IMAGE AND VIDEO ANALYSIS – TRENDS AND CHALLENGES: INTERACTIVE PROCESSING AND VISUALIZATION

Riccardo Leonardi

University of Brescia, Dept. of Electronics for Automation
Via Branze, 38 I-25123 Brescia (BS), Italy. Riccardo.Leonardi@ing.unibs.it

Multimedia research and in particular image and video analysis has been targeting means to better represent information for variety of tasks and applications such as pattern recognition, coding, content browsing, information filtering, ... Adequate modeling of the underlying physical phenomena has been an essential step. In addition new mathematical tools such as wavelets have brought great sophistication to the latest methods. However most successful techniques have been designed to target very well defined objectives, in specialized contexts. Adaptation to evolving contexts still remains far from reach, as a priori knowledge modeling is very elaborate.

This may be eased if the control of the algorithmic process is left to the end user. The perspective that end users or end devices take an active role and guide the algorithmic process has been given little attention so far. Often the objective dynamically evolves as new information becomes available. In our view, signal processing and computer vision tools simply represent means to extend the ability of man to understand and organize visual content. Indeed with the explosion of visual information sources, most of the problem becomes to discriminate between relevant and irrelevant material.

As such, the challenge is not only limited to the design of single effective algorithms for a unique scope. Preprocessing information to extract low-level content description can often facilitate any elaborate task that would take place at a later stage. Raw information can thus be *enriched* with low levels tags (metadata) that ease the processing for more sophisticated goals. More "real-time" recognition processes can then take place once the information is dealt with for higher level semantic characterization, such as expert annotation.

In this framework the ability of users to actually control the visual analysis process additionally requires two complementary conditions:

- adequate user training
- ease in interaction.

For the first scope, the user should have a clear understanding of the semantics of low-level metadata that may be attached to the content, and possibly an indication of the performance of the associated extraction method. If a user can guide any subsequent processing task, he/she should also understand which algorithm is being used, so that he/she can optimally control the processing strategy. The user should also be able to constantly assess the performance of the processing taking place, which imposes that he/she is skilled for the task at hand. As a support, the user can improve his/her skills by having the vision system propose a training-by-example strategy.

Ease of interaction is needed to achieve good control of any semi-automatic processing task. This requires effective switch between point of views for the task at hand, i.e. have a multivariate presentation of the processing results. For example, time evolving information such as image sequences could be alternatively summarized in a single view, enabling an instant perception of information correlation. Conversely static information could be viewed dynamically, enabling the perception of the 3-dimensional structure of a scene. Visualization of data in high dimensional spaces also bring new perspectives in information representation and/or data synthesis (e.g., plenoptic viewing).

Real-time feedback is finally essential to fully extend the human skills in information mining. Indeed this allows the end user to reiterate any on-going processing with a new parameters or to adequately preprocess information.