MPEG-1 Super-Resolution Decoding for the Analysis of Video Still Images

Kyle J. Erickson
University of Notre Dame
kerickso@watt.lisa.nd.edu

Richard R. Schultz
University of North Dakota
Richard_Schultz@mail.und.nodak.edu

Abstract

A digital image sequence coded at a low bitrate using a motion-compensated video compression standard should contain little data redundancy. However, the success of a particular super-resolution enhancement algorithm is predicated on subpixel-resolution overlap (i.e., redundancy) of moving objects from frame-to-frame. If an MPEG-1 bitstream is coded at a relatively high bitrate (e.g., a compression ratio of 15:1), enough data redundancy exists within the bitstream to successfully perform super-resolution enhancement within the decoder. Empirical results are presented, in which decoded pictures from MPEG-1 bitstreams containing both global scene transformations and independent object motion are integrated to generate Bayesian high-resolution video still (HRVS) images. It is shown that additional spatial details can be extracted by integrating several motion-compensated coded pictures, provided that a large number of subpixel-resolution overlaps – such as those captured by a reconnaissance airplane or surveillance satellite – are present among the original digitized video frames.

1. Introduction

As a preprocessing step to the analysis of individual video frames, super-resolution enhancement [5] can be used to extract additional spatial detail from an image sequence through the temporal integration of multiple frames. Due to the intra-coded macroblocks contained within an MPEG-1 image sequence and the overlap of coded, motion-compensated macroblocks between frames, it is possible to perform super-resolution enhancement on compressed video by temporally integrating a short subsequence of decoded pictures [1,6]. The intra-coded I-picture is a logical candidate for enhancement, since it contains the most unique information among the I-, P-, and B-picture types. For an MPEG-1 image sequence coded in the B-B-I-B-B-P-B-B-P display order, the I-picture may be enhanced using a maximum of six integrable decoded frames (B-B-I-B-B-P), if only pictures possessing encoded half-pel motion vectors estimated with respect to the I-picture are considered. P- and B-pictures can also be enhanced using neighboring frames, but different pictures may be integrated. A maximum of seven decoded frames (I-B-B-P-B-B-P) can be integrated to enhance the first P-picture following an I-picture, while a maximum of three decoded frames (I-B-X-P or I-X-B-P) are integrable in the super-resolution enhancement of a B-picture. To calculate a Bayesian high-resolution video still (HRVS) estimate [5] with an up-sampling factor of $\uparrow n$ pixels, $1/(\uparrow n)$-th pel resolution motion vectors are required for all frames with respect to the picture undergoing enhancement [4,5]. Adequate up-sampled displacements can be estimated efficiently by utilizing the half-pel MPEG-1 encoded motion vectors as initial conditions to a reduced-search block matching algorithm [1]. Improvements in visual quality can be quite impressive, provided that global scene transformations such as those generated by a camera pan exist within the sequence.

2. MPEG-1 Super-Resolution Decoder

A digital image sequence coded at a low bitrate using a motion-compensated video compression standard should contain little data redundancy. However, the success of a particular super-resolution enhancement algorithm is predicated on subpixel-resolution overlap (*i.e.*, redundancy) of moving objects from frame-to-frame. If an MPEG-1 bitstream is coded at a relatively high bitrate (*e.g.*, a compression ratio of 15:1), enough data redundancy exists within the bitstream to successfully perform super-resolution enhancement within the decoder.

Embedding the Bayesian HRVS algorithm into an MPEG-1 decoder allows for the super-resolution enhancement of any frame within a compressed image sequence, although many empirical limitations have been discovered which affect the quality of the video stills. As a broad generalization, any intra-coded macroblocks that appear in a frame are capable of improving the visual appearance of the high-resolution video still far more than nonintra-coded macroblocks. This is intuitively obvious, since macroblocks that are only motion-compensated do not contribute additional information to the bitstream. Experimentally, it has been found that integrating more than approximately four pictures from an arbitrary MPEG-1 sequence does not improve the quality of a super-resolution enhanced video still. A law of diminishing returns takes effect; *i.e.*, as additional pictures are integrated, more and more motion estimation errors leak into the enhanced frame, further degrading the image. Mosquito noise seems to be quite prevalent in the super-resolution enhancement of compressed video, apparently due to the integration of macroblocks compensated using inaccurate motion vectors estimated in the MPEG-1 encoder, as well as the integration of similar information. For example, enhancing an I-picture through integration with its following P-picture usually results in a low-quality video still, since the P-picture consists primarily of forward-compensated macroblocks from the I-picture, and thus similar pixels between the two frames must be combined. The integration of I- and B-pictures seems to provide higher-quality video stills than I-P integrations, most likely since the coded compensation differences in B-pictures are more significant than those present in P-pictures.

Since the MPEG encoder uses block matching to compute motion vectors [2,3], the half-pel displacements estimated between the I-, P-, and B-pictures for a particular sequence can be obtained from the MPEG decoder directly. These MPEG half-pel motion vectors serve as an initial condition to a reduced-search block matching algorithm that calculates the $1/(\uparrow n)$-th pel motion vectors required for super-resolution enhancement [1]. To estimate the displacements, the reference decoded I-picture and a second integrable frame are first up-sampled by a factor of $\uparrow n$ using cubic B-spline interpolation. Reduced-search block matching is then applied to these up-sampled pictures, in which a relatively small search area is centered on an MPEG vector up-sampled by a factor of $\uparrow n/2$. This reduces the computation time required to estimate the subpixel motion vectors by up to a factor of 30, when compared to estimating the displacements using full-search block matching [1]. Moreover, using reduced-search block matching seems to result in little apparent degradation in the quality of the super-resolution enhanced frame.

3. Simulations and Generalizations

MPEG-1 super-resolution decoding results are shown in Figures 1-3, for the *Software*, *EELab*, and *Flower Garden* MPEG-1 bitstreams. In each case, the intra-coded frame was enhanced by a factor of $\uparrow 4$. Both *Software* and *EELab* were captured by a panning camera; thus, they represent image sequences containing global scene transformations. *Flower Garden* is a common test sequence containing segmented motion. Experimental results indicate that integrating the I-picture with only its following P-picture produces a blurred, noisy estimate. Integrating the I-picture with its following P-picture and either one or two of the intervening B-pictures, however, generates an estimate with improved definition at the cost of additional mosquito noise artifacts. In general, it was found that I-picture enhancement improves with each additional B-picture and deteriorates with each additional P-picture. Furthermore, MPEG-1 bitstreams coded with compression ratios greater than approximately 40:1 contain very little data redundancy, and thus the super-resolution enhancement algorithm becomes ineffective. For higher bitrates (*e.g.*, compression ratios less than 30:1), however, it appears that enough data redundancy exists within an MPEG-1 bitstream to perform super-resolution enhancement.

4. Conclusion and Future Directions

An MPEG-1 super-resolution decoder was designed to enhance digital video encoded using motion-compensated compression standards. Provided that a particular MPEG-1 bitstream is coded at a relatively high bitrate, enough data redundancy exists within the sequence to allow for the successful super-resolution enhancement of an I-picture within the decoder.

Future research will involve the attenuation of mosquito noise artifacts through the adaptive adjustment of the enhancement algorithm based on the coding classification of each macroblock, as well as tuning the encoder for super-resolution decoding. By understanding the errors that can occur in the super-resolution enhancement of compressed video, it should be possible to tune the MPEG-1 encoder to generate a video bitstream tailored specifically for high-quality super-resolution enhancement within the decoder.

References

[1] D. Chen and R. R. Schultz, “Extraction of High-Resolution Frames from MPEG Image Sequences.” In *Proc. ICIP-98* (on CD-ROM).

[2] B. Furht, J. Greenberg, and R. Westwater, *Motion Estimation Algorithms for Video Compression*. Norwell, MA: Kluwer Academic Publishers, 1997.

[3] J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, and D. J. LeGall, *MPEG Video Compression Standard*. New York, NY: Chapman & Hall, 1996.

[4] R. R. Schultz, L. Meng, and R. L. Stevenson, “Subpixel Motion Estimation for Super-Resolution Image Sequence Enhancement,” *Journal of Visual Communication and Image Representation*, vol. 9, no. 1, pp. 38-50, 1998.

[5] R. R. Schultz and R. L. Stevenson, “Extraction of High-Resolution Frames from Video Sequences,” *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 996-1011, 1996.

[6] Y. Yang, M. Choi, and N. P. Galatsanos, “New Results on Multichannel Regularized Recovery of Compressed Video.” In *Proc. ICIP-98* (on CD-ROM).

Figure 1: *Software* MPEG-1 image sequence, compression ratio 12:1. *Top*: Decoded I-picture – no enhancement, zero-order hold up-sampling factor of ↑4. *Bottom*: Super-resolution enhanced decoding of the I-picture, through the integration of six frames (B-B-I-B-B-P) with an up-sampling factor of ↑4. The original image sequence was acquired by panning a Sony Handycam in a diagonal trajectory across the stationary scene. Note the enhanced detail in the second picture, as well as the presence of some mosquito noise artifacts near object edges.

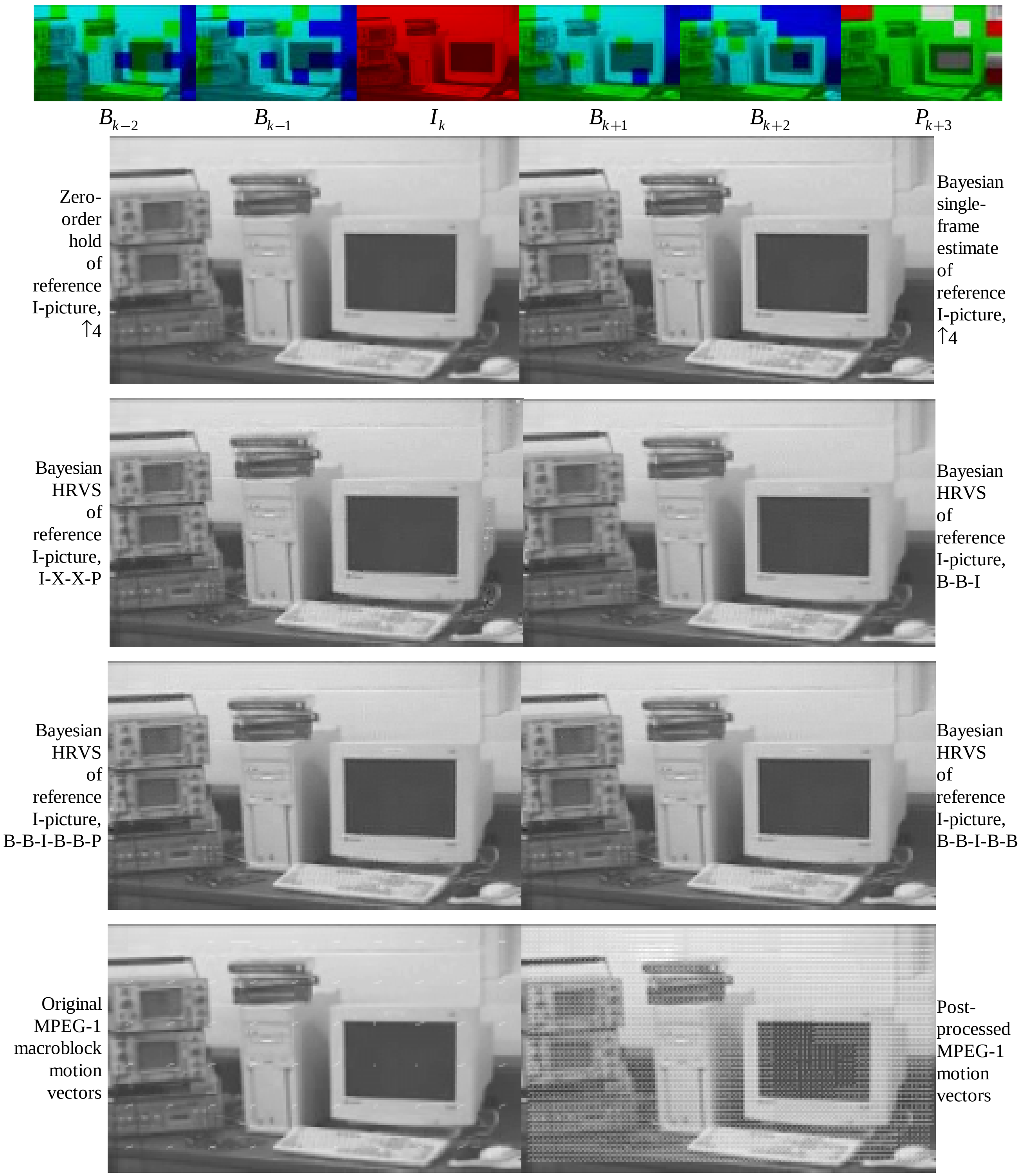

Figure 2: *EELab* MPEG-1 image sequence, compression ratio 20:1. The six uppermost images represent the decomposition of image sequence pictures by coded macroblock type: black – backward-compensated; dark gray – intra; medium gray – forward-compensated; light gray – both forward- and backward-compensated; and white – no compensation. Note the additional mosquito noise and reduced detail in the estimates that integrate P-pictures.

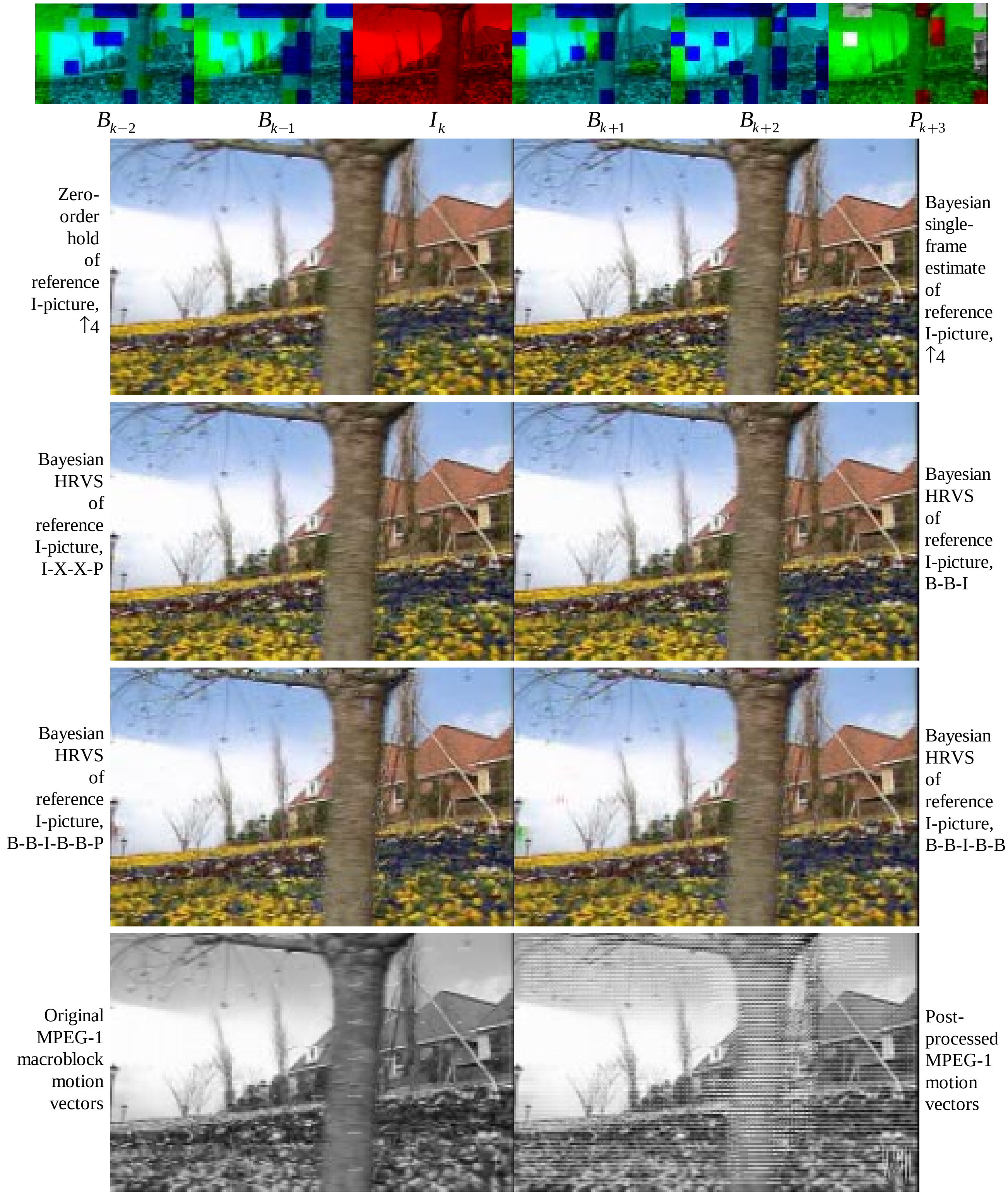

Figure 3: *Flower Garden* MPEG-1 image sequence, compression ratio 20:1. The Bayesian single-frame estimate represents an effective upper-bound on calculable single-frame estimates, and generally contains less detail than HRVS estimates calculated using three or more frames. Note the additional mosquito noise and reduced detail in the I-X-X-P and B-B-I-B-B-P images compared with the B-B-I and B-B-I-B-B images, respectively. Also note the comparatively small amount of extra information found in the B-B-I-B-B image versus the B-B-I image.