A Study on Automatic Color Object Learning and Detection through Acoustic Instructions

Kishor S Jeve
Department of Computer Science,
College of Computer Science and
Information Technology,
Latur

Ashok Gaikwad
Vivekanand College,
Aurangabad

Pravin Yannawar
Department of C.S. & IT,
Dr. Babasaheb Ambedkar Marathwada
University,
Aurangabad

Abstract: - **Now days, there is increasing need of automated video analysis. The proposed system receives acoustic instruction as an input and analyzes the video frames and outputs the location of a moving object within the video frames. This can be viewed as segmenting an object of interest from a video sequence and keeping track of its direction, motion, shape, scale, and occlusion etc. and extract useful information using acoustic instructions. Its main task is to find and track a moving object or several objects in video sequences or image sequences using audio instructions. The objective is to use Automatic color object learning and detection through acoustic instructions (ACOLDAI) for Motion-based recognition of object which has wide range of real-time applications.**

Keywords: Shape, Motion, Scale, Occlusion, Acoustic Instructions.

1. INTRODUCTION:

The aim of ACOLDAI is to estimate the locations of an object in a video sequence using acoustic instructions. Humans perform object detection and recognition effortlessly and instantaneously. An algorithmic description of such functions for implementation on machines has been very difficult. Apart of its difficulty, ACOLDAI has a variety of applications such as Motion-based recognition of human, Human-computer interaction, automated video surveillance, Robot vision, Traffic monitoring Animation, Government or military establishments, vehicle navigation and so on [1][2].

The proposed system is just a combination of audio or speech recognition system and object tracking system. Existing object tracking algorithms can be categorized as either generative algorithms or discriminative algorithms [3]. Generative tracking algorithms [3][4][6] typically learn a target model to represent a target and use the model to search for interesting regions in the next frame with maximum similarity. Discriminative approaches [3][5] construct a model to represent the appearance of a target, and find the decision boundary that best separates the target from the background.

Audio or Speech recognition is used to communicate with a machine using acoustic instruction. It uses algorithms to convert speech or voice signals into a sequence of words or other linguistic units. Automatic Color Object Learning and Detection through Audio Instructions is surprisingly difficult. The proposed method will be applied to a wide range of fields, such as image processing, intelligent machines, multimedia systems, industry production, and military affairs and so on. Accordingly, it is of great real significance and application value to investigate in object tracking.

2. RELATED WORK:

A review of related work on speech recognition with its recognition rate or performance is shown in following table 2.1. A review of related work on video processing with its recognition rate or performance is shown in following table 2.2

TABLE 2.1 A REVIEW OF RELATED WORK ON SPEECH RECOGNITION WITH ITS RECOGNITION RATE OR PERFORMANCE

Year	Topic Name	Researcher Name	Method	Recognition Rate
1996	An Improved Training Algorithm in HMM-based Speech Recognition[7]	Gongjun Li and Taiyi Huang	HMM	Close-Set:96.86% Open-Set:84.93%
1997	HMM-Based Speech Recognition Using State-Dependent, Discriminatively Derived Transforms on Mel-Warped DFT Features[8]	Rathinavelu Chengalvarayan, and Li Deng	HMM and Mel-Warped DFT Features	82.2%

Published by :

http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 04, April-2017

Year	Topic Name	Researcher Name	Method	Recognition Rate
2013	Speaker Recognition System Based On MFCC and DCT [9]	Garima Vyas, Barkha Kumari	MFCC and DCT	99.5%
2013	Speech Recognition and Verification Using MFCC & VQ [10]	Kashyap Patel, R.K. Prasad	MFCC & VQ	87%
2014	Enhancing Speech Recognition Using Improved Particle Swarm Optimization Based Hidden Markov Model[11]	Lokesh Selvaraj and Balakrishnan Ganesan	IP-HMM	97.14%

.

TABLE 2.2 A REVIEW OF RELATED WORK ON VIDEO PROCESSING WITH ITS RECOGNITION RATE OR PERFORMANCE

Year	Topic Name	Researcher Name	Method	Recognition Rate/ Performance
2002	Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering[12]	Lars Bretzner, Ivan Laptev and Tony Lindeberg	Multi-Scale Colour Features, Hierarchical Models and Particle Filtering	No colour prior: 45% colour prior: 86.5%
2014	Compressed-Domain Video Retargeting[13]	Jiangyang Zhang, Shangwen Li,and C.-C. Jay Kuo	compressed-domain video retargeting system	94.81%
2009	Adaptive Mean-Shift Tracking with Auxiliary Particles[14]	Junqiu Wang and Yasushi Yagi	Adaptive Mean Shift	64.11%
2009	Robust Object Tracking Using Joint Color-Texture Histogram[15]	Jifeng Ning	Joint Color-Texture Histogram ,Mean Shift, Local Binary Pattern	Mean Error: JCTH :8.22 Mean Shift:2.83 LBP:10.78
2012	Shape Adaptive Mean shift Object Tracking Using Gaussian Mixture Model[16]	Katharina Quast and Andr´e Kaup	GMM-SAMT	98.49%
2013	Robust Object Tracking via Active Feature Selection[17]	Kaihua Zhang et al	Active Feature Selection	83%
2013	Fast Tracking via Spatio-TemporalContext Learning[18]	Kaihua Zhanga et al	Spatio -Temporal Context	94%

3. ACOLDAI FRAMEWORK

ACOLDAI consist of combination of speech processing and video processing modules as shown in fig 3.1. In which speech recognition module receives percept through voice commands, detect an object in the video and track the object throughout the video.

- *Learning based approaches:* In this approach learning of speech templates or words involves artificial neural network and genetic algorithm based learning.
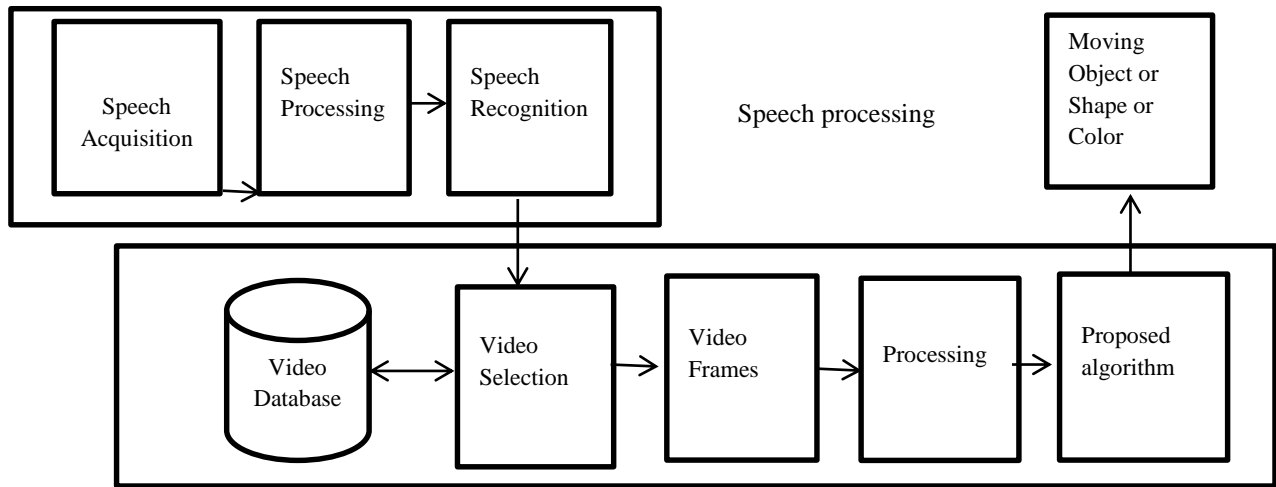
Figure 3.1 General Framework of ACOLDAI

3.1 Speech Recognition:

The aim of speech recognition system is to receive percept, understand and perform the action according to speech information. The sample voice instructions are shown in Table 3.1.1. The different approaches for speech recognition are as follows,

- *Pattern Recognition approach: The* pattern recognition approach involves two essential steps, namely, pattern training and pattern matching.

- *Template-based approaches*: Template-based approaches are used to match unknown speech templates or words against a set of pre-recorded templates or words and find the best match.

- *Statistical based approaches:* Speech templates or sequences are modeled using statistical learning algorithm such as the Hidden Markov Models, or HMM.

- *Dynamic time warping:* This is an algorithm for measuring similarity between two speech templates or sequences which may vary in time or speed.

- *Knowledge based approaches:* An expert knowledge about variations in speech is hand coded into a system. This has the advantage of explicit modeling variations in speech; but unfortunately such expert knowledge is difficult to obtain and use successfully.

The steps in Audio instruction based tracking process are as follows,

Step 1: Speech Acquisition: speech samples (name of the object or color) are obtained from the speaker and store in memory for processing.

Step 2: Speech Preprocessing: preprocess the signal to remove noise, background voice etc.

Step 3: Speech Recognition: Use one of the models (Given in methods) for training and recognition.

Step 4: Speech Text: convert speech in to text.

TABLE 3.1.1 SAMPLE VOICE INSTRUCTION

Sr. No	Voice Instruction	Video Processing
1	Find	Find specific object.
2	Locate	Locate object or to detect particular position of the object.
3	Find "Blue" or "Red" or "Green"	Detect and Track color "Blue" or "Red" or "Green" (specified by Voice Instruction)
4	Find Ellipse or Triangle or Circle or Rectangle or any Shape	Detect and Track Ellipse or Triangle or Circle or Rectangle or any Shape
5	Find any object such as Man, Car, Animal etc.	Detect and Track any object such as Man, Car, Animal etc.

3.2Video processing:

The different approaches for object detection in video processing are as follows,

•*Contour-based object tracking model*: Contour-based object tracking model is used for finding object outline from an image. In the contour-based tracking algorithm, the objects are tracked by considering their outlines as boundary contours.

•*Region-based object tracking model*: The region based object model bases it's tracking of objects on the color distribution of the tracked object. It represents the object based on the color. Hence, it is computationally efficient.

•*Feature-point based tracking algorithm*: In Feature point based model feature points is used to describe the objects such as color, Shape, texture etc.

The steps for object tracking in videos are as follows,

Step 5: Video Selection: Select the video for processing or tracking.

Step 6: Divide the video in to the frames.

Step 7: Preprocess the video frames.

Step 8: Detect the object or color in video based on object or color defined in speech text. (Use one of the method listed in following section)

Step 9: Object Tracking: Track the object through the video

4.CONCLUSION:

The proposed model of ACOLDAI tracks color object accurately based on audio instructions. This model will used for fraud detection, Motion-based recognition of human, Human-computer interaction, automated video surveillance, Robot vision, Traffic monitoring Animation, Government or military establishments, vehicle navigation and so on.

5.REFERENCES:

[1] K. Cannons. A review of visual tracking. Dept. Comput. Sci. Eng., York Univ., Toronto, Canada, Tech. Rep. CSE-2008-07, 2008.

[2] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah. Visual tracking: An experimental survey. PAMI, 36(7):1442–1468, 2014.

[3] J. Ning, J. Yang, S. Jiang, L. Zhang and M-H Yang, "Visual Tracking via Dual Linear Structured SVM and Explicit Feature Map," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016).

[4] X. Mei and H. Ling. Robust visual tracking and vehicle classification via sparse representation. PAMI, 33(11):2259–2272, 2011.

[5] S. Avidan. Ensemble tracking. PAMI, 29(2):61–271, 2007

[6] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Scale and Orientation Adaptive Mean Shift Tracking," IET Computer Vision, vol. 6, no.1, pp. 62-69, 2012

[7] Li, G.; Huang, T, "An improved training algorithm in HMM-based speech recognition", In Proceedings of the 4th International Conference on Spoken Language Processing, Philadelphia,PA, USA, October 3–6, 1996; Volume 2, pp. 1057–1060.

[8] C. Rathinavalu and L. Deng. "HMM-based speech recognition using state-dependent, discriminatively derived transforms on Mel-warped DFT features," IEEE Trans. Speech and Audio Processing, 1997, pp. 243-256.

[9] Garima Vyas, Barkha Kumari , "Speaker Recognition System Based On MFCC and DCT", International Journal of engineering and advanced technology 06/2013; pp.167-169.

[10] Kashyap Patel, R.K. Prasad, "Speech Recognition and Verification Using MFCC & VQ" ,International Journal of Emerging Science and Engineering (IJESE), 2013,Volume-1, Issue-7.

[11] Model Lokesh Selvaraj and Balakrishnan Ganesan ,"Enhancing Speech Recognition Using Improved Particle Swarm Optimization Based Hidden Markov Model" Hindawi Publishing Corporation, e Scientific World Journal, Volume 2014, Article ID 270576, 10 pages.

[12] Lars Bretzner, Ivan Laptev and Tony Lindeberg, "Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering", in Automatic

[13] Junqiu Wang, Yasushi Yagi, Member, "Adaptive Mean-Shift Tracking with Auxiliary Particles, IEEE TRANSACTION ON SYSTEM, MAN AND CYBERNETICS, PART B, March 31, 2009.

[14] NING, J., ZHANG, L., ZHANG, D., AND WU, C.,"Robust object tracking using joint color-texture histogram", International Journal of Pattern Recognition and Artificial Intelligence,2009, pp. 1245–1263.

[15] Katharina Quast and Andr´e Kaup, "Shape Adaptive Mean shift Object Tracking Using Gaussian Mixture Model", in Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 , pp.1-4.

[16] Kaihua Zhang, Lei Zhang, Ming-Hsuan Yang, and David Zhang, "Robust Object Tracking via Active Feature Selection", in Circuits and Systems for Video Technology, IEEE Tran, 2013, pp.1957-1967.

[17] K. Zhang, L, Zhang, M. H.Yang, and D. Zhang, "Fast Tracking via Spatio-Temporal Context Learning," in Computer Vision – ECCV 2014 ,13th European Conference, Zurich, Switzerland, 2014, pp.127-141.