IMAGE AND VIDEO ANALYSIS: TRENDS AND CHALLENGES

Aggelos K. Katsaggelos

Northwestern University
Department of ECE
Evanston, IL 60208
aggk@ece.northwestern.edu

This paper represents a brief position statement addressing the trends and challenges in *image and video analysis*. The topic is very broad and accepts various interpretations. Any image and video analysis task is determined, and it therefore becomes meaningful and its performance measurable, by the specific application it is designed for. It aims, in general, at bridging the semantic gap between *formative information* (as expressed, for example, by shape, texture, spatio-temporal forms or patterns, and motion) and *cognitive information* (pertaining to "knowing" or the interpretation of the data by a user in a specific situation).

Image and video analysis represents one of the "older" topics and therefore there exists a plethora of publications and products. Although in this summary we focus on the analysis of video (images can be considered as a special case), the analysis of other media is an equally active topic. More than that, one of the recent trends is the exploitation of the complimentarity (fusion) expressed by various media, therefore fully justifying the term multimedia.

A number of factors enter a video analysis task. In providing some structure, one can envision a three dimensional space with axes *the video data types, the functional areas, and the research and development objectives* (representation used in BAA03-003 for the Video Analysis and Content Extraction Program). Each application is then represented by a point or a set of points in this three-dimensional volume. Although past accomplishments have addressed a large number of operating points in this volume, there remain considerable uncharted regions.

Examples of video data types are medical, industrial inspection, broadcast, surveillance, and reconnaissance video. A general classification of video data is into *narrow domains* (limited variability in all relevant aspects of its appearance) and *wide domains* (unlimited and unpredictable variability in its appearance even for the same semantic meaning) [A Smeulders *et. al*., "Content-based image retrieval at the end of the early years," *IEEE Trans. PAMI,* no.12, 2000]. A current challenge is to analyze data collected under non-ideal environments (as encountered in numerous applications), that is, for example, under varying lighting conditions, varying or poor spatio-temporal resolution, various degrees of compression, and uncontrolled camera motion. Examples of functional areas are video indexing, retrieval by example, mining, monitoring, browsing, and content-based browsing. Finally the research objectives can be grouped into *core level* (e.g., image enhancement, object detection and tracking, object modeling, and camera parameter estimation), *fusion level* (e.g., multi-modal fusion, analysis of object motion and deformation, object recognition, scene modeling, and collaborative processing in general when dealing with networks of sensors), and *event level* (e.g., event detection, recognition, understanding, and summarization, and event ontology and hierarchy).

Typically the higher level objectives build on capabilities developed by the lower levels (bottom-up approach). An important current challenge in the field is the development of automated video content extraction algorithms for the ultimate goal of human-level understanding and interpretation of video content, addressing all the operating points in this volume. Current efforts either define the parameters of an application and provide efficient solutions and implementations or focus on the development of enabling cross-cutting technologies which can be used for various applications. An example of an application is the abstraction and inferencing about surveillance activities, that is, given the video output of mounted, fixed surveillance video cameras within an industrial, urban, or business setting, the goal is to accurately and reliably monitor the ingress and egress of people and vehicles, to track their movements within the field of vision of the surveillance camera and to establish and understand these patterns of activity and movement well enough to distinguish between normal and abnormal activity. Developing *metrics* to evaluate the performance of a system is another challenge in the field. The development of such metrics is better addressed within the context of a specific application; it becomes a considerably challenging task when dealing with more abstract cross-cutting technologies.