Is Super-Resolution with Optical Flow Feasible?

WenYi Zhao and Harpreet S. Sawhney

Sarnoff Corporation
201 Washington Road, Princeton, NJ 08540
{wzhao, hsawhney}@sarnoff.com

Abstract. Reconstruction-based super-resolution from motion video has been an active area of study in computer vision and video analysis. Image alignment is a key component of super-resolution algorithms. Almost all previous super-resolution algorithms have *assumed* that standard methods of image alignment can provide accurate enough alignment for creating super-resolution images. However, a systematic study of the demands on accuracy of multi-image alignment and its effects on super-resolution has been lacking. Furthermore, implicitly or explicitly most algorithms have assumed that the multiple video frames or specific regions of interest are related through global parametric transformations. From previous works, it is not at all clear how super-resolution performs under alignment with piecewise parametric or local optical flow based methods. This paper is an attempt at understanding the influence of image alignment and warping errors on super-resolution. Requirements on the *consistency* of optical flow across multiple images are studied and it is shown that errors resulting from traditional flow algorithms may render super-resolution infeasible.

1 Introduction

Enhancement of image resolution, called *super-resolution*, by processing multiple video images has been studied by many researchers over the past decade. The majority of super-resolution algorithms formulate the problem as a signal reconstruction problem from multiple samples. These algorithms are based on sampling theorems which state that given enough uniform or non-uniform samples, signals can be reconstructed. In single images captured over discrete grids, super-resolution may not be possible since all parts of the scene may not be adequately sampled by a single image. Multiple images captured using motions of the sensor or objects potentially provide adequate samples for super-resolution of any given image frame. However, ensuring the accuracy of sample locations from multiple images demands adequate alignment between multiple images that may be related through arbitrarily complex motion models. Almost all previous super-resolution algorithms have *assumed* that standard methods of image alignment can provide accurate enough alignment for creating super-resolution images. Implicitly or explicitly most algorithms have assumed that the multiple video frames or specific regions of interest are related through global parametric transformations. However, a systematic study of the demands on accuracy

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2350, pp. 599–613, 2002.

of multi-image alignment and its effects on super-resolution has been lacking. Moreover, when multiple video frames cannot be aligned by global parametric models, local models like optical flow need to be used for alignment. From previous works, it is not at all clear how super-resolution performs under alignment with piecewise parametric or local optical flow based methods.

This paper is an attempt at understanding the influence of image alignment and warping errors on super-resolution. We first present an analysis on how optical flow affects super-resolution algorithms. In particular, we adopt a general motion model composed of a global parametric model plus local flow [15]. To understand what is the impact of flow error on super-resolution, we introduce an image degradation model that explicitly incorporates the motion/flow error. We then convert such a geometric error into image noise. Focusing on the gradient-based flow computation leads us to discover an interesting phenomenon: large/small motion errors are associated with small/large image gradients. This suggests that image warping error is not as catastrophic as the flow error, and implies that image warping process may be well-behaved and hence flow-based super-resolution is feasible.

In order to address the core alignment issue itself, we experiment with novel flow algorithms. Though we show that it is image warping not flow that is directly linked to the super-resolution process, flow is especially critical for reconstruction of high-frequency components in the signal. The flow algorithms we employ address two issues: *flow consistency* and *flow accuracy*. Flow consistency implies that the flow computed from frame A to frame B should be consistent with that computed from B to A. Flow accuracy measures the absolute error in flow. The new algorithms take advantage of multi-image alignment in contrast with traditional flow algorithms devoted to pairwise image alignment. By computing all flows simultaneously, we propose a method that is similar to "bundle adjustment" used in parametric registration. Consistent and bundled flow estimation addresses the issues of flow consistency and accuracy.

2 Related Work

Work on super-resolution can be divided into two main categories: reconstruction-based methods [1-9] and learning-based methods [11-13]. The theoretical foundations for reconstruction methods are (non-)uniform sampling theorems while learning-based methods employ generative models that are learned from samples. The goal of the former is to *reconstruct* the original (super-sampled) signal while that of the latter is to *create* the signal based on learned generative models. In contrast with reconstruction methods, learning-based super-resolution methods assume that corresponding low-resolution and high-resolution training image pairs are available [11-13]. In [12], a general-purpose learning is applied and has demonstrated some good results for super-resolution.

The majority of super-resolution algorithms belong to the signal reconstruction paradigm. Among this category are frequency-based methods [9], Bayesian methods [4-6], BP (back-projection) methods [1], POCS (projection onto con-

vex set) methods [3], non-uniform sampling based methods [7-8] and hybrid methods [2]. These methods are deeply root in sampling theorems. According to these theorems, perfect reconstruction can be achieved as long as adequate sub-samples are available. Image alignment is typically used to ensure the availability of samples. Knowing if the samples are enough is important since in general the condition could vary from point to point in the images. There exist techniques that can be used to handle this issue automatically, e.g., regularization techniques. As for alignment, it is the key for the success of all reconstruction-based super-resolution algorithms. With accurate alignment, the reconstruction task is relatively easy. This is clearly demonstrated in [7].

The emphasis of this work is on the feasibility of super-resolution under general non-parametric motion models. Consistent estimation of *global* motion models (e.g. homographies) are therefore not reviewed here. Accurate alignment demanded by super-resolution may be difficult under general motion models since the system is not heavily over-determined as in the case of global parametric models. Another factor that contributes to the difficulty is that motion is typically estimated from noisy low-resolution images and interpolated to the higher resolution. Ideally, accurate motion estimation should be estimated from high-resolution images that are not available. Restricting the motion models to global parametric limits the application of super-resolution greatly. To work with general image sequences, we have to loosen such requirements. Otherwise, we would have to resort to solving a difficult segmentation problem: segment out moving objects and compute the parametric motion models for the objects. For non-rigid objects, it is even more difficult. In [17], a method for reliable block motion estimation is suggested for super-resolution in which multiple motion vectors are created for each block. However, our work shows that for reconstruction based methods, cross-frame consistency of motion estimates is important for super-resolution.

3 Flow-Based Super-Resolution

We first address the issue of feasibility of flow-based super-resolution.

3.1 Analysis of Flow Error for Super-Resolution

Classical Signal Reconstruction. Without loss of generality, we use 1D signals for analysis. We assume that the original high-resolution digital signal $f(n)$ has bandwidth $(-\omega_0, \omega_0)$ with $\omega_0 > \pi/2$ and two sub-sampled versions of the signal, $f1(n)$ and $f2(n)$, are available. There are two sub-sampling modes considered here: (i) decimation only, and (ii) blurring and decimation. In the decimation-only case: $f_1(n) = f(2n)$ and $f_2(n) = f(2n-1)$. Assume that the two sub-sampled signals are aliased, that is their spectra are created by wrapping the high-frequency part of the original signal into its low-frequency band. Mathematically, their Fourier transforms [18] are related to the Fourier transform of the original high-resolution signal as:

$$F_1(\omega) = \begin{cases} F(\omega/2), & \mid \omega \mid \leq 2\pi - 2\omega_0 \\ F(\omega/2) + A_1(\omega/2), & otherwise \end{cases}$$

and

$$F_2(\omega) = \begin{cases} F(\omega/2)\exp(-j\omega/2), & \mid \omega \mid \leq 2\pi - 2\omega_0 \\ F(\omega/2)\exp(-j\omega/2) - A_1(\omega/2)\exp(-j\omega/2), & otherwise \end{cases}$$

where $A_1(\omega/2) = F((\omega - 2\pi)/2) + F((\omega + 2\pi)/2)$ is the aliased component composed of the original mid-frequency and wrapped high-frequency. To recover the original signal using an estimated signal shift $\tilde{n}_0$, we perform the following operation: $F(\omega) = (F_1(2\omega) + F_2(2\omega)\exp(j\omega\tilde{n}_0))/2$.

In the case of perfect alignment $\tilde{n}_0 = 1$, the original signal can be perfectly recovered since the aliasing items are cancelled. In the case of imperfect alignment, $\delta\tilde{n}_0 \neq 0$, the reconstructed signal is as follows:

$$\tilde{F}(\omega) = \begin{cases} F(\omega)(1 + \exp[j\omega\delta\tilde{n}_0])/2, & \mid \omega \mid \leq 2\pi - 2\omega_0 \\ F(\omega)(1 + \exp[j\omega\delta\tilde{n}_0])/2 + (A_1(\omega) + A_2(\omega)\exp[j\omega\delta\tilde{n}_0])/2, & otherwise \end{cases}$$

Thus, while the reconstructed signal is affected across the whole spectrum by mis-alignment, the effect on the high-frequency components is more severe. In the case of blurring and decimation, we first recover the blurred signal and then perform de-blurring. The same analysis applies as long as the cut-off frequency of the blurring filter is equal to or larger than ω_0. In summary, alignment is critical for perfect reconstruction.

The above analysis does not take into account that alignment is computed from the sub-sampled signals. In the following we study the relationship between signal alignment and signal warping for a 2D signal.

Reconstruction based on Alignment and Warping. In order to model the high-to-low resolution image formation process, we adopt the matrix notations used in [2]. Specifically,

$$Y_k = D_k C_k F_k X + N_k \tag{1}$$

where X is the original high-resolution image, Y_k is the kth low-resolution frame, D_k, C_k , F_k are decimation, blurring and motion-warping matrices, respectively, that embody the corresponding transformations, N_k is the noise model at low-resolution. Note that given a motion representation (including optical flow), F_k can be used to represent the corresponding warping transformation. Based on this model and the assumption of zero-mean Gaussian noise, the ML-estimator of X from Y_k is

$$\tilde{X} = \arg_X \min\{(Y - HX)^T W (Y - HX)\} \tag{2}$$

where W is the weight matrix determined by noise N_k, matrix H is defined as $[H]_k = D_k C_k F_k$. However, this model does not allow for alignment error. In order to account for the uncertainty of estimated alignment, we augment the above model as:

$$Y_k = D_k C_k (F_k + \delta F_k) X + N_k = D_k C_k F_k X + (N_k + N_X^{F_k}) \tag{3}$$

where $N_X^{F_k} \stackrel{\text{def}}{=} \delta F_k X$. Comparing Eqs. (1) and (3), we notice an additional noise term. Now the ML-estimator needs to be modified as:

$$\tilde{X} = \arg_X \min\{(Y - HX)^T W^F (Y - HX)\} \tag{4}$$

where the new weight matrix is computed from both the noise terms $N_X^{F_k}$ and N_k.

Image warping. Before we study the misalignment induced noise term, it is important to review the basic warping technique. Given a source image I^s and the alignment transformation from this image to the warped image I^w, image warping can be formulated as:

$$I^w(\mathbf{p}) = \sum_i \alpha_i I^s(\mathbf{q}_i) \tag{5}$$

where α_i's are the interpolation coefficients, $I^s(\mathbf{q}_i)$ are pixels that surround the center pixel $I(\mathbf{q}_0)$ (very often $\mathbf{q}_0$ is not on the integer grid) in the source image and $I^w(\mathbf{p})$ is the corresponding pixel in the warped image. The given alignment parameter $\mathbf{u}$ defines the relation between $\mathbf{q}_0$ and $\mathbf{p}$: $\mathbf{p} = \mathbf{q}_0 + \mathbf{u}[\mathbf{p}]$.

Model for $N_X^{F_k}$ without N_k. We convert the geometric motion error into additional image noise via image warping. To simplify the analysis, we first assume that the original image noise can be ignored.

Using Eq. 5, the warped image $I^{F_k}(\mathbf{p})$ (matrix notation of $F_k X$) is obtained from I (matrix notation of X). Assuming an error in the estimated motion, the warped image changes to $\tilde{I}^{F_k}(\mathbf{p})$. For *small* motion error $d\mathbf{u}$, we can use a linear approximation:

$$\tilde{I}^{F_k}(\mathbf{p}) \approx I^{F_k}(\mathbf{p}) + I_x^{F_k}(\mathbf{p}) d\mathbf{u}_x + I_y^{F_k}(\mathbf{p}) d\mathbf{u}_y. \tag{6}$$

In the iterative process of estimating motion from the reference image I to a image I_k, at each iteration I is warped to match I_k [15]:

$$I_k(\mathbf{p}) \approx \sum_i \alpha_i I(\mathbf{q}_0 + \mathbf{u}[\mathbf{p}]). \tag{7}$$

Note that the alignment parameters are in the same coordinates $\mathbf{p}$ in Eqs. 6 and 7. Hence the components of $N_X^{F_k}$ can be obtained as:

$$\delta I^{F_k}(\mathbf{p}) \approx I_x^{F_k}(\mathbf{p}) d\mathbf{u}[\mathbf{p}]_x + I_y^{F_k}(\mathbf{p}) d\mathbf{u}[\mathbf{p}]_y. \tag{8}$$

This image noise (warping error) is dependent on pixel location. Now let us model the motion error and image noise across the whole image region. We further assume that the motion error is zero-mean (this might not be true in general but appears to be a valid assumption from our experiments), then the image noise has zero-mean. Its covariance can be computed as:

$$\text{var}(\delta I^{F_k}) = I_x^2(\sum I_y^2) + I_y^2(\sum I_x^2) - 2I_x I_y(\sum I_x I_y),$$

using the covariance matrix for the estimated motion:

$$\mathrm{cov}(d\mathbf{u}_x, d\mathbf{u}_y) = \mathrm{inv}(\begin{bmatrix} \sum I_x^2 & \sum(IxIy) \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix}).$$

It is well-known that larger image gradients lead to more precise motion estimation. Clearly the image noise due to warping error is not only a function of flow error but also of image gradients. In particular, relatively large motion errors due to small gradients are balanced by the small gradients as shown in Eq. (8). In other words, even if the motion is not very accurate, the warped image is still very close to the warped image based on true motion. In summary, in the case of small noise, optical flow can be used for the purpose of super-resolution.

3.2 Computing Consistent and Accurate Flows

Though we showed that it is image warping that directly influences the super-resolution process, flow is especially critical for reconstruction of high-frequency components.

Computing Consistent Flow. Consistent flow [19] between a pair of frames guarantees that the pair of flow fields, from frame 1 to frame 2 and from 2 to 1, will be consistent. In many applications, one-sided traditional flow algorithms are independently applied in the two directions, and points where the two flows are inconsistent are rejected. However, for reconstruction based super-resolution, consistent flow is essential since the maximum-likelihood estimator typically minimizes error between an iteratively reconstructed super-resolved image and the original low resolution images. Depending on the specific iterative technique used for the reconstruction, pixels are mapped both ways between the low and the high-resolution coordinate systems.

Traditional flow estimation algorithms ([15], for instance) compute a flow field between and image pair, I_1 and I_2, using *brightness constancy*:

$$I_1(\mathbf{p}_1) = I_2(\mathbf{p}_2), \tag{9}$$

where $\mathbf{p}_1$ and $\mathbf{p}_2$ are the coordinates of frame 1 and 2 respectively. At each iteration, a linearized approximation to the above equation if employed to solve for increments in the flow field.

$$I_t(\mathbf{p}_2) \approx \nabla I_2(\mathbf{p}_2)^T {J_{12}}^T \mathbf{u}_2[\mathbf{p}_2], \tag{10}$$

where J_{12} is the Jacobian partial derivative matrix of $\mathbf{p}_1$ w.r.t $\mathbf{p}_2$. Equation (10) is the basic equation of iterative multi-grid algorithms that computes the flow field from I_1 to I_2. The following approximation is employed for the Jacobian:

$$J_{12}\nabla I_2(\mathbf{p}_2) \simeq \frac{1}{2}(\nabla I_2(\mathbf{p}_2) + \nabla I_1(\mathbf{p}_2)) \tag{11}$$

The above technique can be used to compute the pair of flow fields from I_2 to I_1 and vice-versa. However, such a computation does not guarantee the following *consistency* constraint:

$$\begin{array}{l} \mathbf{p}_2 \quad\;\; = \mathbf{p}_1 + \mathbf{u}_1[\mathbf{p}_1] \\ \mathbf{u}_2[\mathbf{p}_2] = -\mathbf{u}_1[\mathbf{p}_1] \end{array} \tag{12}$$

To enforce the two-way flow consistency, we propose computing just one flow field, the consistent flow satisfying Eq (12) between any frame pair. Using the consistency constraint and the brightness constraint (Eq. 9), we can derive the *consistent* brightness constraint equation

$$I(\mathbf{p}) = I_1(\mathbf{p} - \frac{1}{2}\mathbf{u}[\mathbf{p}]) = I_2(\mathbf{p} + \frac{1}{2}\mathbf{u}[\mathbf{p}]), \tag{13}$$

where $I(\mathbf{p})$ is the virtual middle frame between the two frames. Using Taylor series expansion, we obtain the following differential form:

$$\begin{array}{rl} I_t(\mathbf{p}) & \stackrel{\text{def}}{=} I_1(\mathbf{p}) - I_2(\mathbf{p}) \\ & \approx \; \frac{1}{2}(\nabla I_1(\mathbf{p}) + \nabla I_2(\mathbf{p}))^T \mathbf{u}[\mathbf{p}]. \end{array} \tag{14}$$

All the coordinates are in the coordinate system of I. An iterative version of this new method can be readily derived. The advantages of computing consistent flow are: 1) only one consistent flow needs to be estimated for an image pair, and 2) the estimated flow by definition guarantees backward-forward consistency and hence may be more accurate.

Computing Bundled Flow Fields. We can generalize the notion of flow consistency over many frames by computing a consistent bundle of flow fields. Suppose consistent flow is to be computed between three frames, I_1 and I_3 to frame I_2 with the flow fields designated as $\mathbf{u}_1$ and $\mathbf{u}_3$, respectively. Frame I_2 is chosen as the reference frame I. Traditional two-frame methods to compute the two flows $\mathbf{u}_1[\mathbf{p}]$ and $\mathbf{u}_3[\mathbf{p}]$ are based on two independent constraints: $I_1(\mathbf{p_1}) = I(\mathbf{p})$ and $I_3(\mathbf{p_3}) = I(\mathbf{p})$. Again, consistency between $\mathbf{u}_1$ and $\mathbf{u}_3$ is not guaranteed if the two are computed independently. A straightforward way to enforce the consistency among flows is to add the following constraint: $I_3(\mathbf{p_3}) = I_1(\mathbf{p_3})$.

The iterative version of these constraints can be expressed in the common coordinate system of $\mathbf{p}$ as:

$$\begin{array}{l} I'_{t1} \;\; = I'_1 - I \;\; \approx \frac{1}{2}(\nabla I + \nabla I'_1))^T \delta\mathbf{u}_1 \\ I'_{t3} \;\; = I_3{}' - I \approx \frac{1}{2}(\nabla I + \nabla I'_3))^T \delta\mathbf{u}_3 \\ I'_{t13} = I'_1 - I'_3 \approx \frac{1}{2}[(\nabla I'_1)^T \delta\mathbf{u}_1 - (\nabla I'_3)^T \delta\mathbf{u}_3] \end{array} \tag{15}$$

where I'_i are the warped version of I_i using motion from the previous iteration, $\delta\mathbf{u}_1[\mathbf{p}]$ and $\delta\mathbf{u}_3[\mathbf{p}]$ are the incremental flows computed at each iteration. The bundled flow can be solved for by formulating the normal equations within a window centered at each pixel by assuming a constant flow model inside the window. For brevity, we omit the normal equations since these can be derived easily from Eq. 15.

Accurate and Consistent Flow Fields. Note that the error minimized above does not take into consideration the consistency between each pair of frames. This is not possible for pairs of frames other than the reference frame since to enforce pairwise consistency we need to use virtual coordinate systems for each pair of frames. To handle this issue, we propose the following algorithm:

- First compute consistent pairwise flow $\mathbf{u}_{i,i+1}$.
- Then cascade the consistent flows to obtain the initial flow estimates $\mathbf{u}_j$ from frame j to the reference frame.
- Bundle adjust the initial flow estimates.

3.3 Flow-Based Super-Resolution

We have analyzed the performance of super-resolution using a popular reconstruction based super-resolution algorithm [1]. However, any of the other algorithms [2-11] would also show a similar behavior since the error measures minimized in all of them are similar. Although in [1] only parametric motion models were employed, we use general flow as the alignment model.

Let us denote $I_h^{(n)}$ as the recovered high-resolution image and $g_k^{(n)}$ as the simulated low-resolution image of the k-th frame, at the n-th iteration. The iterative update of high-resolution image is expressed in [1] as:

$$I_h^{(n+1)} = I_h^{(n)} + \frac{1}{K}\sum_{k=1}^{K}\{[(g_k - g_k^{(n)}) \uparrow s]^{F_k} \cdot p\} \tag{16}$$

where K is the number of low-resolution images, p is a back-projection kernel, $\uparrow s$ denotes a up-sampling operator by a factor s, $[\cdot]^{F_k}$ denotes a forward-warping process. The simulated image $g_k^{(n)}$ is generated as follows:

$$g_k^{(n)} = [I_h^{(n)}]^{B_k} \cdot h\} \downarrow s \tag{17}$$

where $[\cdot]^{B_k}$ denotes a backward-warping process and h is a blurring kernel.

One important quantity embedded in the above formulation is the projection error that measures the quality of the super-resolution

$$E_p = \frac{1}{K}\sum_{k=1}^{K}(g_k - g_k^{(n)}) \tag{18}$$

In Section 4, we will investigate how flow algorithms can affect this important quality.

Estimating motion and performing super-resolution is a chicken-egg problem. To resolve this issue, the following incremental strategy can be used:

Iterative motion estimation and super-resolution.

1. Compute motion from original low-resolution images.

2. Interpolate low-resolution motion and perform super-resolution.
3. Use the super-resolved images to refine the motion at the high-resolution. (This step and the next are optional.)
4. Perform super-resolution using the refined motion and repeat steps 3 & 4 if needed.

Reject Warping Outliers. In order to cover all aspects of motion error, a reconstruction algorithm needs to account for warping errors in spite of accurate and consistent motion estimation procedures. Recall that the error analysis is based on small errors in flow computation. But this may not always be true for general scene or video content. For example, in the case of large object occlusions or scene changes in the video, the alignment is inherently wrong and super-resolution is not possible. In such cases, we need to detect anomalies in flow based on warping. We compute the cross-correlation between a target frame and the warped frame, and if the correlation score at a point is below a certain threshold, the corresponding warped pixels are ignored in the super-resolution process. The correlation scores can also be used as weights with a maximum-likelihood super-resolution estimator.

4 Experiments

We present experimental results with three types of datasets:

- Synthetic data that has synthetic motion and sub-sampled images. This data is generated from one high-resolution image by creating multiple images with synthetic motion.
- Semi-synthetic data that has synthetically sub-sampled images but real motion. This is generated from a real sequence of high-resolution images.
- Real sub-sampled image sequence.

The three different types of data allow us to systematically evaluate flow computation, the relation among flow error, warping error and super-resolution error, and the super-resolution results. We emphasize that for the last two sequences, rigid 3D motion (parallax) and non-rigid motions are present, therefore global motion estimates cannot be utilized.

For the back-projection algorithm, we choose the forward and backward filters as suggested by the authors in [1]. In our experiment, we use the interpolated reference image as the initial high-resolution image.

4.1 Purely Synthetic Data

We select a single high-resolution image from a real video sequence that consists of a rich scene with man-made rigid objects, text, and natural textures. Three synthetic motions are used to create three more frames: $[-0.5, -0.5]$, $[0.5, 1]$, and $[1, 0.5]$. In the following experiments, statistics are computed from all the frames.

For example, the flow error statistics are based on all flows from the reference image to the other images. Note that although a parametric model is chosen for the synthesized motions, for the experiments a non-parametric optical field is computed to assess the various aspects of flow accuracy.

Comparison of Flow Estimation. We compare three different methods for flow computation: 1) Least-squares based flow [15] (LSQ flow), 2) consistent flow (CONS flow), and 3) bundled flow with CONS flow as the initial input (CB flow).

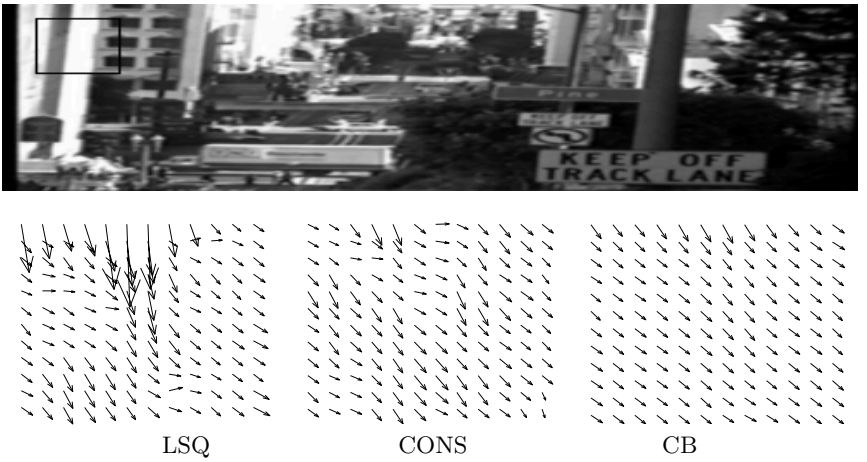

Fig. 1. Top row: the reference image and the selected region of interest (ROI). Bottom row: a close-up look at the computed flows in the selected ROI.

Fig. 1 shows the three flow fields for a region-of-interest (ROI) highlighted in one frame at the top. The qualitative nature of the flow and the relative errors are clear from the pictures. In addition, we compute quantitative errors in the flow in two different ways. If we define the *error flow* as the difference vector between the perfect flow and the computed flow, then the histograms of the two directional components, and the histograms of flow magnitude and angle can be used. To represent the error flow using just one measurement, we define the signed magnitude which is the product of its magnitude and the tangent of its direction. A similar quantity has been used to represent the probability distribution of the flow error in [16].

Figure 2 plots the two histograms, one for x-direction flow and the other for the signed magnitude of flow, for the three different methods. Furthermore, we also compute a measure of flow inconsistency, called *reprojection error flow*, as the difference between the forward and backward flows at corresponding points. That is, $\mathbf{e12}(\mathbf{p}) = \mathbf{p} - \mathbf{u21}(\mathbf{p} + \mathbf{u12}(\mathbf{p}))$, where $\mathbf{e12}, \mathbf{u12}, \mathbf{u21}$ are the reprojection error flow and the flow fields for frames 1-2 and 2-1, respectively. Table 1

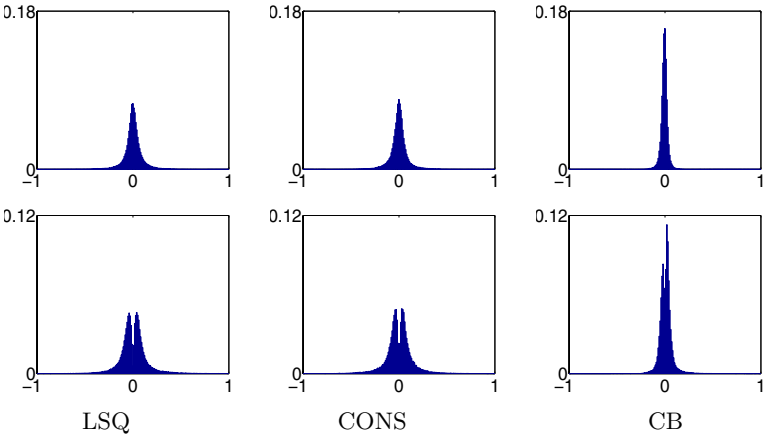

Fig. 2. Comparison of flows computed from synthetic low-resolution images. The top row plots the histogram of flow error in x-direction while the bottom row plots the histogram of the signed magnitude of error flows.

compares the peaks and standard deviations of the histograms of reprojection error flow for the three methods.

Table 1. Histogram peak (left) and standard deviation (right) of error flows corresponding to Fig. 2

	LSQ	CONS	CB
x-dir	.0748/.1685	.0795/.1163	.1601/.0417
signed mag	.0464/.8272	.0493/.1580	.1130/.0769
reproj. signed mag	.0308/.7093	none	small

Gradient-based Flow Error vs. Warping Error. We have computed the statistics for flow error and corresponding errors in image warping based on the gradient of the images. First, regardless of gradient information, if flow error is small, e.g., less than half pixel, then the warping error at the same location has a well-behaved Gaussian-shape distribution. Second, when gradient information is considered, relatively large flow error due to small gradients does not cause large warping error as pointed out in Section 3. This behavior is shown in Table 2 for the three methods using data from regions of three different gradient types..

4.2 Semi-synthetic Data

The second set of results are on a 9-frame video sequence captured with a DV (digital video) camcorder inside an office. There is parallax in the sequence so

Table 2. Standard deviation of flow (left) versus warping (right) errors based on gradient information.

	Low gradient	Medium gradient	High gradient
LSQ	0.3235/0.5802	0.1930/1.9736	0.1686/ 0.7202
CONS	0.1254/0.4836	0.1292/2.0864	0.1602/1.3643

global motion models do not suffice. Dense motion estimation algorithms that incorporate rigidity constraints could have been employed to process this sequence. However, since one of the goals of this paper is to demonstrate the efficacy of consistent dense motion estimation, for uniformity of results, we used the proposed consistent flow estimation algorithm without the rigidity constraints.

Super-resolution Error. We have experimented with the super-resolution algorithm with the three flows computed earlier plus the true flow. In Fig. 3, we plot the projection error (Eq. 18) as a function of the number of iterations for an ROI highlighted in the original frame on the top left in Fig. 4. For comparison purposes, the reconstructed images are also shown in Fig. 3. Note that errors in non-consistent flow tend to increase the reconstruction error as is evident both in the plot as well as the reconstructed image. The error becomes larger with more iterations. However, consistent flow does not suffer from this problem.

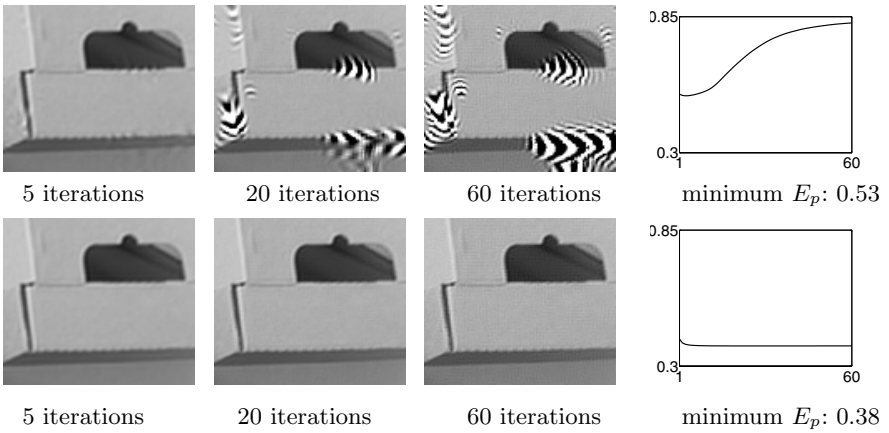

Fig. 3. Comparison of reconstruction results. An ROI is selected (as shown in the top frame in in Fig. 4) and its super-resolved image is zoomed up for display. The figures are arranged in a way that rows from one to two correspond to LSQ flow, and CONS flow, respectively. The first column plots the reconstructed image at iteration 5 while images in the second column are results at iteration 60. The last column plots projection error curves of LSQ and CONS flow for 60 iterations.

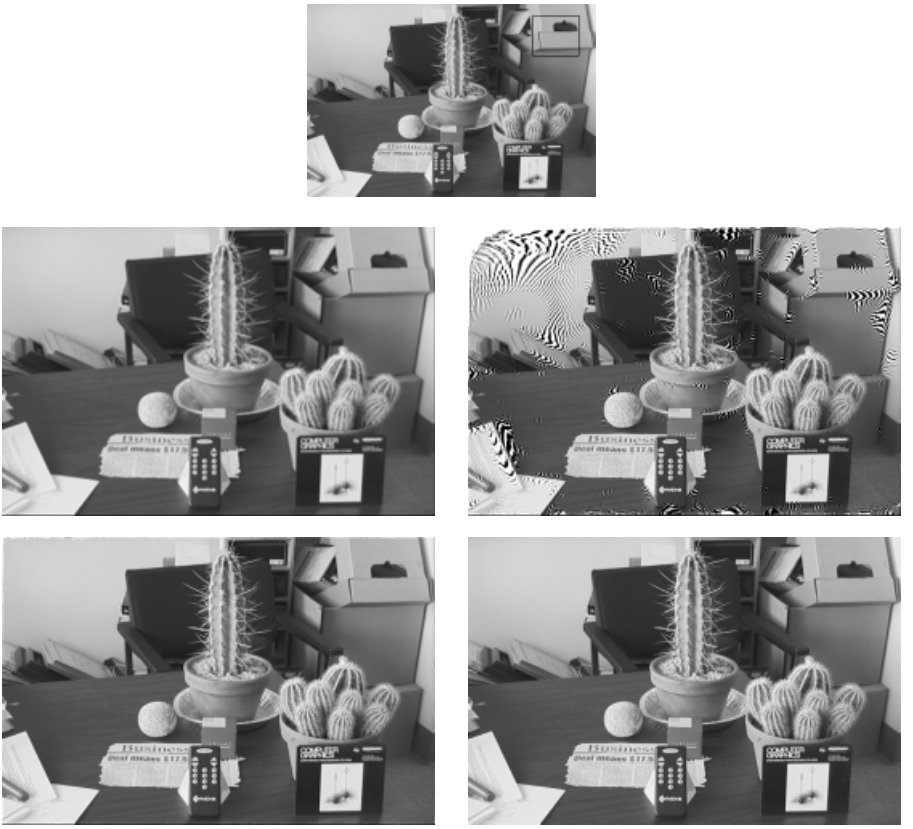

Fig. 4. Reconstruction results using one reference images and eight other images. The top row shows a low-resolution reference image with an ROI marked. Figures plotted in the second row are bicubic interpolation result and LSQ flow based super-resolved image. Finally, figures in the last row are CONS flow based super-resolved image and the true high-resolution image.

4.3 Real Video Sequences

The final results are with one sequence for which there is no ground truth. So the result is for visual quality assessment only. Again, we emphasize that no 2D/3D global motion model is valid for this sequence since the bee and the flower are swaying in the wind while the camera is moving. Fig. 5 shows one original frame and two super-resolved frames, one with LSQ flow and another with CONS flow. The superior quality in terms of sharpness and detail for the latter is evident.

5 Conclusions

We have studied the feasibility of reconstruction-based super-resolution with respect to errors in image alignment. An analysis of errors in optical flow indicates

Low-resolution image

LSQ flow based super-resolved image

CONS flow based super-resolved image

Fig. 5. Comparison of super-resolved images from high-quality DV.

that optical flow based super-resolution may be feasible in the small noise case since warping errors are well-behaved. However, even in the small noise case, flow accuracy is important. We introduced the concept of flow consistency and showed both quantitatively and qualitatively that flow consistency is critical for super-resolution. In the context of flow consistency, we presented algorithms that enforce consistency and demonstrated their efficacy. We plan to test these ideas with other reconstruction based super-resolution algorithms. We are also developing fully consistent bundled flow fields.

References

[1] Irani, M. and Peleg, S. 1993. Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency. *Journal of Visual Comm. and Image Repre.,* Vol. 4, pp. 324-335.

[2] Elad, M. and Feuer, A. 1997. Restoration of a single superresolution image form several blurred, noisy and undersampled measured images. *IEEE Trans. on Image Processing,* pp. 1646–1658.

[3] Patti, A., Sezan, M., and Tekalp, M. 1997. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Trans. on Image Processing,* pp. 1064–1078.

[4] Bascle, B., Blake, A., and Zisserman, A. 1996. Motion deblurring and super-resolution from an image sequence. In *Proc. European Conf. Comp. Vision,* pp 573–581.

[5] Schultz, R.R. and Stevenson, R.L. 1996. Extraction of high resolution frames from video sequences. *IEEE Trans. on Image Processing,* pp. 996–1011.

[6] Hardie, R., Barrard, K., and Armstrong E. 1997. Joint MAP Registration and High-resolution Image Estimation Using a Sequence of Undersampled Images. *IEEE Trans. on Image Processing,* pp. 1621-1633.
Multi-resolution

[7] Marziliano, P. and Vetterlli, M. 1999. Reconstruction of Irregularly Sampled Discrete-Time Bandlimited Signals. *IEEE Trans. on Signal Processing,* pp. 3462-3471.

[8] Shekarforoush, H. and Chellappa, R. 1999. Data-driven multi-channel super-resolution with application to video sequences. *Journal of the Optical Society of America A,* pp. 481–492.

[9] Tsai, R.Y. and Huang, T.S. 1984. Multi-frame Image Restoration and Registration.*Advances in Computer Vision and Image Processing,* JAI Press Inc.

[10] Baker, S. and Kanade, T. 1999. Super-Resolution Optical Flow. CMU-RI-TR-9936.

[11] Baker, S. and Kanade, T. 2000. Limits on Super-Resolution and How to Break Them. In *Proc. Conf. Comp. Vision and Patt. Recog.*

[12] Freeman, W. and Pasztor, E. 1999. Learning low-level vision. In *Proc. Int. Conf. Comp. Vision.*

[13] Bonet, J.S.D. 1997. sampling procedure for analysis and synthesis of texture images. In *Proceedings of SIGGRAPH,* pp. 361-368.

[14] B. D. Lucas and T. Kanade. 1981. An iterative image registration technique with an application to stereo vision. In *Proc. 7th Int. Joint Conf. on Art. Intell.*.

[15] Bergen, J., Anandan, P., Hanna, K., and Hingorani, R. 1992. Hierarchical Model-Based Motion Estimation. In *Proc. European Conf. Comp. Vision,* pp. 237-252.

[16] Simoncelli, E.P. and Adelson, E.H. 1990. Computing Optical Flow Distribution Using Spatio-Temporal Filters. MIT Media Lab Technical Report 165.

[17] Sha, N. R. and Zakhor, A. 1999. Resolution Enhancement of Color Video Sequences. In *IEEE Trans. on IP*, Vol. 8, No. 6, June, 1999, pp. 879–885.

[18] Oppenheim, A. V. and Schafer, R. W. Discrete-Time Signal Processing. Prentice Hall, Englewood Cliffs, NJ, USA, 1989.

[19] Birchfield, S. Derivation of Kanade-Lucas-Tomasi Tracking Equation. Unpublished, May 1996. *http://vision.stanford.edu/ birch/klt/.*