High Frequency 3D Reconstruction from Unsynchronized Multiple Cameras

Yumi Kakumu Fumihiko Sakaue Jun Sato
Nagoya Institute of Technology, Japan
yumi@cv.nitech.ac.jp,sakaue@nitech.ac.jp,junsato@nitech.ac.jp

Kazuhisa Ishimaru
NIPPON SOKEN, INC., Japan
kazuhisa_ishimaru@soken1.denso.co.jp

Masayuki Imanishi
DENSO International America, Inc.

Stereo reconstruction generally requires image correspondence such as point and line correspondences in multiple images. Cameras need to be synchronized to obtain corresponding points of time-varying shapes. However, the image information obtained from synchronized multiple cameras is redundant, and has limitations with resolution. In this paper, we show that by using a set of unsynchronized cameras instead of synchronized one, we can obtain much more information on 3D motions, and can reconstruct higher frequency 3D motions than that with the standard synchronized cameras. As a result, we can attain super resolution 3D reconstruction from unsynchronized cameras.

In the standard reconstruction with multiple cameras, the cameras are synchronized, and observe the same set of M sequential points in 3D space as shown in Fig. 1 (a). Thus, the maximum frequency f_S of 3D motion, which can be recovered from camera images, is $f_S < \frac{1}{2}M$. If we observe the same 3D motion by using K unsynchronized cameras, we observe $K \times M$ different points in the 3D space as shown in Fig. 1 (b). As a result, all the $2KM$ observations from K cameras are independent of one another, unlike those from the synchronized cameras. Thus, we have a possibility of reconstructing 3D points up to $\frac{2}{3}KM$. Therefore, the maximum frequency f_U of recoverable 3D motion is $f_U < \frac{1}{3}KM$. Thus, we find that the unsynchronized cameras have a possibility of reconstructing $\frac{2}{3}K$ times higher frequency 3D motion than the synchronized cameras as follows:

$$f_U = \frac{2}{3}K f_S \qquad (1)$$

For example, if we have 3 unsynchronized cameras, we can recover 2 times higher frequency components than reconstruction with 3 synchronized cameras.

Unfortunately, the image points obtained from a set of unsynchronized cameras are not corresponding to each other as shown in Fig. 1 (b), and thus we cannot reconstruct 3D point by using the existing stereo methods. For reconstructing 3D motions from unsynchronized cameras, we consider a novel camera projection model, which projects 3D point in frequency space to image points in usual metric space.

Suppose the maximum frequency f of 3D motion Σ is less than $\frac{1}{2}N$. Then, 3D motion Σ can be fully represented by N frequency points $\mathbf{Z}_n = [X_n^f, Y_n^f, Z_n^f, \delta_n]^\top$ $(n = 0, \cdots, N-1)$. If we observe 3D motion Σ at M time instants, the observed 3D points $\mathbf{X}_t$ $(t = 0, \cdots, M-1)$ in real space can be represented by frequency points $\mathbf{Z}_n$ as follows:

$$\mathbf{X}_t = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{Z}_n \omega(t,n) \qquad (t = 0, \cdots, M-1) \qquad (2)$$

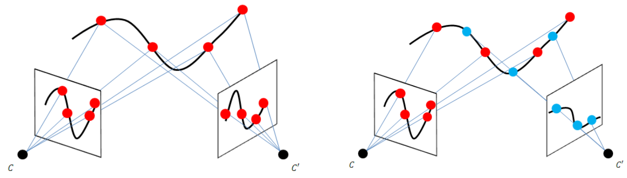
where, $\omega(t,n) = e^{2\pi jnt/N}$. Note, N and M are different in general, and we consider the case where the number of sampling, M, is less than the number of frequency components, N.

Now, if we observe 3D motion Σ by K different unsynchronized cameras at M time instants, their projections can be described as follows:

$$\lambda_t^i \mathbf{x}_t^i = \mathbf{P}^i \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{Z}_n \omega(t + k_i, n) \qquad (i = 1, \cdots, K) \qquad (3)$$

where, k_i denotes sampling delay of the ith camera with respect to the 1st camera. Thus, $k_1 = 0$. $\mathbf{P}^i$, $\mathbf{x}_t^i$ and λ_t^i are a camera projection matrix, a series of image points and their depth in the ith camera.

The equation (3) is a general projection model from 3D motion with N frequency components to K unsynchronized cameras with M time instants. The very important property of this projection model is that the identical 3D frequency point $\mathbf{Z}_n$ is projected to image points $\mathbf{x}_t^i$ in multiple cameras as shown in (3). This property exists in this projection model, since sampling delay $\omega(k_i, n)$ is separated from 3D frequency point $\mathbf{Z}_n$ in frequency space. As a result, 3D frequency point $\mathbf{Z}_n$ can be reconstructed from these image points.



(a) synchronized cameras (b) unsynchronized cameras

Figure 1: Image projection in synchronized and unsynchronized cameras

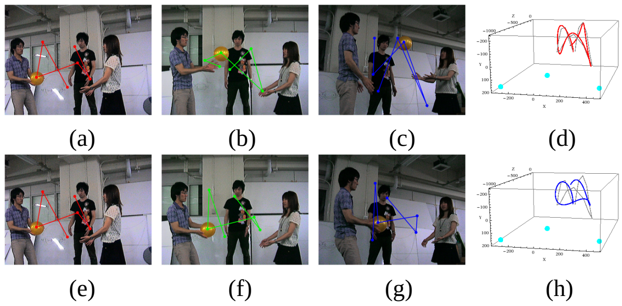


Figure 2: (a), (b) and (c) show images observed by 3 unsynchronized cameras, and (e), (f) and (g) show those observed by 3 synchronized cameras. The three people are passing a ball. The center of the ball is extracted in each camera at 6 time instants, and used for reconstruction. The points and lines show extracted points and their sequences. (d) and (h) show 3D points reconstructed from the proposed method and the standard method respectively.

By taking the vector product of (3) with $\mathbf{x}_t^i$, and reformulating it with respect to the 3D frequency components, $\mathbf{f} = [X_0^f, \cdots, X_{N-1}^f, Y_0^f, \cdots, Y_{N-1}^f, Z_0^f, \cdots, Z_{N-1}^f, 1]^\top$, we have the following system of linear equations:

$$\mathbf{Mf} = \mathbf{0} \qquad (4)$$

where, $\mathbf{M}$ represents a $(2KM) \times (3N+1)$ matrix, whose components consist of $\mathbf{P}$, $\mathbf{x}$ and ω.

Assuming that sampling delays k_i and camera matrices $\mathbf{P}^i$ are known, 3D coordinates $\mathbf{f}$ of N frequency points $\mathbf{Z}_n$ can be computed from M image sequence of K unsynchronized cameras by solving the linear equation (4). Then, 3D point in real space can be recovered by transforming $\mathbf{Z}_n$ to $\mathbf{X}_t$ using (2).

Fig. 2 (d) shows 3D points reconstructed by the proposed method, and (h) shows the results from the standard method. As shown in these figures, the proposed method can fully recover original 3D motion, while the results from the standard method suffer from the aliasing problem and are imperfect.

[1] S. Agarwal, N. Snavely, I. Simon, S.M. Seitz, and R. Szeliski. Building rome in a day. In *Proc. ICCV*, 2009.

[2] O. D. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[3] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.