Single pixel information content

Emanuel Diamant
VIDIA-mant, POB 933 Kiriat Ono, Israel 55100
(emanl@012.net.il)

ABSTRACT

Inspired by the analogy between computer-based visual systems and their biological counterparts, we propose to introduce two new concepts: the concept of a ***pixel's receptive field*** (analogous to neural cells' receptive fields in the eye's retina), and its derivation - the concept of ***information content of a single pixel*** (analogous to the output activity of the retina's neural cells). Exploiting these new concepts, we suggest a quantitative measure for the dissimilarity between a pixel and its surrounding neighbors, which we define as a measure of pixel's information content. With such a measure at the hand, many image processing tasks that usually require (for their successful accomplishment) some information related assumptions could be reformulated and redesigned to gain new and unknown computation efficiency of image features detection, description and discrimination.
On this basis, new image processing techniques could be devised.
Lena images processed by applying some of these techniques are presented for illustration purposes.

Keywords: pixel information content, image information content, image analysis, image understanding, image description.

1. INTRODUCTION

Recent years have witnessed tremendous growth and proliferation of visual information in our surrounding environment. Huge amounts of digital image data are constantly accumulated and, consequently, become publicly available on the World Wide Web, in large specialized archives and libraries (military, medical, geo-science, art, etc.), and even in private repositories. Effective access_to, retrieval_from, management_of and utilization of such immense data masses will be extremely difficult without new image-processing tools specially designed for such purposes. Obviously, it is supposed that such tools will be capable to analyze, interpret and make relevant decisions regarding image information content. That means, be capable to detect, extract and process high-level semantic information contained in the images.

Today's computer-based image-processing systems are definitely lacking such capabilities, (or possess them in very premature and elementary forms). Only humans have the exclusive privilege and the unlimited capacity to perceive and apprehend infinitely variable content of their visual input, (as they do that in their everyday life: ordinary, instantly and effortless). It is clear, that the efforts in developing next generation intelligent image-processing tools must be tightly connected with progress in human visual system peculiarities understanding.

There is another reason, that strengths the need for better understanding of the human vision properties. Before any image-understanding task attempts to be launched, the image must be brought in the system's proprietorship, (by transmission from a remote source or by transfer from a local storage space). It does not matter what way the image will be delivered. What does matter, is: for the sake of proper bandwidth or space utilization, the image is always initially compressed, and then, subsequently, decompressed at the system's input. To achieve reasonable compression rates, part of the source image data must be, eventually, discarded.

How human's ability to comprehend image content is affected by this inevitable data/information loss? Today's image compression technologies were developed regardless to this kind of questions. Most of them rely on abstract mathematical models of image data that have nothing in common with human's visual perception.

Image Processing: Algorithms and Systems II, Edward R. Dougherty, Jaakko T. Astola, Karen O. Egiazarian, Editors, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5014 (2003) © 2003 SPIE-IS&T · 0277-786X/03/$15.00

Meanwhile, within the established modus operandi, humans by themselves become an indispensable part of the computer-based image-processing systems. They are the system's ultimate end-user and, concurrently, the system's teammate that is deeply involved in image content evaluation and management. To resolve this entanglement, the design of next generation compression codes must explicitly take into account the natural requirements of the human visual system.

The above mentioned arguments can easily explain the extreme interest that computer vision society displays now to psychological, psychophysical, neurophysiological, and other biological science studies of human visual perception and human visual system performance. It is generally understood that shaped by millions of years of natural evolution human visual system has achieved performance levels that are orders of magnitude better than the most sophisticated human-made visual systems. By elucidating the computational principles, that make this level of performance possible, it is widely believed that it may be possible to increase and significantly improve the qualities of contemporary computer-based image-processing systems.

2. HUMAN VISUAL SYSTEM'S CUES

When we start to examine the human visual system (essentially, the huge amount of papers resulted from extensive research and scientific work done in the last years[1,2,3,4]), the first surprise was the striking dissimilarity between input front-ends of human and human-made visual systems. Maybe, only the two-dimensional photo-sensor arrangement, found in the CCD camera and in the eye retina as well, can be accounted for a feature that both of the systems possess.

Eye's retina is a highly structured network of neurons that covers the inner surface of the eye. It is composed of three primary layers: the photoreceptors, the bipolar cells, and the ganglion cells layers. Retina's output, arranged as a bundle of one million axon fibers (also called the optic nerve), is linked to the higher sequential layers of the visual processing hierarchy. Thus, anatomically and functionally separated from the rest of the brain, the retina appears as a "distant", "stand alone", almost autonomous (no feedback paths from the higher levels) unit of the human visual system.

Light patterns projected on the retina are sensed by approximately 100 million photoreceptors of its input layer. This is the only place in the system, were the input information appears in an analog form. It is immediately "digitized" by the nearest sequential layer, and from this point on, the input information is encoded and advanced through all the layers of the processing hierarchy as a train of neuronal spikes.

It must be mentioned explicitly: the retina is the first and the only stage in the system where the input information is detected, extracted, evaluated, selected and encoded for further processing. All subsequent, higher level stages in the chain acquire their inputs from the retina. There are no other sources of visual information connected to the system. All, what the high-level part of the system (the cognitive part) has at its disposal, is determined by the mechanisms of early neural processing that take place in the retina. We would like to restate this in other words: the success of high-level image processing is entirely determined by low-level information primitives (and their subsequent transforms/representations) which are the outcome of low-level information processing accomplished by the retina.

It is obvious, that from this point on, our attention is shifted to the retinal processing.

The key principals of retinal processing are hierarchical organization and progressively growing information abstraction accomplished at each hierarchical level. Neuronal interactions are the principal components of this process, and two types of interaction can be distinguished: intra-level and inter-level interactions. Intra-level interaction implies that cells reciprocate between them within a given layer. In the retina, only the nearest neighboring cells are permitted to interact. Long distance connections merely do not exist. Inter-level interaction means that each cell of a given layer takes its input from a spatially localized region in the previous layer, called the receptive field of a cell. The receptive fields usually have a characteristic center/surround organization, have different sizes, and "tile" the retina so that the receptive fields of adjacent cells completely cover the retina without significant overlap.

The question "What is the product of such interplay?" is much more complicated than issues of retina's structural organization. Despite a host of empirical evidences accumulated over the years, a clear understanding what retinal processing actually is has not been attained yet. Classical approaches see it as a set of specified filters probing features in the visual input (intensity, shape, orientation, color, time, motion, etc.). On the other hand, there are arguments (and evidence) that these approaches are not always applicable. It is known, that retina extracts and promotes signals that must be related to local saliency, and which are indifferent to (classical) features' attributes. (For example, the pre-attentive saliency map[5]).

We are ready to incline towards those approaches, that claim that feature's attributes (color, brightness, texture, etc.) are less important than the magnitude of attribute's local change. It seems to us, that the vague concept of salience can be easily replaced by a more palpable concept of "informative importance", which in turn, can be further warped to a concept of an information content (applied to a given place in the retinal map). Information content is a more abstract concept than the magnitude of change in feature's expression. More precisely, it is the only abstract concept, and therefore, far more suitable to fit mechanisms of attention focusing. It is clear, that such an approach can be farther expanded. It is tempting to think that retinal processing generally, from the very beginning attempts to assess information content of the features revealed in the visual input. And those estimations provide the basis for subsequent reasoning about what and when to promote in the processing chain, up to the output to the higher levels. (It is also tempting to think that considering only intra-level interactions among neighboring cells, it will be possible to assess the information content at a local image position.)

Inspired by the above mentioned considerations, we propose to model the low-level image information processing in a way that resembles principal components of the early retinal processing (as we understand and interpret them), and, at the same time, can be applicable to input front-ends of computer-based visual systems. We introduce, thus, two new concepts: a concept of a ***pixel's receptive field*** (analogous to the retinal cells' receptive fields) and a concept of ***information content of a single pixel*** (in brief – **icspe**, pronounced like X-pe), an analog to retinal cells' information extraction activity.

3. ICSPE – IMPLEMENTATION ISSUES

Following retinal analogy, **icspe** can be defined as a measure for information about signal's discontinuity at pixel location. In computer vision context, "signal" is the pixel's gray-level intensity (corresponding to the luminance of the appropriate scene point), and "signal's discontinuity" is the change in intensity levels in the pixel's surrounding. If this change is spatially consistent, presence of an edge is assumed at the pixel location. (For the purpose of our discussion, we consider only monochrome grayscale images. Although, in the real world other image attributes, like color or texture, can contribute to image information content as well, their contributions could be evaluated in a similar manner.)

So, we can define information content of a single pixel as an expression of dissimilarity between it and its nearest neighbors. In other words: discontinuity in intensity levels at the pixel position. In computer vision terminology: "edginess" at the pixel location. A quantitative estimation of this information content can be provided by the measure of uncertainty in pixel-to-neighbors interaction. Two sources of such uncertainty must be considered: luminance uncertainty and topological uncertainty. Therefore, modeling pixel's receptive field as a pixel-centered nxn array of its closest neighbors, **icspe** can be expressed as:

$$I(s) = I(g) \cdot I(t)$$

where $I(s)$ – is the information content of a pixel, $I(g)$ – is the local luminance uncertainty, and $I(t)$ – is the local topology uncertainty.

There are various forms in which luminance uncertainty can be defined. As Van Essen[1] and Meister[3] pointed out, rather than response to an integral luminance level, neural cells primarily respond to the difference in intensity levels between a cell and its surrounding neighbors, (the so-called contrast sensitivity). For our purposes, contrast can be defined as the difference between central pixel gray-level intensity and the mean intensity of its neighbors. Luminance uncertainty, in this case, will be computed as:

$$I(g) = \sum_{1}^{8} |g_i - g_c|$$

Here, g is the gray level intensity, respectively, of the central pixel and of its eight neighbors.

Spatial properties of intensity patterns in an image are asserted by the topology of their borders. To consider the contribution of this information (reduced to the pixel's receptive field level), a map, reflecting spatial interrelations among the neighboring pixels, must be provided. For this purpose, in parallel to $I(g)$ calculation, the mean gray level intensity of the surrounding pixels is computed (essentially, its predecessor – the sum of intensities of the eight surrounding neighbors), and the sign of the difference between it and the intensity of the central pixel (multiplied, correspondingly, by 8) is established:

$$sign = \sum_{1}^{8} g_i - 8g_c$$

The sign is marked as one, when central pixel intensity is equal or greater than the mean intensity, and as zero, if else. This zero/one result is memorized in a special one-bit-depth map, and the topological uncertainty is afterwards figured out using this map:

$$I(t) = p_c(1 - p_c)$$

where p is the probability to find surrounding pixels in the same state (sign) as the central pixel.

Two extreme situations, when all neighbors and the central pixel are in the same state (a flat region, $p=1$) and none of the neighbors is not in the central pixel's state (an outlier, $p=0$), have the same intuitively expected result – $I(t)=0$ and, consequently, $I(s)=0$ – there is no "edginess" at that location.

4. EXPLOITING ICSPE

Following *I(t)* calculations, a compound **icspe** measure *I(s)* for every image pixel can then be computed and stored in an appropriated memory buffer. In parallel, an integral of **icspe** values is accumulated with the purpose to determine (at the end of the round) the **icspe** mean value. Given this mean value, a memory buffer for histogram counting is allocated, with total length of three **icspe** means, partitioned into 20 equal histogram bins. (The width of a single bin does not contain fractions, and so, it is always rounded to the nearest larger integer.) Then, a histogram is accumulated by searching the whole *I(s)* buffer, and in accordance to the following rule: if the current **icspe** value is equal or greater than the bin's lower bound, the value is added to this bin counter.

As a result, a new, previously unattainable knowledge about information content distribution among various image ingredients becomes palpable: one can see exactly what part of the total image information (content) is carried out by pixels with **icspe** values equal or greater than a given bin's floor.

It is clear, that this knowledge can be fruitfully utilized in design of new information-content-distribution-supported image-processing tools. The majority of image-processing tasks, to reach successfully their goals, must be provided with some initial assumptions about information content distribution among image constituents. Unfortunately, such information is usually not available, and the assumptions are made based on heuristics, previous experience, intuition, or are set up *ad hoc*. With **icspe** at the hand, the things look entirely different.

As an example, let us consider such a common image-processing task as edge detection and localization. It is clear from the earlier explanations, that **icspe** is a quantitative expression of "edginess". Thus, it is reasonable to exploit it for such a task. When observed **icspe**'s value exceeds some accepted threshold, that is an evidence of an edge presence. However, establishing an "acceptable threshold" is a hard and a sophisticated problem by itself. Standard edge thresholding techniques produce not stable and ambiguous results because of a wrong assumption that image intensities could be treated as bimodal[6], and "true" edges in an image are really exist and could be defined. Using **icspe** provides a unique opportunity to introduce multi-level thresholding, which does more realistically reflects variable information content of different edges. **Icspe** gracefully endows us with fuzzy thresholding capabilities that are far more suitable for "soft" human-like image analysis. An example of a three-level edge importance grade thresholding can be find in the section of graphical illustrations.

Icspe's contribution to another edge related problem (known as edge localization uncertainty) must be specially emphasized. Because edge is an abstraction, a symbolic line placed between two image parts that does not belong to either of them, the common practice of setting up one-pixel-wide (specially tinned) edge marking line is always a headache and a dilemma. Contemporary edge localization is problematic and controversial. On the contrary, **icspe**-based edge localization certainly outlines edges as two coexisting, closely spaced lines. Because edge is an intensity gradient, a sign of lower/higher gradient side (already available from the sign map of $I(t)$ computations) can be attached to each of the lines. This way, not only the edge localization problem receives a proper solution, but a new sort of edge descriptors appears, with new, previously unknown and unmatched descriptive capacity. An example of double-line edge mapping can be find in the section of graphical illustrations.

5. CONCLUSIONS

We have proposed a measure for estimating the information content of a single pixel. It is a simple, and, thus, computationally inexpensive approach to low-level image-information processing, which leads to effective low-level image features discovery and description.

Evaluation of image information content is a new, extremely important topic in current attempts to provide tools for visual databases/ video collections management that would possess human-like capabilities. Majority of today's approaches to image information content assessment attempt to reveal and to exploit global information content estimations related to the entire image area. Such approaches can be only partially successful and apply to a restricted domain of image content. We think, that like the human visual system, the future tools must primarily rely on low-level (physical) information contained in the images.

We think, we do the right thing, trying first of all to reveal the elementary, low-level feature-information-related representative image components, which will then, later, appropriately serve the higher levels of image understanding. To our knowledge, no other approaches, following such processing manner, have yet been attempted.

REFERENCES

1. D. C. Van Essen, C. H. Anderson, D. J. Felleman, ''Information Processing in the Primate Visual System: An Integrated System Perspective'', *Science*, **255**, pp. 419-423, 1992.

2. D. K. Warland, P. Reinagel, and M. Meister, ''Decoding Visual Information From a Population of Retinal Ganglion Cells'', *The Journal of Neurophys.*, **78** (5), pp. 2336 – 2350, 1997.

3. M. Meister and M. J. Berry II , "The Neural Code of the Retina", *Neuron*, **22**, pp. 435 – 450, Mar. 1999.

4. R. Van Rullen, S. J. Thorpe, "Rate Coding Versus Temporal Order Coding: What the Retinal Ganglion Cells Tell the Visual Cortex", *Neural Computation*, **13** (6), pp. 1255 – 1283, 2001.

5. L. Itti, C. Koch, and E. Niebur, “A Model of Saliency-based Visual Attention for Rapid Scene Analysis”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **20**(11), pp. 1254 – 1259, Nov. 1998.

6. A. Brink, “Using spatial information as an aid to maximum entropy image threshold selection”, *Pattern Recognition Letters*, **17** (1), pp. 29 – 36, Jan. 1996.

Some examples of image information content elicitation using icspe-based techniques.

Upper-left corner – is the original Lena image, 256x256 pixels, 256 gray levels.

Bottom-left side – the map of perceptually important image parts; in dark-gray are the most prominent parts, carrying more than 50% of image information content, in half-gray – the less important parts, carrying between 50% and 70%, in light-gray – the lowest importance parts, 70%-85%.

In the upper-right corner – The edge localization map. In dark-gray are the darker sides of the edges, in light-gray are brighter