Automatic Letter/Pillarbox Detection for Optimized Display of Digital TV

Lúcia Carreira and Maria Paula Queluz
Instituto de Telecomunicações, Lisboa, Portugal
Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

Keywords: Quality Control, Quality Assessment, Active Format Description, Picture Aspect-Ratio, Letterbox Detection, Pillarbox Detection, Subtitles Detection, Subtitles Extraction.

Abstract: In this paper we propose a method for the automatic detection of the true aspect ratio of digital video, by detecting the presence and width of horizontal and vertical black bars, also known as letterbox and pillarbox effects. If active format description (AFD) metadata is not present, the proposed method can be used to identify the right AFD and associate it to the video content. In the case AFD information is present, the method can be used to verify its correctness and to correct it in case of error. Additionally, the proposed method also allows to detect if relevant information (as broadcaster logos and hard subtitles) is merged within the black bars and, in the case of subtitles, is able to extract it from the bars and dislocate it to the active picture area (allowing the letterbox removal).

1 INTRODUCTION

In recent years, automatic quality monitoring and control of multimedia content has become an increasingly important topic, especially due to the transmission of digital video over the internet and mobile networks. Television (TV) is perhaps the most relevant field where numerous examples of digital video systems, as cable and satellite services, IPTV and terrestrial digital TV broadcast, and a wide variety of user displaying devices, can now be found. In order to reach such a diversity of platforms, multiple content transformations, as format conversion, picture aspect-ratio adaptation, and associated metadata update, may be required. At each transformation, the content interacts with diverse systems and technologies, and more content quality issues arise. Hence, there is a critical need for automatic quality control (QC) systems, assuring content quality and content readiness at all points of the video transmission chain.

Most of the research in automatic video quality assessment and control systems has been devoted to the picture quality aspect (Wu & Rao, 2006) (IEEE - SP, 2011); however, to provide the users with a improved quality of experience (QoE), other quality issues, as those related with standards conformance and correctness of metadata associated with content should be also considered (Kumar, 2010).

In television technology, picture aspect ratio may be transmitted in the MPEG video stream or in the baseband SDI video through the use of a standard (SMPTE, 2009) metadata known as *active format description* (AFD). The use of AFD enables both 4:3 and 16:9 television sets to optimally present pictures transmitted in either format, by providing information to video devices about where in the coded picture the active image area is (i.e., the area that needs to be shown). It has also been used by broadcasters to dynamically control how down-conversion equipment formats widescreen 16:9 pictures for 4:3 displays.

In this paper we propose a method for the automatic detection of the true picture aspect ratio, by detecting the presence and width of horizontal and vertical black bars, also known as letterbox and pillarbox effects. If AFD data is not present, the proposed method can be used to identify the right AFD and assign it with content. In the case AFD information is present, the method can be used to verify its correctness and to correct it in case of error. Additionally, the proposed method also allows to detect if any type of information (as broadcaster logos and hard subtitles) is merged within the black bars and, in the case of subtitles, is able to extract it

Carreira L. and Queluz M..
Automatic Letter/Pillarbox Detection for Optimized Display of Digital TV.
DOI: 10.5220/0005064202810288
In *Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2014), pages 281-288
ISBN: 978-989-758-046-8

from the bars and to dislocate it to the active image area (allowing the letterbox removal). Related work has been proposed in (Markandey, 2002) and (Schoner & Neuman, 2008). However, none of these works address subtitles detection partially overlapping the active image area, nor the (re-) placement of subtitles on that area. Also, we propose a lower complexity, yet effective, method for the detection of the black bar borders.

Following this introduction, section 2 presents the advantages of using AFD metadata on TV transmissions. Section 3 details the methods developed for the detection of the pillar/letterbox effects; a method to detect the presence of relevant information (logos, hard subtitles) inlayed on the picture and over the pillar/letterboxing, is also described. In section 4 we present a preliminary approach to the automatic extraction of the subtitles from the black bars and respective placement on the active image area. Results are presented in section 5. Finally, conclusions are drawn in section 6.

2 PICTURE ASPECT RATIO AND THE NEED FOR AFD

The aspect ratio of an image is the proportional relationship between its width and its height, and is commonly expressed as two numbers separated by a colon. The most common aspect ratios used today are 4:3 (or 1.33:1) and 16:9 (or 1.78:1) in television (TV) systems - the former in standard definition (SD) TV and the latter in high definition (HD) TV - and 1.85:1 or 2.39:1 in cinema films. Other formats are also possible. Picture mapping to a display of unequal aspect ratio is typically achieved by adding horizontal black bars (letterboxing) or vertical black bars (pillarboxing) to retain the original format's aspect ratio (Fig. 1). Other possible procedures are enlargement of the original image to fill the receiver format's display area and cutting off (hence loosing information) the exceeding parts of the picture; stretching (hence distorting) the image to fill the receiving format's ratio; transmission of anamorphic content, that is expanded to the whole screen area.

Active Format Description (AFD) is a method of transmitting the aspect ratio of a picture by inserting flags (a four-bit code) on the MPEG video stream or in the baseband SDI video; bar data, indicating the extent of top, bottom, left, and right black bars, may also be transmitted for some AFD codes, as the bar width may be asymmetric. AFD metadata was developed with the purpose of optimizing the

a) b)

Figure 1: a) Pillarbox effect, resulting from a 4:3 picture mapped to a 16:9 format; b) Letterbox effect, resulting from a 16:9 picture mapped to a 4:3 format. (Jones, 2006).

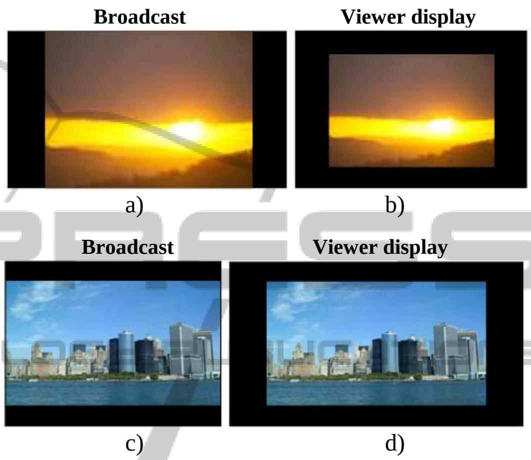

Figure 2: "Postage stamp" effect: a) 4:3 content is mapped to 16:9 by adding pillarboxes; b) display converts the content from 16:9 to 4:3 by adding letterboxes; c) 16:9 content is mapped to 4:3 by adding letterboxes; d) display converts the content from 4:3 to 16:9 by adding pillarboxes. (Jones, 2006).

display of the image for the TV viewer, providing guidance for the format conversion process in the receiver display, and also in post- and pre-broadcast.

Usually, the display of SD content on HD displays is done according to one of the following procedures:

1. full 4:3 content is mapped to 16:9 content by adding pillarboxes (Fig. 1-a);
2. 16:9 letter box content (Fig.1-b) is mapped to HD by removing the bars and displaying the content only.

Similar procedures exist for down conversion from HD content to a SD display - HD 16:9 is either cropped or letterboxed.

Procedure 2. above is possible only if the conversion/display process is AFD sensitive and AFD metadada is present indicating that SD content is 16:9 letterbox or the HD content is 4:3 pillarbox. If AFD information is missing or incorrect, the content becomes a postage stamp content by adding further pillar boxes to it (Fig. 2).

To handle AFD issues, an automatic method for detecting the presence of letterboxes and pillarboxes is required. The procedure should be generic in order to cope with any possible picture aspect ratio, and with any possible bar width setting. Furthermore, it should have a low complexity level, allowing its use on quality control systems placed at any point of the video transmission chain.

3 PILLAR/LETTERBOX DETECTION

3.1 Introduction

This section describes a software tool that analyzes a video sequence, frame by frame, with the aim of detecting the presence, and width, of black bars resulting from pillarbox and/or letterbox effects. Several situations, as pictures with very dark backgrounds or with logos and subtitles inlayed on the video and over the bars (Fig. 3), make this detection a non trivial task. Also, the black level used on the bars may vary from video to video. Although recommendations (ITU-R, 2011) (ITU-R, 2002) define the black level as having the value of 16 on the luminance (Y) component and the value 128 on the chrominance components (C_r and C_b), in many observed cases the used black level has a Y value close to 0. The black level can also present small variations inside the bars, which may result from added noise due to compression or analog-to-digital conversion (in case the video was originally on an analog format). Hence, a pixel is considered as black if conditions (1) and (2) are verified:

$$Y \leq \Delta_{Y1} \text{ XOR } |Y - 16| \leq \Delta_{Y2} \quad (1)$$

$$|C_r - 128|, |C_b - 128| \leq \Delta_C \quad (2)$$

where Δ_{Y1} ,Δ_{Y2} and Δ_C are, respectively, luminance and chrominance differentials that can be user defined (by default, Δ_{Y1}=10, Δ_{Y2}=5 and Δ_C=5). The XOR operator in eq. (1) means that only one of the conditions for the Y component can be verified in a given black bar.

3.2 Bars Detection Procedure

The developed procedure for pillar/letterbox bars detection considers two cases:

- Case 1 – Bars can be assumed as free of any content inlayed on it.
- Case 2 – Bars may have relevant content (logos, subtitles) inlayed on it.

In Case 1, once the bars have been detected, the aspect ratio of the active image area of the video is computed as the quotient between active image columns (frame width minus vertical bars width) and active image lines (frame height minus horizontal bars height) ; if it is within 1% variation of any of the aspect ratios allowed by the AFD codes - 4:3, 14:9, 16:9 and >16:9 - the AFD will be verified, or settled, accordingly. In Case 2, after detecting the bars, the algorithm proceeds by evaluating the presence and position of logos and/or subtitles (Section 3.3); if these are not present, the ADF codes will be settled as in Case 1; if just subtitles are present, we propose in Section 4 a preliminary approach to a method that allows to move the subtitles to the active part of the picture, which would also allow the AFD codes to be settled as in Case 1.

Figure 3: a) Frame with letterbox whose limits are barely perceived; b) Frame with subtitles on the bar.

It is worth to mention that TV programs typically use two types of captioning: soft and hard. Soft subtitles are sent as tiff or bmp graphics, or as a specially marked up text, that is overlaid on the video at the display. For this type of captioning, Case 1 also applies. In hard captioning, subtitles are sent merged in the frame, so Case 2 will apply.

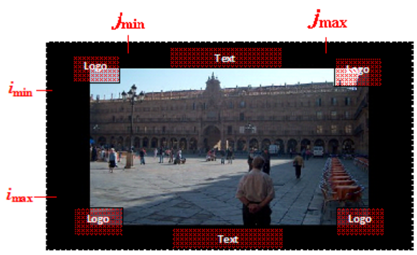

Figure 4: Typical positioning of text and logos on video frames.

In Case 1, and in order to detect the letterbox top bar (its presence and width), the algorithm starts by scanning each frame line, from the top to the bottom of the frame, applying conditions (1) and (2) to each pixel to verify if it corresponds to a black pixel; if line i is the first one for which those conditions are not verified, the top horizontal bar width is set to i-1. To detect a horizontal bar on the bottom of the frame, the procedure is repeated but carrying out the scanning from the bottom to the top of the frame. To detect the bars due to the pillarbox effect, a similar procedure is applied along the horizontal direction of the frame.

For Case 2, consider Fig. 4 where the typical positioning of subtitles and logos is represented. To detect the letterbox top bar, the algorithm starts by scanning each image line (from top to bottom, as in Case 1), but considering only the pixels situated between the limits $j_{min}=0.25\times Width$ and $j_{max}=0.75\times Width$, where *Width* is the horizontal resolution, in pixels, of the video. This strategy reduces the inclusion of pixels from logos. Conditions (1) and (2) are applied to each pixel along the scan line. Let F_b be the fraction of pixels, along the current picture line, that verifies those conditions. The line in question is considered has a potential black bar line if $F_b \geq F_T$, where F_T is a user defined threshold (by default, F_T = 0.8). With this criterion, lines of the image where a certain fraction of pixels is not black due to the existence of subtitles on the black margins (which will be confirmed by the procedure described in section 3.3), can still be considered as belonging to a black border.

When a set of N_c consecutive lines (by default N_c=20), does not check the condition $F_b \geq F_T$, it is considered that the limits of the bar have been overpassed; the width of the bar will be given by the i coordinate of the last line that has verified the condition $F_b \geq F_T$.

To detect a horizontal bar on the frame bottom, the procedure is repeated but carrying out the scanning from bottom to top. To detect the bars due to the pillarbox effect, a similar procedure is applied along the horizontal direction, but with the controlling parameters set to N_c = 1 and F_T = 0, since no text is expected over those bars; i_{min} and i_{max} are respectively set (by default) to $0.25\times$*Heigth* and $0.75\times$ *Height*, where *Height* is the number of lines per frame.

In both cases 1 and 2, and to minimize false detections, it is required that the resulting aspect ratio should be present in a minimum number, N_F, of consecutive frames, before accepting it as valid. By default, N_F = 125 (5 seconds of video for a frame rate of 25 Hz).

3.3 Logo and Subtitles Detection

This section describes the procedure for detecting logos and hard subtitles that may exist over the pillarbox and letterbox black bars. The distinction between logos and subtitles detection can be done by its spatial location, as the subtitles are typically centered on the bottom or on the top of the frame, occupying the space of one or two lines of text, and logos tend to be located in the corners of the frame, as depicted in Fig. 4. Accordingly, logos are searched for on the part of the bars area situated between the frame limits and 1/10 of the height (for vertical bars) and 1/10 of the width (for horizontal bars) of the frame; subtitles (their vertical limits) are searched for in the area of the letterbox bars comprised between j_{min} and j_{max}.

For subtitles detection each line within the search area is scanned on the horizontal, from bottom to top, searching for non-black rows of pixels. The vertical limits (signalized by the red lines in Fig. 5) of the subtitles are considered to be the position of the first and last non-black rows found. The procedure is repeated on the horizontal direction, scanning along the image columns inside the searching area, in order to find the lateral limits of the subtitles (signalized by the yellow lines in Fig. 5). Note that if a subtitle text line intercepts the active image area boundary, only three subtitle limits will be found (Fig. 6). Logo detection is carried out with a similar procedure but with the scan first performed along the image columns, within the logo searching area, and in order to determine the lateral limits of it (signalized by the green lines in Fig. 5). If the search zone contains only a part of the logo (case in which just one of the limits will be found), the search proceeds outside the initial search area, column by column, until the second limit is found. In order to find the vertical limit (signalized by the orange line in Fig. 5), the process is repeated on the perpendicular direction.

Figure 5: Logo (green and orange lines) and subtitle (red and yellow lines) limits.

For a logo and a subtitle to be considered as valid, it should be detected in a minimum number of consecutive frames. For subtitles the minimum recommended value (*m_text*) is 12 frames (about 0.5 seconds for a frame rate of 25 Hz); for logos, which usually remain in the image throughout the whole video shot, a minimum number (*m_logo*) of 125 frames (about 5 seconds for a frame rate of 25 Hz) should be used.

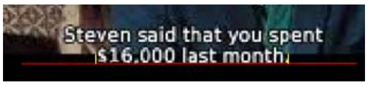

Figure 6: Example of a frame where the subtitle intercepts the active image area boundary; in this case, only three subtitle limits will be found.

4 SUBTITLES PROCESSING

In this section, a preliminary approach to a subtitles detection, extraction and placement algorithm is presented. This could be used by an advanced quality control system, able to detect hard subtitles on the letterbox and to copy them to the active image area, before setting the correct AFD code. The architecture of the developed algorithm is represented on Fig. 7. Parts of this algorithm were inspired by the work on (Tang, Gao, Lin, & Zhang, 2002).

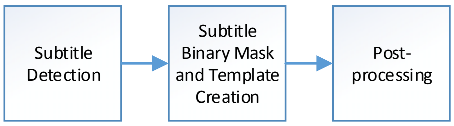

Figure 7: Architecture of the subtitle processing algorithm.

4.1 Subtitle Detection

From the subtitle area whose limits have been found in section 3.3, the subtitle pixels are extracted by applying a thresholding operation to the three color channels. The threshold for each color channel is defined by the highest color component value used in criteria (1) and (2), defined in section 3.1. The processing of the three channels in separate is favorable for the detection of subtitles of colors different from white (e.g., yellow subtitles). The three resulting binary images are combined (bitwise operation OR) into a binary mask, $b_{partial}$, of the pixels that belong to the subtitle (Fig. 8).

Figure 8: Binary mask of the subtitles in the letterbox of the frame in Fig. 3-b), after thresholding and combination of the three color channels.

For each color channel Y and C, with $C = C_r$ or C_b, of the subtitles, the mean value, med_Y, med_C and standard deviation, σ_Y, σ_C, is then computed from the video frame and using the binary mask to assure that only the pixels that belong to the characters are considered. At this point we have obtained one important feature characterizing the subtitles - their color values. This will be used in a first estimate of other text lines of the subtitle that may be over the active image area, by applying (3) to the corresponding color component of each pixel.

$$bm(i,j) = \begin{cases} 1, & \text{if } med_Y - \sigma_Y < Y(i,j) < med_Y + \sigma_Y \text{ and } \\ & med_C - 2\sigma_C < C(i,j) < med_C + 2\sigma_C \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The resulting binary image, *bm*, will have white pixels in the areas with color similar to the subtitles. This binary mask is projected horizontally (Fig. 9) and vertically, and the intersection of both projections results on the image regions, R_{text}, that may (potentially) contain subtitles; the region at the image bottom (corresponding to the horizontal projection signalized by B in Fig.9), results from the subtitle inside the black bar. The upper limit of this subtitle is then refined for the case it intercepts the black bar borders (as in Fig. 6). Once a complete line of subtitle characters was identified, five subtitle features are obtained from it: character high, h_{text}, mean and variance of character lines width, μ_{text} and σ_{text}, mean and variance of the space between character lines, μ_s and σ_s (all measures are in pixels).

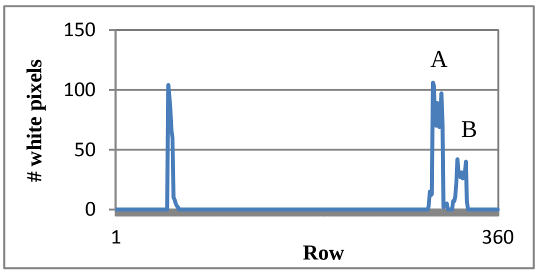

Figure 9: Horizontal projection of the binary mask *bm* corresponding to Fig. 3-b). A and B signalize the projections of the two subtitle lines.

For each line of each image area R_{text} of *bm*, the following parameters are computed: mean of white pixels sequence length, μ_W; mean of the black pixels sequence length, μ_B. The line is kept as a potential subtitle line if condition (4) is verified:

$$\begin{cases} \mu_{text} - 2\sigma_{text} \le \mu_W \le \mu_{text} + 2\sigma_{text} \\ \mu_s - 2\sigma_s \le \mu_B \le \mu_S + 2\sigma_S \end{cases} \quad (4)$$

Let l_{top} and l_{bottom} the top and bottom lines of a set of consecutive lines that have verified condition (4); R_{text} is kept as a potential subtitle region if condition (5) is verified:

$$0.75 \times h_{text} \le l_{top} - l_{bottom} \le 1.25 \times h_{text} \quad (5)$$

Once every region R_{text} has been analyzed, a final procedure is applied to those regions that have passed criterion (5): starting from the subtitle line on the black bar, a region R_{text} is considered as a subtitle region if it is centered with the previously found subtitle region, and if the distance between them is not higher than h_{text}.

4.2 Subtitle Binary Mask and Template Creation

Once the subtitle boundaries are known, a new binary image, *n_bm*, is computed, using condition (7); this condition is similar to (3) but less restrictive in the chrominance components limitation. Note that due to compression artifacts, subtitle colors may vary relatively to their original color.

$$n\_bm(i,j) = \begin{cases} 1, & \text{if } med_Y - \sigma_Y < Y(i,j) < med_Y + \sigma_Y \text{ and} \\ & med_C - 10\sigma_C < C(i,j) < med_C + 10\sigma_C \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The subtitle mask is cropped from the *n_bm*, as illustrated in example of Fig. 10-a). Further processing of this binary mask is needed, to include missing pixels on the characters and on the black contour around it. This may be achieved by expanding the mask through a dilation operator (Serra, 1983), with a disk type structuring element of size 3, as illustrated in Fig. 10-b). The template for the subtitle is created by copying, from the original frame and using the binary mask, the pixels that belong to the subtitle (Fig. 10-c)).

Figure 10: (a) Subtitle binary mask cropped from the mask of the frame of Fig. 3-b). (b) Final binary mask of the subtitle, after dilation.. (c) Template of the subtitle.

4.3 Post-Processing

For situations where part of a subtitle overlaps the active image area, its new placement may leave some holes to fill; this will happen if the line of text that comes from the letterbox is narrower that the line of text in the active image area, as in Fig. 3-b). In this case, it will be necessary to disguise the space left open, which constitutes a video inpainting problem (Abraham, Prabhavathy, & Shree, 2012). In this work, a simple bilinear spatial interpolation was implemented. More elaborated solutions should be investigated to cope with textured backgrounds and objects motion. After inpainting and removing the black bars of letterbox/pillarbox effect, the subtitle is placed within the image on the frame by copying the template to the bottom of the active image area (Fig. 11).

Figure 11: Final frame with black margin removal, inpainting and subtitle placement from the original frame in Fig. 3-b).

5 RESULTS

In this section, the results for the detection of black bars, subtitles and logos are presented. To test the proposed algorithms, the following set of video sequences, with diverse content, was chosen:

- S1) Letterbox (2).
- S2) Sequences (a), (b), (c), (d) and (e) with letterbox and subtitles in the bottom bar (2).
- S3) Letterbox, pillar box and subtitles in both the top and bottom bars (3).
- S4) Letterbox and semi-transparent logo in the bottom right corner (1).
- S5) Letterbox and subtitles in the bottom bar (4) (hard subtitles entered manually); includes "starry sky" frames (Schoner & Neuman, 2008).
- S6) Letterbox and opaque logo in the top left corner (5).

Sequence S5 was chosen because it contains some frames where it is particularly difficult to isolate the subtitle lines, due to their content. As for the sequence in (2), since it is a very long video, some smaller sequences with subtitles ((a) to (e)) were chosen. Table 1 presents the main characteristics of the test sequences. All sequences have a spatial resolution of 640×360 pixels, except sequence 6 which has 480×360 pixels. An image frequency of 25 Hz is assumed.

Table 1: Main characteristics of the test sequences.

Sequence	Starting frame	Duration (s)	Active Image Area
S1	4650	74.8	640 × 264
S2(a)	~ 12100	60.0	640 × 264
S2(b)	~ 4800	43.2	640 × 264
S2(c)	~ 7900	13.5	640 × 264
S2(d)	~ 9650	14.3	640 × 264
S2(e)	~ 22450	33.5	640 × 264
S3	400	5.2	599 × 246
S4	1	352.1	640 × 277
S5	2100	16.0	640 × 270
S6	200	10.0	480 × 270

In all the results the algorithms parameters were set to their default values (indicated along the text); these values were obtained empirically from tests. The hardware used was an x86 machine with an Intel® Core™ i7-4770, at 3.40 GHz, with 16 GB of RAM.

5.1 Black Bar Detection

To evaluate the performance of the simplest algorithm developed for Case 1, three video sequences with letterboxing were chosen: S1, S2(c) and S4. Table 2 presents the black bar boundaries for each sequence and the resulting detection success rate for a frame-by-frame analysis. In the case of sequence S1, where bars are free of logos and subtitles, the detection was correct in every frame. For sequence S2(c) the detection of the bottom bar is incorrect in all the frames where subtitles are inlayed over the bar. For sequence S4, the bottom bar is just correctly detect at the sequence beginning, where the logo is not present; for some frames, the other bars limits are also wrongly detect, due to a very dark background present in some parts of the sequence (Fig. 3-a depicts a frame extracted from S4). The processing time for Case 1 algorithm varies between 0.33 and 0.39 ms per frame.

Table 3 presents the success rate in detecting the correct values of the letter/pillarbox effect boundaries, for Case 2, considering frame-by-frame analysis only, or by applying a temporal filter to the results; a boundary value is considered as correct if the resulting image aspect ratio is within 1% deviation of its true value. As for the temporal filter, a simple majority filter over a time window of 5 s (125 frames) was applied. The processing time for Case 2 algorithm varies between 3.09 and 3.45 ms per frame.

Table 2: Bar boundaries and detection success rate for Case 1, and a frame-by-frame analysis.

Sequence	Correct detections (%) (frame-by-frame analysis)			
	Left	Right	Top	Bottom
S1	100	100	100	100
S2(c)	100	100	100	28.12
S4	84.63	95.72	89.83	0.26

The results for sequence S4 are worse than for sequence S3 because a significant part of the frames in S4 are very dark in the borders of the image, leading to small errors in the detection.

Table 3: Success rate in the black bar boundaries detection for Case 2, with frame-by-frame analysis, and after temporal filtering.

Sequence	Correct detections (%) (frame-by-frame analysis)			
	Left	Right	Top	Bottom
S3	100	100	100	100
S4	73.57	93.03	89.58	73.76
Sequence	Correct detections (%) (after temporal filtering)			
	Left	Right	Top	Bottom
S3	100	100	100	100
S4	99.20	99.88	99.99	84.88

5.2 Subtitles and Logo Detection

The subtitle and logo detection performance is evaluated by the detection rate and correctly detected boundaries rate. Tables 4 and 5 present the obtained results for a frame-by-frame analysis. Note that sequence S4 contains a semi-transparent logo, whose boundaries are more difficult to detect accurately than an opaque logo. If a temporal majority filter is applied to the results, using a time window of 5 s (125 frames), the detection rates for sequence S4 will increase to 100%.

These results are slight superior to those from other authors, namely (Khodadadi & Behrad, 2012), with subtitles detection rates of 96.1/86.3/83.1% for different languages, and (Sang & Yan, 2011), with a subtitles detection rate of 92.2%. However, in our case the subtitle features can be better characterized since they can be extracted from the text over the letterbox, so a direct comparison cannot be made.

Table 4: Results for the detection and location of subtitles and logos for the test sequences.

Sequence	Subtitles or Logo detection (%)	Correctly detected boundaries (%) (inside the black bars)
S2a)	100.00	100.00
S2b)	99.72	100.00
S2c)	97.97	99.55
S2d)	100.00	100.00
S2e)	96.98	99.76
S5)	99.50	90.50
S4)	94.92	92.53
S6)	100.00	100.00

6 CONCLUSIONS

The main objective of this paper was to present an accurate and low complexity technique for the automatic detection of the active image area boundaries, in video sequences with letterbox/pillarbox effects. This method could be used in a quality control (QC) system placed at any point of the TV transmission chain, allowing the verification or settling of the correct AFD flags, and an optimized display of TV pictures.

In the proposed technique, two cases were considered: Case 1, corresponding to the situation where it can be assured that no relevant information (like logos and hard subtitles) are inlayed on the pillar/letterbox black bars; Case 2, to be used when Case 1 could not be guaranteed. In fact, since the objective of the AFD flags detection is to expand the active image to the size of a screen, if relevant information exists over the bars, it will be lost. Accordingly, the algorithm developed for Case 2 also detects the existence and position of subtitles and logos that may exist over the black bars.

Motivated by the results of Case 2, we have initiated the research for a method allowing the automatic extraction of the subtitles from the black bars (or from the active image area) and correspondent placement (or replacement) on the image active area, which could be used in advanced QC system; a preliminary version of it was also presented. Concerning this last method, more elaborated inpainting solutions should be investigated to cope with textured backgrounds and objects motion. Also, further work needs to be done for more accurate subtitle detection on the active image area, like including more discriminative texture information extracted from the subtitles that are inlayed on the letterbox.

ACKNOWLEDGEMENT

This work was developed under project "MOG – QC on the GO", Ref. SIIDT/2013/034075, funded by the European Regional Development Fund / QREN, Portugal.

REFERENCES

1. https://www.youtube.com/watch?v=IM_dOoUXgLE
2. https://www.youtube.com/watch?v=6NvVwA6Er8c
3. https://www.youtube.com/watch?v=ocTUey4mIac
4. https://www.youtube.com/watch?v=rtOvBOTyX00
5. https://www.youtube.com/watch?v=7Y6oRLJvjAo

Abraham, A., Prabhavathy, A., & Shree, J. (2012). A Survey on Video Inpainting. *International Journal of Computer Applications*.

IEEE - SP. (2011). Multimedia Quality Assessment - A World of Applications. *IEEE Signal Processing Magazine - Special Issue, 28*(6).

ITU-R. (2002). *Recommendation ITU-R BT.709-5 - Parameter values for the HDTV standards for production and international programme exchange.*

ITU-R. (2011). *Recommendation ITU-R BT.601-7 - Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios.*

Jones, G. (2006). Metadata for Formatting with Multiple Aspect Ratios. *HPA Technology Retreat.*

Khodadadi, M., & Behrad, A. (2012). Text Localization, Extraction and Inpainting in Color Images. *Iranian Conferance on Electrical Engineering,* (pp. 1035-1040). Tehran.

Kumar, S. (2010). Content Readiness Ensuring Content Quality Across Content Lifecycle. *Australian Broadcast Exhibition (ABE).* Sidney.

Markandey, V. (2002). *Patente N.º US 6,340,992 B1.* United Stated of America.

Sang, L., & Yan, J. (2011). Rolling and Non-Rolling Subtitle Detection with Temporal and Spatial Analysis for News Video. *International Conference on Modelling, Identification and Control*, (pp. 285-288). Shanghai.

Schoner, B., & Neuman, D. (2008). *Patente N.º US 7,339,627 B2.* United States of America.

Serra, J. (1983). *Image Analysis and Mathematical Morphology.* Academic Press.

SMPTE. (2009). *Format for Active Format Description and Bar Data, ST 2016-1.* SMPTE.

Tang, X., Gao, X., Lin, J., & Zhang, H. (July de 2002). A spatio-temporal approach for video caption detection and recognition. *IEEE Trans. on Neural Networks.*

Wu, H. R., & Rao, K. R. (2006). *Digital Video Image Quality and Perceptual Vision.* CRC Press.