

# High Frequency 3D Reconstruction from Unsynchronized Multiple Cameras

Yumi Kakumu<sup>1</sup>  
yumi@cv.nitech.ac.jp

Fumihiko Sakaue<sup>1</sup>  
sakaue@nitech.ac.jp

Jun Sato<sup>1</sup>  
junsato@nitech.ac.jp

Kazuhisa Ishimaru<sup>2</sup>  
kazuhisa\_ishimaru@soken1.denso.co.jp

Masayuki Imanishi<sup>3</sup>

<sup>1</sup> Department of Computer Science and Engineering  
Nagoya Institute of Technology  
Nagoya 466-8555, Japan

<sup>2</sup> NIPPON SOKEN, INC., Japan

<sup>3</sup> DENSO International America, Inc.

---

## Abstract

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. The efficiency of the proposed method is tested by using real image sequences.

## 1 Introduction

Stereo reconstruction methods generally require a set of corresponding points in multiple images [1, 2, 4]. These corresponding points must be projections of the same points in 3D space. Thus, if we want to reconstruct dynamic scenes, multiple cameras must be synchronized. Synchronized cameras can observe the same 3D point at each time instant as shown in Fig. 1 (a), and these 3D points can be reconstructed. Hence, the number of reconstructible 3D points is the same as the number of corresponding points in images, and the maximum reconstructible frequency is a half of the camera sampling frequency. Thus, if the original 3D motion includes higher frequencies than a half of the camera sampling frequency, the reconstructed 3D motion suffers from the aliasing problem, and we obtain wrong 3D motion from images. The traditional solution to this problem is to use high speed cameras, i.e. high sampling rate cameras. However, the amount of light obtained with the high speed cameras is often insufficient. Also, there are limits with sampling frequency even with high speed cameras. Thus, the aliasing problem is inevitable in the standard reconstruction methods.

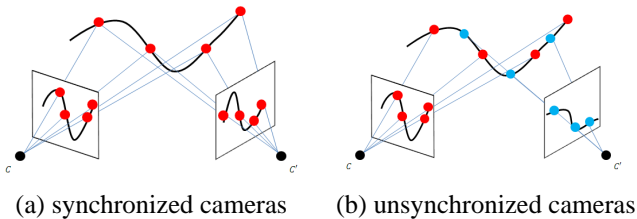


Figure 1: The synchronized and unsynchronized cameras

In this paper, we consider a set of cameras, whose sampling phases are different from each other, as shown in Fig. 1 (b), and propose a method for reconstructing higher frequency components of 3D motion than the sampling frequency of cameras. We call them unsynchronized cameras. If we observe a moving 3D point with a pair of synchronized cameras, we obtain two sets of 2D coordinates, i.e. four coordinates, from a single sampling. However, only three of them are independent, since these two cameras observe the same 3D point. However, if we observe the moving 3D point with a pair of unsynchronized cameras, all four coordinates provide us independent information on the moving 3D point, and we have a possibility of using the information for recovering higher frequency components of 3D motion. Unfortunately, these image points are not corresponding to each other, and thus we cannot reconstruct the 3D point by using the existing stereo methods.

Some efforts have recently been made for reconstructing 3D points from unsynchronized stereo cameras [3, 5, 7]. For recovering 3D motions from unsynchronized cameras, image interpolation techniques are often used to generate corresponding points in unsynchronized images [3, 7, 8]. Svedman et al. [8] proposed a method for creating virtual synchronized images from unsynchronized images based on interpolation. Zhou and Tao [11] proposed a method for generating a virtual synchronized images from projective invariants assuming linear motion. Tresadern and Reid [9] proposed a method of bundle adjustment based optimization for recovering human motions from unsynchronized cameras. Unfortunately, the interpolation used in these methods was an approximation of the real motion in images, and thus the results from reconstruction were not accurate. Furthermore, all these methods had the same limitation. That is, the highest reconstructible frequency of 3D motion was the half of the camera sampling frequency.

In this paper, we propose a method for reconstructing higher frequency components of 3D motion from unsynchronized low frequency cameras, i.e. low sampling rate cameras. Our research is related to recent work on trajectory reconstruction from moving cameras [6]. For reconstructing 3D motion from unsynchronized cameras, we consider a novel camera projection model, which projects 3D points in frequency space to image points in the usual metric space. This new projection model has two major advantages to the standard camera projection model. First, identical 3D points in frequency space are projected to multiple images, even if these images are observed by unsynchronized cameras. Thus, we can obtain image correspondence in multiple unsynchronized images, and can reconstruct 3D points from unsynchronized images. Second, the new projection model enables us to reconstruct higher frequency 3D motion that could not be recovered from the standard stereo method. As a result, we can reconstruct more precise 3D motion than that with the existing 3D reconstruction methods, and can avoid the aliasing problem.

## 2 Unsynchronized Cameras

Let  $\Sigma$  be a continuous 3D motion of a space point, whose highest frequency is  $f$ . Suppose we have  $N$  sampling points  $\mathbf{X}_i = [X_i, Y_i, Z_i, 1]^\top$  ( $i = 0, \dots, N-1$ ) on the 3D motion  $\Sigma$ . Then, according to the sampling theorem, 3D motion  $\Sigma$  can be fully recovered by the  $N$  sampling points  $\mathbf{X}_i$  ( $i = 0, \dots, N-1$ ), if the following inequality holds:

$$N > 2f \quad (1)$$

Suppose  $\mathbf{Z}_i = [X_i^f, Y_i^f, Z_i^f, \delta_i]^\top$  ( $i = 0, \dots, N-1$ ) is the discrete Fourier transform (DFT) of  $\mathbf{X}_i = [X_i, Y_i, Z_i, 1]^\top$  ( $i = 0, \dots, N-1$ ), where  $\delta_i$  takes 1 for  $i = 0$  and takes 0 for the others. Then, we can represent 3D motion  $\Sigma$  by using  $N$  frequency points  $\mathbf{Z}_i$  ( $i = 0, \dots, N-1$ ) in 3D frequency space. Thus, if we can reconstruct  $3N$  frequency components  $\{X_i^f, Y_i^f, Z_i^f\}$  ( $i = 0, \dots, N-1$ ), 3D motion  $\Sigma$  can be reconstructed up to frequency  $f$ .

Now, let us consider  $K$  cameras, which observe 3D motion  $\Sigma$  at  $M$  time instants. Since each camera observes  $M$  image points, we have a total of  $2KM$  observations from  $K$  cameras. In the standard reconstruction with multiple cameras, the cameras are synchronized, and observe the same set of  $M$  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 as follows:

$$f_S < \frac{1}{2}M \quad (2)$$

However, if we observe the same 3D motion  $\Sigma$  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 as follows:

$$f_U < \frac{1}{3}KM \quad (3)$$

From (2) and (3), 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:

$$f_U = \frac{2}{3}Kf_S \quad (4)$$

Since  $K$  is 2 or more than 2 in the reconstruction with multiple cameras, the unsynchronized cameras can recover much higher frequency of 3D motion. Table 1 compares the recoverable maximum frequency of the proposed unsynchronized reconstruction method,  $f_U$ , with that of the standard synchronized reconstruction method,  $f_S$ . For example, if we have 3 unsynchronized cameras, we can recover 2 times higher frequency components than reconstruction with 3 synchronized cameras.

The question is how can we recover these higher frequencies from the unsynchronized cameras? Obviously, no standard stereo methods can be used, since we do not have any point correspondences in multiple images. In the following sections, we propose a linear method for recovering higher frequency components from unsynchronized images.

Table 1: Comparison of maximum frequency recoverable from  $M$  sampling points.  $f_S$  denotes the maximum frequency obtained from synchronized cameras, and  $f_U$  denotes that from unsynchronized cameras.

number of cameras	2	3	4	5	6	...	$K$
$f_S$	$\frac{1}{2}M$	$\frac{1}{2}M$	$\frac{1}{2}M$	$\frac{1}{2}M$	$\frac{1}{2}M$	...	$\frac{1}{2}M$
$f_U$	$\frac{2}{3}M$	$M$	$\frac{4}{3}M$	$\frac{5}{3}M$	$2M$	...	$\frac{1}{3}KM$

### 3 Projection in Frequency Space

We first introduce a brand new camera projection model, which projects 3D points in frequency space to 2D image points in real metric space. In this projection model, a 3D point is represented by frequency coordinates,  $\{X^f, Y^f, Z^f\}$ , while a 2D point is represented by the usual image coordinates,  $\{x, y\}$ .

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

$$\mathbf{X}_t = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{Z}_n \omega(t, n) \quad (t = 0, \dots, M-1) \quad (5)$$

where,  $\omega(t, n) = e^{2\pi j n t / N}$ , and  $j$  denotes the imaginary unit. Note,  $N$  and  $M$  are different in general, and we consider a case where the number of samplings  $M$  is less than the number of frequency components  $N$ . Equation (5) is similar to the inverse discrete Fourier transform (IDFT). However, it is not IDFT, since  $N$  and  $M$  are different. Since image points in the camera are the projection of  $\mathbf{X}_t$  ( $t = 0, \dots, M-1$ ), the image projection from 3D frequency space to the image points  $\mathbf{x}_t = [x_t, y_t, 1]^\top$  ( $t = 0, \dots, M-1$ ) can be described as follows:

$$\lambda_t \mathbf{x}_t = \mathbf{P} \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{Z}_n \omega(t, n) \quad (6)$$

where,  $\mathbf{P}$  denotes the  $3 \times 4$  camera projection matrix, and  $\lambda_t$  denotes a scalar that corresponds to the depth of the point.

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

$$\lambda_t^i \mathbf{x}_t^i = \mathbf{P}^i \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{Z}_n \omega(t + k_i, n) \quad (i = 1, \dots, K) \quad (t = 0, \dots, M-1) \quad (7)$$

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

Equation (7) 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 (7). 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 as we will explain in the next section.

## 4 3D Super Resolution Reconstruction

We next consider 3D super resolution reconstruction. That is, we reconstruct higher frequency 3D motion from low frequency image sampling. In this method, the recoverable frequency of the 3D motion is  $\frac{2}{3}K$  times higher than that of the standard reconstruction method as we have seen in section 2. As a result, reconstructed 3D motion does not suffer from the aliasing problem, even if the 2D images have aliasing.

By taking the vector product of (7) with  $\mathbf{x}_t^i$ , and reformulating it with respect to the 3D frequency components,  $\mathbf{f} = [X_0^f, \dots, X_{N-1}^f, Y_0^f, \dots, Y_{N-1}^f, Z_0^f, \dots, Z_{N-1}^f, 1]^\top$ , we have the following system of linear equations:

$$\mathbf{M}\mathbf{f} = \mathbf{0} \quad (8)$$

where  $\mathbf{M}$  represents a  $(2KM) \times (3N+1)$  matrix as follows:

$$\mathbf{M} = \begin{bmatrix} \mathbf{L}^1(P_{11}^1 - P_{31}^1 \mathbf{T}_x^1) & \dots & \mathbf{L}^1(P_{13}^1 - P_{33}^1 \mathbf{T}_x^1), & \mathbf{L}^1 \mathbf{Z}_4(P_{24}^1 - P_{34}^1 \mathbf{T}_x^1) \\ \mathbf{L}^1(P_{21}^1 - P_{31}^1 \mathbf{T}_y^1) & \dots & \mathbf{L}^1(P_{23}^1 - P_{33}^1 \mathbf{T}_y^1), & \mathbf{L}^1 \mathbf{Z}_4(P_{24}^1 - P_{34}^1 \mathbf{T}_y^1) \\ \vdots & & \vdots & \vdots \\ \mathbf{L}^K(P_{11}^K - P_{31}^K \mathbf{T}_x^K) & \dots & \mathbf{L}^K(P_{13}^K - P_{33}^K \mathbf{T}_x^K), & \mathbf{L}^K \mathbf{Z}_4(P_{24}^K - P_{34}^K \mathbf{T}_x^K) \\ \mathbf{L}^K(P_{21}^K - P_{31}^K \mathbf{T}_y^K) & \dots & \mathbf{L}^K(P_{23}^K - P_{33}^K \mathbf{T}_y^K), & \mathbf{L}^K \mathbf{Z}_4(P_{24}^K - P_{34}^K \mathbf{T}_y^K) \end{bmatrix} \quad (9)$$

Here,  $\mathbf{Z}_4$  denotes the 4th row of  $\mathbf{Z}(n)$ , and  $\mathbf{L}^i$  denotes an  $M \times N$  matrix, which consists of  $\omega(t, n) = e^{2\pi jnt/N}$  as follows:

$$\mathbf{L}^i = \frac{1}{N} \begin{bmatrix} \omega(0+k_i, 0) & \dots & \omega(0+k_i, N-1) \\ \vdots & & \vdots \\ \omega(M-1+k_i, 0) & \dots & \omega(M-1+k_i, N-1) \end{bmatrix}$$

$\mathbf{T}_x^i$  and  $\mathbf{T}_y^i$  are vectors, which consist of  $x$  and  $y$  coordinates of  $M$  time instants in the  $i$ th camera image as follows:

$$\mathbf{T}_x^i = [x_0^i, \dots, x_{M-1}^i]^\top \quad \mathbf{T}_y^i = [y_0^i, \dots, y_{M-1}^i]^\top \quad (10)$$

Assuming that the sampling delays,  $k_i$ , and the camera matrices,  $\mathbf{P}^i$ , are known, the 3D coordinates,  $\mathbf{f}$ , of  $N$  frequency points  $\mathbf{Z}_n$  can be computed from the  $M$  image sequence of  $K$  unsynchronized cameras by solving the linear equation (8). Then, the 3D points in real space can be recovered by transforming  $\mathbf{Z}_n$  to  $\mathbf{X}_t$  by using the following equation:

$$\mathbf{X}_t = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{Z}_n \omega(t+k_i, n) \quad (11)$$

Note, we can reconstruct a different set of 3D points observed by each camera by choosing delay  $k_i$  of each camera in (11).

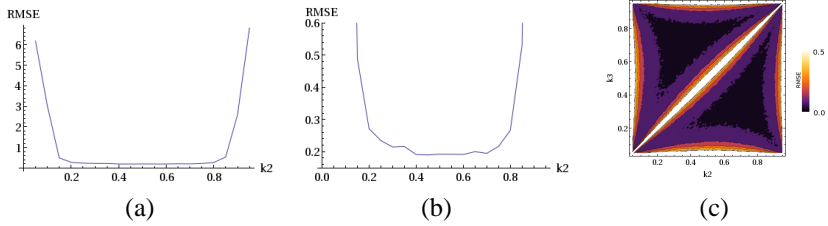


Figure 2: Relationship between sampling delay and RMS error of reconstruction. (a) shows the relationship in the case of 2 cameras. (b) magnifies vertical axes of (a). (c) shows relationship in the case of 3 cameras.  $k_2$  and  $k_3$  denote the sampling delays of camera 2 and 3 respectively.

## 5 Sampling Delay and Stability

Although the higher frequency components of 3D motion can be reconstructed up to  $f_U$  by using the proposed method, the stability of reconstructed 3D motion depends on variations in sampling delay  $k_i$  of multiple cameras. If the variations in sampling delay are small in multiple cameras, the variations in 3D points observed by these cameras will be small. As a result, the reconstruction of high frequency 3D motion becomes unstable.

To clarify the relationship between the variations in sampling delay and the stability of 3D reconstruction, we generated a pair of synthetic unsynchronized images with image noise, and reconstructed high frequency 3D motion by using the proposed method. Sampling delay  $k_2$  of the 2nd camera was changed from 0 to 1, and we evaluated the RMS error of the reconstructed 3D points. Fig. 2 (a) shows the relationship between the magnitude of sampling delay and the RMS error of reconstruction, and Fig. 2 (b) magnifies its vertical scale. As shown in these figures, the reconstruction error is large when sampling delay  $k_2$  of the 2nd camera is close to 0 or 1. However, the stability of reconstruction is drastically improved, when sampling delay  $k_2$  approaches to 0.5.

We also evaluated the stability of reconstruction for 3 unsynchronized cameras changing the sampling delay of the 2nd and 3rd cameras,  $k_2$  and  $k_3$ , from 0 to 1. The results are shown in Fig. 2 (c), where the stability of reconstruction is improved, when  $k_2$  and  $k_3$  approach  $\frac{1}{3}$  and  $\frac{2}{3}$  or vice versa. These results confirm that variations in sampling delay are important for the stability of reconstruction.

## 6 Experiments

We will next present the results from real image experiments as well as synthetic image experiments to show the efficiency of the proposed method.

### 6.1 Reconstruction of High Frequency 3D Motions

We will first show the reconstruction of high frequency 3D motion from low sampling rate images by using the proposed method.

Fig. 3 (a), (b) and (c) show image points obtained by observing 3D motion of a ball at 6 time instants with three unsynchronized cameras, and Fig. 3 (e), (f) and (g) show those

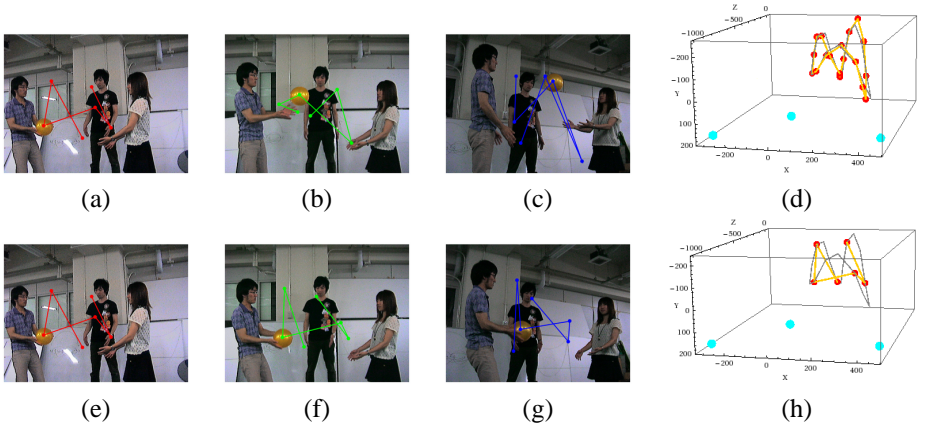


Figure 3: The 3D motion observed by three unsynchronized cameras, (a), (b) and (c), and the three synchronized cameras, (e), (f) and (g). The three people are passing the ball. The center of the ball is extracted in each camera at 6 time instants, and used for reconstruction. The points and lines show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

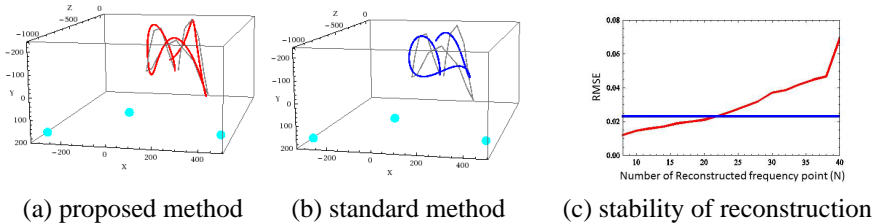


Figure 4: Recovered continuous 3D motions. (a) shows a continuous 3D motion recovered from the proposed method, while (b) shows those from the standard method. (c) shows the stability of reconstruction. The horizontal axis is the number of reconstructed frequency points  $N$ , and the vertical axis is the RMS error of reconstruction. The red and blue lines correspond to stability of the proposed method and the standard method respectively.

obtained from three synchronized cameras. As seen in Fig. 3, the unsynchronized cameras observe different 3D points, while the synchronized cameras observe the same 3D points. The sampling period of all cameras is 0.6 sec, and the sampling delays of the second and the third unsynchronized cameras are  $k_2 = \frac{1}{3}$  and  $k_3 = \frac{2}{3}$ . The internal and the external parameters of these 3 cameras were calibrated in advance with Zhang’s method [10].

Then, we reconstructed 3D motion of the ball by using the proposed method and the standard reconstruction method. In the proposed method, we reconstructed higher frequency components of 3D motion from 6 sequential image points in the unsynchronized images. In the standard method, we reconstructed 6 points from 6 sequential image points in the synchronized images. Since we used 3 cameras, the proposed method could recover twice as many higher frequency components as the standard method.

Fig. 3 (d) shows the 3D points reconstructed by the proposed method. The red points and the yellow lines show the reconstructed points and their sequences, and the gray lines show

the ground truth obtained by the synchronized high speed stereo cameras. The blue points show the viewpoints of 3 cameras. Fig. 3 (h) shows the results from the standard method. As we can see from these figures, the proposed method can fully recover original 3D motion, while the results from the standard reconstruction method are imperfect.

To clarify the advantage of the proposed method, we computed continuous 3D motion from the 3D frequency components recovered by the proposed method. This was achieved by convolving the sinc function to the recovered 3D frequency components. For comparison, we also computed continuous 3D motion from the results of the standard method by transforming it into frequency space and convolving the sinc function. Fig. 4 (a) shows the continuous 3D motion recovered from the proposed method, and Fig. 4 (b) shows that from the standard method. As we can see from these figures, the proposed method could accurately recover motion of the ball, while the results from the standard method suffered from the aliasing problem and have large distortion.

## 6.2 Stable Reconstruction of Low Frequency 3D Motions

As we have seen in section 6.1, the proposed method can recover high frequency 3D motion from low sampling rate cameras. Although this is the most important property of the proposed method, it also has another distinct advantage. That is, it can control the highest frequency of reconstructed 3D motion freely within the range from 0 to  $f_U$ . This means that we can also reduce the number of frequency components reconstructed from the proposed method. The advantage of reducing the number of frequency components is that we can increase the stability of reconstruction, i.e. it becomes less sensitive to image noise.

The highest frequency of reconstruction can be controlled by changing the number of frequency points  $N$  in (8) within the range from 1 to  $\frac{2}{3}KM$ . If  $N$  is equal to  $M$ , the stability of reconstruction with the proposed method is the same as that with the standard method. If  $N$  is smaller than  $M$ , the stability of the proposed method is better than that of the standard method. If  $N$  is larger than  $M$ , our method can recover higher frequency components, but it becomes more sensitive to image noise.

Fig. 4 (c) shows the stability of reconstruction against image noise, which was evaluated by using synthetic images with image noise. In this evaluation, 3 cameras were used for reconstruction, and each of them observed 20 points. Thus, the proposed method could recover up to 40 frequency points, while the standard method recovered 20 points. The horizontal axis is the number of reconstructed frequency points  $N$ , and the vertical axis is the RMS error of reconstruction caused by image noise with a standard deviation of 0.5. As seen in Fig. 4 (c), stability with the proposed method changes according to the number of reconstructed frequency points  $N$ , and it is more stable than the standard reconstruction method, if  $N$  is smaller than 20.

We next applied the proposed method to recovering the distance to the vehicle in front by using narrow baseline stereos. The 3D measurements of narrow baseline stereos are in general very unstable, since disparities in narrow baseline stereos are very small relative to the image noise. Thus it is very important to improve the stability of reconstruction when using narrow baseline stereos. In this experiment, we show that by reducing the number of reconstructed frequency points in the proposed method, we can obtain stable 3D measurements, even if the baseline of the stereo camera is very small.

The baseline of the stereo camera was 5cm, and the distance to the vehicle in front changed from 30 m to a few meters. Fig. 5 (a) shows some of the sequential images used in these experiments. The image size was  $640 \times 480$  pixels. The left and right tail lamps





Figure 5: Real image experiment. The left and right tail lamps were used for reconstruction.

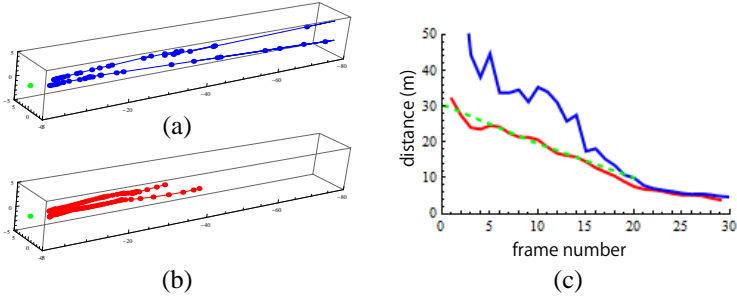


Figure 6: 3D points (tail lamps) reconstructed by the standard method (a) and the proposed method (b). (c) shows the accuracy of depth estimation. The red line shows the distance to the tail lamp recovered from the proposed method, and blue line shows that of the standard method. The green dashed line shows the ground truth.

were tracked with a correlation tracker. Fig. 5 (b) and (c) show the left and right stereo images. The blue points show feature points obtained with the tracker. The 30 synchronized frames were used in the standard method, and the unsynchronized 30 frames were used in the proposed method. Since the proposed method can control the highest frequency components of reconstruction freely, we set  $N = 15$  in the proposed method for increasing the stability of reconstruction.

Fig. 6 (a) shows the 3D points reconstructed by the standard method, and Fig. 6 (b) shows those from the proposed method. As seen in Fig. 6 (a), the 3D points reconstructed from the standard method are unevenly scattered over a large range, while the actual depth of these points is up to 30 m, and they are lined up evenly. Although the results from the standard method are extremely poor, they are reasonable since the baseline of the stereo camera is just 5 cm. However, the results from the proposed method are much more stable as seen in Fig. 6 (b).

For evaluating the accuracy of reconstruction numerically, we compared the depth of reconstructed points with their ground truth, which was obtained from a measuring device. Fig. 6 (c) shows the recovered distance of each 3D point and its ground truth. The horizontal axis shows the frame numbers of images, and the vertical axis shows the distance to the vehicle. The red line shows the results from the proposed method, and the blue line shows those from the standard reconstruction method. The green dashed line shows the ground truth of the vehicle distance at each image frame. As shown in this graph, the proposed method provides us fairly good results even if the baseline of the stereo cameras is very small, while the standard method fails at long distances.

## 7 Conclusion

In this paper, we proposed a method of recovering high frequency 3D motion from low frame rate images. Hence, super resolution 3D reconstruction could be obtained. In particular, we showed that by using a set of unsynchronized cameras, we can recover much more information on 3D motion, and can reconstruct more precise 3D motion than that with the standard synchronized cameras.

For reconstructing 3D motions from unsynchronized cameras, we introduced a new camera projection model, which projects 3D points in frequency space to 2D points in real space. By using the new projection model, we can observe identical 3D frequency points from different unsynchronized cameras, and hence the 3D points can be reconstructed without using any interpolation techniques. The proposed method can recover finer 3D motion as we increase the number of unsynchronized cameras. The efficiency of the proposed method was tested by using real image sequences.

## References

- [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] H. Fujiyoshi, S. Shimizu, Y. Nagasaka, and T. Takahashi. A method of pseudo stereo vision from images of cameras shutter timing adjusted. In *Proc. ACCV*, pages 443–450, 2004.
- [4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.
- [5] N. Hasler, B. Rosenhahn, T. Thormahlen, M. Wand, J. Gall, and H.P. Seidel. Markerless motion capture with unsynchronized moving cameras. In *Proc. CVPR*, 2009.
- [6] H.S. Park, T. Shiratori, I. Matthews, and Y. Sheikh. 3d reconstruction of a moving point from a series of 2d projections. In *Proc. ECCV, Part III*, pages 158–171, 2010.
- [7] A. Seki and M. Okutomi. Simultaneous optimization of structure and motion in dynamic scenes using unsynchronized stereo cameras. In *Proc. CVPR*, pages 1–8, 2007.
- [8] M. Svedman, L. Goncalves, N. Karlsson, M. Munich, and P. Pirjanian. Structure from stereo vision using unsynchronized cameras for simultaneous localization and mapping. In *Proc. International Conference on Intelligent Robots and Systems*, pages 3069 – 3074, 2005.
- [9] P. Tresadern and I. Reid. Uncalibrated and unsynchronized human motion capture: A stereo factorization approach. In *Proc. CVPR*, 2004.
- [10] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.
- [11] C. Zhou and H. Tao. Dynamic depth recovery from unsynchronized video streams. In *Proc. CVPR*, pages 351–358, 2003.