Algebraic Retrieval of Fragmentarily Indexed Video

Katsumi TANAKA, Keishi TAJIMA and Takashi SOGO
Department of Computer and Systems Engineering, Kobe University
Rokkodai, Nada, Kobe 657-8501, JAPAN
{tanaka,tajima,sogoh}@db.cs.kobe-u.ac.jp

Sujeet PRADHAN
Kurashiki University of Science and the Arts
Nishinoura 2640, Tsurejima, Kurashiki, Okayama 712-8505, JAPAN
sujeet@soft.kusa.ac.jp

Received 15 January 2000
Revised manuscript received 3 April 2000

Abstract When dealing with long video data, the task of identifying and indexing all meaningful subintervals that become answers to some queries is infeasible. It is infeasible not only when done by hand but even when done by using latest automatic video indexing techniques. Whether manually or automatically, it is only fragmentary video intervals that we can identify in advance of any database usage. Our goal is to develop a framework for retrieving meaningful intervals from such fragmentarily indexed video data. We propose a set of algebraic operations that includes our glue join operations, with which we can dynamically synthesize all the intervals that are conceivably relevant to a given query. In most cases, since these operations also produce irrelevant intervals, we also define various selection operations that are useful in excluding them from the answer set. We also show the algebraic properties possessed by those operations, which establish the basis of an algebraic query optimization.

Keywords: Video Database, Query, Algebraic Retrieval, Interval Operation, Video Intervals.

§1 Introduction

Video data consists of a sequence of shots. Over the past several years, substantial progress has been made in automatic detection of shot boundaries
based on the changes of visual characteristics. It has enabled automatic segmentation of video data into shots with a high accuracy. There has also been considerable progress in indexing those video shots by automatically extracting keywords using the techniques such as speech and text recognition. A single shot extracted in that way, however, is mostly fragmentary and can hardly be considered as a meaningful interval having self-contained information. Users are interested in semantically meaningful intervals which generally span several consecutive shots. It is enormously difficult to identify all of those intervals beforehand for answering any subsequent queries. Rather than investigating how to identify all the meaningful intervals, we shift our focus on how to synthesize them dynamically from fragmentarily indexed shots, when queries are given.

We achieve our goal by introducing two types of algebraic operations. The former type, called glue operations, synthesize new meaningful intervals by gluing given fragmentary intervals, and the latter one, called filter operations, select a subset of synthesized intervals meeting some given conditions. Since in most cases, video intervals obtained through glue operations contain many irrelevant intervals to a given query, filter operations are provided in order to exclude them from the set of query answers. Both glue operations and filter operations have interesting algebraic properties that are useful for a logical query optimization. We introduce a particular class of filter operations, called optimizable filters with certain properties and show how they are applied for achieving the optimization. We show the necessary and the sufficient condition for any filter to be optimizable. Finally, we also show that Allen's interval logic is expressible by one of the proposed optimizable filters.

§2 Related Work and Motivations

2.1 Video Indexing Models

There are basically two different approaches to video indexing — segmentation and stratification approach. The former one, segmentation approach, is to decompose a video interval to non-overlapping subintervals. Each subinterval has its feature data (such as color, texture, motion and shape) associated with it by some automatic feature extraction techniques. Informedia project is a well-known system to make the best use of segmentation approach. The latter one, stratification approach, is to identify arbitrary subintervals (that may overlap each other) in a given video, and annotate them with some textual information.

Video indexing with segmentation approach, which solely depends on automatic feature extraction techniques suffers from some major drawbacks like

1. lower precision,
2. inability to define the semantics, and
3. fragmentariness of the obtained index.

In this paper, we have paid a special attention to the third drawback. Suppose we use automatic speech recognition techniques for extracting keywords from