

Resolution Enhancement of Color Video

Brian C. Tom and Aggelos K. Katsaggelos

Northwestern University

Department of Electrical Engineering and Computer Science

McCormick School of Engineering and Applied Science

Evanston, IL 60208-3118 USA

Tel: (847) 491-7164

Fax: (847) 491-4455

Email: brian@eecs.nwu.edu, aggk@eecs.nwu.edu

ABSTRACT

In this paper, an approach to improve the spatial resolution of color video is presented. Such high resolution images are desired, for example, in video printing. Previous work has shown that the most important step in achieving high quality results is the accuracy of the motion field. It is well known that motion estimation is an ill-posed problem. However, in processing color video, additional information contained in the color channels may be used to improve the accuracy of the motion field over the motion field obtained with the use of only one channel. In turn, this improvement in the motion field will be shown through several experimental results to significantly improve the estimation of a high resolution image sequence from a corresponding observed low resolution sequence.

1 Introduction

Improved resolution of a still image from a monochrome image sequence has been demonstrated in [8, 9]. In both of these approaches, the frames surrounding the current working frame are first motion compensated to the current frame. In [8] a non-recursive approach using *maximum a posteriori* (MAP) estimation is presented, using a Gibbs prior. In [9], a technique based on [6] is used to improve the resolution of a video sequence. However, work in color video processing has been significantly more limited. In this paper, the extension of [9] to color video resolution enhancement will be discussed. From [7], it is recognized that the motion estimation of color sequences is more accurate than performing motion estimation on each channel alone. This is because of the additional information each channel provides, analogous to multi-channel image restoration, where the cross-correlation terms can be exploited to improve the restored image [4, 10]. Therefore, improved results can be obtained by formulating the algorithm of [9] in a multi-channel fashion.

Towards this end, the resolution of the current (working) frame is increased by incorporating the spatial information contained in the adjacent frames. These adjacent frames need to first be motion compensated to the

current frame. In the case of perfect motion compensation, each frame in the sequence can be considered to be identical to the working frame. This is the assumption made in [6], where multiple distorted versions of the same original image are available.

This paper is organized as follows. In Sec. 2, notation is introduced and the iterative algorithm is presented. Section 3 discusses the importance of the motion estimator, and presents two approaches for motion estimation using all three of the color channels. Experimental results are presented and discussed in Sec. 4, and conclusions are drawn in Sec. 5.

2 Problem Formulation

Let $x_{i,k}$ denote the i th channel of a lexicographically ordered high resolution k th frame of a sequence, and $y_{i,k}$ its ordered low resolution (observed) counterpart, where for color sequences $i = 1, 2, 3$. Typical color schemes used are either RGB or YUV. Following [8], [9], the low resolution frames are obtained from their high resolution counterparts by assuming that the images were integrated and subsampled. Mathematically, this yields

$$y_{i,k} = A^{(k,k)} x_{i,k}, \quad i = 1, 2, 3, \quad (1)$$

where $A^{(k,k)}$ denotes the subsampling and integration matrix. If each low resolution and high resolution frame is $N \times N$ and $PN \times PN$, respectively, where P is an integer determining the degree of subsampling, then $A^{(k,k)}$ is a matrix of size $N^2 \times (PN)^2$. The transpose of A , denoted by A^T , is equivalent to a zero-order interpolation matrix.

In [9], an iterative equation for the estimation of the high resolution frame for monochrome sequences is given, based on the set theoretic approach presented in [6]. In this paper, this equation is extended to include contributions from all three color channels, and is given by

$$\begin{aligned} x_{i,k}^{p+1} &= x_{i,k}^p + b \sum_{j=0}^{M-1} \beta_j C^{(k-j,k)} A^T r_{i,j,k} \\ &- \lambda Q^T Q x_{i,k}^p, \end{aligned} \quad (2)$$

where $r_{i,j,k}$ is the residual error term given by

$$r_{i,j,k} = y_{i,k-j} - A C^{(k,k-j)} x_{i,k}^p, \quad (3)$$

and $C^{(k,k-j)}$ is the motion compensation operator defined by

$$\begin{aligned} x_{i,k-j}(\mathbf{r}) &= C^{(k,k-j)} x_{i,k}(\mathbf{r}) \\ &= x_{i,k}(\mathbf{r} + \mathbf{d}_*^{(k,k-j)}), \quad i = 1, 2, 3, \end{aligned} \quad (4)$$

where $\mathbf{d}_*^{(k,k-j)}$ is the displacement vector field (DVF) between frames $k-j$ and k and \mathbf{r} denotes the spatial coordinates. The subscript $*$ on \mathbf{d} represents that motion field which is chosen in an optimal sense for multi-channel motion estimation. The choice of this \mathbf{d}_* will be discussed in more detail in Sec. 3.

In Eq. (2), M is an integer specifying the number of frames to process, with $M \geq P^2$, λ is the regularization parameter, the superscript p denotes the p th iteration, Q is a high pass filter, β_j is the relative weighting assigned to frame y_{k-j} , and b is a parameter chosen to ensure convergence of the algorithm.

In [6], the contribution of each observed image was inversely proportional to the amount of noise present in that particular image. In Eq. (2), β_i should be inversely proportional to the displaced frame difference (DFD), given by

$$\varepsilon = \left| x_{i,k}(\mathbf{r}) - x_{i,k-j}(\mathbf{r} + \mathbf{d}_*^{k,k-j}) \right|. \quad (5)$$

In analyzing Eq. (2), it can be seen that at each iteration, a residual component is added to the previous estimate. This residual has contributions from each high resolution image that has been estimated and compensated. Each of these contributions is weighted inversely proportionally to the DFD, which means that those frames with not very accurate motion estimation and compensation will be weighted less. The residual component has dimensions of the low resolution image, which is then upsampled by A^T and added to $x_{i,k}^p$. Besides upsampling, the low resolution residual needs to be compensated back to frame k as well, as required by $C^{(k-j,k)}$. The motion estimation can be performed by a variety of approaches. While previous work has used the pel-recursive approach of [3], in this case, more sophisticated methods are available to include the additional cues found in the color channels. Finally, the convergence of the algorithm has been established experimentally, although the approaches for choosing the regularization parameter discussed in [5] is currently under consideration.

3 Motion Estimation of Color Image Sequences

As presented earlier, the most important step in estimating good quality high resolution sequences is that of motion estimation [9]. In monochrome sequences, improvement in resolution is somewhat limited by the fact

that the motion estimates are estimated from bilinearly interpolated frames, which introduce artifacts and propagates the bilinear interpolation errors around the edges (i.e., smoothness). On the other hand, color image sequences offer the benefit of having more information *a priori*, and this leads to improved accuracy of the field (measured in terms of mean square error and smoothness of the field) [7]. The basic observation is that for any color image sequence, the motion between adjacent frames for each channel is exactly the same. In other words, there is only one correct motion field which describes the motion from one frame to the next.

In practice, however, when motion estimation is performed on each channel independently (following a block matching or a pel-recursive approach), the motion fields will tend to differ among the channels. The degree of (dis)similarity among these motion fields depends upon both the complexity of the image sequence (in terms of motion), as well as, the motion estimator itself. In any event, it is desired and advantageous to have only one motion field representing the motion for all three channels.

Two approaches have been investigated in estimating such a motion field. The first one is to estimate the motion field for each channel independently, and then fuse the three fields into one. The median and the mean of the vectors have been considered as a first step. Vectors can be ordered according to the approach given in [2], where it was shown that the median of a sequence of values is the Maximum Likelihood (ML) estimate of a bi-exponential probability density function, assuming that the data originated from such a source. In a similar manner, the vector mean is the ML estimate for a Gaussian distribution. The vector median of a series of vectors, $\vec{x}_1, \dots, \vec{x}_N$, is \vec{x}_{vm} such that

$$\vec{x}_{vm} \in \{ \vec{x}_i \mid i = 1, \dots, N \}, \quad (6)$$

and for all $j = 1, \dots, N$,

$$\sum_{i=1}^N \|\vec{x}_{vm} - \vec{x}_i\|_2 \leq \sum_{i=1}^N \|\vec{x}_j - \vec{x}_i\|_2. \quad (7)$$

The second approach is to choose the motion field by optimizing an error term that is a weighted function of the DFD's of all three channels, or

$$\mathbf{d}_* = \arg \min \{ F (w_1 \varepsilon^{red} + w_2 \varepsilon^{green} + w_3 \varepsilon^{blue}) \}, \quad (8)$$

where F indicates some function, ε for each channel is given by Eq. (5), and $w_i, i = 1, 2, 3$ denote the weights of each channel. This approach is expected to yield better results, in terms of motion compensation error, than the first approach which fuses the results of the three independent motion estimators.

4 Experimental Results

The above algorithm was tested on the RGB *flower garden* sequence, whose size was cropped to 360×240 pixels.

The frames from the early part of the sequence was chosen, and subsampled to 240×120 for all three channels. These subsampled frames were then bilinearly interpolated back up to 360×240 for each channel, and this was used as the initial starting frames for the algorithm. The second frame of the original sequence is shown in Fig. 1, while the subsampled frame is shown in Fig. 2, and the bilinear interpolated image (of Fig. 2) is shown in Fig. 3. It should be mentioned that all images are not displayed in their true size, but have been scaled by 50% in order to fit this paper.

In the first experiment, the median vector field was chosen using Eqs. (6) and (7), and a high resolution version of the second frame was estimated. The reconstructed image after 10 iterations is shown in Fig. 4. The stopping criterion used was the step size difference was less than 2.0×10^{-5} . Note that the general 'blurriness' of the bilinear interpolated image has been removed. Interestingly, there is additional edge enhancement in the foreground (flowers) over the original image as well. The trade-off with such enhancement is additional noise, which is seen to be enhanced as well.

The next experiment optimized the motion field by minimizing Eq. (8), with all of the weights being equal to 1.0. The resulting high resolution image is shown in Fig. 5. By comparing Fig. 5 with Fig. 4, it can be seen that the images are very similar. In fact, in terms of PSNR, they differ only by 0.1 dB (Fig. 5 has the higher PSNR).

Finally, a comparison was made with those images reconstructed from a single channel approach. Fig. 6 shows the second frame of the red channel that was reconstructed by the algorithm in [9]. In terms of mean square error, the values were 1522.16 and 1615.25 for the multi-channel and single-channel approaches, respectively. This in turn translated to a difference of 0.13 dB PSNR.

5 Conclusions

In this paper, a method to improve the definition of a video sequence has been proposed. The proposed algorithm is iterative and utilizes adjacent frames which have been motion compensated to the current working frame. A particular zero-order hold matrix is used to model the subsampling process, and the motion is estimated between bilinear interpolated versions of these observed frames.

In addition, the inclusion of color cues also benefits the reconstruction process over that obtained from independent, single-channel approaches. This is similar to the advantages gained in performing multi-channel image restoration over independent, single-channel image restoration. As reported in earlier work, the critical factor in the reconstruction process is the accuracy of the motion estimator. Two approaches for color motion estimation have been presented here, in the context of im-

age resolution enhancement. Temporally recursive resolution enhancement techniques are also currently under investigation.

References

- [1] K. Aizawa, T. Komatsu, and T. Saito, "A Scheme for Acquiring Very High Resolution Images Using Multiple Cameras," *IEEE Proc. ICASSP-92*, San Francisco, CA, vol. III, pp. 289-292, 1992.
- [2] J. Astola, P. Haavisto, Y. Neuvo, "Vector Median Filters," *Proceedings of the IEEE*, vol. 78, no. 4, pp. 678-689, April 1990.
- [3] J. C. Brailean and A. K. Katsaggelos, "A Recursive Nonstationary MAP Displacement Vector Field Estimation Algorithm," *IEEE Trans. Image Processing*, vol. 4, no. 4, pp. 416-429, April 1995.
- [4] N. P. Galatsanos and R. T. Chin, "Digital Restoration of Multichannel Images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, No. 3, pp. 415-421, March 1989.
- [5] M. G. Kang and A. K. Katsaggelos, "General Choice of the Regularization Functional in Regularized Image Restoration," *IEEE Trans. Image Proc.*, vol. 4, no. 5, pp. 594-602, May 1995.
- [6] A. K. Katsaggelos, "A Multiple Input Image Restoration Approach," *J. Visual Commun. Image Represent.*, vol. 1, pp. 93-103, Sept. 1990.
- [7] J. Konrad and E. Dubois, "Use of Colour Information in Bayesian Estimation of 2-D Motion," *IEEE Proc. ICASSP-90*, Albuquerque, NM, vol. 4, pp. 2205-2208, April 1990.
- [8] R. R. Schultz and R. L. Stevenson, "Improved Definition Video Frame Enhancement," *IEEE Proc. ICASSP-95*, Detroit, MI, vol. 4, pp. 2169-2171, May 1995.
- [9] B. C. Tom, and A. K. Katsaggelos, "An Iterative Algorithm for Improving the Resolution of Video Sequences," *Proc. 1996 SPIE Visual and Communications Image Processing* Orlando, FL, pp. 1430-1438, March 1996.
- [10] B. C. Tom, K. T. Lay, and A. K. Katsaggelos, "Multi-Channel Image Identification and Restoration Using the Expectation-Maximization Algorithm," *Optical Engineering*, vol. 35, no. 1, pp. 241-254, Jan. 1996.
- [11] B. C. Tom, A. K. Katsaggelos, and N. P. Galatsanos, "Reconstruction of a High Resolution from Registration and Restoration of Low Resolution Images," *IEEE Proc. International Conference on Image Processing*, Austin, TX, vol. 3, pp. 553-557, Nov. 1994.



Figure 1: Original frame #2, red channel



Figure 2: Subsampled frame #2, red channel



Figure 3: Figure #2, bilinear interpolated



Figure 4: Reconstructed high resolution image, frame #2, red channel, median



Figure 5: Reconstructed high resolution image, frame #2, red channel



Figure 6: Reconstructed high resolution image, single channel method