ELSEVIER

Signal Processing: *Image Communication* 8 (1996) 327–351

SIGNAL PROCESSING: IMAGE COMMUNICATION

Efficient representations of video sequences and their applications

Michal Irani*, P. Anandan, Jim Bergen, Rakesh Kumar, Steve Hsu

David Sarnoff Research Center, CN5300, Princeton, NJ 08530, USA

Abstract

Recently, there has been a growing interest in the use of mosaic images to represent the information contained in video sequences. This paper systematically investigates how to go beyond thinking of the mosaic simply as a visualization device, but rather as a basis for an *efficient* and *complete* representation of video sequences. We describe two different types of mosaics called the *static* and the *dynamic* mosaics that are suitable for different needs and scenarios. These two types of mosaics are unified and generalized in a mosaic representation called the *temporal pyramid*. To handle sequences containing large variations in image resolution, we develop a *multiresolution mosaic*. We discuss a series of increasingly complex alignment transformations (ranging from 2D to 3D and layers) for making the mosaics. We describe techniques for the basic elements of the mosaic construction process, namely sequence *alignment*, sequence *integration* into a mosaic image, and *residual analysis* to represent information not captured by the mosaic image. We describe several powerful video applications of mosaic representations including *video compression*, *video enhancement*, *enhanced visualization*, and other applications in *video indexing*, *search*, and *manipulation*.

Keywords: Video representation; Mosaic images; Motion analysis; Image registration; Video databases; Video compression; Video enhancement; Video visualization; Video indexing; Video manipulation

1. Introduction

Video is a very rich source of information. Its two basic advantages over still images are the ability to obtain a continuously varying set of views of a scene, and the ability to capture the temporal (or 'dynamic') evolution of phenomena.

A number of applications that involve processing the entire information within video sequences have recently emerged. These include digital libraries, interactive video analysis and softcopy exploitation environments, low-bitrate video transmission, and interactive video editing and manipulation systems. These applications require efficient representations of large video streams, and efficient methods of accessing and analyzing the information contained in the video data.

There has been a growing interest in the use of a panoramic 'mosaic' image as an efficient way to represent a collection of frames (e.g., see Fig. 1) [17, 21, 22, 16]. Since successive images within a video sequence usually overlap by a large amount, the mosaic image provides a significant reduction in the total amount of data needed to represent the scene.

* Corresponding author. E-mail: michal@sarnoff.com.

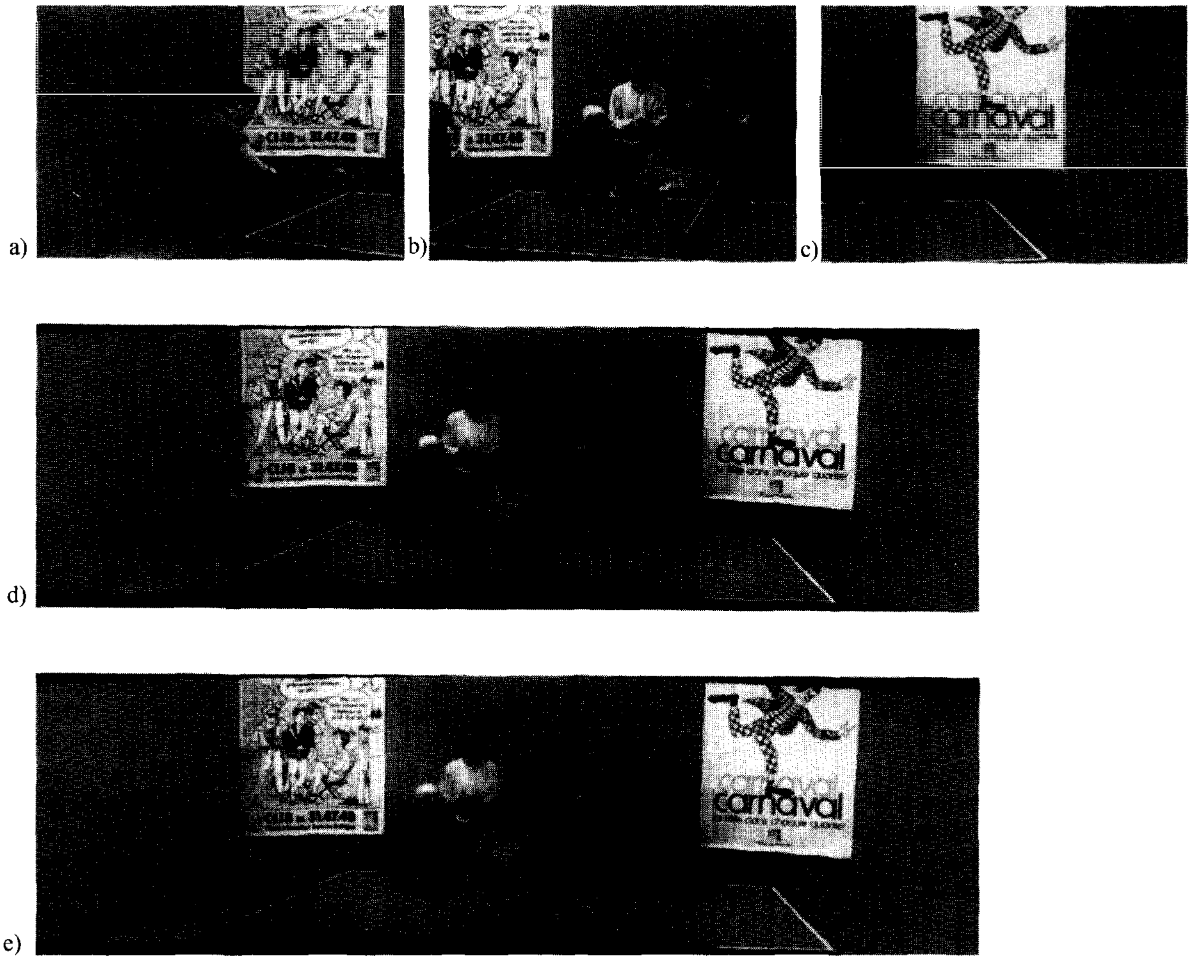

Fig. 1. Static mosaic image of a table-tennis game sequence. (a)–(c) Three out of a 300 frame sequence obtained by a camera panning across the scene; (d) the static mosaic image constructed using a temporal median; (e) the static mosaic image constructed using a temporal average.

Although the idea of the mosaic and even some of its applications have been recognized, there has not been a systematic approach to the characterization of what the mosaic is, or even an attempt to develop any type of standard terminology or taxonomy. In practice, a single type of mosaic, such as a static mosaic image obtained from all the frames of a contiguous sequence, is suitable for only a limited class of applications. Different applications such as video database storage and retrieval and real-time transmission and processing require different types of mosaics.

Also, while mosaics have been recognized as efficient ways of providing 'snapshot' views of scenes, the issue of how to develop a *complete* representation of scenes based on mosaics has not been adequately treated. Specifically, we refer to the question of how to represent the details *not* captured by the mosaics, so that the sequence can be fully recovered from the mosaic representation.

The purpose of this paper is to develop a taxonomy of mosaics by carefully considering the various issues that arise in developing mosaic representations. Once this taxonomy is available, it can be readily seen how the various types of mosaics can be used for different applications. The paper includes examples of several applications of mosaics, including video compression, video visualization, video enhancement, and other applications.

The remainder of the paper is organized as follows. Section 2 presents various types of mosaic representations, and discusses their efficiency and completeness in terms of sequence representation. Section 3 describes the techniques that we use to align the images, construct the mosaics, and detect the significant 'residuals' not captured in the mosaics from the input video stream. Section 4 outlines a number of powerful video applications of the mosaic representations with examples and experimental results. Finally Section 5 discusses the salient issues for future research on this topic.

2. The mosaic representation

A mosaic image is constructed from all frames in a scene sequence, giving a panoramic view of the scene. Although the idea of a mosaic image is simple and clear, a closer look at the definition reveals a number of subtle variations. For instance, since the different images that comprise a mosaic spatially overlap with each other, but are taken at different time instances, there is a choice regarding how the different grey values available for the same pixel are combined. Similarly, the variations in the pixel resolution between images leads to the issue of choosing the resolution of the mosaic image. Finally, there are also choices regarding the geometric transformation model used for aligning the images to each other. The different choices in these various issues is typically a result of the type of application for which the mosaic is intended.

In this section we describe different 'types' of mosaics that arise out of the types of considerations outlined above.

2.1. Static mosaic

The static mosaic is the common mosaic representation [17, 22, 21, 16, 14], although it is usually not referred to by this name. It has been previously referred to as mosaic or as 'salient still' (e.g., see Figs. 1 and 2). It will be shown (in Section 4) how the static mosaic can also be extended to represent temporal subsamples of key events in the sequence to produce a static 'event' mosaic (or 'synopsis' mosaic).

The input video sequence is usually segmented into contiguous *scene subsequences* (e.g., see [23]), and a static mosaic image is constructed for each scene subsequence to provide a snapshot view of the subsequence. This is done *in batch mode*, by aligning all frames of that subsequence to a *fixed* coordinate system (which can be either user-defined or chosen automatically according to some other criteria). The aligned images are then integrated using different types of temporal filters into a mosaic image, and the significant residuals are computed for each frame of relative to the mosaic image. The details of the mosaic construction process are described in Section 3. Note that after integration, the moving objects either disappear or leave 'ghost-like' traces in the panoramic mosaic image.

Examples of static mosaic images are shown in Figs. 1 and 2. In Fig. 1 a static mosaic image of a table-tennis game sequence is constructed, once using a temporal median, and once using a temporal average. In this sequence, the player and the crowd move with respect to the background, while the camera pans to the right. The constructed mosaic image displays a sharp background, with blurry crowd, and a ghost-like player. Fig. 2 shows a static mosaic image of a baseball game sequence produced using a temporal median. In this sequence two players run across the field (from right to left), while the camera pans to the left and zooms in on the players. The constructed mosaic image in this case displays a sharp image of the background with no trace of the two players. In both examples, a 2D motion model was sufficient to align the images (see Section 3).

The static mosaic image exploits long term *temporal* redundancies (over the entire scene subsequence) and large *spatial* correlations (over large portions of the image frames), and is therefore an efficient scene representation. For examples, in Figs. 1 and 2, the *entire* video sequence can be represented by the mosaic image of the background scene with the appropriate transformations that relate each frame to the mosaic image. The only information in the sequence *not* captured by the mosaic image and needing additional representation are the changes in the scene with respect to the background (e.g., moving players). These residuals can either be represented independently for

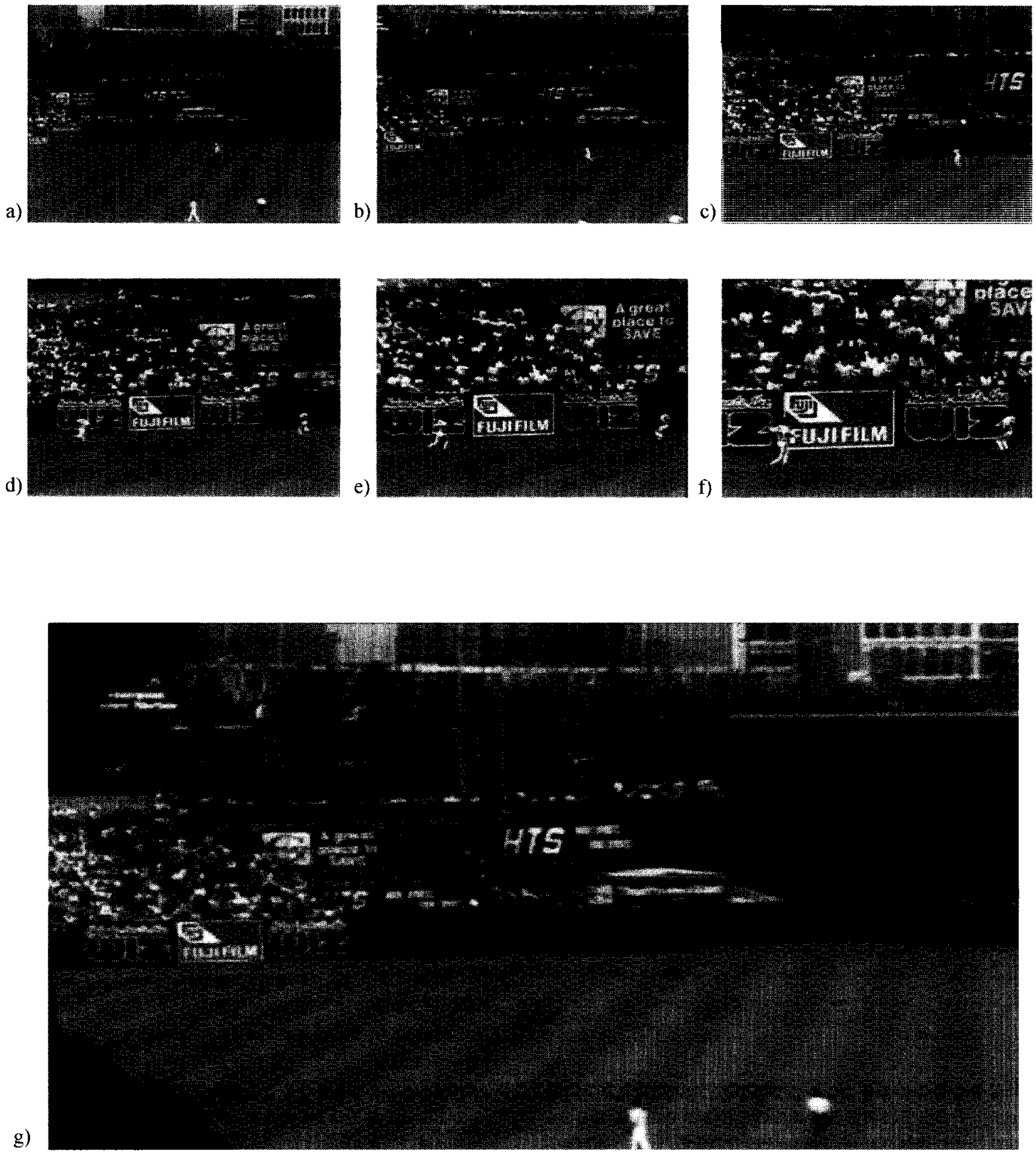

Fig. 2. Static mosaic image of a baseball game sequence. (a)–(f) Six out of a 90 frame sequence obtained by a camera panning from right to left and zooming in on the runners. (g) The static mosaic image constructed using a temporal median. The black regions are scene parts that were never imaged by the camera (since the camera zoomed-in on the scene).

each frame, or can frequently be represented more efficiently as another layer using yet another mosaic [1] (see Section 2.5).

The issue of representing residuals which are not captured by the mosaic image has frequently been overlooked by handling sequences with no scene activity [21, 16, 14]. The mosaic image, along with the frame alignment transformations, and with the residuals together constitute a *complete* and *efficient* representation, from which the video sequence can be *fully* reconstructed. These issues have been addressed to a limited extent with respect to video compression in [1], although that work does not consider how to assign a significance measure to the residuals or how to handle *non-rigid* layers.

The static mosaic, being an efficient scene representation, is ideal for *video storage and retrieval*, especially for *rapid browsing* in large digital libraries and to obtain efficient access to individual frames of interest. It can also be used to increase the efficiency of content-based indexing into a video sequence, to reduce the tedium associated with video manipulation and analysis. Last but not least, it can be used for enhanced visualization in the form of panoramic views, as well as a tool for enhancing the contents of the images. These applications are described in greater detail in Section 4.

2.2. Dynamic mosaic

Since the *static* mosaic is constructed in *batch mode*, it cannot completely depict the dynamic aspects of the video sequence. This requires a *dynamic* mosaic, which is a *sequence* of evolving mosaic images, where the *content* of each new mosaic image is updated with the most current information from the most recent frame. The sequence of dynamic mosaics can be visualized either with a stationary background (e.g., by completely removing any camera induced motion), or in a manner such that each new mosaic image frame is aligned to the corresponding input video image frame. In the former case, the coordinate system of the mosaic is fixed (see Fig. 3), whereas in the latter case the mosaic is viewed within a moving coordinate system (see Fig. 4). In some cases a third alternative may be more appropriate, wherein a portion of the camera motion (e.g., high frequency jitter) is removed or a preferred camera trajectory is synthesized.

When a *fixed* coordinate system is chosen for the dynamic mosaic, each new image frame is warped towards the current dynamic mosaic image, and the information within its field of view is updated according to the update criterion (e.g., most recent, average, weighted average, etc. (see Section 3.2)). When the coordinate system of the mosaic is chosen to be *dynamically* updated to match that of the input sequence, the current dynamic mosaic image is warped towards each new frame, and then the information within the current field of view is updated according to the update criterion. When a *virtual coordinate system* is chosen (either predetermined by the user, or computed according to some criterion), both the dynamic mosaic and the current frame are warped towards that coordinate system. Note that the definition of the coordinate system and the warping mechanism will vary according to the world and motion model (see Section 3).

Figs. 3 and 4 show examples of the evolution of some dynamic mosaics. Fig. 3 shows an evolving dynamic mosaic image of a table-tennis game, where the player and the crowd move with respect to the background, while the camera pans to the right. In this example we chose to construct the mosaic in a *fixed* coordinate system (that of the first frame). Note that in the dynamic mosaic the crowd and the player do not blur out (as opposed to the static mosaic shown in Fig. 1), and are constantly being updated.

Fig. 4 shows an evolving dynamic mosaic image of a baseball game sequence, where two players run across the field (from right to left), while the camera pans to the left and zooms in on the players. In this example we chose to construct the mosaic in a *dynamic* coordinate system that matches that of the input video (i.e., changes with each new frame). Note that in the dynamic mosaic the players do not disappear (as opposed to the static mosaic in Fig. 2), but are constantly being updated.

The *complete* dynamic mosaic representation of the video sequence consists of the *first* dynamic mosaic, and the *incremental* alignment parameters and the *incremental* residuals that represent the changes. Note that the difference in mosaic content between the static and dynamic mosaics implies a difference in the residuals that are not represented by the mosaic. In the dynamic case, since the content of the mosaic is

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.