

Video Object Segmentation

Ee Ping Ong

Institute for Infocomm Research, Singapore

Weisi Lin

Nayang Technological University, Singapore

INTRODUCTION

Video object segmentation aims to extract different video objects from a video (i.e., a sequence of consecutive images). It has attracted vast interests and substantial research effort for the past decade because it is a prerequisite for visual content retrieval (e.g., MPEG-7 related schemes), object-based compression and coding (e.g., MPEG-4 codecs), object recognition, object tracking, security video surveillance, traffic monitoring for law enforcement, and many other applications. Video object segmentation is a nonstandardized but indispensable component for an MPEG4/7 scheme in order to successfully develop a complete solution. In fact, in order to utilize MPEG-4 object-based video coding, video object segmentation must first be carried out to extract the required video object masks. Video object segmentation is an even more important issue in military applications such as real-time remote missile/vehicle/soldier's identification and tracking. Other possible applications include home/office/warehouse security where monitoring and recording of intruders/foreign objects, alarming the personnel concerned or/and transmitting the segmented foreground objects via a bandwidth-hungry channel during the appearance of intruders are of particular interest. Thus, it can be seen that fully automatic video object segmentation tool is a very useful tool that has very wide practical applications in our everyday life where it can contribute to improved efficiency, time, manpower, and cost savings.

BACKGROUND

For segmentation of objects from video sequences, temporal and spatial information and their appropriate combination have been extensively exploited (Aach & Kaup, 1993; Bors & Pitas, 1998; Castagno, Ebrahimi,

& Kunt, 1998; Chen, Chen, & Liao, 2000; Chen & Swain, 1999; Chien, Ma, & Chen, 2002; Cucchiara, Onfiani, Prati, & Scarabottolo, 1999; Kim, Choi, Kim, Lee, Lee, Ahn, & Ho, 1999; Kim, Jeon, Kwak, Lee, & Ahn, 2001; Koller, Weber, Huang, Malik, Ogasawara, Rao, & Russel, 1994; Li et al., 2001; Li, Tye, Ong, Lin, & Ko, 2002; Li, Gu, Leung, & Tian, 2004; Liu, Hong, Herman, & Chellappa, 1998; Liu, Chang, & Chang, 1998; Mech & Wollborn, 1997; Meier & Ngan, 1999; Mester & Aach, 1997; Neri, Colonnese, Russo, & Talone, 1998; Odobez & Bouthemy, 1998; Ong & Spann, 1999; Shao, Lin, & Ko, 1998a, 1998b; Toklu, Tekalp, & Erdem, 2000). Fully automatic extraction of semantically meaningful objects in video is extremely useful in many practical applications but faces problems like limited domain of application, *ad hoc* approaches, need of excessive parameter/threshold setting and fine-tuning, and overly complicated algorithms. With current level of development of algorithms, in general, only supervised segmentation approaches (e.g., Castagno et al., 1998; Toklu et al., 2000) are capable of detecting semantic objects more accurately from video. Supervised approaches can be found in applications such as studio editing and content retrieval.

Nevertheless, automatic video object segmentation is essential and possible for some specific but important scenarios like traffic monitoring, video-conferencing, and online/mobile security systems. For digital image and video storage and transmission, object-based coding improves the compression and coding efficiency. In low bit rate applications, object segmentation enables transmitting objects of interest in higher quality and objects of less importance in lower quality (e.g., allocating smaller number of bits for objects of less importance or only refreshing them occasionally, as used in sprite-based video coding), while a conventional frame-based video coding scheme would have to suffer an across-the-board reduction in video quality in order to fit into the bandwidth available.

The requirements for a practical video object segmentation system to be useful in the above-mentioned scenarios are: (a) fully automatic operations; (b) efficient and fast (preferably real-time) processing; and (c) robustness with noisy data. Automatic segmentation is necessary because user intervention is not feasible for real-time online applications. Appropriate domain knowledge (e.g., head and shoulder scene is mostly expected as foreground in video-conferencing) can be utilized in the process, besides spatial and temporal information. Real-time realization is usually needed for such applications and therefore low computational complexity is the key. Fast processing itself also aids in obtaining a better segmentation since frame difference due to the motion becomes too big for good segmentation if a segmentation process fails to catch up with the speed of live video camera input (unlike the case with prestored video sequences). Moreover, the segmentation module in an integrated system may need to share the available computing power with the video compression and coding module (e.g., Lin, Tye, Ong, Xiong, Miki, & Hotani, 1999) which is also computationally expensive. Furthermore, algorithms developed should be reasonably resilient to noise for reliable segmentation with live video. Possibilities can be exploited to incorporate appropriate stationary regions (e.g., human trunk that does not move all the time) with moving regions (like head and hands in the example) to form a semantic object.

Liu et al. (1998a) describes an object boundary extraction algorithm based on optical flow that is real-time implemented on a HyperSparc workstation. For the more restricted applications, real-time vehicle detection systems have been developed with dedicated hardware design (Cucchiara et al., 1999; Koller et al., 1994) and a PC-based platform (Siyal & Fathy, 1995). Such systems are based upon interframe difference analysis. Methods using optical flow calculation and other motion parameter estimators are generally too computationally expensive for economical real-time PC/DSP implementation.

There are various techniques for video object segmentation, but the faster video object segmentation techniques (Aach & Kaup, 1993; Cheung & Ramath, 2004; Kim et al., 1999; Liu et al., 1998b; Mester & Aach, 1997; Neri et al., 1998; Shao, Lin, & Ko, 1998b) are based on the change detection approach (with and without preprocessing to cater for global motion) followed by further postprocessing. However, the

different video object segmentation methods differ in how the change detection and the postprocessing are being performed. These have a significant influence on the actual speed, accuracy, and reliability of the final segmentation results.

VIDEO SEGMENTATION: NEW AND OLD PROBLEMS

A general solution is still elusive in practice for automatic video object segmentation due to its well-acknowledged inherent difficulties: (i) the difficulty in estimating objects' motions and other distinguishing features accurately in order to facilitate the separation of differently moving foreground and background objects; (ii) the inevitable presence of noise/clutters that affects accuracy in segmentation; (iii) the effects of occlusions and dis-occlusions, causing dis-occluded regions to be wrongly labeled as foreground (moving) objects; (iv) the lack of generosity and robustness of an algorithm, especially in the situations in which no human intervention or fine-tuning of the segmentation parameters is possible; (v) high computational complexity involved which makes it difficult to design an algorithm that is robust enough to run in real-time for real-life applications; and (vi) for change detection methods, the stationary background frame obtained is nonrobust and not possible to obtain a good stationary background frame in real-life practical applications.

Spatiotemporal segmentation is more likely to obtain good segmentation for digital video when compared to methods using only motion information because spatial features usually provide more information on precise object boundary. There were many spatiotemporal approaches that have been proposed using change detection (Aach & Kaup, 1993; Cheung & Kamath, 2004; Kim et al., 1999; Liu et al., 1998b; Mester & Aach, 1997; Neri et al., 1998; Shao et al., 1998b), instead of motion model/parameter estimation (Bors & Pitas, 1998; Li et al., 2001; Liu et al., 1998; Meier & Ngan, 1999; Odobez & Bouthemy, 1998; Shao, Lin, & Ko, 1998a; Ong et al., 1999) since the former aims at real-time processing and different motion in a foreground object does not need to be distinguished in the targeted applications (that is, a scene is separated into only two classes of objects: foreground and background). Temporal information is extracted via change detection as simple and straightforward hints of foreground

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/video-object-segmentation/13438

Related Content

Improving English Teaching Strategies From the Perspective of College Students' Mental Health
Xinhui Zhang (2024). *Journal of Cases on Information Technology* (pp. 1-18).

www.irma-international.org/article/improving-english-teaching-strategies-from-the-perspective-of-college-students-mental-health/342090

New ICTs for Conflict Prevention and Management

Ángela-Jo Medina (2008). *Information Communication Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 3784-3803).

www.irma-international.org/chapter/new-icts-conflict-prevention-management/22915

Exploring the Use of an Integrated Financial Management Information System (IFMIS) in the Local Government Institutions in Ghana

Lexis Alexander Tetteh, Paul Muda, Daniel Susuawu, Prince Sunuand Thompson Kubaje Aneyire (2022). *Information Resources Management Journal* (pp. 1-25).

www.irma-international.org/article/exploring-the-use-of-an-integrated-financial-management-information-system-ifmis-in-the-local-government-institutions-in-ghana/298973

A Methodology for the Auditing of Technological Knowledge Management

Enrique Paniagua Arisand Belén López Ayuso (2010). *Information Resources Management: Concepts, Methodologies, Tools and Applications* (pp. 566-588).

www.irma-international.org/chapter/methodology-auditing-technological-knowledge-management/54503

Healthcare IT Project Failure: A Systems Perspective

Rajneesh Chowdhury, Ruth E. Butler and Steve Clarke (2007). *Journal of Cases on Information Technology* (pp. 1-15).

www.irma-international.org/article/healthcare-project-failure/3209