Towards Efficient and Reliable Video Query Systems

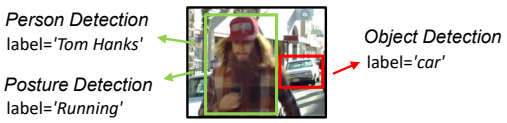
Daren Chao

Department of Computer Science
University of Toronto
drchao@cs.toronto.edu

1 Introduction

The rapid proliferation of video data across various platforms has created a pressing need for systems that can efficiently and reliably query and retrieve relevant video content. Advancements in video analytical models, including computer vision models and multimodal large models, have significantly enhanced the capability to process video data. These technologies have been applied in diverse fields such as traffic analysis, security and surveillance, sports analysis, and news preparation. Nevertheless, a lingering problem is that the inherent uncertainty and computational demands of video analytical models, coupled with their complexity for non-expert users, often prevent their integration into existing application requirements.

Video Query Systems (VQS, a.k.a. VDBMS) have emerged as indispensable tools in this data-rich landscape, enabling non-expert users to query and retrieve specific content from vast video databases. As an example, Figure 1 presents a SQL-like query statement used in our proposed VQS to retrieve frames from "Forrest Gump" featuring Tom Hanks running alongside a car. VQS is meticulously engineered to interpret, optimize, and execute video queries in a declarative[1] manner. This talk will discuss the development of reliable and efficient VQS as a critical response to the burgeoning demand for processing the rapidly growing volumes of video data.




```
-- Q1: Retrieve running Tom Hanks with a car in
      the scene from Movie Forrest Gump.
WITH V AS LOAD_VIDEO('forrest_gump.mp4');
SELECT DISTINCT P.fid
  FROM PERSON_ID(V) AS P, OBJ_DET(V) AS D
    ON (P.fid = D.fid)
  WHERE P.label = 'Tom Hanks' AND D.label =
    'car' AND PosDet(P.box) = 'running';
```

Figure 1: Example Video Query Q1.

[1] In declarative systems, users specify what data they want to retrieve or manipulate without detailing how the system should perform the operation.

2 Talk Content

This talk will discuss a series of papers dedicated to developing more efficient video query systems. The talk is organized into three sections: First, we introduce a fundamental pipeline of VQS that leverages user knowledge to optimize queries, as detailed in our work [2]. Second, we explore optimization approaches for specific query predicates, such as object detection [5], object tracking [1], action recognition [3], and event detection [4]. For each predicate type, we delineate the associated research challenges concerning the reliability and efficiency of query processing and introduce our innovative solutions. Finally, the talk will pivot to future research avenues, particularly on the efficient and reliable integration of multi-modal large models into novel VQS frameworks to manage extensive video databases.

References

[1] Daren Chao, Yueting Chen, Nick Koudas, and Xiaohui Yu. Track merging for effective video query processing. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 164–176, 2023.

[2] Daren Chao, Yueting Chen, Nick Koudas, and Xiaohui Yu. Optimizing video queries with declarative clues. *Proc. VLDB Endow.*, 17(11):3256–3268, 2024.

[3] Daren Chao and Nick Koudas. Querying for actions over videos. In *Proceedings 27th International Conference on Extending Database Technology, EDBT 2024*, pages 162–174, 2024.

[4] Daren Chao, Nick Koudas, and Xiaohui Yu. Marshalling model inference in video streams. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 1379–1392, 2023.

[5] Daren Chao, Nick Koudas, Xiaohui Yu, and Yueting Chen. Ensembling object detectors for effective video query processing. In *Proceedings 28th International Conference on Extending Database Technology, EDBT 2025*, pages 66–79, 2025.