

A Novel System for Automatic Detection of TV Commercial Breaks

Nicolás Rondan , Marcos Juayek , Jose Joskowicz , *Senior Member, IEEE*, Rafael Sotelo , *Senior Member, IEEE*, Santiago Quincke , Andrés Patrone , Maximiliano Aguerre , and Gastón Gonzalez 

Abstract— This paper explores the problem of automating the detection of commercial breaks in live broadcasts and video recordings, distinguishing them from regular channel content by focusing on visual signals, commonly referred to as “bumpers”, that clearly delineate commercial content from regular programming. Accurate measurement of commercial break duration is crucial for regulatory compliance, ensuring timely airing for advertisers, and maintaining verifiable records for broadcasters. Through an examination of regulatory frameworks in various countries—such as Spain, Uruguay, and the United States—this paper discusses how these regulations are applied and explores the potential for automated detection technologies to aid broadcasters and regulators in monitoring compliance. This paper details the design, implementation, and performance analysis of a novel system for automatic TV advertising detection.

Link to graphical and video abstracts, and to code: <https://latam.ieceer9.org/index.php/transactions/article/view/9985>

Index Terms— TV Commercial Break Detection, Image Hashing, Video Hashing, TV Logo Detection.

I. INTRODUCTION

AUTOMATIC detection of commercial breaks within live broadcasts and recorded videos is a growing focus in media research, driven by the needs of regulators, advertisers, and broadcasters alike. The objective of automatic advertising detection is to detect which sections of a live broadcast, or a video recording, correspond to commercial breaks and which correspond to the main program of the channel. Accurate identification of the start, end, and duration of commercial segments is essential for various stakeholders, including regulators, advertisers, and broadcasters: regulators want to control if the current legislation is applied, advertisers want to make sure that their spots are shown at the appropriate time, and broadcasters need to have evidence of the break’s duration.

Many countries have strict rules regarding the duration of commercials on free-to-air TV. For example, the European Advertising Standards Alliance (EASA) stipulates that the

The associate editor coordinating the review of this manuscript and approving it for publication was Roberto S. Murphy (*Corresponding author: Jose Joskowicz*).

This work was supported by ANII/URSEC under Grant DPU_S_2022_1_172807.

N. Rondan, M. Juayek, Jose Joskowicz, R. Sotelo, S. Quincke, A. Patrone, M. Aguerre, and G. Gonzalez are with University of Montevideo (e-mails: nrondan@correo.um.edu.uy, mjuayek@correo.um.edu.uy, jjoskowicz@correo.um.edu.uy, rsotelo@ieee.org, squincke@correo.um.edu.uy, apatrone@correo.um.edu.uy, maguerre1@correo.um.edu.uy, and ggonzalez14@correo.um.edu.uy).

proportion of television advertising and teleshopping spots from 6:00 AM to 6:00 PM must not exceed 20% of that timeframe. Similarly, from 6:00 PM to midnight, the proportion of advertising and teleshopping spots should also not exceed 20% of the period [1]. The USA and other countries also have different rules that define the types of advertising and establish time limits for each type.

In addition to time limits, television advertising regulations emphasize the importance of distinguishing commercial content from regular programming to avoid confusion for viewers and to ensure transparency. This is often achieved through the use of *bumpers* — short visual or auditory announcements that signal the start and end of a commercial break. These bumpers play a crucial role in ensuring that both broadcasters and advertisers comply with regulations, which often include limits on advertising duration within a set timeframe.

This paper describes the design, development, implementation, and results of a novel system for automatic detection of TV commercial breaks based on the detection of start and end bumpers. The project originated in a public challenge called by the Uruguayan Regulatory Unit of Communication Services (URSEC) in collaboration with the Uruguayan National Research and Innovation Agency (ANII) [2]. The challenge sought an innovative technological solution for monitoring and control of advertising segment lengths and ensuring compliance with national regulations. The objective of this system is to measure the total duration of the commercial segments using an automatic advertising detection solution that can run in near real-time. The developed system can process many different TV channels at the same time, requires low processing power, and is easily extensible to different types of regulations and broadcasting technologies. The detection of non-traditional TV advertising, station promotional material (promos), or public service announcements was out of the scope of the developed system.

The paper is organized as follows: Section II reviews the regulatory frameworks governing TV advertisements and TV Commercial breaks across multiple countries, underscoring the global need for automated compliance solutions. Section III describes different works related to the topic. Section IV examines various techniques for automatic advertising detection, including both signal-based and content-based approaches, and provides insights into the technologies and algorithms that enhance detection accuracy. Section V focuses on the application of these techniques in free-to-air television, followed by Section VI, which presents the system architecture for the proposed automatic advertisement detection solution. Section VII evaluates the system's

performance through experimental results, and finally, Sections VIII and IX discuss potential future advancements and conclude the findings.

II. TV COMMERCIALS GUIDELINES AND REGULATIONS

Across various countries, free-to-air television channels are subject to regulations that govern both the duration of commercial breaks and the manner in which these breaks are distinguished from regular programming. These regulations are crucial for maintaining transparency and ensuring that broadcasters, advertisers, and viewers are all in compliance with national broadcasting standards.

For instance, in Uruguay, free-to-air TV is regulated by the recent Law No. 20,383 [3], which stipulates that each commercial break must be accompanied by station announcements at both the beginning and the end, often in the form of bumpers. These bumpers typically consist of brief animations or fixed frames with messages like “Start of Commercial Segment” and “End of Commercial Segment” to clearly indicate when the viewer is moving from program content to advertising.

Spain, in Law 7/2010, states that both, television advertising messages and teleshopping, must be clearly differentiated from the programs by means of acoustic and optical mechanisms [4].

In the United States, the Federal Communications Commission (FCC) has long mandated that advertising, especially in children's programming, be clearly separated from the main content. Since 1974, the FCC has required the use of bumpers—often a simple phrase like “And now, a word from our sponsors”—to help viewers, particularly children, differentiate between programming and advertisements. The FCC also enforces the host selling policy, which prohibits program characters from appearing in commercials during their own shows, further emphasizing the need for separation between content and advertising by intervening and unrelated program material [5].

In Canada, the Canadian Radio-television and Telecommunications Commission (CRTC) enforces similar rules, requiring broadcasters to clearly inform viewers from commercial advertising, though bumpers are not explicitly required [6].

India's Ministry of Information and Broadcasting (MIB) caps advertising at 12 minutes per hour under the Cable Television Networks Rules. Advertisements must be clearly distinguishable from content, with an emphasis on ethical practices. To ensure compliance, the MIB has established monitoring mechanisms, including the Electronic Media Monitoring Centre (EMMC), which oversees the content broadcasted by television channels [7].

In Australia, the Australian Communications and Media Authority (ACMA) regulates advertising under the Commercial Television Industry Code of Practice, requiring clear separations between content and commercials. ACMA also restricts certain product advertisements, such as unhealthy foods, during specific programming [8].

In Argentina, the Audiovisual Services framework (Law 26,522 and its implementing Decree 1,225/2010, as amended by Decree 678/2016) requires that advertising be kept separate from programming; historically, TV ad blocks also had to be explicitly labeled as “*espacio publicitario*”, and although Decree 678/2016 removed the labeling obligation, the separation mandate and advertising time limits remain in force [9]. In Brazil, statutory provisions are complemented by industry self-regulation: the Brazilian Advertising Self-Regulation Code (CONAR) requires that advertisements be clearly distinguishable as such and that sponsors be clearly identifiable across media, discouraging concealed or “ambush” practices [10].

In practice, these requirements are operationalized on-air through routine bumpers and identifiers that delineate commercial breaks and avoid confusion with editorial content, aligning with the separation criteria surveyed in this section.

These regional regulations collectively uphold standards for transparency in advertising, ensuring viewers can easily identify and differentiate commercials from programming. These rules typically mandate that audio volume must not increase at the start of a commercial break (in some cases following ATSC recommendation established in [11]), each break must begin and end with a clear announcement to separate it from the program, and there are usually time limits on the duration of commercial breaks during programming.

While the specific rules may vary by country, a common feature is the requirement for a clear separation between commercials and regular programming, often achieved through bumpers. It consists of short animations or fixed frames that last for a few seconds. The bumper displays a text like “Start of Commercial Segment” or “Publicity” when moving from the program to a commercial break, and another that indicates the end of commercials with a text similar to “End of Commercial Segment”. Figure 1 shows screenshots of one frame of different bumpers. Traditional commercials are allowed to be broadcasted in the time segment between bumpers.

III. RELATED WORK

Historically, TV ad or commercials detection relied on manual reviews or basic rule-based methods, which were labor-intensive and error-prone. However, recent advancements in machine learning and computer vision have significantly improved the ability to automatically detect and classify advertisements. Modern approaches employ a variety of techniques, including machine learning classification, clustering, and advanced feature extraction, to recognize ads in complex broadcast environments. For example, [12] compared clustering and classification algorithms to assess their effectiveness in segmenting commercial content, highlighting the precision and scalability of machine learning in ad detection.

Current methodologies for automatic advertisement detection fall into two main categories: signal-based techniques and content-based techniques. Signal-based

techniques analyze specific properties of the audio-visual signal, such as the presence of black frames or audio depressions, which can signal the beginning and end of commercial segments. On the other hand, content-based techniques utilize elements within the video content itself, such as the detection of channel logos, bumpers, and recurring video sequences, often leveraging advanced video retrieval and hashing methods to achieve precise identification [13].



Fig. 1. Example of commercial bumpers from different TV broadcaster from Uruguay (up and middle) and Spain (bottom) pointing the start of commercial segment (at the left), and its end (at the right).

Advancements in machine learning further enhance these techniques, with algorithms like support vector machines (SVM) and deep learning models improving the accuracy of both logo and content recognition. For instance, logo detection frameworks can identify brand-specific content even without a pre-existing database, enabling adaptive recognition across varying logo designs and transparency levels [14]. Additionally, video and image hashing techniques allow systems to efficiently index and retrieve ad segments based on short, unique binary representations of video content, as demonstrated by recent work in robust video [13], [15].

These and other technologies are presented and discussed in the next section.

IV. AUTOMATED ADVERTISING DETECTION METHODS

As mentioned in the previous sections, the approaches to make automatic advertisement detection can be classified into two broad groups: techniques based on signal characteristics and techniques based on the content. Signal-based techniques focus on analyzing abstract characteristics of the signal without interpreting the audio or video content. For example, these methods may evaluate frame color patterns at specific moments or detect variations in audio volume. In contrast, content-based techniques examine the actual content of the signal, such as identifying specific markers like channel logos or matching segments against a database of known

advertisements. Among content-based methods, Optical Character Recognition (OCR) can also be employed to detect and interpret text associated with specific advertisements.

A. Techniques based on Signal Characteristics

Two common signal-based characteristics are frequently used for advertisement detection: variations in audio volume, which often signal the start or end of commercial segments, and the presence of black frames, typically marking transitions between programming and advertisements.

a) Audio Silence Detection

In certain instances, audio silence, also referred to as “audio depression”, commonly occurs during transitions between commercials and regular programming, such as a brief silence lasting around 240 milliseconds. This silence can be identified by analyzing the channel's audio signal. Detecting these audio depressions can signal either a scene change within the programming or a transition between consecutive advertisements. In , the proposal is based on detecting a certain frequency of occurrence of silences within the advertising spots.

b) Black Frame Detection

In some countries such as Turkey and Ireland, TV channels employ a technique known as the “black frame code” to signal the start and end of advertising segments. This method involves inserting a brief sequence of black frames into the video signal just moments before an advertisement begins. While these frames are typically imperceptible to viewers, they can be reliably detected through automated signal analysis algorithms [16], [17].

There are multiple techniques to detect black frames in a video signal. For example, converting each frame to grayscale and if the average of the pixels value is less than a certain threshold it will be counted as a black frame. If a certain number of consecutive black frames is detected, an advertising segment can be inferred. Another technique uses an average of the continuous component (DC) of the Discrete Cosine Transform (DCT) of several consecutive frames to determine whether it is a sequence of black frames or not.

When black frames and audio depression are used to indicate the start of the commercial break, the combination of the two above-mentioned detection methods results in good detection accuracy, as described in [17]. In [18], the authors proposed a robust audio-only method that uses regression models to distinguish true advertising boundaries from non-boundary silences, achieving high accuracy on real broadcast datasets.

B. Techniques based on Signal Content

These techniques focus on analyzing audio and video content to identify specific elements within the signal, such as channel logos, recognized video segments (e.g., previously known advertisements), familiar images (e.g., station bumpers), or text that appears in the video frames.

a) Logo Detection

Television broadcasters commonly display a station logo on-screen to identify their channel. As a standard practice, this logo is often removed during commercial breaks. Consequently, various authors have explored methods for using these logos as markers to detect the start and end of advertising segments [14], [19], [20].

One approach for logo detection is using a database of channel logos that need to be identified within a video [21], [22]. However, this method faces challenges due to the dynamic nature of logos, which frequently changes. Furthermore, prior knowledge of logos is necessary for accurate identification. Even the same logo might undergo variations in transparency, color, and size for special occasions or different times of the year. To address these issues, alternative systems aim to eliminate the reliance on a predefined database. These systems focus on recognizing logos based on their spatial location and fundamental characteristics, thereby circumventing the need to store logos in a database [20]. According to the referenced literature, the accuracy of logo detection based on logo databases is between 95% and 98%, while the methods that do not require a logo database are between 88% and 95%.

b) Video Retrieval

Video retrieval refers to the process of locating and extracting relevant videos from a database using content-based indexing and retrieval methods. This approach typically involves segmenting the video, extracting meaningful features, and applying dimensionality reduction techniques to optimize search efficiency and accuracy [23].

Machine learning techniques have been widely applied in the video retrieval literature to support video segmentation, indexing, and retrieval. The most common algorithms include K-means clustering, Support Vector Machine (SVM), and Nearest Neighbors. On the other hand, the use of deep learning for video retrieval and advertisement detection is growing [24], [25], [26]. A literature review of video processing using deep learning techniques can be read in [27], and a survey on deep learning for video objects detection can be found in [28].

Video segmentation is used to separate a video into multiple units which can improve indexing. It is a technique that generally consists of extracting information to describe the frames and identify the limits of the segments. Shot Boundary Detection (SBD) techniques are commonly used. Segmentation can also benefit from machine learning algorithms, such as clustering techniques, to group frames into cohesive units based on visual or audio characteristics [29].

Feature extraction involves identifying and capturing perceptual characteristics of a video, such as color, texture, and motion. These features serve as indexes, enabling the comparison and retrieval of similar videos. The generation of these indexes or characteristics is usually known as video fingerprinting, a compact representations of video frames, often as unique hashes, enabling rapid matching with known content. Robust video hashing methods, such as those

proposed by [13] developed a geometrically robust hashing method to improve the reliability of detecting repeated ad content across different broadcast scenarios. Recent advances in deep learning have introduced convolutional neural network (CNN) based feature extractors, which outperform traditional methods by learning high-level semantic features from video content. For example, [30] utilized transfer learning with pre-trained CNN models to extract audio-visual features for commercial break detection, achieving high accuracy across diverse broadcast formats.

Video retrieval generally depends on similarity calculation based on indexes (dimensions). This search can be complex if the dimension space is too large. Dimensionality reduction is a technique used to reduce the number of indexes and/or create new ones that are more relevant and not redundant. Techniques such as principal component analysis (PCA) and T-distributed stochastic neighbor embedding (t-SNE) are commonly used to reduce the dimensionality of extracted features, improving the efficiency of similarity calculations.

Video retrieval depends heavily on similarity metrics to match query video segments with reference content. Common approaches include calculating Euclidean distances, cosine similarity, or correlation coefficients between feature vectors. Advanced methods such as video hashing further optimize these comparisons. [31] introduced a Base64 Encoding technique that transforms video frames into encoded strings, enabling rapid comparisons through string matching. While not as robust as hashing, this method offers a lightweight and computationally simple alternative for detecting recurring video segments.

c) Image Retrieval

Videos are essentially sequences of frames, allowing image retrieval techniques to be applied to individual frames for analysis. One widely adopted method is image hashing, a technique that generates compact binary representations based on key visual features extracted from the image content. These binary hashes act as unique fingerprints, enabling efficient comparison and detection of similarities and differences between images, even after transformations such as resizing, compression, or slight modifications [32]. The process relies on hash functions—mathematical algorithms that convert variable-sized data into fixed-length outputs—to create these image fingerprints. Similarity is typically measured using the Hamming distance, which counts the number of differing elements between two hashes. Values close to zero indicate high similarity. Although hash values are not strictly unique, they are sufficiently distinct for reliable matching. Common approaches include Average Hash (aHash) [33], Difference Hash (dHash) [34], Perceptual Hash (pHash) [35], and Wavelet Hash (wHash) [36], which differ in their feature extraction strategies.

In [37], an advertising detection method for broadcast environments applies dHash to each frame of segmented video sequences, concatenates the results, and then applies the Discrete Cosine Transform (DCT) to produce a fingerprint that is matched against a database of known advertisements.

When paired with a reference database of known advertisement frames, image hashing becomes a powerful tool for accurately identifying commercials. This method proves particularly effective in scenarios where bumpers are used to signal the beginning and end of advertising segments.

d) Text Content

The technology of text content detection on video has been used for different purposes. In early and subsequent works, automatic optical character recognition (OCR) is used to read news titles and lower-thirds for indexing and retrieval in broadcast archives [38], [39], [40], [41]. Recommendation and cataloging pipelines also exploit OCR to derive searchable metadata and build video catalogs at scale, as illustrated by industrial deployments and cloud services [42], [43], [44]. A comprehensive survey of video text detection, tracking, and recognition—covering methods, datasets, and evaluation protocols—can be found in [45], while community benchmarks such as the ICDAR Robust Reading Competitions include dedicated “Text in Videos” tracks [46].

Methodologically, contemporary pipelines separate detection from recognition. Widely adopted detectors (e.g., EAST, CRAFT, DB/DBNet) localize multi-oriented or curved text efficiently [47], [48], [49], and recognizers range from convolutional recurrent neural network (CRNN) baselines to language-aware models such as ABINet and PARSeq that improve robustness under blur, compression, and partial occlusion typical of broadcast signals [50], [51], [52]. In the video setting, prior work reports consistent improvements from temporal aggregation (e.g., majority voting across frames), text tracking/stabilization, and region-of-interest constraints over typical on-screen areas for overlays [45], [46].

In advertising detection, OCR can be employed to determine whether specific text corresponds to a start or end bumper. A key advantage of this technique is its ability to identify text content independently of the bumper's visual design (layout, color palette, or animation), enabling discrimination among near-duplicate visuals used for different purposes. Practical implementations frequently combine lexicon or rule-based matching with edit-distance tolerance to handle noisy reads and multilingual variants [45]. This flexibility allows OCR to effectively differentiate start and end bumpers from other types of bumpers used for different purposes.

C. Hybrid and Advanced Techniques

Hybrid methods combine signal-based and content-based techniques to achieve higher accuracy and robustness. For instance, systems integrating black frame detection with video hashing or OCR can cross-verify results, reducing false positives and negatives. Recent advances in deep learning, such as segment-level autoencoders, have also enabled better multimodal representation of advertisements, improving detection performance in varied contexts [53].

As described in this section signal-based methods such as black frames or audio drops offer low complexity but often

fail when broadcasters do not use these cues consistently. Logo-based techniques can be accurate, yet they rely on stable on-screen graphics, which may persist during ads or be absent altogether. Repetition and hashing approaches effectively detect recurring ad content but usually require reference databases and do not provide exact break boundaries. Deep audio-visual CNNs achieve greater robustness but demand large training datasets and high computational resources, which limits near real-time, multi-channel deployment. In contrast, our system exploits the mandated start and end bumpers: dHash enables fast candidate detection, OCR disambiguates visually similar bumpers, and SBD supports the automatic discovery of new bumper variants, as described in the next sections.

V. SYSTEM DESIGN FOR AUTOMATIC ADVERTISING DETECTION IN FREE-TO-AIR TV

Considering the different existing techniques for automatic detection of TV commercial segments and the particularities of the existing regulations, some possible solutions for designing the system were analyzed. After a thorough inspection of many hours of different TV channels from different sites, it was determined that in many signals there are no black frames when advertising segments are added. Even when in some cases, the audio volume is louder in some advertisement segments, the regulation of some countries states that the audio volume should not be increased when an advertising space is started. So, audio-based techniques are disregarded for the design of the system. With these considerations, a system based on signal characteristics will not work well.

Although it is not mandatory, inspecting many hours of different TV channels from different sites, it was verified that several channels include a logo to identify the station while programs are on air, and these logos are usually removed during commercial segments. In general, these logos are located in the upper or bottom right corner, as shown in Figure 2. This is a general practice. However, using logo detection as an indicator of the start and end of the advertising spaces is not optimal since the use is not consistent. Adding the logo on the transmitted signal is often done manually by the broadcasters, and we have verified that stations do not perform this action systematically. As the use of logos is not mandatory by regulators, sometimes the logo will continue to appear during commercial breaks while sometimes it may not be present during program times.

Considering the variability in logo usage and the inconsistent presence of audio markers, an alternative approach was explored. According to the regulations of different countries, each advertising space must begin and end with an identifying bumper. These distinctive bumpers emerge as potential markers for determining the commercial break durations. This approach allows the use of image retrieval techniques, specifically using image hashing [54]. Image hashing techniques are algorithms used for image comparison, that are based on features from an image's content to create a fingerprint representation, as described in section IV.

Particularly, a difference hashing (dhash) algorithm, known to be fast and accurate [55], was selected to compare video frames to reference fingerprints of known bumpers. This approach allows the use of image retrieval techniques, specifically using image hashing [54]. Image hashing techniques are algorithms used for image comparison, that are based on features from an image's content to create a fingerprint representation, as described in section IV. Particularly, a difference hashing (dhash) algorithm, known to be fast and accurate [55], was selected to compare video frames to reference fingerprints of known bumpers. Figure 3 shows the steps of the dhash algorithm on a sample image: a) the original image is selected (i.e., a bumper frame); b) the image size is converted to grayscale and reduced to 9×8 pixels; c) gradients are computed to account for differences between adjacent pixels. This converts the 9×8 image into an 8×8-pixel representation and d) bits are assigned to each resulting pixel, checking whether the left pixel is brighter than the right pixel, to generate a 64-bit (8-hex characters) representation.



Fig. 2. Example of channel logo, at the upper-right corner of the screen.

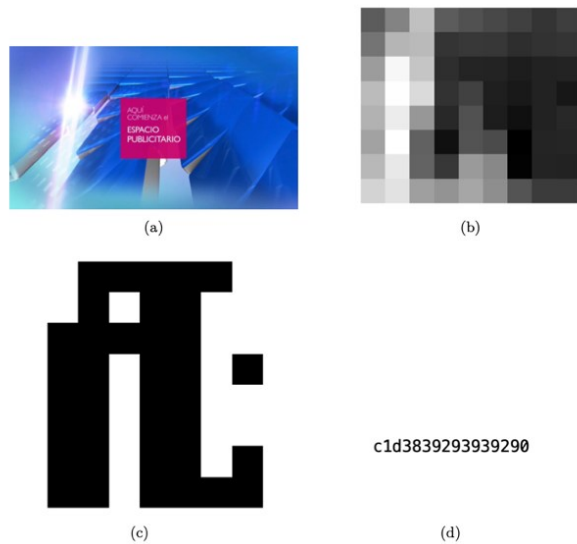


Fig. 3. Difference hashing (dhash) algorithm steps: (a) Original Image 1280×720. (b) Grayscale 9×8 representation. (c) 8×8 representation, binarized difference between left and right pixels. (d) Hex image hash representation.

To measure the similarity between two different images, a

simple distance metric between the 64-bit dhash representation can be used (i.e., Hamming distance). Figure 4 shows examples of hashes, distances between similar and dissimilar images that contain bumper examples.

When start and end bumpers share near-identical designs, their down-sampled dHash fingerprints can differ by only a few bits (e.g., Hamming distance 1 in Fig. 4, top), which may cause start/end confusions. Additionally, similar bumpers may be used for other purposes, such as indicating the end of “child protection hours”.

To address this limitation, text content analysis becomes essential. In this approach, after a bumper is detected using the hashing algorithm, an OCR technique is applied to interpret the text within the image. This step helps accurately classify the bumper as a “start”, “end” or another type, ensuring more reliable identification. This OCR technique was implemented using EasyOCR library which uses deep learning algorithms such as CRAFT [48] for text detection and a convolutional recurrent neural network (CRNN) [56] for text recognition.



Fig. 4. Examples of dhash distance between visually similar and dissimilar images. Hamming distance is applied.

When extracting frames from video content, captured frames can be blurry, and OCR may not always be flawless. Some mistaken letters can be erroneously recognized. Besides, different channels use different texts to describe the start and end of the commercial break. Therefore, a spell-checking process and a straightforward dictionary were employed to account for OCR imperfections and the variations in texts across different channels. The dictionary was crafted based on the most common words that appeared on each of the start and end bumpers in the sample videos used for training and testing the system. Nevertheless, this dictionary could be built for any other scenarios if sufficient sample videos are provided.

The dhash + OCR algorithm relies on having prior knowledge of the bumpers used by the channels. It is then necessary to establish a repository containing recognized bumper instances and subsequently employ a fast image retrieval method to locate them. Nevertheless, these bumpers can change over time, and the image repository must be updated regularly. In order to avoid this inconvenience, an automatic bumper discovery algorithm was also designed and implemented. This algorithm is based on the integration of shot boundary detection (SBD) techniques, combined with

OCR. SBD is a technique used to segment videos into scenes by detecting transitions such as cuts or fades. This process allows the system to isolate short video segments that likely contain bumpers, as these typically last between 1 and 5 seconds. Once these potential scenes are identified, OCR is applied on multiple frames within these scenes to detect common words found in bumper texts. By doing so, the system can automatically detect and catalog new bumpers, which are saved on the system's bumper database. The algorithm used for SBD is a content based algorithm [57] which is available in `pySceneDetect` SBD library. The content-based algorithm was used to select scenes with duration between 1.3 and 5 seconds, and the OCR technique was applied with cadence of 10 fps on the selected scenes.

VI. SYSTEM ARCHITECTURE

The general system architecture, as shown in Figure 5, has been designed to address the task of automatic TV commercial break detection. It captures and records TV signals, processes them, identifies commercial breaks, and reports the findings to a main program.

The process begins when TV broadcasts are syntonized and the video signals are recorded and segmented in one-hour chunks for subsequent processing. Every hour on the hour, a local service called "Media monitor" selects the corresponding chunk of the video signals recorded and sends them to the "Bumper detection" process. At the core of this process is located a database that stores the dhash representation of potential bumpers for each TV channel. The system selects one out of every ten frames and compares its dhash representation with the dhashes stored in the database. If a match is found, an OCR process is run. This process extracts the text from the frame, spell-checks it and compares it with a stored dictionary, in which the keywords associated with different bumper types are stored. If the text of the frame matches with the keywords of a "start" or "end" bumper, the timestamp, channel name and a start or end flag is informed to the main program (not shown in Figure 5) for further processing.

If no bumpers are detected for an entire day, it initiates a process to search for new bumpers, assuming the channel has changed its previous bumper. In this case, new bumpers must be identified and added to the dhash database. This process is shown in Figure 6. During it, the video is analyzed using SBD techniques to segment it into scenes, keeping only those lasting between 1.3 and 5 seconds, which typically corresponds to bumper durations. OCR is then applied to these scenes. If any text matches pre-defined terms in the dictionary, the new bumper is identified. The system calculates the dhash for the new bumper and stores it in the bumper dhash database. Finally, the program reanalyzes the recorded video to detect the newly identified bumpers.

VII. RESULTS

The proposed solution was implemented and tested on a dataset of 114 hours of video from live recordings of four free-

to-air TV stations from Uruguayan channels², recorded in the period between July 9 and October 26, 2023.

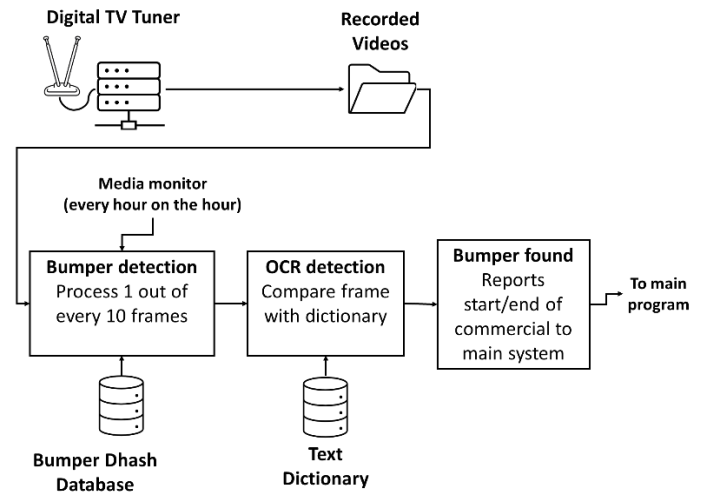


Fig. 5. Automatic commercial breaks detection system.

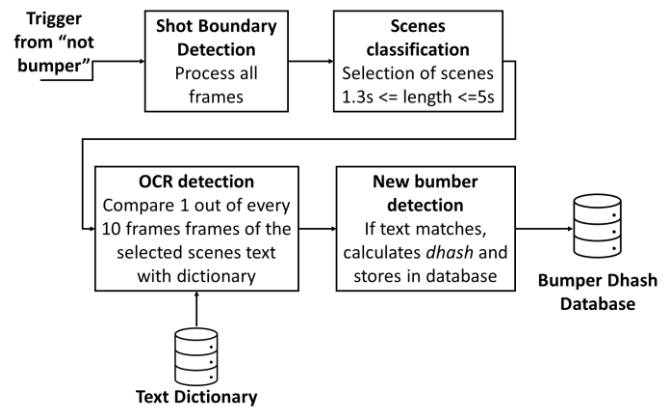


Fig. 6. New bumper identification process.

The image-retrieval stage (dHash) successfully matched frames to reference bumpers; see Sections IV–VI for methodological details. Table I summarizes detection performance.

The 114 hours of video were manually annotated and contained 264 commercial bumpers. Since these videos were recorded from live transitions these were manually inspected and none of them contained decoding artifacts or visual errors. All of them were properly detected in this test, corresponding to 100% precision and 100% recall. For the definition of precision and recall a 4-second window was used. Therefore, if a bumper detection is found within a 4-second window of an annotation of the same class a True Positive (TP) is accounted. The same window criteria was used for False Negatives (FN)

² The TV recordings come from the Uruguayan free-to-air TV signals: Channel 4 "Montecarlo", Channel 10 "Saeta TV", Channel 12 "La Tele" and "TV Ciudad", of which 18 hours are from Channel 4, 23 from Channel 10, 46 from TV Ciudad and 24 correspond to Channel 12.

and False Positives (FP). Detailed results are shown in Table I.

Regarding the automatic discovery of new bumpers, several experiments were performed, and the proposed algorithm was evaluated in a subset of the original 114 video hours corresponding to 49 hours from 3 different signals. The results are presented in Table II. In all tests, perfect precision was achieved. Given that bumper discovery will be only executed on times where no commercial breaks are found, and that the newly discovered bumpers will then be used in the normal operation of the algorithm, it is very important that the approach does not create false positive bumpers that then will remain in the system generating consistent errors.

TABLE I
COMMERCIAL BREAK DETECTION EVALUATION

Signal	Start Bumper			End Bumper			Overall		
	TP	FP	FN	TP	FP	FN	TP	FP	FN
Ch. 4	23	0	0	22	0	0	45	0	0
Ch. 10	21	0	0	21	0	0	42	0	0
Ch. 12	20	0	0	20	0	0	40	0	0
TV Ciudad	68	0	0	69	0	0	137	0	0

TABLE II
AUTOMATIC DETECTION OF UNSEEN BUMPER

Signal	Precision	Recall	# Videos
Ch. 4	1.0	1.0	17
Ch. 10	1.0	0.95	19
Ch. 12	1.0	1.0	13

The server used to run the developed system has a CPU with an Intel CORE i7-13700 chipset (with 16 cores and 24 threads), 32GB of RAM and a Nvidia RTX 3090 GPU. The average processing time for a 1-hour video was 353 seconds, with a standard deviation of 127 seconds, primarily associated with the variable duration of OCR and spell-checking operations performed during video processing. As an additional performance evaluation, the system successfully executed the processing of up to three TV stations simultaneously, showing its capability to handle multiple video signals in parallel.

Although perfect results were obtained while analyzing pre-recorded videos, the system was also tested in live mode showing some sporadic errors. Live mode implies that the full system pipeline is executed: free-to-air TV signals are tuned, videos are recorded and segmented in one-hour chunks, commercial break detection is executed, and results are reported to a main system. During these “live” tests the system failed to detect some bumpers. The exact frequency of these errors is hard to evaluate given that the errors were sporadic and there are no ground truth annotations for live transmissions, thus we computed a fair estimate of the error using the available data. As a reference the system was tested over 15 days on the same channel 4 as on offline tests but with

new different bumpers from the ones used in the offline tests. The bumper discovery algorithm was executed and successfully found the correct start and end bumper. During this test, 360 hours of video were recorded, and 258 commercial breaks were found. The evaluated signal usually shows programming during 18 of the 24 hours of the day, totaling $18 \times 15 = 270$ programming hours in the tested period. Even if the sample size from our offline tests is small, we used this to estimate the frequency of commercial breaks per hour. Typically, these signals show on average 1.25 commercial breaks per programming hour, so 337 commercial breaks detection were expected. These results suggest that at least more than 76% of the commercial breaks were effectively detected (258/337).

After the examination of a few test samples, in some cases, bumpers failed to be detected, and no commercial breaks were reported. In such cases, alarms were triggered to report irregularities on commercial break times. It was detected that these errors mainly occurred when reception signals were low and recorded videos were corrupted. The system discards videos with more than a certain number of corrupt frames. Another explanation for these errors is that the new evaluated bumpers included fast animations and text was displayed faster than in the offline tests. Even if the dhash sampling frequency was set to 1 every 8 frames instead of 1 every 10 frames as used during offline tests this still might not have been fast enough to handle the diversity of bumpers that can appear live broadcasting scenarios. Increasing the processing frame rate improves the efficacy of the algorithm in complex scenarios, however, it slightly increases the computing power needed to process a video.

VIII. FUTURE WORK

To enhance the current approach and address its limitations, some improvements can be studied and implemented.

Logo detection, while not foolproof, could complement bumper-based detection. By cross-referencing results from both methods, a refined advertising space identification can be achieved and enhance overall precision, at least for the channels that eliminate the logo during commercial break times.

The designed algorithm can consistently detect commercial breaks relying on mandatory visual cues in the form of visual bumpers. Although the approach is consistent, it is only applicable in countries where such cues are required by regulations. A search for a more generalist approach to advertisement detection needs to be conducted in the case of countries where such regulations are not present. With the advancement of generative AI and in particular Large Language Models (LLMs) it seems that a semantic analysis of TV content by transcribing audio could be performed to distinguish advertisements from regular programming.

Another promising direction for future research is the detection of non-traditional advertising content, such as product placements or promotional material embedded within TV programs. Unlike standard commercial breaks, this type of

advertising is often subtle and lacks explicit visual or textual boundaries, making it more challenging to identify. Non-traditional advertisements are harder to detect since they are not delimited by explicit cues and are usually blended within the real TV programs. Machine learning and new deep learning models could be used to detect this type of advertising. These techniques excel at finding complex patterns, enabling us to monitor a broader range of advertising strategies, and ensuring regulatory compliance. Several approaches seem appropriate to perform this task. One is to exploit repetition in the case of non-traditional advertisements that are prerecorded and inserted in programs. Although the idea of exploiting repetition to detect advertisements is not novel and has been already proposed by [58] it can be updated with more novel deep learning algorithms. In the case of content that cannot be detected by repetition, anomaly detection in video sequences could be applied to detect advertisements that are not in concordance with the video content usually shown in specific TV programs. To implement these techniques, a conceptual strategy similar to the one illustrated in Figure 6 can be employed. SBD can be used to segment scenes, followed by scene classification. Video segment hashing can be used, with an unsupervised learning technique, to detect repeated content typically associated with advertising. Additionally, OCR can complement these methods by extracting text from scenes, which is often present in advertisements. Using an expanded dictionary or artificial intelligence, the system can infer if the intent behind the detected text is for advertisement purposes.

IX. CONCLUSIONS

In this work, we have provided examples of commercial regulations in different countries and a detailed description of techniques for the automatic detection of commercials. The design, implementation, and evaluation of a system for detecting TV commercial breaks automatically are described. The system addresses the challenge of near-real-time automatic commercial breaks detection, adaptable to various countries and regulatory frameworks, particularly where text-based bumper messages are required by regulations on TV channels.

While conventional techniques relying on signal characteristics are more flexible and could be applied in broader scenarios, they face limitations due to variability in logo usage and audio features. We proposed a combined approach utilizing image retrieval and OCR that effectively leverages commercial bumpers that mark the start and end of commercial segments, ensuring accurate determination of their duration. Detection of new bumpers was also implemented, using SBD techniques, along with OCR.

While perceptual binary hashing, shot-boundary detection and OCR are individually well-known techniques, our work's contribution is an engineering synthesis and deployment-oriented pipeline that (i) uses fast frame-level dHash sampling for broad coverage, (ii) employs automatic bumper discovery via SBD to adapt to changing on-air graphics, and (iii) applies

OCR as a lightweight but effective disambiguation stage to distinguish visually similar bumpers — validated on real free-to-air recordings with per-channel performance metrics and live-mode tests. This integration (and the associated empirical evaluation in real broadcast conditions) is not, to our knowledge, presented in the recent literature.

Test results demonstrated the feasibility of this approach. By integrating image retrieval, shot boundary detection and OCR techniques, we have established a promising path for automatic advertising detection, especially valuable in regions where textual bumpers are a regulatory requirement.

ACKNOWLEDGEMENTS

This work was supported by ANII/URSEC under Grant DPU_S_2022_1_172807: “Time control of advertising guidelines and national music by analyzing audio signals with deep learning techniques”.

REFERENCES

- [1] “Directive - 2018/1808 - EN - EUR-Lex.” Accessed: Jan. 12, 2025. [Online]. Available: <https://eur-lex.europa.eu/eli/dir/2018/1808/oj/eng>
- [2] N. Rondan, M. Juayek, J. Joskowicz, and R. Sotelo, “Automatic Detection of TV Commercial Breaks,” *2024 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2025, To be published.*
- [3] “Ley N° 20383.” Accessed: Jan. 09, 2025. [Online]. Available: <https://www.impo.com.uy/bases/leyes/20383-2024>
- [4] “BOE-A-2010-5292 Ley 7/2010, de 31 de marzo, General de la Comunicación Audiovisual.” Accessed: Jan. 09, 2025. [Online]. Available: <https://www.boe.es/buscar/act.php?id=BOE-A-2010-5292>
- [5] “Children’s Educational Television | Federal Communications Commission.” Accessed: Jan. 09, 2025. [Online]. Available: <https://www.fcc.gov/consumers/guides/childrens-educational-television>
- [6] C. R. and T. C. (CRTC) Government of Canada, “TV and Radio Advertising Basics.” Accessed: Jan. 09, 2025. [Online]. Available: <https://crtc.gc.ca/eng/television/publicit/publicit.htm>
- [7] “Acts, Policies & Guidelines | Ministry of Information and Broadcasting.” Accessed: Jan. 09, 2025. [Online]. Available: <https://mib.gov.in/documents/notification/acts-policy-guidelines>
- [8] A. C. and M. Authority, “Advertising rules for broadcasters | ACMA.” Accessed: Jan. 09, 2025. [Online]. Available: <https://www.acma.gov.au/advertising-rules-broadcasters>
- [9] “Decreto 678/2016 — Servicios de Comunicación Audiovisual,” Argentina.gob.ar. Accessed: Aug. 16, 2025. [Online]. Available: <https://www.argentina.gob.ar/>
- [10] “Código Brasileiro de Autorregulamentação Publicitária,.” Accessed: Aug. 16, 2025. [Online]. Available: <https://www.gov.br/secom/pt-br/acao-a-informacao/legislacao/ca2digobrasdeautoregulanovo.pdf>
- [11] “A/85, Techniques for Establishing and Maintaining Audio Loudness for Digital Television,” ATSC : NextGen TV. Accessed: Jan. 09, 2025. [Online]. Available: <https://www.atsc.org/atsc-documents/a85-techniques-for-establishing-and-maintaining-audio-loudness-for-digital-television/>

- [12] E. Abdelfattah and S. Joshi, "Comparison of Machine Learning Classification and Clustering Algorithms for TV Commercials Detection," *IEEE Access*, vol. 11, pp. 116741–116751, 2023, doi: 10.1109/ACCESS.2023.3325888.
- [13] W. Tang, Y. Wo, and G. Han, "Geometrically robust video hashing based on ST-PCT for video copy detection," *Multimedia Tools Appl.*, vol. 78, no. 15, pp. 21999–22022, Aug. 2019, doi: 10.1007/s11042-019-7513-8.
- [14] A. Gomes, M. P. Queluz, and F. Pereira, "Automatic detection of TV commercial blocks: A new approach based on digital on-screen graphics classification," in *2017 11th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Dec. 2017, pp. 1–6. doi: 10.1109/ICSPCS.2017.8270456.
- [15] P.-C. Su and C.-S. Wu, "Efficient copy detection for compressed digital videos by spatial and temporal feature extraction," *Multimed Tools Appl.*, vol. 76, no. 1, pp. 1331–1353, Jan. 2017, doi: 10.1007/s11042-015-3132-1.
- [16] D. A. Sadlier, S. Marlow, N. O'Connor, and N. Murphy, "Automatic TV advertisement detection from MPEG bitstream," *Pattern Recognition*, vol. 35, no. 12, pp. 2719–2726, Dec. 2002, doi: 10.1016/S0031-3203(01)00251-5.
- [17] T. Arsan, E. E. Bulut, B. Eren, A. Uzgor, and S. Yolcu, "A novel IPTV framework for automatic TV commercials detection, labeling, recognition and replacement," *Multimed Tools Appl.*, vol. 82, no. 6, pp. 8561–8579, Mar. 2023, doi: 10.1007/s11042-021-11563-y.
- [18] A. Ramires, D. Cocharro, and M. Davies, "An audio-only method for advertisement detection in broadcast television content," *ArXiv*, Nov. 2018, Accessed: Aug. 16, 2025. [Online]. Available: <https://www.semanticscholar.org/paper/An-audio-only-method-for-advertisement-detection-in-Ramires-Cocharro/8a59c36396f35a06b0c18850788165d3b55b8337>
- [19] N. Özay and B. Sankur, "Automatic TV logo detection and classification in broadcast videos," in *2009 17th European Signal Processing Conference*, Aug. 2009, pp. 839–843. Accessed: Aug. 16, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/7077387>
- [20] P. Carvalho, A. Pereira, and P. Viana, "Automatic TV Logo Identification for Advertisement Detection without Prior Data," *Applied Sciences*, vol. 11, no. 16, p. 7494, Jan. 2021, doi: 10.3390/app11167494.
- [21] A. Tüzkö, C. Herrmann, D. Manger, and J. Beyerer, "Open Set Logo Detection and Retrieval," Oct. 30, 2017, *arXiv: arXiv:1710.10891*. doi: 10.48550/arXiv.1710.10891.
- [22] "Deep learning for logo recognition - ScienceDirect." Accessed: Jan. 09, 2025. [Online]. Available: <https://www.sciencedirect-com.proxy.timbo.org.uy/science/article/pii/S0925231217305660>
- [23] N. Spolaôr, H. D. Lee, W. S. R. Takaki, L. A. Ensina, C. S. R. Coy, and F. C. Wu, "A systematic review on content-based video retrieval," *Engineering Applications of Artificial Intelligence*, vol. 90, p. 103557, Apr. 2020, doi: 10.1016/j.engappai.2020.103557.
- [24] S. Minaee, I. Bouazizi, P. Kolan, and H. Najafzadeh, "Ad-Net: Audio-Visual Convolutional Neural Network for Advertisement Detection In Videos," June 22, 2018, *arXiv: arXiv:1806.08612*. doi: 10.48550/arXiv.1806.08612.
- [25] Y. Li, D. Wang, and L. Tang, "Robust and Secure Image Fingerprinting Learned by Neural Network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 362–375, Feb. 2020, doi: 10.1109/TCSVT.2019.2890966.
- [26] Z. Liu, "A Deep Neural Framework to Detect Individual Advertisement (Ad) from Videos," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2023, pp. 3567–3576. doi: 10.1109/WACV56688.2023.00357.
- [27] V. Sharma, M. Gupta, A. Kumar, and D. Mishra, "Video Processing Using Deep Learning Techniques: A Systematic Literature Review," *IEEE Access*, vol. 9, pp. 139489–139507, 2021, doi: 10.1109/ACCESS.2021.3118541.
- [28] L. Jiao *et al.*, "New Generation Deep Learning for Video Object Detection: A Survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3195–3215, Aug. 2022, doi: 10.1109/TNNLS.2021.3053249.
- [29] M. Li, Y. Guo, and Y. Chen, "CNN-based Commercial Detection in TV Broadcasting," in *Proceedings of the 2017 VI International Conference on Network, Communication and Computing*, in ICNCC '17. New York, NY, USA: Association for Computing Machinery, Dec. 2017, pp. 48–53. doi: 10.1145/3171592.3171619.
- [30] M. Z. P. Wardana and M. E. Wibowo, "Audio-Visual CNN using Transfer Learning for TV Commercial Break Detection," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 3, pp. 291–300, July 2023, doi: 10.22146/ijccs.76058.
- [31] Waseemullah, M. F. Hyder, M. A. Siddiqui, and M. Mukarram, "TV Ad Detection Using the Base64 Encoding Technique," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7605–7609, Oct. 2021, doi: 10.48084/etasr.4337.
- [32] F. Vega, J. Medina, D. Mendoza, V. Saquicela, and M. Espinoza, "A robust video identification framework using perceptual image hashing," in *2017 XLIII Latin American Computer Conference (CLEI)*, Sept. 2017, pp. 1–10. doi: 10.1109/CLEI.2017.8226396.
- [33] S. F. C. Haviana and D. Kurniadi, "Average Hashing for Perceptual Image Similarity in Mobile Phone Application," *Journal of Telematics and Informatics (JTI)*, vol. 4, no. 1, 2016.
- [34] "Kind of Like That - The Hacker Factor Blog." Accessed: Aug. 16, 2025. [Online]. Available: <https://www.hackerfactor.com/blog/?/archives/529-Kind-of-Like-That.html>
- [35] C. Zauner, "Implementation and Benchmarking of Perceptual Image Hash Functions," 2010. Accessed: Aug. 16, 2025. [Online]. Available: <https://www.semanticscholar.org/paper/Implementation-and-Benchmarking-of-Perceptual-Image-Zauner/635e1b5261ad1545aab7acde48efa267ae428fc3>
- [36] D. Petrov, "Wavelet image hash in Python," Medium. Accessed: Aug. 16, 2025. [Online]. Available: <https://fullstackml.com/wavelet-image-hash-in-python-3504fdd282b5>
- [37] T. J. Rodrigues Pereira, "Detecção de Publicidade em Conteúdos de Difusão." FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO, 2018.
- [38] T. Sato, T. Kanade, E. K. Hughes, M. A. Smith, and S. Satoh, "Video OCR: indexing digital news libraries by recognition of superimposed captions," *Multimedia Systems*, vol. 7, no. 5, pp. 385–395, Sept. 1999, doi: 10.1007/s005300050140.
- [39] S. S. and P. M. Kamade, "Video OCR for Video Indexing," *IJET*, vol. 3, no. 3, pp. 287–289, 2011, doi: 10.7763/IJET.2011.V3.239.
- [40] J. Chaloupka, K. Paleček, P. Červa, and J. Žďánský, "Optical Character Recognition for Audio-Visual Broadcast Transcription System," in *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*,

- Sept. 2020, pp. 000229–000232. doi: 10.1109/CogInfoCom50765.2020.9237867.
- [41] B. Rozsa and M. Mocofan, “TV News Database Indexing System with Video Structure Analysis, Representative Images Extractions and OCR for News Titles,” in *2022 International Symposium on Electronics and Telecommunications (ISETC)*, Nov. 2022, pp. 1–4. doi: 10.1109/ISETC56213.2022.10010319.
- [42] F. Borisyuk, A. Gordo, and V. Sivakumar, “Rosetta: Large Scale System for Text Detection and Recognition in Images,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, in KDD '18. New York, NY, USA: Association for Computing Machinery, July 2018, pp. 71–79. doi: 10.1145/3219819.3219861.
- [43] bandersmsft, “Get optical character recognition (OCR) insights.” Accessed: Aug. 17, 2025. [Online]. Available: <https://learn.microsoft.com/en-us/azure/azure-video-indexer/ocr-insight>
- [44] “Recognize text | Cloud Video Intelligence API Documentation | Google Cloud.” Accessed: Aug. 17, 2025. [Online]. Available: <https://cloud.google.com/video-intelligence/docs/text-detection>
- [45] X.-C. Yin, Z.-Y. Zuo, S. Tian, and C.-L. Liu, “Text Detection, Tracking and Recognition in Video: A Comprehensive Survey,” *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2752–2773, June 2016, doi: 10.1109/TIP.2016.2554321.
- [46] D. Karatzas *et al.*, “ICDAR 2015 competition on Robust Reading,” in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, Aug. 2015, pp. 1156–1160. doi: 10.1109/ICDAR.2015.7333942.
- [47] X. Zhou *et al.*, “EAST: An Efficient and Accurate Scene Text Detector,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 2642–2651. doi: 10.1109/CVPR.2017.283.
- [48] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, “Character Region Awareness for Text Detection,” Apr. 03, 2019, *arXiv:1904.01941*. doi: 10.48550/arXiv.1904.01941.
- [49] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, “Real-time Scene Text Detection with Differentiable Binarization,” Dec. 03, 2019, *arXiv:1911.08947*. doi: 10.48550/arXiv.1911.08947.
- [50] B. Shi, X. Bai, and C. Yao, “An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017, doi: 10.1109/TPAMI.2016.2646371.
- [51] S. Fang, H. Xie, Y. Wang, Z. Mao, and Y. Zhang, “Read Like Humans: Autonomous, Bidirectional and Iterative Language Modeling for Scene Text Recognition,” Mar. 11, 2021, *arXiv:2103.06495*. doi: 10.48550/arXiv.2103.06495.
- [52] D. Bautista and R. Atienza, “Scene Text Recognition with Permuted Autoregressive Sequence Models,” in *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVIII*, Berlin, Heidelberg: Springer-Verlag, Oct. 2022, pp. 178–196. doi: 10.1007/978-3-031-19815-1_11.
- [53] K. Somandepalli, V. Martinez, N. Kumar, and S. Narayanan, “Multimodal Representation of Advertisements Using Segment-level Autoencoders,” in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, in ICMI '18. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 418–422. doi: 10.1145/3242969.3243026.
- [54] F. Vega, J. Medina, D. Mendoza, V. Saquicela, and M. Espinoza, “A robust video identification framework using perceptual image hashing,” in *2017 XLIII Latin American Computer Conference (CLEI)*, Sept. 2017, pp. 1–10. doi: 10.1109/CLEI.2017.8226396.
- [55] “Kind of Like That - The Hacker Factor Blog.” Accessed: Nov. 19, 2024. [Online]. Available: <https://www.hackerfactor.com/blog/?archives/529-Kind-of-Like-That.html>
- [56] B. Shi, X. Bai, and C. Yao, “An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition,” July 21, 2015, *arXiv:1507.05717*. doi: 10.48550/arXiv.1507.05717.
- [57] I. S. Gruzman and A. S. Kostenkova, “Algorithm of scene change detection in a video sequence based on the threedimensional histogram of color images,” in *2014 12th International Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, Oct. 2014, pp. 1–1. doi: 10.1109/APEIE.2014.7040826.
- [58] M. Covell, S. Baluja, and M. Fink, “Advertisement Detection and Replacement using Acoustic and Visual Repetition,” in *2006 IEEE Workshop on Multimedia Signal Processing*, Oct. 2006, pp. 461–466. doi: 10.1109/MMSP.2006.285351.



Nicolas Rondan received his Master's degree in Artificial Intelligence from the University of Edinburgh awarded in 2018 and a Bachelor's degree in Telematics Engineering from Universidad de Montevideo, 2014. Currently, he serves as a Senior R&D Team Leader at Digital Sense, specializing in computer vision and machine learning, with a strong focus on deep learning and optimization algorithms. He is also a lecturer in computer vision at Universidad de Montevideo, has contributed as a researcher to several academic projects and is enrolled in a Ph.D program in Electrical Engineering at Universidad Católica del Uruguay.



Marcos Juayek received his Master's degree in Applied Research in Engineering from Universidad de Montevideo, Uruguay, in 2020, and a degree in Telematics Engineering from the same university in 2012. He is currently a Solution Architect and Business Development Lead at Quantum-South and a Director at Viento Fortuna.

Throughout his career, he has held various leadership roles in software development, logistics, and business development, combining technical expertise with strategic vision to deliver innovative solutions. His research and professional work include optimization algorithms, video quality assessment, and applications in digital television and artificial vision. He has co-authored several scientific publications and participated in projects related to digital television for the Uruguayan government. He is also a Scrum Master (PSMI), certified by Scrum.org.



Jose Juskowicz (M'09, SM'10) received his Ph.D. degree in Telematics Engineering from Universidad de Vigo, Spain, in 2012. He is an Electronics Engineer specialized in Telecommunications from Universidad de la Republica, Uruguay since 1995. He is an Associate Professor with the Faculty of Engineering, Universidad de Montevideo

and Universidad de la Republica, Montevideo, Uruguay, a member of the National Researchers System, and a Consulting Engineer at Isabel. Throughout his professional career, he has conducted various consultancy projects and has led the design and implementation of various national and international projects in the field of information and communication technologies. His research interests include multimedia applications and quality of experience. Dr. Juskowicz is internationally certified as a Project Management Professional (PMP) by the Project Management Institute.



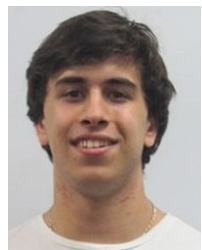
Maximiliano Aguerre received his Telematics Engineer degree in 2024 from the Universidad de Montevideo. He is currently a Technical Engineer at Foxsys.



Gastón Gonzalez received his Telematics Engineer degree in 2024 from Universidad de Montevideo. He is currently a Full Stack AI Developer at Effectus Software, where he specializes in building and integrating artificial intelligence solutions into full-stack applications



Rafael Sotelo (M'05, SM'13) received his Ph.D. in Engineering from the University of Vigo, Spain, an MBA from Universidad de Montevideo, Uruguay, and an Electrical Engineer degree from Universidad de la Republica, Uruguay. He is the Dean of Engineering at Universidad de Montevideo. Cofounder and President at Quantum-South. He has been an Administrative Committee Member at the IEEE Broadcast Technology Society, and a Distinguished Lecturer and Board of Governors Member at the IEEE Consumer Technology Society.



Santiago Quincke received his Telematics Engineer degree in 2024 from Universidad de Montevideo. He has extensive experience in software development, specializing in front-end and back-end technologies. Currently, he works as a Software Developer at Quanam, contributing to innovative solutions in a hybrid work environment.



Andres Patrone received his Telematics Engineer degree in 2024 from Universidad de Montevideo, with a strong interest in information and communication technologies (ICT), programming, data networks, and telecommunications infrastructure. He is currently an Industrial Automation Engineer at Secoin. Previously, he worked

as a Software Developer at ANII and Qualabs.