# Salient Video Stills:

# Content and Context Preserved

**Laura Teodosio and Walter Bender**
MIT Media Laboratory, Cambridge, MA 02139 USA

## ABSTRACT

A new class of images called *salient stills* is demonstrated and a software development platform for their creation is discussed. These images do not represent one discrete moment of time, as do a photograph or single video frame. Rather, one image reflects the aggregate of the temporal changes that occur in a moving image sequence with the salient features preserved. By the application of an affine transformation and non-linear temporal processing, multiple frames of an image sequence, which may include variations in focal-length or field-of-view, are combined to create a single still image. The still image may have multi-resolution patches, a larger field-of-view, or higher overall resolution than any individual frame in the original image sequence. It may also contain selected salient objects from any one of the sequence of video frames. The still can be created automatically or with user intervention. A by-product of the salient still process is a structured representation of moving image data.

[*Keywords*: image processing, optical flow, structured video, image reproduction]

## INTRODUCTION

Photography is a discrete medium. It suspends the world for one instant, rendering it silently within the two-dimensional boundaries of an image frame. There exists in that photographic representation just one visual point of view and one level of spatial resolution. Video, by contrast, is a temporal medium. The progression of images suggest motion. A sequence of frames may consist of zooms, tilts, pans and other camera movement that over time change the viewer's vantage point, field-of-view and perceived resolution. Changes in the camera settings may cause the viewer's attention to shift from one part of a scene to another. Objects can move in and out of the scene relative to the camera motion.

Any single frame of a continuous sequence of frames accurately portrays only one discrete moment of time, camera focal-length and orientation. While this frame may articulate a moment, it is an incomplete representation of the space/time continuum recorded by the sequence of moving images. The single frame cannot capture the intended expression of the multiple frames.

However, the information in a series of video frames can be culled and transformed into a resulting single still image. The process described below facilitates the creation of a variety of synthetic and composite still images from the multiple discrete frames of a zoom and/or pan sequence. Images are created that are a synthesis of the information extracted from a portion of the entire image sequence. Derived from an intersection of photography and video, the processed images are called *salient stills*. Salient stills attempt to retain much of the original content (detail) and spatial and temporal extent (context) of the original sequence while condensing data.

One example discussed in this paper is an image with variable resolution patches. The overall scene is captured at low-resolution. High-resolution patches correspond to areas in the scene where the camera zoomed in. This variation in resolution, rather than being disconcerting, is a narrative tool. Much like the focus-pull of cinema, it draws the user's attention to those salient parts of the scene that commanded the attention of the camera operator or were deemed important by the creator of the salient still. Other examples include images which contain composited objects from the scene, and images with a field-of-view greater than that of any one of the individual video frames – essentially a synthesized panoramic image.

A salient still from zoom is created when multiple frames from a video zoom are combined into a single "multi-resolution" image. A salient still from pan is created when multiple frames from a video pan are combined into a single "panoramic" still image. The methodology does not need to distinguish between the two shot types, so zooms and pans can be intermixed freely in the resulting still image.

### Redundancy

Usually when a video camera samples the world in front of it, much of the information captured is repeated in a number of frames. For example, in a sequence in which the camera "zooms in", locations in the physical world are captured many times, but objects central in the scene are imaged with greater resolution in each successive frame. The relative position of the objects in each successive frame may also change, for instance, some objects will be out of the field-of-view and no longer will be imaged. Likewise, since there is usually a great deal of overlap in a pan sequence, the same locations are imaged many times. As in the zoom, the relative position of the objects to the frame changes frame-to-frame; some objects fall out of the field-of-view, while new objects become revealed.

The typical sampling of a scene by a video camera results in a redundant representation, e.g., the same area in the scene may be sampled multiple times, but the redundant samples are likely not to be aligned from frame to frame. The salient still process, which

From the Proceedings of the ACM Multimedia Conference, Anaheim, August, 1993.

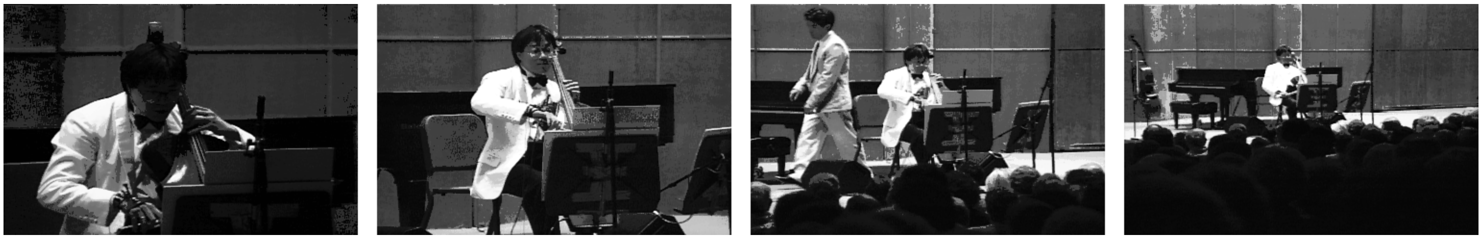

Figure 1: Four frames from a video zoom featuring the cellist Yoyo Ma.

aligns all of the frames in a sequence, is able to exploit this redundancy in four ways: (1) Transient noise can be reduced in the image by examining multiple samples of the same point over time, (2) Additional image resolution can be achieved by combining samples that have been taken at phase shifts that differ either by changes in camera orientation or focal-length, (3) When the camera orientation is changing, an increased field-of-view can be achieved by compositing images that have been correlated by regions of overlap, (4) Transient objects can be located and extracted by comparing the composite image to individual frames.

One result of the salient still process is the extraction of "resolution-out-of-time". In image processing terms, this increased resolution corresponds to a higher bandwidth signal, and, in cognitive terms, it corresponds to an image with more content information than any individual frame in the sequence.

**An Example**

Four frames from a 12-second zoom sequence of the cellist Yoyo Ma on stage during a performance at Tanglewood are shown in figure 1. Imagine the task at hand is to choose a single image to represent this performance, perhaps to print the image in a newspaper.

One may initially be inclined to chose the close-up frame of Mr. Ma since this is rendered with the most detail per pixel per "area of interest". But this image does not provide the contextual information of the ambiance of Tanglewood, nor does it relay the fact that there is a live audience. In contrast, the far shot retains the ambiance, but it does not provide enough spatial detail of the performer to recognize him as Mr. Ma or to see that he is playing a hypercello. A frame selected from the middle of the zoom contains the musical assistant walking across the stage, detracting from Mr. Ma as the center of focus.

The solution to finding an image with both enhanced visual quality and semantic saliency is not to use any single frame, but to use all of them. The close-up frames render Mr. Ma with the needed detail, the far shot gives context, and middle frames provide enough redundancy in data to assist in the automatic removal of the non-salient musical assistant. Figure 2 compares the resolution of the far shot of the sequence with the resolution of the resulting still.

Figure 3 shows the resulting image of using data from all the video frames. Notice that the musical assistant is nowhere to be seen. Additionally, the resultant image is much larger than any of the individual frames. The single digitized frames are 640 x 480 pixels, whereas the resulting salient still is 3000 x 1500 pixels. The central portion of the image has the same detail per area of interest as the close-up frames and the resolution decreases towards the edge of the frame.

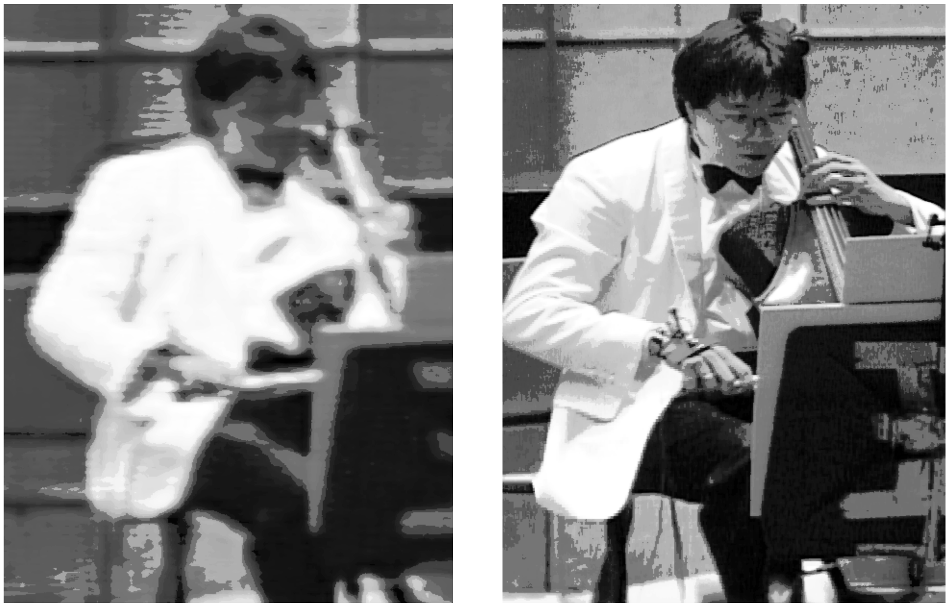

Figure 2: On the left, the "area of interest" extracted from a frame in the original sequence (the right hand image in figure 1). On the

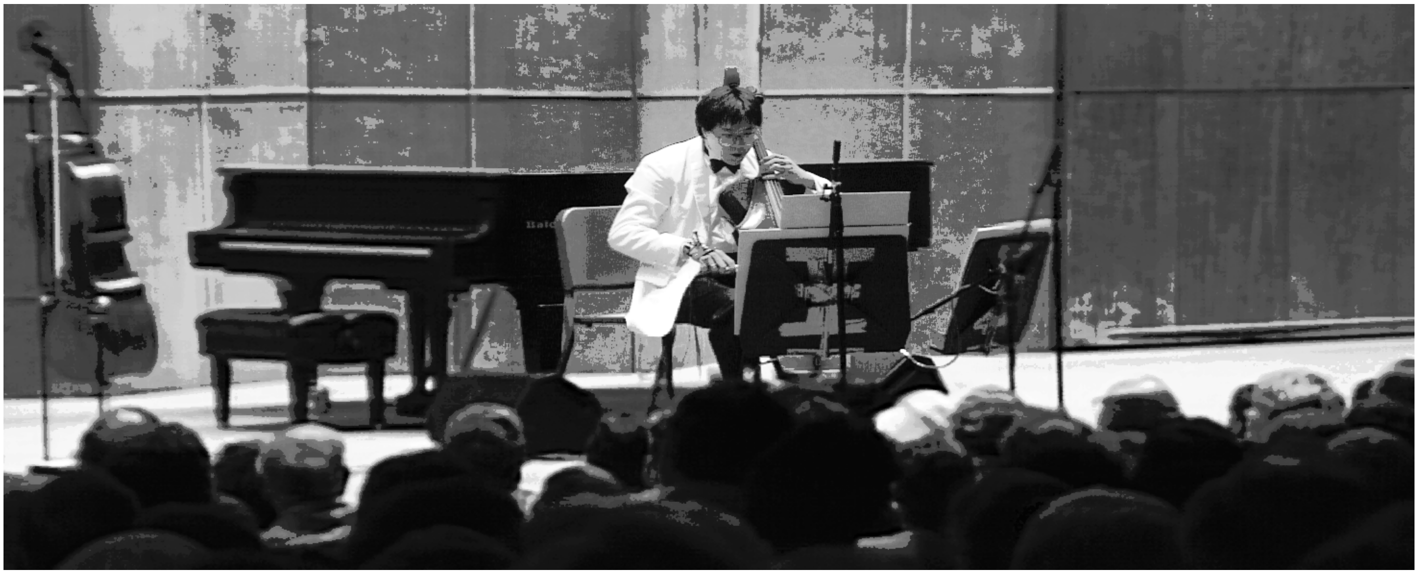

Figure 3: The salient still.

The resulting picture takes the best features of both an established wide shot and a high detail close-up and merges them into one. Additionally, the high-resolution patch in the center of the image draws the user's attention to the area of interest.

If it is deemed that the musical assistant is a salient feature of the scene, he can be composited into the final still image. Figure 4 shows a still of the same sequence where the musical assistant was added back into the image multiple times to show his traversal across the stage. In order to convey the temporal nature of the assistant's actions, the degree of translucency used to render him is a function of time in the scene, i.e., the earlier he appeared in the scene, the more transparent he is in the composite.

### Structured Representation

In order to create salient stills we manipulate video in a structured manner moving beyond the frame and into the elements, e.g., foreground and background. Objects and movement get modeled. This decomposition of the signal from its original frame-by-frame packaging allows new degrees of freedom in the compression, manipulation and visualization of this video data. This representation of a visual scene as a series of manipulable objects is the basic model of synthetic images in computer graphics. It has only begun to be explored in the world of 'real' images as discussed in [13], [16], [7], [17], and [11].

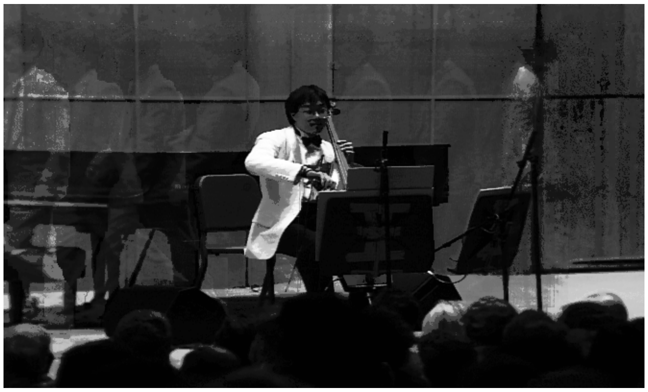

Figure 4: Temporal Musical Assistant

The system for building the still consists of determining global camera motion, building a 3-D space/time representation of the video data, constructing a high-resolution background scene from which other moving objects can be detected, extracting objects, and finally, composing the still. The user is able to assist in the determination of salient features which are to be included in the still.

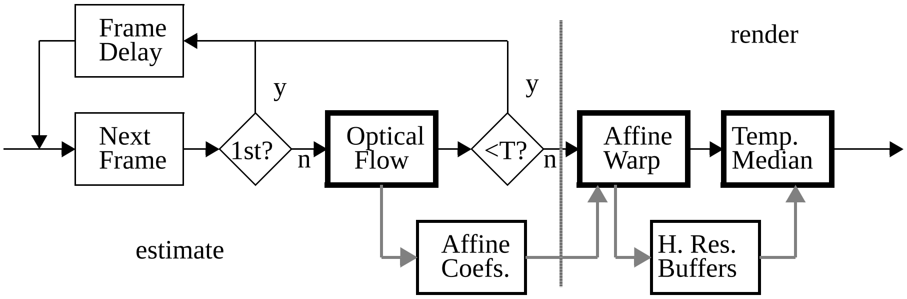

Figure 5: System block diagram.

## METHODOLOGY

There are three steps to building the high-resolution images: (1) The optical flow between successive frame pairs is calculated. (2) Successive affine transformations, calculated from the flow, are applied. These translate, scale and warp each frame into a single, high-resolution raster. (3) A weighted temporal median filter is applied to the high-resolution image data, resulting in the final image.

### Optical Flow

The output of a video camera is a set of discrete samples of a space/time continuum of the world in front of the camera. The creation of a salient still allows the recovery of this continuum. Camera movement between frames is modeled using an optical flow analysis [6], [8], [9], [10]. By characterizing this frame to frame difference as an affine transformation, zooms, as well as translational camera movement, are recovered.

The optical flow can be modeled by a continuous variation of image intensity as a function of position and time:

$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$

Assuming that the motion field is continuous everywhere, the right hand side of the equation can be expanded using a Taylor series, (the higher order terms are ignored):

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} = -\frac{\partial I}{\partial t}$$

This equation usually is solved by modeling image motion as translational in the *x*, horizontal, and *y*, vertical, directions. This is sufficient to model the pan and tilt of a camera, but it does not model zoom. However, if motion is modeled as an affine transformation, zooms in combination with other camera movement are recovered. The specific algorithm and implementation comes from Bergen et al.[2]. Following their notation, if a pattern is moving with a velocity *p(x,y)* then modeling only translational changes:

$$p(x, y) = p_x(x, y), p_y(x, y)$$

So, if velocity *p* is described by six parameters, $a_x$, $b_x$, $c_x$, $a_y$, $b_y$, $c_y$, which describe the affine movement, the velocity equations become:

$$p_x(x, y) = a_x + b_x x + c_x y$$
$$p_y(x, y) = a_y + b_y x + c_y y$$

where $a_x$ and $a_y$ are pure translation terms in the *x* and *y* directions in units of pixels, $b_x$ is a percentage scaling factor for *x* in the *x* direction, $c_x$ is a percentage rotation factor for *x* in the *y* direction, $b_y$ is a percentage rotation factor of *y* in the *x* direction, and $c_y$ is a scaling factor for *y* in the *y* direction. From the above equations, it is observed that a change of focal-length of a camera lens is described by the $b_x$ and $c_y$ terms. Pans and tilts are described by the $a_x$ and $a_y$ terms.

To determine these affine parameters, first a Gaussian pyramid is constructed for each pair of frames (*t* and *t+1*). Computation begins at the lowest level and is refined at each successive level by computing the residual motion.

This method was used principally to detect global camera motion and works well for static scenes where there is only camera motion; unfortunately, the real world of video is not filled with many shots of this type. Therefore a few different methods of reliability were built into the affine estimates, including masking noisy regions, using a priori knowledge of shot type expected to adjust estimates, and smoothing the data by passing a spline through the samples. For a more detailed explanation of these methods see [16]. Figure 6 shows a vector field representing the optical flow from successive frames in the Ma sequence.

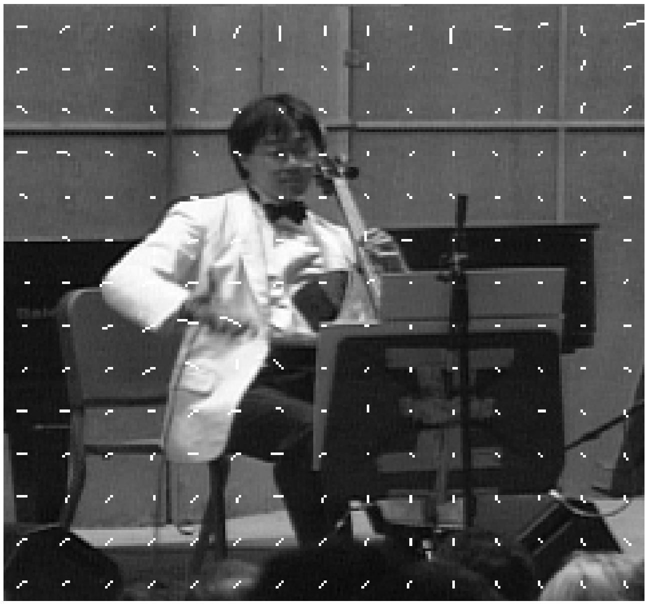

Figure 6: The vector field representing the optical flow between successive frames in the Ma sequence.

### Warping Into Continuous Space/time Raster

A 3-D space/time continuum is built for the video sequence. The result is a video volume where spatial location in the world is on the X and Y axes, and time is on the Z axis. A vector passing through the volume perpendicular to the first image plane will pierce the same spatial location in the world of each image. For example, the second image of a pan left sequence is adjusted right so that the two frames line up; the second image of a zoom sequence is scaled so that it appears the two images were captured with the same focal-length lens (see figure 7).

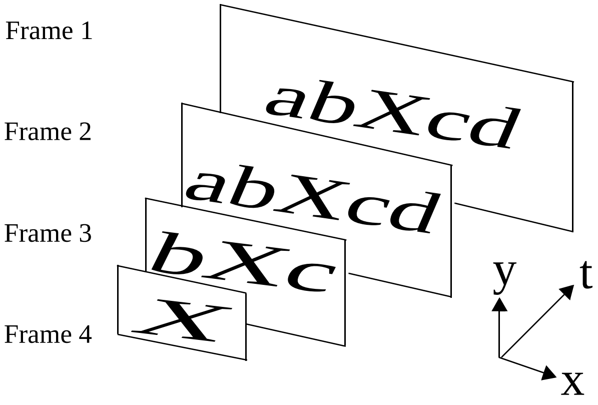

Figure 7: The motion between the images is recovered and used to warp all the images into a common raster.

The affine transform$_{(t \rightarrow t+1)}$ is used to warp image$_{(t)}$ into the space occupied by image$_{(t+1)}$. Image$_{(t)}$ is further warped into the space occupied by image$_{(t+2)}$ by applying the affine transform$_{(t+1 \rightarrow t+2)}$. The following equations are used to sum the affine parameters:

$$\begin{aligned} ax_{new} &= ax_1 \times bx_2 + ay_1 \times cx_2 + ax_2 \\ ay_{new} &= ax_1 \times by_2 + ay_1 \times cy_2 + ay_2 \\ bx_{new} &= bx_1 \times bx_2 + by_1 \times cx_2 \\ by_{new} &= bx_1 \times by_2 + by_1 \times cy_2 \\ cx_{new} &= bx_2 \times cx_1 + cx_2 \times cy_1 \\ cy_{new} &= by_2 \times cx_1 + cy_1 \times cy_2 \end{aligned}$$

The result of the recursion of this procedure is that all of the frames in the sequence are mapped into a common raster. It is important to note that, following the mapping, the images are no longer at a uniform resolution. Images that were taken with a short focal-length were scaled more than images taken with a long focal-length, thus the former are of lower resolution than the latter. However, there is now a global correspondence between pixel positions and locations in the original scene.

**Extraction Of High-Resolution Or Extended Buffer**

From this intermediate space/time volume representation, an extended scene or high-resolution buffer can be created. Various approaches can be taken in the creation of this raster. The still image can be pieced together bit-by-bit using the pixels unaltered as they are rendered in the space/time volume temporally. For example, in a pan sequence, the first occurrence of each point in the world is used for the value of the corresponding point in the final altered image.This method works fine for static images, but for scenes containing moving objects or lighting changes the resulting still may be a chaotic rendering of multiply imaged objects. Another approach would be to calculate the average value of every sample corresponding to a location in the scene. This method results in "ghosts" with no dramatic increase in image resolution. A potential remedy is to apply an operator over a region of pixels as they change over time to determine the corresponding pixel value in the final image. Ideally, this operator will preserve all of the detail found throughout the sequence, while eliminating noise and extraneous motion in front of the camera. We experimented with temporally weighted median operators, and found them effective.

A simple median operation is generally effective at eliminating ghosts [3]. Any transient data, such as a moving object that is stationary for less than half of the frames, are eliminated from the final image. However, a simple median operation is not optimal when zoom sequences are being manipulated, since it gives equal weight to both pixels interpolated between coarse samples (short focal-length) and pixels originating from fine samples (long focal-length). A linear weighting based on the position of each frame in the frame sequence depreciates interpolated pixels. Assuming the zoom is directed outward, the first occurrence of the pixel is weighted more that the last. A refinement, which accounts for variations in the pace and direction of the zoom, assigns a weight according to the scale terms, $b_x$ and $c_y$, from the affine estimator. Figure 8 shows conceptually from where in the image sequence the data in the still image have been derived.

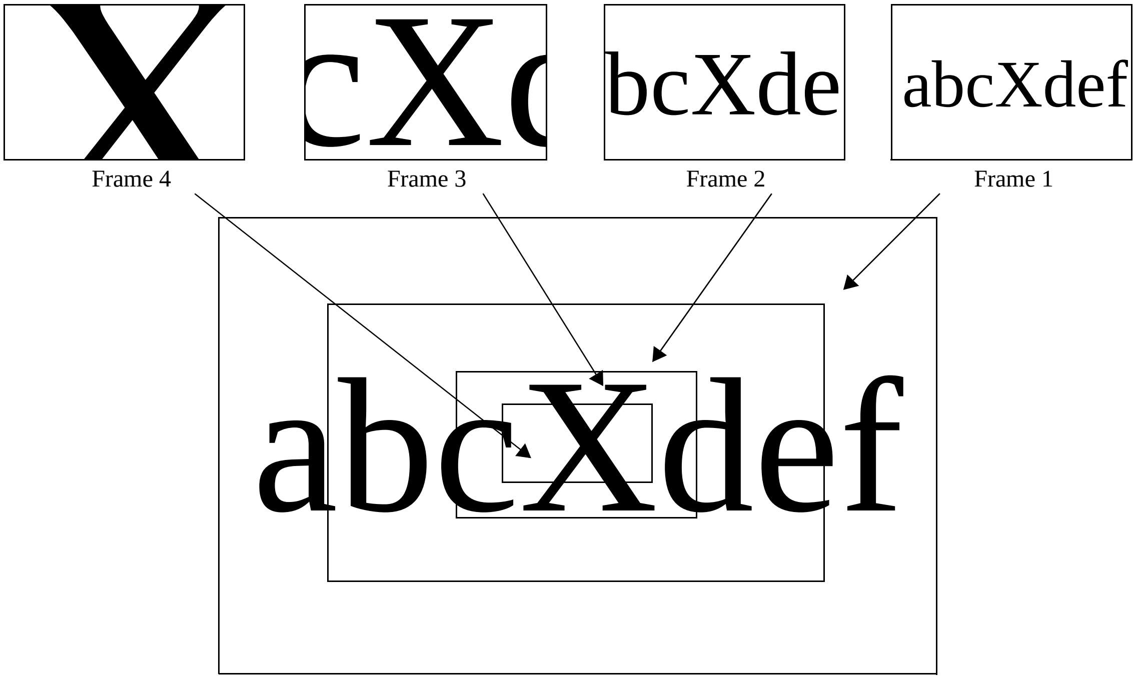

Figure 8: Synthetic image created from zoom sequence. The four boxed regions in the composite image are taken from the warped frames shown above. The high-resolution image is created by applying a temporal median over every pixel in the warped frames; i.e., pixels from frames 1 to 4 contribute to the final determination of any pixel in the composite. Since frames 1 through 3 are scaled by

In this resulting image, the center of the salient still is the area of highest resolution. One could imagine, however, a situation where the camera operator zoomed in on an object, zoomed back out, panned to the right, then zoomed back in again. The resulting salient still could have two high-resolution patches.

## ENHANCEMENTS

There have been many enhancements made to the basic salient still process described in section 2, including improvements in the estimation of frame-to-frame correspondence, better image segmentation and compositing, and efficiency considerations.

### Auto-Aperture Compensation

Illumination change is one factor in determining the correspondence between frames in an image sequence that is not accounted for by either the optical flow estimation or the affine transformation. By calculating an overall illumination level change between frames, a more accurate estimation of flow is achieved. The normalization of the illumination levels between frames eliminates some discontinuities in the composited still image.

### Reduction of Affine Coefficients

Under typical circumstances, the affine shear coefficients, $c_x$ and $b_y$, are negligible and can be eliminated from the calculations. Also, under typical circumstances, the affine scale coefficients, $b_x$ and $c_y$, should be identical. In practice, the mean of these coefficients is used.

### Memory Usage Reduction Due To Inverse Affine

As described in discussion of the methodology, each frame in the image sequence is scaled to a high-resolution buffer from which the salient still is then extracted. Although this works, it is very expensive in terms of image storage; every image in the sequence must be stored at the resolution of the final output image. An alternative is to again inverse the direction of the transformation. After having determined the mapping of each image into the high-resolution space, it is possible to scan this space, and map it back to the appropriate pixels in the original sequence. This method requires the same number of calculations, but only one high-resolution buffer.

$$p'_x(x, y) = \frac{a_y c_x - a_x c_y + c_y x - c_x y}{b_x c_y - b_y c_x}$$

$$p'_y(x, y) = \frac{-(a_y b_x) + a_x b_y - b_y x + b_x y}{b_x c_y - b_y c_x}$$

### Chromatic Redundancy

Since the optical flow algorithm is “color blind”, all the computations of optical flow, warping and temporal median were done on the Y, or achromatic channel, of a YUV color representation. The same manipulations are then performed on the corresponding pixels in the chromatic channels.

### Character Extraction

Since the composite image is a result of a median operator, any object that is stationary relative to the camera motion for more than some duration of time (50% of the frames for a simple median) is incorporated in the composite, whereas objects that are moving relative to the camera “disappear.” This somewhat arbitrary selection of foreground objects which are discarded or retained is not always appropriate. Additional contextual cues needed to determine the object’s saliency can come from the salient still creator, a log, or description of the video footage.

In the case of the Ma sequence, the character extraction was performed by comparing frames of the sequence with the “actorless” background scene that was extracted from the temporal median process. The regions that differed were considered foreground objects and extracted. There were a few problems with this method, most significant was the temporal variation in the sampled pixels due to the methods of recording, (e.g., aperture change) or changes in the environment such as shadows cast from moving actors. In most cases a noisy segmentation was extracted and then the user was asked to assist in the final segmentation. Other methods of extracting multiple motions can be found in [2] and [5].

### Compositing

Once a background scene is extracted, characters or objects considered salient can be composited back into the set. These regions can be composited at higher resolution or with changed color or luminance. In order to ensure a seamless blend of regions as in the Ma temporal composite, a laplacian pyramid blending method is used [1].

## EVALUATION OF METHODOLOGY

The results depend upon three factors: the accuracy of the flow vector, the camera attributes and operation, and scene characteristics.

Optical flow incorporates an incomplete model that assumes a world made of a two-dimensional rubber sheet, e.g. all motion is affine (non-perspective) and rigid. While this model is sufficient for many circumstances, it does not work well if there is the possibility of excessive image noise, transient image features, (e.g., actors moving relative to the camera), or changes in perspective between images. Truck and track shots, both of which involve moving the focal plane of the camera relative to the scene, will cause errors due to changes in perspective. The former type of shot will cause the image to bow out from the image plane, whereas the latter will cause objects to unwrap onto the surface of the image plane. In the case of pans and tilts, if the objects in the field-of-view are proximal, there is some perspective distortion. Object occlusion is another source of violation of the optical flow model; whenever the focal-plane of the camera or objects in front of the camera move relative to the scene, occluded or revealed elements distort the estimation of flow.

The theoretical limit of how much resolution can be added to the salient still due to redundant sampling has to do with both the camera modulation transfer function (MTF) and its operation. While there is no theoretical restriction on using zooms to enhance resolution of points-of-interests, and pans and tilts to cover a wider field-of-view, there is a hard limit on how much additional resolution can be extracted due to tracking the phase shifts between frames. This is discussed in [14].

## DISCUSSION OF APPLICATIONS

Salient stills and their methodology are useful in the areas of compression and visualization of temporal video data.

Sometimes it is desirable to have the ability to print a still image from video. But there are inherent problems in this task. First, the spatial resolution of video is below that of a print medium. Additionally, the video image collector is capturing for a temporal medium; following movement often takes precedence over frame composition. As a result, there might not be one frame composed well enough to stand on its own as a still, or perhaps, there is information in the moving image that would be distracting or misleading in a still frame (the multiresolution Yoyo Ma image is an example of such a still). The salient still process can be used to create still images with increased spatial resolution that are suitable for print publication. The use of variable spatial resolution to indicate regions of interest and temporal transitions is a mechanism whereby the narrative devices of the cinematography can be "transcoded" into those of photography. The patches of high resolution which correspond to the camera's changes in focal-length need not be restricted to a single region within the still image, as in the Yoyo Ma example. If the camera zoomed in on a region and then panned, the result would be a "stripe" of high resolution. If the camera zoomed in and out while panning, the result would be multiple regions of high resolution.

The salient still technology allows for the easy creation of panoramic images without the use of expensive specialized hardware such as a Glubuscope camera. Spaces may be imaged not only 360 degrees out from the focal center of a lens, but also, they may be imaged 360 degrees pointing into some focal center, around a building, for example. The resulting still from this process can also be used as an interface device for the individual frames of video data. This latter process was done in a real-space viewing system of navigable video data of a Russian palace [15].

Salient stills can be used as icons in a digital video database. A video data icon is the synthesis of the salient features of some temporal duration of footage into one image. Saving time, the user of the digital video database would only have to glance at a single salient still rather than view the whole sequence of data. This can be very useful if the database is quite large, for example containing hundreds or thousands of hours of footage. Additionally, the intermediate space/time volume can be used to represent this temporal data as well. The camera motion of the sequence (pans, tilts and zooms) can easily be ascertained from the contours of the space/time volume. An interesting variant on visualizing video databases is described in [4].

The methodology affords new flexibility in the creation of traditional photomontages. From the intermediate space/time volume, the user can traverse the video data and select sections of the volume both in space and time to be used in the synthesis of the final image. Pieces of individual frames can be used for some sections, and temporally processed data can be used in other sections.

This technology can be exploited to create not only single frame representations of moving image data but also more dynamic systems. One could imagine a "video radar", where a camera will continuously film an entire room by panning around 360 degrees [12]. In this scenario, whenever an object is encountered that is not a permanent fixture in the room, the object will be imaged onto the static background image of the space; the system effectively tracks all new activity in the room. The compositing methodology also could be used to highlight characteristics about the activity – the time a person entered the room and the urgency with which she should be noticed.

One by-product of most digital video compression systems is the generation of an inter-frame motion model. Although this model is typically optimized for prediction rather than interpolation, it may be possible to use the model to estimate the inter-frame affine transform required in the salient still process. This would greatly reduce the computational overhead of the process.

## CONCLUSION

We have developed the notion of salient stills as well as methods for constructing them. This work is part of a larger body of work in Media Transcoding which is aimed at automatic conversion of one medium into another. In contrast to McLuhan's maxim, when the message is news or information, the medium is not the message. With the salient still, we are trying to create a high-resolution still image of reproduction quality from video. In this process, we are trying to translate a narrative told in the language of the cinematographer into the language of the photographer. This ability to adapt media is requisite in a heterogeneous communications environment.

## REFERENCES

1. E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden. Pyramid methods in image processing, *RCA Engineer*, Nov/Dec 1984.

2. J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. Computing Two Motions from Three Frames, David Sarnoff Research.

3. T. Doyle and P. Frencken. Median Filtering of Television Images, *Technical Papers Digest, International Conference on Consumer Electronics*, June 1986.

4. Edward Elliot. Multiple Views of Digital Video, MIT Media Lab, Interactive Cinema Working Paper, March 1992.

5. Bernd Girod and David Kuo. Direct Estimation of Displacement Histograms, *Optical Society of America Meeting on Understanding and Machine Vision Proceedings*, Cape Cod, MA, June 1989.

6. D. J. Heeger. Optical Flow Using Spatiotemporal Filters, International *Journal of Computer Vision*, 279-302 (1988).

7. Henry Holtzman. Three-Dimensional Representations of Video using Knowledge Based Estimation, Masters Thesis, Massachusetts Institute of Technology, September 1991.

8. B. Horn and B. Schunk. Determining Optical Flow, *Artificial Intelligence*, 17, 1981.

9. J. Jain and A. Jain. Displacement measurement and its application in interframe image coding, *IEEE Transactions on Communications*, COM-29, 1981.

10. J. Lim and J. Murphy. Estimating the velocity of moving images in television signals, *Computer Graphics and Image Processing*, 4, 1975.

11. Andrew Lippman. Feature Sets for Interactive Images, *Communications of the ACM*, April 1991.

12. Steve Mann. Compositing Multiple Pictures of the Same Scene, Proceedings IS&T Annual Meeting, May 1993.

13. Patrick McLean. Structured Video Coding, Masters Thesis, Massachusetts Institute of Technology, June 1991.

14. A. Shepp. Introduction to Images and Imaging, *Image Processes and Materials*, Van Nostrand Reinhold, 1989.

15. Laura Teodosio and Michael Millis. Panoramic Overviews for Navigating Real-World Scenes, Proceedings ACM Multimedia, August 1-6 1993.

16. Laura Teodosio. Salient Stills, Master's Thesis, Massachusetts Institute of Technology, September 1992.

17. John Watlington. Synthetic Movies, Master's Thesis, Massachusetts Institute of Technology, September 1989.

## ACKNOWLEDGEMENTS

This work was sponsored in part by International Business Machines, Inc. The authors would like to thank the Media Laboratory's Entertainment and Information Group, as well as Massimiliano Poletto, Edmond Chalom, John Wang, and Edward Adelson for their assistance.