

Research on multimedia network audio stream synchronization algorithm based on recognition and detection

Li Ruicheng

Hainan College of Software Technology, Qionghai. Hainan, 276400

Keywords: multimedia network; audio synchronization; audio fingerprinting

Abstract. This study is focused on audio streaming synchronization problems of multimedia network. Audio streaming synchronization issues of multimedia network directly affect the development of multimedia technology. To this end, this paper presents a multimedia network audio stream synchronization method based on the whole node identification, with feature detection method to identify endpoint of audio streams in perception node, and an audio fingerprint is conducted for information of each node, then, mark the starting point of sync, and sync tasks are decomposed at each node, and finally the audio stream synchronization is achieved in the cluster head node, error correction adjustment and information correction are performed to reduce the energy consumption of cluster head node, enhance correction capability of information loss. The simulation results show that the synchronization delay of proposed algorithm is significantly reduced, which enhance the performance of the multimedia network.

Introduction

Multimedia network becomes an important branch of Internet of things for its rich sensation information and flexible deployment [1]. In recent years, with the cost of audio sensing components and CMOS camera hardware gradually reducing, multimedia networks is developing rapidly, and widely used in all areas of society, has a huge space for development [2,3]. Multimedia network has big data volume, high time relevant, and strict requirement for real-time, therefore, the synchronization of multimedia data is the basis of network collaboration, accurate data synchronization can choose the best media streaming, suppress redundant information and reduce energy consumption, valid correction is processed for packet loss information of multimedia network [4-6].

For the problem that low error correction capability for information loss of the nodes in multimedia network, this paper presents a multimedia network audio synchronization technology based on all node identification, with feature detection method to identify endpoint of audio streams in perception node, and an audio fingerprint is conducted for information of each node, then, mark the starting point of sync, and sync tasks are decomposed at each node, and finally the audio stream synchronization is achieved in the cluster head node, error correction adjustment and information correction are performed, the proposed method have higher accuracy for synchronization, which can effectively reduce the energy consumption of cluster head node, and the effect for synchronization precision of internet packet loss.

The principle of audio stream synchronization method based on task decomposition

Multimedia network often compose network by the approach of tuft or tree, if all the nodes of collected data event are part of the same nodes cluster, the nodes of the cluster head performs all the synchronization behavior, if not in the same node cluster, hierarchical operation will be performed for node cluster and synchronized process is followed, the study object of this paper is focused on the synchronization structure of the same cluster.

With different distance from the node to the sound source in a multimedia network, the propagation delay will seriously affect the synchronization of audio stream. Strong correlation process of audio stream signal is utilized to deal with the audio flow problems due to propagation delay, with the correlation of the multi-channel audio streams from the same sound source can

process signal loss, noise and other problems similarly. Audio fingerprint is constructed after endpoint detection for the audio stream, the synchronized starting point of fingerprint is marked, and the audio fingerprint is adjusted at the cluster head, this method first to achieve a coarse synchronization, as long as implementing a simple adjustment in the cluster head node in the end, audio stream synchronization can be achieved. Thus, synchronization task shared by the nodes and cluster head, effectively reducing the cluster head node energy consumption and prolong the life cycle of the multimedia network.

Endpoint detection of the audio stream.

By detection of the endpoint of the audio stream can determine starting point of the audio, first buffer and sub-frame are processed for captured audio, the frame length can be set by the parameter, through the comparison of extracted characteristics and threshold of self-adaptive to detect the endpoint information of audio stream, build audio fingerprinting, and determine the starting point for audio synchronization. Endpoint detection is a key step for pretreatment, time domain analysis is utilized to analyze the characteristics of the audio signal, framing process is performed after audio stream buffer in the node, so as to form the collection. After extracting features, thresholds are compared to form endpoint information and construct an audio fingerprint.

Zero-crossing rate is calculated as follows:

$$Z = \frac{1}{2} \sum_{n=2}^N |\text{sgn}(S_n) - \text{sgn}(S_{n-1})| \tag{1}$$

Wherein, S_n is the sample value, N is the number of samples.

The absolute value of sampling point within a frame of audio represents short-term energy, its value can be expressed as follows:

$$E = \sum_{n=1}^N |S_n| \tag{2}$$

Zero-crossing rate and audio short-term energy threshold can be expressed as:

$$Z_{th} = \text{Max}(Z) \times \alpha \tag{3}$$

$$E_{th} = (\text{max}(E) - \text{min}(E)) / \beta + \text{min}(E) \tag{4}$$

Among them, α, β are dynamic factors, the maximum short-term energy-frame in the sequence of consecutive frames is determined firstly, and gradually the frame is divided into smaller frames, the maximum sample value in the sequence of sub-frames in defined as the endpoint. Sub-frame length is set to 5ms, because the too small sub-frame size will increase the load of a multimedia network node, an excessive load will affect the synchronization accuracy of audio stream, the endpoint of audio is defined as starting point of synchronization, so as to achieve coarse synchronization of audio stream in the cluster head.

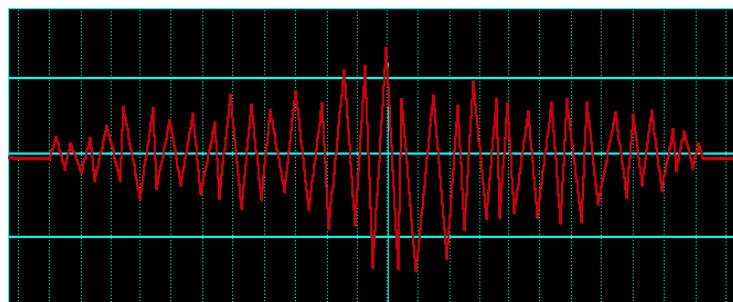


Figure 1 detection figure of endpoint

During the audio stream transmission of different nodes, there will be some delay when different node arrives the cluster head, chaotic situation of adjacent points will be produced by node. Construction of the audio fingerprint for analyzing, the amplitude of the audio is averaged as M aliquot, the ratio of number of the aliquot sampling points and the overall number of sampling

points are normalized to obtain amplitude ratio variable, thereby completing the endpoint detection.

Achievement of audio streaming synchronization.

Sync header structure is designed, data synchronization is achieved in the cluster head. In multimedia networks, because of low bandwidth and data burst will generate a lot of packet loss, audio streaming synchronization techniques can correct packet loss effectively, synchronization header composed of the leader sequence, sequence number, an audio fingerprint, leader sequence tag data packet and sequence number can correct fingerprint information effectively under the packet loss situation, the fingerprint information is able to identify the different endpoints. As shown below, leader sequence in the three audio segments of message is set to be same with fingerprint parameters, the serial number is incrementing. First, the audio fingerprint are compared to find the synchronization symbol, followed by the alignment operation according to the same sequence number, the cluster head receives the audio stream of multimedia network node 1 and the multimedia network node 3, the alignment standard is sequence 1, but two-way audio appears without audio fingerprints of same sequence, then the sequence number of the header is adopted to detect the synchronized audio stream effectively.

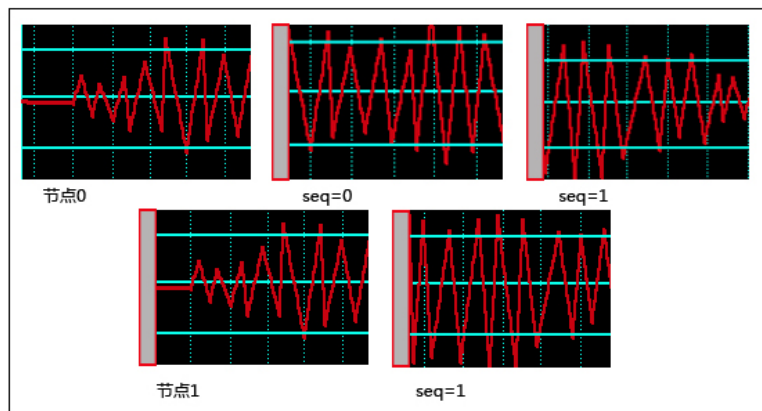


Figure 2 parsing map of synchronous packet header

All audio streams stored in the cluster head buffer, with the following methods to identify synchronized starting point of audio streams:

$$\left| R_i^m / R_i^n - 1 \right| \leq \zeta \tag{5}$$

$$\left| R_j^m / R_j^n - 1 \right| \leq \zeta \tag{6}$$

R_i^m in the equation above represents the magnitude ratio of the audio stream in m -th node of the multimedia network at position i , first to calculate the audio frame sequence magnitude near synchronized starting point of each audio stream, the Euclidean distance between different audio streams is calculated, searching the minimum Euclidean distance to obtain the synchronized starting point. When aiming at characteristic change of the audio signal, self- adaptive adjustment for the parameters of endpoint detection can have good results. When the number of nodes detected is not same, it is easy to find out which way the signal is weak, so that you can effectively avoid the confusion caused by inaccurate endpoint, α , β value can be adjusted to increase the value of Z_{th}, E_{th} , in order to perform error adjustment and information correction. If the detected packet loss rate is too high, the parameters α , β can be adjusted to enable the detection for more endpoints, reducing the impact of packet loss on data synchronization.

Simulation Analysis

In order to verify the performance of the synchronization of the proposed algorithm, it is added in the actual multimedia network, audio sensors composed of Atmel AT91SAM7X256 processor chip, microphone, radio equipment, this processor has 64KB of SRAM, has sufficient bearing

capacity for the algorithm. Based on CC2430 chip to build multimedia networks, sensors and ordinary collection nodes have the same hardware configuration.

During the experiment, the paper conducted test for different sound sources, the duration of each sound source is one minute, the sampling frequency is 8Khz, the length of the audio frame is 5ms. The traditional method is to detect the starting point of every word, multi-task decomposition algorithm in this paper have more than one word existing in an audio sequence, less number of endpoint need to be detected, the comparison of synchronization performance of the proposed algorithm and the traditional algorithm is as follow:

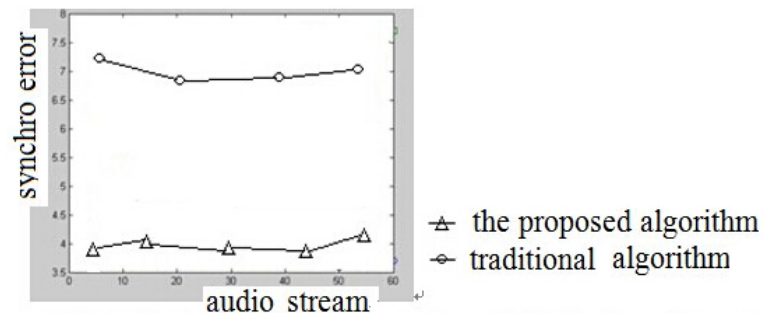


Figure 3 comparison of synchronization performance under the packet loss situation

The figure shows, the network synchronization delay is reduced a lot by adopting the proposed algorithm compared to the traditional algorithm, which still remains near 3.5ms, but the conventional method have longer delay and poor performance.

Conclusion

Aiming at the centralized synchronization process which is used for audio stream synchronization method of multimedia network, all audio stream synchronization are processed by the cluster head node, which have the drawback like energy consumption, thus an audio streaming synchronization method of multimedia network based on a full node identification is proposed. A synchronization header is designed for the data packet of each node, all nodes carrying synchronization information achieves audio synchronization in the cluster head. The simulation results show that the synchronization delay of proposed algorithm is significantly reduced, which enhance the performance of the multimedia network.

References

- [1] A. Ben Ali, I. Lengliz and F. Kamoun, "Congestion control for multimedia flows in multihop wireless networks: case of chain topology". In Proc. of International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, 2009:188-193.
- [2] BOW YER K W. CHANG K. FLYNN P. A survey of approaches and challenges in 3D And multimodal 2d+3d face recognition [J]. Computer vision and image understanding 2006 101(1):1-15.
- [3] Zhang Hongli. A Data Management Framework Based on MA for Wireless Sensor Network [J]. Journal of Tongling College, 2009.8(5):67-69.
- [4] Guo Xudong, Feng Ping, Kang Jichang, Xu Wei. Research and Implementation of Communication Performance Optimization Based on FC Switch Network [J]. Computer measurement and control, 2009.1:135-137.
- [5] Wang Fubao, Shi Long, Ren Fengyuan. Self-Localization Systems and Algorithms for Wireless Sensor Networks [J]. Journal of software, 2005.16:857-868.
- [6] Kim, H. Cost-effective and dispersion-tolerant cascade modulation formillimeter-wave-over-fiber applications. Optics Communications, 2008, 281:1108-1112.