

# High-precision synchronization of video cameras using a single binary light source

Qi Zhao<sup>a</sup> and Yan Qiu Chen<sup>b</sup>

<sup>a</sup>Fudan University, School of Information Science and Engineering, Shanghai, China

<sup>b</sup>Fudan University, School of Computer Science, Shanghai, China

E-mail: cheniyq@fudan.edu.cn

**Abstract.** Camera synchronization is necessary for multicamera applications. We propose a simple and yet effective approach termed random on-off light source (ROOLS) to synchronize video sequences. It uses a single light source such as an LED to generate a random binary valued signal that is captured by the video cameras. The captured binary-valued sequences are then matched and the temporal offset of the cameras is computed up to subframe interval precision. We test the proposed method on synchronizing video sequences captured under a variety of illumination conditions and the results are verified against the ground truth provided by an LED array clock. The main contribution of the proposed method is that it reliably achieves high-precision synchronization at a low cost of only adding a simple light source. In addition, it is suited for synchronization in both laboratory and outdoor environments. © 2009 SPIE and IS&T. [DOI: 10.1117/1.3247860]

## 1 Introduction

For multicamera systems, synchronization is a must to provide accurate temporal correlation for incorporating image information from multiple viewpoints.

Synchronicity can be achieved through real-time hardware synchronization<sup>1</sup> or by establishing a time relationship between sequences recorded by unsynchronized video cameras.<sup>2</sup> While ensuring high precision synchronization, hardware solutions are costly and complex.

In a scenario where synchronous video sequences provided by hardware are not feasible, it is still possible to obtain synchronicity using image features.<sup>3-5</sup> These feature-based methods depend on the existence of salient and robust features in the scene. The failure of such features to exist in the scene and the error of detecting, tracking, and matching them would lead to incorrect synchronization.

In this paper, we present a simple and yet effective method, termed random on-off light source (ROOLS), to recover the temporal offset at subframe accuracy. It utilizes an auxiliary light source such as an LED to provide temporal cues. Compared to special-purpose hardware approaches, our method is far less complex and is inexpensive. Compared to feature-based approaches, ROOLS is more robust since it is completely independent of scene properties.

Paper 09041LRR received Apr. 3, 2009; revised manuscript received Aug. 25, 2009; accepted for publication Aug. 26, 2009; published online Oct. 20, 2009.

1017-9909/2009/18(4)/040501-3/\$25.00 © 2009 SPIE and IS&T.

## 2 Problem Statement

Without loss of generality, we consider the case of two video cameras. Let the time instances of the video frames taken by the  $\alpha$ 'th camera be denoted by

$$\mathbf{T}^\alpha = (t_1^\alpha, t_2^\alpha, \dots, t_{N_\alpha}^\alpha), \quad \alpha \in \{1, 2\}, \quad N_\alpha \in \mathbb{N}, \quad (1)$$

where  $N_\alpha$  denotes the length of the  $\alpha$ 'th sequence, and  $t_k^\alpha$  denotes the time of the  $k$ 'th frame in the  $\alpha$ 'th sequence. Note that  $\mathbf{T}^1$  and  $\mathbf{T}^2$  are measured by a common clock.

In a typical situation, identical video cameras of constant frame interval  $\Delta T$  are used, when

$$\mathbf{T}^\alpha = (t_1^\alpha, t_1^\alpha + \Delta T, \dots, t_1^\alpha + (N_\alpha - 1) \times \Delta T). \quad (2)$$

Synchronizing two video sequences in such situation is equivalent to measuring the temporal offset between their initial frames

$$T_{\text{diff}} = t_1^1 - t_1^2. \quad (3)$$

## 3 Proposed Method

### 3.1 Formulation

We propose to use a single temporally coded light source such as an LED as the signal to be captured by the cameras for synchronization. The light signal is essentially a time-continuous binary-valued function denoted as

$$f: R \mapsto \{0, 1\}. \quad (4)$$

It is sampled at  $\mathbf{T}^\alpha$  by the  $\alpha$ 'th camera, producing a time-discrete binary-valued sequence

$$[f_{\text{cam}}^\alpha(n)]_{n=1}^{N_\alpha} = [f(t_n^\alpha)]_{n=1}^{N_\alpha} = \{f[(n-1)\Delta T + t_1^\alpha]\}_{n=1}^{N_\alpha}. \quad (5)$$

Since  $f$  is binary valued, it can be characterized by time instants where its function value rises from 0 to 1 or drops from 1 to 0. We term each of these instants a transition event. Let  $\Phi$  denote a subsequence of all transition events in  $f$ ,

$$\Phi = (\phi_1, \phi_2, \dots, \phi_N). \quad (6)$$

For each  $\phi_k$ , we have

$$\forall \epsilon > 0, f(\phi_k - \epsilon) \oplus f(\phi_k + \epsilon) = 1, \quad (7)$$

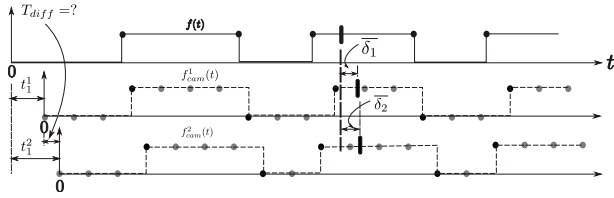
where  $\oplus$  denotes the exclusive or operator, and  $\epsilon$  is an arbitrarily small real positive number. For the  $\alpha$ 'th camera, let  $\Phi^\alpha$  denote the transition events corresponding to  $\Phi$ . Obviously, as part of  $\mathbf{T}^\alpha$ ,  $\Phi^\alpha$  can be expressed as follows:

$$\Phi^\alpha = [t_{\mathbf{I}^\alpha(1)}^\alpha, t_{\mathbf{I}^\alpha(2)}^\alpha, \dots, t_{\mathbf{I}^\alpha(N)}^\alpha], \quad (8)$$

where  $\mathbf{I}^\alpha$  is a subsequence of  $(1, 2, \dots, N_\alpha)$ . For each  $\mathbf{I}^\alpha(k)$ , it satisfies

$$f_{\text{cam}}^\alpha[\mathbf{I}^\alpha(k)] \oplus f_{\text{cam}}^\alpha[\mathbf{I}^\alpha(k) - 1] = 1. \quad (9)$$

This notation and their relationship are illustrated in Fig. 1.



**Fig. 1** Achieving subframe-interval estimation: (1) the transition events of  $f$ ,  $f_{cam}^1$ , and  $f_{cam}^2$  are highlighted by black dots; (2) black vertical bars denote the mean position of corresponding transition events in  $f$ ,  $f_{cam}^1$ , and  $f_{cam}^2$ .

### 3.2 Achieving Sub-Frame-Interval Precision

Given  $\Phi$  and  $\Phi^\alpha$ , consider the difference between a pair of corresponding transition events:

$$\delta_k^\alpha = \Phi_k^\alpha - \Phi_k = t_{I^\alpha(k)}^\alpha - \phi_k. \quad (10)$$

Provided that  $\delta_k^\alpha$  is an independent and identically distributed (i.i.d.) random variable uniformly distributed over  $[0, \Delta T]$  with mean and variance being  $\mu = \Delta T/2$  and  $\sigma^2 = \Delta T^2/12$ . The averaged difference  $\bar{\delta}_\alpha = 1/N \sum_{k=1}^N \delta_k^\alpha$  is a random variable with mean and variance being  $\bar{\mu} = \mu$  and  $\bar{\sigma}^2 = (1/N)\sigma^2$ . However,

$$\bar{\delta}_\alpha = \frac{1}{N} \sum_{k=1}^N [t_{I^\alpha(k)}^\alpha - \phi_k] = \bar{t}_1^\alpha - \bar{\phi}, \quad (11)$$

where  $\bar{t}_1^\alpha$  and  $\bar{\phi}$  are the mean positions of transition events with respect to  $f_{cam}^\alpha$  and  $f$ , as illustrated by Fig. 1. According to central limit theorem,<sup>6</sup> the averaged sum of a sufficiently large number of i.i.d. random variables each with finite mean and variance approximates normal distribution. Hence,  $\bar{\delta}_\alpha$  takes a normal distribution with mean of  $\mu$  and variance of  $\Delta T^2/12N$ . From Eq. (11) we obtain

$$\bar{t}_1^\alpha - \bar{t}_2^\alpha = \bar{\delta}_1 - \bar{\delta}_2. \quad (12)$$

Since  $\bar{t}_1^\alpha = (1/N) \sum_{k=1}^N \{t_1^\alpha + [I^\alpha(k) - 1]\}$ , Eq. (12) can be further written as

$$\delta' = T'_{diff} - \underbrace{\frac{1}{N} \sum_{k=1}^N [I^2(k) - I^1(k)]}_{\text{estimated temporal offset}}, \quad (13)$$

where  $T'_{diff} = T_{diff}/\Delta T$ ,  $\delta' = (\bar{\delta}_1 - \bar{\delta}_2)/\Delta T \sim \mathcal{N}(0, 1/6N)$ . When  $N$  is sufficiently large, the variance of  $\delta'$  will be negligibly small, leading to high precision estimation of  $T'_{diff}$ .

### 3.3 Transition Detection Accuracy

In real-world applications, the binary sequence is obtained through quantifying the image intensity of the light source by certain threshold  $\tau$ . For samples crossing transition events, the quantified binary value might flip, causing the transition event to shift one frame backward. If we take a sample right before the edge where signal rises from 0 to 1 for instance, its intensity is close to 1 and incorrectly quantified to 1. Equation (13) tells us that the shift will introduce

**Table 1** Results of 10 experiments.

Illum.	Estimated	Ground Truth	Error
Daylight	-3.410	-3.2653	0.1447
Daylight	0.6150	0.6000	0.0150
Daylight	0.5600	0.6122	0.0522
Fluorescent	-0.4900	-0.6122	0.1222
Fluorescent	0.1100	0	0.1100
Darkness	-2.9200	-2.9605	0.0405
Darkness	1.1150	1.1475	0.0325
Darkness	-2.8850	-2.7961	0.0889
Darkness	1.1900	1.3158	0.1258
Darkness	-2.7300	-2.6273	0.1027

additional error to the estimation. Suppose the light source intensity is normalized and  $\tau=0.5$  is chosen as the threshold so that the probabilities for transitions to flip from 0 to 1 and 1 to 0 are identical. Let  $x_i^\alpha$  denote a single shift event in a video sequence  $\alpha$ . Its probability density function (pdf) is  $p(x_i^\alpha = -1) = p(x_i^\alpha = 0) = 0.5$ , with expectation  $\mu_{x_i^\alpha} = -0.5$  and variance  $\sigma_{x_i^\alpha} = 0.25$ . Let  $\bar{x}^\alpha = (1/N) \sum_{i=1}^N x_i^\alpha$  denote averaged transition shift. Because  $x_i^\alpha$  are i.i.d. random variables, once again by the use of central limit theorem, we have  $\bar{x}^\alpha \sim \mathcal{N}(-0.5, 1/4N)$ . According to Eq. (13), the extra error  $\delta_s$  introduced by transition shift in two video sequences turns out to be  $\bar{x}^1 - \bar{x}^2$ . It can be proved that  $\delta_s$  takes a normal distribution with mean of 0 and variance of  $1/2N$ . Counting in  $\delta_s$ , the total error of temporal offset estimation is

$$\delta = \delta_s + \delta' \sim \mathcal{N}\left(0, \frac{1}{1.5N}\right). \quad (14)$$

### 3.4 Random Binary Sequence Design

The proposed method requires  $\delta_k^\alpha = t_{I^\alpha(k)}^\alpha - \phi_k$  to be i.i.d. random variables uniformly distributed in  $[0, \Delta T]$ . To achieve this, we set the transition time

$$\phi_k = \phi_{k-1} + \chi, \quad (15)$$

where  $\phi_{k-1}$  is the time of the previous transition and  $\chi$  is uniformly distributed in  $[\iota\Delta T, (\iota + \kappa)\Delta T]$ ,  $\iota, \kappa \in \mathbb{N}$ . Transition time generated in this way can be proved to ensure  $\delta_k^\alpha$  meeting the requirement.

### 3.5 Transition Matching

The estimation of the temporal offset requires the transition events of the two cameras be matched. We refer to this process as transition matching. Let the segment

**Table 2** Average temporal offset error of various approaches

	Method[3]	Method[4]	Method[5]	ROOLS
Average error	0.1	1	0.2	<b>0.08</b>

between two consecutive transition events be denoted as  $\mathbf{D}^\alpha(k)$ :

$$\mathbf{D}^\alpha(k) = \{f_{\text{cam}}^\alpha[\mathbf{I}^\alpha(k+1)] - f_{\text{cam}}^\alpha[\mathbf{I}^\alpha(k)]\} \times [\mathbf{I}^\alpha(k+1) - \mathbf{I}^\alpha(k)]. \quad (16)$$

The binary sequence  $[f_{\text{cam}}^\alpha(n)]_{n=1}^{N_\alpha}$  can be equivalently represented by a sequence of transition segments. Let  $\lambda(i, j) = |\mathbf{D}^1(i) - \mathbf{D}^2(j)|$  denote the difference between two segments. Transition matching can be equivalently achieved by matching two sequences of transition segment. Based on the observation that the difference of corresponding segments would be small, optimal transition matching can be obtained by solving the following formula:

$$\arg \min_{l,i,j} \left( \frac{\sum_{\gamma=0}^{l-1} \lambda(i+\gamma, j+\gamma)}{\max \left[ \sum_{\gamma=0}^{l-1} |\mathbf{D}^1(i+\gamma)|, \sum_{\gamma=0}^{l-1} |\mathbf{D}^2(j+\gamma)| \right]} + \exp \left\{ - \frac{\sum_{\gamma=0}^{l-1} |\mathbf{D}^1(i+\gamma)| + \sum_{\gamma=0}^{l-1} |\mathbf{D}^2(j+\gamma)|}{2 \max \left[ \sum_{k=1}^{M_1-1} |\mathbf{D}^1(k)|, \sum_{k=1}^{M_2-1} |\mathbf{D}^2(k)| \right]} \right\} \right), \quad (17)$$

where  $l$  denotes the number of overlapping segments, and  $M_1$  and  $M_2$  are the lengths of the two segment sequences. The first term assesses the similarity of two overlapping segment sequences. However, considering only the first term possibly leads to erroneous matching due to short overlapping length  $l$ . To avoid this, the second term is introduced to give a large penalty for small  $l$ . The optimal solution is sought by evaluating all combinations of  $i$  and  $j$ .

## 4 Experiments

### 4.1 Hardware and Configuration

The experiment system is made up of two Sony HVR-V1 high-definition (HD) video cameras, an LED array clock providing the ground truth, and a single temporally encoded LED light source. The video cameras operate at 200 frames per second. The values of  $\iota$  and  $\kappa$  in Eq. (15) are set to 2 and 6, which ensures there are at least two frames

between adjacent transition events and avoids ambiguity in transition matching. We selected  $\tau=0.5$  to quantize the image intensity of the LED a binary value.

### 4.2 Experiment Results

We conducted three groups of experiments under illumination conditions including daylight, fluorescent lighting, and darkness. The results are shown in Table 1. In all the tests, only 200 transition events were used. The average estimation error was about 0.08 frame intervals. We observe that all estimation errors are less than 0.2 frame intervals. This would be explained later.

### 4.3 Comparison with Other Methods

The proposed method was compared with the feature-based approaches<sup>3-5</sup> and the results are summarized in Table 2. The comparison indicates ROOLS achieves higher estimation accuracy than existing approaches.

### 4.4 Analysis and Discussions

The property of normal distribution states that 3 standard deviations from the mean account for about 99.7% of the distribution. When  $N=200$  transitions are used, according to Eq. (14), the standard deviation is about 0.05. The estimation error is bounded in  $3 \times 0.05=0.15$  frame intervals. This explains why the estimation error in Table 1 are bounded in 0.2 frame intervals. The performance of the proposed method can be improved by increasing  $N$ .

## 5 Conclusion

We presented an innovative approach toward synchronizing commercial video cameras. It achieved high-precision synchronization at the low cost of adding only a simple temporally coded light source. The proposed method requiring the video cameras to have identical frame rates is not a serious limitation since using identical video cameras for one task is convenient and typical.

### Acknowledgments

The research work presented in this paper is supported by National Natural Science Foundation of China, Grant No. 60875024.

### References

1. T. Kanade, H. Saito, and S. Vedula, *The 3D Room: Digitizing Time-Varying 3D Events by Synchronizing Multiple Video Streams*, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (1998).
2. A. Whitehead, R. Laganieri, and P. Bose, "Temporal synchronization of video sequences in theory and in practice," in *Proc. IEEE Workshop on Motion and Video Computing*, Vol. 2, pp. 132-137 (2005).
3. Y. Caspi and M. Irani, "Spatio-temporal alignment of sequences," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 1409-1424 (2002).
4. C. Rao, A. Gritai, M. Shah, and T. Syeda-Mahmood, "View-invariant alignment and matching of video sequences," in *Proc. 9th IEEE Int. Conf. on Computer Vision*, pp. 939-945 (2003).
5. S. N. Sinha and M. Pollefeys, "Synchronization and calibration of camera networks from silhouettes," in *Proc. 17th Int. Conf. on Pattern Recognition ICPR 2004*, Vol. 1, pp. 116-119, IEEE Computer Society, Washington, D.C. (2004).
6. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, New York, (1971).