Available online at www.sciencedirect.com
ScienceDirect
Signal Processing: *Image Communication* 22 (2007) 583–590
SIGNAL PROCESSING: IMAGE COMMUNICATION
www.elsevier.com/locate/image

Modeling human-like intelligent image processing: An information processing perspective and approach

Emanuel Diamant

VIDIA-mant, P.O. Box 933, 55100 Kiriat Ono, Israel

Received 29 May 2007; accepted 30 May 2007

Abstract

We live in the Information Age, and information has become a critically important component of our life. The success of the Internet made huge amounts of it easily available and accessible to everyone. To keep the flow of this information manageable, means for its faultless circulation and effective handling have become urgently required. Considerable research efforts are dedicated today to address this necessity, but they are seriously hampered by the lack of a common agreement about "What is information?" In particular, what is "visual information"—human's primary input from the surrounding world. The problem is further aggravated by a long-lasting stance borrowed from the biological vision research that assumes human-like information processing as an enigmatic mix of perceptual and cognitive vision faculties. I am trying to find a remedy for this bizarre situation. Relying on a new definition of "information", which can be derived from Kolmogorov's complexity theory and Chaitin's notion of algorithmic information, I propose a unifying framework for visual information processing, which explicitly accounts for the perceptual and cognitive image processing peculiarities. I believe that this framework will be useful to overcome the difficulties that are impeding our attempts to develop the right model of human-like intelligent image processing.

Keywords: Visual information; Physical and semantic image information; Knowledge and information relationship

1. Introduction

The explosive growth of visual information in our surroundings has raised an urgent demand for effective means for organizing and handling these immense volumes of information [1]. Because humans are known to be very efficient in such tasks, it is not surprising that computer vision designers are trying again and again to get answers for their worrying problems among the solutions that human visual system has developed in course of millions of years of natural evolution. Nearly half of our cerebral cortex is busy with processing visual information [2], but how it is done "in vivo" remains a puzzle for many generations of thinkers, philosophers, and contemporary scientific researchers.

Nevertheless, a working theory of human visual information processing has been established about 25 years ago by the seminal works of Marr [3], Treisman [4], Biederman [5], and a large group of their associates and followers. Since then it has become a classical theory, which dominates today in all further developments in the field. The theory considers human visual information processing as

E-mail address: emanl@012.net.il

0923-5965/$ - see front matter
doi:10.1016/j.image.2007.05.007

an interplay of two inversely directed processing streams. One is an unsupervised, bottom-up directed process of initial image information pieces discovery and localization. The other is a supervised, top-down directed process, which conveys the rules and the knowledge that guide the linking and binding of these disjoint information pieces into perceptually meaningful image objects.

In modern biological vision research, this duality is referred to as perceptual and cognitive faculties of vision. In computer vision terminology, these are the low-level and high-level paths of visual information processing. Although Treisman's theory [4] definitely positions itself as "A Feature-Integration Theory", the difficulties in defining proper rules for this feature integration have impelled a growing divergence between perceptive and cognitive fields of image processing [6]. Obviously, that was a wrong and a counter-productive development, and human vision researchers were always aware of its harmful consequence [7]. For this reason, the so-called "binding problem" has been announced as a critical exploration goal, and massive research efforts have been directed to its resolution [8]. Unfortunately, without any discernable success.

In computer vision, the situation is even more bizarre. Thus far, computer vision community was so busy with its everyday problems that there was no time to raise basic research ventures. Principal ideas as well as their possible solutions are usually borrowed from biological vision research. Therefore, following the trends in biological vision, for decades computer vision R&D has been deeply plunged into bottom-up pixel-oriented image processing. Low-level image computations have become its prime and persistent goal, while the complicated issues of high-level processing were just neglected and disregarded.

However, it is impossible to ignore them completely. It is generally acknowledged that any kind of image processing is unfeasible without incorporation into it the high-level knowledge ingredients. For this reason, the whole history of computer-based image processing is an endless saga on attempts to seize the needed knowledge in any possible way. The oldest and the most common ploy is to capitalize on the domain expert's knowledge and adapt it to each and every application case. It is not surprising, therefore, that the whole realm of image processing has been, and continues to be, fragmented according to the high-level knowledge competence of the experts in the corresponding domains. That is why we have today: medical imaging, aerospace imaging, infrared, biologic, underwater, geophysics, remote sensing, microscopy, radar, biomedical, X-ray, and so on "imagings".

The advent of the Internet, with huge volumes of visual information scattered over the web, has demolished the long-lasting custom of capitalizing on the expert's knowledge. Image information content on the Web is unpredictable and diversified. It is useless to apply specific expert knowledge to a random set of distant images. To meet the challenge, the computer vision community has undertaken an enterprise to develop the so-called content-based image retrieval (CBIR) technologies [9,10]. However, deprived of any reasonable sources of the desired high-level information, computer vision designers were forced to proceed in only one possible direction of trying to derive the high-level knowledge from the available low-level information pieces [11,12].

In doing so, computer vision designers have once again demonstrated their reliance on biological vision trends and fashions. In biological vision, a rank of theoretical and experimental work has been done in order to support and to justify this above-mentioned tendency. Two ways of thinking could be distinguished in this regard: chaotic attractors modeling [13,14], and saliency attention map modeling [15,16]. We will not review these approaches in details. We will only note that both of them presume low-level bottom-up processing as the most proper way for high-level information recovery. Both are computationally expensive. Both definitely violate the basic assumption about the leading role of high-level knowledge in the low-level information processing.

It will be a mistake to say that computer vision people are not aware of these discrepancies. On the contrary, they are well informed about what is going on in the field. However, they are trying to justify their attempts by promoting a concept of a "semantic gap", an imaginary gap between low- and high-level image features. They sincerely believe that they would be able to bridge it some day [17].

It is worth to mention that all these developments—feature binding in biological vision and semantic gap bridging in computer vision—are evolving in atmosphere of total indifference to prior claims about high-level information superiority in the general course of visual information processing. Such indifference seems to stem from a very loose understanding about what is the concept of