Racing Bib Numbers Recognition

Idan Ben-Ami
idan.benami@gmail.com
Tali Basha
talib@eng.tau.ac.il
Shai Avidan
avidan@eng.tau.ac.il

School of Electrical Engineering,
Tel Aviv University,
Tel Aviv 69978, Israel

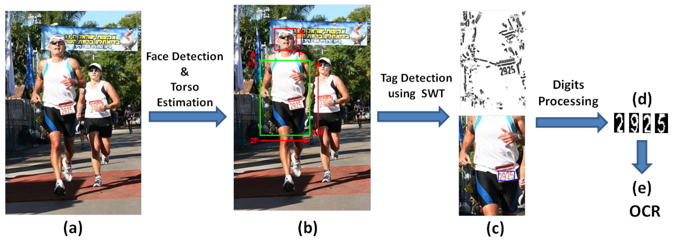

Figure 1: **Method Outline:** **(a)** the input image; **(b)** face detection results in red; the hypothesis estimated region of the RBN in green; **(c)** the stroke-width map of the hypothesis region (top) and the detected tag (bottom); **(d)** the detected tag after processing is fed to the OCR **(e)**.

Figure 2: **(a)** The result of applying SWT on the entire image; the detected text regions are marked in blue inside the yellow dashed regions; **(b)** zoom-in of the yellow dashed regions; **(c)** the stroke width map computed on the entire image; **(d)** the detected face in red, and the hypothesis tag region in green;

Running races, such as marathons, are broadly covered by professional as well as amateur photographers. This leads to a constantly growing number of photos covering a race, making the process of identifying a particular runner in such datasets difficult. Today, such identification is often done manually. In running races, each competitor has an identification number, called the Racing Bib Number (RBN), used to identify that competitor during the race. RBNs are usually printed on a paper or cardboard tag and pined onto the competitor's T-shirt during the race. We introduce an automatic system that receives a set of natural images taken in running sports events and outputs the participants' RBN.

This specific application can be studied in the wider context of detecting and recognizing text in natural images of unstructured scenes. Existing methods that fall into this category fail to reliably recognize RBNs (as demonstrated in our experiments), due to the large variability in their appearance, size, and the deformations they undergo. The RBNs usually cover only a small portion of the image and are surrounded by complex backgrounds. Moreover, the images often contain irrelevant text such as sponsor billboards, signs, or text printed on people's clothes. Therefore, text detection methods are expected to be inefficient and to produce many false detections. This is demonstrated in Fig. 2(a)-(c), showing the results of applying SWT [1] on the entire image.

In this paper, we propose a method specifically designed for RBN recognition. Our method can be applied without any adjustments to images taken at various running races by different photographers. We show that by using prior information - the spatial relation between a person's face and the tag he/she wears, we obtain an effective RBNR system that outperforms text detection methods and state of-the-art commercial LPR software.

Method Outline: The input to our method is a collection of images covering a running race. The images are generally different in size, resolution, and viewpoint, and may be taken using different cameras. Each image is assumed to capture one or more participants. The RBN tags are allowed to have any color, font and scale. The only assumption is that the RBN tag is located on the front torso of the participant. The outline of our method is shown in Fig. 1. First, we use a face detector [2] to generate hypotheses regarding the RBN location and scale. We then adapt and enhance the stroke width transform (SWT) [1] to detect the location of the tag, which is then processed and fed to a standard optical character recognition (OCR) engine [3].

Results in a Glance: To test our method, we collected images from running races found on the Web. Our database includes 217 color images divided into three sets, each taken from a different race. The tag dimensions vary between 13x28–120x182 pixels while digit stroke widths vary from 13 pixels to as few as 2 pixels in the smallest tags. To verify and compare our results, we manually generated the ground truth RBNs.

Comparison to Conventional Approaches: In this experiment, we evaluated the contribution of each component in our pipeline. To do so, we compared our full pipeline (**"Face Detector+E-SWT+OCR"**) to the following two sub-systems: (1) **"SWT+OCR":** the SWT is applied on the entire image to locate the RBNs, followed by standard OCR on the detected text regions; (2) **"Face Detector+E-SWT+OCR":** the face detector is added to generate hypothesis search regions. The SWT combined with our enhancements (noted by E-SWT) is applied on the hypothesis regions, followed by standard OCR. Fig. 3 presents the results of the two sub-systems compared to our full pipeline on one of the datasets.

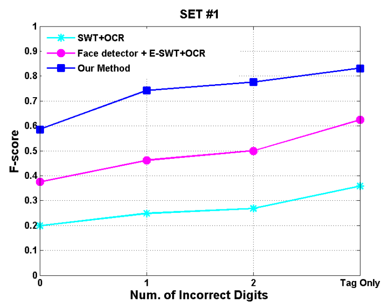

Figure 3: The graph show the computed F-score for each dataset, for the cases of perfect detection, one wrong digit, two wrong digits, and only tag (the tag area is correctly detected but with more than 2 wrong digits).

Comparison to LPR: We compared our performance to the CARMEN FreeFlow LPR system. CARMEN is a leading, commercial, general-purpose LPR system, that provides high rate plate recognitions in a large variety of image scenes and plate types. To adapt CARMEN for our purpose, a special parameter adjustment was required. The performance of CARMEN on our datasets is shown in Table 1. For each dataset, the precision, recall and F-score are measured for perfectly correct recognition (i.e., all digits are correct). The results indicate that our system achieved higher F-score than CARMEN (we achieved higher recall than CARMEN and similar precision rate).

[1] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. *CVPR*, 2010.

[2] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–900. Ieee, 2002.

[3] R. Smith. An overview of the tesseract ocr engine. *ICDAR*, 2007.

	Set #1			Set #2			Set #3		
	Prec.	Rec.	F	Prec.	Rec.	F	Prec.	Rec.	F
Our method	**0.66**	**0.50**	**0.57**	**0.75**	**0.45**	**0.56**	**0.65**	**0.62**	**0.63**
CARMEN	0.67	0.37	0.48	0.68	0.38	0.49	0.73	0.47	0.57

Table 1: The computed precision, recall and F-score (w.r.t the ground truth) of our results compared with the results of CARMEN LPR system.