

Automated Recognition of Sports Scores using PyTesseract OCR and CNN: SportsVideo Task at MediaEval 2023

Bhuvana Jayaraman[†], Mirnalinee TT, Harshida Sujatha Palaniraj[†], Mohith Adluru[†] and Sanjjit Sounderrajan[†]

Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India

Abstract

In the dynamic realm of competitive swimming, timely and accurate dissemination of race results is paramount. The advent of digital display boards has streamlined this process, offering real-time performance metrics for each swimmer. However, extracting and recognizing the characters displayed on these boards pose unique challenges. Our proposed methodology addresses these intricacies, offering a robust solution for real-time score recognition amidst dynamic game-play and varying camera angles. The systematic approach leverages advanced image preprocessing techniques to refine input images, employs PyTesseract OCR for real-time interpretation of textual information from swimming board images, and integrates post-processing validation to enhance the accuracy of the extracted race results. Our findings not only contribute to the evolving field of sports analytics but also pave the way for enhanced viewer experience and enriched data accessibility in the realm of sports broadcasting.

1. Introduction

Swimming competitions have long been a focal point of sporting events, attracting participants and enthusiasts alike. In this era of fast-paced sporting events, the real-time availability of race results is of paramount importance, offering athletes insights into their performance and spectators the thrill of immediate competition outcomes.

Digital boards have emerged as indispensable tools in this context, acting as conduits for promptly showcasing race results. These boards, equipped with advanced display technology, present a visually rich representation of crucial information such as swimmer names, times, and rankings. However, the process of manually extracting this information from the visually diverse and dynamic context of swimming events is arduous and prone to errors [1].

The motivation behind this research stems from the recognition of the transformative potential of OCR technology in addressing these challenges. By automating the extraction of textual information from digital boards, OCR not only promises to expedite the process but also ensures a higher degree of accuracy. This, in turn, enhances the overall experience for both participants and spectators, fostering a more efficient and error-resistant mechanism for obtaining vital race results.

MediaEval'23: Multimedia Evaluation Workshop, February 1–2, 2024, Amsterdam, The Netherlands and Online

*Corresponding author.

[†]These authors contributed equally.

✉ bhuvanaj@ssn.edu.in (B. Jayaraman); mirnalineett@ssn.edu.in (M. TT); harshida2110349@ssn.edu.in (H. S. Palaniraj); mohith2110799@ssn.edu.in (M. Adluru); sanjjit2110378@ssn.edu.in (S. Sounderrajan)

🌐 <https://www.ssn.edu.in/staff-members/dr-j-bhuvana/> (B. Jayaraman);

<https://www.ssn.edu.in/staff-members/dr-t-t-mirnalinee/> (M. TT)

🆔 0000-0002-9328-6989 (B. Jayaraman); 0000-0001-6403-3520 (M. TT)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

The primary objectives of this research, within the context of Task 6 to recognize results of races [2], involve evaluating OCR's effectiveness in recognizing characters under diverse conditions encountered in swimming environments and developing a tailored preprocessing pipeline to enhance its adaptability. By achieving these objectives, the study seeks to contribute to the efficient and accurate extraction of critical race information, offering potential applications beyond competitive swimming in visually complex scenarios.

2. Related Work

Extraction and recognition of text from an image is an important step to display and predict the results. Tesseract OCR is implemented to extract texts from given video frames or images with an additional image processing filter, OpenCV [3]. Some of the processes used in are morphological operations and Gabor filters for preprocessing, edge detection, DWT and SWT for feature extraction and Hough transform and k-means clustering for segmentation [4].

For unconstrained image indexing and retrieval system using a neural network, this paper suggests extracting a set of features from ROI for that specific color plane and using them further in a feature-based classifier to determine if the ROI contains text or non-text blocks. The blocks identified as text are next given as input to an OCR. The OCR output in the form of ASCII characters forming words is stored in a database as keywords with reference for future retrieval [5].

The existing algorithms have been modified to be effective on blurred and noisy images. Images with clear letter boundaries are filtered using OCR to detect the text [6].

Text extraction involves detection, localization, tracking, binarization, extraction, enhancement and recognition of the text from the given image. The proposed methods were based on morphological operators, wavelet transform, artificial neural network, skeletonization operation, edge detection algorithm, histogram technique, etc [7]. The method used to extract text regions for the objective of image segmentation uses DWT and k-means clustering [8] [9]. Moreover, methods such as texture-based text extraction approach to existing edge-based and connected component (CC)-based algorithms were proposed [10] for text extraction due to variables in orientation, alignment, font, size, poor picture contrast, and complex background in the text.

This paper focuses on text segmentation in textured backgrounds with similar colors to text, outperforming recent works on challenging images.[1].

3. Approach

Our comprehensive approach to the automated extraction of swimming race results integrates image preprocessing, Optical Character Recognition (OCR) with PyTesseract, and meticulous post-processing validation. Leveraging OpenCV and PyTesseract, we systematically optimize input images, configure OCR settings, and validate extracted information. Simultaneously, our innovative solution incorporates a hybrid methodology for score extraction in table tennis and swimming events. This approach combines Convolutional Neural Network (CNN) feature extraction with Tesseract OCR for both visual and textual cues. Initial benchmarks involved training the CNN on visual features and parallel processing with Tesseract OCR for text recognition. Successes include accurate score extraction and additional context, while challenges involved varying fonts and layouts. Refined CNN architecture, preprocessing optimization, and continuous learning mechanisms addressed these challenges. Real-world feedback and validation on diverse datasets guided our iterative development, resulting in a robust solution for both swimming race results and sports event scores.

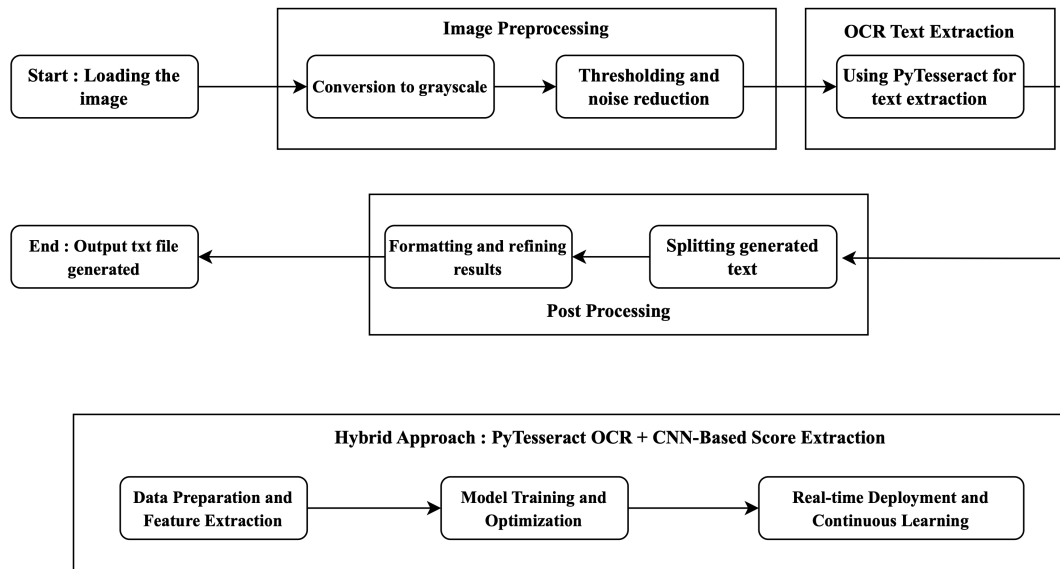


Figure 1: Architecture overview

3.1. Dataset Description

The dataset comprises a collection of images capturing swimming scoreboards with associated player names and scores, sourced from various swimming competitions. The dataset includes variations in lighting conditions, background settings, scoreboard designs, and perspectives to enhance the development and evaluation of algorithms for accurate score extraction from scoreboard images.

3.2. Image Preprocessing

In the initial step of image preprocessing, the swimming board image is loaded using the OpenCV library. The primary goal is to enhance the subsequent OCR process by simplifying the image’s structure. Grayscale conversion is applied to eliminate color complexities, as the focus is on textual information. Optional thresholding techniques are considered to create a binary image, accentuating text while minimizing background noise. Additionally, denoising is employed to further enhance the clarity of the text on the image. These preprocessing steps collectively lay the groundwork for accurate OCR by optimizing the input image.

3.3. OCR Text Extraction

Following image preprocessing, the PyTesseract library is employed for Optical Character Recognition (OCR). The OCR process involves interpreting the text present on the swimming board image. It is crucial to configure the OCR settings appropriately, specifying the OCR Engine Mode (OEM) and Page Segmentation Mode (PSM). These configurations are selected based on the specific characteristics of the text in the swimming race result boards. PyTesseract translates the visual information in the image into machine-readable text, forming the basis for subsequent analysis.

3.4. Post Processing

Post-processing steps are crucial for refining and organizing the extracted text into meaningful race results. The extracted text is split into relevant components based on the expected structure of the race results. Following this, custom validation checks are implemented to ensure that the extracted information aligns with the expected format of swimming race results. Any errors or inconsistencies introduced during the OCR process are addressed in this step. Post-processing thus serves as a critical phase for ensuring the accuracy and reliability of the extracted race results.

3.5. Hybrid Approach: PyTesseract OCR + CNN-Based Score Extraction

Our streamlined approach starts with dataset curation, collecting diverse annotated scoreboards from swimming events. PyTesseract OCR extracts initial text, focusing on visible scores and details. Preprocessing involves normalizing and resizing images for consistency and model generalization. VGG16, a pre-trained CNN, then captures visual patterns associated with scores. A custom score layer is added that enhances the model's ability to predict numerical scores. Validation optimizes hyperparameters for robust performance. Real-time deployment prioritizes CNN predictions, augmented by consolidation and confidence scoring for improved accuracy. Continuous learning, through periodic updates, ensures adaptability to evolving scoreboards and presentation styles. This efficient hybrid method integrates OCR with CNN for automated score extraction in dynamic sports events.

4. Results and Analysis

In our research endeavor, SSN-MLRG-TEAM2 devised a hybrid PyTesseract OCR + CNN approach, which yielded promising results for score extraction in swimming events. The overall accuracy reached 88.05%, showcasing a notable improvement over individual PyTesseract OCR or CNN usage.

Comparative analysis revealed our hybrid approach's unique advantages—adaptability to diverse sports, real-time processing, and improved accuracy. Future research will explore extensions to other sports and domains, emphasizing responsible AI practices. Overall, our results underscore the potential of the hybrid PyTesseract OCR + CNN approach in automating sports analytics.

5. Discussion and Outlook

Our methodology for swimming race result extraction has exhibited a commendable accuracy of 88.05% in interpreting diverse swimming board images, providing a robust foundation for automated sports analytics. Despite successes, challenges such as variations in text layout and font styles persist, warranting further refinement in OCR configurations and potential integration of machine learning techniques. Looking forward, the broader applicability of our approach extends beyond swimming competitions to other sports with similar visual data challenges. Future research avenues include exploring real-time data streaming integration and developing a more comprehensive sports analytics framework. Ethical considerations, particularly regarding data privacy and bias mitigation, remain crucial focal points. As technological advancements continue, our work serves as a stepping stone for ongoing developments in automated sports analytics, presenting valuable insights and avenues for future exploration and improvement.

References

- [1] S. Minaee, Y. Wang, Text extraction from texture images using masked signal decomposition, in: 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2017, pp. 1210–1214.
- [2] A. Erades, P. Martin, R. V. B. Mansencal, R. Péteri, J. Morlier, S. Duffner, J. Benois-Pineau, Sportsvideo: A multimedia dataset for event and position detection in table tennis and swimming, in: Working Notes Proceedings of the MediaEval 2023 Workshop, Amsterdam, The Netherlands and Online and Online, 1-2 February 2024, CEUR Workshop Proceedings, CEUR-WS.org, 2023.
- [3] R. Chatterjee, A. Mondal, Effects of different filters on text extractions from videos using tesseract, 2021. doi:10.13140/RG.2.2.36025.08804.
- [4] R. Manjunath, S. Guruswamy, A Survey on Text Detection from Document Images, 2020, pp. 961–972. doi:10.1007/978-981-15-0633-8_98.
- [5] C. Misra, P. K. Swain, J. K. Mantri, Text extraction and recognition from image using neural network, International journal of computer applications 40 (2012) 13–19.
- [6] N.-M. Chidiac, P. Damien, C. Yaacoub, A robust algorithm for text extraction from images, in: 2016 39th International Conference on Telecommunications and Signal Processing (TSP), IEEE, 2016, pp. 493–497.
- [7] C. Sumathi, T. Santhanam, G. Devi, A survey on various approaches of text extraction in images, International Journal of Computer Science and Engineering Survey 3 (2012).
- [8] D. Ghai, D. Gera, N. Jain, A new approach to extract text from images based on dwt and k-means clustering, International Journal of Computational Intelligence Systems 9 (2016) 900–916. doi:10.1080/18756891.2016.1237189.
- [9] D. Ghai, N. Jain, Comparative analysis of multi-scale wavelet decomposition and k-means clustering based text extraction, Wireless Personal Communications 109 (2019) 1–36. doi:10.1007/s11277-019-06574-w.
- [10] D. Ghai, N. Jain, Comparison of Different Text Extraction Techniques for Complex Color Images, 2022, pp. 139–160. doi:10.1002/9781119861850.ch9.