Preface

Humans observe, move, and act in the physical world, and interact with objects, other humans, other living beings, and the whole environment to achieve their goals and satisfy their needs and their desires for a fully satisfactory life. A broad spectrum of intertwined mental activities, from fundamental instincts to high-level intellectual reasoning, drives the human choices, actions, and the interactions with the environment, especially in complex situations. Replying to actions only when they have been completed allows for observing their complete effects and have a complete picture of the whole environment, but may create serious problems in the case their outcomes negatively influence the beings and the environment. To prevent these situations a more dynamic interpretation of the outcomes is needed in real time, directly while the actions themselves are still evolving. Interacting with humans implies therefore deducing the goals and the motivations beyond the actions so as to understand their implications and impact and, thus, to decide which reaction is more appropriate.

Understanding the human behavior and the related motivations, goals, and intentions by visually observing movements and actions is a complex cognitive task which involves various aspects of the vision process, the behavior abstraction, and the behavior knowledge-based interpretation. Many neurophysiologists and psychologists have studied the various aspects of the human behavior and its understanding and created conceptual models and descriptions motivating the human actions and interactions.

In the modern society, the information and communication technologies play a key role in supporting our life and our interactions, in augmenting our abilities, and in replicating—at least partially—several of our skills and activities. In this book we focus therefore our attention on the use of these technologies for human motion analysis, recognition, and interpretation, also in real time, as enabling tools for assisting in human behavior understanding and, thus, for building advanced applications based on such understanding. The human motion is typically captured in video streams by observing the scene with appropriate video cameras and other

possible supporting devices. Video analytics techniques are used to analyze the video streams to detect and extract events which are related to variations of images in the stream, and not only to the individual images, like the visual cortex does in the human brain on the sequence of images captured by the eyes. Information extracted from the video sequence is then interpreted at a higher abstraction level to achieve the desired understanding of the observed aspects of the human behavior.

Human behavior understanding from video streaming analysis has a huge and continuously growing market. In some cases the impact is so relevant that the expansion of the market is exploding. Currently, this technology is used by a variety of applications, covering a broad spectrum of sectors, e.g., encompassing advanced user interfaces, motion analysis, health care, surveillance, and virtual reality. According to recent studies, video analytics had globally in fact a significant market of US$13.2 billions at the end of 2009 and will reach US$28 by 2013 with a compound annual growth rate of about 22 %. The sector of surveillance systems (including intelligence, surveillance, and reconnaissance) saw government investments worldwide for about US$8 billions in 2011 and is expected to increase of 9 % annually in the USA till 2014. The videogames industry valued US$66 billions worldwide in 2010 and is expected to grow up to US$81 billions by 2016: a significant reason of the current value and the expected grow is due to the motion capture and understanding technologies which are the basis of highly interactive and immersive games and entertainments. These examples show the paramount importance and the humongous and growing economical value of human behavior understanding from the analysis of video streaming, especially for real-time operation. Besides, the increasing availability of capturing devices and high-performance computers at lowering costs make these technologies appealing to a number of new application areas and even open the doors to the invention of new applications, thus increasing even more their market value and impact.

Several applications may greatly benefit from observing human behavior and its understanding through the analysis of video streaming. In the area of advanced user interfaces, understanding human motions and complex actions will allow for creating gesture- and motion-driven interfaces for commanding the operating systems and the applications. This will greatly simplifying the interactions between users and systems since they will become more natural and will not need any specialized device (keyboard, mouse, or similar). This modality will complement and, possibly, integrate the speech interaction, also significantly facilitating people with disabilities. Vision systems may be used in conjunction with speech interactions to solve ambiguous situations. These advanced interfaces greatly simplify and offer more natural the interactions by making computer systems to comprehend the natural human methods of communication, e.g., in entertainment products (especially for advanced interactive games), home automation, social networks, automotive, and transportation.

The understanding of human motion and gestures is also extremely helpful in applications which make extensive use of models and, specifically, which use

model-based encoding. For example, in telecommunications video compression for teleconferencing is essential to reduce dramatically the required bandwidth by compacting the image representation and, thus, focusing transmission on the moving parts. Concise representation is helpful also in variable animation control for entertainment and publicity: a few parameters transmitted over the network will allow for describing the whole moving object and controlling its movements. Motion modeling as performance capture is also extremely attractive in games and the entertainment industry to develop products and environments with highly realistic motions both of humans and cartoons by analyzing the motion of real actors. In turn, the reduced required bandwidth will make feasible the use of advanced, interactive applications also in geographical areas, typically the economically less-favored regions, still affected by the digital divide which limits the real-time operations.

Motion analysis is another wide area of application in which capturing and analyzing the human behavior has a significant value and innovation, especially to simplify the observation and to facilitate the development of advanced products. One of the primary market segments will be health care, in which clinical studies of human locomotion, medical diagnosis, medical monitoring (e.g., pre- and post-treatment functional evaluation of patients), and rehabilitation will play a significant role in the market by exploiting the advanced technological support offered by human motion analysis for deepening the understanding of the outcomes of the medical practice. A strictly related segment is assisted living for elders and people with disabilities: motion analysis and understanding will allow for facilitating the interactions and monitoring their activities, thus, for example, detecting possible critical or dangerous situations, medical monitoring, emergencies, and need for help. Another segment related to the medical area is athletes monitoring, with specific focus, e.g., on assisted sport training as well as athletes' motion analysis and optimization: analysis of the sport activities allows for detecting improvements in performance and understanding possible critical aspects of the athlete's movements leading to performance reduction. Motion analysis can be used also in the entertainment industry, e.g., for improving the quality of human motion in choreography of dance and theatre, and for automated indexing of television programs for content-based footage.

Physical security is a very sensitive application area which greatly benefits from human behavior analysis and understanding. In smart surveillance systems, the video streams are analyzed automatically by intelligent approaches which mimic the operation and the skills of surveillance personnel, possibly with constant level of attention and homogeneous (high) level of quality and care, in detecting patterns of unusual behavior and thus deducing the malevolent intentions of people monitored by the video cameras. Indoor and outdoor scenes can be continuously monitored, by assuring safety and protection of persons and goods and, thus, more confidence and trust for a better quality of the daily life. These techniques can be used to protect sensitive locations (e.g., public places, main streets and squares,

sports arenas, theatres, cinemas, public meeting places), critical buildings and services (e.g., airports, railways stations, ports, justice and governmental buildings, military buildings, supermarkets, big retail shops, sports halls, universities), and private buildings (e.g., homes). Human behavior analysis for physical security and safety can be exploited also in military, peace-keeping, and humanitarian situations, especially in dangerous areas and during military operations, thus providing support for monitoring the environment and early detection of possible dangers. Intelligence, surveillance, and reconnaissance based on behavior monitoring and understanding are typical application areas for national protection, forensics, and danger and terrorism prevention. In addition to deducing critical or dangerous behavior, human motion observation can be expanded to support person recognition, specifically by using gait recognition, which is sufficiently effective under some environmental and clothing constraints.

Finally, a fruitful market segment, with unique potentials for an even more pervasive and valuable expansion is virtual reality, encompassing a broad variety of specific applications. Motion understanding can be used to create realistic interactive virtual worlds, populated by animated characters which naturally mimic the real counterparts or empower the users to create models very similar to real living beings or behaving in a way very similar to theirs. Animators capture the behavior of human actors and then incorporate their motions in the characters for the virtual environment. Teleconferencing can benefit from a realistic perspective while limiting the bandwidth. Virtual movies and television productions can incorporate characters and special effects based on virtualization and models derived from the analysis of the human behavior and motion. Motion modeling for games and entertainment, especially performance capture, can recreate a realistic environment through virtualized models and interactions; fundamental characteristics may be in fact extracted from the analysis of the human behavior of significant subjects. Virtual and augmented reality (e.g., visual perception test, performing virtual walkthroughs in three-dimensional environments, and training simulations) can be significantly boosted by the extensive use of models and interaction techniques based on information and knowledge provided by human behavior analysis and understanding: in this perspective, robotics, industrial automation, manufacturing, and design can use the virtual environments as a new dimension for interacting with the real world.

This book focus the attention on natural and intuitive human–computer interaction and, more specifically, to the problem of real-time, vision-based behavior analysis and human motion recognition and interpretation. This will constitute the underlying methodology to capture knowledge from the external environment and exploit information and models within advanced applications, like those mentioned above.

Simplifying the behavior analysis by means of a symbolic representation will make viable the scene monitoring and the human behavior understanding in online operations. The research methodology developed in this book is in fact based on

the hierarchical decomposition of the human behavior. Scene analysis and behavior understanding are performed by using a symbolic interpretation of the visual input: images from the captured video streams are described by means of strings of symbols and the grammar-based analysis methodology allows for hierarchical decomposition and parsing of the visual input. The proposed framework is flexible enough to provide analysis at various abstraction levels of the observed scenes (encompassing the scene level, the level of the group of objects, and the level of the individual object), thus offering a structured approach suited for various points of view at different granularity.

http://www.springer.com/978-1-4614-5485-4

Semantic Analysis and Understanding of Human Behavior in Video Streaming

Amato, A.; Di Lecce, V.; Piuri, V.

2013, XII, 108 p., Hardcover
ISBN: 978-1-4614-5485-4