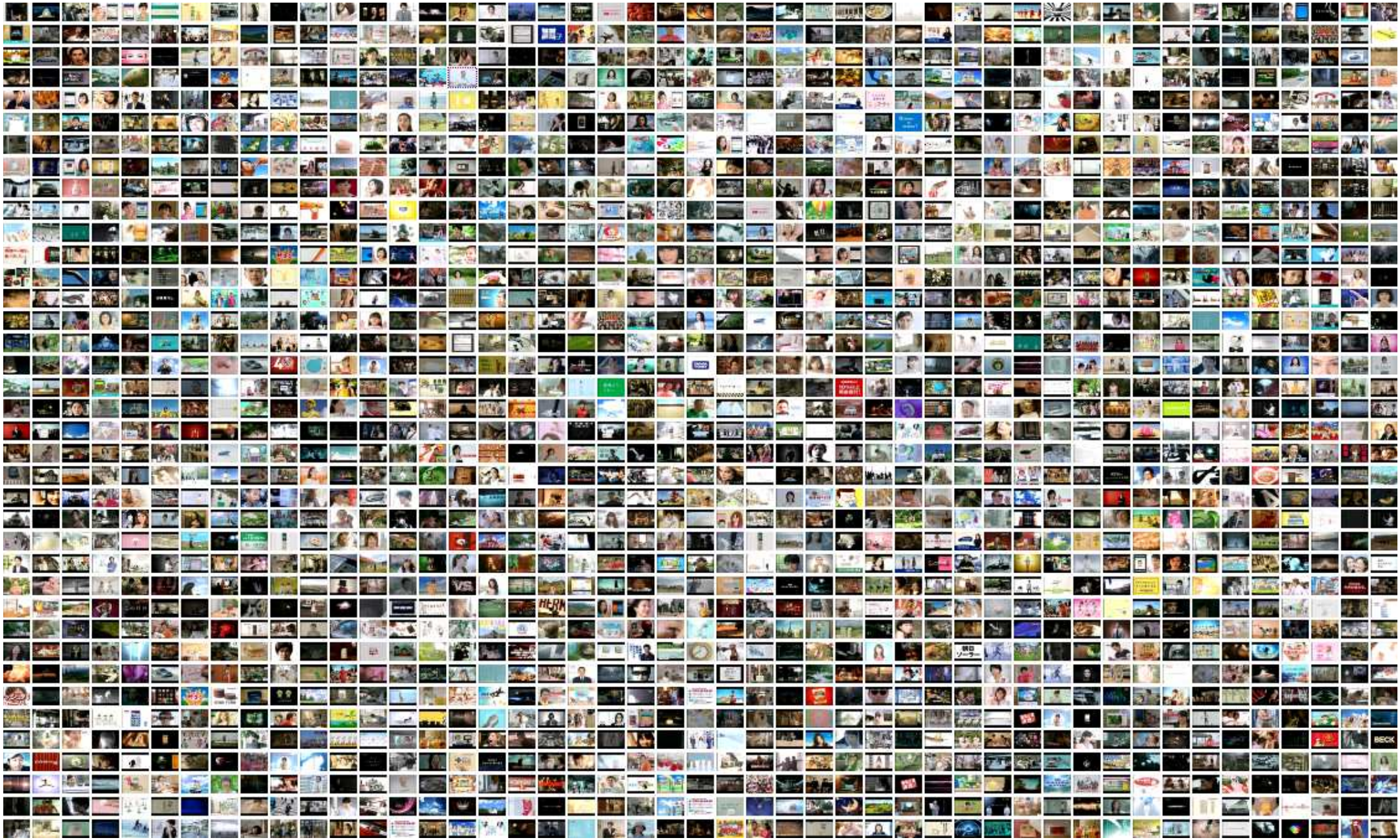
# Temporal Recurrence Hashing Algorithm for Mining Commercials from Multimedia Streams

Xiaomeng Wu, Shin'ichi Satoh
National Institute of Informatics, Japan

# Introduction



# Knowledge-Based CF Mining

- The intrinsic characteristics of CF (Commercial Film)
  - Monochrome Frame
  - Silence Frame
  - Product Frame
  - Duration
  - Shot Information
    - The frequency or style of cut, etc.
  - Motion Information
    - Edge change ratio, motion vector length, etc.

- Discussion
  - Efficient
  - Data-Dependent

# Repetition-Based CF Mining

- CF can be regarded as repetition
  - Hashing-Based
  - Clustering-Based

- Discussion
  - Unsupervised
  - Generic
  - Large computation burden

# Dual-Stage Hashing Algorithm

- Fully Unsupervised
  - No training databases nor queries provided beforehand
- Generic
  - Does not depend on any prior knowledge that might vary with countries or times
- Ultrahigh-Speed
  - 10-hour stream: 4 seconds
  - 1-month stream: 42 minutes
  - 5-year stream: 21 hours by parallel computing

# 1st-Stage Hashing

- Near Duplicate
  - Pairs or sets of approximately identical videos
  - Derived from an original video, usually by means of various transformations including cam cording, picture in picture, etc.

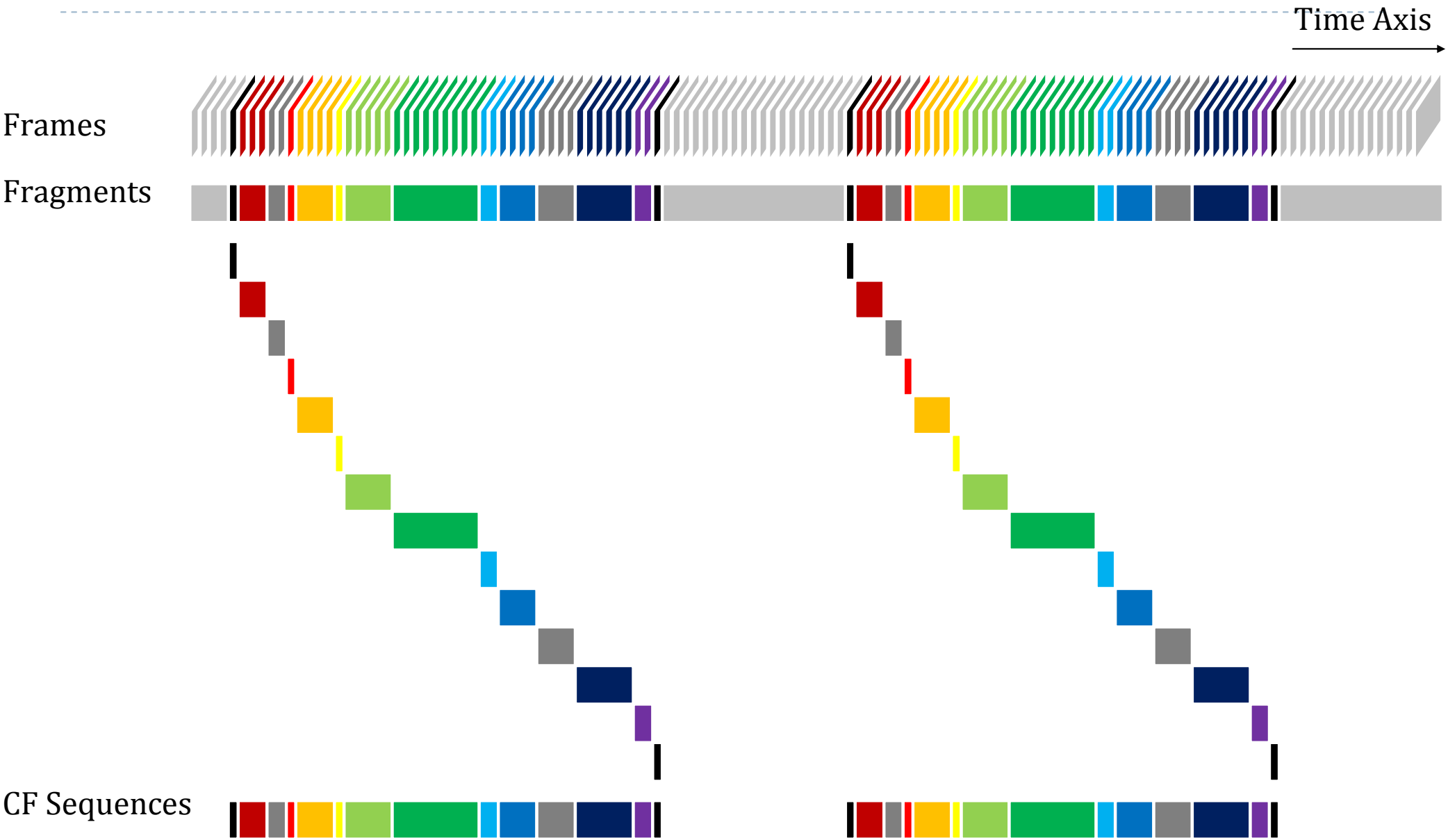
- CF
  - Exact duplicates derived from the original video without any transformations
  - Fragments of the videos can be translated into an exceptionally compact fingerprint so that identical fragments across the two or more exact duplicates can be mapped to exactly the same fingerprint
  - If we insert each fragment into a hash table by regarding the corresponding fingerprint as an inverted index, a hash collision will occur in the corresponding hash bucket

# 1st-Stage Hashing

- Video Hashing
  - Luminance-Based Fingerprint Strategy

- Audio Hashing [5]
  - It focuses on the energies of 33 bark-scaled bands of each audio frame, and uses the sign of the energy band differences both on the time and frequency axes as a 32-bit fingerprint

# Temporal Consistency Analysis (2nd-Stage)



# Temporal Consistency Analysis (2nd-Stage)

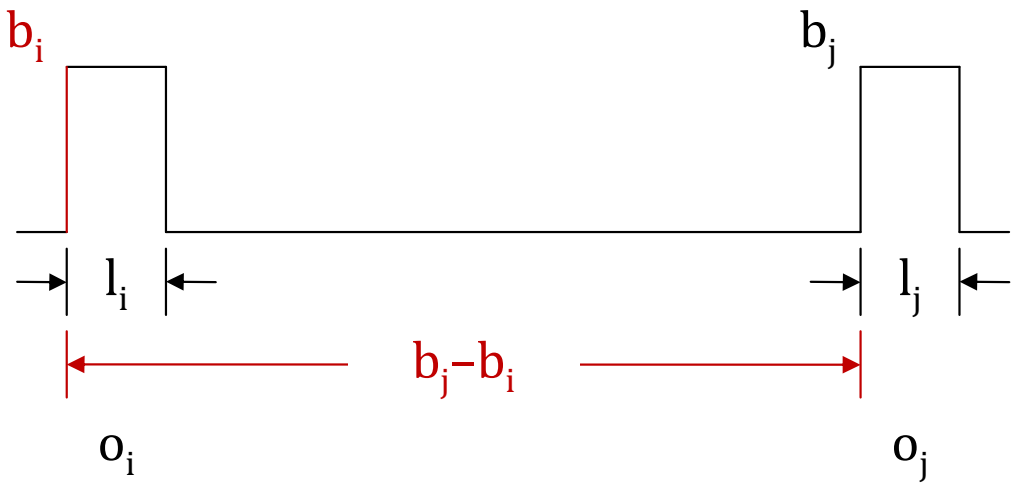
- [9]
  - Apply pairwise matching based on temporal consistency to all pairs of duplicate fragments derived from the whole stream
  - $\Theta(n_p^2) = 78{,}900{,}000^2$ (1-Month Stream)

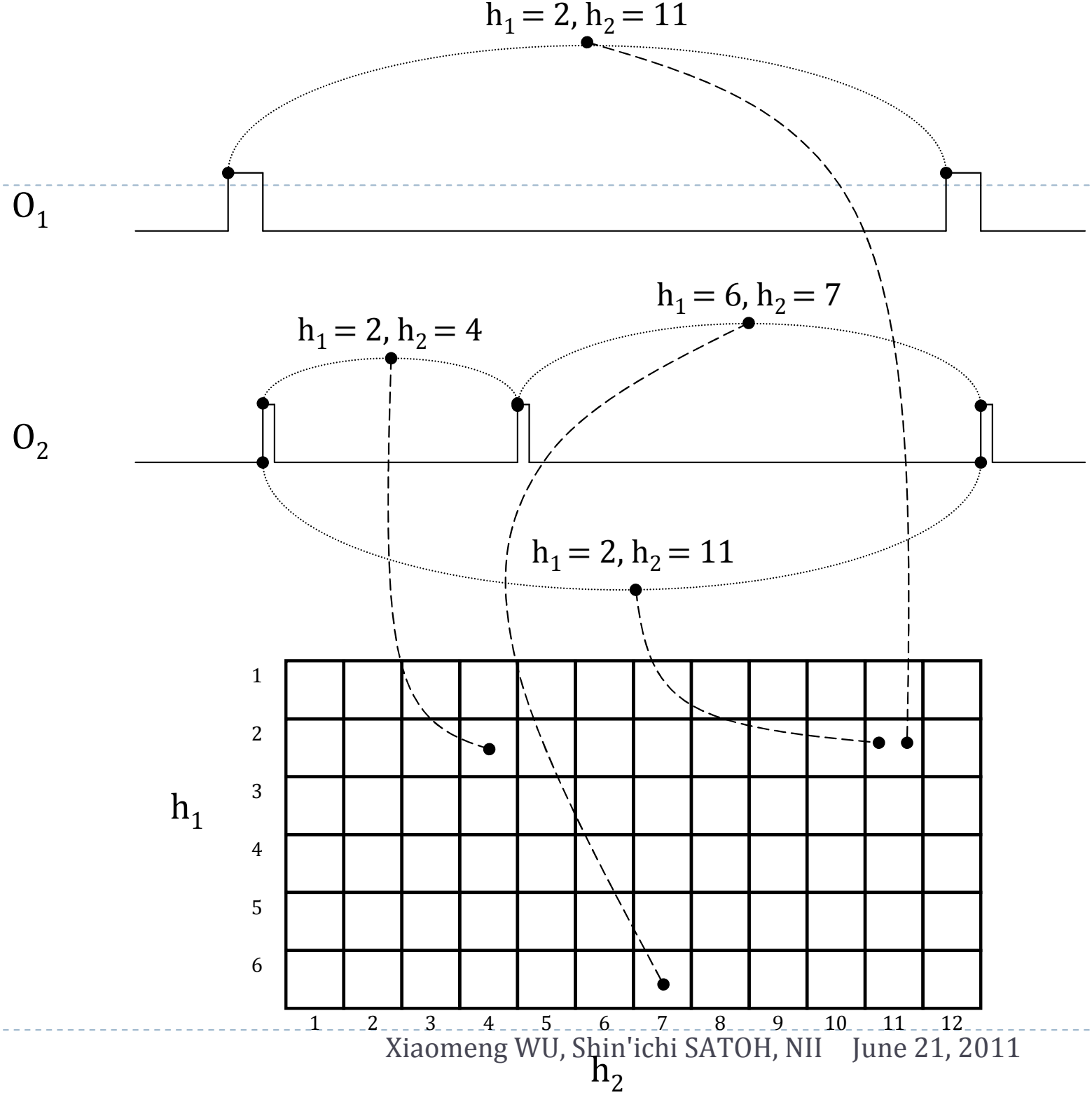
- [8, 10]
  - Apply pairwise matching based on temporal consistency to all sets of duplicate fragments derived from the whole stream
  - $\Theta(n_o^2) = 59{,}800{,}000^2$ (1-Month Stream)

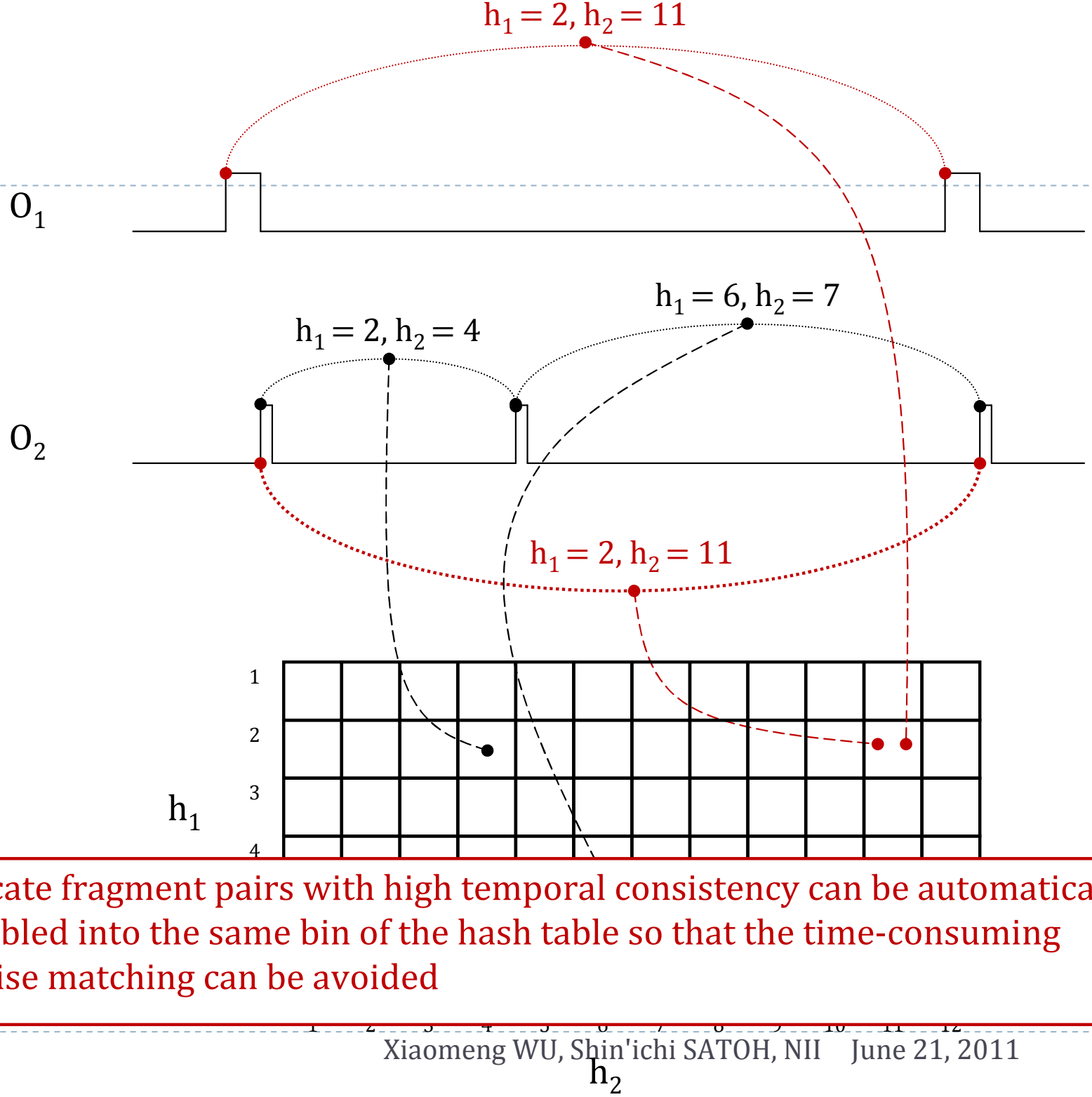
- [5]
  - Apply a fragment growing strategy to the duplicate fragment pairs
  - $\Theta(n_p)$ with high single operation cost

# Temporal Recurrence Hashing (2nd-Stage)

- $p=\{o_i,o_j\}$
  - $h_1(p)=h_1(o_i,o_j)=b_i$ minute
  - $h_2(p)=h_2(o_i,o_j)=b_j-b_i$ second




$h_1 = 2, h_2 = 11$
$O_1$
$h_1 = 6, h_2 = 7$
$h_1 = 2, h_2 = 4$
$O_2$
$h_1 = 2, h_2 = 11$
1
2
3
4
5
6
$h_1$
1 2 3 4 5 6 7 8 9 10 11 12
$h_2$

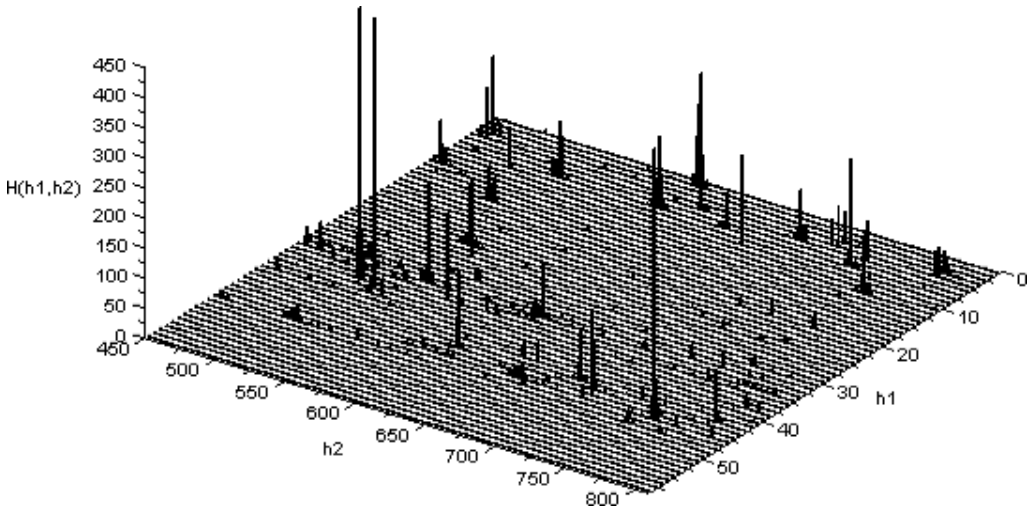


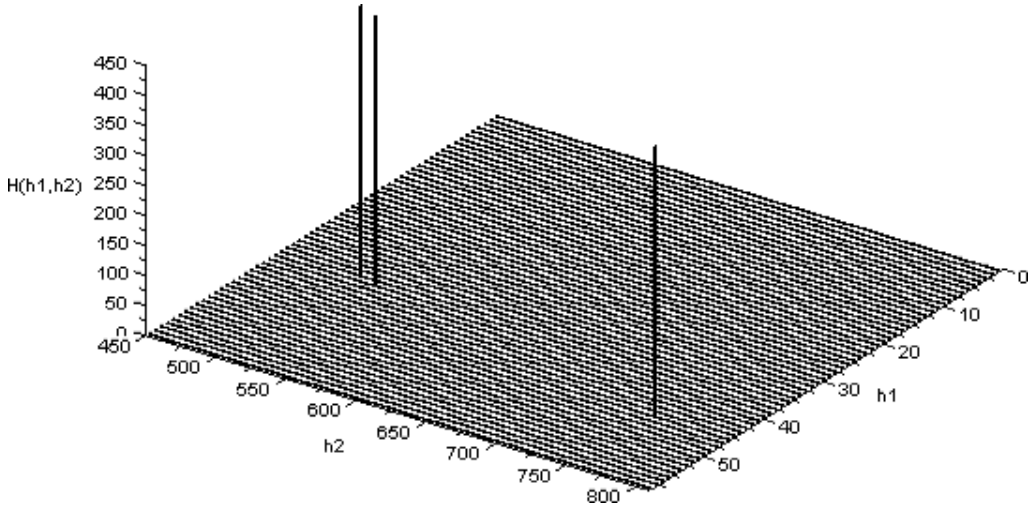

Duplicate fragment pairs with high temporal consistency can be automatically assembled into the same bin of the hash table so that the time-consuming pairwise matching can be avoided

# Temporal Recurrence Hashing (2nd-Stage)

- $H(a,b)=\sum_{h1(p)=a,h2(p)=b} l(p)$
- $l(p)=l(o_i,o_j)=\min(l_i,l_j)$

- Since the duplicate fragment pairs with high temporal consistency have been assembled into the same bin, this bin normally form a local maximum in the histogram
  - The bin value indicates the temporal duration of the duplicate sequence



# Temporal Recurrence Hashing (2nd-Stage)

- The computation cost of the temporal recurrence hashing algorithm is linear to $\Theta(n_p)$, which is much lower than that of related studies, e.g. $\Theta(n_p^2)$ [9] and $\Theta(n_o^2)$ [8, 10]

# Experimental Setting

- 10-Hour Stream
  - 66 commercials
  - 202 CF sequences
  - Accuracy Evaluation
- 1-Month (720-Hour) Stream
  - Efficiency Evaluation
- 5-Year (43,200-Hour) Stream
  - Efficiency Evaluation
- Linux Server
  - Intel Xeon 2.66-GHz CPU
  - 128 GB of main memory

# Evaluation Criteria

- In sequence level
  - How precisely the algorithm can detect and identify the CF sequences?

- In frame level
  - How precisely the algorithm can localize the CF sequences, i.e. from which frame the CF sequence starts and to which frame the CF sequence ends

# Experimental Results (10-Hour Stream)

| | $P_s$ (%) | $R_s$ (%) | $F_s$ (%) | $P_f$ (%) | $R_f$ (%) | $F_f$ (%) | t (sec.) |
|---|---|---|---|---|---|---|---|
| TRHA-V | 90.58 | 100 | 95.06 | 97.24 | 96.33 | 96.78 | 4 |
| DOHRING | 74.42 | 95.05 | 83.48 | 98.19 | 94.40 | 96.26 | 108 |
| BERRANI | 72.22 | 83.66 | 77.52 | 99.18 | 83.47 | 90.65 | 40 |

- DOHRING [9]
  - Apply pairwise matching to all pairs of duplicate fragments derived from the whole stream
  - □ $\Theta(n_p^2)$
- BERRANI [8]
  - Apply pairwise matching to all sets of duplicate fragments derived from the whole stream
  - □ $\Theta(n_o^2)$

# Experimental Results (10-Hour Stream)

| | $P_s$ (%) | $R_s$ (%) | $F_s$ (%) | $P_f$ (%) | $R_f$ (%) | $F_f$ (%) | t (sec.) |
|---|---|---|---|---|---|---|---|
| TRHA-V | 90.58 | 100 | 95.06 | 97.24 | 96.33 | 96.78 | 4 |
| DOHRING | 74.42 | 95.05 | 83.48 | 98.19 | 94.40 | 96.26 | 108 |
| BERRANI | 72.22 | 83.66 | 77.52 | 99.18 | 83.47 | 90.65 | 40 |

| | $P_s$ (%) | $R_s$ (%) | $F_s$ (%) | $P_f$ (%) | $R_f$ (%) | $F_f$ (%) | t (sec.) |
|---|---|---|---|---|---|---|---|
| TRHA-A | 93.93 | 99.50 | 96.63 | 99.43 | 92.03 | 95.59 | 6 |
| HAITSMA | 82.16 | 98.02 | 89.39 | 97.89 | 86.69 | 91.95 | 40 |

- HAITSMA [5]
  - Apply a fragment growing strategy to the duplicate fragment pairs
  - $\Theta(n_p)$ with high single operation cost

# Experimental Results (10-Hour Stream)

| | $P_s$ (%) | $R_s$ (%) | $F_s$ (%) | $P_f$ (%) | $R_f$ (%) | $F_f$ (%) | t (sec.) |
|---|---|---|---|---|---|---|---|
| TRHA-V | 90.58 | 100 | 95.06 | 97.24 | 96.33 | 96.78 | 4 |
| DOHRING | 74.42 | 95.05 | 83.48 | 98.19 | 94.40 | 96.26 | 108 |
| BERRANI | 72.22 | 83.66 | 77.52 | 99.18 | 83.47 | 90.65 | 40 |

| | $P_s$ (%) | $R_s$ (%) | $F_s$ (%) | $P_f$ (%) | $R_f$ (%) | $F_f$ (%) | t (sec.) |
|---|---|---|---|---|---|---|---|
| TRHA-A | 93.93 | 99.50 | 96.63 | 99.43 | 92.03 | 95.59 | 6 |
| HAITSMA | 82.16 | 98.02 | 89.39 | 97.89 | 86.69 | 91.95 | 40 |

| | $P_s$ (%) | $R_s$ (%) | $F_s$ (%) | $P_f$ (%) | $R_f$ (%) | $F_f$ (%) | t (sec.) |
|---|---|---|---|---|---|---|---|
| V+A | 96.63 | 99.50 | 98.05 | 97.34 | 97.44 | 97.39 | 10 |

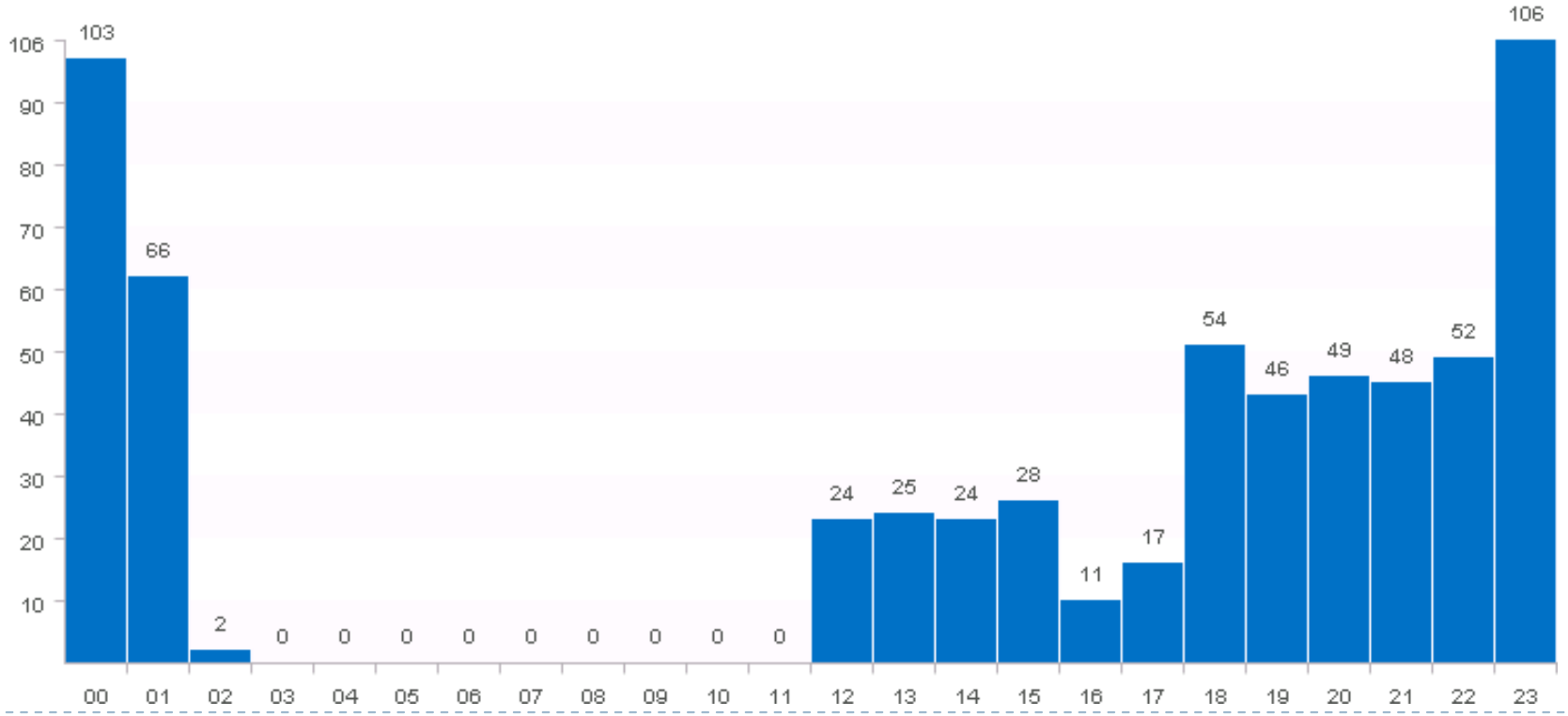
# Experimental Results (1-Month Stream)

| | Video (mm:ss) | Audio (mm:ss) |
|---|---|---|
| 1st-Stage Hashing | 03:07 | 07:29 |
| 2nd-Stage Hashing | 09:39 | 02:36 |
| Hashing Histogram Analysis | 12:22 | 01:03 |
| Recurring Fragment Assemblage | 23:21 | 29:13 |
| Post-Processing | 00:24 | 00:51 |
| Total Time | 49:43 | 41:12 |

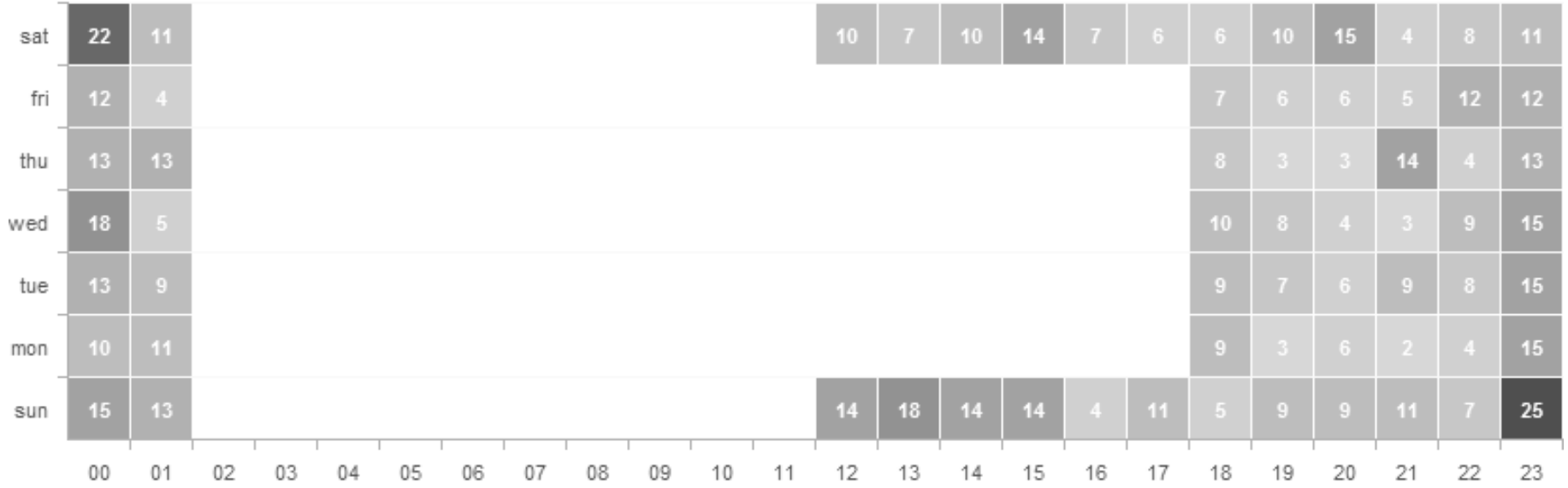
# Experimental Results (5-Year Stream)

- The 5-year stream was divided into 60 1-month streams
  - TRHA-V was individually applied to each 1-month stream
- 4 threads
  - 15 months per thread
- Processing Time
  - 21 hours

# Asahi Breweries

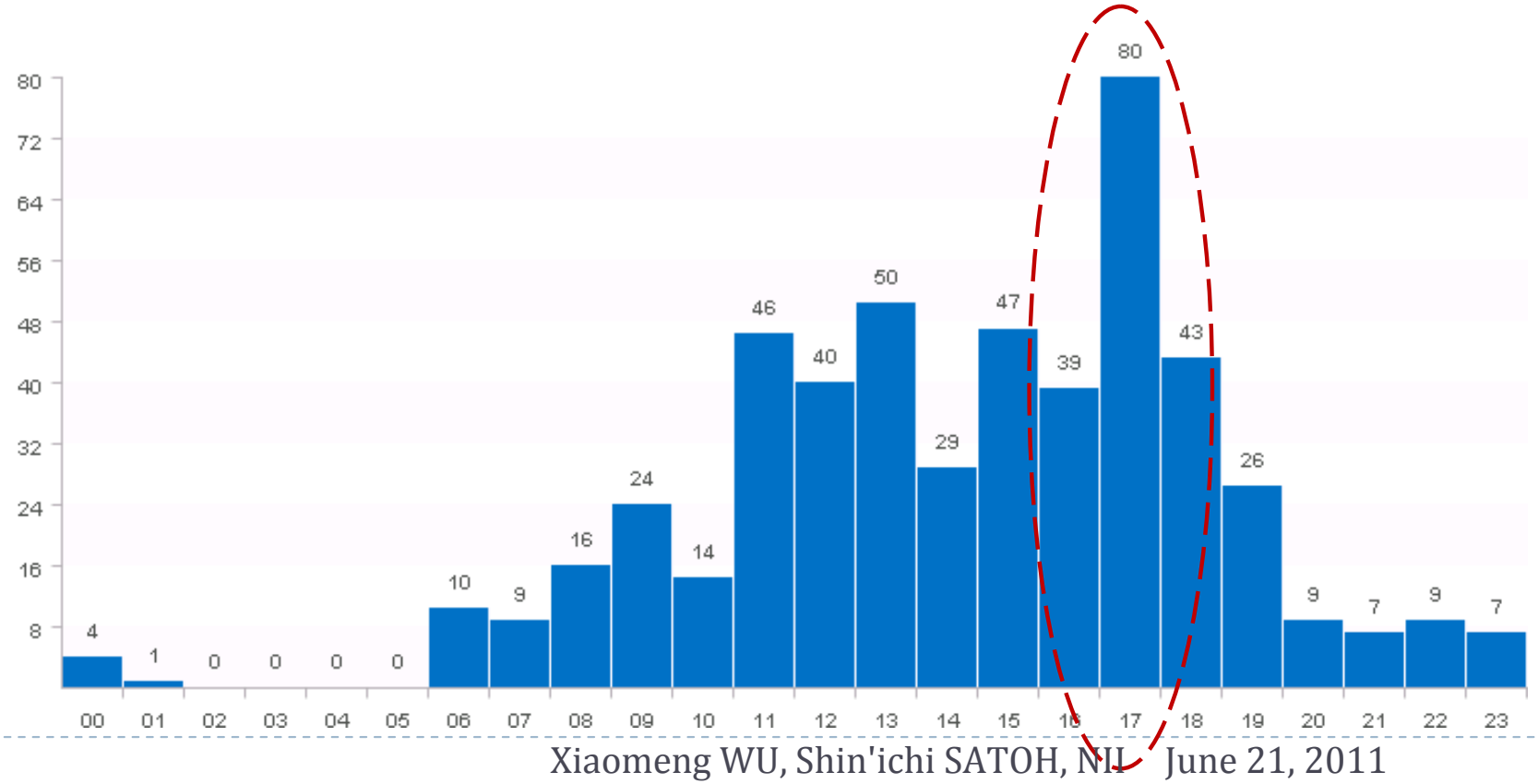



# Asahi Breweries

# Pizza-La

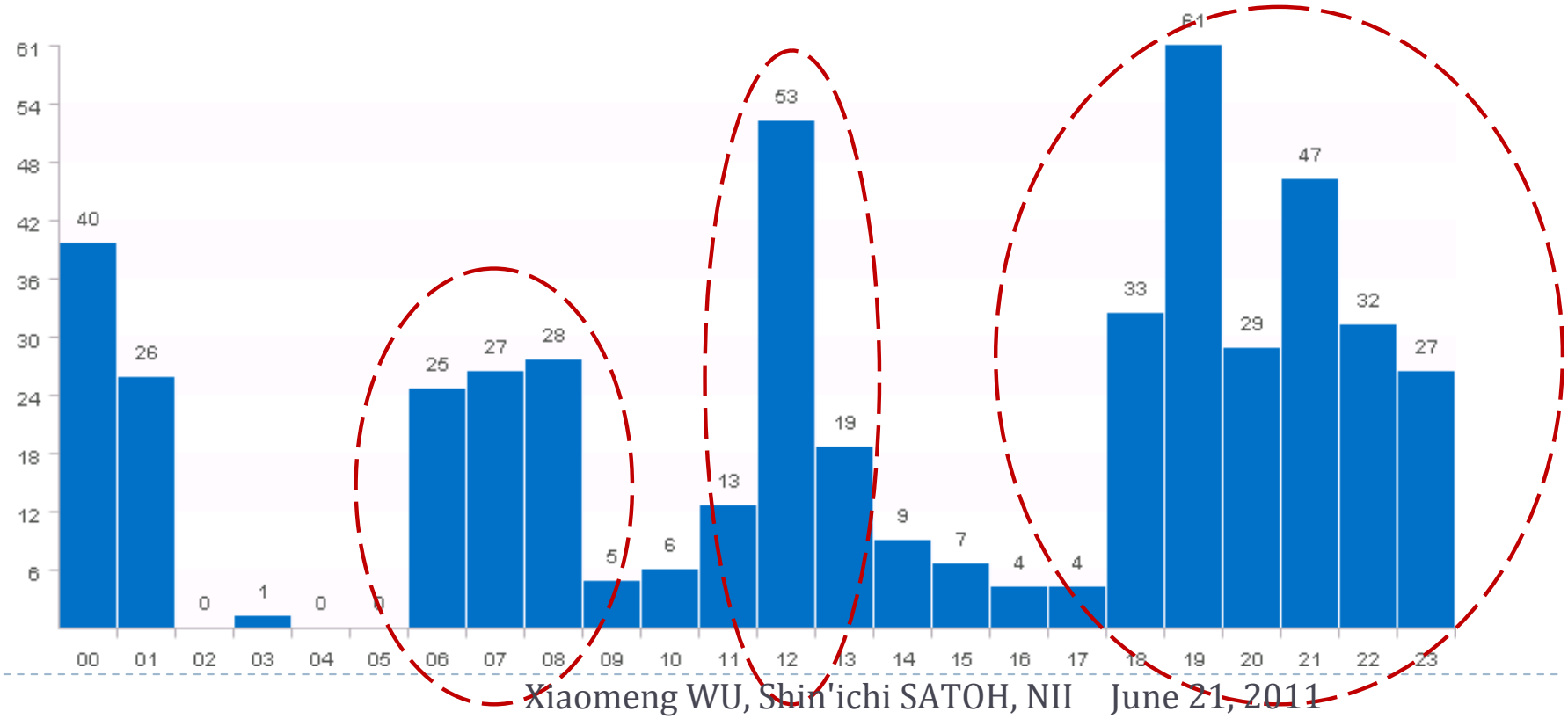

PIZZA-LA
PIZZA-LA
PIZZA-LA


80
72
64
56
48
40
32
24
16
8
4
1
0
0
0
0
10
9
16
24
14
46
40
50
29
47
39
80
43
26
9
7
9
7
00
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23

# Daihatsu Tanto Exe



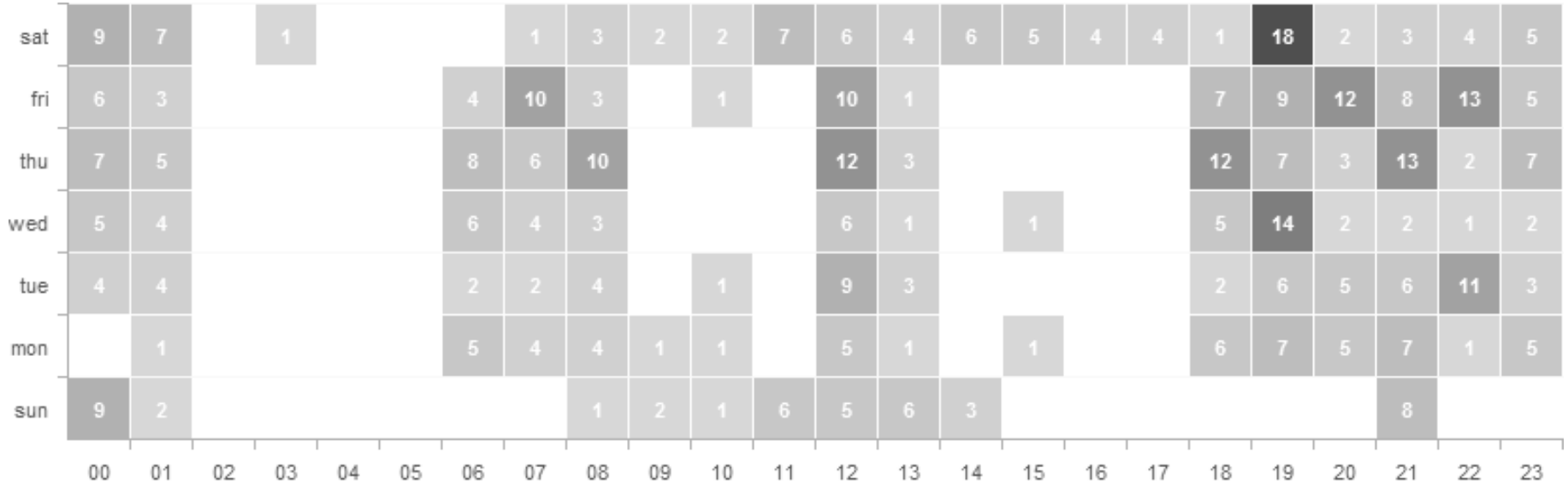




# Daihatsu Tanto Exe







# Conclusions

- Advantage
  - Fully Unsupervised
  - Generic
  - Ultrahigh-Speed

- The algorithm mined commercials, from the 1-month stream in less than 42 minutes, and from the 5-year stream in less than 21 hours, with a 98.1% sequence-level and 97.4% frame-level accuracy

- Limitation
  - Unfitted for mining near duplicates with severe transformations, e.g. translation, scaling, noising, picture in picture, etc.

- Extension
  - Investigating the performance consistency on other video fingerprinting techniques
  - Investigating the possibility of a feature-level or fingerprint-level integration of the video and audio streams

# Thank you for your kind attention.

Xiaomeng Wu, Shin'ichi Satoh
National Institute of Informatics