Demo Abstract: FOCUS: Clustering Crowdsourced Videos by Line-of-Sight

Puneet Jain
Duke University
puneet.jain@duke.edu

Justin Manweiler
IBM T. J. Watson
jmanweiler@us.ibm.com

Arup Achary
IBM T. J. Watson
arup@us.ibm.com

Kirk Beaty
IBM T. J. Watson
kirkbeaty@us.ibm.com

ABSTRACT

We present a demonstration of FOCUS [1], a system to appear in the *SenSys 2013* main conference. FOCUS is a video-clustering service for live user video streams, indexed automatically and in realtime by shared content. FOCUS uniquely leverages visual, 3D model reconstruction and multimodal sensing to decipher and continuously track a video's line-of-sight. Through spatial reasoning on the relative geometry of multiple video streams, FOCUS recognizes shared content even when viewed from diverse angles and distances. We believe FOCUS can enable a new family of applications, such as instant replay, augmented reality, citizen journalism, security breach detection, and disaster assessment.

In the demonstration, we will show 325 video clips taken at Duke University Wallace Wade Stadium being processed in real-time via FOCUS pipeline. The recorded video clips contain one of three spots in the stadium: East Stand, Scoreboard, and West Stand. The demo shall be shown in the form of a web interface, first showing randomly clustered video clips. Later, on a button click, FOCUS shall process displayed videos in real-time, outputting clusters of videos, containing the common shared subject in each of them. For each successfully processed video clip in a cluster, we will further show similar clips from near, medium, and wide angle. To display performance and accuracy of FOCUS in indoor environments, a similar demonstration will be shown for an office space. FOCUS shall run on multi-node Hadoop cluster built on top of IBM SmartCloud platform.

General Terms

Algorithms, Design, Experimentation, Performance

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software

Keywords

Video Analytics, Localization, Structure from Motion

1. DEMONSTRATION

FOCUS is implemented in two parts: (1) a mobile app (prototyped for Android) (2) a distributed service, deployed on infrastructure-as-a-service cloud using Hadoop. The app lets users to capture and upload live video streams, annotated with time-synchronized sensor data, from GPS, compass, and gyroscope. The cloud service receives video streams in parallel, analyzing each, leveraging Computer Vision and sensing to continuously model and track the video's line-of-sight. FOCUS performs a series of steps on cloud before inferring a video's line-of-sight: (1) Frame decoding, (2) Feature extraction, (3) Features to model alignment. The above steps estimate relative camera pose of any given frame, with respect to the provided model; We define line-of-sight as a ray emerging from camera position, extending toward the infinity in the outward facing direction of camera. The estimated line-of-sight is used to quantify the shared physical space between any two videos and an spatio-temporal matrix of similarity is constructed. Finally, clustering using modularity maximization is performed to construct groups of similar clips. The similar clips in a group are further divided into subgroup by looking at 3D-angular separation in their line-of-sight.

2. SETUP

We shall demo FOCUS on a web browser. The participants will be asked to select a set of clips from a pre-captured pool of videos. The selected video clips shall be processed via FOCUS pipeline and results will be displayed in three column format from different angle of views. A tentative screenshot of our demo is shown in Figure 1. FOCUS estimates video's line-of-sight with the help of a 3D model. To display the significance of the model, along with each video, we shall also show the corresponding images used in the model construction.

Figure 1: Random unclustered videos (left): showing video previews, Videos clustered by line-of-sight (right): columns show similar videos from various angular separation.

3. REFERENCES

[1] P. Jain, J. Manweiler, A. Acharya, and K. Beaty.
    Focus: Clustering crowdsourced videos by line-of-sight.
    In *Proceedings of the 11th ACM Conference on
    Embedded Networked Sensor Systems.* ACM, 2013.