Enhancing and Experiencing Spacetime Resolution with Videos and Stills

Ankit Gupta[1] Pravin Bhat[1] Mira Dontcheva[1] Michael Cohen[1,2] Brian Curless[1] Oliver Deussen[3]

[1]University of Washington [2]Microsoft Research, Redmond [3]University of Konstanz, Germany

Abstract

We present software solutions for enhancing spatial and/or temporal resolution of videos. Our algorithm targets the emerging consumer-level hybrid cameras that can capture video with intermittent high-resolution stills. Our technique can best be described as reconstructing the high-resolution spacetime videos using a few high resolution images for rendering and a low-resolution video to guide the reconstruction and the rendering. Our simple framework integrates two existing algorithms, namely a high-quality optical flow algorithm [Sand and Teller 2006] and a high-quality image-based-rendering algorithm [Bhat et al. 2007] to enable a variety of applications that were previously unavailable to the amateur photographer, such as the ability to automatically create videos with high spatiotemporal resolution, to easily retime the flow of a video, and shift a high-resolution still to nearby points in time to better capture a missed event.

1 Introduction

Still cameras are capturing increasingly high-resolution images. Video resolution, on the other hand has not increased at nearly the same rate. Unfortunately, capturing high-resolution images at a high frame rate is a difficult and expensive hardware problem. Video produces enormous amounts of data, which must be captured quickly using smaller exposure times and pushed onto the storage device. The problem of capturing videos with high spatiotemporal resolution is further compounded by the constant push in the consumer market for miniaturization and integration with other products (e.g., cell phones, PDAs). Hence, high spatial resolution imagery is often incompatible with high frame rate imagery, especially in the case of consumer level cameras. [1] In the face of these realities, we investigate software solutions that can increase the spatial and/or temporal resolution of imagery captured by a combination of low-resolution medium frame rate (15-30 fps) video and high-resolution stills captured at very low frame rates (1-5 fps).

A few commercially available cameras (e.g., Sony HDR-HC7, Canon HV10, and Canon MVX330), claim to allow people to capture videos while simultaneously capturing high-resolution stills; however, the number of stills that they capture during a video session is currently limited to three. In addition, from our research, we found either the high-resolution stills or the video lacking in the quality one would hope for with these cameras. Several researchers have proposed and even built prototypes of such cameras [Cohen and Szeliski 2006; Nagahara et al. 2005; Ben-Ezra and Nayar 2004] to begin to explore this space and we can expect rapid advances from manufacturers in this area.

Keeping such hybrid cameras in mind we propose a simple framework that integrates two existing algorithms, namely a high-quality optical flow algorithm [Sand and Teller 2006] and a high-quality image-based-rendering algorithm [Bhat et al. 2007] to enable a variety of applications that were previously unavailable to the amateur photographer, including:

- automatically producing high spatiotemporal resolution videos using low-resolution, medium-frame-rate videos and intermittent high-resolution stills,
- time shifting high-resolution stills to nearby points in time to better capture a missed event,
- retiming of high-resolution videos for dramatic emphasis or other artistic effects,
- and using retiming to explore the continuum between still images and videos with a new media type coined *Cliplets*.

We performed a user study to validate that users would indeed be interested in at least some of these applications. Our subjects were positive about the idea of a hybrid camera and using our technology to capture better images and videos.

The contributions of this paper include the integration of optical flow and (IBR) algorithms. We also describe a straightforward two step flow algorithm that improves previous work on optical flow in our setting. Finally, we identify a new set of consumer applications as noted above that can leverage this work.

2 Related work

The problem of enhancing low-resolution images and video has received a lot of interest in the past. In this section we review some of the previous approaches to this problem.

Spatial and temporal super-resolution using multiple low-resolution inputs One class of these techniques aligns multiple low-resolution images of a scene at sub-pixel accuracy to a reference view [Irani and Peleg 1991; Schultz and Stevenson 1996; Elad and Feuer 1997]. The aligned images can then be used to reconstruct the reference view at higher resolution. Another class of techniques perform temporal super-resolution by using sub-frame-rate shifts across multiple low temporal resolution videos of a scene [Shechtman et al. 2005; Wilburn et al. 2005]. (Schechtman et al additionally perform spatial super-resolution.) These techniques rely on the assumption that the low resolution inputs are undersampled and contain aliasing. However, most cameras usually bandlimit the high frequencies in the scene to minimize such aliasing, which severely limits the amount of resolution enhancement that can be achieved using these techniques. Lin et al. [2004] study and find that these reconstruction methods can enhance the spatial super-resolution by a factor of at most 1.6 in practical conditions. Also, the reliance of these methods on static scenes or multiple video cameras currently limits the practicality of these methods.

Temporal resolution enhancement For temporal resolution enhancement, most techniques [Castagno et al. 1996; Cafforio et al. Feb 1990; Wong and Goto 1995; Choi et al. 2000] compute motion correspondences and perform a weighted average of motion compensated frames assuming linear motion between correspondences. However, since the intermediate frame is generated as a weighted average of the two warped images, any errors in the correspondences can result in artifacts such as ghosting, blurring and flickering in the final result. We use a similar technique for generating the correspondences, but employ a spacetime, gradient domain compositing process to reduce these artifacts.

[1]High-end DSLR (Digital Single Lens Reflex) cameras capable of fairly high frame rates for short burst durations are beginning to appear (e.g., Canon EOS-1D Mark-III).

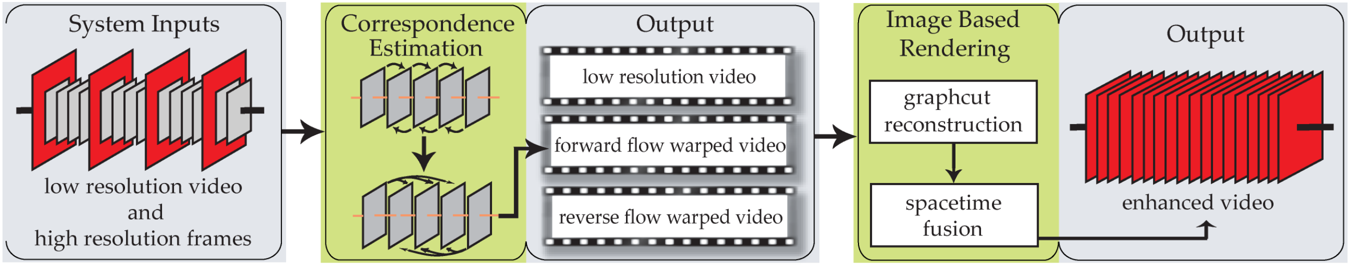

Figure 1: Our system consists of two main components. We use a two step optical flow process to compute correspondences between pixels in a low-resolution frame and nearby high-resolution stills. We employ the spacetime fusion algorithm of Bhat et al. to generate the final result.

Learning-based techniques These regression based techniques learn a mapping between a patch of low-resolution pixels and the corresponding high-resolution pixel at the center of the patch. The high resolution image is synthesized using the patch of low-resolution pixels surrounding every pixel in the input to infer the corresponding high-resolution pixel. Often, the synthesis also incorporates some smoothness prior to avoid artifacts that commonly arise from relying on regression alone. Freeman et al. [2000] and Baker et al. [2002] were among the first to take this approach for spatial resolution enhancement. Hertzmann et al. [2001] and Freeman et al. [2002] used nearest neighbor based regression to enhance the spatial resolution of images. Bishop et al. [2003] extended this work to videos using temporal regression in addition to spatial regression. Cheung et al. [2005] present a more general learning based framework in the form of video epitomes, which can be used for spatiotemporal resolution enhancement. In comparison to these, our method takes advantage of excellent exemplars (high resolution images taken nearby in space and time) and flow-based correspondence when copying in higher resolution detail. All of these methods generally result in seams, which we mitigate through spacetime gradient domain fusion.

Combining stills with videos Our technique can best be described as reconstructing the high-resolution spacetime video using a few high-resolution images for rendering and a low-resolution video to guide the reconstruction and the rendering. Bhat et al. [2007] proposed a similar approach to enhance low-resolution videos of a static scene by using multi-view stereo to compute correspondences between the low-resolution video and the high-resolution images. In contrast, our method uses optical flow to compute correspondences and can therefore handle dynamic scenes as well and allows us to enhance temporal resolution also.

The method proposed by Watanabe et al. [2006] works on similar input data as ours – a low-resolution video with high frame rate and a high-resolution video with low frame rate. Each high resolution frame is used to flow the high frequency information to the low-resolution frames using a DCT fusion step. This method does not flow the high frequencies in both directions for generating an intermediate frame which is generally necessary to compensate for imperfect motion correspondence. Nagahara et al. [2006] also take a similar approach but use feature tracking instead of motion compensation. These methods generate the frames separately and are prone to temporal incoherence artifacts. These techniques only look into increasing the temporal resolution upto the captured rate from the hybrid camera. We go further in this work to enhance resolution in both space and time by unconstrained factors.

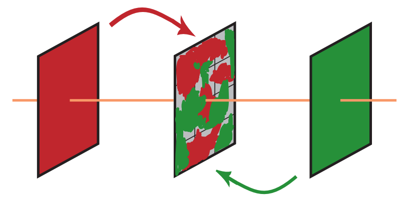

Figure 2: The high-resolution boundary frames are warped along flow lines and composited along with the upsampled low-resolution frame using a graph-cut optimization.

3 System overview

Figure 1 gives a visual description of our system for performing spatial and/or temporal resolution enhancement of videos using high-resolution stills when available.

3.1 Spatial resolution enhancement

Correspondences Using bicubic interpolation, the system upsamples the low-resolution frames to match their frame size to that of the high-resolution stills. Then, the system computes correspondences between every low-resolution frame and its two nearest high-resolution stills. Unfortunately, computing correspondences between temporally distant images of a dynamic scene is a hard problem. Most optical flow algorithms can compute correspondences for motion involving only tens of pixels. In our case the system needs to compute correspondences between a high-resolution still and a low resolution frame that might contain objects that have displaced over hundreds of pixels.

One approach is to compute compute optical flow between the high-resolution boundary stills to each of the interval frames. This approach, however, can produce errors because of the differences in image resolution. Alternatively, we could use the low-resolution boundary frames instead of the high-resolution ones. This improves matching, but, without good initialization, optical flow often fails to find good correspondences between frames that are far apart in time. We could also compute pairwise optical flow and sum the flow vectors to estimate correspondence between distant frames. This approach performs still better but flow errors tend to accumulate.

Our approach is to use a two step process. First we approximate the overall motion by summing the flow between consecutive frames, f_i and f_{i+1}. Then we use this sum as initialization for a second flow computation between the low-resolution boundary frames, s_{right} and s_{left}, and intermediate frames, $f_{1\ldots i}$. S_{right} and S_{left} are the corresponding high-resolution boundary frames for s_{right} and s_{left}. To actually compute optical flow we use the algorithm proposed by Sand et al. [2006].

Spatial reconstruction Once the system has computed correspondences from s_{left} to f_i and s_{right} to f_i, it warps S_{left} and S_{right} to bring them into alignment with f_i thus producing two warped images, w_{left} and w_{right}. Then it reconstructs f_i using regions from w_{left} and w_{right}. We compute the reconstruction with a multi-label graph-cut optimization with a metric energy function [Boykov et al. 2001]. Each pixel in f_i is given three candidate labels – w_{left}, w_{right}, and f_i as shown in Figure 2. We use the standard energy function used for graphcut compositing with a data cost that is specialized for our problem and the smoothness cost proposed by Kwatra et al. [2003]. Kwatra's smoothness cost encourages the reconstruction to use large coherent regions that transition seamlessly from one patch to another. Our data cost encourages the reconstruction to prefer labels that are likely to produce a high-resolution reconstruction while trying to avoid artifacts caused by errors in the correspondences.

The formal definition of our data cost function D for computing the cost of assigning a pixel p to a given label l is as follows:

$$D(p,l) = \begin{cases} c & \text{if } l = f_i \\ \infty & \text{if } w_l(p) \text{ undefined} \\ D_c(p,l) \cdot D_f(p,l) \cdot D_d(l,f_i) & \text{otherwise} \end{cases}$$

$$D_c(p,l) = ||w_l(p) - f_i(p)||$$
$$D_f(p,l) = 1 - \text{motion\_confidence}(w_l(p))$$
$$D_d(l,f_i) = \frac{|\text{frame\_index}(w_l) - i|}{n}$$

Here, c is the default cost for assigning a pixel to the low-resolution option (i.e, f_i); w_l is the warped image corresponding to the label l; D_c encourages color consistency between a pixel and its label; D_f factors in the confidence of the motion vector that was used to generate $w_l(p)$; and D_d favors labels that are closer in temporal distance to the current frame number i. The data cost components are normalized to 1 before taking the product. All examples in this paper were generated by setting c to 0.3. The confidence of the motion vectors is generated by Sand's optical flow algorithm [2006]in the process of computing the correspondences.

Spacetime fusion When each individual frame in the video $f_{1...n}$ is reconstructed using the graph-cut compositing step just described, the resulting video has high spatial resolution, but it suffers from the types of artifacts common to videos reconstructed using pixel regions – that is, the spatial and temporal seams between the regions tend to be visible in the result. These spatial seams can often be mitigated using the 2D gradient domain compositing technique described by Pérez et al. [2003]. However, the temporal seams that arise due to errors in the motion vectors and exposure/lighting differences in the high-resolution stills can be difficult to eliminate. We use spacetime fusion [Bhat et al. 2007], a 3D gradient domain compositing technique to eliminate both the spatial and temporal seams. Spacetime fusion tries to maintain the spatial gradients of the high-resolution reconstruction and the temporal gradients along the computed flow lines of the original input video. Thus, the temporal coherence seen in the low-resolution video is reproduced in the final high resolution result. We can assign relative weights to the spatial and temporal gradients. Using only spatial gradients leads to high spatial but poor temporal quality. Using only temporal gradients leads to too much blurring in the spatial domain. All of our results have been generated using a relative weight of 0.85 between spatial and temporal gradients, which we found experimentally. The reader is referred to the spacetime fusion paper [Bhat et al. 2007] for the low-level details.

3.2 Temporal resolution enhancement

To increase the temporal resolution of a video we insert the appropriate number of intermediate frames between existing frames. To create a frame between two existing frames the system assumes the motion between the corresponding pixels is linear. The system first computes forward and reverse optical flow between the two frames. Then the two frames are warped to the appropriate point in time and composited using a graph-cut optimization as described in Section 3.1 to construct the intermediate frame. Occlusions will cause holes in the reconstruction. Previously, we used the low-resolution frames, f_i, to fill these holes. Here, we force a complete labeling of the composite frame, but this causes artifacts. However, the spacetime fusion step helps to mitigate these artifacts. We generate temporal gradients for spacetime fusion by sampling the temporal gradients between the original frames.

3.3 Performance

The current processing speed of our system is quite slow with six minutes of processing for each video frame (resolution: 800 x 450); where five minutes are spent on the optical flow computation, and the last minute is spent on IBR. Since runtime performance was not our primary focus, our unoptimized research code can be improved. The optical flow computation time can be decreased by an order of magnitude using GPU acceleration and a multi-grid solver as shown by Bruhn et al. [2005]. Our implementation of the spacetime fusion algorithm, which uses a simple software based conjugate gradient solver, can be significantly faster with GPU acceleration and a preconditioner similar to the one proposed by Szeliski [2006].

4 Applications and results

Combining low-resolution video with high-resolution stills to produce a high-resolution spatiotemporal video will enable a number of applications previously unavailable to the amateur photographer.

4.1 Spatial resolution enhancement

We demonstrate our approach on a number of complex scenes that include non-rigid and fast motion, deformations, and changing lighting effects. Since hybrid cameras do not yet generate the type of quality we would like, we simulate their output by downsampling a high-resolution video and keeping high-resolution frames from the original video at even intervals. Figure 3 shows three examples. Please see the supplementary video for additional examples.

As discussed in Section 3.1 there are a number of techniques for computing the motion correspondences between the boundary and intermediate frames. We show a comparison between our two-step approach and the three alternatives in Figure 4. We also compare our graph-cut compositing step to a naive morphing composite. Our approach, although not perfect, has fewer artifacts than any of the others.

We also perform a quantitative analysis of our spatially enhanced results by measuring the overall peak signal-to-noise ratio (PSNR) of the result video with respect to the original high-resolution video (Figure 5). PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is widely used as a compression quality metric for images and videos. For this analysis we use the dolphin sequence shown in Figure 4. We explore PSNR variation with respect to the downsampling factor we use to simulate the hybrid input (Figure 5A) and the sampling of the high resolution stills (Figure 5B). Figure 5A shows that as the resolution of the input video

Figure 3: The left column shows the low-resolution input video frame. The right column shows our result. We suggest zooming in to see improvements at the actual scale of the images.

decreases the PSNR also decreases. This is not surprising as the resolution of the input will invariably affect the quality of the output. Figure 5B shows that as the sampling of the high-resolution stills increases, i.e., there are more high resolution stills, the PSNR also increases. The figures also show two additional methods. The morphing approach uses our motion correspondence method and then computes a weighted average of the boundary frames. We also compare to our intermediate results before the spacetime fusion step. Spacetime fusion has a considerable effect, as it guides the resulting video based on the temporal information in the initial low resolution video and thus removes many of the artifacts and noise present in the high-resolution composites.

4.2 Time shift imaging

Although advances in digital cameras have made it easier to capture aesthetic images, photographers still need to know when to press the button. Capturing the right instant is often elusive, especially when the subject is an exuberant child, a fast-action sporting event, or an unpredictable animal. Shutter delay only exacerbate the problem. As a result, users will often capture many images or use the motor drive setting of their cameras.

When the user "takes a picture" with a hybrid camera, the camera stores the video interval between the last and next high resolution still and also three high-resolution stills (two periodic captures and the one clicked by the user). Using our spatial enhancement approach, we can propagate the high-resolution information from the high-resolution stills to the surrounding low-resolution frames, thus producing a very short high-resolution video around the high-resolution still captured by the user. This high-resolution image collection enables users to choose a different frame as their still image than the one originally captured. We expect that this ability to shift a still backward or forward in time will make it easier

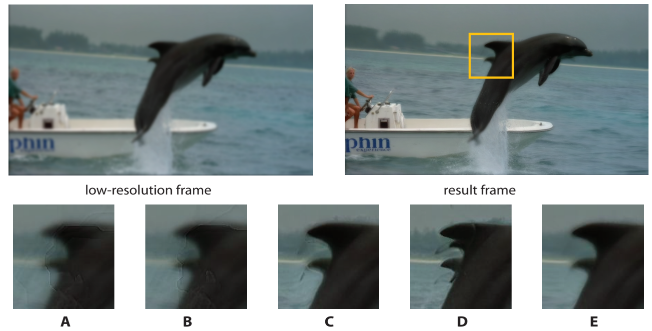

Figure 4: We compare our approach to existing alternatives. (A) Computing flow directly between the high-resolution boundary frames and low-resolution intermediate frames produces ghosting and blurring artifacts. (B) Computing flow between the lower resolution boundary frames and intermediate frames still results in similar artifacts due to long range motion. (C) Summed up pairwise optical flow to estimate the motion between distant frames still has artifacts such as motion trails. (D) Morphing the boundary frames to composite the intermediate frames results in tearing artifacts. (E) Our approach produces a video that is visually consistent and has relatively few artifacts. We suggest zooming in to see improvement at the actual scale of the images. Note that (A)-(C) and (E) use our graph-cut spacetime fusion system.

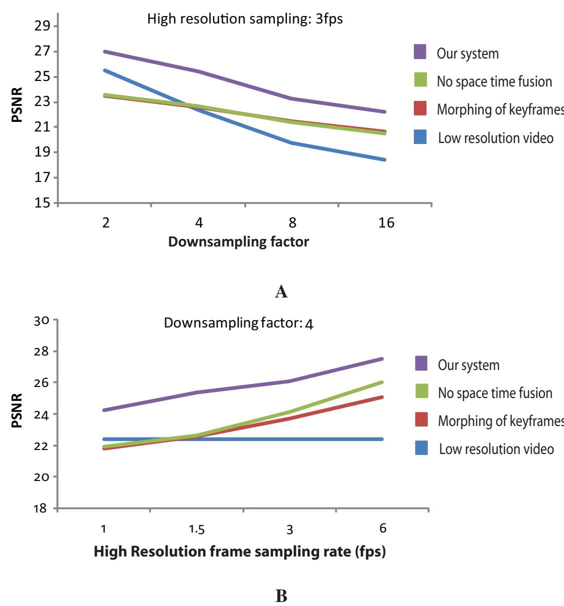

Figure 5: Our approach produces better images as compared to other approaches using peak signal-to-noise ratio (PSNR).

to capture that "blowing out the candles" moment, or that "perfect jumpshot." Figure 6 shows a result for this application.

To confirm whether users would be interested in shifting their images forward or backward in time, we performed a user study. We recruited 17 students from our university for a 30 minute user study on new digital camera design. The study had three parts. The subjects first completed a survey on their current image and video capture practices. Then, they were presented with a software application that simulates a camera interface. The application plays videos in the camera's display. The user then "takes pictures" by pressing a button, just as they would if they were holding a real camera. The participants were shown 22 video sequences and asked to take representative pictures of the scene. They were instructed to take pictures as they do typically. In the last phase of the experiment the subjects were asked to go through and "edit" the photographs they

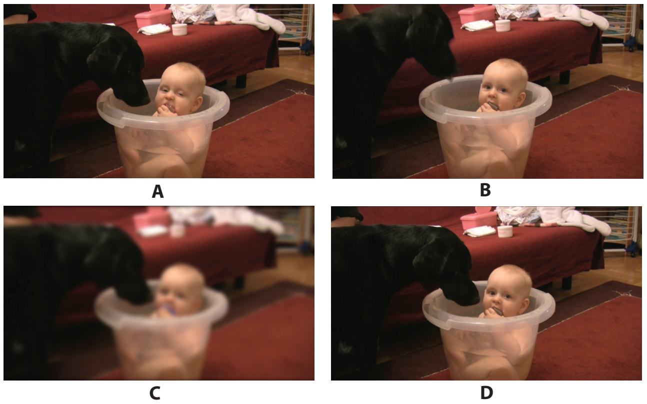

Figure 6: Images (A) and (B) show the high-resolution frames captured by the hybrid camera. The eyes of the baby are closed in (A), and the dog's snout is far away in (B). A better photograph may be (C), where the eyes are open and the snout is close to the face. Our system generates a high spatial resolution frame for (C) as shown in (D) by flowing and compositing high-resolution information from (A) and (B).

had captured. To edit the photographs the subjects were allowed to replace the captured shot with a frame from a buffer of 10 frames to either side of the one they originally chose. Their stated goal was to select their favorite frame in the sequence around their originally captured image. The subjects did not know how they were going to "edit" their photographs until they got to the third phase. Finally, we concluded the study with a discussion about this new camera model.

The overall user feedback was positive. Fifteen of the seventeen participants were eager to use a hybrid camera and our framework for enhancing images and video. One subject said, "If the adjacent frames are hi-res, I would totally love this. Sometimes I was just a few frames off. Other times I was way off. If you can time shift, that's awesome!" Our original hypothesis was that users would more often select frames preceding the one originally captured due to the delay between seeing what they wanted and clicking the "shutter". However, Figure 7 shows a roughly normal distribution of finally selected frames around the original one. Perhaps users over-anticipate the right instant as much as they "miss it."

4.3 Temporal resolution enhancement

To enhance the temporal resolution of a video an algorithm would ideally compute perfect motion correspondences and then simply morph the frames to generate the right result. Unfortunately, computing perfect correspondences for most real world videos is hard. Imperfect motion correspondences result from hardware constraints (e.g., inability to capture fast motion at an appropriate frequency) and the complexity of natural scenes (e.g., textureless regions and occlusions). Existing techniques [Vatolin and Grishin 2006] use weighted averaging of warped images to hide artifacts resulting from bad correspondences. In comparison, our method (i.e., graph-cut compositing plus spacetime fusion) focuses on preserving the overall shape of objects and hence leads to stroboscopic motion in regions where the motion correspondences are extremely bad. Therefore, the artifacts of our technique are different from those of previous techniques. Our technique may be preferred for enhancing videos that involve objects like human faces where distortion and tearing artifacts would be jarring. On the other hand, the distortion and tearing artifacts of previous methods look like motion blur for certain objects, which in turn makes their results look temporally smoother for some videos. Like most morphing-based methods, our method is unable to get rid of motion blur present in the input for fast moving objects, which can seem perceptually odd when seen in slow motion.

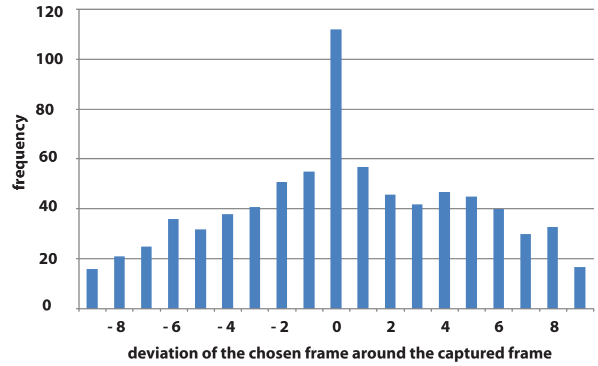

Figure 7: This graph shows users' preferences for the frames around the frame they actually capture in the user study.

We combine the spatial and temporal steps of our approach to produce spatiotemporally enhanced videos. We show examples of this type of enhancement in the supplementary video.

4.4 Video retiming and cliplets

Dynamic retiming is an effective method for enhancing the suspense or meaning in a scene. With our framework, we can create retimed videos simply by varying the temporal resolution. To enable users to create their own retimed videos, we created a simple curve editor that allows users to specify the video retiming and preview an approximate result in real time using simple frame averaging. Once the user is happy with the flow of the retimed video, our system generates the high quality retimed video.

Cohen and Szeliski [2006] proposed the idea of cliplets as a media between stills and videos to enhance the perceptual experience of a moment. Cliplets are described by a still frame, a very short segment of motion, and another still frame. Cliplets are simply a special case of retiming, and users can generate them with our curve interface.

5 Conclusions

We have demonstrated the power of combining optical flow with graph-cut-based optimization and spacetime fusion to achieve a number of enhancements to the spatial and temporal resolution of captured imagery. We hope that our work will encourage camera manufacturers to provide more choices of hybrid video and still cameras. The combination of good image capture and appropriate software can provide amateur photographers and videographers a wide variety of new applications. Currently the capabilities of our framework depend on the quality of motion correspondence. As motion correspondence algorithms improve, we will be able to apply our framework to a broader set of scenarios, such as videos with large and fast motion. We envision using our framework to generate high spatiotemporal videos from motor-drive imagery. Additionally, capability to generate high spacetime resolution videos from hybrid input could possibly be used to improve video compression systems.

References

Baker, S., and Kanade, T. 2002. Limits on super-resolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell. 24*, 9, 1167–1183.

BEN-EZRA, M., AND NAYAR, S. K. 2004. Motion-based motion deblurring. *IEEE Trans. Pattern Anal. Mach. Intell. 26*, 6, 689–698.

BHAT, P., ZITNICK, C. L., SNAVELY, N., AGARWALA, A., AGRAWALA, M., CURLESS, B., COHEN, M., AND KANG, S. B. 2007. Using photographs to enhance videos of a static scene. In *Rendering Techniques 2007 (Proc. of Eurographics Symposium on Rendering)*, Eurographics, 327–338.

BISHOP, C., BLAKE, A., AND MARTHI, B. 2003. Super-resolution enhancement of video. *Proc. of the 9th Conference on Artificial Intelligence and Statistics (AISTATS)*.

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell. 23*, 11, 1222–1239.

BRUHN, A., WEICKERT, J., FEDDERN, C., KOHLBERGER, T., AND SCHNORR, C. 2005. Variational optical flow computation in real time. 608–615.

CAFFORIO, C., ROCCA, F., AND TUBARO, S. Feb 1990. Motion compensated image interpolation. *Communications, IEEE Trans. on 38*, 2, 215–222.

CASTAGNO, R., HAAVISTO, P., AND RAMPONI, G. 1996. A method for motion-adaptive frame rate up-conversion. *IEEE Trans. Circuits and Systems for Video Technology 6*, 5 (October), 436–446.

CHEUNG, V., FREY, B. J., AND JOJIC, N. 2005. Video epitomes. In *CVPR '05: Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, IEEE Computer Society, Washington, DC, USA, 42–49.

CHOI, B., LEE, S., AND KO, S. 2000. New frame rate up-conversion using bi-directional motion estimation. *IEEE Trans. on Consumer Electronics 46*, 3, 603–9.

COHEN, M. F., AND SZELISKI, R. 2006. The moment camera. *Computer 39*, 8, 40–45.

ELAD, M., AND FEUER, A. 1997. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. on Image Processing 6*, 12 (Dec.), 1646–1658.

FREEMAN, W. T., PASZTOR, E. C., AND CARMICHAEL, O. T. 2000. Learning low-level vision. *International Journal of Computer Vision 40*, 1, 25–47.

FREEMAN, W. T., JONES, T. R., AND PASZTOR, E. C. 2002. Example-based super-resolution. *IEEE Comput. Graph. Appl. 22*, 2, 56–65.

HERTZMANN, A., JACOBS, C. E., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. In *ACM Trans. on Graphics (SIGGRAPH 2001)*, ACM, New York, NY, USA, 327–340.

IRANI, M., AND PELEG, S. 1991. Improving resolution by image registration. *CVGIP: Graph. Models Image Process. 53*, 3, 231–239.

KWATRA, V., SCHŐDL, A., ESSA, I., TURK, G., AND BOBICK, A. 2003. Graphcut textures: Image and video synthesis using graph cuts. *ACM Trans. on Graphics, SIGGRAPH 2003 22*, 3 (July), 277–286.

LIN, Z., AND SHUM, H.-Y. 2004. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *Pattern Analysis and Machine Intelligence, IEEE Trans. on 26*, 1, 83–97.

NAGAHARA, H., HOSHIKAWA, A., SHIGEMOTO, T., IWAI, Y., YACHIDA, M., AND TANAKA, H. 2005. Dual-sensor camera for acquiring image sequences with different spatio-temporal resolution. 450–455.

NAGAHARA, H., MATSUNOBU, T., IWAI, Y., YACHIDA, M., AND SUZUKI, T. 2006. High-resolution video generation using morphing. In *ICPR '06: Proc. of the 18th International Conference on Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, 338–341.

PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. In *ACM Trans. on Graphics (SIGGRAPH 2003)*, ACM Press, New York, NY, USA, 313–318.

SAND, P., AND TELLER, S. 2006. Particle video: Long-range motion estimation using point trajectories. In *CVPR '06: Proc. of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, 2195–2202.

SCHULTZ, R., AND STEVENSON, R. 1996. Extraction of high-resolution frames from video sequences. *IP 5*, 6 (June), 996–1011.

SHECHTMAN, E., CASPI, Y., AND IRANI, M. 2005. Space-time super-resolution. *IEEE Trans. on Pattern Analysis and Machine Intelligence 27*, 4, 531–545.

SZELISKI, R. 2006. Locally adapted hierarchical basis preconditioning. In *ACM Trans. on Graphics (SIGGRAPH 2006)*, ACM, New York, NY, USA, 1135–1143.

VATOLIN, D., AND GRISHIN, S. 2006. N-times video frame-rate up-conversion algorithm based on pixel motion compensation with occlusions processing. Graphicon: International Conference on Computer Graphics and Vision, conference, 112–119.

WATANABE, K., IWAI, Y., NAGAHARA, H., YACHIDA, M., AND SUZUKI, T. 2006. Video synthesis with high spatio-temporal resolution using motion compensation and spectral fusion. *IEICE - Trans. Inf. Syst. E89-D*, 7, 2186–2196.

WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E.-V., ANTUNEZ, E., BARTH, A., ADAMS, A., HOROWITZ, M., AND LEVOY, M. 2005. High performance imaging using large camera arrays. *ACM Trans. Graph. 24*, 3, 765–776.

WONG, W. F., AND GOTO, E. 1995. Fast evaluation of the elementary functions in single precision. *IEEE Trans. on Computers 44*, 3, 453–457.