

Figure 5: Transmission compression: results of dynamic mosaic-based-compression vs. standard MPEG compression on a storage-house surveillance sequence. *Left column:* Some representative frames of a 24 second sequence. *Middle column:* The reconstructed frames after using dynamic mosaic-based-compression at a constant bit rate of 32 Kbits/sec. *Right column:* For comparison: The reconstructed frames after using standard MPEG compression of the sequence at the same bit rate, i.e., 32 Kbits/sec. Note the differences in the reconstructed quality of the running soldiers in the images of the bottom row.

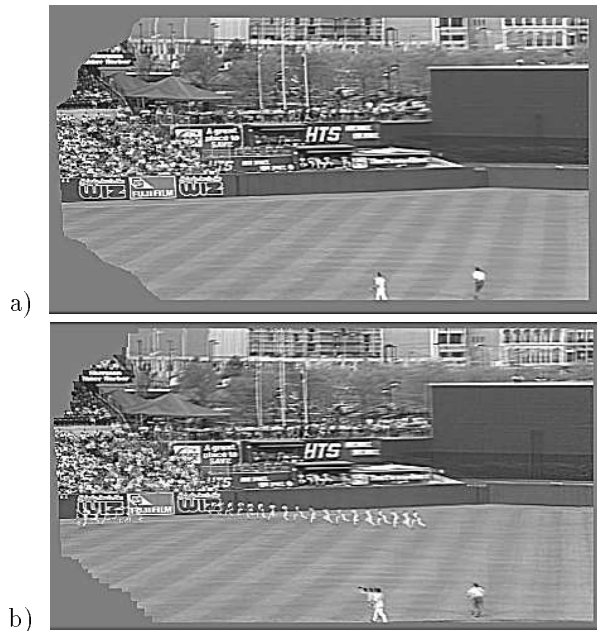


Figure 6: Synopsis mosaic image of the baseball game sequence. (a) The static *background* mosaic of the scene *without* the event. (b) The synopsis mosaic of the baseball sequence showing the event that occurred in the scene on top of the background mosaic (i.e., showing the trajectories of the two runners).

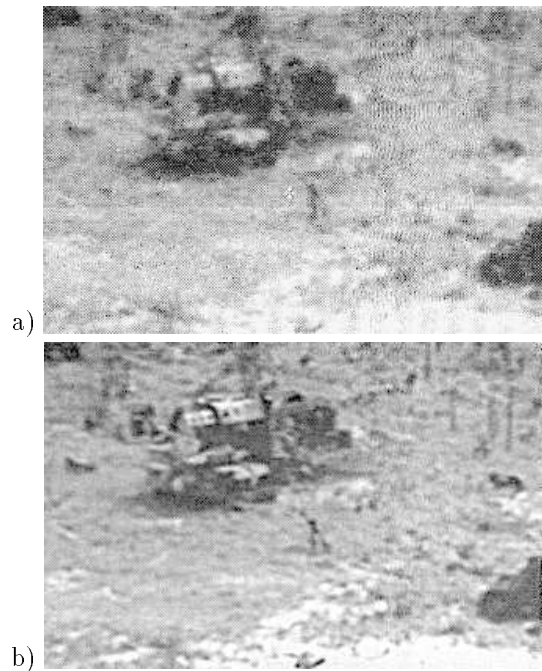


Figure 7: Mosaic-based video enhancement from a surveillance sequence of a deserted truck. (a) One out of 30 frames (all frames are of the same quality). (b) The corresponding enhanced frame in the enhanced video sequence. (All the frames in the enhanced video is of the same quality.)

resolution is even more pronounced. This method is known as Super-resolution [3, 4, 9].

The *efficiency* of using *mosaics* for *video* enhancement is due to the fact that the mosaic is an *efficient* representation of the video sequence. Rather than enhancing the frames one-by-one (as is suggested in [4]), the enhancement of the entire sequence (or layer) is done in a single step within the mosaic coordinate system, and only then are the enhanced frames retrieved from the enhanced mosaic.

Fig. 7 shows an enhanced frame from a sequence of 30 frames of a deserted truck imaged from a remote helicopter surveillance video. In this example, all the input frames were of very poor quality and very noisy. The sequence contained a single static scene that could be completely aligned using 2D alignment. The entire video sequence was enhanced by constructing a single enhanced 2D static mosaic, and then retrieving the frames from the mosaic back into their original coordinate systems (according to the inverse 2D parametric transformations).

4.4 Other Mosaic Based Applications

The benefit of mosaic images for various other applications has been recognized [10, 12]. Some of these relate to managing large digital libraries [12], and with respect to manipulating and editing video in video post-production environments.

For interactive video editing and manipulation applications, the mosaic (e.g., the key-frame or the synopsis mosaic) provides a way to significantly reduce the redium associated with the process. The editing process can be applied to the mosaic image instead of each individual frame. Also, large surfaces that are not completely visible within a single frame (e.g., signs and slogans) are available as a coherent whole in the mosaic and can be operated on directly. Since the geometric transformation between the mosaic and each individual frames is known, the results of the editing process can be used to generate a new video which contains the necessary effects.

Key frame mosaics are particularly useful for rapid browsing, since each such keyframe captures the contents of an entire subsequence. This process can replace the tedious fast-forwarding/rewinding process done in manual video browsing today.

Mosaic images are also useful for reducing the cost of indexing and search operations. This results directly from the efficiency of the representation provided by the mosaics. For instance, instead of searching for an object or feature in each frame, the search can be limited to the static mosaic and the residuals. As in the case of manipulation, the wider field of view provided by the mosaics overcomes some of the difficulties in search due to objects being split across many frames.

5 Conclusion

The problem addressed by this paper is that of developing efficient and complete representation of large video streams and efficient methods for accessing and analyzing the information contained in the video data. In this paper, we have systematically explored the issues that arise when considering how such a complete

representation may be developed. We have also described a number of different applications of the mosaic representations and illustrated them with real examples.

References

- [1] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. European Conference on Computer Vision*, pages 237–252, Santa Margarita Ligure, May 1992.
- [2] M. Irani, S. Hsu, and P. Anandan. Mosaic based video compression. In *Proceedings of SPIE Conference on Electronic Imaging*, February 1995.
- [3] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53:231–239, May 1991.
- [4] M. Irani and S. Peleg. Using motion analysis for image enhancement. *Journal of Visual Communication and Image Representation*, 4(4):324–335, December 1993.
- [5] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16, January 1994.
- [6] R. Kumar, P. Anandan, and K. Hanna. Direct recovery of shape from multiple views: a parallax based approach. In *Proc 12th ICPR*, 1994.
- [7] Rakesh Kumar, P. Anandan, and K. Hanna. Shape recovery from multiple views: a parallax based approach. In *DARPA IU Workshop*, Monterey, CA, November 1994.
- [8] Rakesh Kumar, P. Anandan, M. Irani, J. R. Bergen, and K. J. Hanna. Representation of scenes from collections of images. In *submitted to Workshop on Representations of Visual Scenes '95*.
- [9] S. Mann and R.W. Picard. Virtual bellows: Constructing high quality stills from video. In *Proc. IEEE Int. Conf. on Image Proc.*, November 1994.
- [10] P.C. McLean. Structured video coding. Master's thesis, MIT, June 1991.
- [11] Richard Szeliski. Image mosaicing for tele-reality applications. Technical Report CRL 94/2, Digital Equipment Corporation, 1994.
- [12] L. Teodosio and W. Bender. Salient video stills: Content and context preserved. In *Proc. ACM Multimedia Conf.*, 1993.
- [13] H.-J. Zhang, A. Kankanhalli, and S.W. Smoliar. Automatic partitioning of full-motion video. *Multimedia Systems*, 1(1):10–28, 1993.

Section 3.3 is used to mask out image regions that are accurately predicted from the mosaic, and to weight and prioritize the residuals in other regions before coding them.

Very-low Bitrate Transmission: Transmission requires online processing, hence, the dynamic mosaic is the natural choice for this application. The major components in the transmission codec are incremental *dynamic* mosaic construction, incremental residual estimation by comparison to the reconstructed mosaic from the previous time instance, the computation of significance measures of the residuals, and spatial coding and decoding. As is typical of any predictive coding system, the coder maintains a decoder within itself in order to be in synchrony with the receiver. The spatial coding of the images and the residuals can be based on any available technique, e.g., Discrete Cosine Transform (DCT) or wavelets.

Compression for Storage: For storage applications, it is important to provide random access to individual frames. However, the coding process need not be done in real time and can be done in an off-line mode. Therefore, the static mosaic is a natural choice for this application.

In the storage codec, the sequence is processed in batch mode, with these major steps being: *static* mosaic construction, residual estimation for each frame, significance analysis, and spatial coding and decoding of the mosaic image and the individual residuals. During retrieval, the decoded individual residuals are composed with the decoded mosaic and after performing appropriate inverse motion transformation and image window selection the individual frames can be displayed.

Fig. 5.a shows some representative frames of a surveillance video of a building viewed from a flying helicopter. The video sequence was *temporally* sampled by four (i.e., 7.5 frames/sec). The sequence was then coded at the constant bit rate of 32 Kbits/sec using mosaic based compression with with DCT spatial coding of the first frame and of the detected residuals (Fig. 5.b). For comparison, the sequence was compressed by MPEG (without mosaic pre-processing), which is the existing standard video compression method to date, at the *same bitrate* (Fig. 5.c), which resulted in significantly poorer visual quality. Note the reconstruction quality of the running soldiers on the right hand side of the building. More experimental results and details are found in our paper on mosaic based video compression [2].

4.2 Mosaic Based Visualization

One of the key benefits of mosaics is as a means of enhanced visualization. The panoramic view of the mosaic provides the scene context necessary for the viewer to better appreciate the events that take place in the video [10] (e.g., see Fig. 2). Several different types of panoramic visualizations are possible, each highlighting a different type of information. In the following set of Figures we will present examples of three

useful ways of visualizing the same video sequence using mosaic representations. Each visualization has its own use:

Key frame mosaic: Given a video sequence segmented into contiguous clips of *scene sequences* (e.g., see [13]), a static mosaic image of the most salient features in the scene can be constructed for each scene sequence. The static mosaics images (e.g., Figs. 1 and 2) represent their scenes better than any single frame, and can therefore be used as “key frames” for rapid browsing through the entire digitally-stored video sequence [12]. Other applications of the key frame mosaic are describe in detail Section 4.4.

Synopsis mosaic: While the key frame mosaic is useful for capturing the *background*, in some cases, it may be desirable to get a synopsis of the event that takes place within the video sequence. This can be achieved through a mosaic that captures the *foreground event*. The synopsis (or event) mosaic is constructed by using the residual maps (e.g., Fig. 4) as weights during the integration process, allowing for foreground moving objects to be retained within the mosaic. Fig. 6.b shows an example of a synopsis mosaic for the baseball sequences. Note that *regular* averaging of the aligned frames will *not* maintain the foreground moving objects, but will rather make them either completely disappear or significantly fade out.

Mosaic video: The panoramic visualization provided by the mosaics is useful not only as a static image, but for dynamic video visualization as well. In this case, a new video sequence is generated (called the “mosaic video”) which is a sequence of (dynamic) mosaic images. This type of visualization simulates the output of a virtual camera with desired features. The simplest example of this is stabilized video mosaic display, in which case the camera motion is completely removed. The previously shown Fig. 3 shows an example of video mosaics. Such a display has uses in various applications such as *remote navigation*, and *remote surveillance*. Similar mosaic Visualizations have also been suggested by [10].

4.3 Mosaic Based Video Enhancement

Mosaic representations can serve as a useful and efficient tool for producing *high quality stills* from video as well as enhancing an entire video sequence.

The resolution of an image is determined by the physical characteristics of the camera: the optics, the density of the detector elements, and their spatial response. An increase in the sampling rate could, however, be achieved by obtaining more samples of the imaged scene/object from a sequence of images in which the scene/object appears moving at subpixel displacements. Therefore, aligning the sequence frames over a *finer* mosaic grid can provide higher sampling rate of the background scene, and hence integrating over that grid provides higher spatial resolution. When the blur function of the camera is also known or can be computed and used for deblurring, the increase in

aligned images:

(i) A regular temporal average of the intensity values of the aligned images.

(ii) A temporal median filtering of the intensity values of the aligned images.

(iii) A *weighted* temporal median or a *weighted* temporal average where the weights can correspond to one of several choices, yielding very different types of mosaics. E.g., the weights can be chosen to decrease with the distance of a pixel from its frame center (to account for alignment inaccuracies near image boundaries); the weights can be the outlier rejection maps computed in the motion estimation process of the dominant “background” [5] (see also Fig. 4), yielding a more complete mosaic image of the background scene (i.e., less “ghost-like” traces of “foreground” objects); the weights can also correspond to the *inverse* of these outlier rejection maps, yielding a mosaic image which contains a panoramic image not only of the scene, but also of the foreground *event* that took place in that scene sequence (see Fig. 6).

(iv) Integration in which the *most recent information*, i.e., that which is found in the most recent frame, is used for updating the mosaic (see Fig. 3).

(v) Alternative integration schemes for image enhancement, such as Super-resolution [3], to produce mosaic image whose resolution and image quality surpasses those of any of the original image frames. See more details in Section 4.

3.3 Significant Residual Estimation

The complete sequence representation includes the mosaic image, the transformation parameters that relate the mosaic to each individual frame, and the residual differences between the mosaic image and the individual frames. To reconstruct any given frame in its own coordinate system, the mosaic image is warped using the corresponding mosaic-to-image transformation and composed with the residuals for that frame. In the case of the static mosaic, the differences are directly estimated between a single reference (static) mosaic and each frames, and the reconstruction is straight forward, whereas in the case of the dynamic mosaic, however, the residuals are incremental, being with respect to the previous mosaic image frame. In this case the reconstruction proceeds sequentially from frame to frame.

Residuals between the current frame and the mosaic-based predicted frame occur for several reasons: object or illumination change, residual misalignments, interpolation errors during warping, and noise. Of these the object changes are the most semantically significant, and in some cases the illumination changes are as well.

The efficiency of the representation can be maximized by assigning a significance measure to the residuals, and using those to weight the residuals. An effective way of determining semantically significant residuals is to consider not only the residual intensity but also the the magnitude of *local residual motions* (i.e., the local misalignments) between the frame predicted from the mosaic and the actual frame. To approximate the magnitudes of the residual motions, an estimate

$S_t(x, y)$ of the normal flow magnitude at each pixel (x, y) at time t is computed:

$$S_t(x, y) = \frac{\sum_{(x_i, y_i) \in N(x, y)} |I_t(x_i, y_i) - I_t^{Pred}(x_i, y_i)|}{\sum_{(x_i, y_i) \in N(x, y)} |\nabla I_t(x_i, y_i)| + C}$$

where: I_t is the frame at time t , I_t^{Pred} is the predicted frame from the mosaic at time t , $\nabla I_t(x, y)$ is the spatial intensity gradient at pixel (x, y) in frame I_t , $N(x, y)$ is a small neighborhood of pixel (x, y) (typically a 3×3 neighborhood), C is used to avoid numerical instabilities and to suppress noise. Fig. 4 shows an example of significant frame residuals detected for the static and the dynamic representations in the table-tennis sequence.

Although the same significance measure is used with the static and the dynamic mosaic, the locations and magnitudes of the significant residuals differ between the two schemes even when applied to the same sequence. In the case of the static mosaic, the significance measures in regions of objects that move with respect to the background are usually larger than in the case of the dynamic mosaic, as moving objects tend to blur out or even disappear in the static mosaic, and hence the changes will be significant. In the dynamic case, the mosaic is constantly being updated with the most recent information, and therefore, the changes in image regions that correspond to independently moving objects will be smaller between the predicted and actual frame. In the dynamic mosaic, however, more residuals will be obtained at image boundaries than in the static case.

4 Mosaic Applications

The most obvious applications of mosaic representation are *video compression* (as mosaics are efficient scene representations) and as a means of *visualization* (as mosaics provide a wide and stabilized field of view). These will be discussed in sections 4.1 and 4.2. However, mosaics are also useful in other applications, such as scene change detection, efficient video search and video indexing, efficient video editing and manipulation, and others.

4.1 Mosaic Based Video Compression

Since mosaics provide an efficient means of representing a video sequence, the most natural application to consider is video image compression. The differences between static and dynamic mosaic representations that were outlined in the previous sections lead to differences in the two types of codecs, one for real-time transmission over very-low bitrate channels, and the other for efficient storage of video sequence.

The static mosaic, or the first frame in the dynamic case, may be compressed by any known method for lossy still image coding. All subsequent frames are predicted by the computed 2D parametric transformation from the static or dynamic mosaic and only the significant missing residuals are coded. For each frame, such motion information, coded in a lossless or lossy manner, needs to be stored or transmitted along with the residuals. The significance mask described in



Figure 3: Evolution of the dynamic mosaic images of the table-tennis game sequence. *Left column:* Three frames from the original sequence. *Right column:* The corresponding dynamic mosaic images. Note that the position of the player and the crowd are constantly being updated to match the current frame.

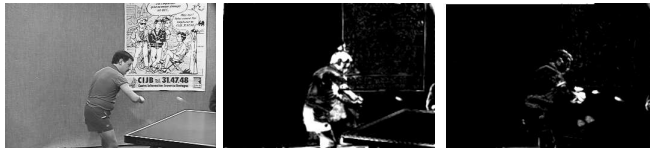


Figure 4: The residual maps of static vs. dynamic cases. *Left:* A single frame from the table-tennis sequence. *Middle:* The residual map computed for the corresponding frame in the static representation. The brighter values signify more significant residuals. *Right:* The residual map computed for the corresponding frame in the dynamic representation.

2.4 Multiresolution Mosaic

In order to handle the variations in image resolution that occur (due to camera zoom and other reasons), we introduce a *multi-resolution* mosaic data structure, which captures information from each new frame at its closest corresponding resolution level in a *mosaic pyramid*. It is a sparse (*spatial*) pyramid in the sense that the resolution levels are not complete (certain mosaic regions may be represented at high resolution, others only at low resolution).

2.5 Other Mosaic Representations

To handle scenes and sequences with additional complexity, the 2D mosaic can be extended to *parallax based mosaics* to handle scenes containing significant 3D parallax, and to *layered mosaics* to handle transparency and multiple motions. These extensions are described in greater detail in [8].

3 Mosaic Construction

A mosaic based representation is constructed from all frames in a scene sequence, giving a panoramic view of that scene. Three steps are involved in this process: the *alignment* of the images in the sequence, the *integration* of the images into a mosaic image, and the *computation of significant residuals* between the mosaic and the individual frames.

3.1 Image Alignment

Image alignment depends on the chosen world model and motion model. The alignment can be limited to 2D parametric motion models, or can utilize more complex 3D motion models and layered representations. The examples in this paper utilize 2D motion models, in particular a 6-parameter affine transformation and an 8-parameter quadratic transformation, to approximate the motions between two images. Work on 3D image alignment is currently in progress and described in [7, 6, 8].

The alignment of *all* image frames in the video sequence *to form the mosaic image* can be performed in one of the following ways: (i) Successive images are first aligned, then the computed 2D motion parameters are cascaded to determine the alignment parameters between any frame to a chosen reference frame. (ii) Each image is aligned directly to the current composite mosaic image using the constructed mosaic image as the reference (i.e., fixed coordinate system), or (iii) The current mosaic image is aligned to the new image, using the new image as the reference (i.e., dynamic coordinate system).

To align two images we use the hierarchical direct registration technique described in [1, 5]. This technique first constructs a Laplacian pyramid from each of the two input images, and then estimates the motion parameters in a coarse-fine manner. Within each level the Sum of squared difference (SSD) measure is used as a match measure: $E(\{\mathbf{u}\}) = \sum_{\mathbf{x}} (I(\mathbf{x}, t) - I(\mathbf{x} - \mathbf{u}(\mathbf{x}), t - 1))^2$ where $\mathbf{x} = (x, y)$ denotes the spatial image position of a point, I the (Laplacian pyramid) image intensity and $\mathbf{u}(\mathbf{x}) = (u(x, y), v(x, y))$ denotes the image velocity at that point, and the sum is computed over all the points within the region and $\{\mathbf{u}\}$ is used to denote the entire motion field within that region.

This measure is minimized with respect to the quadratic image motion parameters:

$$\begin{aligned} u(\mathbf{x}) &= p_1 x + p_2 y + p_5 + p_7 x^2 + p_8 xy \\ v(\mathbf{x}) &= p_3 x + p_4 y + p_6 + p_7 xy + p_8 y^2 \end{aligned}$$

The objective function $E(\{\mathbf{u}\})$ is minimized via the Gauss-Newton optimization technique. After iterating a certain number of times within a pyramid level, the process continues at the next finer level. More details can be found in [1, 5].

3.2 Image Integration

Once the frames are aligned (or, in the dynamic case, the current mosaic and new frame are aligned), they can be integrated to construct the mosaic image (or *update* the mosaic, in the dynamic case). One of several schemes can be chosen for integrating the



Figure 1: Static mosaic image of a table-tennis game sequence. *Top row*: Three out of a 300 frame sequence obtained by a camera panning across the scene. *Bottom row*: The static mosaic image constructed using a temporal median.

image displays a sharp image of the background with no trace of the two outfielders. In both examples, a 2D motion model was used to align the images.

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). Section 3.3 presents a method for detecting such “residuals”. 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. Applications of the static mosaic are described 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, whereas in the latter case the mosaic is viewed within a moving coordinate system. 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. Note that in the dynamic mosaic the moving objects do not blur out or disappear (as opposed to the static mosaics in Figs. 1 and 2), but are constantly being updated.

The *complete* dynamic mosaic representation of the video sequence constitutes 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 infers a difference in

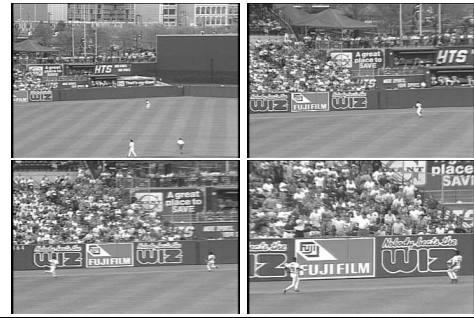


Figure 2: Static mosaic image of a baseball game sequence. *Top rows*: Four out of a 90 frame sequence obtained by a camera panning from right to left and zooming in on the runners. *Bottom row*: The static mosaic image constructed using a temporal median.

the “residuals” that are not represented by the mosaic. Fig. 4 shows an example of frame residuals detected for the static and the dynamic representations in the table-tennis sequence. In general, since changes between successive frames are relatively small, the amount of “residual” information in the dynamic mosaic will be smaller than that in the static case. The dynamic mosaic is therefore a more *efficient* scene representation than the static mosaic. However, due to its *incremental* frame reconstruction, it lacks the capability of *random access* to individual frames, which is essential for video manipulation and editing. The dynamic mosaic, however, is an ideal tool for low bitrate transmission (see Section 4).

2.3 Temporal pyramid

A natural extension of the static and dynamic mosaics is the temporal pyramid, which is a hierarchy of static mosaics whose levels corresponds to different amounts of temporal integration. This hierarchical organization is similar to spatial image pyramid representation. The finest level corresponds to the temporal sampling of the input sequence. The successive coarser levels are based on successively increasing temporal integration and downsampling. An efficient way to represent such a pyramid is in the form of a Laplacian pyramid. The *coarsest* level consists of a single static mosaic, and the succeeding levels represent residuals estimated over various time scales.

Mosaic Based Representations of Video Sequences and Their Applications

Michal Irani, P. Anandan, and Steve Hsu
David Sarnoff Research Center
CN5300, Princeton NJ 08543-5300
Email: michal@sarnoff.com

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 the 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 mosaic 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 transformation (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 analysis of residual analysis not captured by the mosaic. 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.

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 of the temporal (or “dynamic”) evolution of phenomena. A number of applications that have recently emerged that involve processing the entire information within video sequences. 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) [10, 11, 12, 9]. 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. 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, so that the sequence can be fully recovered from the mosaic image, has not been adequately treated.

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 to video compression, video visualization, video enhancement, and a other applications.

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. 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 [10, 12, 11, 9, 7], 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 [13]), and a static mosaic image is constructed for each scene, by aligning all frames of that subsequence to a *fixed* coordinate system. The aligned images are then integrated using different types of temporal filters into a mosaic image, and the significant residuals (i.e., data not captured by the mosaic) are computed for each frame of relative to the mosaic image.

Examples of static mosaic images are shown in Figs. 1 and 2. In Fig. 1, the constructed mosaic image displays a sharp background, with blurry crowd, and a ghost-like player. In Fig. 2 the constructed mosaic