Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 58195, Pages 1–12
DOI 10.1155/ASP/2006/58195

Super-Resolution for Synthetic Zooming

Xin Li

Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown WV 26506-6109, USA

Received 1 December 2004; Revised 3 March 2005; Accepted 4 March 2005

Optical zooming is an important feature of imaging systems. In this paper, we investigate a low-cost signal processing alternative to optical zooming—synthetic zooming by super-resolution (SR) techniques. Synthetic zooming is achieved by registering a sequence of low-resolution (LR) images acquired at varying focal lengths and reconstructing the SR image at a larger focal length or increased spatial resolution. Under the assumptions of constant scene depth and zooming speed, we argue that the motion trajectories of all physical points are related to each other by a unique vanishing point and present a robust technique for estimating its 3D coordinate. Such a line-geometry-based registration is the foundation of SR for synthetic zooming. We address the issue of data inconsistency arising from the varying focal length of optical lens during the zooming process. To overcome the difficulty of data inconsistency, we propose a two-stage Delaunay-triangulation-based interpolation for fusing the LR image data. We also present a PDE-based nonlinear deblurring to accommodate the blindness and variation of sensor point spread functions. Simulation results with real-world images have verified the effectiveness of the proposed SR techniques for synthetic zooming.

1. INTRODUCTION

Image resolution is a critical factor affecting the quality of image and video. To increase the spatial resolution, we can increase the sampling density or the focal length of CCD sensors [1]. However, such physics-based approaches are limited by the cost of manufacturing high-precision or power-zoom optics. Recently, signal-processing-based alternatives, that is, to reconstruct a high-resolution (HR) image from multiple low-resolution (LR) images, have attracted much attention. Super-resolution (SR) techniques exploit the fundamental tradeoff between space and time—the multiple LR images are acquired separately in time but fused together to enhance the resolution in space.

The literature of SR image reconstruction is large. Interested readers are referred to several recent reviewing articles [2–4]. As our understanding of SR improves, new techniques with relaxed assumptions about motion model [5] and point spread function (PSF) [6] appear. The performance of SR algorithms has also improved in terms of robustness [7] and efficiency [6]. A natural generalization of SR, which deals with resolution enhancement of video (also-called dynamic SR in [3]), has been widely studied for both uncompressed [8, 9] and compressed [10, 11] video.

In this paper, we investigate the problem of SR for synthetic zooming. The zoom capability of optical lens, determined by the range of focal length, is a primary factor affecting the manufacturing cost. Synthetic zooming is an alternative low-cost approach of enhancing the imaging system's zoom capability by processing images acquired within the range of focal length of existing imaging systems. Synthetic zooming has various important applications from law enforcement to geological surveillance. One salient feature of synthetic zooming is the relaxed assumptions about the acquisition process. For example, in some scenarios such as video surveillance, security camera is mounted at some place and cannot easily perform panning or rotation due to mechanical constraints, but is still able to zoom in. Another example is when an object of interest moves in front of a camera, object motion often causes scene depth changes, which will generate zoom-like motion into the images (variation of scene depth is geometrically equivalent to the change of focal length).

Unlike translational or rotational motion assumed by most existing SR techniques, zoom motion [12] has been studied much less in the literature of motion estimation. Under the assumption that the region of interest has the same scene depth (e.g., a flat surface parallel to the imaging plane) and the zooming speed is approximately constant, we argue that all images are linked by a simple line-geometric model—the projections of any point in the physical scene at different focal lengths lie along a ray, and the rays corresponding to different physical points intersect at a unique point called "vanishing point" (VP). Such observation motivates us to

solve the registration problem for synthetic zooming by estimating the 3D coordinate of VP. We present a spatio-temporal analysis technique for tracking feature points along the rays and a robust nonlinear optimization approach for locating their intersection (VP). We also provide arguments about the robustness of such VP-based registration based on a rigorous analysis of error bounds in the Euclidean space.

Another issue at the heart of SR techniques is the sensor PSF. The PSF of an optical system is affected by a number of factors such as focal length, object distance, distance of the point from the center, and so on [13]. One particular challenge with SR for synthetic zooming is that the sensor PSF is not only unknown but also varying along the temporal axis. Such observation gives rise to the issue of *data consistency* in SR image reconstruction; that is, if LR image data correspond to different PSFs, how do we fuse them together? We propose to divide the collection of LR frames into consecutive groups (to alleviate inconsistency) and employ Delaunay-triangulation (DT)-based interpolation [14] to fuse the data for each group separately. The interpolated images from each group are then linearly merged into the final result. We use experimental results to support the effectiveness of such idea.

It should be noted that due to the nonlinearity behind the zooming process (VP-related geometry and varying PSF), it is extremely difficult to relate the LR and HR images by a linear system such as the warping matrix used in [4, equation (1)]. Therefore, many effective regularized SR image reconstruction techniques, such as maximum a posteriori (MAP) [9] and projection-onto-convex-set (POCS), [8] are not applicable. Instead, we advocate a PDE-based nonlinear deblurring approach that originated from shock filters [15]. Our model is spiritually similar to [16] except the use of a mean curvature diffusion flow term [17] in order to effectively suppress the artifacts associated with interpolation errors.

The rest of this paper is organized as follows. Section 2 introduces SR for synthetic zooming and emphasizes the role of line geometry for image registration and the issue of PSF for image deblurring. Section 3 describes a robust technique for estimating VP via tracing the rays of feature points in the spatio-temporal space and provides theoretical analysis of error bounds. Section 4 covers the DT-based interpolation and PDE-based nonlinear deblurring for synthetic zooming. Simulation results with real-world image sequences are reported in Section 5. We make several concluding remarks in Section 6.

2. PROBLEM STATEMENT

The problem of synthetic zooming can be formulated as follows. It is well known that the capability of optical zooming is determined by the range of focal length of optical lens. For example, the lens with a focal length of 35–105 mm have the power of 3 × zoom. Suppose $\{I_k\}_{k=1}^{K}$ are the K consecutive frames captured by a video camera as we zoom in (or equivalently, as the object moves closer to the camera). Without loss of generality, we can assume that I_1 and I_K are the furthest and closest shot of the scene.

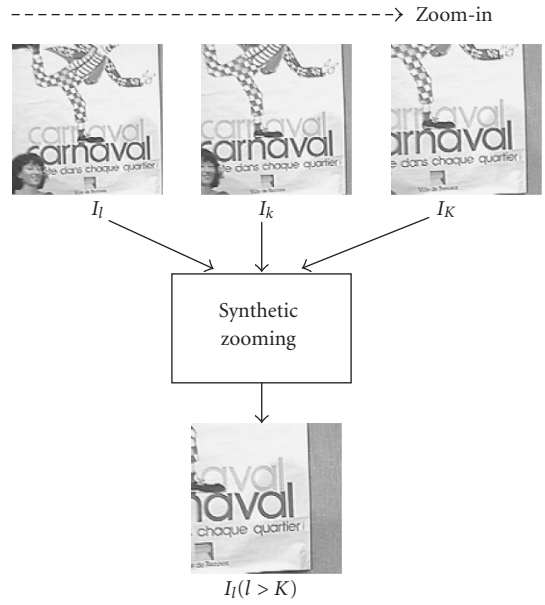


Figure 1: Synthetic zooming problem.

Synthetic zooming refers to generating a new image I_l ($l > K$) from $\{I_k\}_{k=1}^{K}$ as if it were captured at a larger focal length, as shown in Figure 1. Note that although the sampling density of I_l is the same as that of I_k, the actual pixel size is reduced due to the zoom-in along the temporal axis (we call it "temporal mode" SR, refer to Figure 2(a)). Along with the increased focal length, we can also artificially increase the spatial sampling density (we call it "spatial mode" SR, refer to Figure 2(b)). In practice, the sensor PSFs in temporal and spatial modes would differ even if they are tuned to exactly reach the same pixel size (resolution). However, under the context of synthetic zooming, we do not need to distinguish temporal mode and spatial mode because they will be shown equivalent from a geometric perspective in the next section.

To make the problem of synthetic zooming tractable, we need to make the following assumptions.

(1) Constant scene depth—that is, we assume that the object of interest has a flat surface parallel to the imaging plane. Just like translational motion models adopted by most existing SR techniques, such assumption is for the reason of simplifying the analysis. Otherwise, scene depth discontinuities need to be taken into account, which dramatically increases the complexity of motion estimation. We believe that studying such simplified case is the first step to solve SR for synthetic zooming in more general cases where motion segmentation is required.

(2) Constant zooming speed—under such assumption, motion trajectories of any physical point would arguably vanish to a single point, called "vanishing point" (VP) as shown in Figure 3. VP can be viewed as the limiting case of zooming as the focal length goes to zero or the scene depth goes to infinity. Such property greatly simplifies the geometric modeling of zoom images, which serves as the basis for our SR image reconstruction techniques. In practice,

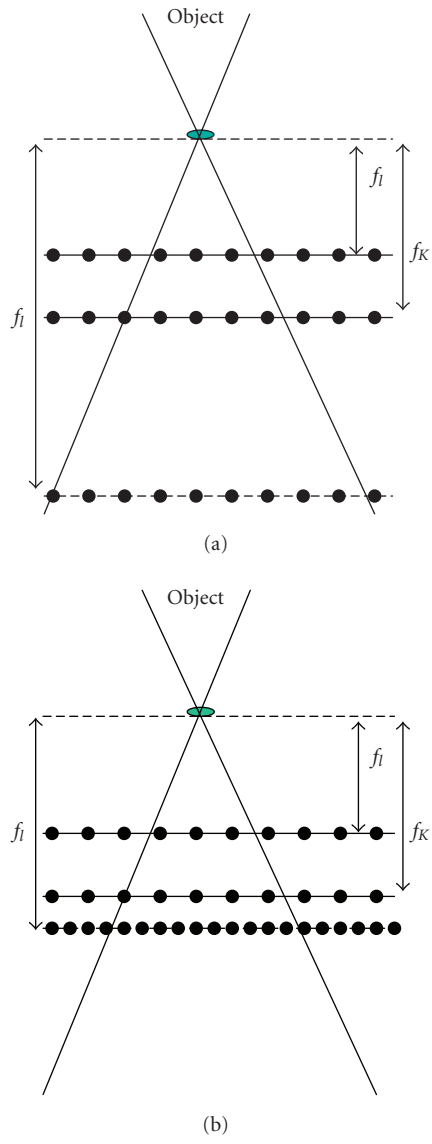


Figure 2: (a) Temporal mode. (b) Spatial mode. Note that temporal mode requires a larger focal length to reach the same pixel size (resolution) as spatial mode.

the speed of optical zooming determined by the lens mechanics or object motion is approximately constant at least for a short time interval, which justifies the validity of such assumption.

(3) Small K values—due to either limited zoom capability of the lens or limited time interval T (usually a fraction of a second) available for acquiring the object of interest. Such assumption is made to both reflect practical constraints (e.g., a fast moving car passes the camera quickly) and simplify the problem (as we will discuss in detail next). When video is captured at 30 fps, we might assume that only a dozen or so frames ($K = 12$, less than 0.5 second) are available for SR image reconstruction.

The basic idea behind SR for synthetic zooming is illustrated (see Figure 3). In the continuous space, the projections

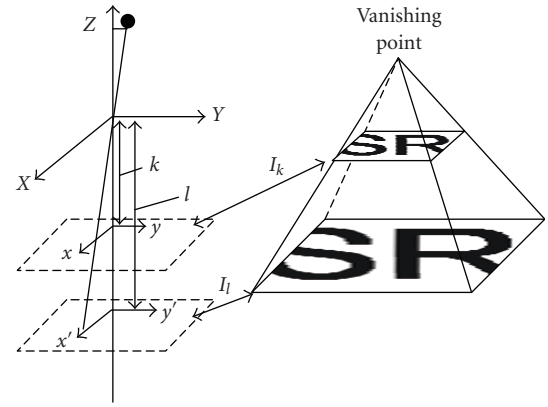


Figure 3: Geometry of zoom images and illustration of vanishing point.

of a physical point at (X, Y, Z) at two different frames I_k and I_l are linked by a simple geometric relationship as follows:

$$\frac{x}{X} = \frac{k}{Z}, \qquad \frac{y}{Y} = \frac{k}{Z},$$
$$\frac{x'}{X} = \frac{l}{Z}, \qquad \frac{y'}{Y} = \frac{l}{Z}. \tag{1}$$

Here (x, y) and (x', y') are the coordinates of the projected point in frame k and l, respectively. As Z goes to infinity at a constant speed, all projections will arguably converge to a single point—the VP $= (x_0, y_0, z_0)$. The zooming ratio of frame I_k is then defined by

$$a = \frac{|z_0| + k}{|z_0|}. \tag{2}$$

Note that the zooming ratio is opted to be normalized at the origin of temporal axis (frame number), though we do not have an image at $k = 0$.

If we ignore the dissipation of light irradiance crossing a short distance in the space, the following identity would hold:

$$\iint_S R(x, y)dxdy = \iint_{S'} R(x', y')dx'dy', \tag{3}$$

where R denotes the incidence spectral irradiance (*watts/unit area*) and S, S' are the projected areas of the same physical point onto I_k, I_l, respectively. After discrete sampling, the pixel intensity value at a specific location $(i, j) \in [1, H] \times [1, W]$ is given by

$$I_k(i, j) = Q\left[\iint_A h_k(i - x', j - y')R(x', y')dx'dy'\right]. \tag{4}$$

Here $A = [(i - 1/2)d, (i + 1/2)d] \times [(j - 1/2)d, (j + 1/2)d]$ and $Q[\cdot]$ is the nonlinear quantization operator. We note that $h_k(x, y)$, the PSF of optical lens while acquiring I_k could change along with the variation of focal length f_k [13]. We also remark that the sampling distance d is related to the actual pixel size D by

$$\frac{d}{D} = \frac{f_k}{Z}. \tag{5}$$

Therefore, a larger focal length f_k gives rise to a smaller D, which implies a reduced pixel size or imaging at a higher resolution. Synthetic zooming essentially reconstructs an image with a large focal length from the given K images with smaller focal lengths.

The above geometric and radiometric analysis of optical zooming raises a fundamentally new challenge with SR for synthetic zooming. In translational motion, it is often convenient to explicitly write out the forward imaging process by a linear system (e.g., the warping matrix W in [4, equation (2)]). However, such linear system modeling would become highly cumbersome and inaccurate for zoom motion due to the following two reasons. First, the fractional part of the displacement among LR images is plausibly uniformly distributed over $(0, 1)$ due to the modulation with scene depth Z. Second, the PSF of LR images varies along the temporal axis due to its dependency on f_k. Without linear system models, many well-established SR image reconstruction techniques, such as MAP [9] and POCS [8], are not applicable any more. The intrinsic nonlinearity caused by VP-related geometry and time-varying PSF distinguishes SR for synthetic zooming from other SR techniques.

3. ESTIMATING VANISHING POINT OF MOTION TRAJECTORIES

Image registration [18] or motion estimation (ME) [19] was traditionally formulated as a problem with two frames only. Such formulation enjoys conceptual simplicity and matches the time-varying characteristics of motion in the real world. However, when the motion model is constrained (e.g., constant-speed zooming), it is advantageous to employ multiple frames during ME for the reason of robustness as well as computational efficiency. In this section, we present a multi-frame ME technique for zoom images based on tracking a selected group of feature points in the spatio-temporal domain.

3.1. Multi-frame motion estimation for zooming

Theoretically it is sufficient to use just two intersecting straight lines (motion trajectories) to locate VP. However, due to discrete sampling as well as potential errors with matching feature points, it is wise to employ more than two lines. Since any line in 3D Euclidean space is determined by two points, we need to identify a group of matched feature points $\{(x_n^1, y_n^1, z_n^1), (x_n^2, y_n^2, z_n^2)\}_{n=1}^N$.

To locate N pair of points, we propose to select a group of feature points from the last frame ($z_n^2 = K$) and trace them back to the first frame ($z_n^1 = 1$). The selection of good feature points for object tracking have been widely studied in the literature of computer vision (e.g., [20–22]). It is often suggested that points with large local variance, indicating strong texturedness or cornerness, are good candidates for tracking. We adopt this idea in this paper but additionally put a constraint that the selected feature points be spatially uniformly distributed. Such requirement is useful to avoid the potential bias in the estimation of VP.

Specifically, the input image (I_K) is filtered by the difference-of-Gaussian operator in the scale space [20]. Then the candidates of feature points are taken to be the local maximum in the filtered image. Unlike [20], we do not resample the image at different scales because we have found that a handful of feature points is sufficient for estimating VP. One way of enforcing the uniformly-distributed constraint is to structure the input image into nonoverlapping blocks and pick out at most one feature point from a block if the variance of that block is above a chosen threshold.

To track the location of selected feature points in the first frame, we propose an exhaustive search strategy within a local window based on the following distortion criterion:

$$\begin{aligned}\mathrm{SAD}_{\mathrm{avg}}(x,y) &= \frac{1}{K-1} \\ &\times \sum_{k=2}^{K}\sum_{i=-T}^{T}\sum_{j=-T}^{T} \left| I_k\left([i+x_n^1+(k-1)\delta_x], [j+y_n^1+(k-1)\delta_y]\right) \right. \\ &\qquad \left. - I_1\left(i+x_n^1, j+y_n^1\right) \right|, \end{aligned} \tag{6}$$

where $\delta_x = (x - x_n^1)/(K-1)$, $\delta_y = (y - y_n^1)/(K-1)$, and $[\cdot]$ denotes the rounding operator. The best matching result is therefore given by

$$(x_n^2, y_n^2) = \min_{x,y} \mathrm{SAD}_{\mathrm{avg}}(x,y). \tag{7}$$

The above search strategy is a straightforward extension of sum-of-absolute-distance-(SAD) based tracking [23] from two-frame into multi-frame.

We note that assumptions (1) and (3) are critical to the success of the above matching procedure. Assumption (1) tells us that no occlusion occurs except around the image border. Since I_K is the frame closest to the camera, we are guaranteed that the selected feature points from the last frame will not go outside the image border while tracking. Due to assumption (3), we do not need to use sophisticated affine-invariant distortion measures (e.g., [24]). However, just like any feature-based vision algorithms, completely accurate matching results are difficult to obtain. Instead, we resort to robust statistical tools such as outlier rejection to eliminate incorrectly matched pairs.

With a collection of straight lines $L_n = \{\vec{p}_n = [x_n^1, y_n^1, z_n^1], \vec{q}_n = [x_n^2, y_n^2, z_n^2]\}$, the initial estimation of VP $\vec{r} = (x_0, y_0, z_0)$ can be found by solving the following nonlinear optimization problem:

$$\min J(\vec{r}) = \frac{1}{N}\sum_{n=1}^{N} d(\vec{r}, L_n)^2, \tag{8}$$

where $d(\vec{r}, L_n)$ is the point-to-line distance given by

$$d(\vec{r}, L_n) = \frac{\left|(\vec{q}-\vec{p})\times(\vec{p}-\vec{r})\right|}{\left|\vec{q}-\vec{p}\right|}. \tag{9}$$

There exist numerous techniques for solving the above nonlinear optimization problem. For example, we adopt the

Nelder-Mead simplex method [25] whose implementation is directly available from Matlab optimization toolbox. With the initial estimate $\vec{r}^{(0)}$, we can obtain the distance profile $d(\vec{r}^{(0)}, L_n)$. If its maximum is larger than a preselected threshold (outlier detection), we reject the pair of feature points above the threshold and update the VP estimation with the remaining $N' < N$ pairs. Such rejection procedure is continued until no more outlier is found.

3.2. Error analysis and implications

As mentioned above, discrete sampling and feature point matching both contribute to the potential errors in VP estimation. Assuming a small disturbance with the feature points, we can derive an upper bound on the distance metric of (9). We summarize our analysis into the following lemmas and their derivations can be found in the appendices.

Lemma 1. *Given three consecutive points $\vec{r}$, $\vec{p}$, $\vec{q}$ along a line whose z-coordinates are $z_0 = Z < 0$, $z_1 = 0$, $z_2 = K > 0$, respectively. If the middle point is perturbed by $\delta\vec{p} = (\delta_x, \delta_y, 0)$, then the point-to-line distance ($\vec{r}$-to-$\{\vec{p}, \vec{q}\}$) is bounded by $d \leq (|Z|\sqrt{\delta_x^2 + \delta_y^2})/K$.*

Such theoretical bound is useful to check the validity of our estimation. If the minimal distortion achieved by nonlinear optimization is above this bound, the registration simply fails due to either too few feature points or invalid assumptions.

Similar error analysis can also be applied to bound the potential derivation of any point from its ground truth in a synthetic frame I_l ($z_2 = l$) caused by the errors in VP $= (x_0, y_0, z_0)$, $z_0 = Z < 0$. Suppose the point of interest is located at $\vec{p} = (x_1, y_1, z_1)$, $z_1 = k$ ($1 \leq k \leq K$), then its projected location at $z_2 = l > k$ is

$$x_2 = x_1 + (x_0 - x_1)\Delta, \qquad y_2 = y_1 + (y_0 - y_1)\Delta, \tag{10}$$

where $\Delta = (z_2 - z_1)/(z_0 - z_1)$ is the scaled sampling distance. From (10), we can see the equivalence between temporal mode SR and spatial mode SR. For example, spatially doubling the sampling density (i.e., integer-pel to half-pel) is equivalent to temporally increasing the value of z_2 to $2z_2 - z_1$ (doubling the zoom ratio). Such geometric duality between temporal and spatial mode can also be intuitively justified with respect to Figure 2. The following lemma relates the error bound of a projected location $\vec{q} = (x_2, y_2, z_2)$ to that of VP.

Lemma 2. *Suppose the VP is perturbed by $\delta\vec{r} = (\delta_x, \delta_y, \delta_z)$, then the projected location given by* (10) *will be disturbed by at most*

$$\begin{aligned} |x_2' - x_2| &\approx \left|\frac{\delta_x(l-k)}{Z-k}\right|, \\ |y_2' - y_2| &\approx \left|\frac{\delta_y(l-k)}{Z-k}\right|. \end{aligned} \tag{11}$$

Lemmas 1 and 2 together provide us guidance about the tradeoff between space (pixel location) and time (zooming speed) in SR for synthetic zooming. For example, the slower the camera zooms in, the larger is $|z_0| = -Z$ (the slope of motion trajectories decreases). According to Lemma 1, a larger $|Z|$ has the potential of increasing the errors with VP estimation; meantime, the pixel location in a synthesized frame I_l becomes more robust to the errors in VP based on Lemma 2. Conversely, a fast zooming is beneficial to reduce the estimation errors of VP, which is compensated by the increased sensitivity of pixel location to the errors of VP. Although it is easy to see that a large K is always preferred from the geometric perspective, we note that collecting a large number of LR frames is not always feasible in practice, as we argued while presenting assumption (3). Additionally, when K increases, the variation of sensor PSF becomes more serious, as we will detail next.

4. SR IMAGE RECONSTRUCTION FOR SYNTHETIC ZOOMING

With the estimated VP, we can map any pixel in I_k to a location in the synthetic frame I_l by (10). However, the mapped locations could overlap with others and often do not exactly align with the target HR grid points, which calls for the need of *interpolation*. Even if the alignment occurs by coincidence, the pixel intensity does not correspond to the desirable but blurred version of HR image, which calls for the need of *deblurring*. In this section, we study the implications of PSF variation on interpolation and deblurring in SR for synthetic zooming.

4.1. Two-stage interpolation via Delaunay triangulation

In most existing SR techniques with translational motion, registered data points might not exactly align with the HR grid points but still have uniform density in the spatial domain. In zoom motion, the spatial density of projected data points is nonuniform and often irregular due to the modulation with Z (see Figure 4). For such kind of data, DT has been shown to be an effective interpolation technique [14]. The basic idea behind DT is to use triangular patches to locally fit the available data. Due to the simplicity and convexity of triangles, DT-based interpolation enjoys low complexity and suitability for hardware implementation [14].

Due to the varying PSF of zoom lens, one potential risk with DT-based interpolation is data inconsistency, especially for two distant frames (e.g., I_1 and I_K). If the LR images are all obtained via an artificial warping operator (blurring followed by downsampling) of a HR image, we will not have the problem of data consistency. However, such artificial operator often does not faithfully reflect the imaging system in practice due to the ignorance of various factors (e.g., lighting variation, nonuniform motion, varying scene depth, etc.). Synthetic zooming represents a scenario where data inconsistency is easy to observe due to the variation of focal length (and therefore PSF). The consequence of ignoring data inconsistency is annoying artifacts in the interpolated image (refer to Figure 10 and Section 5 for further illustrations).

Figure 4: Nonuniform and irregular sampling points while projecting LR pixels to I_l according to (10).

To overcome the difficulty with data inconsistency, we propose a two-stage interpolation scheme. In the first stage, we divide K LR frames into K_g consecutive groups and apply DT-based interpolation within each group to produce K_g copies of HR image I_l^c $(c = 1, \ldots, K_g)$. The basic criterion for group division is that the data points within each group cover the desired HR grid points as uniformly as possible. A rule of thumb is based on the target frame number $z_2 = l$ and the coordinate of VP $z_0 = -Z$, which jointly determines the scaled sampling distance Δ in (10). The smaller the change of the fractional part of Δ is, as k increases, the larger the group size needs to be chosen.

In the second stage, K_g copies of HR image are fused through a linear weighting strategy:

$$I_l(i,j) = \sum_{n=1}^{K_g} w_n(i,j) I_l^n(i,j), \tag{12}$$

and the weight is given by

$$w_n(i,j) = \frac{1_{\{I_n(i,j)>0\}}}{\sum_{n=1}^{K_g} 1_{\{I_n(i,j)>0\}}}, \tag{13}$$

where $1_{\{I_n(i,j)>0\}}$ is a binary-value function indicating the validity of the interpolated data. More sophisticated weighting strategy is possible—for example, $w_n(i,j)$ can be chosen to reflect the confidence about the interpolated value at a given location. The closer the projected data point to the desired HR grid point, the higher confidence can be set. For simplicity, we have only considered the ad hoc weighting strategies of (13) in the simulation so far.

4.2. PDE-based nonlinear deblurring

There are primarily two challenges with deblurring in SR for synthetic zooming. First, the blurring kernel is unknown and could vary as the focal length varies. Therefore, blindness and variation of PSF jointly increase the difficulty of deblurring. Second, interpolation errors need to be taken into account; otherwise, interpolation errors would easily get amplified and become annoying artifacts in reconstructed images. The above two challenges make it difficult to formulate image deblurring as an inverse problem under the traditional linear filtering framework. Instead, we advocate the approach of PDE-based nonlinear deblurring [15, 16].

One of the early pioneering works on PDE-based image deblurring is shock filter [15], which does not require the knowledge of PSF but assumes a piecewise smooth function for the target image. Later, the concept of shock filter is combined with anisotropic diffusion [26, 27] in [16] to achieve simultaneous directional filtering (for suppressing noise) and edge enhancement (for deblurring). There also exists forward-and-backward nonlinear diffusion for SR [28], in which forward and backward diffusion handle noise suppression and image deblurring, respectively.

The PDE model employed in this work is conceptually similar—that is, we want to deblur the HR image without blowing up the noise. However, the noise model here is different from that in previous works where additive white Gaussian noise is often assumed. The errors introduced by DT-based interpolation are not Gaussian and are often edge-dependent. This is because the triangle patch model become less effective around edge areas (in other words, more data points are needed for an edge to be accurately reconstructed than smooth regions). To suppress such signal-dependent noise during the deblurring process, we propose the following PDE:

$$\frac{\partial I}{\partial t} = \alpha F_{\text{diffusion}} + \beta F_{\text{shock}}, \tag{14}$$

where

$$F_{\text{diffusion}} = \frac{I_{xx}\left(1+I_y^2\right) - 2I_x I_y I_{xy} + I_{yy}\left(1+I_x^2\right)}{2\left(1+I_x^2+I_y^2\right)^{3/2}},$$
$$F_{\text{shock}} = \nabla g(I) \cdot \nabla I, \tag{15}$$

where $g(\cdot)$ is the smooth nonincreasing edge-stopping function as proposed in [29]. The anisotropic diffusion flow $F_{\text{diffusion}}$ in (15) is essentially the mean curvature diffusion (MCD) [17]. Since it is known that the MCD flow converges to the surface of minimal area, interpolation errors around edges, which have the tendency of increasing the surface area, are effectively suppressed. We have found that a similar type of diffusion, affine invariant anisotropic diffusion [30], gives similar results. The constants (α, β) are the relaxation parameters controlling the balance between forward (anisotropic diffusion) and backward (image deblurring) diffusion. Empirical studies show that $\alpha = \beta$ often gives good result for small zoom ratio, but $\alpha > \beta$ is preferred in the case of large zoom ratio where interpolation errors become more severe.

5. SIMULATION RESULTS

In this section, we report our simulation results with two popular video sequences: *tennis* (SIF resolution) and *mobile* (CIF resolution). Both sequences contain a segment of camera zooming and we crop out the background portions (sized 120×120) of 50 zoom frames as the test material. Moreover,

Figure 5: Sample LR frames used in our experiments. Left-to-right and top-down: $k = 1, 3, 5, 7, 9, 11$.

we intentionally reverse their order in the temporal domain to simulate the zoom-in operation (original sequences are zoom-out). Among the extracted 50 LR frames, we only employ the first 12 frames ($K = 12$) to test our SR algorithms (the rest are used as ground truth). A few sample frames are shown in Figure 5. Such preprocessing is for the purpose of validating assumptions (1) and (3). In particular, the object of interest in our SR for synthetic zooming is assumed to be the textual or texture information in the image.

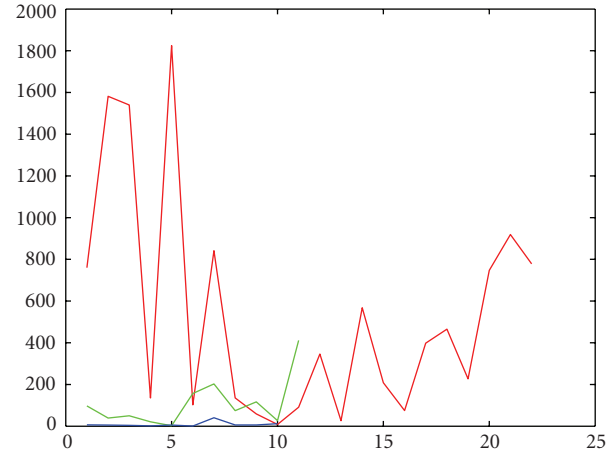
5.1. Vanishing point estimation

In our implementation of feature point extraction from the last frame, image is structured into nonoverlapping 10×10 blocks. Only when the standard deviation of a block is above a chosen threshold, it is eligible for producing a feature point. For each eligible block, its local maximum is marked to be the feature point and the corresponding feature point in the first frame is searched with a template sized by 7×7 ($T = 3$).

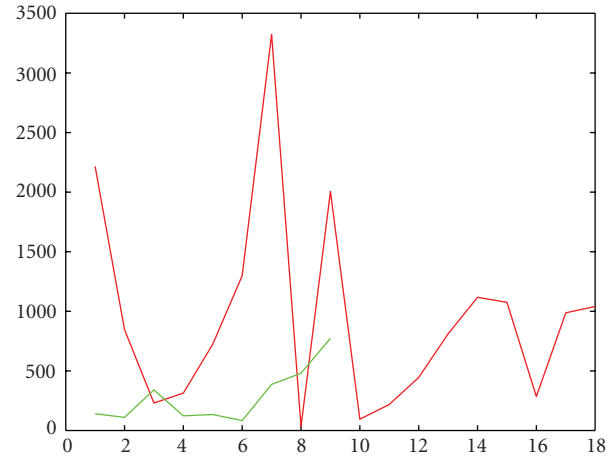
(a)

(b)

FIGURE 6: Initially extracted feature points for the last frames of *tennis* (a) and *mobile* (b) sequences.



(a)



(b)

FIGURE 7: The evolution of distance profile in (9) for *tennis* (a) and *mobile* (b) during iterative estimation of VP (red, green, and blue colors denote 1st, 2nd, and 3rd iterations, resp.). Note how the outliers get rejected (N decreases) as the iteration proceeds.

Figure 6 shows the initial extracted 22 and 18 pairs of feature points for *tennis* and *mobile* image sequences, respectively. We note that no special effort is devoted to eliminate the mismatched pairs. The outliers will be automatically rejected as we iteratively refine the estimate of VP (refer to Figure 7).

After iteratively eliminating the outliers, we find 10 and 9 pairs left for VP estimation, respectively (refer to Figure 8). Solving (8) gives the coordinates of VP: $(128, 173, -45)$ for *tennis* ($J_{\min} = 0.78$) and $(118, 340, -302)$ for *mobile* ($J_{\min} = 49.96$). The large registration error with *mobile* is not surprising because of its large $|z_0|$ value. We remark that such results should be interpreted properly due to the analysis given at the end of Section 3.2. A large $|z_0|$ value gives rise to large registration errors according to Lemma 1, but improves the robustness of pixel locations in synthetic zooming according to Lemma 2.

To see this clearly, we have designed the following simple experiment. Since the 12 frames are cropped out from the original sequence, we can perform synthetic zooming and compare it with the actual optical zooming result (ground truth). Furthermore, we can intentionally disturb the estimated (x_0, y_0, z_0) by additive white Gaussian noise and evaluate its impact on the synthesized images. Figure 9 shows the synthesized 20th frame (with and without disturbance) and the actual frames for both sequences. It can be observed that for *mobile* sequence, its large $|z_0|$ value leads to highly robust reconstruction. Though the noise in VP geometrically distorts the synthesized image, the visual quality does not appear to be affected. By contrast, due to a small $|z_0|$ value of *tennis* sequence, it becomes more sensitive to the noise in VP.

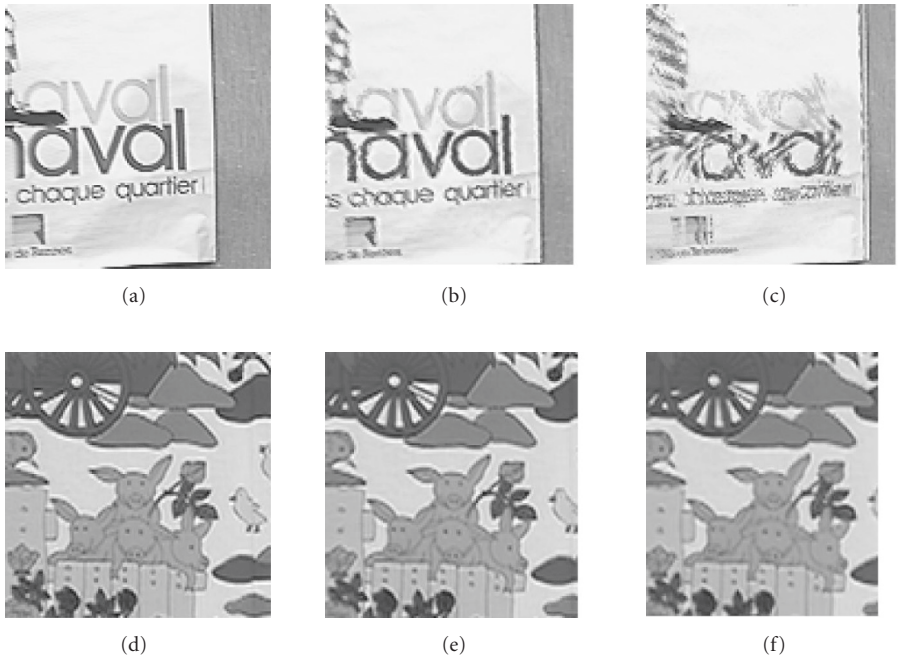
5.2. Super-resolution for synthetic zooming

The problem of data inconsistency is easier to observe with *tennis* than *mobile* due to its faster zooming speed. Figure 10 compares two synthesized frames of *tennis* in the temporal mode when $l = 85$ (zooming ratio is approximately three). One is to merge all 12 frames into one frame and apply DT-based interpolation; the other is to adopt the two-stage interpolation strategy proposed in Section 4.1 ($K_g = 4$). The inconsistency among LR frames causes annoying jittering-like artifacts around edges. By contrast, the

(a) (b)

(c) (d)

FIGURE 8: Feature points after outlier removal for the first (left) and last (right) frames of *tennis* (top) and *mobile* (bottom) sequences.



(a) (b) (c)

(d) (e) (f)

FIGURE 9: Comparison among the original 20th frame (left), synthesized without noise (middle) and synthesized with noise $N(0, 25)$. Top: *tennis*; bottom: *mobile*. Note that the white strip on the right border (pixel values undefined) is due to the fact that VP is located outside the right boundary of cropped portions.

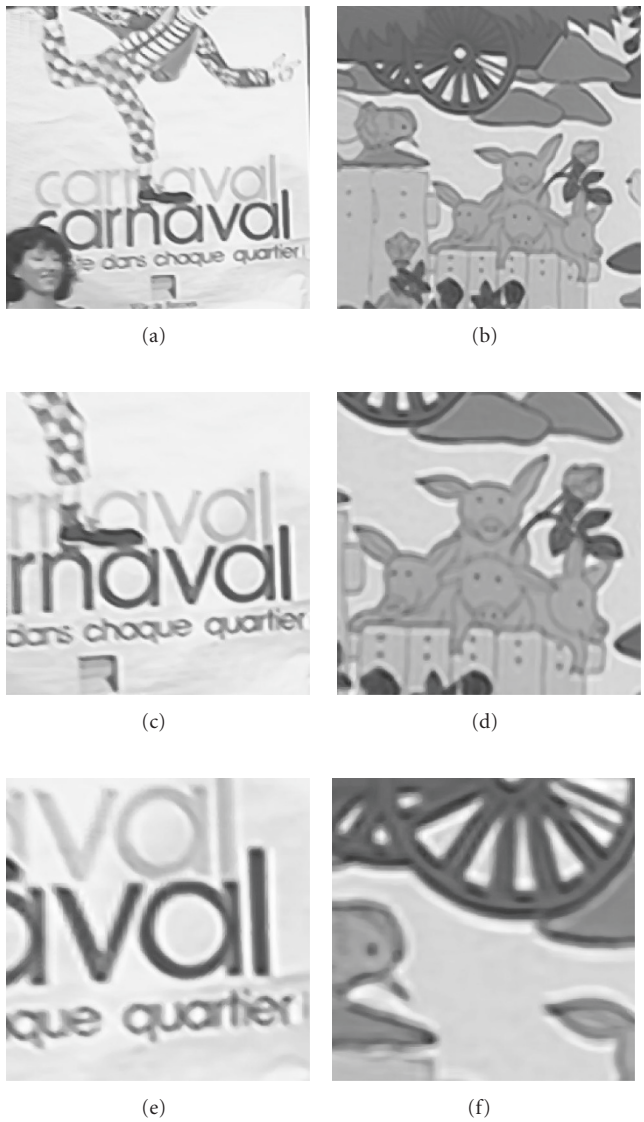
(a) (b)

Figure 10: Impact of data inconsistency on SR images. (a) interpolated image by fusing all 12 LR frames; (b) interpolated image by the proposed two-stage fusion strategy.

(a) (b)

(c) (d)

Figure 11: SR images before (left) and after (right) PDE-based nonlinear deblurring. Top: synthetic double-size 20th frame of *tennis*; bottom: synthetic double-size 20th frame of *mobile*.



(a) (b)

(c) (d)

(e) (f)

Figure 12: SR reconstructed images with different zooming ratios. Top: $l = |z_0|$; middle: $l = 2|z_0|$; bottom: $l = 4|z_0|$.

proposed two-stage interpolation strategy effectively alleviates the problem of data inconsistency.

In our implementation of PDE-based nonlinear deblurring, the relaxation parameters are chosen to be $\alpha = \beta = 0.125$. We stop the nonlinear diffusion after five iterations. The SR images before and after deblurring of synthesized 20th frame for two sequences are shown in Figure 11. Enhanced edge sharpness can be observed; though the deblurred images suffer from loss of naturalness. Such weakness with PDE-based approaches has been known—the limiting solution of shock filters tends to be a piecewise smooth function. However, we remark that in some applications such as license plate detection in law enforcement, the target object does satisfy the piecewise smooth condition, which supports the adoption of PDE-based nonlinear deblurring.

We also want to demonstrate the performance of the proposed SR algorithms at different zoom ratios. By setting target frames index l to be $|z_0|$, $2|z_0|$, $4|z_0|$, we obtain varying zooming ratios of 2, 3, 5 according to (2). Figure 12 compares the reconstructed SR images with different zoom ratios, respectively. Note that as the zoom ratio increases, field of view also increases. So we opt to only show portions of SR images (240×240) when the zooming ratio is more than 2. The visual quality of reconstructed SR images appear to be satisfactory.

Finally, we comment on the computational cost of the proposed SR algorithm. Feature extraction, outlier removal, and estimating VP do not appear to be computationally demanding. The bottleneck lies in the tracking of feature points across frames because of exhaustive search. On a 1.6 GHz Pentium-IV laptop, we have found that it takes about 15 seconds to track the extracted twenty or so feature points under Matlab (faster search algorithms can be used to reduce the computational cost). When the zoom ratio is set to be two, another 10 seconds are required for DT-based two-stage interpolation and 5 seconds for nonlinear deblurring.

6. CONCLUSIONS AND PERSPECTIVES

In this paper, we formulate the problem of SR for synthetic zooming—that is, to simulate optical zooming from a sequence of zoom images. We present a robust line-geometry-based algorithm for registering zoom images by estimating their VP and analyze the tradeoff between registration errors and reconstruction robustness. We address the issue of data inconsistency in fusing multiple LR image data and propose a two-stage DT-based interpolation scheme. We also propose a PDE-based nonlinear deblurring technique to accommodate the blindness and variation of sensor PSF.

One open problem beyond the scope of this work is synthetic zooming for nonconstant scene depth (i.e., assumption (1) is violated). When objects are located at different scene depth, their VPs will vary. How to estimate multiple VPs with an image sequence? How to segment objects based on their corresponding VPs? How to handle the occluded regions associated with scene depth discontinuities? These questions are left for future work. Another related problem is synthetic zooming for a large number of frames (i.e., assumption (3) is violated). For a long sequence, the zooming speed might vary, which renders multiple VPs. How to segment a sequence based on their varying VPs and how to fuse multiple SR reconstructed images both deserve further study.

APPENDICES

A. PROOF OF LEMMA 1

If we write $\vec{e} = \vec{p} - \vec{r} = (e_x, e_y, e_z)$, then,

$$e_x = -\frac{\delta_x(|Z| + K)}{K}, \quad e_y = -\frac{\delta_y(|Z| + K)}{K}, \quad e_z = Z. \tag{A.1}$$

Note that $|\vec{q} - \vec{p}| \geq K$. It follows from (8) that the disturbed distance is given by

$$\begin{aligned} d &= \frac{|\delta\vec{p} \times \vec{e}|}{|\vec{q} - \vec{p}|} \leq \frac{\sqrt{e_z^2(\delta_x^2 + \delta_y^2) + (\delta_x e_y - \delta_y e_x)^2}}{K} \\ &= \frac{|Z|\sqrt{\delta_x^2 + \delta_y^2}}{K}. \end{aligned} \tag{A.2}$$

B. PROOF OF LEMMA 2

The disturbance of VP moves $\vec{q} = (x_2, y_2, l)$ to $\vec{q}' = (x_2', y_2', l)$, where

$$\begin{aligned} x_2' &= x_1 + \frac{(x_0' - x_1)(z_2 - z_1)}{z_0' - z_1}, \\ y_2' &= y_1 + \frac{(y_0' - y_1)(z_2 - z_1)}{z_0' - z_1}. \end{aligned} \tag{B.1}$$

Comparing (10) and (B.1), we obtain the movement along x-axis:

$$|x_2' - x_2| = \left| \left(\frac{x_0' - x_1}{z_0' - z_1} - \frac{x_0 - x_1}{z_0 - z_1} \right)(z_2 - z_1) \right|. \tag{B.2}$$

Assuming that δ_z is small, we have

$$|x_2' - x_2| \approx \left| \frac{x_0' - x_0}{z_0 - z_1}(z_2 - z_1) \right| = \left| \frac{\delta_x(l - k)}{Z - k} \right|. \tag{B.3}$$

Analysis along y-axis is similar.

ACKNOWLEDGMENT

This work was partially supported by the NASA WV-EPSCoR Award.

REFERENCES

[1] G. C. Holst, *CCD Arrays, Cameras and Displays*, SPIE-International Society for Optical Engine, Bellingham, Wash, USA, 1998.

[2] E. Choi, J. Choi, and M. G. Kang, "Super-resolution approach to overcome physical limitations of imaging sensors: an overview," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 36–46, 2004, Special issue on high resolution image reconstruction.

[3] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, 2004, Special issue on high resolution image reconstruction.

[4] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[5] Y.-W. Wen, M. K. Ng, and W.-K. Ching, "High-resolution image reconstruction from rotated and translated low-resolution images with multisensors," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 75–83, 2004, Special issue on high resolution image reconstruction.

[6] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient super-resolution image reconstruction algorithm," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 573–583, 2001.

[7] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super-resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[8] A. J. Patti, M. I. Sezan, and A. Murat Tekalp, "Super-resolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1064–1076, 1997.

[9] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, 1996.

[10] L. D. Alvarez, J. Mateos, R. Molina, and A. K. Katsaggelos, "High-resolution images from compressed low-resolution video: Motion estimation and observable pixels," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 58–66, 2004, Special issue on high resolution image reconstruction.

[11] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of compressed video using transform-domain statistics," *IEEE Transactions on Image Processing*, vol. 13, no. 1, pp. 33–43, 2004.

[12] R. Jin, Y. Qi, and A. Hauptmann, "A probabilistic model for camera zoom detection," in *Proceedings of IEEE 16th International Conference on Pattern Recognition (ICPR '02)*, vol. 3, pp. 859–862, Quebec City, Quebec, Canada, August 2002.

[13] R. Kingslake, *Applied Optics and Optical Engineering*, Academic Press, New York, NY, USA, 1965.

[14] S. Lertrattanapanich and N. K. Bose, "High resolution image formation from low resolution frames using Delaunay triangulation," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1427–1441, 2002.

[15] S. Osher and L. I. Rudin, "Feature-oriented image-enhancement using shock filters," *SIAM Journal on Numerical Analysis*, vol. 27, no. 4, pp. 919–940, 1990.

[16] L. Alvarez and L. Mazorra, "Signal and image restoration using shock filters and anisotropic diffusion," *SIAM Journal on Numerical Analysis*, vol. 31, no. 2, pp. 590–605, 1994.

[17] A. I. El-Fallah and G. E. Ford, "Mean curvature evolution and surface area scaling in image filtering," *IEEE Transactions on Image Processing*, vol. 6, no. 5, pp. 750–753, 1997.

[18] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[19] A. Murat Tekalp, *Digital Video Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.

[20] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of IEEE 7th International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1150–1157, Kerkyra, Greece, September 1999.

[21] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530–535, 1997.

[22] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pp. 593–600, Seattle, Wash, USA, June 1994.

[23] C. Yen, P. J. Burt, and X. Xu, "Local correlation measures from motion analysis: a comparative study," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '82)*, pp. 269–274, Las Vegas, Nev, USA, June 1982.

[24] Y.-H. Gu and T. Tjahjadi, "Coarse-to-fine planar object identification using invariant curve features and B-spline modeling," *Pattern Recognition*, vol. 33, no. 9, pp. 1411–1422, 2000.

[25] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[26] L. Alvarez, P.-L. Lions, and J.-M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion. II," *SIAM Journal on Numerical Analysis*, vol. 29, no. 3, pp. 845–866, 1992.

[27] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM Journal on Numerical Analysis*, vol. 29, no. 1, pp. 182–193, 1992.

[28] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Forward-and-backward diffusion processes for adaptive image enhancement and denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 7, pp. 689–703, 2002.

[29] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[30] P. J. Olver, G. Sapiro, and A. Tannenbaum, "Affine invariant detection: edge maps, anisotropic diffusion, and active contours," *Acta Applicandae Mathematicae*, vol. 59, no. 1, pp. 45–77, 1999.

Xin Li received the B.S. degree with highest honors in electronic engineering and information science from the University of Science and Technology of China, Hefei, in 1996, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, New Jersey, in 2000. He was a Member of Technical Staff with Sharp Laboratories of America, Camas, Washington, from August 2000 to December 2002. Since January 2003, he has been a Faculty Member in Lane Department of Computer Science and Electrical Engineering. His research interests include image/video coding and processing. Dr. Li received the Best Student Paper Award at the Conference of Visual Communications and Image Processing, San Jose, California, in January 2001.