Medical Computer Vision – Quo Vadis?

Or: Towards Bridging the Gap Between Theory and Practice

H. Siegfried Stiehl

1 What is it all about?

MCV or else in programmatic terms: CV for medicine - is it i) ... a key technology which is believed to contribute to a nation's economic leading edge in an ever increasing way in the next millenium? or ii) ... a camouflage of military computer vision since medical high tech is relevant for the so-called battlefields of the future as conceived and thus funded by the military-industrial complex? or iii) ... solely the application of well-understood methods from e.g. applied mathematics and mathematical physics to medical imagery for the sake of formal notations only? or iv) ... an opportunity to upgrade high-tech medical imaging products in a year-wise manner thus allowing companies to keep a share in the market? Or is it - coaxingly spoken - the good side of CV solely devoted to humanity? or ... A scientific discipline erected to crack the theoretical tough nuts associated with processing and analyzing digital representations of the continuous human body with the aid of a Turing machine? or ... The missing computational link between noisy measurements of physical properties of continuous phenomena in patients on a regular discrete grid (vulgo, raw image data) and task-specific image representations tailored to be better suited to match or even enhance the perceptual and cognitive abilities of medical users than raw image data alone?

Obviously, no single, unique answer to the genuine what question can be given due to the rather large number of facets like e.g. individual perspective, institutional context, and sciento-sociological embedding. However, viewing MCV from the ivory tower's vantage point one feels obliged to come up with an answer to the what question be it rather provisional or not.

What is MCV then? Seen from a consensual academic perspective the scientific crux certainly is the quest for (in the sense of Marr's seminal contribution) computational theories - a key issue in CV along with the design of appropriate low-, intermediate-, and high-level representations as well as the incorporation of explicitely encoded knowledge in the different levels of a representational hierarchy. In this context computational theories nicely play the role of formal transformations between representation levels while incorporating level-specific knowledge. According to Marr one crucial level of understanding complex information processing systems e.g. for visual data is the one of a computational theory - an approach which tacitly implies the primacy of theory from which algorithmization, implementation, and experiment strictly derive. In the ideal case a computational theory X is a synonym for some lines (or pages) of maths along with a proof revealing that X holds for the entire space of admissible images, which in turn is also a key to achieving genericity of algorithms for visual data. Genericity as a principal desideratum of CV research in this context means: By appropriately orthogonalizing the multi-dimensional space of problems to be solved in MCV (e.g. according to clinical disciplines, incidence of diseases, anatomical domains, problem classes, complexity of clinical cases, image degradation characteristics, types of potential users, degree of automation, level of interaction, etc.), the state-of-the-art in terms of prevalent single, successful solutions of practical problems may well be seen as scattered points only in that space, whereas desired generic solutions may be conceived of as possibly overlapping hypervolumes in the sense that e.g. algorithms may be applied to a variety of domains, a number of organs, a range of image classes, etc.

Seen from a programmatic point-of-view CV in general is a methodical approach to bridging the so-called signal-symbol gap, viz relating raw image data through computational theories across the levels of representations to meaningful, task-supportive image representations up to the level of symbols denoting objects and their relations in the physical world from which the image was acquired. In this view a computational theory transforms a representation, solves a visual problem e.g. edge detection or motion recovery, and supports a visual task e.g. finding organ boundaries or measuring heart motility. In the domain of MCV such representations may serve a variety of purposes and tasks e.g. accurate measurement of lesion shape and volume, generation of probabilistic anatomy atlasses, surgery planning in an augmented reality environment, content-based image retrieval from medical mass image data archives, control of a surgical robot acting in three-dimensional patient space. For the derivation of symbolic representations from the level of signals the exploitation of prior knowledge in various forms and at numerous levels of complexity is crucial. Knowledge is involved through models of image formation, local models of intensity functions, knowledge of the visually salient structure of objects, differential geometry models of anatomical objects, models of the spatial layout of the human body, etc. In view of research in the next millenium it will be thrilling to see how the necessary theoretical and methodical connex between close-to-signal, low-level methods and symbolic, high-level approaches - both assumed to be theoretically well-understood then - will evolve, a connex being one of the grand challenges of CV.

Following the credo that nothing is more practical than having a good theory for the purpose of mastering the complexity of designing reliable working systems, the very starting point for an engineer faced with the particular problem of designing a CV system with predictable performance should be the existence of proven computational theories in conjunction with data and spec sheets characterizing the performance of resulting algorithms under well-controlled but domain-relevant experimental conditions. Put in other words, designing for instance an as-automated-as-admissible object recognition system on the basis of less-understood, or even non-understood, low-level methods runs at risk of unnecessarily increasing system complexity, e.g. through the necessity of complex processes of reasoning about success and failure of low-level processes which in turn requires probably evenly complex knowledge representation schemes. By the way, the same is true also in the realm of designing user-centred interactive visual information systems for e.g. medical image analysis since it is an impediment to the human user to be forced to painstakingly correct the errors of, say, dumb and dull low-level MCV methods. Evidently, even from an ergonomic point-of-view computational theories are indispensable since they contribute to the design of reliable and easy-to-control medical image analysis systems with predictable performance which fully support the specific problem solving task of a cognizant user, a task which in the case of the medical domain is always part of a complex human perception-to-action cycle.

2 So, where to go?

About one decade ago Yannis Aloimonos complained, "Unfortunately, there is a disconcerting lack of visual systems which perform well in real-world environments, particularly when compared to the amount of mathematical theory published on the subject.? - a nowadays still valid complaint which not only holds for CV in general but in particular also for the safety-critical case of MCV. One reason for this unfortunate situation is clearly the fact that the experimental basis of CV per se is still rather weak. The issue of establishing a scientifically grounded experimental methodology - which also touches upon sciento-theoretical and epistemological problems as well - has

Künstliche Intelligenz, Heft 3/00, Seiten 30-31,
ISSN 0933-1875, arenDTaP Verlag, Bremen