HIGH RESOLUTION STANDARDS CONVERSION OF LOW RESOLUTION VIDEO

Andrew J. Patti, M. Ibrahim Sezan† and A. Murat Tekalp

Department of Electrical Engineering and Center for Electronic Imaging Systems
University of Rochester, Rochester, NY 14627
†Electronic Imaging Research Labs, Eastman Kodak Company, Rochester, NY, 14650-1816

ABSTRACT

With the advent of frame grabbers capable of acquiring multiple video frames, a great deal of attention is being directed at creating high-resolution (hi-res) imagery from interlaced or low-resolution (low-res) video. This is a multi-faceted problem, which generally necessitates standards conversion and hi-res reconstruction. Standards conversion is the problem of converting from one spatio-temporal sampling lattice to another, while hi-res image reconstruction involves increasing the spatial sampling density. Also of interest is removing degradations that occur during the image acquisition process. These tasks have all received considerable, yet separate, treatment in the literature. Here, a unifying video formation model is presented which addresses these problems simultaneously. Then, a POCS-based algorithm for generating high-resolution imagery from video is delineated. Results with real imagery are included.

1. INTRODUCTION

Video standards refer to the format used to store, transmit, and display video signals. A standard can be described by a particular spatio-temporal sampling lattice [1]. Since various video systems ranging from High-Definition television (HDTV) to videophone have different spatial and temporal resolution requirements, there are a variety of standards in use today. The task of converting from one of these standards to another is referred to as standards conversion. Two examples are deinterlacing and frame rate conversion.

High-resolution (hi-res) standards conversion involves simultaneously solving the problems of standards conversion, and what has previously been referred to as the hi-res reconstruction problem. The hi-res reconstruction problem refers to reconstructing a good quality still image from a sequence of low-res images that suffer from one or more of the following degradations:

- aliasing (due to undersampling which may be over an arbitrary spatio-temporal lattice),
- sensor blur (due to sensor integration and relative scene-sensor motion during a finite aperture time),
- focus blur (due to defocused lenses), and
- noise (sensor and quantization noise).

This work is supported in part by a National Science Foundation IUCRC grant and a New York State Science and Technology Foundation grant to the Center for Electronic Imaging Systems at the University of Rochester, and a grant by Eastman Kodak Company.

Hi-res reconstruction aims at obtaining a still image sampled over a denser sampling grid that is free from the effects of these degradations [2, 3, 4, 5].

Our goal in this paper is to propose an algorithm that will take an input low-res video signal sampled on an arbitrary spatio-temporal lattice, which is effected by the previously listed degradations, and simultaneously solve the standards conversion and hi-res reconstruction problems. Note that the resulting hi-res progressive images can be subsampled on another lattice, if higher resolution video on a lattice other than progressive (e.g. interlaced) is desired.

Solutions to this problem are useful in many applications. One is converting from NTSC interlaced video to interlaced video for HDTV. Another application is the creation of a synthetic "video zoom". Here, a region of the video display is enlarged by some factor and played. Printing hi-res stills from video is also an important application. In this case it is often desirable to upsample a given low-res image while increasing the detail. Because video signals are commonly interlaced, the processes of deinterlacing and removal of acquisition degradations must be combined to produce the desired image.

There are a variety of methods used to solve the hi-res problem when progressive low-res video is available. The proposed solutions have 3 basic components: (i) motion compensation, (ii) interpolation, and (iii) blur and noise removal. Motion compensation is used to map the pixels from the available low-res frames to a common hi-res grid. The motion vectors can be computed on a pixel by pixel basis by using a technique such as block matching [4], or motion models such as pure translations, rotations, affine transformations [2], and perspective transformations [6] can be used. The second component, interpolation, refers to combining the pixels that have been mapped back from the low-res grid to the hi-res grid, to produce a hi-res image sampled on a rectangular grid. The third component, blur and noise removal, is needed when the low-res frames have not been acquired using an ideal sensor. In this paper, we are interested in algorithms that implement components (ii) and (iii) simultaneously.

A frequency domain formulation for simultaneous interpolation and removal of blur and noise, has been proposed by Kim and Su [3]. Their method, however, is restricted to pure translational motion between low-res frames. An affine motion model is used by Irani and Peleg [2], and an iterative method similar to the Landweber iteration is used to form the hi-res image. The Landweber iteration is also used by Komatsu et. al. [4]. Rather than using a motion model, however, they interpolate the low-res images to the

0-7803-2431-5/95 $4.00

hi-res size, and use block matching to compute a vector for every pixel location. Mann [6] uses the same iterative scheme as Irani and Peleg, except that a perspective motion model is used. These methods all take into account the degradations caused by the physical dimensions of the low-res sensor elements and focal blur, but do not address the effect of a non-zero aperture time. In a previous paper, [5], we take this into account and use the method of projections onto convex sets (POCS) in solving the hi-res problem. There, motion is modeled for the case of pixel-independent translations following an arbitrary path, with small rotation.

In this paper we provide an extension of the work carried out in [5], where the problem of still frame hi-res reconstruction is addressed, to the case where the input low-res video signal is sampled on an arbitrary space-time lattice. We model the following problem: video is generated using a given low-res camera or cameras, of some continuous scene. During the low-res imaging process we account for the following: (i) movement of the low-res camera, (ii) movement or changes in the contents of the scene, (iii) a non-zero sensor aperture time, (iv) non-zero physical dimensions for each individual sensor element (i.e. giving rise to a particular image pixel), (v) blurring caused by the imaging optics, (vi) sensor noise, (vii) sampling of the continuous scene on an arbitrary space-time lattice. We refer to this model as the video formation model. The modeling is then used in conjunction with the method of Projections onto Convex Sets (POCS) for reconstructing a hi-res version of this low-res video.

2. A UNIFYING MODEL

In this section we present a model that is used to unify the problems of standards conversion and hi-res image reconstruction. The video formation model is first presented. Then motion trajectories are applied to the model, and lastly, a discrete model is given.

2.1. The video formation model

The video formation model we propose is depicted in Fig. 1. In the figure, the input signal $f(x_1, x_2, t)$ denotes the continuous video signal in the focal plane coordinate system (x_1, x_2). At each point within the (x_1, x_2) system, the aperture time of the sensor is modeled by an integrating function with offset in time of t_0. The output of the integrator is given by

$$g_1(x_1, x_2, t) = \int_{t+t_0}^{t+t_0+T_a} f(x_1, x_2, \tau) d\tau. \quad (1)$$

The effects of the physical dimensions of the low-res sensor, and the out-of-focus blur of the optical system are modeled in the second stage of the figure. The input to this stage, $g_1(x_1, x_2, t)$, is convolved with both the kernel representing the shape of the sensor, $h_a(x_1, x_2, t)$, and the kernel representing the focal blur, $h_o(x_1, x_2, t)$. These are both functions of time, but we restrict them to be constant over the aperture time. The focus blur and aperture dimensions are thus allowed to differ from frame to frame.

The third stage in Fig. 1 models sampling with the arbitrary space-time lattice Λ_s. The output of this stage is $g_3(m_1, m_2, k)$. As a matter of convention, integer values that appear as a function argument are interpreted as in

$$g_3(m_1, m_2, k) = g_3(x_1\ x_2\ t)|_{[x_1\ x_2\ t]^t = V_s[m_1\ m_2\ k]^t}, \quad (2)$$

where V_s denotes the matrix that specifies the sampling lattice, and ^{t} denotes the transpose operation. In the last modeling step, additive noise due to the low-res sensor is added to the sampled video signal.

2.2. The motion model

The utility of the model is realized when object motion within $f(x_1, x_2, t)$ is taken into account. By using the concept of a motion trajectory, we can express the result of the first modeling stage in the form

$$g_1(x_1, x_2, t) = \quad (3)$$
$$\int\int f(u_1, u_2, t_r) h_1(u_1, u_2; \vec{c}(t_r; x_1, x_2, t)) du_1 du_2,$$

where $h_1(u_1, u_2; \vec{c}(t_r; x_1, x_2, t))$ is the LSV blur function kernel, and $\vec{c}$ denotes the motion path (defined below). We briefly summarize the derivation of the LSV kernel here.

The motion trajectory of an object within the continuous video signal is represented by the path through the continuous sensor space (x_1, x_2),

$$\vec{c}(t_r; x_1, x_2, t) = (c_1(t_r; x_1, x_2, t), c_2(t_r; x_1, x_2, t)) \quad (4)$$

This path specifies the location of the intensity value $f(x_1, x_2, t)$, at time t_r. The motion trajectory of a particular intensity value is often referred to as the optical flow. The motion paths will have a temporal beginning and end, as objects enter and leave the scene. When objects cross in front of others, occlusions and uncovered background will result, which can also begin or terminate a path. This concept has been fully developed by Dubois in [1]. Using the path $\vec{c}$, the optical flow within the continuous video signal can now be expressed as

$$f(x_1, x_2, t) = f(\vec{c}(t_r; x_1, x_2, t), t_r). \quad (5)$$

By substituting equation (5) into the video formation model and carrying out the resulting line integrals, we arrive at the desired form for the LSV PSF in (3). With this result, the remainder of the image formation model can be worked out in a straight-forward manner.

2.3. The discrete model

As in [5], the proposed POCS method for hi-res reconstruction, requires a discretized version of $f(u_1, u_2, t_r)$. Thus, a discrete superposition summation of the form

$$g(m_1, m_2, k) = \sum_{(n_1, n_2)} f(n_1, n_2, t_r) h_{t_r}(n_1, n_2; m_1, m_2, k), \quad (6)$$

will be formulated, where it is assumed that the continuous image $f(n_1, n_2, t_r)$ is sampled on the 2-D lattice Λ_{t_r} (i.e. (n_1, n_2) are integers that specify a point in Λ_{t_r}). By appropriately choosing t_r and Λ_{t_r}, sampling of $f(n_1, n_2, t_r)$ can be formed over an arbitrary space-time lattice.

An individual hi-res sensor element (giving rise to a single hi-res image pixel) is assumed to have physical dimensions which can be used as a unit cell $\mathcal{U}_{t_r}$ for the lattice Λ_{t_r}.

Thus, the entire space of the focal plane is completely covered by the hi-res sensor. The term $U_{t_r}(n_1, n_2)$ is used to denote the unit cell U_{t_r} shifted to the location specified by (n_1, n_2). With this definition, and with the assumption that $f(u_1, u_2, t_r)$ is approximately constant over $U_{t_r}(n_1, n_2)$, the results from the video formation model can be posed in the discrete formulation of (6).

It is interesting to note that even when the aperture time is zero, if the input lattice is not progressive, the blur function will be LSV. In this case, the solution method must be capable of processing LSV blurs. Both the Landweber iteration and POCS methods have this property. The POCS solution delineated in the next section, however, has a mechanism for adapting to the properties of the additive noise, whereas the Landweber iteration does not.

3. THE POCS SOLUTION

As in [5], we propose a POCS based solution to the hi-res reconstruction problem. The method of POCS requires the definition of closed convex constraint sets, within a well-defined vector space, that contain the actual hi-res image. An estimate of the hi-res image is then defined as a point in the intersection of these constraint sets, and is determined by successively projecting an arbitrary initial estimate onto the constraint sets.

Associated with each constraint set is a projection operator, P, mapping an arbitrary point within the space to the closest point within the set. Relaxed projection operators, $T \doteq (1-\lambda)I + \lambda P;\ 0 < \lambda < 2$, can also be defined and used in finding an estimate in the intersection set.

We define the following closed, convex constraint set, for each pixel within low-res image sequence $g(m_1, m_2, k)$:

$$C_{t_r}(m_1, m_2, k) = \{y(n_1, n_2, t_r) : |r^{(y)}(m_1, m_2, k)| \leq \delta_0\}, \quad (7)$$

$$r^{(y)}(m_1, m_2, k) \doteq \quad (8)$$

$$g(m_1, m_2, k) - \sum_{(n_1, n_2)} y(n_1, n_2, t_r) h_{t_r}(n_1, n_2; m_1, m_2, k),$$

is the residual associated with an arbitrary member, y, of the constraint set. Note that sets $C_{t_r}(m_1, m_2, k)$ can be defined only where the motion information is valid. The quantity δ_0 is an *a priori* bound reflecting the statistical confidence with which the actual image is a member of the set $C_{t_r}(m_1, m_2, k)$. Since $r^{(f)}(m_1, m_2, k) = v(m_1, m_2, k)$, where f denotes the actual hi-res image, the statistics of $r^{(f)}(m_1, m_2, k)$ are identical to those of $v(m_1, m_2, k)$. Hence the bound δ_0 is determined from the statistics of the noise process so that the actual image (i.e., the ideal solution) is a member of the set within a certain statistical confidence.

The projection $\mathbf{P}_{t_r}(m_1, m_2, k)[x(n_1, n_2, t_r)]$ of an arbitrary $x(n_1, n_2, t_r)$ onto $C_{t_r}(m_1, m_2, k)$ is defined similarly to that in [5]. The difference in this case is that this formulation only defines sets at the locations of the arbitrary lattice, as opposed to over a progressive lattice. Additional constraints such as bounded energy, positivity, and limited support can be utilized to im prove the results. Here, we also use the amplitude constraint set, C_A (having T_A).

Given the above projections, an estimate $\hat{f}(n_1, n_2, t_r)$ of the hi-res image $f(n_1, n_2, t_r)$ is obtained iteratively from all low-res images $g(m_1, m_2, k)$ where constraint sets can be defined, as

$$f_{k+1}(n_1, n_2, t_r) = \mathrm{T}_A \tilde{\mathrm{T}}[f_k(n_1, n_2, t_r)] \quad (9)$$

where $\tilde{\mathbf{T}}$ denotes the composition of the relaxed projection operators, projecting onto the family of sets $C_{t_r}(m_1, m_2, k)$, and bilinearly interpolated low-res images can be used as the initial estimate $\hat{f}_0(n_1, n_2, t_r)$.

4. RESULTS

We conduct two experiments. The first demonstrates the application of the algorithm to real low-res images that are not sampled on a progressive lattice. In this case, we verify the efficacy of the latter modeling stages under the condition of zero-aperture time. Real data from a digital camera is used in this experiment. The second experiment is a simulation, which demonstrates convergence when the aperture time is non-zero, and the low-res sampling lattice is interlaced.

In the first experiment, 6 pictures are taken using a digital camera. The camera uses a color filter array (CFA) that samples the green signal component using a diamond lattice. We assume that the focal plane of the camera is completely covered by the CFA array, and all elements have equal size and uniform response. We process the green channel. A conversion from the diamond lattice to progressive, and further upsampling of the progressive grid by a factor of two, is carried out simultaneously using the proposed algorithm. The focal blur was assumed to be Gaussian, with a variance of 1, and a 5x5 support (units are relative to the hi-res spatial sampling period). Motion was computed between the the low-res pictures using hierarchical block matching [7] with quarter pixel accuracy (relative to a progressive low-res grid). As previously mentioned, the aperture time was assumed to be zero. The results of this experiment are shown in Fig. 2. At the top in the figure is the hi-res image estimated using bilinear interpolation, and at the bottom is the POCS result.

In the second experiment, we simulate the formation of an interlaced low-res sequence. Five low-res fields are generated using velocities $v_{1_i} = v_{2_i}$ of 4.25, 4.25, 4.5, 5 and 6, in units of pixels per high-res sample spacing, from the Nuke image. The aperture time is set to 0.5, and a square low-res sensor geometry is used, where a side of the low-res sensor is 2 times the length of a side of the hi-res sensor. For this configuration, only $\frac{1}{2}$ the area of the low-res sensor focal plane is sampled for each field. White Gaussian noise with a 30dB SNR is added to each low-res image, and we use the actual motion information. In practice, global motion between low-res images can be estimated fairly accurately at sub-pixel resolution using block-matching or phase-correlation techniques. Simulation results are shown in Fig. 3. At the top in the figure is the hi-res image estimated using bilinear interpolation, and at the bottom is the POCS result.

5. CONCLUSION

We have proposed a method for modeling video sampled on an arbitrary lattice, that takes into account motion blurring, focus blurring, and additive noise. We have applied the method of POCS in the context of this model to simultaneously solve the problems of standards conversion and high resolution image reconstruction. The effectiveness of

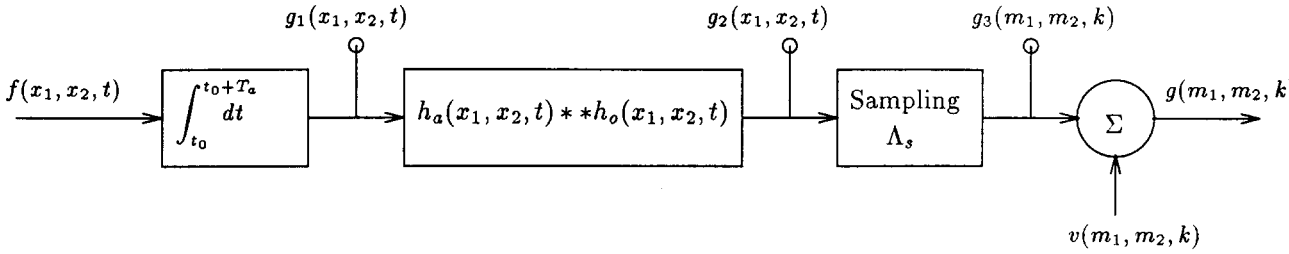

Figure 1: The video formation model is presented.

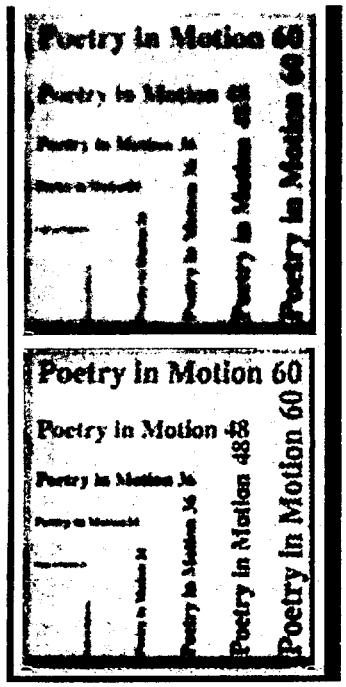

Figure 2: Results from the digital camera experiment. At the top is the hi-res estimate using bilinear interpolation, and at the bottom is the result using the proposed algorithm

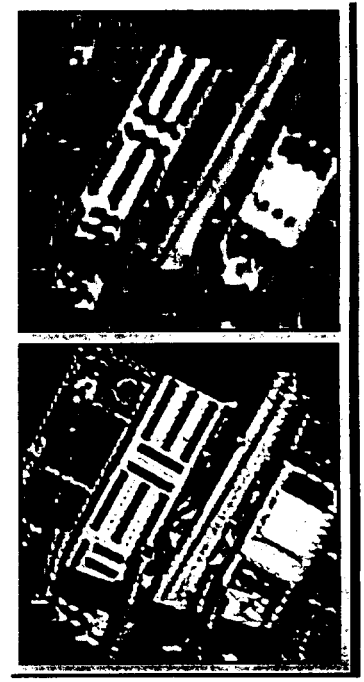

Figure 3: Results from interlaced video simulation. At the top is the hi-res estimate using bilinear interpolation, and at the bottom is the result using the proposed algorithm.

the proposed algorithm has been demonstrated by applying it to real data, and conducting a simulation.

6. REFERENCES

[1] E. Dubois, "Motion-compensated filtering of time-varying images," *Multidimensional Systems and Signal Processing*, vol. 3, pp. 211–239, 1992.

[2] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. of Visual Comm. and Image Representation*, vol. 4, pp. 324–335, December 1993.

[3] S. P. Kim and W.-Y. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Processing*, vol. 2, pp. 534–539, October 1993.

[4] T. Komatsu, T. Igarashi, K. Aizawa, and T. Saito, "Very high resolution imaging sheme with multiple different-aperature cameras," *Signal Proc.: Image Comm.*, vol. 5, pp. 511–526, December 1993.

[5] A. Patti, M. Sezan, and A. Tekalp, "High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur," in *IEEE Int. Conf. Image Proc.*, (Austin, Texas, USA), November 13-16 1994.

[6] S. Mann and R. W. Picard, "Virtual bellows: constructing high quality stills from video," in *IEEE Int. Conf. Image Proc.*, (Austin, Texas, USA), November 13-16 1994.

[7] M. Bierling and R. Thoma, "Motion-compensated field interpolation using a hierarchically structured displacement estimator," *Signal Processing*, vol. 11, pp. 387–407, 1986.