Time-based Calibration: A Way to Ensure that Stitched Images are Captured Simultaneously

Ziwei Song*, Xiaowen Cai, Xiaoying Zhang, Jiaxiang Zong, Gangyong Jia

School of Compute Science and Technology, Hangzhou Dianzi University, China
{191050059, caixiaovn}@hdu.edu.cn, 13185092062@163.com, 928273034@qq.com, gangyong@hdu.edu.cn

Abstract

With the rapid development of modern science and technology, people's demand for such information as images and videos is also growing, and the requirements for image and video quality, clarity and view range are also increasing. Therefore, the construction of high-resolution, wide-view panoramic video has also gradually become a hot research topic. Video stitching technology is an increasingly popular research direction in the field of graphics, and the approach addresses the problem of a limited range of views due to a single device capture. Many researchers have proposed various algorithms for video stitching and achieved good stitching results, but the research on time synchronization calibration of different video sources is not yet well developed. This thesis proposes a multi-source video frames calibration technique based on external information sources for solving the problems of ghosting and cutting when stitching with different video sources. The proposed method calibrates the video stitching by introducing an information source and calculating the time difference of different devices. The error of the calibrated video stitching is less than 33 ms, which can guarantee the quality of the spliced video.

Keywords: Image stitching, Video stitching, Video time calibration

1 Introduction

With the development of technology, people's access to information is shifting from newspapers to images and videos. As people's demand for image and video information grows, so do their requirements for video quality, clarity and viewing range. Constructing high-resolution and wide-view videos has gradually become a hot research topic. Most of the current videos use a monocular camera as the recording device, but it has a limited field of view of the screen and low resolution. Although the fisheye camera has a wide field of view, due to the short focal length, there is a picture distortion of objects at the edge of the imaging field of view. And professional ultra-wide angle lens shooting equipment is expensive and cannot be applied to daily life.

Video stitching [1] capable of many spatially complex scenes. the main application scenarios of video stitching technology are intelligent monitoring [2], assisted driving [3], virtual reality [4], biomedicine [5], aerospace [6-7] and so on [8-9]. Video stitching is capable of processing videos with overlapping areas taken by multiple devices to generate a panoramic video with a wider perspective without reducing the video resolution, effectively solving the problem of not being able to capture large-view images, and providing people with richer information about the outside world. Therefore, the study of video stitching technology has theoretical and practical value.

In the study of video stitching, many methods use motion information of video frames to identify estimates to match different video frames. Yang et al [10] use the temporal context of video frames to identify moving objects and estimate their motion vectors, determine whether they can reach the seams, selectively update the seams, and avoid distortions when moving objects cross the seams. Yang [11] proposed a semantic segmentation-based dynamic video stitching algorithm that uses simultaneous image stabilization and stitching to obtain stable and seamless large-field stitching videos. Lan [12] considered the pixel differences on both sides of the image stitch line after stitching to make the transition of the stitched image at the stitch line more natural and reduce the appearance of texture breakage and ghosting. But those methods are computationally intensive and some of those cannot meet the requirements of video stitching for real-time. These methods are used after the video image has been shot to make the stitching more natural and do not take into account some of the processing that takes place before the video is shot.

For video stitching of video data from different devices, this paper proposes a calibration method based on external information sources is proposed. The method is based on an external time source to calibrate the multi-source video data, which solves the problem of unsynchronized different video data during video stitching. The research contribution of this paper is to provide a new time calibration method for video stitching, and the method calibrates the time difference of physical clocks of different devices by introducing external time source. It ensuring the video frames captured by all devices can be synchronized based on the calculated time difference.

2 Related Work

As time goes by, many methods have been proposed [13-15]. In video stitching from different devices, the video streams captured by different devices are affected by network delays, or different degrees of physical time shifts caused by the long running time of the device [16], making the video captured by different devices inconsistent in time, and this

*Corresponding Author: Ziwei Song; E-mail: 191050059@hdu.edu.cn
DOI: 10.53106/160792642022112306025

temporal inconsistency in the quality of video stitching is fatal damage. Time deviation in video stitching causes the images to be inaccurately matched, which lead to object ghosting and cutting in the stitching area. The key to video stitching is time synchronization calibration technology. Many researchers have proposed many video stitching calibration methods.

Caspi [17] proposed matching between image sequences using temporal and spatial information in 2000. Velipasalar et al [18] used the image of a video sequence to calculate the interframe offset in the video sequence by selecting the object in the foreground and using projection invariance to obtain its corresponding position in another sequence. Rao et al [19] used rank constraints on the corresponding points in two views to measure the similarity between trajectories. Wang et al [20] proposed an efficient algorithm to compute motion trajectories by the longest common subsequence, matching the video sequences obtained from asynchronous cameras.

Xu et al [21] extracted the feature parameters of the video for correlation calculation and judged the correspondence between video frames by the value of the correlation coefficient. Yin [22] used the method of motion detection to detect the amount of motion in the video at different times, and calculated the correlation coefficient at high motion points to determine the corresponding video frames at the same time point. Liu [23] used an affine transformation matrix combined with a moving object identification and estimation algorithm to amplify the temporal errors between different video sources by displacement differences, and used an inter-frame weighted interpolation algorithm for eliminating the temporal errors. Cao [24] proposed a video synchronization method based on temporal information, which can effectively calculate the time deviation to achieve synchronization by calculating the similarity of the contours of the moving objects extracted and applying the DTW algorithm to the matching calculation of video frames.

Although there are many calibration algorithms, they are all calibrated from the software side.

3 Video Frame Calibration Method

Some existing video stitching methods ignore the time difference of different devices and directly stitch the video captured by different devices, which is applicable to static scenes or videos that do not require high accuracy. This paper mainly explores how to improve the matching accuracy of video images from different devices, and proposes a time calibration method for the physical time difference of different devices, by introducing an external unified information source as the standard value, to calculate the time difference between the devices.

3.1 Traditional Stitching Method

Traditional multi-source video stitching is to integrate the images captured by cameras with overlapping areas into a wide-field image, and synchronize the calibration of different video sources by controlling the cameras to be turned on at the same moment. However, when fast moving objects appear in the scene, it will cause motion ghosting, object tearing, object blurring and cutting in the stitched video. They do not consider the error of different video sequences in time. Due to the difference of hardware, the camera on time of different video sources can not be turned on at the same moment, so in order to improve the quality of stitching must take into account the time information of video sequences.

As shown in Figure 1, video sequence A and video sequence B are different video sequences from different devices. Usually, there is a time interval between different video sources in the actual shooting. The error is not obvious when shooting a still scene, but when there are moving objects in the overlapping area, it can cause ghosting in the stitching result. The video sequences need to be calibrated so that images that are consistent above the time axis are used for stitching.

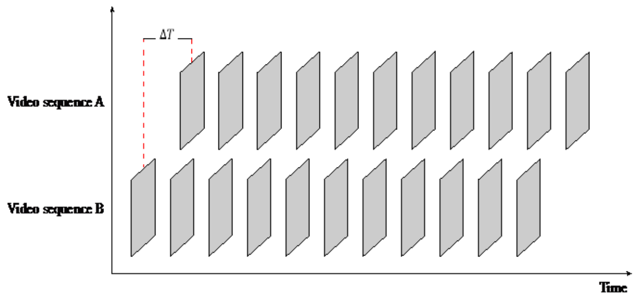


Figure 1. Video sequences with time lag

3.2 Overview of Video Frame Calibration Method

The time-based calibration is a calibration method before video stitching. Firstly, the cameras frome different devices are placed as shown in Figure 2, so that there are overlapping areas in their field of view.

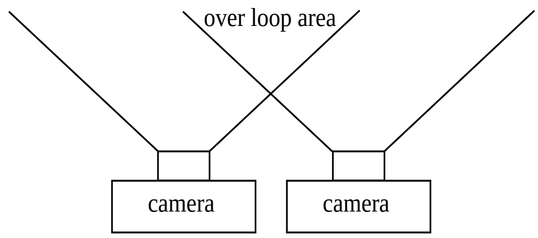


Figure 2. Camera position

The calibration method is to use different devices to capture the same information source at the same time, calculate the time difference of different devices according to the information captured in different video sequences, and use the calculated time difference to match the videos captured by different devices, and the matched video frames can be regarded as the video frames captured at the same moment after the time calibration. Its calibration method is shown below.

1) The calibration is performed by first setting a millisecond information source in the overlapping area of the camera. There is a number in the overlapping area which increases every 4 ms.

2) When the cameras from different devices capture the images of the information source, recording the local time when each frame of the corresponding video sequence was shot. At the same time, the different devices send the obtained video sequences and the recorded local time to server.

3) The server processes the collected images and local time records, assuming two sequences of images A, B. One image

is randomly selected from sequence A and sequence B, and the images are recognized using Optical Character Recognition (OCR) technology [25] for milliseconds t_A and t_B captured in the images, respectively. because the external information source changes every 4 ms, the difference time etween the two images $t = (t_A - t_B) \times 4$.

4) According to the recorded time, take the time obtained from the image at the source device at this time, subtract the time obtained from the 2 images at the source device and then subtract the above calculation to obtain t to obtain the time offset.

5) After the time offset is obtained, the calculation is repeated several times and the median of all the results is taken as the time difference between the two devices to calibrate.

3.3 Device Time Calibration Method Implementation

In this subsection, the proposed calibration method will be described in detail. The method calibrates video before stitching for videos using different devices, and the method calibrates video sequences from different systems by introducing an external time source, and the process is shown in Figure 3.

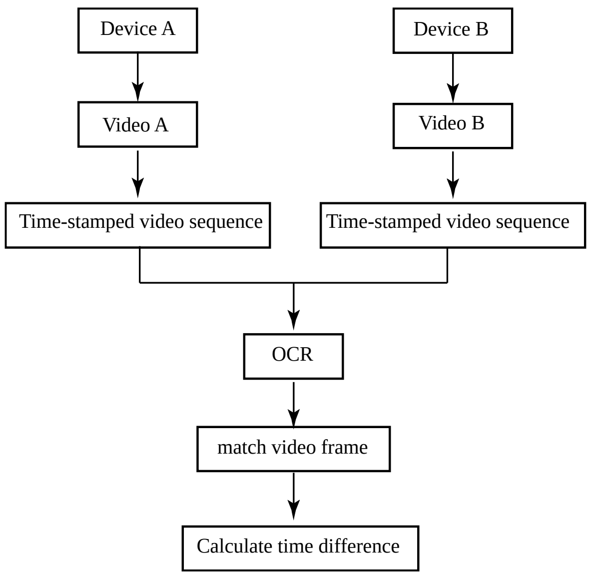


Figure 3. Overall process

3.3.1 Video Frame Matching for Different Devices

The proposed video stitching method with time-based video frame calibration is described in detail here.

1) Millisecond digital clock display

The millisecond information source captured in this paper is programmed to display a millisecond time. This information source is displayed on a display in real time and increases every 4 ms.

2) Digital recognition

For video sequences obtained in different devices, recognition techniques needed to applied to recognize the time of the milliseconds number. Among them, OCR is a text recognition technology. OCR is used to recognize the text of each frame captured by the device to obtain digital information.

3) Matching video frames

After using OCR technology to identify the digital information in the video frame images. First, two frames are randomly selected from video sequences A and sequence B from two different devices, noted as A_i and B_j, while the device times $L(A_i)$ and $L(B_j)$ at the time the two frames were captured are taken out. Then the images are used separately for OCR text recognition technology to identify the millisecond information sources t_A and t_B captured in the images, and the two images differ in time $t = (t_A - t_B) \times 4$. The difference between the device time and the captured time is made to finally arrive at the time difference between the two frames captured before calibration, which is calculated as follows.

$$\Delta T = \left(L(A_i) - L(B_j) - t\right). \quad \textbf{(1)}$$

3.3.2 Video Frame Matching for Different Devices

The refresh rate of the display used in the experiment was 60 Hz, which indicates that the display was refreshed every 16.67 ms, and the frame rate of the camera equipment used was 30 fps, which means that an image was taken every 33.33 ms. In the ideal case, if the cameras, and the screen are started at the same time, the camera has performed the capture of the picture at every two times the display is refreshed by the camera. Then, there will be no time interval for different video sequences when video stitching.

However, in the actual shooting process, there are fluctuations in both the screen refresh and camera shots, this leads to a time interval between the screen display refreshing and the camera capturing the image.When a device takes a picture of the screen, if the screen of the display has not been refreshed, the information captured may be the same time image as the previous device, but the actual physical time of the capture is inconsistent. In this case, the time difference between the acquired image information and the calculated time is inconsistent, making the results fluctuate.

As shown in Figure 4, ideally, there will be no errors between all images after calibration. However, the shooting interval of the camera and the refresh interval of the screen fluctuate to a certain extent, which makes the calibration method have a certain error. In the extreme case, the camera has taken a shot when the display is about to refresh and the other camera takes a shot after the display, the calculated physical time difference will have an extra error of 16.66 ms. Therefore, there are camera fluctuations in the case of error within 16.66 ms. When there is a screen refresh fluctuation, because the shooting interval of two cameras is 16.66 ms, when the time difference between the two cameras is close to 16.66 ms, if there is a screen refresh fluctuation at this time, it will lead the image captured by one camera is a previous frame, and the image captured by the other camera is the next frame, which makes their calculated time difference is increased by 33.33 ms.

Therefore, when using this method for calibration, the calibration will fluctuate within [0, 33.33], which is acceptable for video stitching.

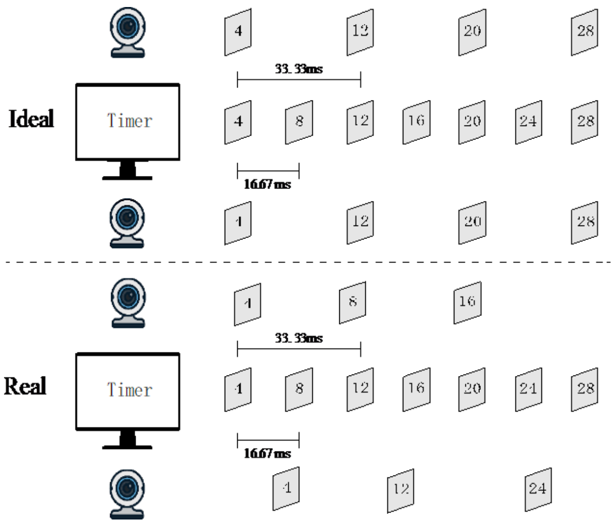


Figure 4. Error range of the algorithm

3.3.3 The Impact of Video Ghosting and Processing

According to the steps of Equation (1) to calibrate the video frames from diderent device, the calibration results are shown in Figure 5, and it is found that there are some points of anomalous cases in the calibration results. This is because the external source of information has been changing rapidly, making the camera imaging some of the video frames appear ghosting. This leads to errors in the identification of temporal information in the video image, which prevents accurate matching of video frames, and these video frames need to be excluded at the time of matching.

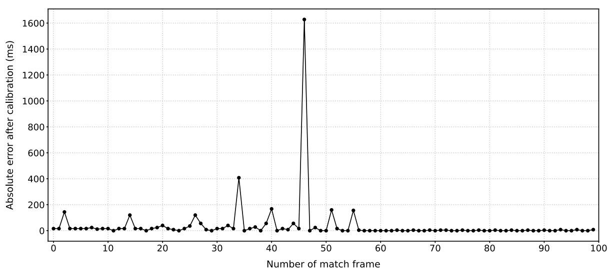


Figure 5. Matching results with ghost images

The camera frame rate used in the experiment is 30fps, so the interval between each frame should be between 32-36 ms. The digital timer on the display is self-increasing every 4 ms, so the difference between the numbers of adjacent frames is in the range 7-9. After the OCR digital recognition is performed, it is judged whether the time difference between the currently detected value and the valid values detected in the previous frame and the next frame is within a fixed interval. As shown in Table 1, starting from the captured image recognition results, there are several data anomalies representing the problems caused by digital conversion overlap in images as follows.

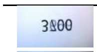
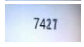
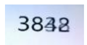
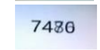
According to Table 1, the situation of ghosting is analyzed and summarized, and Equation (2) and (3) is designed to filte the recognition result, if the calculation result satisfies the Equation (2) and (3), it is considered that the recognition error is caused by shooting ghosting. And this video image will not taken into acount.

$$t_i - t_{i-1} > 9 \; or \; t_i - t_{i-1} < 7. \tag{2}$$

$$t_{i+1} - t_i > 9 \text{ or } t_{i+1} - t_i < 7. \tag{3}$$

where the t_i represent the time that recognized by OCR.

Table 1. Identification of error cases

Video frame	OCR result	Difference with previous frame	Difference with next frame
	3300	-392	409
	7427	24	3
	3838	4	48
	106738	9	-8
	7480	-2	8

The calculation results after the error data elimination are shown in Figure 6. It shows that after filting the anomalies images, the fluctuation of the absolute value of the time difference of the device can be kept within 33 ms.

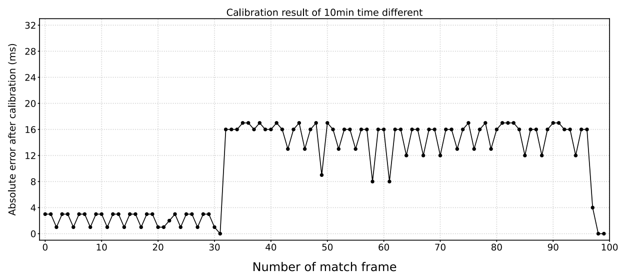


Figure 6. Calibration time difference error after eliminating ghosting frames

4 Analysis of Experimental Results

4.1 Experimental Results

In this subsection, experiments are designed to test the calibration method for different devices, in which two aoni cameras are used to capture video sequences with a camera frame rate of 30 fps captured with an image resolution of 680 x 480. Two desktop computers with 2.50 GHz and 3.20 GHz main frequencies were used as different devices, and a desktop computer with 3.4 GHz main frequency as the server. The calibration results its results are shown in Figure 7.

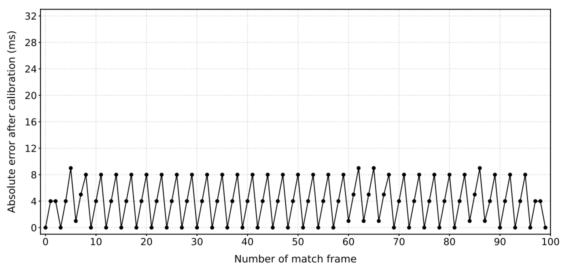


Figure 7. Calibration results

It can be found that the proposed calibration method in this paper has a good performance, and the errors after calibration all fluctuate within the range of 33 ms, which can meet the needs of video stitching.

4.2 The Impact of Video Ghosting and Processing

To find the effect of different video duration on stitching, videos of 10s, 20s, 30s, 40s, and 50s were taken, and 300, 600, 900, 1200, and 1500 images were obtained, and the time difference between the two devices was calculated.

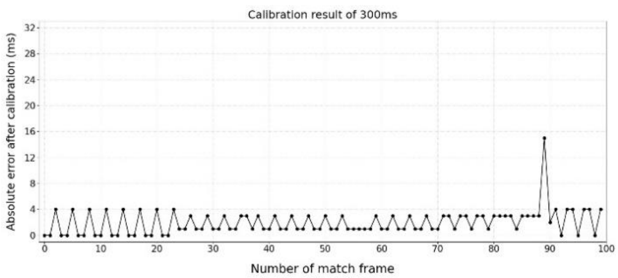
This experiment is mainly processed by shooting different lengths of video corresponding to different frame rates, and in order to ensure the consistency of the experiment, when obtaining each set of experimental data, the camera will be rebooted. the wrong data were filted from the token video data, and the remaining valid data in the two devices were randomly combined to calculate the time difference according to Equation (1). Then, the median of all time difference values is used as the device time difference., and the median of the time differences obtained in each group is shown in Table 2.

The matching results are shown in Figure 8. The figure shows the results for 100 random sets of data in each experiment. From the figure, it can be seen that the time difference error range is guaranteed to be no more than 33 ms when changing the number of frames taken, whether it is 300, 600, 900, 1200, and 1500 video frames.

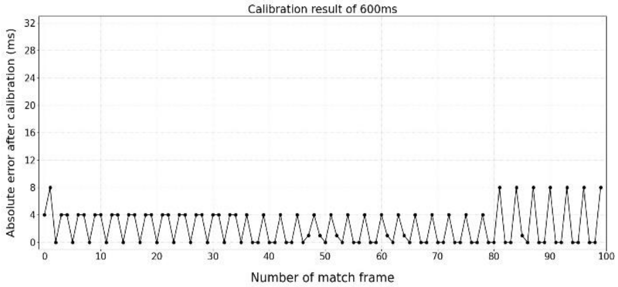
Table 2. Median of time difference calculated for various frames of the same device

Video frames	Device time difference (ms)
300	20433
600	20433
900	20430
1200	20434
1500	20431

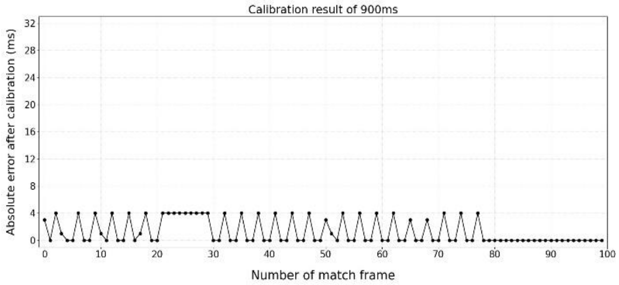
The average value of time difference error fluctuation for 300 video frames is 1.8 ms; the average value of time difference error fluctuation for 600 video frames is 2.2 ms; the average value of time difference error fluctuation for 900 video frames is 1.6 ms; the average value of time difference error fluctuation for 1200 video frames is 4.5 ms; the average value of time difference error fluctuation for 1500 video frames is 4.4 ms. And for each video length, the average value of the absolute value of the error fluctuation is within 5 ms for each group. The experiment shows that the time calibration method can keep the fluctuation range of its time difference within 33 ms under the condition of different number of video frames.



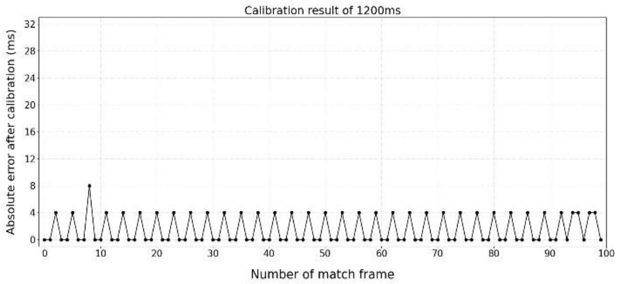
(a) Calibration of 300 ms



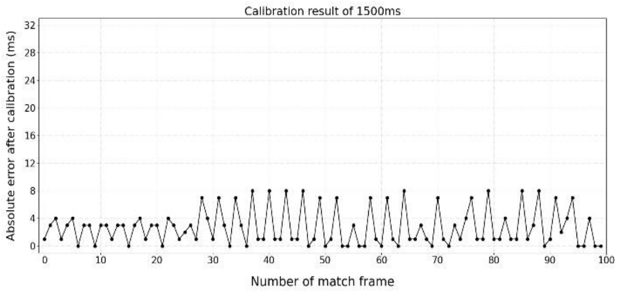
(b) Calibration of 600 ms



(c) Calibration of 900 ms



(d) Calibration of 1200 ms



(e) Calibration of 1500 ms

Figure 8. Time calibration results at different video length

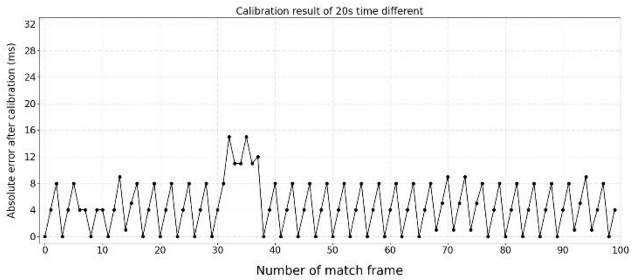
4.3 Effect of Different Physical Time Differences

There are different time intervals between different devices, and the same device has a time slip in its physical time at different time periods. In this experiment, the time difference variables of the two devices were adjusted to about 20 s, 2 min, 5 min and 10 min, and 600 video frames were taken by each group. Calibrate the time of each group.

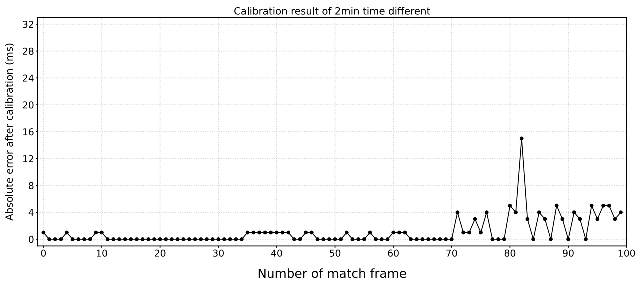
Each set of 100 data was finally selected randomly and the difference between the calculated results and the median shows the distribution of the absolute value of the error. The results are shown in Figure 9 and Table 3. For devices with time difference of 20 s, 2 min, 5 min and 10 min, the increasing of video time will not cause the calibration error to increase, and the time difference error range can be guaranteed not to exceed 33 ms.

Table 3. Calibration error of different video length

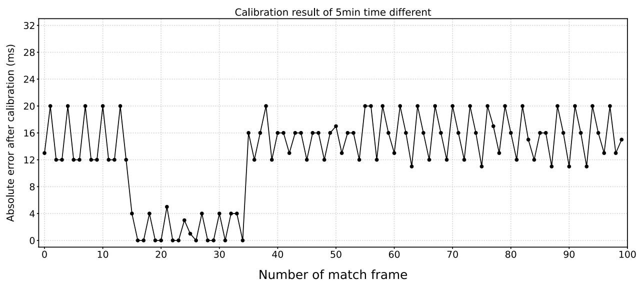
Video length	Calibration error
20 s	8.3 ms
2 min	6 ms
5 min	4.5 ms
10 min	5.5 ms



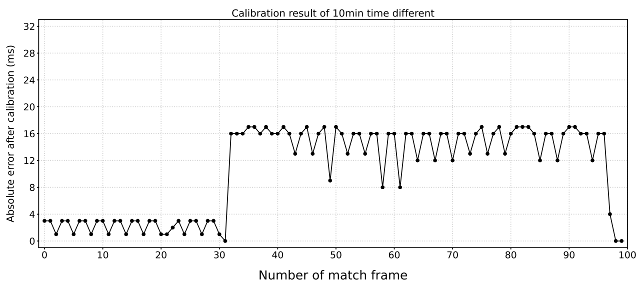
(a) Calibration of 20 s time different



(b) Calibration of 2 min time different



(c) Calibration of 5 min time different



(d) Calibration of 10 min time different

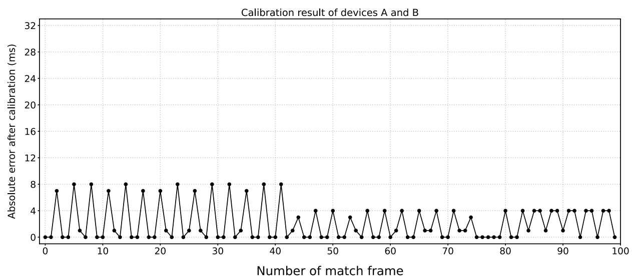
Figure 9. Time calibration results for different devices with physical time differences

4.4 Impact of Different Equipments

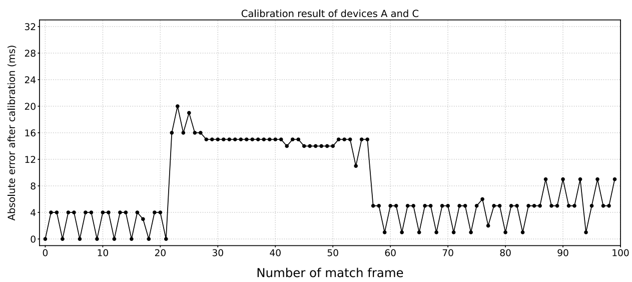
The above experiments were conducted on the same two devices. In order to verify the robustness and applicability of the proposed method, this experiment was conducted using different devices to calibration. Therefore, in order to avoid the difference between the devices, the next two calibrations were performed using three devices to check whether the error of the experimental results would change. In the experiment, device A is a desktop computer with a CPU frequency of 2.50 GHz and 24 GB memory; device B is a desktop computer with a CPU frequency of 3.20 GHz and 16 GB memory; device C is a laptop with a CPU frequency of 3.60 GHz and 16 GB memory.

Each set of 100 data was finally selected randomly and the difference between the calculated results and the median shows the distribution of the absolute value of the error. The results are shown in Figure 10. It can be seen that after the time calibration of different devices, the fluctuation range of the calculated time difference still does not exceed 33 ms. The average value of the time difference fluctuation between device A and device B is 2.2 ms; the average value of the time difference fluctuation between device A and device C is 7.5 ms; the average value of the time difference fluctuation between device B and device C is 5.8 ms.

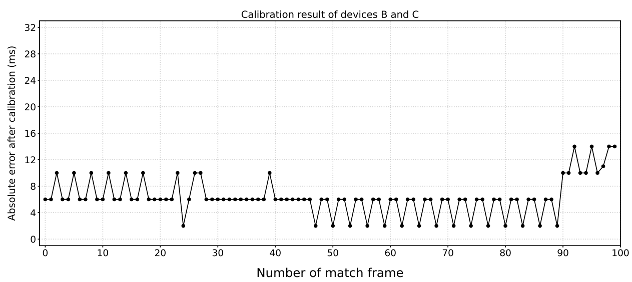
It can be found that the average value of the absolute value of the error in each group is within 8 ms. The experiment also confirms that the time calibration method is capable of performing a reasonable calibration between different devices, ensuring that the fluctuation range of the calculated time difference does not exceed 33 ms.



(a) Calibration of devices A and B



(b) Calibration of devices A and C



(c) Calibration of devices B and C

Figure 10. Time calibration results for different equipment cases

5 Conclusion

In this paper, a calibration method for video stitching using videos from different devices is proposed. Before executing the video stitching algorithm, an external information is introduced as a benchmark, and the real time difference between devices is calculated by capturing the content of a constantly refreshed digital text on a display. The calculated time difference is used to match the videos captured by different devices. Then, a theoretical fluctuation range of the calculated results is analyzed. Finally, the robustness of the method is experimentally verified by setting different influencing factors.

Acknowledgement

The work was supported by the National Key Research and Development Program under Grant No. 2019YFC0118404, National Natural Science Foundation of China under 455 Grant No. U20A20386, Zhejiang Key Research and Development Program under Grant No.2020C01050, the Key Laboratory fund general project under Grant No. 6142110190406, the key open project of 32 CETC under Grant No. 22060207026.

References

[1] Q. Liu, X. Su, L. Zhang, H. Huang, Panoramic video stitching of dual cameras based on spatio-temporal seam optimization, *Multimedia Tools and Applications*, Vol. 79, No. 5-6, pp. 3107-3124, February, 2020.

[2] S. Vasuhi, A. Samydurai, M. Vijayakumar, Multicamera Video Stitching for Multiple Human Tracking, *International Journal of Computer Vision and Image Processing*, Vol. 11, No. 1, pp. 17-38, January-March, 2021.

[3] J. Chalfoun, M. Majurski, T. Blattner, W. Keyrouz, P. Bajcsy, M. Brady, MIST: Microscopy Image Stitching Tool, *2015 IEEE International Conference on Bioinformatics and Biomedicine*, Washington, DC, USA, 2015, pp. 1757-1757.

[4] P. Fuchs, G. Moreau, P. Guitton, *Virtual reality: concepts and technologies*, CRC Press, 2019.

[5] R. Aktar, V. H. Huxley, G. Guidoboni, H. AliAkbarpour, F. Bunyak, K. Palaniappan, Mosaicing of dynamic mesentery video with gradient blending, *2020 IEEE International Conference on Image Processing*, Abu Dhabi, United Arab Emirates, 2020, pp. 563-567.

[6] X. Wang, The application of video image stitching technology in the intelligent monitoring system of flight area, *Electronic Technology & Software Engineering*, Vol. 18, No. 2, pp. 99-100, September, 2016. http://www.cnki.com.cn/Article/CJFDTotal-DZRU201618081.htm

[7] J. Wang, Y. Zhang, J. Lu, B. Zhou, W. Xu, Y. Li, Feature-based mosaic method for aerial video sequences, *Journal of PLA University of Science and Technology (Natural Science Edition)*, Vol. 13, No. 2, pp. 159-164, April, 2012.

[8] H. Gao, J. Xiao, Y. Yin, T. Liu, J. Shi, A Mutually Supervised Graph Attention Network for Few-shot Segmentation: The Perspective of Fully Utilizing Limited Samples, *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, pp. 1-13, 2022, DOI: 10.1109/TNNLS.2022.3155486

[9] H. Gao, K. Xu, M. Cao, J. Xiao, Q. Xu, Y. Yin, The Deep Features and Attention Mechanism-Based Method to Dish Healthcare Under Social IoT Systems: An Empirical Study With a Hand-Deep Local–Global Net, *IEEE Transactions on Computational Social Systems*, Vol. 9, No. 1, pp. 336-347, February, 2022.

[10] D. Yang, J. Chen, X. Zhu, Real-time Video Stitching Algorithm Based on Time-Series Features Prediction, *Journal of Network New Media Technology*, Vol. 10, No. 4, pp. 32-38, July, 2021. https://www.cnki.com.cn/Article/CJFDTotal-WJSY202104005.htm

[11] X. Yang, *Research on dynamic video image stitching technology in large parallax scene*, Master's thesis, China University of Mining and Technology, Jiangsu, China, 2020.

[12] X. Lan, *Research on key technologies of panoramic video stitching*, Master's thesis, University of Electronic Science and Technology of China, Sichuan, China, 2020.

[13] H. Gao, W. Huang, T. Liu, Y. Yin, Y. Li, PPO2: Location Privacy-Oriented Task Offloading to Edge Computing using Reinforcement Learning for Intelligent Autonomous Transport Systems, *IEEE Transactions on Intelligent Transportation Systems (T-ITS)*, pp. 1-14, 2022, DOI: 10.1109/TITS.2022.3169421.

[14] H. Gao, B. Qiu, R. J. D. Barroso, W. Hussain, Y. Xu, X. Wang, TSMAE: A Novel Anomaly Detection Approach for Internet of Things Time Series Data Using Memory-Augmented Autoencoder, *IEEE Transactions on Network Science and Engineering (TNSE)*, pp. 1-14, 2022, DOI: 10.1109/TNSE.2022.3163144.

[15] Y. Yin, Q. Huang, H. Gao, Y. Xu, Personalized APIs Recommendation with Cognitive Knowledge Mining for Industrial Systems, *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 9, pp. 6153-6161, September, 2021.

[16] Z. Wang, *Design and realization of network clock synchronization algorithm evaluation system based on clock calibration*, Master's thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2018.

[17] Y. Caspi, M. Irani, A step towards sequence-to-sequence alignment, *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000*, Hilton Head Island, SC, USA, 2000, pp. 682-689.

[18] S. Velipasalar, W. Wolf, Frame-level temporal calibration of video sequences from unsynchronized cameras by using projective invariants, *IEEE Conference on Advanced Video and Signal Based Surveillance, 2005*, Como, Italy, 2005, pp. 462-467.

[19] C. Rao, A. Gritai, M. Shah, T. Syeda-Mahmood, View-invariant alignment and matching of video sequences, *IEEE International Conference on. Computer Vision*, Nice, France, 2003, pp. 939-939.

[20] Y. Wang, S. Velipasalar, Frame-level temporal calibration of unsynchronized cameras by using Longest Consecutive Common Subsequence, *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, China, 2009, pp. 813-816.

[21] W. Xu, C. Feng, J. Li, L. Huang, Y. Yao, A new algorithm for frame alignment in video quality evaluation, *Journal of Xiamen University (Natural Science)*, Vol, 51, No. 2, pp 185-188, February, 2012.

[22] X. Yin, Frame alignment method for video mosaic based on relative coefficient, *Computer Engineering and Design*, Vol. 33, No. 3, pp. 1037-1041, March, 2012.

[23] J. Liu, *Study on video Mosaic with synchronic adjustment technology*, Master's thesis, Tianjin University, Tianjin, China, 2012.

[24] S. Cao, *Research on object tracking based on multi camera stitched video*, University of Electronic Science and Technology of China, Sichuan, China, 2017.

[25] N. V. Rao, A. S. C. S. Sastry, A. S. N. Chakravarthy, P. Kalyanchakravarthi, Optical Character Recognition Technique Algorithms, *Journal of Theoretical and Applied Information Technology*, Vol. 83, No. 2, pp. 275-282, January, 2016.

Biographies

Ziwei Song gets master degree from Hangzhou Dianzi University. Now is a PhD student in Yamanashi University. His research interserts include Computer Vision, object detection and Deep learning and so on.

Xiaowen Cai, a master student in the Department of Computer, Hangzhou Dianzi University. And her research interests include unmanned systems, intelligent transportation systems and so on.

Xiaoying Zhang, major in computer science and technology at univercity, a master student in the College of Computer and software, Hangzhou Dianzi University. Her current research mainly includes edge computing and cloud computing.

Jiaxiang Zong, major in computer science and technology at university, a master student in the College of Computer and software, Hangzhou Dianzi University. His research mainly includes edge computing and object detection.

Gangyong Jia got PhD in Computer System Architecture from the University of Electronic Science and Technology of China. He is now working at Hangzhou Dianzi University, and his main research interests are operating systems, edge intelligence and so on.