Chapter 63
Advances in Region-of-Interest Video and Image Processing

Dan Grois
Ben-Gurion University of the Negev, Israel

Ofer Hadar
Ben-Gurion University of the Negev, Israel

ABSTRACT

The advent of cheaper and more powerful devices with the ability to play, create, and transmit video content has led to a dramatic increase in the multimedia content distribution on both wireline and wireless networks. Also, the reduction of cost of digital video cameras along with the development of user-generated video sites (e.g., iTunes™, YouTube™) stimulated a new user-generated video content sector and made unprecedented demands for high-quality and low-delay video communication. The Region-of-Interest (ROI) is a desirable feature in many future scalable video coding applications, such as mobile device applications, which have to be adapted to be displayed on a relatively small screen; thus, a mobile device user may wish to extract and track only a predefined ROI within the displayed video. At the same time, other users having a larger mobile device screen may wish to extract other ROIs to receive higher video stream resolution. Therefore, to fulfill these requirements, it would be beneficial to simultaneously transmit or store a video stream in a variety of ROIs, as well to enable efficiently tracking of the predefined Region-of-Interest. This chapter presents recent advances in Region-of-Interest video and image processing techniques for multimedia applications, while making a special emphasis on a scalable extension of the H.264/AVC standard. The detailed observations and conclusions, which are presented in this chapter, are supported by authors' personal experience in this field, thereby presenting a variety of experimental results.

INTRODUCTION

The number of video applications has been dramatically increased in the last decade, due to many reasons, such as rapid changes in the video coding standardization process driven by the increase of the computing power and significant developments of network infrastructures. Nowadays, the most common video applications include wireless and wired Internet video streaming, high-quality video conferencing, High-Definition (HD) TV broadcasting, HD DVD storage and Blu-ray storage, while employing a variety of video transmission and storage systems (e.g., MPEG-2

for broadcasting services over satellite, cable, and terrestrial transmission channels, or H.320 for conversational video conferencing services (Schwarz et al., 2007)).

The H.264/AVC (ISO/IEC MPEG-4 Part 10) video coding standard (Wiegand & Sullivan, 2003), which was officially issued in 2003, has become a challenge for real-time video applications. Compared to the MPEG-2 standard, it gains about 50% in bit rate, while providing the same visual quality. In addition to having all the advantages of MPEG-2 (ITU-T & ISO/IEC JTC 1, 1994), H.263 (ITU-T, 2000), and MPEG-4 (ISO/IEC JTC 1, 2004), the H.264 video coding standard possesses a number of improvements, such as the content-adaptive-based arithmetic codec (CABAC), enhanced transform and quantization, prediction of "Intra" macroblocks, and others. H.264/AVC is designed for both constant bit rate (CBR) and variable bit rate (VBR) video coding, useful for transmitting video sequences over statistically multiplexed networks, the Ethernet, or other Internet networks. This video coding standard can also be used at any bit rate range for various applications (e.g., typically from 100kb/sec to 15Mb/sec), varying from wireless video phones to high definition television (HDTV) and digital video broadcasting (DVB). In addition, H.264 provides significantly improved coding efficiency and greater functionality, such as rate scalability, "Intra" prediction and error resilience in comparison with its predecessors, MPEG-2 and H.263. However, H.264/AVC is much more complex in comparison to other coding standards and to achieve maximum quality encoding, high computational resources are required (Grois et al., 2010c; Kaminsky et al., 2008).

Most access networks are usually characterized by a wide range of connection qualities, and a wide range of end-user devices with different capabilities, starting from cell phones/mobile devices with relatively small displays and limited computational resources to powerful Personal Computers (PCs) with high-resolution displays (Schwarz et al., 2007). As a result, due to the continuous need for scalability, much of the attention in the field of video processing and coding is currently directed to the Scalable Video Coding (SVC), which was standardized in 2007 as an extension of H.264/AVC (Schwarz et al., 2007), since the bit-stream scalability for video is currently a very desirable feature for many multimedia applications (e.g., video conferencing, video surveillance, telemedical applications, etc.). The need for the scalability arises from the need for spatial formats, bit-rates or power (Wiegand & Sullivan, 2003; Grois & Hadar, 2011a; Grois & Hadar, 2011b). To fulfill these requirements, it would be beneficial to simultaneously transmit or store video in a variety of spatial/temporal resolutions and qualities, leading to video bit-stream scalability. Major requirements for the Scalable Video Coding are to enable encoding of a high-quality video bitstream that contains one or more subset bitstreams to provide video services with lower temporal or spatial resolutions, or to provide reduced reliability, while retaining reconstruction quality that is highly relative to the rate of the subset bitstreams. Therefore, Scalable Video Coding provides important functionalities, such as the spatial, temporal and fidelity/quality (i.e. Medium Grained Scalability (MGS) and Coarse Grain Scalability (CGS)) scalabilities (Schwarz et al., 2007; Schierl et al., 2007), as schematically presented in Figure 1. In turn, these functionalities lead to enhancements of video transmission and storage applications.

SVC has achieved significant improvements in coding efficiency compared to the scalable profiles of prior video coding standards due to the largely increased flexibility and adaptability (e.g., SVC provides a graceful degradation in lossy transmission environments as well as in the bit-rate and power adaptation). Also, in addition to the temporal, spatial and fidelity/quality scalabilities, the SVC supports ROI scalability, thereby enabling to define a particular portion of each picture/frame as a desired ROI, as schematically presented in

Related Content

Image Compression Technique for Low Bit Rate Transmission
Shaimaa A. El-said, Khalid F. A. Husseinand Mohamed M. Fouad (2011). *International Journal of Computer Vision and Image Processing (pp. 1-18).*
www.irma-international.org/article/image-compression-technique-low-bit/64182

Invariant Model Combining Geometry and Appearance for Facial Detection and Gender Classification From Arbitrary Viewpoints
Mokhtar Taffarand Serge Miguet (2018). *Handbook of Research on Advanced Concepts in Real-Time Image and Video Processing (pp. 196-231).*
www.irma-international.org/chapter/invariant-model-combining-geometry-and-appearance-for-facial-detection-and-gender-classification-from-arbitrary-viewpoints/186279

An Efficient Method for Video Shot Transition Detection Using Probability Binary Weight Approach
Nandini H. M., Chethan H. K.and Rashmi B. S. (2021). *International Journal of Computer Vision and Image Processing (pp. 1-20).*
www.irma-international.org/article/an-efficient-method-for-video-shot-transition-detection-using-probability-binary-weight-approach/282058

Recognition of Humans and Their Activities for Video Surveillance
Alok Kumar Singh Kushwahaand Rajeev Srivastava (2018). *Computer Vision: Concepts, Methodologies, Tools, and Applications (pp. 2290-2306).*
www.irma-international.org/chapter/recognition-of-humans-and-their-activities-for-video-surveillance/197052

A Novel Approach to Improve Face Recognition Process Using Automatic Learning
Yacine Gafour, Djamel Berrabahand Abdelkader Gafour (2020). *International Journal of Computer Vision and Image Processing (pp. 42-66).*
www.irma-international.org/article/a-novel-approach-to-improve-face-recognition-process-using-automatic-learning/245669