

Project Title: Sports Video Annotation and Summarization**Team # 9****Members:** Manikanta Maddula**Motivation:**

The implementations and applications based on Big data tools especially real time applications are increasing day by day. Auto annotation, summarization, content analysis etc. from live video are interestingly growing in real time video analysis. Auto news article making, assisting for physically challenged people, task recommendations etc. can be said as applications of the video analysis. Combining video and audio analysis will have even more great applications which will be useful but it will be out of scope for this project. Annotation and summarization will be first step towards a better solution. Content analysis has wide range of applications in surveillance domain.

Objectives:

1. Categorize video into predefined sports categories
2. Identify specific format in that game
3. Object recognition in each video (People, places etc.).
4. Background and Foreground objects identification.
5. Real time analysis for all of the above. Fast response of within mille seconds.

System:

The system architecture will make use of apache Spark, Apache Storm, Apache Kafka, and video transferring client (android mobile or API). For the training purpose of model, spark's machine learning library is used. And to extract features from the video or image frames in video OpenImaj library is used. SIFT implementation is OpenImaj is used to achieve different objectives. Apache Storm is used to achieve parallel task distribution for real time data. Apache Kafka is used to communicate between different modules like Android and Storm, spark and storm, storm and android client. For a live video analysis, video stream data is sent to Strom using Apache Kafka as a Broker in between. Features of video stream is extracted using OpenImaj library. Using the model developed in spark machine learning library and with the extracted features, categorization, identification etc. are executed. Using Apache Kafka relevant information is sent to client application (e.g. Android mobile). SigSpace architecture can be used for the model training purpose. SigSpace architecture has advantages of space and time reduction, incremental learning, fuzzy matching, dimension reduction, distributed learning etc. YouTube videos of Olympics 2016 at Rio can be used for training purpose. For audio analysis, Android or web API's can be used to convert Speech to text and this text can be used but this application is trying to analyze significant events in audio data.

Significance and Uniqueness:

There are several related projects in the domain of real time video analysis and sports. But this kind of architecture is never implemented at UMKC – CSEE. And usage of SigSpace is first to implement in Real time and video analysis. Sports domain is a billion-dollar industry which will have huge impact if application can do some useful analysis with the available data.

Back UP:

As a backup solution, System architecture can be modified little bit by using Apache spark for real time analysis taking Storm out of picture. Spark streaming provides batch and micro batch processing. This can be used instead of Storm with a little more delay in response.

Bibliography:

1. Pradyumna Doddala, SigSpace: Class-base Feature representation for Scalable and Distributed Machine Learning, Thesis Spring 2016 at University of Missouri-Kansas City.
2. <https://spark.apache.org/>
3. <https://storm.apache.org/>
4. <http://kafka.apache.org/>
5. <http://openimaj.org/>
6. <http://www.infoworld.com/article/2854894/application-development/spark-and-storm-for-real-time-computation.html>
7. Ma, Shuangmei, and Zhengli Liang. "Design and Implementation of Smart City Big Data Processing Platform Based on Distributed Architecture." *Intelligent Systems and Knowledge Engineering (ISKE), 2015 10th International Conference on*. IEEE, 2015.
8. Morshed, Sarwar Jahan, Juwel Rana, and Marcelo Milrad. "Open source initiatives and frameworks addressing distributed real-time data analytics." *2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE, 2016.