast is applied to take into consideration the spatial properties of the image. The source image is separated into regions, and the saliency of the region is a weighted sum of all region pixels saliency to all other regions. Distant regions have less weight than close regions.

First method is faster, but the second produces higher quality results. Both are inefficient when provided with a highly textured image, but on a "natural image" they behave appropriate, being fast, simple and efficient.

Summary

Various existing saliency examination methods try to provide accurate machine vision which emulates human eye and nervous system. The first algorithm of Itti/Koch is used as a guideline for most if not all research, as a tool to compare or contrast other methods.

There exist methods derived from biologically plausible architectures like Itti/Koch, but also those without a biological component, based on pure mathematics. Both image space and frequency space, but also less popular sources like compressed JPEG streams can be sampled for saliency information. Most of the methods provide low resolution maps, but there are some which purposefully produce full-res results. Some of methods concentrate on salient pixels, or parts of image, when others output salient regions or even fully separated salient objects with well-defined boundaries.

All authors present experiments providing results which show that their method is on par or better than Itti/Koch in terms of accuracy, speed, resources consumed or ease of implementation, frequently more than one of those.

References

[1] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry.," Human Neurobiology 4, pp. 219–227, 1985.

[2] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis.," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, p. 1254–1259, 11 1998.

[3] X. Hou and L. Zhang, "Dynamic Visual Attention: Searching for coding length increments," in Advances in Neural Information Processing Systems, D. Koller, D. Schuurmans, Y. Bengio and L. Bottou, Eds., 2009, pp. 681–688.

[4] L. Zhang, T. K. Marks and M. H. Tong, "SUN: A Bayesian Framework for Saliency Using Natural Statistics," Journal of Vision, vol. 8, no. 7, pp. 1–20, December 2008.

[5] A. G. Leventhal, Ed., The Neural Basis of Visual Function: Vision and Visual Dysfunction, vol. 4., Boca Raton, Fl: CRC Press, 1991..

[6] S. Engel, X. Zhang and B. Wandell, "Colour Tuning in Human Visual Cortex Measured With Functional Magnetic Resonance Imaging," Nature, vol. 388, no. 6,637, p. 68–71, July 1997.

[7] F. Liu and M. Gleicher, "Region enhanced scale-invariant saliency detection," in ICME '06: Proceedings of IEEE international conference on multimedia and expo, 2006.

[8] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," Journal of Vision, vol. 9, no. 3, pp. 1–24, March 2009.

[9] J. Harel, C. Koch and P. Perona, "Graph-Based Visual Saliency," Proceedings of Neural Information Processing Systems (NIPS)., 2006.

[10] X. Hou, J. Harel and C. Koch, "Image Signature: Highlighting Sparse Salient Regions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 1, pp. 194–201, January 2012.

[11] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," International Journal of Computer Vision, vol. 42, no. 3, pp. 145–175, 2001.

[12] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach," in IEEE Conference on Computer Vision and Pattern Recognition, 2007.

[13] R. Achanta, F. Estrada, P. Wils and S. Susstrunk, "Salient region detection and segmentation.," in International Conference on Computer Vision Systems, 2008.

[14] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, "Frequency-tuned Salient Region Detection," IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009), pp. 1597–1604, 2009.

[15] Y. Fang, Z. Chen, W. Lin and C.-W. Lin, "Saliency-based image retargeting in the compressed domain," in Proceedings of the 19th ACM international conference on Multimedia, New York, 2011.

[16] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang and S.-M. Hu, "Global contrast based salient region detection," in Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, 2011.