Indexing Motion Detection Data for Surveillance Video

Vind, Søren Juhl; Bille, Philip; Gørtz, Inge Li

Published in:
Proceedings of the IEEE International Symposium on Multimedia (ISM2014)

Link to article, DOI:
10.1109/ISM.2014.36

Publication date:
2014

Document Version
Peer reviewed version

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Indexing Motion Detection Data for Surveillance Video

Søren Vind, Philip Bille †, and Inge Li Gørtz †
Department of Applied Mathematics and Computer Science
Technical University of Denmark
Copenhagen, Denmark
sovi@dtu.dk, phbi@dtu.dk, inge@dtu.dk

Abstract—We show how to compactly index video data to support fast motion detection queries. A query specifies a time interval $T$, a area $A$ in the video and two thresholds $v$ and $p$. The answer to a query is a list of timestamps in $T$ where $\geq p\%$ of $A$ has changed by $\geq v$ values.

Our results show that by building a small index, we can support queries with a speedup of two to three orders of magnitude compared to motion detection without an index. For high resolution video, the index size is about 20% of the compressed video size.

Keywords—motion detection index; motion detection data structure; surveillance video; video analysis; data structure;

I. INTRODUCTION

Video data require massive amounts of storage space and substantial computational resources to subsequently analyse. For motion detection in video surveillance systems, this is particularly true, as the video data typically have to be stored (in compressed form) for extended periods for legal reasons and motion detection requires time-consuming decompressing and processing of the data. In this paper, we design a simple and compact index for video data that supports efficient motion detection queries. This enables fast motion detection queries on a selected time interval and area of the video frame without the need for decompression and processing of the video file.

A. Problem & Goal

A motion detection query $MD(T, A, v, p)$ specifies a time range $T$, an area $A$, and two thresholds $v \in [0, 255]$ and $p \in [0, 100]$. The answer to the query is a list of timestamps in $T$ where the amount of motion in $A$ exceeds thresholds $v$ and $p$, meaning that $\geq p\%$ of the pixels in $A$ changed by $\geq v$ pixel values. Our goal is build an index for video data that supports motion detection queries. Ideally, the index should be small compared to the compressed size of the video data and should support queries significantly faster than motion detection without an index.

B. Related Work

Several papers have considered the problem of online motion detection, where the goal is to efficiently identify movement in the video in real time, see e.g. [1], [2], [3], [4], [5]. Previous papers [6], [7] mentions indexing movement of objects based on motion trajectories embedded in video encoding. However, to the best of our knowledge, our solution is the first to show a highly efficient index for motion detection queries on the raw video.

C. Our Results

We design a simple index for surveillance video files, which support motion detection queries efficiently. The performance of the index is tested by running experiments on a number of surveillance videos that we make freely available for use. These test videos capture typical surveillance camera scenarios, with varying amounts of movement in the video.

Our index reduces the supported time- and area-resolution of queries by building summary histograms for the number of changed pixels in a number of regions of frames succeeding each other. Histograms for a frame are compressed and stored using an off-the-shelf compressor. Queries are answered by decompressing the appropriate histograms and looking up the answer to the query in the histograms.

The space required by the index varies with the amount of motion in the video and the region resolution supported. The query time only varies slightly. Compared to motion detection without an index we obtain:

- A query time speedup of several orders of magnitude, depending on the resolution of the original video. The choice of compressor has little influence on the time required to answer a query by the index.
- A space requirement which is $10 – 90\%$ of the compressed video. The smallest relative space requirement occur for high resolution video. Quadrupling the region resolution roughly doubles the space use.

Furthermore, as the resolution of the video increases, the time advantage of having an index grows while the additional space required by the index decreases compared to the compressed video data. That is, the index performs increasingly better for higher resolution video.
II. THE INDEX

A \( MD(T, A, v, p) \) query spans several dimensions in the video file: The time dimension given by \( T \) and two spatial dimensions given by \( A \). However, as high-dimensional data structures for range queries typically incur high space cost, we have decided to not implement our index using such data structures. Instead, we create a large number of two-dimensional data structures for the pixel value difference for each successive pair of frames, called a difference frame. Answering a query then involves querying the data structures for all difference frames in \( T \).

We restrict the query area \( A \) to always be a collection of regions, \( r_1, \ldots, r_k \). The height and width of a region is determined by the video resolution and the number of regions in each dimension of the video (if other query areas are needed, the index can be used as a filter). For simplicity, we assume that the pixel values are grey-scale.

The index stores the following. For each region \( r \) and difference frame \( F \), we store a histogram \( H_{F,r} \), counting for each value \( 0 \leq c \leq 255 \) the number of pixels in the region changed by at least \( c \) pixel values. Clearly a histogram can be stored using 256 values only. While this may exceed the number of pixels in a region when storing many regions per frame, it generally does not. However, because modern video encoding is extremely efficient, the raw histograms may take more space than the compressed video (especially for low video resolutions). Thus, we compress the histograms using an off-the-shelf compressor before storing them to disk.

To answer a \( MD(T, A, v, p) \) query, we decompress and query the histograms for each region in \( A \) across all difference frames in \( T \). Let \( |r| \) denote the number of pixels in region \( r \). For a specific difference frame \( F \), we calculate \( p' = \sum_{v \in A} H_{F,r}[v] / \sum_{v \in A} |r| \), which is exactly the percentage of pixels in \( A \) changed by \( v \) pixel values. Thus, if \( p' \geq p \), frame \( F \) is a matching timestamp.

III. EXPERIMENTS

A. Experimental setup

All experiments ran on an Apple Macbook Pro with an Intel Core i7-2720QM CPU, 8GB ram and a 128GB Apple TS128C SSD disk, plugged into the mains power. All reported results (both time and space) were obtained as the average over three executions (we note that the variance across these runs was extremely low).

B. Data sets

We tested our index on the following three typical video surveillance scenarios, encoded at 29.97fps using H264/MP4 (reference [8]). We use different video resolutions (1920 \( \times \) 1080, 1280 \( \times \) 720 and 852 \( \times \) 480 pixels). See Table I.

1) Office: Recording of typical workday activities in a small well-lit office with three people moving. The image is almost static, only containing small movements by the people. There is very little local motion in the video.

2) Students: Recording of a group of students working in small groups, with trees visible through large windows that give a lot of reflection. People move about, which gives a medium amount of motion across most of the frame.

3) Rain: A camera mounted on the outside of a building, recording activities occurring along the building and looking towards another building. It is windy and raining, which combined with many trees in the frame creates a high amount of motion across the entire frame.

C. Implementation

The system was implemented in Python using bindings to efficient C/C++ libraries where possible. In particular, OpenCV and NumPy were used extensively, and we used official python bindings to the underlying C/C++ implementations of the compressors. The implementation uses a number of tunable parameters, see Table II. The source code can be found at [8].

IV. MAIN RESULTS

We now show the most significant experimental results for our index compared to the trivial method. We show results on both query time and index space when applicable. Unless otherwise noted, the index was created for a video size of 1080p, storing 1024 regions/frame, 3 frames/second, 1 frame/file, using linear packing and zlib-6 compression. We will only give detailed results for the students scenario, as the index performs relatively worst in this case.

A. Regions Queried

The first set of experiments show the influence of the number of regions queried in the image on the total query time and also check if one scenario deviates significantly from

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Length (s)</th>
<th>Motion amount</th>
<th>Size (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office</td>
<td>60</td>
<td>Low</td>
<td>9.0</td>
</tr>
<tr>
<td>Students</td>
<td>60</td>
<td>Medium</td>
<td>27.3</td>
</tr>
<tr>
<td>Rain</td>
<td>60</td>
<td>High</td>
<td>67.6</td>
</tr>
</tbody>
</table>

Table II

TUNABLE PARAMETERS FOR USE IN EXPERIMENTS.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frames/Second</td>
<td>The frame rate of the difference frames to index.</td>
</tr>
<tr>
<td>Regions/Frame</td>
<td>The number of regions to divide a frame into.</td>
</tr>
<tr>
<td>Compressor</td>
<td>The compressor used to compress the histograms.</td>
</tr>
<tr>
<td>Frames/File</td>
<td>The number of frames for which the histograms should be stored in the same file on disk.</td>
</tr>
<tr>
<td>Packing</td>
<td>Which strategy should be used when storing histograms for more than one frame in same file on disk.</td>
</tr>
</tbody>
</table>
Table III
INDEX QUERY TIME VERSUS VIDEO QUERY TIME FOR 1 AND 1024 REGIONS QUERIED. SIZE OF INDEX COMPARED TO THE VIDEO SIZE.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Query Reg.</th>
<th>Time (s)</th>
<th>Speedup</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.17</td>
<td>463.51</td>
<td>2726</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>0.81</td>
<td>467.17</td>
<td>576</td>
</tr>
<tr>
<td>Office</td>
<td>1</td>
<td>0.20</td>
<td>249.76</td>
<td>1248</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>0.83</td>
<td>253.86</td>
<td>305</td>
</tr>
<tr>
<td>Students</td>
<td>1</td>
<td>0.20</td>
<td>351.40</td>
<td>1757</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>0.83</td>
<td>355.98</td>
<td>428</td>
</tr>
</tbody>
</table>

Table IV
SPEEDUP FOR INDEX QUERY TIME COMPARED TO VIDEO QUERY TIME FOR STUDENTS SCENARIO, WITH VARYING INPUT VIDEO RESOLUTIONS AND NUMBER OF REGIONS QUERIED.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Speedup / Query Reg.</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>480p</td>
<td>1 454× 368× 102×</td>
<td>75.8%</td>
</tr>
<tr>
<td>720p</td>
<td>903× 745× 219×</td>
<td>47.4%</td>
</tr>
<tr>
<td>1080p</td>
<td>1248× 1060× 305×</td>
<td>20.1%</td>
</tr>
</tbody>
</table>

Table V
SIZE OF INDEXES FOR VARYING INPUT VIDEO RESOLUTIONS.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Index Size (MB)</th>
<th>Index Size (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1080p</td>
<td>720p</td>
</tr>
<tr>
<td>Office</td>
<td>2.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Students</td>
<td>5.5</td>
<td>3.7</td>
</tr>
<tr>
<td>Rain</td>
<td>5.4</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Table VI
INDEX SIZE FOR VARYING NUMBER OF STORED REGIONS AND RESOLUTIONS IN THE STUDENTS SCENARIO.

<table>
<thead>
<tr>
<th>Store Reg.</th>
<th>Index Size (MB)</th>
<th>Index Size (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>1.1 0.8 0.6</td>
<td>4.0% 10.2% 18.2%</td>
</tr>
<tr>
<td>256</td>
<td>2.3 1.7 1.2</td>
<td>8.4% 21.8% 36.4%</td>
</tr>
<tr>
<td>1024</td>
<td>5.5 3.7 2.5</td>
<td>20.1% 47.4% 75.8%</td>
</tr>
</tbody>
</table>

always the same (varies between 0.2s and 1s), while the video query times decrease from around 250s at 1080p to 75s at 480p, and we thus only report the relative speedup for different numbers of regions queried. The relative index performance compared to the video approach improves in both space and time with larger video resolution. The index query time varies very little with the resolution (which is as expected, since the number of histogram values to check does not change).

C. Regions Stored

Clearly, the number of regions stored by the index has an influence on the index size (as this directly corresponds to the number of histograms to store). However, from Table VI, it is clear that the influence is smaller than would be expected. In fact, due to the more efficient compression that is achieved, quadrupling (or×) the number of regions only causes the index to about double (2×) in size. That is, it is relatively cheap to increase the resolution of regions.

V. OTHER RESULTS

In this section, we review the index performance when varying the different parameters listed in Table II. Changing the index parameters had insignificant influence on query times, so we only show results for the index space when varying the parameters. The largest contributor to the index advantage over the video decompression method is the idea of storing an index, and thus we only briefly review the results when varying the parameters.

A. Compressor

Table VII shows the size of the index when the histograms are compressed using a number of different compressors, and Table VIII shows the time spent compressing the index in total (remember the input is a 60s video file). In all of the tests in the previous section, we have used the zlib-6
compressor, as it gives a good tradeoff between compression time and space use. If one can spare the computing resources, there is hope for almost halving the index space if switching to bzip2 compression. Note, however, that the query time increases with a slower decompressor (especially when querying few regions, as shown in Section V-B).

B. Query Time Components

To see the influence of the compressor used, we determined how much of the total query time is spent checking the decompressed values, compared to the amount of time spent decompressing the video frames or histograms. The results are in Table IX. It is evident that the video resolution and our chosen index compressor only has a small influence on the total query time when the number of regions queried is large: Around 75% of the time is spent on the actual query. As for the video query time, > 95% of the total time is spent decompressing the video, with the actual query time being very insignificant in all cases. In absolute terms, the query times for the index and the decompressed video approach are comparable (difference around 10×) after decompression.

C. Frames/File & Histogram Packing Strategies

We tested the influence on the index size if storing more than one difference frame per file on disk. We tested four different value packing strategies: linear, binned, reg-linear, reg-binned. Consider a frame $F$ with two region histograms $r_1, r_2$. In the linear strategy, we just write all values from $r_1$ followed by all values from $r_2$. In the binned strategy, we interleave the value for the same histogram index from all regions in a frame, i.e. we write $r_1[0] r_2[0] r_1[1] r_2[1] \ldots r_1[255] r_2[255]$ on disk. When storing multiple frames $F_1, F_2$ in a file, assume $r_1, r_2$ has the same spatial coordinate in both frames. Then the reg-linear strategy writes $r_1$ followed by $r_2$, while the reg-binned strategy interleaves the values as before.

The hope is that this would result in more efficient compression, since the histogram for the same region may be assumed to be very similar across neighbouring frames. However, our results show that storing more frames per file and changing the packing strategy had very little effect on the index efficiency for storing many regions. One exception is when storing few regions (less than 64), increasing the number of frames per file decreases the index size due to the added redundancy available for the compressor.

VI. CONCLUSION

We have shown an index for motion detection data from surveillance video cameras that provides a speedup of at least two orders of magnitude when answering motion detection queries in surveillance video. The size of the index is small compared to the video files, especially for high resolution video.

REFERENCES


