ANALYSIS

37(168), September 1, 2015

ISSN 2278–5469 EISSN 2278–5450

Discovery

Tracking objects using multiple acoustic video systems

Syed Navaz AS[1], Ravishankar S[2], Sudarson M[3]

1. Asst.Professor, Department of Computer Application, Muthayammal.College of Arts & Science, Namakkal, India; Email:- a.s.syednawaz@gmail.com
2. Asst.Professor, Department of Computer Science, K.S.Rangasamy College of Arts & Science, Namakkal, India; Email:- ravirohan83@gmail.com
3. Asst.Professor, Department of B.Tech IT, K.S.Rangasamy College of Engineering, Namakkal, India; Email:- sudarson86@gmail.com

Publication History

Received: 09 June 2015

Accepted: 24 July 2015

Published: 1 September 2015

Citation

Syed Navaz AS, Ravishankar S, Sudarson M. Tracking objects using Multiple Acoustic Video Systems. *Discovery*, 2015, 37(168), 59-64

ABSTRACT

To impart the academic knowledge gain through this two years course to practical application was the main objective to do this paper. The Paper entitled "Tracking objects using Multiple Acoustic Video Systems" Is to track the objects Using joint acoustic and video system measurements, we propose a particle filtering solution that can handle multiple sensor modalities. We formulate the tracking problem using a particle filter based on a state-space approach. We first discuss the acoustic state-space formulation whose observations use a sliding window of direction-of-arrival estimates. We then present the video state space that tracks a target's position on the image plane based on online adaptive appearance models. The joint operation of the filter, we combine the state Vectors of the individual modalities and also introduce a time-delay Variable to handle the acoustic-video data synchronization issue, Caused by acoustic propagation delays. The Kullback-Leibler divergence measure, it is shown that the joint operation of the filter decreases the Worst case divergence of the individual modalities. The resulting Joint tracking filter is quite robust against video and acoustic occlusions Due to our proposal strategy. Computer simulations are presented with synthetic and field data to demonstrate the filters Performance.

Index Terms: Multiple sensor, Acoustic, Image plane, Kullback-Leibler, Joint tracking filter

1. INTRODUCTION

Audio and video signals originating from the same source tend to be related. To achieve optimal performance, a tracking system must exploit not just the statistics of each modality alone, but also relationships between the two. Currently, most crowded, public locations are monitored by dozens, sometimes even hundreds of cameras. Airports, shopping malls, casinos, stores, and even traffic intersections are populated with digital imaging equipment, usually for security, but sometimes for marketing or simple behavioral observation. Since the cost of this infrastructure has dropped significantly with the introduction of digital video systems, their proliferation has exploded beyond the point where humans can realistically observe all feeds concurrently.

Recently hybrid nodes that contain an acoustic array collocated with a camera was proposed for vehicle tracking problems. To intelligently fuse information coming from both modalities, novel strategies for detection and data association have to be developed to exploit the multimodal information. Moreover, the fused tracking system should be able to sequentially update the joint state vector that consists of multiple target motion parameters and relevant features (e.g., shape, color and so on), which is usually only partially observable by each modality. It is well known that acoustic and video measurements are complementary modalities for object tracking. Individually, the acoustic sensors can detect targets, regardless of the bearing with low power consumption, and the video sensors can provide reliable high-resolution localization estimates, regardless of the target range, with high power consumption.

2. LITERATURE REVIEW

In tracking objects, fusing multi-modal sensor data under a power-performance is becoming increasingly important. Proper fusion of multiple modalities can help in achieving better tracking performance while decreasing the total power consumption. In this paper, we present a framework for tracking a target given joint acoustic and video observations from a co-located acoustic array and a video camera. We demonstrate on field data that tracking of the direction-of-arrival of a target improves significantly when the video information is incorporated at time instants when the acoustic signal-to-noise ratio is low.

Tracking people in known environments has recently become an active area of research in computer vision. Several person tracking systems have been developed to detect the number of people present as well as their 3D position over time. These systems generally use a combination of foreground/background classification, clustering of novel points, and trajectory estimation in one or more camera views. It uses a Kullback filter for tracking multiple talkers using a microphone array. There has been less work done in the audio-visual tracking domain. This showed that by using a particle filter, sound and vision can be used effectively to achieve a more robust tracking of a single object than any of the modalities on their own.

When tracking multiple people, we have found that rendering an orthographic vertical projection of detected foreground pixels is a useful representation. A “plan view” image facilitates correspondence in time since only 2D search is required. Previous systems would segment foreground data into regions prior to projecting into a plan view, followed by region-level tracking and integration, potentially leading to sub-optimal segmentation and/or object fragmentation. Instead, we developed a technique that altogether avoids any early segmentation of foreground data. We merge the plan-view images from each view and estimate over time a set of trajectories that best represents the integrated foreground density. Trajectory estimation is performed by finding connected components in a spatio temporal filtered volume.

Localizing and tracking speakers in enclosed spaces using AV information has increasingly attracted attention in signal processing and computer vision given the complementary characteristics of each modality. Broadly speaking, the differences among existing works arise from the overall goal (tracking single vs. multiple speakers), the specific Detection/tracking framework and the AV sensor configuration. Much work has concentrated on the single-speaker case, assuming either single-person scenes, or multi-person scenes where only the location of the current speaker is tracked. Many of these works use simple sensor configurations (e.g. one camera and a microphone pair). Among the existing techniques, probabilistic generative models based on exact or approximate inference methods (both variation and sampling-based) appear to be the most promising, given their principled formulation and demonstrated performance.

In fact, although audio-based multi-speaker tracking and vision-based multi-person tracking have been studied for a few years as separate problems in signal processing, and computer vision, respectively, the AV multi-speaker tracking problem has been studied relatively recently, making use of more complex sensor configurations. While single cameras are useful for remote conferencing applications, multi-person conversational settings like meetings often call for the use of multiple cameras and microphones to cover an entire workspace (table, whiteboards, etc.).

More specifically, Cutler et al. described a system based on a device that integrates a small circular microphone array and several calibrated cameras, whose views are composed into a panorama. The tracking system, in which each person is tracked independently, consists of three modules: AV auto-initialization, (using either a standard acoustic source localization algorithm or visual cues), visual tracking using a Hidden Markov Model (HMM), and tracking verification.

Kapralos et al. described a non-probabilistic multi-speaker detection algorithm using an omni-directional camera (which has limitations of resolution) and a microphone array, calibrated with respect to each other. At each video frame, the method extracts skin-color blobs by traditional techniques, and then detects a sound source using standard beam forming on the small set of directions indicated by the skin-blob locations.

Chen and Rui used the same calibrated sensor setup as Cutler et al, and tracked multiple speakers with a set of independent PFs, one for each person. Each PF uses a mixture proposal distribution, in which the mixture components are derived from the output of single-cue trackers (based on audio, color, and shape information). This proposal increases robustness in case of tracking failures in single modalities. As we describe in the remainder of the paper, our work substantially differs from previous work in AV multi speaker tracking with respect to the choices for the multi-person dynamical model, the AV observation model, and the sampling mechanism.

Building on the model recently proposed by Khan et al, our model has two advantages over the works of Check a et al and Chen and Rui. First, unlike both and, we use a multi person dynamical model that explicitly incorporates a pairwise person interaction prior term. This model is especially useful to handle person occlusion. Second, unlike and, we use efficient MCMC sampling techniques that allow to jointly track several objects in a tractable manner (effectively close to the case of independent PFs), while preserving the rigorous joint state-space formulation.

R. Cutler and L. Davis, an extensive amount of research has investigated the topic of sensor networks for acoustic target localization. The types of acoustic sensors vary – some of the sensors are able to detect the distance to an acoustic target, while others can detect the direction to an acoustic target, and still others assume only acoustic volume is detectable. Some techniques involve the use of a known strength of signal at the source in order to estimate the distances to the target and combine them in real time. The approach in assumes a time synchronized microphone array and uses time difference of arrival between microphones and microphone arrays to localize targets. In all of these cases, the sensor net integrates the detected information to localize the target.

Our real world constraints also prevent us from using a known energy output level from the target, both because of practical issues of varying sound levels at the source and declining microphone sensitivity as battery power declines. Sophisticated algorithms for target localization under these constraints have been developed in, which uses mathematical analysis to estimate the target position based upon the strengths of the signals heard by neighboring nodes. However, these analytical models were designed for open air environments where acoustic signal propagation models are known.

Murphy's Trulla et al, since this modeling is not relevant to our application, it is not meaningful to try to predict a target's future path. Our approach, therefore, is to plan a path to the current estimation of the target position, and then dynamically replan the path as the target position changes. While an extensive number of path planning techniques are possible, we use a dual wavefront propagation algorithm to plan a path to the target position. Several variations of the wavefront propagation algorithm have been implemented. Our approach is similar to the work of Behring, *et al*.

Malaka Walpola et al, in the localization problem is analyzed and geometric dilution of precision (GDOP) and normalized geometric dilution of precision (NGDOP) measures are introduced and analyzed for maximum likelihood estimation. A node selection method that selects the best three nodes that minimize the GDOP measure is developed. In a Kalman based global node selection method is introduced and a decentralized extended Kalman filter based tracking method is integrated to perform the target tracking. In a more energy efficient variant of above tracking algorithm, a combination of a local node selection method and a decentralized extended Kullback filter based tracking is developed.

3. EXISTING SYSTEM

Early work in this area focused to fuse information coming from both modalities has been applied to problems such as tracking of humans under surveillance and smart videoconferencing. Typically, the sensors are a video camera and an acoustic array (not necessarily collocated). The acoustic time-delay-of arrivals derived from the peaks of the generalized cross-correlation function, are used along with active contours to achieve robust speaker tracking with fast lock recovery. The tracking humans using audio visual cues, based on foreground detection and image-differencing. The visual appearance models should be calculated online as opposed to using trained models for tracking. Although fixed image templates (e.g., frames) are very useful for face tracking, they are not effective for tracking vehicles in outdoor environments. In vehicle-tracking problems, acoustics and video synchronization causes biased localization estimates that can lead to filter divergence. This is because the bias in the fused cost function increases the video's susceptibility to drift in the background. Previous implementations make use of the basic bootstrap particle filter, whereas a more general approach involves the concept of importance sampling.

None of the above works can handle the problems are continuously inferring, from audio and video data, the location and speaking status for several people in a realistic conversational setting. Many of these works use simple sensor configurations (e.g. one camera and a microphone pair). Among the existing techniques, probabilistic generative models based on approximate inference methods (both variation and sampling-based) appear to be the most promising, given their principled formulation and demonstrated performance.

4. PROPOSED SYSTEM

The hybrid nodes that contain an acoustic array collocated with a camera was proposed for vehicle tracking problems. To intelligently fuse information coming from both modalities, novel strategies for detection and data association have to be developed to exploit the multimodal information. Moreover, the fused tracking system should be able to sequentially update the joint state vector that consists of multiple target motion parameters and relevant features (e.g., shape, color and so on), which is usually only partially observable by each modality. By using particle filters, whose proposal function uses audio cues, have better speaker tracking performance under visual occlusions.

To track vehicles using acoustic and video measurements, we propose a particle filtering solution that can handle multiple sensor modalities. We use a fully joint tracker, which combines the video particle filter tracker and a modified implementation of the acoustic particle filter tracker at the state-space level. We emphasize that combining the output of two particle filters is different from formulating one fully joint filter or one interacting filter.

The resulting filter has a lower Kullback–Leibler distance to the true target posterior than any output combination of the individual filters. Hence, by fusing the acoustic and video modalities,

- Achieve tracking robustness at low acoustic range
- Improve target confirmation.

5. SYSTEM DESCRIPTION

Tracking objects using Multiple Acoustic Video Systems is a System to track the objects using joint acoustic and video system measurements; We formulate the tracking problem using a particle filter based on a state-space approach. A particle filtering solution that can handle multiple sensor modalities.

The acoustic state space formulation whose observations use a sliding window of direction-of-arrival estimates. Then we present the video state space that tracks a target's position on the image plane based on online adaptive appearance models. The joint operation of the filter, we combine the state Vectors of the individual modalities and also introduce a time-delay. Variable to handle the acoustic-video data synchronization issue. A novel particle filter proposal Strategy for joint state-space tracking. Using the Kullback-Leibler divergence measure, that the joint operation of the filter decreases the Worst case divergence of the individual modalities. The resulting Joint tracking filter is quite robust against video and acoustic occlusions. Computer simulations are Presented with synthetic and field data to demonstrate the filters Performance.

6. SYSTEM ANALYSIS

System analysis can be defined, as a method that is determined to use the resources, machine in the best manner and perform tasks to meet the information needs of an organization.

System Description & Modules

Preprocessing

1. Frame Extraction
2. Edge Detection

In this module, we convert the input video into the frames, which means we extract the frames from the images, which is converted from the input video. And we use the image enhancement to convert images into the frames. The edges can be detected from the frames, which are extracted from the videos. Detection of edges is based on the grayscale. Grayscale is used to convert the images into black and white. Canny edge algorithm is used to detect the edges.

Object detection

1. Object Specification
2. Object Tracking

In this module, we are going to find out the moving actions. This moving action can be detected from the input given. Input is given in AVI format; it extracts the images from AVI and also extracts the frames from the images. From the images, we are going to find moving object.

Classification

1. Acoustic Processing
2. Target Fixing
3. Target with Sound

In this module, we are going to find out the audios. This audio action can be detected from the input given. Input is given in AVI format; it extracts the audios from AVI.

Report Module

In this module we have given the report for the previous module shows that the video tracking (Moving object) in a pixel format and the audio tracking in a graphical format.

7. IMPLEMENTATION

Implementation is the most crucial stage in achieving a successful system and giving the user's confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one. This type of conversation is relatively easy to handle, provide there are no major changes in the system.

Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the programs specification, the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user. And so the system is going to be implemented very soon. A simple operating procedure is included so that the user can understand the different functions clearly and quickly.

Initially as a first step the executable form of the application is to be created and loaded in the common server machine which is accessible to the entire user and the server is to be connected to a network. The final stage is to document the entire system which provides components and the operating procedures of the system.

Comparative Study

In the existing system, the Target tracking of humans under surveillance and smart video conferencing. Typically, the sensor are a video camera and acoustic array, In vehicle-tracking problems, acoustics and video a synchronization causes biased localization estimates that can lead to filter divergence. So they are not effective for tracking vehicles in outdoor environments. In the proposed system, the hybrid nodes that contain an acoustic array collocated with a camera was proposed for vehicle tracking problems. Hence the proposed system would help the tracking robustness at low acoustic range and Improve target confirmation.

8. CONCLUSION

We presented a particle filter tracker that can exploit acoustic and video observations for target tracking by merging different state-space models that overlap on a common parameter. By the construction of its proposal function, the filter mechanics render the particle filter robust against target occlusions in either modality, when used with Huber's robust statistics criterion function. The presented filter also demonstrates a scheme for adaptive time-synchronization of the multimodal data for parameter estimation. The time-delay variable is incorporated into the filter and is modeled as multiplicative. It is the authors' observation that without the time-delay variable, the joint filter is susceptible to divergence.

FUTURE ISSUES

We presented a probabilistic framework for the joint tracking of multiple people and their speaking activity in a multi-sensor meeting environment. Our framework integrated a novel AV observation model, a principled mechanism to represent proximity-based interactions, and an efficient sampling strategy that overcomes some of the problems faced by traditional PFs in high-dimensional state-spaces. In principle, the sensor calibration algorithm we defined puts few constraints on the sensors' location, so cameras and microphones could potentially be placed in various configurations. Several issues remain open. First, more refined interaction models could be proposed, making use of the speaking activity variable in the MRF prior, and introducing an occlusion variable in the state-space, which could explicitly define a set of switching occlusion modes. Second, although our model can reflect simultaneous speaking activity from multiple people, it is based on a limiting single-audio-source assumption. We are currently developing truly multi-speaker detection techniques and plan to integrate them in our framework in the future. Third, the auto-initialization mechanism could be enhanced by using audio-based localization and/or face detection, whose integration in the MCMC-PF is conceptually direct. Finally, the evaluation on more dynamic data, including more complex cases of object birth/death, is also part of future work. A particle filter based tracking algorithm is used for additional performance fine-tuning to obtain high quality tracking.

REFERENCE

1. Leichter, M. Lindenbaum, and E. Rivlin, "A probabilistic framework for combining tracking algorithms," in *Proc. CVPR 2004*, Jun./Jul. 2004.
2. V. Cevher and J. H. McClellan, "Proposal strategies for joint state space tracking with particle filters," in *Proc. ICASSP 2005*, Philadelphia, PA, Mar. 18–23, 2005.
3. S. Mallat and S. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1998.
4. V. Cevher and J. H. McClellan, "An acoustic multiple target tracker," in *Proc. IEEE SSP 2005*, Bordeaux, France, Jul. 17–20, 2005.
5. Y. Zhou, P. C. Yip, and H. Leung, "Tracking the direction-of arrival of multiple moving targets by passive arrays: Algorithm," *IEEE Trans.Signal Process.*, vol. 47, no. 10, pp. 2655–2666, Oct. 1999.
6. V. Cevher and J. H. McClellan, "General direction-of-arrival tracking with acoustic nodes," *IEEE Trans. Signal Process.* vol. 53, no. 1, pp. 1–12, Jan. 2005.