ROBUST CONTENT IDENTIFICATION AND DE-DUPLICATION WITH SCALABLE CONTENT HASH FROM FISHER AGGREGATION

Lakshmi Prasanna Gadiparthi, Candidate for the Master of Science Degree

University of Missouri-Kansas City, 2017

# ABSTRACT

Robust content identification and de-duplication of video content in networks and caches have many important applications in content delivery networks. In this work, we propose a scalable hashing scheme based Fisher Vector aggregation of selected key point features, and a frame significance function based non-uniform temporal sampling scheme on the video segments, to create a very compact binary representation of the content fragments that is agnostic to the typical coding and transcoding variations. The key innovations are a key point repeatability model that selects the best key point features, and a non-uniform sampling scheme that significantly reduces the bits required to represent a segment, and scalability from PCA feature dimension reduction and Fisher Vector freaturesgation, and Simulation with various frame size and bit rate video contents for DASH streaming are tested and the proposed solution have very good performance of precision-recall, achieving 100% precision in duplication detection with recalls at 98% and above range.