International Journal of Multimedia Data Engineering and Management
Volume 10 • Issue 2 • April-June 2019

FaceTimeMap:
Multi-Level Bitmap Index for Temporal Querying of Faces in Videos

Buddha Shrestha, University of Alabama in Huntsville, Huntsville, USA
Haeyong Chung, University of Alabama in Huntsville, Huntsville, USA
Ramazan S Aygün, University of Alabama in Huntsville, Huntsville, USA

ABSTRACT

In this article, the authors study bitmap indexing for temporal querying of faces that appear in videos. Since the bitmap index is originally designed to select a set of records that satisfy a value in the domain of the attribute, there is no clear strategy for how to apply it for temporal querying. Accordingly, the authors introduce a multi-level bitmap index that the authors call "FaceTimeMap" for temporal querying of faces in videos. The first level of the FaceTimeMap index is used for determining whether a person appears in a video or not, whereas the second level of the index is used for determining intervals when a person appears. First, the authors analyze the co-appearance query where two or more people appear simultaneously in a video, and then examine next-appearance query where a person appears right after another person. In addition, to consider the gap between the appearance of people, the authors study eventual- and prior-appearance queries. Queries are satisfied by applying bitwise operations on the FaceTimeMap index. The authors provide some performance studies associated with this index.

KEYWORDS

Allen's Intervals, Co-Appearance, Eventual-Appearance, Face Search, Next-Appearance

1. INTRODUCTION

Concurrent with the growing use of online social networks is the significant increase in the number of videos that are uploaded to the Internet. Consider, for example, the number of video hours watched daily on YouTube reached one billion hours with more than 70% of watching time spent on mobile devices (YouTube, 2019). Many videos are uploaded for sharing experiences, knowledge, and entertainment. When a video is posted, users who have viewed similar content in the past may receive notification of new uploads. However, if videos had been uploaded some time ago, users need to search using retrieval engines to locate the relevant videos. Nevertheless, even plain video retrieval that seeks to locate an object, event, or action is a challenging task due to the complexity of query building, utility gap (Hanjalic, 2013), and subjectivity. Content-based video retrieval requires efficient index structures that support both spatial and temporal queries. Analyzing videos, extracting features,

DOI: 10.4018/IJMDEM.2019040103

indexing content and classifying video data represents an increasingly important and active research area (Ashgar, Hussain, & Manton, 2014). The problem remains, however, that the gap between what the user seeks when he or she initiates a query, and what the retrieval system is capable of returning, currently hinders the broader use of these retrieval tools.

Face search/retrieval is one type of video retrieval that has many applications: locating a video in which a single person appears, two people appear together (as in a sporting game or event,) a person appears after another person, or some other temporal constraints. For surveillance, it may be important to determine if two people are exchanging an item or not. For crime-scene investigations using surveillance, it may be possible to track people before and after an event. Querying faces in videos requires face detection, face recognition, and then retrieving video clips based on the user query. In particular, modern methods utilizing deep learning for face recognition (Ashgar et al., 2014; Schroff, Kalenichenko, & Philbin, 2015; Taigman, Yang, Ranzato, & Wolf, 2014) are proving to be nearly as accurate as human perception. Nonetheless, indexing video content, even just for face searches, is quite challenging due to the large volume of data. As indicated earlier, the number of online videos and the number of people who appear in these videos are significantly higher compared to 15.

In this paper, we propose a new technique, FaceTimeMap, for indexing videos for face searches using a bitmap index (Shrestha, Chung, & Aygun, 2019). The bitmap index has recently been used for column-based retrieval in big-data systems (Chen et al., 2015). Since the bitmap index was originally designed to select a set of records that satisfy a value in the domain of the attribute, there is no clear strategy as to how to apply it for temporal querying. We utilized a multi-level bitmap index by creating two types of matrices. The first bitmap matrix has a bit set if a person appears in a video. The second level of the bitmap index is built for each video whereby a video is represented as a sequence of intervals. In the second level matrix, a bit is set for a person if it appears in that interval. Whenever a query based on appearance or temporal ordering of faces is submitted to the system, our retrieval engine first finds the relevant videos using the first level of the index, and then the intervals are only checked for those relevant videos based on the user query. There are three types of queries considered in this paper: (a) co-appearance: intervals where multiple people (or faces) appear simultaneously in a scene of a video; (b) next-appearance: intervals where a person appears right after another person disappears; and (c) eventual (or prior)-appearance: videos where a person appears sometime after (or before) another person disappears.

FaceTimeMap index utilizes two types of bitmaps. Specifically, we use bit manipulation strategies for computing the query components and then returning the query results. The major advantage of the bitmap index is to be able to use bitwise operators. While bitwise operators are suitable for appearance or co-appearance, its application for other types of temporal queries (e.g., next or eventually operators from linear temporal logic) is not clear. Thus, we present how bitwise operators could be used to represent such temporal operators. As the number of query elements or intervals increase in a query, it becomes harder for users to build or formulate such queries. In this study, the user needs to provide sample images of people in whom the user is interested, and then choose the relevant temporal operator to indicate the temporal relationships between these people. In our prior work (R. Aygun & Benesova, 2017), studies on multimedia information retrieval have been grouped into (a) domain of media & cross-domain, (b) detection & recognition, (c) retrieval & summarization, and (d) interfaces. In this study, we develop a simple interface for temporal querying of faces, use the available libraries for face detection and recognition, and propose FaceTimeMap index for indexing faces in videos. Although FaceTimeMap index is designed for temporal querying of faces, the technique can be used for other query elements that can be represented as intervals.

This paper is organized as follows. The following section describes the related work on face retrieval and indexing for temporal data. Section 3 explains how the bitmap index is utilized in FaceTimeMap index. The user experiments we conducted are detailed in Section 4, while the final section provides concluding information.