Resolution Improvement from Stereo Images with 3D Pose Differences

Siu Hang Or[+], Ying Kin Yu[+], Kin Hong Wong[+] and Michael Ming Yuen Chang[*]
Computer Science & Engineering Department[+]   Information Engineering Department[*]
The Chinese University of Hong Kong

Abstract

Resolution improvement from several images is typically restricted to simple planar rotations and translations. In this paper, a super-resolution algorithm that allows 3D ego-motion of the camera system is proposed. By exploiting the trifocal tensor constraint of a stereo camera system, the intermediate step of scene structure recovery is effectively skipped. 3D motion is estimated recursively through an extended Kalman filter and used by a novel image warping procedure to perform resolution enhancement. Real image experiment with comparison to well known Irani-Peleg approach confirms the validity of the algorithm.

Index Terms --- image enhancement, image registration, motion analysis, stereo vision, kalman filtering.

1. Introduction

Super resolution, which refers to the construction of an image with higher resolution from several images of the same scene, is a topic that receives much attention recently. The seminal work by Irani and Peleg[8] described a reconstruction based algorithm which iteratively refines a higher resolution image by re-projecting it to produce several lower ones. These lower resolution images are compared with the input samples and the difference is used to update the higher resolution image. A number of later methods are inspired by Irani's approach[9,11,12,13]. The idea is also expanded into time axis[14]. However almost all of these approaches are rather restricted to planar features, and assuming that the distance from the scene to the camera is large compared with the variation in depth of the scene. As a result, almost all the proposed algorithms are being applied to outdoors distant scene or simple planar data.

In this paper, we propose an algorithm to incorporate a more general motion model, in particular rigid transformation, into the Irani-Peleg formulation. We demonstrate the advantage of our approach by applying it to upgrade a sequence of real images

2. Theory

In the formulation of Irani and Peleg[8] and other approaches[11,12,13] thereafter developed, the image formation process is represented as

$$\rho = \left[H_{cam} * F\left(H_{atm} * \mathrm{P}\right)\right]\downarrow + \varepsilon \tag{1}$$

where ρ is the captured noisy image, P is original high resolution image of the scene, H_i is the point spread transfer function(*psf*) which models different effects – subscript *cam* represents the camera capture effect and *atm* models the atmospheric transfer. Finally F denotes the motion which the camera undergoes and ε is system noise function, and $\downarrow$ corresponds to down sampling.

The original formulation requires that the camera undergoes a rigid planar transform which is modeled as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x'\cos\theta - y'\sin\theta + t_x \\ x'\sin\theta + y'\cos\theta + t_y \end{pmatrix} \tag{2}$$

where $(x,y)^{\mathrm{T}}$ is the captured image coordinates. $(x',y')^{\mathrm{T}}$ is the original scene point coordinates. $(t_x, t_y)^{\mathrm{T}}$ represents the translational and θ represents the rotational component of the planar transformation.

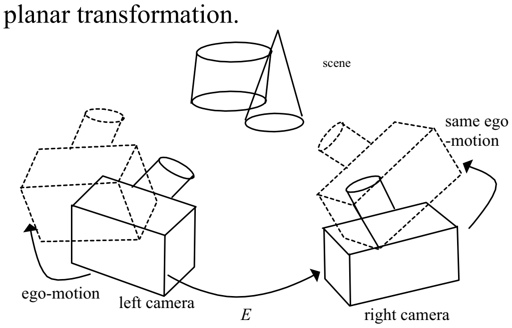


Figure. 1 System set up for image capture

During the image capture, we assume the camera undergoes an ego-motion and use a stereo set up as in Figure 1. In this case, a scene point with world coordinates $(x^w y^w z^w)^T$ will project onto the left and right camera plane as $(u,v)^T$ and $(u',v')^T$, respectively:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = K[I_{3\times3} \quad 0_{3\times1}]M \begin{bmatrix} x^W \\ y^W \\ z^W \\ 1 \end{bmatrix}, \begin{bmatrix} \tilde{u}' \\ \tilde{v}' \\ \tilde{w}' \end{bmatrix} = KEM \begin{bmatrix} x^W \\ y^W \\ z^W \\ 1 \end{bmatrix} \tag{3}$$

$$p = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \tilde{u}/\tilde{w} \\ \tilde{v}/\tilde{w} \end{bmatrix}, p' = \begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \tilde{u}'/\tilde{w}' \\ \tilde{v}'/\tilde{w}' \end{bmatrix} \tag{4}$$

where K is a 3x3 matrix that encodes the *intrinsic* parameters, say focal length f, of the camera. It is fixed and can be found by the camera calibration process[10]. For simplicity, the two cameras used in our stereo system are assumed to be identical. E is a 3x4 matrix describing the rigid transformation between the two cameras in the stereo system. M is a 4x4 matrix representing the rigid transformation in $\Re^3$. It transforms the 3-D structure from the world frame to the reference camera. The relationship between equation (3) and (1) can be summarized by the following formula $F = KE'M$, where E' equals to [I 0] and E for left and right camera respectively. In general K can assume to be identity matrix for simplicity, or lumped into H_{atm} in equation (1). In addition, E is fixed throughout the capture process and is thus constant. Thus without lost of generality, we take $F = CM$ for constant C and focus primarily on M hereafter.

The proposed system set up obtains a pair of image sequences $\{p_t, t = 1,2,..n\}$ and $\{p'_t, t = 1,2,..n\}$, where t is the time instant the image taken. To perform resolution enhancement, the typical process is to take several images of the same scene at different poses and carry out the refinement. We use the same reconstruction framework as that of Irani:

$$\rho^{i+1} = \rho^i + \frac{1}{L}\sum_{\forall t} F_t^{-1}\left[H_{atm} * \left(\hat{\rho}_t - \rho_t\right)\uparrow\right], \qquad (5)$$

where L is a normalizing constant, $\hat{\rho}_t$ is the reconstructed version of the captured frame ρ_t from the high resolution one through the image formation process described in equation (1), finally $\uparrow$ is the up-sampling process.

2.1 Image Warping

The reconstruction process requires the estimation of both F_t and F_t^{-1}, which account for the warping of an image to another due to geometrical transformation. For the relaxed condition of general 3D motion described by M, the estimation process would be more involved since the image now probably has some occlusions as well as newly appeared part of the scene. An accurate estimation of these regions as well as the original correspondences among scenes are very important as they determine the quality of the final reconstructed high resolution image.

On the other hand, applying the approach of planar motion cannot correctly register the images as the motion model is unable to handle the warping of image regions when rotations into the image plane occurs, as shown in the result later in Figure 2. To correctly account for the 3D motion, we propose to first recover the accurate 3D camera motion by a pose estimation algorithm, and then estimate the dense warping between reference frame and candidate frame.

2.2 3D motion Tracking

3D motion estimation, also called pose tracking, is a well-studied area in computer vision research. The task can be accomplished depending on whether the knowledge about the structure of the scene is known beforehand: *model-based*[2,15] in which advance knowledge of the scene is known and *structure from motion*[3,6] where both structure of the scene and motion information are to be extracted simultaneously. In most situations, an accurate estimation of the scene geometry is difficult to achieve. However in our problem of resolution enhancement, recovering the scene structure, i.e. using structure from motion method, to perform image registration would further complicate the devised solution. To solve the problem, we apply a novel extended Kalman filter[10,17] algorithm, together with the trifocal tensor[17], so as to fully exploit the advantages of the stereo camera system. The significance is that the 3D motion can now be *directly estimated* i.e. eliminating the intermediate step of scene structure recovery.

Firstly a number of features are being tracked throughout the left and the right image sequences independently using the Kanade-Lucas-Tomasi (KLT) tracker[6]. Stereo correspondences among feature points are setup afterwards by first estimating the fundamental matrix, followed by correlation using epipolar geometry[5]. For the details of design of the EKF in our implementation, the readers are referred to [16,17] .

2.3 Image Warping by Surface Fitting

The tensor-based Kalman filter approach we used[17] can accurately estimate the 3D rigid transformation of the camera. Theoretically given a full correspondences between frame p_1 and p'_1, which are the initial left and right frame respectively, together with the tensors relating p_1 and p'_1 and p_n i.e. the left frame at time n, the mapping of every pixels from p_1 to p_n can be uniquely determined. However it would thus require an accurate and dense correspondences between p_1 and p'_1. To avoid the tedious task of dense correspondence estimation, below we propose a simple but effective approach.

As we already got a number of features correspondences through the **KLT** tracker, we can use them to compute the dense correspondences. The idea is to first compute the correspondences of those **KLT** features between p_1 and p_n using the trifocal tensor. A scattered fitting is then performed to reconstruct the estimation for the *x*- and *y*-mapping of other pixels between the two frames. Bicubic interpolation is used in the fitting. The fitted surface can thus provide a dense correspondence for the two frames. The advantage of using surface fitting is that we are using the reliably tracked features to guide the interpolation. The correspondence results are thus free from spurious mismatches that occurred in typical optic flow based dense reconstruction. The correspondence map will take the role of the global image registration step in that of Irani and Peleg. To further ensure the correct registration of the two images, we apply the original image registration step as in equation (2). The image warping parameters F_t^{-1} thus consists of both the feature correspondences predicted by our motion estimation algorithm, together with the planar rotation and translation parameter in equation (2). Equation (5) is then applied to perform the super resolution.

3. Experiments and results

We apply the above procedures to a stereo laboratory image sequence that consists of a series of rotations in addition to translations in both *x*- and *z*-axis. The testing images consist of two 115-frame sequences of 640 by 480 pixels captured by both the left and right camera. We first apply the proposed 3D motion estimation algorithm as described in section 2.2 to extract the motion information. Then all the features, (approximately 70 in most cases), which are reliably tracked in the selected frames are being reprojected to the frame p_n by the tensor to establish the scattered set of correspondences. Image warping parameter estimation described previously are then applied to produce the full correspondences between p_n and p_1. This procedure is repeated for a few frames centered around the frame that resolution enhancement is being performed. We found that usually 4 frames are sufficient to produce good results in this process. The Irani-Peleg reprojection algorithm is then performed to iteratively refine the images until no improvement can be seen.

Selected results are shown in Figure 2. The result of using the proposed warping method to generate the reconstructed version of reference frame is shown in Figure 2f. We also implemented the original Irani-Peleg method using motion described by equation (2) as a comparison for the selected frame and the result is also shown in Figure 2e. As can be seen from the result, the original Irani-Peleg implementation cannot correctly enhance the zoomed region due to the rotational camera motion as well as the movement along the *z*-axis. For example, significant aliasing can be seen in the blow up region, especially at the interior of the file label with “student” printed on it. On the contrary, our proposed reconstruction method can correctly enhance the image quality as shown in the same figure.

The results above illustrate the validity of our 3D motion-based super resolution algorithm. Although we do not show the result for the right camera sequence here, but the idea of resolution enhancement can equally be applied to the right camera sequence with just minor changes in frame reference and the tensor computation. An obvious advantage is that using the proposed set up, one can perform super resolution on both left and right image sequence and thus produce a stereo sequence with enhanced resolution.

4. Conclusion

By relaxing the simple motion model assumed in the resolution enhancement algorithm by Irani and Peleg[8], robust performance has been achieved for an image sequence with more general 3D motion. The key idea is to exploit the trifocal tensor constraint in stereo vision so that scene structure recovery is no longer needed. The 3D motion information is thus used to guide a novel image warping in the super-resolution process. The effectiveness of the algorithm is demonstrated through a real image sequence.

5. Acknowledgment

The work described in this paper was supported by a grant (Project No.: 4204/04E) from the Research Grant Council of Hong Kong Special Administrative Region and a direct grant (Project Code: 2050350) from the Faculty of Engineering of the Chinese University of Hong Kong.

6. References

[1] A.Azarbayejani and A.P.Pentland, “Recursive estimation of motion, structure, and focal length”, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 6, pp. 562-575, Jun. 1995.

[2] V.Lippiello, B.Siciliano and L.Villani, “Objects motion estimation via BSP tree modeling and Kalman filtering of stereo images”, in *Proc. IEEE Int. Conf. Robotics Autom.*, pp. 2968-2973, Washington DC, 2002

[3] Y.K.Yu, K.H.Wong and M.M.Y.Chang, “Recursive three-dimensional model reconstruction based on Kalman filtering”, *IEEE Trans. Syst., Man, Cybern. B,* vol. 35, no. 3, pp. 587-592, Jun. 2005.

[4] M.S.Grewal, A.P.Andrews, Kalman Filtering Theory and Practice, Prentice Hall, 1993.

[5] R.Hartley and A.Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2000.

[6] C.J.Poelman, T.Kanade, “A paraperspective factorization method for shape and motion recovery”, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 3, pp. 206-218, Mar. 1997.

[7] S.Avidan and A.Shashua, “Threading fundamental matrices”, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 1, pp. 73-77, Jan. 2001.

[8] M. Irani and S. Peleg, "Improving resolution by Image Registration," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 231--239, Mar. 1991.

[9] M. Elad and A. Feuer. “Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images”. In *IEEE Trans. on Image Processing*, pages 1646--1658, December 1997.

[10] Y.K.Yu, K.H.Wong and M.M.Y.Chang, *“*Merging artificial objects with marker-less video sequences based on the interacting multiple model method*”, IEEE Trans. Multimedia,* pp. 521-528, vol. 8, no. 3, Jun. 2006*.*

[11] David Capel, Andrew Zisserman. "Super-Resolution Enhancement of Text Image Sequences," *icpr*, p. 1600, 15th

International Conference on Pattern Recognition (ICPR'00) - Volume 1, 2000.

[12] Farsiu, S. , D. Robinson, M. Elad, and P. Milanfar, "Fast and Robust Multi-frame Super-resolution", *IEEE Trans. on Image Processing, vol. 13, no. 10, pp. 1327-1344 , October 2004*

[13] W. Zhao and H. Sawhney, "Is Super-Resolution with Optical Flow Feasible?" In Proc. *European Conf. Computer Vision*, pp. 599-613 Vol. I, 2002.

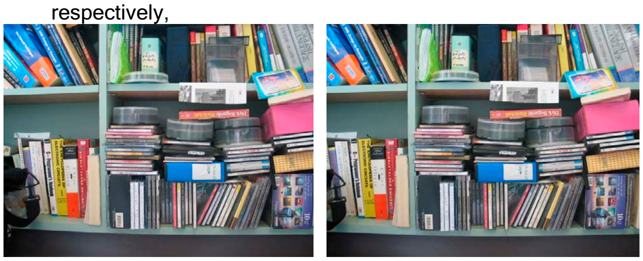
[14] E. Shechtman, Y. Caspi, and M. Irani, "Increasing Space-Time Resolution in Video", *European Conference on Computer Vision (ECCV)*, May 2002.

[15] S.H.Or, W.S. Luk, K.H.Wong & Irwin King, "An Efficient Iterative Pose Estimation Algorithm"*, Image & Vision Computing*, 16(5):355-364, 1998.

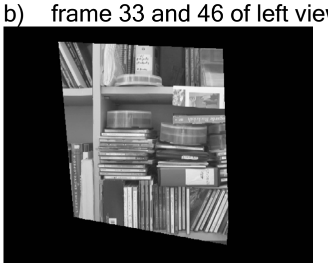
[16] Y.K.Yu, K.H.Wong, M.M.Y.Chang and S.H.Or, "Recursive camera motion estimation with the trifocal tensor", *IEEE Trans. Syst., Man, Cybern. B, Cybern., to be published.*

[17] Y.K.Yu, K.H.Wong, S.H.Or and M.M.Y.Chang, "Recursive recovery of position and orientation from stereo image sequences without three-dimensional structures"*, in Proc. IEEE CVPR,* New York, Jun. 2006.

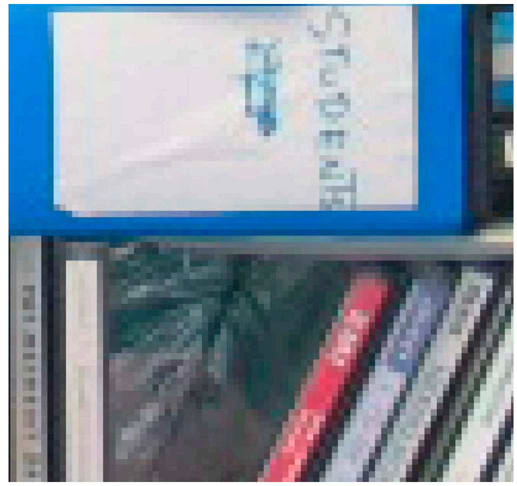
a) laboratory image sequence, frame 1 of left and right view respectively,



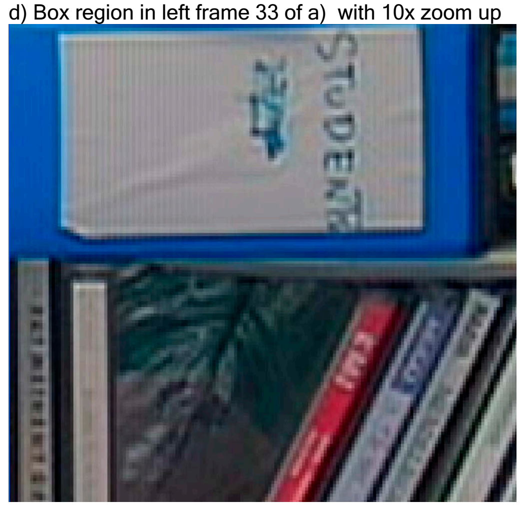
b) frame 33 and 46 of left view respectively,



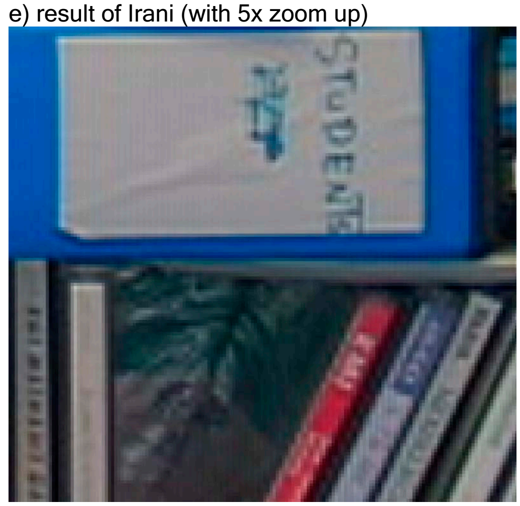
c) frame 46 being warped back to frame 33,



d) Box region in left frame 33 of a) with 10x zoom up



e) result of Irani (with 5x zoom up)



f) result of our algorithm (with 5x zoom up). Note the better anti-aliasing effect at the interior of white file label.

Figure 2. Testing image sequence and comparison of results of our proposed super-resolution algorithm with Irani-Peleg.