Preface

With the development of silicon technologies, consumer electronics devices, such as personal computers, HDTV, mobile phones, digital cameras, and game consoles, have become ubiquitous for people's daily life. These devices can provide multimedia sources for entertainment, communication, and so on. To interact with these equipments, consumers currently rely on devices such as remote controls, keyboards, or control panels, which are often inconvenient, ambiguous, and noninteractive. How to design user interfaces of CE products that enable natural, intuitive and fun interaction is one of the main challenges the CE industry is facing. Many companies and institutes are working on the advanced user interfaces.

User interface technologies have been studied in various disciplines for decades. Considering that modern CE products are usually supplied with both microphones and cameras, how to employ both audio and visual information for interactive multimedia has recently received much attention in both academia and industry. But interactive multimedia is still an under-explored field. Many challenges exist when moving to multimodal interaction. For example, how to annotate and search huge data acquired by using multiple sensors, especially in the unconstrained end-user environments? How to effectively extract and select representative multimedia features for human behavior recognition? And how to select the fusion strategy of multimodal data for a given application? To address these challenges, we must adapt the existing approaches or find new solutions suitable for multimedia interaction.

This book brings together high-quality and recent research advances on multimedia interaction, user interfaces, and particularly applications on consumer electronics. The targeted readers are researchers and practitioners working in the areas of multimedia analysis, human–computer interaction, and interactive user interfaces from both academia and industry. It can also be used as a reference book for graduate students studying computer vision, pattern recognition, or multimedia. In the following we summarize all the chapters.

In "*Retrieving Human Actions Using Spatio-Temporal Features and Relevance Feedback*", Jin and Shao present the solution of human action retrieval with local spatio-temporal features based on the bag-of-words model. Brightness gradient and 3D shape context are combined to increase the discriminative power of feature descriptors. Relevance feedback is then applied to refine retrieved action sequences

and is demonstrated to be effective in highly complex scenarios, such as actions in movies.

In "*Computationally Efficient Clustering of Audio-Visual Meeting Data*", Hung et al. present novel computationally efficient algorithms to extract semantically meaningful acoustic and visual events related to each of the participants in a group discussion. Their methods can be used as a principal component that enables many higher-level semantic analysis tasks needed in search, retrieval, and navigation.

In "*Cognitive-Aware Modality Allocation in Intelligent Multimodal Information Presentation*", Cao et al. emphasize that modality allocation in intelligent multi-modal presentation systems should also take into account the cognitive impacts of modality on human information processing. After presenting a user study, they show a possible way to integrate cognitive theories into a computational model that can predict the suitability of a modality choice for a given presentation task.

In "*Natural Human–Computer Interaction*", D'Amico et al. introduce the theorization and development of natural human–computer interaction systems. After reviewing the state of the art, a case study, a Smart Room with Tangible Natural Interaction, was discussed.

In "*Gesture Control for Consumer Electronics*", Shan presents an overview on gesture control technologies for Consumer Electronics. Different sensing technologies are discussed; existing researches on vision-based gesture recognition are extensively reviewed, covering face/hand detection, tracking, and gesture recognition. The latest developments on gesture control products and applications are also introduced.

In "*Empirical Study of a Complete System for Real-Time Face Pose Estimation*", Gritti focuses on the task of fully automatic real-time face 3D pose estimation. A complete system is developed, which is capable of self-initializing, estimating the pose robustly, and detecting failures of tracking. A robust tracking methodology is also introduced.

In "*Evolution-Based Virtual Content Insertion with Visually Virtual Interactions in Videos*", Chang and Wu present an evolution-based virtual content insertion system which can insert virtual contents into videos with evolved animations according to predefined behaviors emulating the characteristics of evolutionary biology. The system would bring a new opportunity to increase the advertising revenue for video assets of the media industry and online video-sharing websites.

In "*Physical Activity Recognition with Mobile Phones: Challenges, Methods, and Applications*", Yang et al. introduce a novel system that recognizes and records the physical activity of a person using a mobile phone. With the data collected by a built-in accelerometer, the system recognizes five everyday activities in real-time, i.e., stationary, walking, running, bicycling, and in vehicle.

In "*Gestures in an Intelligent User Interface*", Fikkert et al. investigate which hand gestures are intuitive to control a large display multimedia interface from a user's perspective. Numerous gesture possibilities are evaluated for a set of commands that can be issued to the interface. A working prototype is then implemented with which the users could interact with both hands and the preferred hand gestures with 2D and 3D visualizations of biochemical structures.

As video summarization techniques have attracted increasing attention for efficient multimedia data management, quality evaluation of video summary is required. In "*Video Summary Quality Evaluation Based on 4C Assessment and User Interaction*", Ren et al. propose a novel full-reference evaluation framework to assess the quality of the video summary according to various user requirements.

In "*Multimedia Experience on Web-Connected CE Devices*", Tretter et al. discuss the features that characterize the new generation of CE and illustrate this new paradigm through an examination of how web services can be integrated with CE products to deliver an improved user experience. Choosing digital photography as a case study, they introduce AutoPhotobook, an automatic photobook creation service, and show how the collection of technologies is integrated into a larger ecosystem with other web services and web-connected CE devices to deliver an enhanced user experience.

Sheffield, UK   Ling Shao
Eindhoven, The Netherlands   Caifeng Shan
Rochester, NY, USA   Jiebo Luo
Kanagawa, Japan   Minoru Etoh

Springer