A Workstation for Real-Time Processing of Multi-Channel TV

Mathieu Delalandre
mathieu.delalandre@univ-tours.fr
LIFAT Laboratory
Tours, France

ABSTRACT

This papers presents the architecture of a workstation for the real-time processing of multi-channel TV. A first application for real-time detection of video copy is discussed.

KEYWORDS

Workstation; real-time; multi-channel TV; video detection

ACM Reference Format:
Mathieu Delalandre. 2019. A Workstation for Real-Time Processing of Multi-Channel TV. In *1st International Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV '19), October 21, 2019, Nice, France*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3347449.3357489

1 INTRODUCTION

Over the last ten years, connected TVs (smart TVs, TV apps on mobile) became dominant in the industry and market. Compared to traditional TVs, the devices offer new services to users as interactive media, on-demand content and networking access. The recent trends for parallel computing, advances in real-time image processing, machine learning and artificial intelligence, embedded and hardware solutions, make possible the design of systems for real-time video analysis. This could offer a wide range of applications for TV industry and users as the video-based soccer analysis [1] or the advertising, logo and text detection [2-4].

In this paper, we present a workstation for real-time processing of multi-channel TV. This workstation is desiged at the LIFAT Laboratory. It takes part of a global research program called *ImageStream* with industrial and academic partnerships. The *ImageStream* program aims to promote the activities at the LIFAT Laboratory on the topics of real-time image processing, machine learning and artificial intelligence.

The rest of this paper is organized as follows. In section 2, the workstation is presented. Section 3 details a first application for video copy detection. Section 3 will conclude and propose some perspectives.

AI4TV '19, October 21, 2019, Nice, France

ACM ISBN 978-1-4503-6917-6/19/10.
https://doi.org/10.1145/3347449.3357489

2 THE WORKSTATION

Fig.1 presents our workstation. We will discuss different issues related to the capture of multi-channel TV and the hardware architecture for real-time processing.

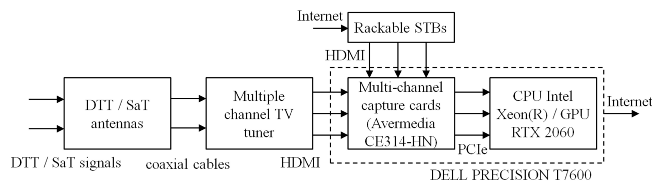

Figure 1: The workstation

The TV streams are broadcasted over different networks including the IPTV, Internet TV, SaT and DTT signals. IPTV and Internet TV services have gained much interest over the last years. They present a main benefit for the users in term of connectivity. However, the IP based TVs suffer from a big latency and jitter while broadcasting, compared to DTT and SaT signals [5]. This is a critical point for real-time applications, where the notification of events must respect a deadline to fit with the requirements of users. Another core problem is to access the video and image data. The IP and Internet TVs are delivered with no open-source softwares where the client-side APIs are designed for user experience. They require a large network bandwidth for acquisition.

To solve these problems, the main capture in our workstation is driven from the DTT signal. The DTT signal is processed with a multiple channel TV tuner. This tuner demodulates the DTT signal into multiple channels. The connexion with the workstation is ensured with HDMI interfaces using the MPEG format. Not all the TV channels are delivered through the DTT. For coverage, our workstation processes with additional entries. The first is the SaT signal. The second is obtained from Internet with rackable set-top-boxes (STBs).

Tab. 1 gives the performances of our workstation. It is a DELL PRECISION T7600 workstation having an Intel Xeon(R) CPU E5 – 2620 2GHz. This CPU has a near performance of 68 GFLOPS SP and is supported with a GPU RTX 2060 of 7180 GFLOPS SP. The workstation embeds multi-channel capture cards. These cards drive the video acquisition at the hardware level to control the FPS, the video downscaling, the color-space conversion, the video decoding and transfer to the main memory.

Several cards are available on the market. The main providers are AVerMedia, Blackmagic Design and Magewell. We have selected the Avermedia CE314−HN cards processing 4 Full

HD channels at 30 FPS. These cards are plugged to the PCIe bus 3.0×4 in the workstation having a maximum bandwidth of 3.94 GB.s. For the need of performance, we use the YUY2 color space to process the frames in gray level. The workstation embeds 8 cards able to process $8 \times 4 = 32$ TV channels requiring a bandwidth of 3.7 GB.s. For scalability, a cluster of workstations must be considered.

Ch	**FPS**	**Res**	**CS**	**BW** GB.s	**Data** TB.day	**CPU/GPU** GFLOPS SP
32	30	Full HD	YUY2	3.7	312	68/7180

Channels, **CS** color space, **BW** bandwidth, SP is single precision

Table 1: Performances of the workstation

3 VIDEO COPY DETECTION

We present in this section a first application of our workstation for real-time detection of video copy in multi-channel TV. The video copy detection is key topic in image processing and computer vision [6]. It can be applied to detect advertising spots, news, interviews or jingles in the TV streams. The detection deals with exact duplicate images under noisy conditions due to the MPEG Video Coding [7]. The overall process must fit with a real-time constraint.

In our approach we have considered the template matching for detection. Template matching applies a direct comparison requiring no image fingerprinting. The detection problem is not concerned with rotation and scale invariance. We have applied the normalized cross-correlation (NCC) Eq. (1) deviating from the sum of squared differences (SSD) [8].

$$NCC = \frac{\sum_{(x,y)\epsilon W}(I_{x+i,y+j} - \overline{I_{i,j}}).(T_{x,y} - \overline{T})}{\sqrt{\sum_{(x,y)\epsilon W}(I_{x+i,y+j} - \overline{I_{i,j}})^2}.\sqrt{\sum_{(x,y)\epsilon W}(T_{x,y} - \overline{T})^2}} \quad (1)$$

In (1), $T_{x,y}$ is a discrete function of a template taking values in a window W with $(x, y) \in W$. The template is compared with a sub-image of I at a given position (i, j). $\overline{T}$ and $\overline{I_{i,j}}$ are the means of the template and the sub-image, respectively. $NCC \in [-1, 1]$ with 1 is the perfect correlation. The detection of a duplicate frame can be achieved with the comparison of one to three templates as an average. The selection of discriminative templates can be done manually or automatically with artificial intelligence methods [9].

Tab. 1 gives the maximum number of template supported in our workstation for a real-time detection. The experiments have been driven using 32 channels for capture using only the CPU for processing. The matching and NCC metric Eq. (1) have been implemented with a full parallelism support[1]. The maximum number of template has been fixed in order to respect a rate of 30 FPS with a hard deadline. None matching of a template with a frame can exceed 1/30 s.

As highlighted in Tab. 1, our workstation supports from 2K to 20K templates at a mean size for a real-time comparison. These performances will be improved while using the GPU. Next, optimization can be obtained with reformulation of the matching problem. Indeed, in Eq. (1), the right terms of the numerator and denominator are related to the template T and can be computed off-line. The left terms comparing I and T can be supported with image integral for acceleration [8]. At last, the NCC can be controlled with upper bounding [10]. In that case, a fast elimination condition is verified at every matching. This elimination condition relies on the evaluation of an upper-bound for the NCC function. The workstation could be then designed to process millions of templates.

[1]Multi core and thread, vectorization, SSE and GPU.

Template size	32^2	64^2	96^2
Template number	22.2K	4.3K	2K

K is thousand

Table 2: Maximum numbers of templates matched at 30 FPS in hard real-time with the CPU

4 CONCLUSIONS

This papers has presented the architecture of a workstation for real-time processing of multi-channel TV. This workstation can capture 32 channels in gray level at 30 FPS in real-time. It has a near performance of 7 TFLOPS SP fully dedicated to the post-capture processing. A first application for real-time detection of video copy is discussed. We have applied template matching to solve the problem. Our workstation can support from 4K to 20K templates for the detection of duplicate frames in real-time with the CPU. Large improvements could be obtained while considering a processing with the GPU and a reformulation of the matching problem.

BIBLIOGRAPHY

[1] M. Schlipsing and al, Adaptive Pattern Recognition in Real-time Video-based Soccer Analysis, JRTIP (2017).

[2] Y. Liao and al, Mutual Enhancement for Detection of Multiple Logos in Sports Videos, ICCV (2017).

[3] M. Hossar and al, ADNet: A Deep Network for Detecting Adverts, CEUR Workshop (2018).

[4] R. Kanna and al, Overlay Text Extraction From TV News Broadcast, INDICON (2015).

[5] M. Dabrowski and al, Analysis of video delay in Internet TV service over adaptive HTTP streaming, FedCSIS (2015).

[6] D.A. Phalke and al, A Systematic Review of Near Duplicate Video Retrieval Techniques, IJPAM (2018).

[7] M.A. Ali and al, Efficient Spatiotemporal Matching For Video Copy Detection in H.264/AVC Video, IJCA (2012).

[8] G. Facciolo and al, Integral Images for Block Matching, IPOL (2013).

[9] A. Penate-Sanchez and al, Matchability Prediction for Full-Search Template Matching Algorithms, 3DV (2018).

[10] G. Adhikari and al, Fast real time object tracking based on normalized cross correlation and importance of thresholding segmentation, ICRTIT (2016).