Super-resolution reconstruction using spatio-temporal filtering

N. Goldberg,[a] A. Feuer,[a,*] and G.C. Goodwin[b]

[a] *Department of EE, Technion—Israel Institute of Technology, Haifa, Israel*
[b] *CIDAC, Department of EE, University of Newcastle, Newcastle, Australia*

Abstract

It has been known for some time that temporal dependence (motion) plays a key role in the super-resolution (SR) reconstruction of a single frame (or sequence of frames). While the impact of global time-invariant translations is relatively well known, the general motion case has not been studied in detail. In this paper, we discuss SR reconstruction for both motion models from a frequency-domain point of view. A noniterative algorithm for SR reconstruction is presented using spatio-temporal filtering. The concepts of motion-compensated windows and sinc interpolation kernels are utilized, resulting in a finite impulse response (FIR) filter realization. In the simulations, we assume a priori knowledge of the motion (optical flow), which is commonly done throughout much of the SR reconstruction literature. The proposed process is localized in nature, and this enables the selective reconstruction of desired parts of a particular frame or sequence of frames.
© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Super-resolution; Aliasing

1. Introduction

The problem of super-resolution reconstruction in image processing has been of much interest among researchers and practitioners for some time (see e.g., Bose et al., 1993; Hardie et al., 1997; Huang and Tsai, 1984; Irani and Peleg, 1993; Kim et al., 1990; Patti et al., 1994, 1995, 1997; Schultz and Stevenson, 1994; Tekalp,

---

* Corresponding author.

1995; Tekalp et al., 1992 and others). Recently, an in depth survey and discussion on this problem was published by Baker and Kanade, 2002.

Roughly speaking, this problem can be stated as follows:

Given a number of low resolution frames of the same scene, construct a single frame of improved resolution.

This can be generalized to a sequence of frames as well (Elad and Feuer, 1999a,b) such as a video clip.

In most published results the problem is formulated as an estimation problem and using different optimality criteria and different a priori information, an assortment of algorithms are suggested. In a number of cases this resulted in algorithms which were computationally infeasible and simplifications were proposed to ease the computational load while attempting to minimize the performance degradation (see Elad and Feuer, 1997, 1999a; Peleg et al., 1987). The problem as stated, can be viewed as an inverse problem and, in the tradition to many such problems, when the transformation to be inverted is singular or ill conditioned, various regularizations are proposed. As a result, a solution is always generated. However, the question of whether one generates a truly higher-resolution image which contains more *true* information or just a better looking image, has hardly been addressed. The answer to this question can be traced to the question of aliasing in the sampling process associated with the data image generation. Our focus here is on this aspect only of the data image generation process.

Basically, the assumption behind the statement of the problem is that the sequence of data frames contains information which is not present in a single data frame. In other words, when each sampled frame is viewed individually, aliasing has occurred. While, when viewing the whole sequence no aliasing has occurred (or, alternatively, the degree of aliasing is significantly smaller).

To formalize this we introduce the following notation. Let $I(x, y, t)$ denote the intensity at point (x, y) on the image plane at time t. We restrict our discussion here to monochromatic images. This $2D + T$ function (we use this terminology to distinguish between $2D$, the spatial dimensions and T, the time dimension) will be referred to as "image" while, when viewed at a fixed time, it will be referred to as "frame."

The dependence on time can result from either motion or temporal changes in the data generation process (or both). We discuss here only motion related time dependence. In fact, we assume that the perceived motion, or, as commonly referred to in the literature, the *optical flow*, has been estimated by some preprocessing of the data. Namely, using the *optical flow constraint* equation

$$\frac{\partial I(x, y, t)}{\partial x} \frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial I(x, y, t)}{\partial y} \frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial I(x, y, t)}{\partial t} = 0, \tag{1}$$

where

$$\frac{\mathrm{d}x}{\mathrm{d}t} = V_x(x, y, t); \quad \frac{\mathrm{d}y}{\mathrm{d}t} = V_y(x, y, t) \tag{2}$$

are the known optical flow components.

The data available are given by

$$I_d(x, y, t) = I(x, y, t)s(x, y, t), \tag{3}$$

where

$$s(x, y, t) = \sum_l \sum_m \sum_n \delta(x - l\Delta x, y - m\Delta y, t - n\Delta t) \tag{4}$$

defines the sampling lattice (separable rectangular in our case), Δx, Δy, and Δt are the sampling intervals in the respective directions.

Since the aliasing phenomenon is clearly observed in the frequency domain, much of our discussion will be in that domain. Let us first restate the relationship between the image $I(x, y, t)$ and any of its frames, say $I_0(x, y) = I(x, y, 0)$. With $\widehat{I}(\omega_x, \omega_y, \omega_t)$ and $\widehat{I}_0(\omega_x, \omega_y)$ the Fourier Transform (FT) of $I(x, y, t)$ and $I_0(x, y)$, respectively, we have

$$\widehat{I}_0(\omega_x, \omega_y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{I}(\omega_x, \omega_y, \eta) \mathrm{d}\eta. \tag{5}$$

Namely, the support of $\widehat{I}_0(\omega_x, \omega_y)$ is the projection of the support of $\widehat{I}(\omega_x, \omega_y, \omega_t)$ on the (ω_x, ω_y) plane. Hence, one can readily imagine a situation where, when sampled at $(\Delta x, \Delta y, \Delta t)$, $\widehat{I}(\omega_x, \omega_y, \omega_t)$ has no aliasing, while, the single frame, $\widehat{I}_0(\omega_x, \omega_y)$, when sampled at $(\Delta x, \Delta y)$ has aliasing. This is demonstrated in Fig. 1.

In Section 2 we discuss the case where the motion in the image is spatio-temporal invariant. This motivates the discussion in Section 3 dealing with general motion and leading to the super-resolution (SR) reconstruction process we propose. The process is demonstrated in Section 4 and some concluding comments are provided in Section 5.
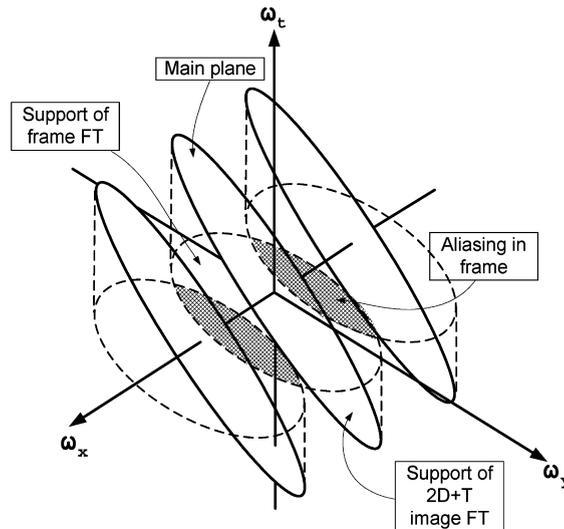

Fig. 1. Supports of samples $2D + T$ image FT and their relations to the supports of a sampled single frame FT.

2. Spatio-temporal invariant motion

Let us consider first the case where the motion is spatio-temporal invariant. Namely,

$$V_x(x, y, t) = V_x, \quad V_y(x, y, t) = V_y. \tag{6}$$

This case has been discussed in the literature. It is well known that in this case (using (1) and (2)) we have

$$\begin{aligned} I(x, y, t) &= I(x - V_x t, y - V_y t, 0) \\ &= I_0(x - V_x t, y - V_y t) \end{aligned} \tag{7}$$

and

$$\widehat{I}(\omega_x, \omega_y, \omega_t) = \widehat{I}_0(\omega_x, \omega_y)\delta(\omega_x V_x + \omega_y V_y + \omega_t). \tag{8}$$

Clearly, in this case, the energy of $\widehat{I}(\omega_x, \omega_y, \omega_t)$ is restricted to the plane $\omega_x V_x + \omega_y V_y + \omega_t = 0$ (which we refer to as the main plane).

Using (3) and (4) we get then

$$\begin{aligned} \widehat{I}_d(\omega_x, \omega_y, \omega_t) &= \sum_\ell \sum_m \sum_n \widehat{I}\left(\omega_x - \ell\frac{2\pi}{\Delta x}, \omega_y - m\frac{2\pi}{\Delta y}, \omega_t - n\frac{2\pi}{\Delta t}\right) \\ &= \sum_\ell \sum_m \sum_n \widehat{I}_0\left(\omega_x - \ell\frac{2\pi}{\Delta x}, \omega_y - m\frac{2\pi}{\Delta y}\right)\delta\left(\left(\omega_x - \ell\frac{2\pi}{\Delta x}\right)V_x\right. \\ &\quad \left. + \left(\omega_y - m\frac{2\pi}{\Delta y}\right)V_y + \left(\omega_t - n\frac{2\pi}{\Delta t}\right)\right). \end{aligned} \tag{9}$$

Namely, the sampled data energy is restricted to the main plane and its shifted versions. Clearly, as long as the main plane contains no elements of the *reciprocal lattice* (see e.g., Dubois, 1985)[1] other than $(0, 0, 0)$ we are guaranteed not to have aliasing (even if $I_0(x, y)$ is not band limited). That is, iff

$$\ell\frac{2\pi}{\Delta x}V_x + m\frac{2\pi}{\Delta y}V_y + n\frac{2\pi}{\Delta t} \neq 0 \quad \forall(\ell, m, n) \neq (0, 0, 0). \tag{10}$$

To further investigate this condition let us denote

$$\widetilde{V}_x = \frac{\Delta t}{\Delta x}V_x; \quad \widetilde{V}_y = \frac{\Delta t}{\Delta y}V_y \tag{11}$$

which are the velocities (or, rather, the optical flow) expressed in the units (pixels/frame). Then (10) can be rewritten as

$$\ell\widetilde{V}_x + m\widetilde{V}_y + n \neq 0 \quad \forall(\ell, m, n) \neq (0, 0, 0).$$

Or, equivalently,

---

[1] The sampling lattice here is $\{(\ell\Delta x, m\Delta y, n\Delta t) : \ell, m, n \in \mathcal{Z}\}$ and the reciprocal lattice is then $\{(\ell(2\pi/\Delta x), m(2\pi/\Delta y), n(2\pi/\Delta t) : \ell, m, n \in \mathcal{Z})\}$.

$$\ell\widetilde{V}_x + m\widetilde{V}_y \neq \left[\ell\widetilde{V}_x + m\widetilde{V}_y\right] \quad \forall(\ell, m) \neq (0, 0), \tag{12}$$

where $[a] = \text{round}(a)$ (the closest integer to $a \in \Re$).

Observing (12) we note that if either $\widetilde{V}_x$ or $\widetilde{V}_y$ or $\widetilde{V}_x/\widetilde{V}_y$ are rational numbers (12) does not hold. This means that, for $I_0(x, y)$ which is not band limited, aliasing is unavoidable. When $I_0(x, y)$ is band limited, say that

$$\widehat{I}_0(\omega_x, \omega_y) = 0 \quad \text{for all } (\omega_x, \omega_y) \text{ such that } |\omega_x| > \frac{W_x}{2}, \text{ or } |\omega_y| > \frac{W_y}{2}$$

(12) can be relaxed to

$$\ell\widetilde{V}_x + m\widetilde{V}_y \neq \left[\ell\widetilde{V}_x + m\widetilde{V}_y\right], \tag{13}$$

$$\forall(\ell, m) \neq (0, 0) \text{ such that } |\ell| \leqslant \frac{W_x\Delta x}{2\pi} \text{ and } |m| \leqslant \frac{W_y\Delta y}{2\pi}. \tag{14}$$

Velocities for which (13) does not hold are referred to in the literature as 'critical velocities' (see Tekalp, 1995).

Another constructive way of stating (14) is through defining

$$\beta = \min_{\substack{(\ell,m)\neq(0,0) \\ |\ell| \leqslant \frac{W_x\Delta x}{2\pi} \text{ and } |m| \leqslant \frac{W_y\Delta y}{2\pi}}} \left|\ell\widetilde{V}_x + m\widetilde{V}_y - \left[\ell\widetilde{V}_x + m\widetilde{V}_y\right]\right|. \tag{15}$$

Then, (13) is equivalent to the requirement $\beta > 0$. (Note that we have postulated that there is aliasing in a single frame, namely, we have both $(2\pi/\Delta x) < W_x$ and $(2\pi/\Delta y) < W_y$—otherwise there is no need of SR reconstruction).

The significance of β will be further discussed later. It is however clear, that it provides a measure of how far one is from aliasing in the ideal case discussed here. Hence, the larger β is the more robust the assumption of ideal case. Namely, when, as a result of finite image size, the energy of image Fourier Transform (FT) is not restricted to a plane anymore, the size of β is an indication of what combinations of data size, sampling intervals and velocities will avoid aliasing. Clearly, from (15) there are velocities which are preferable as far as aliasing and SR reconstruction is concerned. It can readily be shown that $\beta = 0$ for the critical velocities mentioned in (Tekalp, 1995), in which case reconstruction is impossible.

Furthermore, we observe that the pair (ℓ, m) leading to β identify the closest relevant shifted plane. Hence, $(2\pi/\Delta t)\beta$ is the distance between the closest relevant shifted plane and the main plane. This leads to the following possible reconstruction procedure.

Assuming that $\beta > 0$ the original image can be reconstructed by passing the data through the filter defined by its frequency response

$$\widehat{h}(\omega_x, \omega_y, \omega_t) = \text{rect}_{W_x}(\omega_x)\text{rect}_{W_y}(\omega_y)\text{rect}_{W_t}(\omega_x V_x + \omega_y V_y + \omega_t), \tag{16}$$

where

$$\text{rect}_a(x) = \begin{cases} 1 & \text{for } |x| \leqslant \frac{a}{2}, \\ 0 & \text{otherwise}, \end{cases} \tag{17}$$

and, using our observation earlier, we allow half the minimal distance around the main plane. Namely, we choose,

$$W_t = \frac{2\pi}{\Delta t}\beta. \tag{18}$$

The result will then be (see (9))

$$\begin{aligned} \widehat{I}_{\text{est}}(\omega_x, \omega_y, \omega_t) &= \widehat{h}(\omega_x, \omega_y, \omega_t)\widehat{I}_d(\omega_x, \omega_y, \omega_t) \\ &= \widehat{I}_0(\omega_x, \omega_y)\delta(\omega_x V_x + \omega_y V_y + \omega_t) \\ &= \widehat{I}(\omega_x, \omega_y, \omega_t). \end{aligned}$$

Namely, we have achieved perfect reconstruction. This type of filter is referred to in the literature as 'motion compensated filter' (see Tekalp, 1995).

The dependence of β on W_x, W_y, $\widetilde{V}_x$ and $\widetilde{V}_y$ (as presented in Eq. (15)) is quite intriguing. One could ask, for given Δx, Δy, Δt, W_x and W_y, what is the largest β possible. Namely, pose the following problem: Find

$$\beta^* = \max_{\widetilde{V}_x, \widetilde{V}_y} \min_{\substack{(\ell, m) \neq (0,0) \\ |\ell| \leqslant \frac{W_x \Delta x}{2\pi} \text{ and } |m| \leqslant \frac{W_y \Delta y}{2\pi}}} \left| \ell \widetilde{V}_x + m \widetilde{V}_y - \left[\ell \widetilde{V}_x + m \widetilde{V}_y \right] \right|. \tag{19}$$

The solution is given in the following proposition.

Proposition 1. *With β^* as defined in* (19) *we have*

$$\beta^* = \frac{1}{(L+1)(M+1)}, \tag{20}$$

where $L = fix((W_x \Delta x)/2\pi)$ and $M = fix((W_y \Delta y)/2\pi)$. (By fix(a) we mean, round real a to the nearest integer towards the origin.) This maximum is achieved with

$$\widetilde{V}_x^* = \beta^* = \frac{1}{(L+1)(M+1)}; \quad \widetilde{V}_y^* = \frac{1}{M+1}$$

($\widetilde{V}_x^*$ and $\widetilde{V}_y^*$ are not necessarily unique).

Proof. (The proof of this proposition appears in the appendix.) □

Note that, since we assumed that both $W_x > (2\pi/\Delta x)$ and $W_y > (2\pi/\Delta y)$, Proposition 1 implies that $\beta < 0.25$.

3. General motion

In this section we let the motion (as reflected by the optical flow) be general but assume it is known and has some continuity properties to be specified in the sequel.

Our approach to this general case is motivated by the discussion in the previous section. It is described in Fig. 2.

From Fig. 2 we have

$$
\begin{aligned}
I_{\mathrm{est}}^{0}(x,y,t) &= h^{0}(x,y,t) * * * \left(w^{0}(x,y,t)s(x,y,t)I(x,y,t) \right) \\
&= h^{0}(x,y,t) * * * I_{d}^{0}(x,y,t),
\end{aligned}
\tag{21}
$$

where

$$
\begin{aligned}
I_{d}^{0}(x,y,t) &= w^{0}(x,y,t)I_{d}(x,y,t) \\
&= w^{0}(x,y,t)s(x,y,t)I(x,y,t)
\end{aligned}
\tag{22}
$$

and $* * *$ denoted the $2D + T$ convolution. Defining

$$
I^{0}(x,y,t) = w^{0}(x,y,t)I(x,y,t),
\tag{23}
$$

we can rewrite the above as

$$
I_{d}^{0}(x,y,t) = s(x,y,t)I^{0}(x,y,t).
\tag{24}
$$

By doing this the process described in Fig. 2 can be viewed as consisting of two parts (see Fig. 3): (1) Windowing the *image* and reconstructing the original image from its *windowed versions* and (2) sampling the *windowed images* and reconstructing the *sampled windowed images* from their sampled data.

This observation helps us in the choice of the windows and filters to be used and is discussed next.
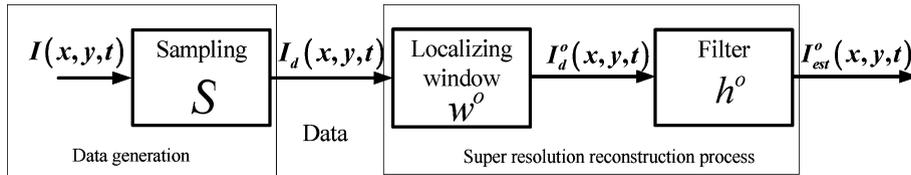

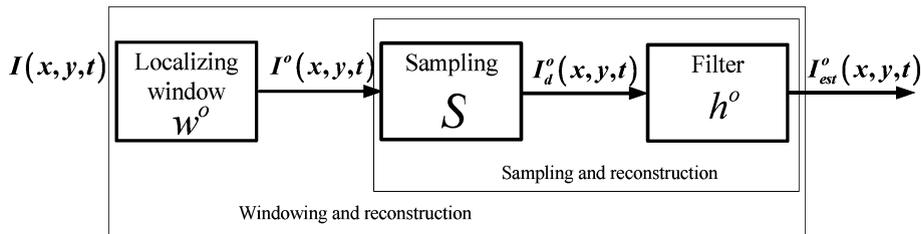Fig. 2. General super-resolution reconstruction.

Fig. 3. Super-resolution reconstruction as a two part process.

3.1. Motion compensated windowing

Consider a window $w(x, y, t)$ symmetric around the origin, with $w(0, 0, 0) = 1$ and $w(x, y, t) \approx 0$ outside a finite support Ω which contains the origin. Then we can write

$$I(x, y, t) = \int^{\infty} \int_{-\infty} \int I^0(x, y, t)\delta(x - x_0, y - y_0, t - t_0)\, dx_0\, dy_0\, dt_0, \qquad (25)$$

where

$$I^0(x, y, t) = w(x - x_0, y - y_0, t - t_0)I(x, y, t). \qquad (26)$$

Namely, if for every (x_0, y_0, t_0) we have the windowed image $I^0(x, y, t)$, $I(x, y, t)$ can be reconstructed.

Our purpose in introducing windowed images is to create a friendlier environment from a sampling and reconstruction point of view. Motivated by the results of the previous section we would like to have

$$I^0(x, y, t) \approx w(x - x_0, y - y_0, t - t_0)I_0(x - V_x^0 t, y - V_y^0 t) \qquad (27)$$

in which case $I_0(x_0 - V_x^0 t_0, y_0 - V_y^0 t_0) \approx I(x_0, y_0, t_0)$.

Observing the optical flow constraint Eq. (1), we note that (27) will follow from (26) if we have

$$w(x - x_0, y - y_0, t - t_0)\begin{pmatrix} V_x(x, y, t) \\ V_y(x, y, t) \end{pmatrix} \approx w(x - x_0, y - y_0, t - t_0)\begin{pmatrix} V_x(x_0, y_0, t_0) \\ V_y(x_0, y_0, t_0) \end{pmatrix} \qquad (28)$$

and take $V_x^0 = V_x(x_0, y_0, t_0)$, $V_y^0 = V_y(x_0, y_0, t_0)$ so that (28) holds exactly at $(x, y, t) = (x_0, y_0, t_0)$.

Assuming sufficiently smooth optical flow and small enough support of $w(x, y, t)$ the approximation (27) holds. Namely, *a small support, Ω, of $w(x, y, t)$ is a desired property of the window*.

The RHS of the approximation in (27) contains $I_0(x - V_x^0 t, y - V_y^0 t)$ for which we have shown that its energy in the frequency domain is restricted to the plane $\omega_x V_x^0 + \omega_y V_y^0 + \omega_t = 0$. However, as we well know, the window multiplying it will cause the FT energy to disperse away from the plane (since $\widehat{I^0} = \widehat{w} * * * \widehat{I_0}$). The larger the support of $\widehat{w}$ the further away from the plane the energy dispersion. This dispersion clearly increases the danger of aliasing in the sampled data of I^0. Hence, *a small support of $\widehat{w}(\omega_x, \omega_y, \omega_t)$ is a desired property of the window*.

Recalling Heisenberg's Uncertainty Principle we realize that the (practical) supports of both w and $\widehat{w}$ cannot be decreased simultaneously. To guarantee as little aliasing in I_d^0 as possible we choose first the support of $\widehat{w}$ and the support of w is then determined. To improve the approximation for the resulting support of w we use *motion compensated windowing*. Namely, for every point (x_0, y_0, t_0) we use the window

$$w^0(x - x_0, y - y_0, t - t_0) = w(x - x_0 - V_x^0(t - t_0), y - y_0 - V_y^0(t - t_0), t - t_0). \qquad (29)$$

To demonstrate the benefit from using motion compensated windowing let us consider a Gaussian window:

$$w(x, y, t) = e^{-(1/2)(x^2/\sigma_x^2 + y^2/\sigma_y^2 + t^2/\sigma_t^2)} \Longleftrightarrow \widehat{w}(\omega_x, \omega_y, \omega_t)$$
$$= (\sqrt{2\pi})^3 \sigma_x \sigma_y \sigma_t e^{-(1/2)(\sigma_x^2 \omega_x^2 + \sigma_y^2 \omega_y^2 + \sigma_t^2 \omega_t^2)}. \tag{30}$$

Clearly, the choices of σ_x, σ_y, σ_t determine the shape and size of the supports of the window in both domains. Let us define the support of the Gaussian function by choosing $q > 0$ such that for any $|z| > q$, $e^{-(z^2/2)} \approx 0$ (typically, $q \geqslant 3$). Then, it can be shown that, for the window in (30) to avoid aliasing (see discussion in previous section) we need to satisfy

$$q\sqrt{\left(\frac{V_x^0}{\sigma_x}\right)^2 + \left(\frac{V_y^0}{\sigma_y}\right)^2 + \left(\frac{1}{\sigma_t}\right)^2} \leqslant \frac{2\pi}{\Delta t} \beta. \tag{31}$$

Using now the motion compensated window, $w^0(x, y, t) = w(x - V_x^0 t, y - V_y^0 t, t)$, the condition becomes

$$\frac{q}{\sigma_t} \leqslant \frac{2\pi}{\Delta t} \beta. \tag{32}$$

The constraint in (32) is clearly less restrictive than in (31). Furthermore, it determines a lower bound on σ_t only while σ_x, σ_y can still be chosen sufficiently small to give a small support for $w^0(x, y, t)$ and thus, making the approximation in (27) more accurate.

We conclude that for every type of window one would consider, making it motion compensated would improve the performance of the SR reconstruction process proposed here.

3.2. Motion compensated filter

Going back to Eq. (21) we note that as far as reconstruction from sampled data is concerned we can concentrate on the windowed data $I^0(x, y, t)$ and its sampled version $I_d^0(x, y, t)$. We design the reconstruction filter for this data and using the approximation in (27), we can use the same (motion compensated) filter structure as in the previous section

$$\widehat{h^0}(\omega_x, \omega_y, \omega_t) = \text{rect}_{W_x}(\omega_x)\text{rect}_{W_y}(\omega_y)\text{rect}_{W_t^0}(\omega_x V_x^0 + \omega_y V_y^0 + \omega_t). \tag{33}$$

Note however, that V_x^0, V_y^0, and W_t^0 depend on (x_0, y_0, t_0) and generally, will be different at different points. Hence, the filter we will use is both spatial and temporal varying.

3.3. SR reconstruction process—implementation issues

The impulse response of the filter we have is given by (see (33))

$$h^0(x,y,t) = \frac{W_x W_y W_t^0}{8\pi^3} \text{Sinc}\left(\frac{W_x}{2\pi}\left(x - V_x^0 t\right)\right) \text{Sinc}\left(\frac{W_y}{2\pi}\left(y - V_y^0 t\right)\right) \text{Sinc}\left(\frac{W_t^0}{2\pi} t\right)$$
$$= h(x,y,t) *** \delta\left(x - V_x^0 t, y - V_y^0 t\right). \tag{34}$$

Using (22), (29), and (34) we can write

$$I_d(x,y,t)\delta(x - x_0, y - y_0, t - t_0)$$
$$= \left\{h^0(x,y,t) *** \left(w^0(x - x_0, y - y_0, t - t_0)I_d(x,y,t)\right)\right\}\delta(x - x_0, y - y_0, t - t_0)$$
$$= \left\{\left(h^0(x,y,t)w^0(x,y,t)\right) *** I_d(x,y,t)\right\}\delta(x - x_0, y - y_0, t - t_0) \tag{35}$$
$$= \left\{h_p^0(x,y,t) *** I_d(x,y,t)\right\}\delta(x - x_0, y - y_0, t - t_0), \tag{36}$$

where we have also used the symmetry of the window around the origin. Namely, we can combine the motion compensated filter and motion compensated window into a modified filter

$$h_p^0(x,y,t) = h^0(x,y,t)w^0(x,y,t)$$
$$= h(x,y,t)w(x,y,t) *** \delta(x - V_x^0 t, y - V_y^0 t). \tag{37}$$

This last modification is very significant for our implementation. Recognizing that any implementation will be done in the discrete domain, the modified filter is (practically) an FIR filter hence, easily implemented.

At this point we are ready to outline the actual implementation steps for our SR reconstruction process

Algorithm description

Available are:
1. Sampling intervals Δx, Δy, Δt
2. Resolution increase ratios R_x, R_y, R_t
3. Choice of window type, $w(x,y,t)$.
4. Data set

$$I_d[\ell, m, n] = I(\ell\Delta x, m\Delta y, n\Delta t)$$

$$0 \leqslant \ell \leqslant N_L; \quad 0 \leqslant m \leqslant N_M \quad 0 \leqslant n \leqslant N_N$$

Reconstruction steps:
Step 1: Upsample $I_d[\ell, m, n]$ to get

$$\widetilde{I}_d[\widetilde{\ell}, \widetilde{m}, \widetilde{n}] = I\left(\widetilde{\ell}\frac{\Delta x}{R_x}, \widetilde{m}\frac{\Delta y}{R_y}, \widetilde{n}\frac{\Delta t}{R_t}\right)$$

$$0 \leqslant \widetilde{\ell} \leqslant R_x N_L; \quad 0 \leqslant \widetilde{m} \leqslant R_y N_M \quad 0 \leqslant \widetilde{n} \leqslant R_t N_N$$

by zero insertion.
Step 2: Estimate optical flow at all points $(\widetilde{\ell}\frac{\Delta x}{R_x}, \widetilde{m}\frac{\Delta y}{R_y}, \widetilde{n}\frac{\Delta t}{R_t})$.

Step 3: For any desired point

$$(x_0, y_0, t_0) = \left(\widetilde{\ell}_0 \frac{\Delta x}{R_x}, \widetilde{m}_0 \frac{\Delta y}{R_y}, \widetilde{n}_0 \frac{\Delta t}{R_t} \right)$$

use $W_x = \frac{2\pi}{\Delta x} R_x$, $W_y = \frac{2\pi}{\Delta y} R_y$ and:

(*a*) Calculate (see (15))

$$\beta = \min_{\substack{|\ell| \leqslant R_x - 1 \\ |m| \leqslant R_y - 1}} \left| \ell \widetilde{V}_x^0 + m \widetilde{V}_y^0 - \left[\ell \widetilde{V}_x^0 + m \widetilde{V}_y^0 \right] \right|$$

and

$$W_t^0 = \frac{2\pi}{\Delta t} \beta$$

where

$$\widetilde{V}_x^0 = \frac{\Delta t}{\Delta x} V_x(x_0, y_0, t_0)$$

$$\widetilde{V}_y^0 = \frac{\Delta t}{\Delta y} V_y(x_0, y_0, t_0)$$

(*b*) Calculate the window $w^0(x, y, t)$.

(*c*) Calculate the impulse response of the modified filter $h_p^0[\widetilde{\ell}, \widetilde{m}, \widetilde{n}]$ using Eqs. (34) and (38).

(*d*) Calculate

$$I_{\mathrm{est}}^0[\widetilde{\ell}, \widetilde{m}, \widetilde{n}] = h_p^0[\widetilde{\ell}, \widetilde{m}, \widetilde{n}] * * * \widetilde{I}_d[\widetilde{\ell}, \widetilde{m}, \widetilde{n}]$$

(*e*) Take

$$I_{\mathrm{est}}[\widetilde{\ell}_0, \widetilde{m}_0, \widetilde{n}_0] = I_{\mathrm{est}}^0[\widetilde{\ell}_0, \widetilde{m}_0, \widetilde{n}_0]$$

4. SR reconstruction—examples

To demonstrate our reconstruction process we present two examples. One with spatio-temporal invariant translation and one with rotation. In both cases, in order to avoid the need of optical flow estimation we have generated the data from a single high resolution frame with predetermined motion. This motion information was then used in the algorithm. The algorithm was constructed according to the outline in the previous section and then applied to the data in each case.

The window chosen was of the form

$$w(x, y, t) = \mathrm{Hann}_{W_1}(x)\mathrm{Hann}_{W_1}(y)\mathrm{Hann}_W(t),$$

where

$$\mathrm{Hann}_{W_1}(x) = \begin{cases} \frac{1}{2}\left(1 + \cos(\frac{\pi x}{W_1})\right) & \text{for } |x| \leqslant W_1, \\ 0 & \text{otherwise.} \end{cases}$$

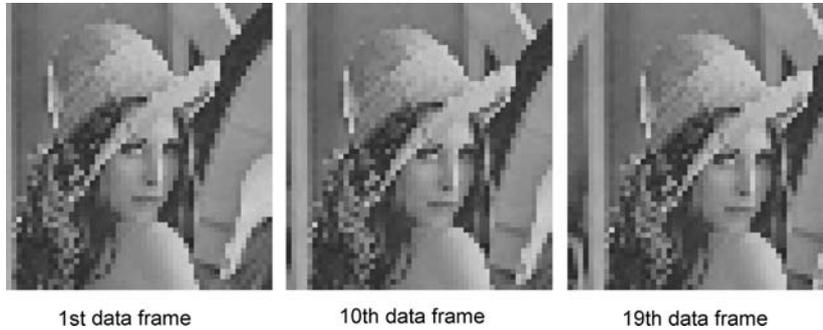1st data frame   10th data frame   19th data frame

Fig. 4. Data frames for Experiment 1.

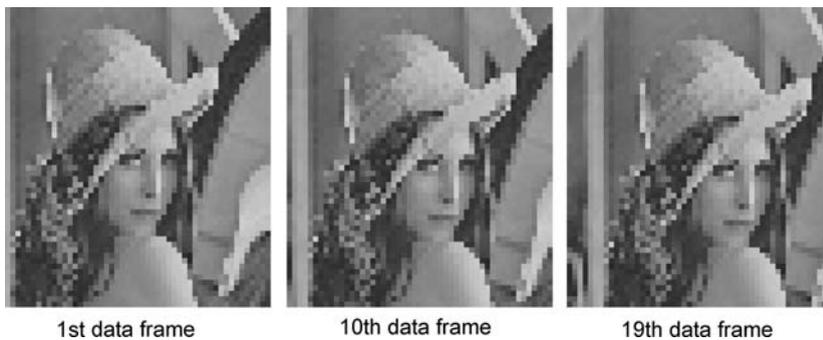1st data frame   10th data frame   19th data frame

Fig. 5. Data frames for Experiment 2.

Fig. 6. The spectrum of a single data frame.

The data in both examples was generated from a 512×512 frame of Lenna. In each case the motion was used to create data frames of 55×51 pixels each. The resolution improvement we generated for both examples is $R_x = R_y = 3$ and $R_t = 1$ (i.e., no resolution improvement in the temporal direction). Since our choice for the window size for both experiments was $W_1 = 7$ *samples* and $W = 9$ *frames* the number of data frames generated was $19 (= 2 \times 9 + 1)$.

In Figs. 4 and 5 we show data frames for Experiments 1 and 2, respectively (zoomed in to 300%). The lack of detail is apparent and in the spectrum of one such frame, seen in Fig. 6, we can clearly observe aliasing.

Fig. 7. An interpolated (lowpass filtered) single data frame.

Fig. 8. The spectrum of the $(2D + T)$ image at $\omega_t = 0$ (Example 1).

An attempt to reconstruct a frame of higher resolution using a lowpass filter (interpolation) on a single data frame is presented in Fig. 7. While somewhat more 'pleasant' to look at, this frame contains no additional information compared to original data frames (Fig. 4) hence, is not a true SR reconstruction.

Example 1. Using $\widetilde{V}_x = \frac{1}{9}$ and $\widetilde{V}_y = \frac{1}{3}$ we generated the 19 frames constituting the sampled version of the $2D + T$ image. The $2D + T$ spectrum of this data is presented in Fig. 8—actually, in the figure we present the spectrum at the plane $\omega_t = 0$. We can clearly observe in the figure the concentration of energy around the main plane $\omega_x + 3\omega_y + 9\omega_t = 0$ and its shifted versions (in fact, we see the lines of intersection of these planes with $\omega_t = 0$). Hence, as predicted by our analysis, there is (almost) no aliasing in the data and properly carried our SR reconstruction is possible.

For the velocities given we have calculated $\beta = \frac{1}{9}$ and the result of our reconstruction is the frame in Fig. 9. Comparing with Fig. 4 (or Fig. 7) we observe clearly the added details in the reconstructed frame.

In order to get a quality measure for the reconstructed frame we scaled its intensity to the range and calculated the MSE (compared to the true frame down sampled by a factor of 3). The reconstructed frame has MSE = 0.0003466 (with both frames intensities scaled to $[0, 1]$).

Example 2. To generate the data for this example we have used the optical flow as given in Fig. 10 (for illustrational purposes we present in the figure the optical flow values only for a downsampled set of pixels). This optical flow is the result of a

Fig. 9. SR reconstructed frame from the spatio-temporal invariant motion data (Example 1).
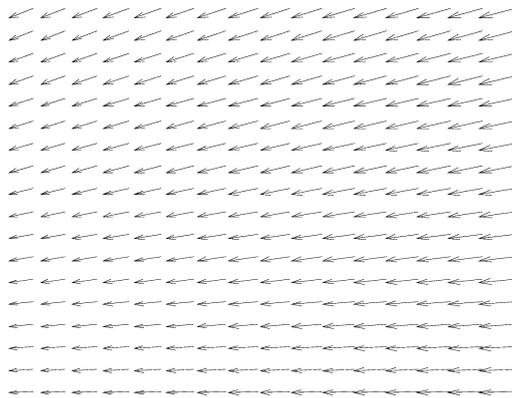

Fig. 10. Optical flow (at a subsampled set of pixels) of the rotational data (Example 2).

rotation around a point outside the frame. Again, 19 frames were generated at the low resolution as our data. Using our algorithm we have carried out an SR reconstruction and the resulting frame is presented in Fig. 11. To demonstrate the local

Fig. 11. SR reconstructed frame from the rotational data (Example 2).

nature of the algorithm we have reconstructed only a section of 40×40 pixels, highlighted in the figure by a white frame around it. Again, the results are quite encouraging.

We again calculated the MSE (only for the reconstructed section) and got MSE = 0.002033. This is clearly worse than the result in our first experiment but hardly surprising. Experiments we conducted with a more drastic changes in optical flow gave even worse results. As we pointed out, the reconstruction proposed does depend on the assumption that within a localizing window the optical flow is almost constant. The further we are from this assumption the worse the results. Clearly, increasing the frame rate will improve the results.

5. Conclusion

We have presented here an SR reconstruction method which uses spatio-temporal varying filtering. In the process, we have investigated the conditions under which the reconstruction is possible from the sampling and aliasing point of view. By viewing the sampled data in the $2D + T$ dimensional space much insight and understanding have been gained and are described here.

In our reconstruction, as is common to most such existing algorithms, we assume knowledge of the optical flow. The general procedure we outline is motivated by the simple case of spatio-temporal invariant optical flow (constant translational motion) and uses *motion compensated* filters and *motion compensated* localizing windows and filters. Hence, it is localized in nature and can be used to reconstruct any desired part of a single frame or image (video clip).

We should point out that in this paper we have just presented the approach and its feasibility. The question of optimizing the choices of the different design parameters and window shapes is beyond the scope here and is currently under investigation

Acknowledgments

Appendix

Proof. In this appendix we prove Proposition 1. The proof will consist of two parts. First we will show that for every real $\widetilde{V}_x$ and $\widetilde{V}_y$

$$\min_{\substack{(\ell,m)\neq(0,0) \\ |\ell| \leqslant L \text{ and } |m| \leqslant M}} \left| \ell \widetilde{V}_x + m \widetilde{V}_y - \left[\ell \widetilde{V}_x + m \widetilde{V}_y \right] \right| \leqslant \frac{1}{(L+1)(M+1)} \tag{A.1}$$

and then, that for $\widetilde{V}_x^* = 1/((L+1)(M+1))$ and $\widetilde{V}_y^* = 1/(\max\{(L+1),(M+1)\})$ we get

$$\min_{\substack{(\ell,m)\neq(0,0) \\ |\ell|\leqslant L \text{ and } |m|\leqslant M}} \left|\ell\widetilde{V}_x^* + m\widetilde{V}_y^* - \left[\ell\widetilde{V}_x^* + m\widetilde{V}_y^*\right]\right| = \frac{1}{(L+1)(M+1)} \tag{A.2}$$

which establishes that the RHS is indeed the maximum.

Let us first define, for any given $\widetilde{V}_x$ and $\widetilde{V}_y$, the sequence

$$a_k = \ell_k\widetilde{V}_x + m_k\widetilde{V}_y - \left[\ell_k\widetilde{V}_x + m_k\widetilde{V}_y\right], \tag{A.3}$$

where (ℓ_k, m_k) are all the distinct integer pairs with $0\leqslant\ell_k\leqslant L$ and $0\leqslant m_k\leqslant M$. We will assume that the sequence $\{a_k\}$ is ordered. Namely,

$$-\frac{1}{2}\leqslant a_1\leqslant a_2\leqslant\cdots\leqslant a_{K-1}\leqslant a_K\leqslant\frac{1}{2}, \tag{A.4}$$

where $K = (L+1)(M+1)$ is the total number of distinct pairs (ℓ_k, m_k), hence number of elements in the sequence. Using the "pigeon-hole" principle (see e.g., Schrijver, 1986) we can readily argue that

$$\min_{1\leqslant k\leqslant K-1}(a_{k+1} - a_k)\leqslant\frac{a_K - a_1}{K-1}. \tag{A.5}$$

Then, either

$$0\leqslant a_K - a_1\leqslant\frac{K-1}{K} \tag{A.6}$$

in which case, from (A.5)

$$0\leqslant\min_{1\leqslant k\leqslant K-1}(a_{k+1} - a_k)\leqslant\frac{1}{K} \tag{A.7}$$

and

$$\min_{1\leqslant k\leqslant K-1}\left|(\ell_{k+1} - \ell_k)\widetilde{V}_x + (m_{k+1} - m_k)\widetilde{V}_y - \left[(\ell_{k+1} - \ell_k)\widetilde{V}_x + (m_{k+1} - m_k)\widetilde{V}_y\right]\right|$$
$$\leqslant\min_{1\leqslant k\leqslant K-1}\left|(\ell_{k+1} - \ell_k)\widetilde{V}_x + (m_{k+1} - m_k)\widetilde{V}_y - \left[\ell_{k+1}\widetilde{V}_x + m_{k+1}\widetilde{V}_y\right]\right.$$
$$\left.+\left[\ell_k\widetilde{V}_x + m_k\widetilde{V}_y\right]\right| = \min_{1\leqslant k\leqslant K-1}|a_{k+1} - a_k|\leqslant\frac{1}{K}. \tag{A.8}$$

Or

$$\frac{K-1}{K} < a_K - a_1\leqslant 1$$

in which case

$$0\geqslant a_K - a_1 - [a_K - a_1] > \frac{K-1}{K} - 1 = -\frac{1}{K}. \tag{A.9}$$

Then, since

$$\left| (\ell_K - \ell_1)\widetilde{V}_x + (m_K - m_1)\widetilde{V}_y - \left[(\ell_K - \ell_1)\widetilde{V}_x + (m_K - m_1)\widetilde{V}_y \right] \right|$$
$$\leqslant \left| (\ell_K - \ell_1)\widetilde{V}_x + (m_K - m_1)\widetilde{V}_y - N \right|$$

for any integer N, we have from (A.9)

$$\left| (\ell_K - \ell_1)\widetilde{V}_x + (m_K - m_1)\widetilde{V}_y - \left[(\ell_K - \ell_1)\widetilde{V}_x + (m_K - m_1)\widetilde{V}_y \right] \right|$$
$$\leqslant |a_K - a_1 - [a_K - a_1]| \leqslant \frac{1}{K}. \tag{A.10}$$

Since for any distinct pairs $0 \leqslant \ell_j, \ell_n \leqslant L$ and $0 \leqslant m_j, m_n \leqslant M$ we have $|\ell_j - \ell_n| \leqslant L$, $|m_j - m_n| \leqslant M$ and $((\ell_j - \ell_n), (m_j - m_n)) \neq (0,0)$ it follows from (A.8) and (A.10) that

$$\min_{\substack{(\ell,m)\neq(0,0) \\ |\ell| \leqslant L \text{ and } |m| \leqslant M}} \left| \ell\widetilde{V}_x + m\widetilde{V}_y - \left[\ell\widetilde{V}_x + m\widetilde{V}_y \right] \right| \leqslant \frac{1}{K}$$

as required.

Let us now choose $\widetilde{V}_x = 1/((L+1)(M+1))$ and $\widetilde{V}_y = 1/(M+1)$, then $\ell\widetilde{V}_x + m\widetilde{V}_y = (\ell + m(L+1))/((L+1)(M+1))$ and since $|\ell\widetilde{V}_x + m\widetilde{V}_y - [\ell\widetilde{V}_x + m\widetilde{V}_y]| = |-\ell\widetilde{V}_x - m\widetilde{V}_y - [-\ell\widetilde{V}_x - m\widetilde{V}_y]|$ w.l.o.g. we can assume that $\ell\widetilde{V}_x + m\widetilde{V}_y > 0$. Then

$$\frac{1}{(L+1)(M+1)} \leqslant \ell\widetilde{V}_x + m\widetilde{V}_y = \frac{\ell + m(L+1)}{(L+1)(M+1)} \leqslant \frac{L + M(L+1)}{(L+1)(M+1)}$$
$$= 1 - \frac{1}{(L+1)(M+1)}$$

which directly implies that

$$\left| \ell\widetilde{V}_x + m\widetilde{V}_y - \left[\ell\widetilde{V}_x + m\widetilde{V}_y \right] \right| \geqslant \frac{1}{(L+1)(M+1)}.$$

Hence, one can either choose $(\ell, m) = (1, 0)$ or $(\ell, m) = (L, M)$ and in both cases the lower bound $1/((L+1)(M+1))$ is reached so that (A.2) is proven which completes the proof of the proposition. $\square$

References

Baker, S., Kanade, T., 2002. Limits on super-resolution and how to Breal Them. IEEE Trans on Pattern Analysis and Machine Intelligence 24 (9), 1167–1183.

Bose, N.K., Kim, S.P., Valenzuela, H.M., 1993. Recursive implementation of total least square algorithm for image reconstruction from noisy, undersampled multiframes, in: Proceedings of the IEEE International Conference ASSP (ICASSP), vol. V, Minneapolis MN, pp. 269–272.

Dubois, E., 1985. The sampling and reconstruction of time-varying imagery with application in video systems. Proceeding of IEEE 73 (4), 502–522.

Elad, M., Feuer, A., 1997. Restoration of a single super-resolution image from several blurred, noisy and under-sampled measured images. IEEE Transactions on Image Processing 6 (12), 1646–1658.

Elad, M., Feuer, A., 1999a. Super-resolution restoration of image sequence: adaptive filtering approach. IEEE Transactions on Image Processing 8 (3), 387–395.

Elad, M., Feuer, A., 1999b. Super-resolution reconstruction of image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 21, 817–834.

Hardie, R.C., Barnard, K.J., Armstrong, E.E., 1997. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. IEEE Transactions on Image Processing 6 (12), 1621–1633.

Huang, T.S., Tsai, R.Y., 1984. Multiframe image restoration and registration, in: T.S. Huang (Ed.), Advances in Computer Vision and Image Processing, Greenwich, CT, JAI (Chapter 7).

Irani, M., Peleg, S., 1993. Motion analysis for image enhancement: resolution, occlusion and transparency. Journal of Visual Communication Image Representation 4, 324–335.

Kim, S.P., Bose, N.K., Valenzuela, H.M., 1990. Recursive reconstruction of high resolution image from noisy undersampled multiframes. IEEE Transactions on ASSP 38, 1013–1027.

Patti, A.J., Sezan, M.I., Tekalp, A.M., 1994. High resolution image reconstruction from low resolution image sequence in the presence of time-varying motion blur, in: Proceedings of the ICIP, Austin, Texas, pp. 343–347.

Patti, A.J., Sezan, M.I., Tekalp, A.M., 1995. High resolution standards conversion of low resolution video, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Detroit, MI, vol. II, pp. 2197–2200.

Patti, A.J., Sezan, M.I., Tekalp, A.M., 1997. Super-resolution video reconstruction with arbitrary sampling lattices and nonsero aperture time. IEEE Transactions on Image Processing 6 (8), 1064–1076.

Peleg, S., Keren, D., Schweitzer,Improving, L., 1987. Image resolution using sub-pixel motion. Pattern Recognition Letters 5, 223–226.

Schrijver, A., 1986. Theory of Linear and Integer Programming. John Wiley & Sons, New York.

Schultz, R.R., Stevenson, R.L., 1994. A bayesian approach to image expansion for improved definition. IEEE Transactions on Image Processing 3 (3), 233–242.

Tekalp, A.M., 1995. Digital Video Processing. Prentice Hall PTR, NJ.

Tekalp, A.M., Ozkan, M.K., Sezan, M.I., 1992. High resolution image reconstruction from lower resolution image sequences and space varying image restoration, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), San-Francisco, CA, vol. III, pp. 169–172.