The TV Workstation project: a research scope

Keynote talk at the VNU-ITI

Mathieu Delalandre

University of Tours (UT), LIFAT Laboratory, RFAI group
Tours city, France
mathieu.delalandre@univ-tours.fr
Talk available at http://mathieu.delalandre.free.fr/talks.html

Hanoï city (Vietnam)
1^{st} of November, 2023

Summary

CV in short

Introduction

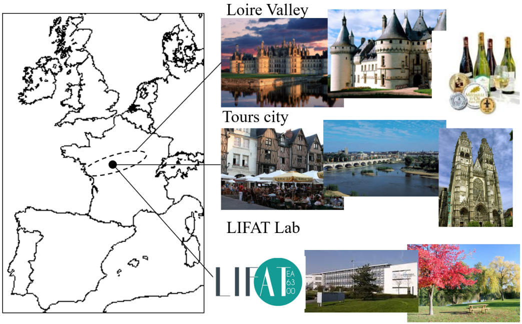
TV video capture

Real-time TV video processing

Conclusions and perspectives

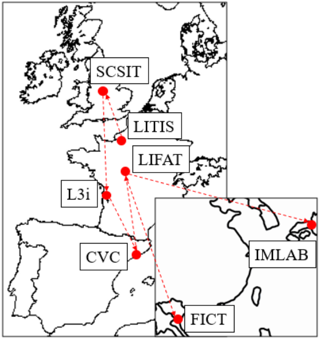
CV in short - Mathieu Delalandre (1/2)

- PhD in computer science with > 20 years of experience,
- Associate Professor at the LIFAT Lab - UT (Tours, France),
- fields of image processing and machine learning,
  - local detectors, processing in the transform domain, template matching,
- application domains,
  - video copy detection, scene text detection, document image networking, manga copyright protection, symbol/logo detection and recognition,
- journals and conferences/workshops,
  - JRTIP, TIP, PR, PRL, IJDAR,
  - CBMI, ICIAP, VISAPP, CAIP, ICPR, ICDAR, DAS, GREC.



CV in short - Mathieu Delalandre (2/2)

- international experience as ($<$ 2009) research fellows in Europe ($>$ 2013) visiting positions in Asia,
- PhD supervisor of T.A. Pham, C. Nguyen, V.H. Le and G. Vu
- head of international LIFAT-RFAI, TV Workstation project
- teacher at the Polytech school,
  - operating systems, real-time systems, distributed systems & computing,
  - head of international exchanges, networking and systems program
- more about myself: http://mathieu.delalandre.free.fr/ .



Summary

CV in short

Introduction

TV video capture

Real-time TV video processing

Conclusions and perspectives

Introduction

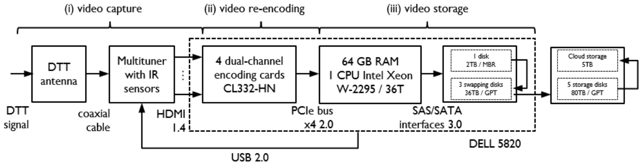
- Television (TV) is a huge source of multimedia data[1],
  - $\simeq 27,000$ channels worldwide,
  - $\simeq 55\%$ in Europe, Russia, China, USA,
  - provided with DTT, SaT, Cable TV, IPTV and InternetTV,
  - e.g. France / Vietnam ($\simeq 210$ channels), USA ($\simeq 1,760$ channels),
- Computer Vision and AI could be applied to TV,
  - Social TV, Sync2Ad, SmartZapping, fact-cheking, catchup TV, . . . ,
- A Workstation has to support the scalability / real-time issues, this leads us to develop the TV Workstation since 2017.


[1] audio/video & metadata

Summary

The DELL 5820 computer

The DELL 5820 **computer** processes 8 channels (HD, 30 FPS, 24h/day), with real-time audio / video encoding, control of tuners with IR sensors, internal / external storage of 38 + 80 TB.



Resolution		Audio/ Video	CPU rate	Video Mbps	TB/ month	Audio Kbps	GB/ month
HD	1280 × 720	asyn	20 %	3	7.23	256	621
SD	720 × 576		12 %	1.6	3.89	160	384
Low	320 × 240		8 %	0.56	1.36	128	308

Partial video copy detection (1/2)

Partial video copy detection (PVCD) aims at finding short segment(s) which have transformed in long video(s):

- it is a key topic with application domains (copyright, retrieval),
- existing datasets (VCDB, VCSL) offer no scalability, control of degradation, frame-level annotation, concistency,
- a TV-based protocol was proposed to design the STVD-PVCD dataset on the task, public available[2] with an agreement, published at [**ORASIS2021**,**ICIAP2022**] referred in the main research portails[3].

[2] `https://dataset-stvd.univ-tours.fr/pvcd/`

[3] e.g. cove.thecvf.com, datasets.visionbib.com, homepages.inf.ed.ac.uk, kaggle.com, opendatalab.com, paperswithcode.com, . . .

Partial video copy detection (2/2)

The STVD-PVCD is compared to the state-of-the-art.

Datasets	**VCDB** 2016	**STVD** 2021	**VCSL** 2022
References	28	**243**	122
Positive videos	528	**19,280**	9,207
Positive pairs	9K	**1,688K**	281K
Negative videos	100,000	64,040	N/A
Duration (h)	2,030	10,660	17,416
Noise characterization	real noise	**noise-free**	real noise
Consistency	yes	**yes**	no
Annotation cost (m-h)	700	**105**	20,000
Timestamping	1s	$\frac{1}{30}$ **s**	1s

(h): hours, (s): seconds, (m-h): man-hours and N/A: not available

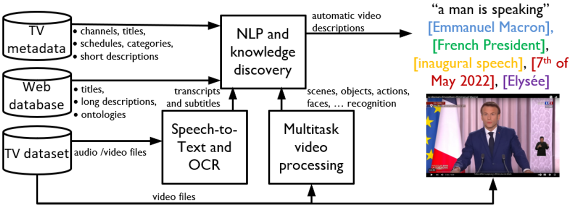
Set A Set B Set C Set D Set E Set F

STVD-PVCD allows deeper characterization tasks e.g. characterization of 2D CNN features [**CAIP2023**].

Automatic video description with knowledge representation

Automatic video description aims to tell a story about events happening in a video:

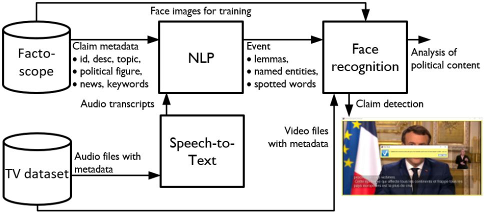
- it is a key problem in the computer vision field [**IEEE2021**],
- datasets[4] suffer from heterogeneity, scalability, lack of context and multimodal information, timing accuracy, black-box characterization,
- a TV protocol could offer a video normalization, a scalability, a knowledge representation [**PT2016**] for a robust and contextual video description.



[4] MSVD, MSR-VTT, ...

Multimodal audio/video analysis for fact-checking

Fact-checking is the process that check the veracity of claims from various media (print, TV, SNS). There is none multimodal / scalable dataset. We have designed the largest dataset STVD-FC:



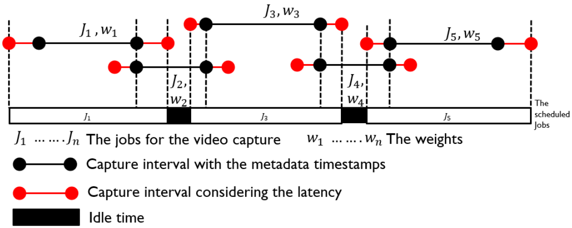
- containing $6,730$ news / political TV programs ($6,540$ h) of the French presidential election 2022[5] ($\simeq 50$ Mwords, $\simeq 706$ Mimages, 1.96 TB),
- linked to $1,300$ claims collected over 6 years (200 political figures, 241K words, 24K named entities) scraped from the Factoscope[6],
- public available[7] with an UT agreement, published at [**CBMI2022**].

[5] 1^{st} of February to 1^{st} of May 2022
[6] `https://rattrapages-actu.epjt.fr/factoscope`
[7] `https://dataset-stvd.univ-tours.fr/fc/`

Parallel machine scheduling (PMS) for video capture

Problem statement: a largest capture (e.g. 32 channels) has an important cost (32k€ + 5k€ a year for storage[8]) not needed for the applications[9]. A partial capture with PMS can be handled:

- as an off-line / no preemptive scheduling using static execution times,
- it is Weighted Interval Selection Problem (WISP) NP-hard having polynomial approximation algorithms (e.g. $GREEDY_\alpha$ [**JA2003**]),
- the latency $L(t)$ is a key parameter of the scheduling problem,
- a public available dataset STVD-PMS[10] published with an UT agreement (170 days, 26 channels, 99k jobs, 5,615 hashcodes, offline/online latency).



[8] Desktop version without maintenance and hosting / 186 TB a year (SD)

[9] Not repeated / idle, political, entertainment TV programs, . . .

[10] https://dataset-stvd.univ-tours.fr/pms/

Summary

CV in short

Introduction

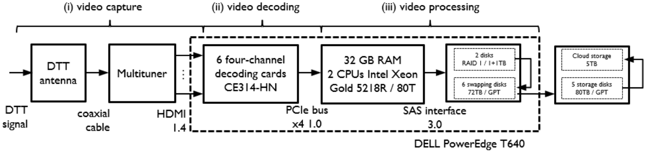
TV video capture

Real-time TV video processing

Conclusions and perspectives

The DELL PowerEdge T640 computer

The DELL PowerEdge T640 **computer** processes 24 channels for real-time video decoding and processing with high-performance CPUs[11] and having an internal / external storage of $72 + 80$ TB.



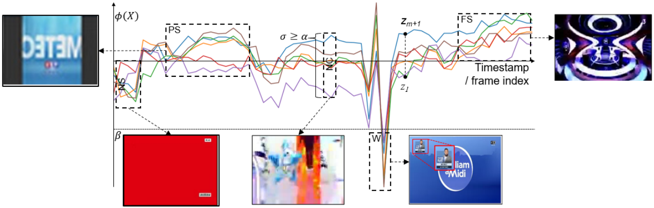
Ch	BPP	Res	FPS	Images	Bandwidth		
24	32	SD	$600 = 24 \times 25$	51.8 M/day	0.69 GB/s	57.9 TB/day	34%
		HD	$528 = 24 \times 22$	45.6 M/day	1.81 GB/s	152.9 TB/day	91%
		Full HD	$240 = 24 \times 10$	20.7 M/day	1.85 GB/s	156.4 TB/day	93%

[11] 2×40 threads with AVX 512 Vector Neural Network Instructions

Real-time PVCD

Real-time PVCD processes with a deadline Δ (e.g., 1-3s) and can be applied to multiple video streams [**CBMI2021**,**CAIP2023**]:

- with real-time video decoding using hardware on the Workstation,
- with rigid (ZNCC) and no-rigid (2D CNN) features for matching,
- with key-frame selection methods using goodness criteria.



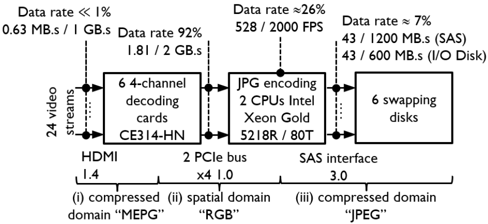
Time optimization for real-time deep learning to investigate:

- acceleration [12] with INT8 and VNNI [**CCIS2020**],
- soft real-time with adaptive inference [**PR2020**].

[12] $\simeq \times 15$ acceleration on *ResNet-50* (OpenVino vs. TensorFlow).

Real-time frame capture, IQA and NDD (1/2)

Real-time frame capture decodes videos into frames re-encoded as image files (e.g. jpeg). The workstation can process 24 streams (22 FPS / HD) in real-time with its hardware architecture.



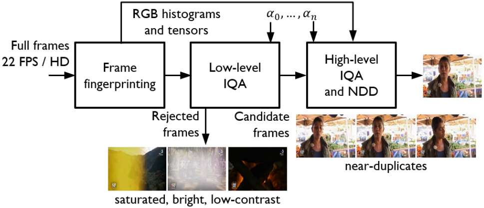
No bottleneck appears, the problem comes from the storage cost.

	Day	Month	Year
image	45.6 M	1.4 B	16.7 B
data	3.4 TB	103.2 TB	1.22 PB

M, B, TB, PB stand for Millions, Billions, Terabyte, Petabyte

Real-time frame capture, IQA and NDD (2/2)

Image Quality Assesment (IQA) and Near-Duplicate Detection (NDD) filter out low quality and duplicate images.



- standard video processing supports low-level IQA and NDD, high-level IQA requires time-efficient blur detection methods [**CIS2023**],
- parameters $\alpha_0, \ldots, \alpha_n$ are set for storage requirements (e.g. $\simeq$ 12 FPM).

Summary

CV in short

Introduction

TV video capture

Real-time TV video processing

Conclusions and perspectives

Conclusions and perspectives

- project launched in 2017, specific / ready-to-use platform,
- 2 PhD grants in progress (V.H. Le, J. Vu), applications to other grants (VIED 89, French Embassy),
- $\simeq$ 40 k€ of investment, 8 researchers working on,
- 6 publications[13] and 3 public available datasets STVD[14],
- cross-disciplinary project (CV, NLP, OR),
- perspectives with key research topics (video description, real-time deep learning, ...),
- projects in the queue (social TV, Fact-Checking).

[13] **[AI4TV2019, CBMI2021, ORASIS2021, ICIAP2022, CBMI2022, CAIP2023]**

[14] `https://dataset-stvd.univ-tours.fr/`

References I

- **[JA2003]** T. Erlebach and F.C.R. Spieksma. Interval selection: Applications, algorithms, and lower bounds. Journal of Algorithms, vol. 46(1), pp. 27-53, 2003.
- **[PT2016]** J.L. Redondo García. Semantically Capturing and Representing Contextualized News Stories on the Web. PhD Thesis, TELECOM ParisTech, 2016.
- **[AI4TV2019]** M. Delalandre. A Workstation for Real-Time Processing of Multi-Channel TV. Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV), pp. 53-54, 2019.
- **[CCIS2020]** E.P. Vasiliev and al. Performance Analysis of Deep Learning Inference in Convolutional Neural Networks on Intel Cascade Lake CPUs. Mathematical Modeling and Supercomputer Technologies (MMST), Communications in Computer and Information Science (CCIS), vol. 1413, 2020.
- **[PR2020]** N. Passalis and al. Efficient adaptive inference for deep convolutional neural networks using hierarchical early exits. Pattern Recognition (PR), vol. 105, pp. 107346, 2020.
- **[IEEE2021]** M. Rafiq and al. Video Description: Datasets & Evaluation Metrics. IEEE Access, vol. 9, 2021.
- **[CBMI2021]** V.H. Le, M. Delalandre and D. Conte. Real-time detection of partial video copy on TV workstation. Conference on Content-Based Multimedia Indexing (CBMI), pp. 1-4, 2021.
- **[ORASIS2021]** V.H. Le, M. Delalandre and D. Conte. Une large base de données pour la détection de segments de vidéos TV. Journées Francophones des Jeunes Chercheurs en Vision par Ordinateur (ORASIS), 2021.
- **[CBMI2022]** F. Rayar, M. Delalandre and V.H. Le. A large-scale TV video and metadata database for French political content analysis and fact-checking. Conference on Content-Based Multimedia Indexing (CBMI), 2022.
- **[ICIAP2022]** V.H. Le, M. Delalandre and D. Conte. A large-Scale TV Dataset for partial video copy detection. International Conference on Image Analysis and Processing (ICIAP), Lecture Notes in Computer Science (LNCS), vol. 13233, pp. 388-399, 2022.
- **[CAIP2023]** V.H. Le, M. Delalandre and H. Cardot. Performance characterization of 2D CNN features for partial video copy detection. International Conference on Computer Analysis of Images and Patterns (CAIP), Lecture Notes in Computer Science (LNCS), vol. 14184, pp. 205-215, 2023.
- **[CIS2023]** X. Wang and X. Liang and S. Li and J. Zheng. Efficient image blur detection via hierarchical edge guidance and region complementation. Journal: Complex & Intelligent Systems, 2023.