Image Processing Theory, Tools and Applications

Semi-Automatic News Video Annotation Framework for Arabic Text

Oussama Zayene[1, 2], Sameh Masmoudi Touj[2], Jean Hennebert[1, 3], Rolf Ingold[1] and Najoua Essoukri Ben Amara[2]

[1] DIVA group
Department of Informatics, University of Fribourg (Unifr)
Fribourg, Switzerland
[3] Institute of Complex Systems
HES-SO, University of Applied Science Western Switzerland
{oussama.zayene, jean.hennebert, rolf.ingold}@unifr.ch

[2] SAGE group
National Engineering School of Sousse (Eniso)
Sousse, Tunisia
najoua.benamara@eniso.rnu.tn
samehmasmouditouj@yahoo.fr

***Abstract*—In this paper, we present a semi-automatic news video annotation tool. The tool and its algorithms are dedicated to artificial Arabic text embedded in video news in the form of static text as well as scrolling one. It is performed at two different levels. Including specificities of Arabic script, the tool manages a global level which concerns the entire video and a local level which concerns any specific frame extracted from the video. The global annotation is performed manually thanks to a user interface. As a result of this step, we obtain the global xml file. The local annotation at the frame level is done automatically according to the information contained in the global meta-file and a proposed text tracking algorithm. The main application of our tool is the ground truthing of textual information in video content. It is being used for this purpose in the Arabic Text in Video (AcTiV) database project in our lab. One of the functions that AcTiV provides, is a benchmark to compare existing and future Arabic video OCR systems.**

***Keywords*—Benchmarking VideoOCR systems; annotation; artificial Arabic text; data sets.**

I. INTRODUCTION

Nowadays, the size of the available digital video content is increasing rapidly to satisfy the extremely diverse demands of the public. Broadcast news is considered as one of the most important information source. It provides a picture of what is happening at the local, national, and international levels. Analysis of public newscast by national as well as foreign news TV channels is of capital importance for media analysts in several domains such as politics, economy, law enforcement and internal security. This situation creates an urgent need for fast and effective information retrieval algorithms for multimedia contents. Text displayed in news video is one of the most important high-level information of the video content. In general, it can be classified into scene text and artificial text. Compared with scene text, the artificial text can provide brief and direct description of video content such as speaker's name, subtitles, location, event information, etc. Therefore, in this context we mainly focus on artificial text annotation in news videos.

Compared to images, text annotation in videos presents some difficulties. One challenging aspect of this problem is the tracking of dynamic text through the video stream. We aim in this work to conceive and implement a semi-automatic annotation framework for the artificial text (static and dynamic) in news video with a high level reliability thanks to human contribution in both annotation and verification process.

Our contribution concerns specifically Arabic video news text annotation. Analysis of Arabic documents and recognition/indexing of Arabic text actually became an attracting research domain in the recent years [4, 5, 6, 7]. Major contributions have already been made in the field of printed and handwritten Arabic text data sets and OCR systems [4, 6, 8, 9, 10, 11].

However, to the best of our knowledge, few attempts have yet been made on the development of Arabic text data sets for videos. Previous related works such as [2, 3] deal directly with annotation of text in video and video frames. Siddiqi and Raza [2] use a somewhat similar idea of annotation. However, the approach differs from the algorithm developed in this paper. The authors proposed a semi-automatic text line labelling scheme which is targeted towards the evaluation of Urdu text localization systems. The limitations of this method include the absence of ground truth data to evaluate the OCR systems performances, and inability to track and annotate the scrolling text. Another method presented in [3], uses a free ground truth authoring tool ViPER [1] (Video Performance Evaluation Resource) to annotate text objects in video content. However, ViPER annotation methodology is dedicated to Latin and cannot be replicated for Arabic texts which is the focus of our work. Specific traits of Arabic script include non-uniform intra/inter word distances, diacritics and the cursive nature of the strokes.

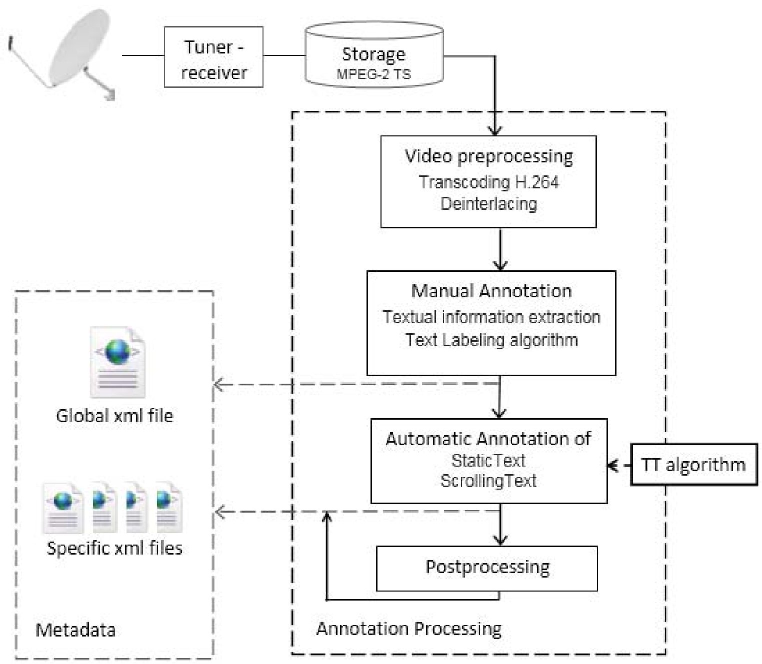

Fig. 1. The semi-automatic process for news video text annotation

This work falls within the framework for the development of a standard database of Arabic Text in Video (AcTiV) that will be used as benchmark to compare and evaluate existing and future video OCR systems [12, 13, 14, 15, 16, 17].

Figure 1 depicts the architecture of our semi-automatic algorithm for Arabic news video annotation. This paper is organized as follows: data acquisition and video preprocessing steps are presented in section II. In section III, we will describe the global video annotation. The automatic frame annotation process is described in section IV. Results are presented and discussed in Section V.

II. DATA ACQUISITION & VIDEO PREPROCESSING

The broadcast of new channels includes a wide variety of programs going from talk shows and interviews to documentaries and weather reports. News reports were specifically chosen for the present study.

In order to ensure maximum diversities of the content and avoid repetition, recordings from the same channel were spaced by one week and the duration of a single video was limited to less than 10 minutes.

The broadcast streams were captured from a Direct Broadcast Satellite (DBS) system. The video stream was initially saved unaltered on the hard drive (MPEG-TS).Then, a transcoding process took place to convert the interlaced (MPEG2/MPEG4) video to a de-interlaced MPEG4-AVC using an x264 based encoder and applying a YADIF filter. The goal of this, is both to prepare the video to frame-by-frame analysis and to lower the video bitrate without perceived quality loss. In the present work, two types of video stream were chosen: Standard-Definition (720x576, 25 fps) and High-Definition (1920x1080, 25fps).

III. GLOBAL VIDEO ANNOTATION

This stage involves the annotation of text displayed as overlay in news video stream. Two main types of text can be identified: static and dynamic (also called scrolling text).

A text that does not undergo a change in its content, position, size, or color within its display interval is considered as static text. This group usually includes event information, speaker's name, subtitles, etc. Dynamic text targeted in our study, refers to the horizontal scrolling text that usually resides in the lower third of the television screen. In Arabic channels, dynamic text moves from left-to-right.

The global labeling process of text regions in videos is performed manually thanks to a user interface which includes a set of functionalities especially dedicated for ground truthing of textual information in video content. The figure 2 demonstrates different types of texts from 2 different Arabic channels displayed in our software.

A. Static text

Identification and selection of static text regions present a challenge given the variability of texts in video broadcast (in alignment, size, color, etc). Therefore, we established a methodology to identify and locate elemental text regions:

- A single rectangle is drawn over texts that belong to the same semantic unit and have uniform size, alignment and spacing.
- Multiple lines of text can be selected within the same rectangle.

Once a rectangle has been selected (as depicted in Figure 2-a), a new set of information is created. It contains content data: text, textColor, backgroundColor, backgroundType<transparent, opaque> and time stamps for its apparition (first frame, last frame). This set of information along with the rectangle's information (x and y coordinate, width, height) is saved in a global xml file.

(a)

(b)

Fig. 2. The user interface of the ground truthing software displaying labeled frames

Arabic letters can be written in different shapes depending on their position in the word. In order to have an easily accessible representation of Arabic text for future processing, it is transformed into a set of labels with a suffix that refers to the letter's position in the word (B: Begin, M: Middle, E: End and I: Isolate). We use the same labels used in the published work [10] to standardize the character labels for Arabic text. An example of Arabic sentence and its corresponding Latin labels is shown in Figure 3.

Fig. 3. An Arabic sentence and its corresponding labels

A transcription label is generated for every Arabic text stored in the xml file (static or dynamic); it is saved under the attribute transcriptionLabel within the same element that contains the Arabic text.

In addition to these data, other informations are stored in the global xml file such as the total number of textBox in the video, the text font, the number of apparition intervals for each textBox, etc. An extract of a global xml file is illustrated in Figure 7.

B. *Dynamic text*

Dynamic text is formed by continuous scrolling series of tickers that can be delimited by a separator like the channel logo. Examples are shown on Figures 4 and 6. For each ticker, we noted its content, the first frame where the ticker appears, and the initial offset in the first frame which is calculated using a virtual line as shown in Figure 2-b. This information is stored in the global xml file.

We can use the information entered in this step along with other video-specific information (as detailed later) to implement a tracking algorithm that can generate specific xml files describing precisely the text for each frame.

IV. AUTOMATIC FRAME LEVEL ANNOTATION

The final goal of this step is to automatically generate descriptors at the frame level. As for the global level, specific xml files are used and contain an accurate description of the displayed text on the screen (dynamic and static).

Given that we have already, in the global xml file, specified the appearance interval of every static text, it is here easy to extract for each frame its corresponding static texts and record them in specific xml files.

Dynamic text is more challenging. We need to develop a so-called *tracking algorithm* that will take into account channel specific information such as scrolling speed and characters widths.

We calculate the scrolling speed (1) based on the assumption that scrolling texts move at a constant pace. We track the progression of an element in the scrolling band through Δf frames as shown in Figure 5. Δp denotes the distance -in pixels- that the element made in Δf frames.

$$\text{Scrolling speed} = \frac{\Delta p}{\Delta f} \ \textit{pixels/frame} \qquad (1)$$

Several tests were performed on various video segments from the same channel to validate the scrolling speed formula.

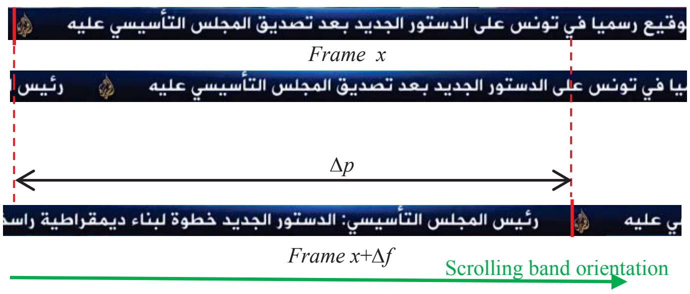

Fig.4. Estimation of the scrolling speed for the AljazeeraHD channel

The dynamic text is usually generated with the same font and size, allowing the use of a table containing character width information of the used font (see Table I for an example).

A. *Text tracking algorithm*

The main idea of our algorithm is to track the dynamic text based on the scrolling speed and the character's width table (see the TT algorithm block of Figure 1).

Figure 5 illustrates the steps of the algorithm. VisibleText is the principal function of the tracking algorithm which is further described below in Algorithm 1. Its output is the part of the ticker

TABLE I. A PART OF THE CHARACTER'S WIDTH TABLE OF ALJAZEERAHD FONT

Character Labels	**Isolate**	**Begin**	**Middle**	**End**	**Width (pxl)**			
					_I	*_B*	*_M*	*_E*
Alif	Alif_I ا		Alif_E ـا		7		8	
Baa	Baa_I ب	Baa_B بـ	Baa_M ـبـ	Baa_E ـب	31	18	15	31
Taaa	Taaa_I ت	Taaa_B تـ	Taaa_M ـتـ	Taaa_E ـت	31	20	21	30
TaaaClosed	TaaaClosed_I ة		TaaaClosed_E ـة		18		20	
Thaa	Thaa_I ث	Thaa_B ثـ	Thaa_M ـثـ	Thaa_E ـث	30	20	21	31
Jiim	Jiim_I ج	Jiim_B جـ	Jiim_I ـجـ	Jiim_E ـج	26	28	24	26
Haaa	Haaa_I ح	Haaa_B حـ	Haaa_M ـحـ	Haaa_E ـح	26	28	24	26
Xaa	Xaa_I خ	Xaa_B خـ	Xaa_M ـخـ	Xaa_E ـخ	26	27	24	26
Daal	Daal_I د		Daal_E ـد		17		17	
Thaal	Thaal_I ذ		Thaal_E ـذ		18		19	
Raa	Raa_I ر		Raa_E ـر		15		15	
Zaay	Zaay_I ز		Zaay_E ـز		15		15	
Siin	Siin_I س	Siin_B سـ	Siin_M ـسـ	Siin_E ـس	45	31	33	45
Shiin	Shiin_I ش	Shiin_B شـ	Shiin_M ـشـ	Shiin_E ـش	45	31	32	45
Saad	Saad_I ص	Saad_B صـ	Saad_M ـصـ	Saad_E ـص	48	35	36	48
Daad	Daad_I ض	Daad_B ضـ	Daad_M ـضـ	Daad_E ـض	48	35	37	48
Thaaa	Thaaa_I ط	Thaaa_B طـ	Thaaa_M ـطـ	Thaaa_E ـط	31	31	30	32

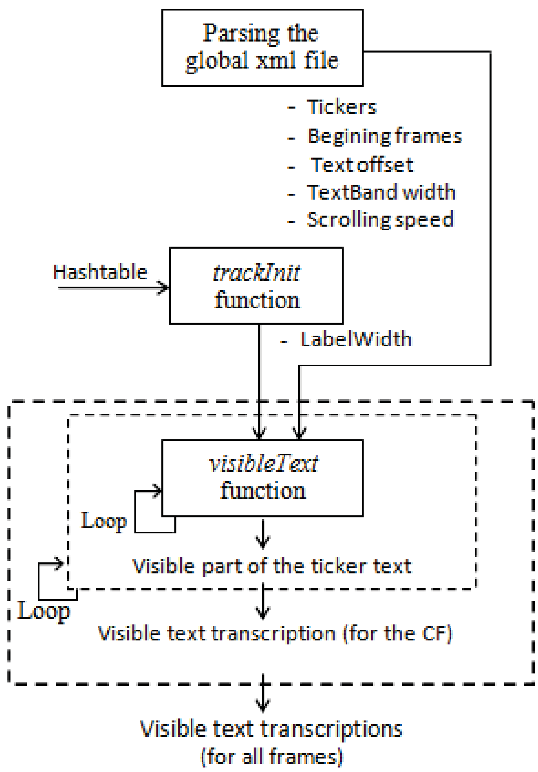

Fig.5. Principal steps of the Tracking text algorithm

that is visible at a given frame. The width label is retrieved from the character's width table using a public function (trackInit).

Algorithm 1: **visibleText**

Input: news ticker *text*, Beginning Frame *BF*, *offset*, Current Frame *CF*
Output: visible text on the screen *outputText*
Require: *TextBandWidth*, *scrollingSpeed*
Split the ticker text to an array of strings (textArray)
// TWB: visible Text Window Beginning
TWB = (CF−BF)× scrollingSpeed+ (offset−TextBandWidth)
//TWE: visible Text Window End
TWE= (CF−BF) × scrollingSpeed+ offset
Cursor=0
for all element label in the textArray **do**
    Add the width of label to cursor
If cursor>=TWB **then** // Beginning of visible text region
        Concatenate label to outputText
    if cursor >TWE **then** break// End of visible text region
    end if
end if
end for

TWB and TWE are used as virtual lines to delimit and define the visible part of the ticker text in the current frame CF.

Initially, the apparition interval of every ticker text is computed based on the scrolling speed. Every frame is processed separately. To determine the dynamic text visible at a given frame, we use the apparition interval to select a set of ticker texts that are partially/totally visible at that frame, and then we use visibleText function to get a precise output, as illustrated in Figure 5.

V. PRELIMINARY EXPERIMENTS

For experimentation, we took 5 news clips, two SD videos from the "France24 arabic" channel and three HD videos from "AljazeeraHD" channel as shown in table II, in order to ensure a level of diversity on the characteristics of processed static and dynamic texts. We then applied the methods as explained in section II and IV on them. The program is implemented in Java.

TABLE II. CHARACTERISTICS OF THE USED NEWS VIDEOS

Resolution	Channel	Video ID	Duration
SD 720x576	France24 arabic	1	00:05:57
		2	00:10:18
HD 1920x1080	Aljazeera HD	3	00:06:57
		4	00:09:17
		5	00:10:17

The performance measure for the tracking task is based on inserting and deleting errors at the character level. A manual checking was performed on a subset of frames (1% of the total frames) selected from each video clip, as shown in table III. Each character fully added or deleted in each frame is considered as error. The average observed Character Error Rate (CER) of the Tracking Text algorithm is less than 0,4 characters per frame. Such insertion/deletion errors would probably have few or no impacts in evaluation protocols that would consider full word recognition scenario.

This fault rate is caused by imprecision in the character's width table. In fact, the values in this table were manually measured and saved for every channel. The difficulties of creating high precision character tables is very probably related to inconsistent fonts, font rendering techniques and video stream post processing between channels.

TABLE III. CHARACTER ERROR RATE ESTIMATION

Video ID	Total Number of frames	Subset of frames	Overall-Character Errors
1	8934	90	41
2	15458	155	56
3	10425	104	32
4	13925	139	61
5	15425	154	47
Total	**64167**	**642**	**237**
CER	**0,36**		

Besides, we notice that the ticker based approach in the adopted TT algorithm prevents from the repercussion of any gap error to the following tickers.

Figure 6 shows a frame and its specific xml file that was automatically generated.


```xml
<video id="2">
  <Frame id="425">
   <staticText nbTextBox="1">
    <textBox id="1">
     <position x="1145" y="910" width="296" height="70" />
     <content nbTextLines="1" textColor="51,65,90" bgColor="240,242,237"
    bgType="opaque">
        <textLine id="1" transcription="ليلى الشيخلي " transcriptionLabel=" Laam_B Yaa_M
    Laam_M AlifBroken_E Space Alif_I Laam_B Shiin_M Yaa_M Xaa_M Laam_M
    Yaa_E" />
    </content>
   </textBox>
  </staticText>
  <scrollingText nbOfvisibleTicker="2">
    <bandPosition x="0" y="977" width="1432" height="65" />
    <content textColor="251,251,255" bgColor="" bgType="opaque">
     <visibleTicker id="1" transcription="لديمقراطية في تركيا" transcriptionLabel=" Laam_B
    Daal_E Yaa_B Miim_M Gaaf_M Raa_E Alif_I Thaaa_B Yaa_M TaaaClosed_E
    Space Faa_B Yaa_E Space Taaa_B Raa_E Kaaf_B Yaa_M Alif_E" />
    <visibleTicker id="2" transcription="استمرار مباحثات أديس أبابا بين طرفي النزاع في دولة ج"
    transcriptionLabel="Alif_I Siin_B Taaa_M Miim_M Raa_E Alif_I Raa_I Space
    Miim_B Baa_M Alif_E Haaa_B Thaa_M Alif_E Taaa_I Space HamzaAboveAlif_I
    Daal_I Yaa_B Siin_E Space HamzaAboveAlif_I Baa_B Alif_E Baa_B Alif_E Space
    Baa_B Yaa_M Nuun_E Space Thaaa_B Raa_E Faa_B Yaa_E Space Alif_I Laam_B
    Nuun_M Zaay_E Alif_I Ayn_I Space Faa_B Yaa_E Space Daal_I Waaw_I Laam_B
    TaaaClosed_E Space Jiim_B" />
    </content>
   </scrollingText>
  </Frame>
</video>
```


Fig. 6. A sample frame and its correspondent specific xml file

```
<video    id="3"    channel="AljazeeraHD"    resolution="1080p"    duration="00:09:17"    fps="25"
    nbOfFrames="13925">
 <staticText  nbOftextBox="49"  font="aljazeeraFont">
  <textBox  id="1"  nbOfaInterval="1">
    <aInterval id="1" frame_S="134" frame_E="376" />
    <position x="482" y="889" width="912" height="90" />
    <content nbTextLines="1" textColor="252, 252,250" bgColor="" bgType="transparent">
        <textLine  id="1"  transcription="ترحيب بالدستور التونسي الجديد" transcriptionLabel="Taaa_B Raa_E Haaa_B
        Yaa_M Baa_E Space Baa_B Alif_E Laam_B Daal_E Siin_B Taaa_M Waaw_E Raa_I Space Alif_I
        Laam_B Taaa_M Waaw_E Nuun_B Siin_M Yaa_E Space Alif_I Laam_B Jiim_M Daal_E Yaa_B
        Daal_E" />
    </content>
   </textBox>
     ...

  <textBox  id="11"  nbOfaInterval="1">
    <aInterval id="1" frame_S="2979" frame_E="3043" />
    <position x="1070" y="849" width="360" height="64" />
    <content nbTextLines="1" textColor="51, 65,90" bgColor="240, 242,237" bgType= "opaque" >
        <textLine  id="1"  transcription="تقرير عمر خشرم" transcriptionLabel="Taaa_B  Gaaf_M  Raa_E  Yaa_B
        Raa_E Space Ayn_B Miim_M Raa_E Space Xaa_B Shiin_M Raa_E Miim_I" />
    </content>
   </textBox>

     ...

   </staticText>
```


(a)


```
 <scrollingText    orientation="left-right"    textColor="251,251,255"    bgColor=""    bgType="opaque"
     runningSpeed="6,770 pixel/frame">
    <bandPosition x="0" y="977" width="1432" height="65"/>
    <content nbOftickerInformation="56">
      <tickerInformation id="1" frame_S="252" offset="4"  transcription="مصر: المجلس الأعلى للقوات المسلحة يفوض السيسي
     الترشح للرئاسة" transcriptionLabel="Miim_B  Saad_M  Raa_E  Colon  Space  Alif_I  Laam_B  Miim_M  Jiim_M
     Laam_M  Siin_E  Space  Alif_I  Laam_EHamzaAboveAlif_E  Ayn_B  Laam_M  AlifBroken_E  Space  Laam_B
     Laam_M  Gaaf_M  Waaw_E  Alif_I  Taaa_I  Space  Alif_I  Laam_B  Miim_M  Siin_M  Laam_M  Haaa_M
     TaaaClosed_E Space Yaa_B Faa_M Waaw_E Daad_I Space Alif_I Laam_B Siin_M Yaa_M Siin_M Yaa_E
     Space  Alif_I  Laam_B  Taaa_M  Raa_E  Shiin_B  Haaa_E  Space  Laam_B  Laam_M  Raa_E
     HamzaAboveAlifBroken_B Alif_E Siin_B TaaaClosed_E"/>
      <tickerInformation id="2" frame_S="443" offset="0"  transcription="المجلس العسكري: لم يكن بوسعنا إلا الاستجابة لرغبة
     الجماهير في ترشيح السيسي" transcriptionLabel="Alif_I  Laam_B  Miim_M  Jiim_M  Laam_M  Siin_E  Space  Alif_I
     Laam_B Ayn_M Siin_M Kaaf_M Raa_E Yaa_I Colon Space Laam_B Miim_E Space Yaa_B Kaaf_M Nuun_E
     Space  Baa_B  Waaw_E  Siin_B  Ayn_M  Nuun_M  Alif_E  Space  HamzaUnderAlif_I  Laam_EAlif_E  Space
     Alif_I Laam_EAlif_E Siin_B Taaa_M Jiim_M Alif_E Baa_B TaaaClosed_E Space Laam_B Raa_E Ghayn_B
     Baa_M TaaaClosed_E Space Alif_I Laam_B Jiim_M Miim_M Alif_E Haa_B Yaa_M Raa_E Space Faa_B
     Yaa_E  Space  Taaa_B  Raa_E  Shiin_B  Yaa_M  Haaa_E  Space  Alif_I  Laam_B  Siin_M  Yaa_M  Siin_M
     Yaa_E"/>
     ...
 </scrollingText>
</video>
```


(b)

Fig.7. Global xml file- (a) part of the static text (b) part of the scrolling text

A manual post-processing verification step can also be applied out to reduce the number of deletion and insertion errors (see Postprocessing block of Figure 1).

VI. Conclusion

In this paper, we have presented a semi-automatic algorithm for artificial Arabic text annotation in news video. The motivation of this work is grounded in the fact that the number of public databases for Arabic news videos is limited. Our algorithm will be used as a tool for the development of standard databases. For this purpose, we have targeted Arabic OCR systems and text detection systems which require the text transcription and the coordinates of all text regions as ground truth data. A specificity of our proposal is to treat dynamic "scrolling" text found typically in news channels.

The performance of our annotation algorithm for dynamic text could potentially be enhanced by fine-tuning the character's width tables or possibly by obtaining accurate font information from the source. However, the obtained precision at the frame level is good enough to generate ground-truth with relatively few manual works and to allow for larger scale benchmarking of video OCR.

A future work will be in the diffusion of a news video corpus with ground truth annotation and in the definition of benchmarking protocols for Arabic text in video.

References

[1] D. Doermann and D. Mihalcik, "Tools and Techniques for Video Performance Evaluation". *International Conference on Pattern Recognition (ICPR 2000), pp.167–170.*

[2] I. Siddiqi and A. Raza, "A Database of Artificial Urdu Text in Video Images with Semi-Automatic Text Line Labeling Scheme", *The Fourth International Conferences on Advances in Multimedia (MMEDIA'12)*.

[3] P. Soundararajan, M. Boonstra, V. Manohar, V. Korzhova, D. Goldgof, R. Kasturi, S. Prasad, H. Raju, R. Bowers and J. Garofolo, "Evaluation Framework for Video OCR", *Indian Conference on Computer Vision, Graphics and Image Processing* (*ICVGIP'06)* .

[4] F. Slimane, O. Zayene, S. Kanoun, A. M. Alimi, J. Hennebert, R. Ingold, "New features for complex Arabic fonts in cascading recognition system", *International Conference on Pattern Recognition (ICPR'12), pp.738-741*.

[5] S. Touj, N. Ben Amara et H. Amiri, "A hybrid approach for off-line Arabic handwriting recognition based on a Planar Hidden Markov modeling", *International Conference on Document Analysis and Recognition (ICDAR'07)*.

[6] R. El-Hajj , C. Mokbel, L. Likforman-Sulem, "Recognition of Arabic handwritten words using contextual character models", *15th Document Recognition and Retrieval Conference (DRR'08), San Jose, CA, USA, January 2008.*

[7] Al-Muhtaseb H. A., Mahmoud Sabri A., Qahwaji R. S., "Recognition of off-line printed Arabic text using Hidden Markov Models", *European Signal Processing Conference, Lausanne, Switzerland 2008, vol. 88, Issue 12, pp. 2902 – 2912*.

[8] N. Kharma, M. Ahmed , R. Ward, "A new comprehensive database of handwritten arabic words, numbers, and signatures used for ocr testing", *Canadian Conference on Electrical and Computer Engineering (CCECE'99), pp.766–768.*

[9] M. Pechwitz, S. Maddouri, V. Maergner, N. Ellouze, H. Amiri, "Ifn/enit database of handwritten Arabic words", *Colloque International Francophone sur l'Écrit et le Document (CIFED'02), Hammamet, Tunisie, octobre 2002*.

[10] F. Slimane, R. Ingold, S. Kanoun, A. M. Alimi, J. Hennebert, "A New Arabic Printed Text Image Database and Evaluation Protocols", *International Conference on Document Analysis and Recognition (ICDAR'09), Barcelona, Spain, July 2009*.

[11] M. A. Charrada, N. Essoukri Ben Amara, "Development of a database with ground truth for old documents analysis", *International Multi-Conference on Systems, Signals and Devices (SSD'13), Hammamet, Tunisia, March 2013*.

[12] R. Lienhart, "VideoOCR: A Survey and Practitioner's Guide," *in Video Mining, Kluwer Academic Publisher, October 2003, pp. 155-184.*

[13] H. Karray, A. M. Alimi, Detection and Extraction of the Text in a video sequence, *International Conference on Electronics, Circuits and Systems (ICECS'05), Gammarth, Tunisia, December 2005, pp. 474–478*.

[14] M. Ben Halima, H. Karray, A.M. Alimi, "A Comprehensive Method for Arabic Video Text Detection, Localization, Extraction and Recognition", *Advances in Multimedia Information Processing - PCM 2010*.

[15] M. Moradi, S. Mozaffari, A. Asghar Orouji, "Farsi/Arabic Text Extraction from Video Images by Corner Detection", *Iranian Conference on Machine Vision and Image Processing (MVIP'10).*

[16] M. Ben Halima, H. Karray, A.M. Alimi, A. Fernandez Vila,"NF-SAVO: Neuro-Fuzzy system for Arabic Video OCR", *International Journal of Advanced Computer Science & Applications (IJACSA), vol. 3 Issue 10, October 2012, pp.128-136*.

[17] Abbas H. AL-Asadi and Thika Ali H. Subber, "Arabic- text Extraction from Video Images", *Journal of Basrah Researches- Sciences (JBR- S), vol.39, 2013.*