Video Enhancement Using Reference Photographs

Cosmin Ancuti Tom Haber Tom Mertens Philippe Bekaert
Hasselt University, firstname.lastname@uhasselt.be

(a) original (b) bicubic interpolation (c) using reference photo

Figure 1: *Enhancement of resolution using our technique. The input is a small 320x240 video frame (a). Naive bicubic upsampling (b), is not able to recover high frequency information. Our technique borrows details from high resolution reference photographs (not shown), and copies this information to the low resolution frame, yielding a detailed 1024x768 frame (c). A single photograph was used for this example. Please see the supplemental for the full videos.*

Introduction Handheld digital video cameras have become increasingly popular and cheaper in recent years. Even still cameras offer additional functionality for shooting videos. Unfortunately, the small form factor of these devices limits the light sensitivity, and often the the lens and sensor do not allow for satisfactory image quality. Matters become worse as each frame can be exposed for only a fraction of a second. Even for such short exposures, motion blur is still noticeable and may destroy visual details. In addition, the internal bandwidth of the storage unit inside the camera also puts a limit on resolution. Image quality may also suffer from the ability of the person that operates the camera. The dynamic range of current cameras is not high enough to correct over- and under-exposure afterward, and excessive shake increases motion blur. On the other hand, when taking photographs with a still camera, we are much less subject to these issues. We wish to process a video into a more aesthetically pleasing version, by borrowing information from high quality reference photographs of the same scene. Since the process of taking a photograph is not time-critical, we can afford a longer exposure for reducing noise, and record more information to increase resolution. Also, photographs are less prone to motion blur if a tripod is used.

Our Approach In this sketch, we introduce a new technique that transfers image data from a high resolution photograph $\mathcal{H}$ (or multiple) to a video frame $\mathcal{L}$ in a patch-wise fashion. We basically copy small windows of image data, guided by a set of pair wise correspondences.

Constrained texture synthesis [Hertzmann et al. 2001] could be used to synthesize a high resolution image from $\mathcal{H}$, guided by $\mathcal{L}$. However, differences in scale and rotation makes direct copying impossible. The reference image $\mathcal{H}$ could be extended with scaled and rotated copies of itself, in order to account for such transformations [Kwatra et al. 2003]. However, it is intractable to search for every possible variations. Instead of searching and enumerating every possible transformation, we use a rotation and scale invariant feature descriptor.

Video-based superresolution methods compound detail over several frames, but cannot go far beyond the original resolution. In contrast, our technique is mostly limited by the resolution of the reference photograph. Our approach is also closely related to example-based superresolution [Hertzmann et al. 2001; Freeman et al. 2002], but we use a scene-specific database of examples. Also, our technique accounts for changes in scale and rotation.

Our technique relies on pair-wise correspondences between $\mathcal{H}$ and $\mathcal{L}$. We start by detecting salient feature points in the images [Lowe 2004], and complete the set with uniformly distributed points in uncovered regions. Correspondences are established by comparing image descriptors in a local window around each feature point. We opted for the SIFT descriptor [Lowe 2004], for its scale and rotation invariance property. We improve SIFT's distinctiveness by complementing it with the mean and variance of each color channels over the local window. For each established pair, we estimate the change in scale and rotation directly from the SIFT descriptor, and warp each window accordingly before copying it into the frame $\mathcal{L}$. Finally, we account for fine translation differences between the patches by computing an optimal offset with normalized SSD (sum of squared differences).

Our technique is particularly useful to increase resolution. See figure Figure 1 for a result. Please see the supplemental material for the complete videos. Aside from superresolution, our technique may also be used to reduce noise, motion blur and exposure issues. In the future, we would like to improve results by exploiting temporal coherence, and experiment with adaptive windows.

References

FREEMAN, W. T., JONES, T. R., AND PASZTOR, E. C. 2002. Example-based super-resolution. *IEEE Computer Graphics and Applications 22*, 2 (March/April), 56–65.

HERTZMANN, A., JACOBS, C. E., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. In *SIGGRAPH 2001, Computer Graphics Proceedings*, ACM Press / ACM SIGGRAPH, E. Fiume, Ed., 327–340.

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

LOWE, D. 2004. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, vol. 20, 91–110.