This dissertation addresses the problem of video summarization, that studies creating algorithms to enable fast information extraction from large video databases. Most video summarization methods generate a summary of a video by selecting key frames or shots from it.

One of the main obstacles for the research in video summarization is that no single video summarizer fits every user’s needs. Therefore, we develop a probabilistic model, Sequential and Hierarchical Determinantal Point Process (SH-DPP), for query-focused video summarization. Given a user query and a video sequence, our algorithm returns a summary by selecting key shots from the video that are either relevant to the query or important in the context of the video.

Next, to evaluate the performance of video summarizers, we develop an novel metric. Additionally, we propose a memory network parameterized sequential DPP in order to attend the user query onto different video frames and shots. Compared to the hierarchical model, this framework requires less supervision to train.

SeqDPP based approaches to video summarization are trained by maximizing the likelihood of user summaries. At the test time, however, the model generates output by searching over the output space in a greedy fashion. We alleviate this issue by incorporating inference technique into the objective function used for training. We also improve the SeqDPP model by a novel probabilistic distribution in order to allow users to control the lengths of system-generated video summaries.

Finally, to enable faster browsing of large databases, we develop a framework to generate text synopsis for a given video. To this end, the video is divided into shots and a descriptive sentence for each shot is generated via a video caption generation network. Our visual-language content matching unit distinguishes between the informative and uninformative sentences. Next, our purport network selects sentences that convey important information.