Izvirni znanstveni članek

TEHNIKA - računalniške tehnologije

Datum prejema: 12. december 2018

ANALI PAZU 9/ 2019/ 1-2: 30-36

Video annotation with metadata: A case study of soccer game video annotation with players

Označevanje videa z metapodatki: analiza primera označevanja video posnetka nogometne tekme s podatki o igralcih

Danilo Zimšek*, Luka Banfi, Mirjam Sepesy Maučec

Faculty of Electrical Engineering and Computer Science , University of Maribor, Koroška c. 46, 2000 Maribor

E-Mails: danilo.zimsek@um.si, luka.banfi@student.um.si, mirjam.sepesy@um.si

* Avtor za korespondenco;

Abstract: In recent years, the use of convolutional neural networks for video processing has become very attractive. The reason lies in the computational power for data processing which is available today. There are many well-defined research areas where neural networks have brought higher reliability than other conventional approaches; for example, traffic sign recognition and isolated number recognition. In this paper, we will describe the architecture and the implementation of the process of soccer game annotation. The game is annotated with data about players. The technology of convolutional neural networks is used for number recognition. The process runs in real-time on a streaming video. Content enriched with metadata is given to the user in parallel with the real-time video. In the paper, we will describe in some detail the following modules: Image binarization, shot localization, the selection and recognition of numbers on players` jerseys.

Key words: video annotation; convolutional neural networks; football

Povzetek: V zadnjih letih je uporaba konvolucijskih nevronskih mrež pri procesiranju video vsebin doživela velik razcvet. Razlog je predvsem v povečani računski moči, ki je danes na voljo za procesiranje podatkov. Zlasti se na tem področju v zadnjem času veliko uporabljajo grafične kartice, ki z velikim številom grafičnih jeder izvajajo operacije konvolucije v paraliziranem načinu. Te operacije pa so pri tovrstnih procesih ključnega pomena. Na posameznih dobro definiranih področjih, kot so na primer razpoznava obrazov, razpoznava prometnih znakov in izoliranih števil, zagotavlja tehnologija nevronskih mrež visoko zanesljivost, ki je primerljiva z zanesljivostjo izmerjeno pri človeku ali jo celo presega. Tehnologija se uporablja na vse več področjih, kar odpira veliko možnosti implementacij novih storitev, tudi na področju IP televizije s stališča obogatenih vsebin. Obogatene vsebine lahko uporabniku izboljšajo izkušnjo ogleda vsebin, zlasti ko gre za informacijo, ki uporabnika zanima in je ta podana v za vsebino videa ključnih časovnih odsekih. Poleg navedenega takšna dodana informacija odpira možnosti za implementacijo dodatnih storitev, ki so vezane na interakcijo uporabnika z vsebino in lahko vključujejo povezave tudi do socialnih omrežij ter omogočajo uporabniku, da izrazi mnenje o vsebini. Hkrati takšne storitve odpirajo nove možnosti ogleda in komentiranja vsebin v mikro socialnih omrežjih ter tako omogočijo doživljanje izkušnje skupinskega ogleda neke vsebine na daljavo. S tem lahko takšna storitev izboljša socialno vključenost posameznika, še posebej pri gibalno oviranih in starejših osebah. Storitev se lahko ponuja z uporabo vmesnikov, kot je televizija, ki je široko razširjena in uporabnikom, tudi starejšim, dobro poznana.

V prispevku bomo opisali primer označevanja videa nogometne tekme. Predstavili bomo arhitekturo in implementacijo procesa označevanja nogometne tekme s podatki o imenih igralcev z uporabo tehnologije nevronskih mrež. Proces se izvaja realno-časovno nad videom razpršene oddaje. Obogatena vsebina je ponujena uporabniku na ogled sočasno z video posnetkom, pri čemer se dodane informacije prikažejo samo v ključnih trenutkih na stalnem mestu ter imajo tako kar se da majhen vpliv na video z osnovno vsebino. V prispevku bomo opisali celoten proces, ki je sestavljen iz večih korakov. Najprej se izvede binarizacija slike za katero so potrebni vhodni podatki o barvah dresov ekip, ki jih lahko pridobimo iz zunanjega aplikacijskega programskega vmesnika. Sledi lokalizacija dresov nogometašev, ki za nadaljnje procesirtanje iz slike izloči področje, na katerem se dresi nahajajo. Sledi proces izločanja in razpoznavanja števil, zapisanih na dresih. Nato se z uporabo zunanjega aplikacijskega programskega vmesnika pridobijo informacije o igralcih, katerim pripadajo izločene številke dresov. Opisani sistem smo evalvirali na testnem setu, ki je vključeval pretežno posnetke igralcev od blizu. Analizirali smo 14 nogometnih tekem, pri čemer smo pri vsaki tekmi analizirali eno od obeh moštev na izbranem številu okvirjev. V prispevku podamo tudi rezultate evalvacije sistema. Z opisanim sistemom smo dosegli 85,60% natančnost označevanja. V raziskavi smo se posvetili tudi časovni analizi postopka označevanja videa. Vsi koraki celotnega postopka skupaj trajajo 0,825 sekunde, kar dokazuje uporabnost sistema v realnem času.

V prihodnosti nameravamo opisani sistem integrirati v IMS omrežje in ga izpostaviti preskusu s stopnjevanim obremenjevanjem, v katerem se lahko pokažejo pomanjkljivosti sistema, ki jih v nadaljevanju še lahko izboljšamo.

.

Ključne besede: označevanje videa; konvolucijske nevronske mreže; nogomet

1. Introduction

Soccer is one of the most popular sports worldwide, especially in Europe and South Africa, and in other regions its popularity is still increasing. Based on historical data, it is expected that its popularity will continue to grow worldwide. Because of its popularity, it is an interesting type of content for telecommunication providers to provision new services for enhanced video viewing. Services such as content annotation and group video viewing using video conference service can enable users to view a soccer game interactively in a micro social network, where each user is able, at some critical moments depending on content, to interact with content in a way that emotional symbols can be associated with player actions, which is an additional way of expressing an opinion regarding player action. It is, therefore, an interesting topic for researchers to find ways to annotate broadcast video data automatically and deliver it to the user.

The rest of the work is structured as follows: In the next section, related work in this area is presented, in section three we describe the overall architecture of the system: Methods used for shirt localization, number identification and number recognition. The annotation process is also presented. The evaluation of results is described in section four. In the last section, ideas are given for future work.

2. Related work

Automatic annotation of broadcast sport videos is an interesting area for researchers. Some approaches deal with ball and field detection [1], and many papers focus on player detection and tracking [2, 3, 4, 7, 10]. The approach in [2] is a semi automatic approach for tracking and detecting an important player in a shot. It requires operator assistance, which adds time delay. In [3], player regions are detected using Multilayer Perceptron Neural Network as the classifier for player and non-player regions. The approach in [4] uses dominant colour region detection for player detection. HOG features combined with an SVM classifier are used for player detection in [7]. Approaches like [10] use jersey colours for player localization.

Some researches are using face recognition techniques for player identification [6], whereas in [5, 8, 9] number recognition techniques are used for the same purpose. In [5], player localization is performed using region adjacency graph and picture trees; afterwards, optical character recognition is used for number recognition. Paper [8] describes two different approaches for number recognition, both of them using convolutional neural networks. This approach uses its own database and augmentation technique to produce a training dataset. The approach in [9] uses Generic Fourier descriptor for the number recognition process.

Our approach differs from the approaches used in literature in the following manner:

- It is a hybrid approach using hand-crafted criteria for shirt localization and convolutional neural network for number recognition;
- It doesn't require an annotated database of football matches to be trained on;
- It can be used for real-time applications and services.

3. System design

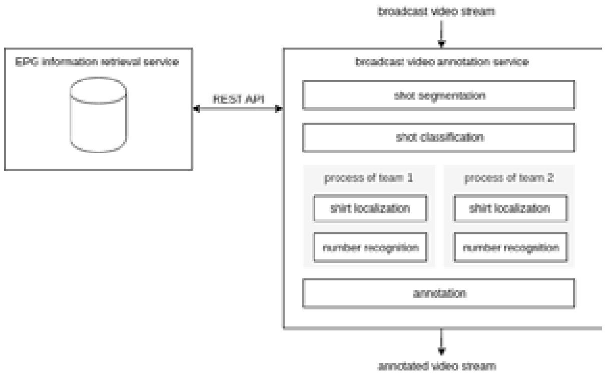


Figure 1: The architecture of the proposed system

The annotation service that we developed is intended for online use on broadcast video. Because of that consideration, we use techniques which add a small portion of delay. This also requires fully automatic operation of the described system. Its architecture is presented in Figure 1.

The input of the proposed system is a broadcast video stream. Typically, in telecommunication systems for video broadcasting, videos are encoded using MPEG-2, MPEG-4 or H264 codecs. The frame-rate of a video stream is commonly 25 frames per second. There are some systems using higher frame rates, but these are not yet largely implemented. Broadcast video content for sports events is nowadays delivered in HD, Full HD or 4K resolution. The resolution of the video is not critical for our system, as it also performs well using SD resolution. Our proposed service uses the broadcast video stream as its input. It decodes the encoded video and does not start video processing after every decoded frame. The process needs additional input about player names, their numbers, team names and jersey colour information.

The additional information required by the service is gathered using the EPG information retrieval service. The service updates its cached EPG information file daily in xml format from a telecommunication provider. It provides a REST API interface for other services to get information about channel programmes at a specified time. For the purpose of the broadcast video annotation service it parses data about soccer video broadcasts. It extracts soccer match broadcasts using keywords in the programme title. It gathers team names from it,. Using external open access API it also gathers information about player names and jersey numbers and team jersey colours. The information is stored in an EPG information retrieval service database, and offered to broadcast video annotation services through REST API.

The output of the system is an annotated video with a list of player names displayed on the side of the screen. To achieve this output, each processing job handles shot segmentation and classification for every decoded video frame. For shots which are classified as close-up shots, shirt localization and, later, number recognition processes, are executed. Two instances of these processes are executed simultaneously, each for one team, so each recognises the numbers on the jerseys of a team and provides it to the annotation process. The process adds the graphical elements to the input video frame based on information from the EPG information retrieval service and streams it to the end user equipment.

The presented system is implemented in Python programming language using scipy, numpy and opencv modules. In the following subsections, each module is explained in some detail.

3.1. Shot segmentation

In many applications of video processing, the first process is intended to break the video stream into shots. In literature, there are many different approaches for shot segmentation [11, 13, 14, 15, 16]. The approach described in [11] is a hybrid shot boundary detection method, which integrates a High-Level Fuzzy Petri Net (HLFPN) model with key-point matching. First, the HLFPN model, with histogram difference, is executed on consecutive frames, then the speeded-up robust features` algorithm is used to detect gradual transitions and eliminate false shots, based on the assumption of the HLFPN model. The method needs 5 consecutive video frames for the analysis of a video. When dealing with a video frame-rate of 25 frames per second, this adds an additional time delay of 0.2 seconds, which is not ideal for real-time IPTV system implementations.

Paper [13] describes the fast shot boundary detection method, which first uses adaptive thresholds to predict shot boundaries and gradual transition lengths. At this point, it discards most not non-boundary frames. The colour histogram in hue-saturation-value colour space is calculated for each candidate segment, which forms a frame-feature matrix on which video shot segmentation is performed to reduce the feature dimension. Based on the gathered metric, shot transitions are identified using the pattern matching method. The approach is not suitable for real-time applications requiring two stage operations on a video.

There are approaches, like in [14, 15], which rely only on two consecutive frames to detect a shot boundary. The authors in [14] describe a three-stage approach based on the Multilevel Difference of Colour Histograms. In the first stage, two self-adapted thresholds are used to detect candidate boundaries. In the second stage, noise filtering takes place, which uses the local maximum difference of the MDCH, generated by shot boundaries. In the third stage, a voting mechanism makes the final detection.

For the purpose of shot segmentation, we used the method described in [15]. It does not require threshold selection. Its criteria are based on colour coherence change, and findings are then combined with a machine learning technique. The shot boundary detection process segments the video stream into a number of shots. When a new shot is detected, the shot classification process starts.

3.2. Shot classification

For shot classification we adopted the approach from [11]. The approach was adopted in a way to classify only in-field close-up shots. Our approach first eliminates out-field frames. In the next stage, three features are extracted, which are based on a number of connected components and shirt colour percent. In the third stage, an SVM (Support Vector Machine) classifier is employed to extract close-up shots. In the continuation of the process, we focus only on close-up shots. The decision for using only close-up shots is based on the assumption that close-up shots occur right after some important event happens. Shot classification is triggered only on the first five frames after a shot boundary is detected, otherwise this step is skipped.

3.3. Shirt localization

When the close-up shot appears, we want to recognise which player is on the shot. The first step in this process is shirt localization. It is the process of identifying pixels of a close-up shot in a video frame which belong to the player. The process of shirt localization consists of image binarization, blob removal, and connected component extraction.

For minimising the false positive detections, we propose the following object adequacy decision. We can make the assumption that each team`s members have shirts of a certain colour. The video annotation service gets these data from the information retrieval service. Based on the gathered colour specified in the HSV colour model, the colour range is specified using the following formula:

$$\mathrm{Hue}_{min} = \mathrm{Hue}_{gathered} - 5°$$
$$\mathrm{Hue}_{max} = \mathrm{Hue}_{gathered} + 5°$$
$$\mathrm{Saturation}_{min} = \mathrm{Saturation}_{gathered} - 50\%$$

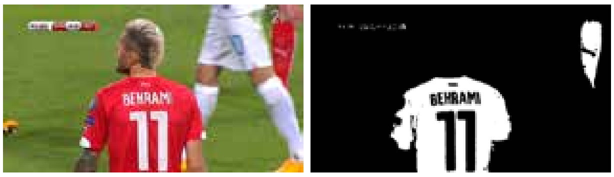


Figure 2: Video frame before (on the left) and after binarization (on the right)

$$\mathrm{Hue}_{min} = \mathrm{Hue}_{gathered} - 5°$$
$$\mathrm{Hue}_{max} = \mathrm{Hue}_{gathered} + 5°$$
$$\mathrm{Saturation}_{min} = \mathrm{Saturation}_{gathered} - 50\%$$

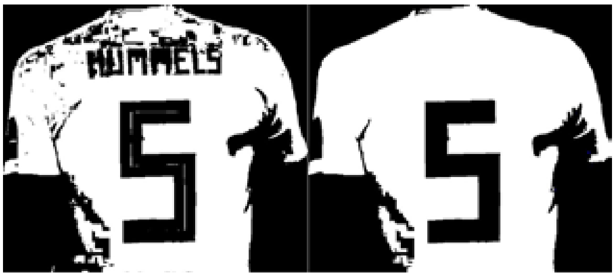


Figure 3: Player jersey area before and after blob removal.

Where Hue_{min} and Hue_{max} specify the minimum and maximum values of a Hue parameter in the HSV colour space. Intervals for other parameters of the HSV colour space are specified accordingly. Two colour ranges are calculated, one for each team. The specified colour range of a team has to include the majority of colours found on players` shirts. The process of frame binarization is run for both teams separately. Figure 2 shows an example of an input video frame of a close-up shot and its binarization using specified colour ranges.

Connected components` regions are labelled after binarization. Two pixels are connected when they are neighbours and have the same value. They can be neighbours in either horizontal, vertical or diagonal directions. After the operation, regions are extracted presenting object candidates.

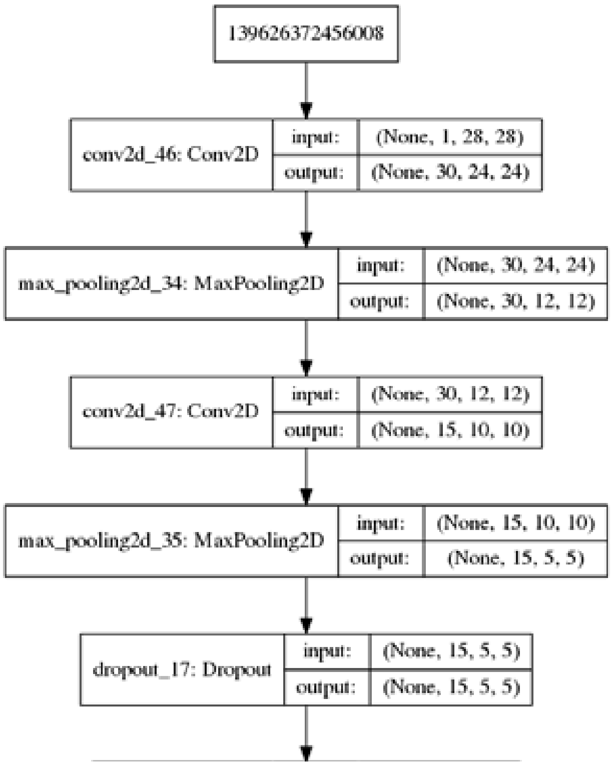
Blob removal is performed for the purpose of eliminating small object candidates. Empirically we set the threshold of area of the largest object candidate, which presents an area of to 5% of a player's jersey. Figure 3 shows an example of a player's jersey area before and after blob removal.

In recent years, new trends have appeared in teams` clothing. Recent work in that domain [9] assumed that the colour of a player`s shirt is different from the colour of pants and socks. In recent years, in the case of some teams, the colour of shirts and pants is the same, which resulted in many false positive detections. To eliminate those detections, we measured object candidates regarding adequacy criteria: Minimum object width and height and allowed range of aspect ratio. Based on [16], we conducted criteria for object selection based on aspect ratio. The allowed aspect ratio between length and height for a desired object, has to be from 2:1 to 1.5:1. Criteria for minimum object width and height are considered relative to the overall extracted area. Object height and width have to be less than 75% of the height of extracted area. Using those considerations, all noisy objects are eliminated from object candidates.

3.4. Number recognition

A neural network is used for the number recognition process. Because of the non-existent freely available large dataset of player shirts with annotated number, and because of the fact that a pre-trained network on a small dataset can lead to overfitting, we trained our network for number recognition using an existing MNIST [18]. The dataset consists of 10 classes; every class represents one digit. The use of this dataset brings the limitation of classifying only single digits, while numbers on players shirts can consist of two digits. In this case, both of the numbers are fed into the network separately, and, after the classification process, they are combined into a two digit number.

Usually approaches which apply a convolutional neural network for object recognition tasks, consist of a training stage and a runtime stage. In the training stage, weights of the neural network are adopted to a specific domain, which is specified by the training dataset. Because the different dataset is used, the network is not optimised to that domain, but can still perform well enough to be used in the described service. Using an existing dataset can, on the other hand, reduce the implementation time of an annotation service drastically.



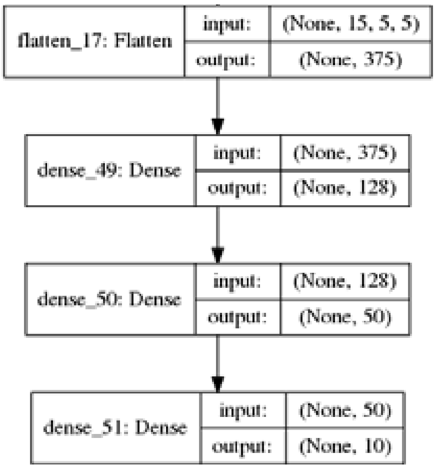


Figure 4: The architecture of a convolutional neural network.

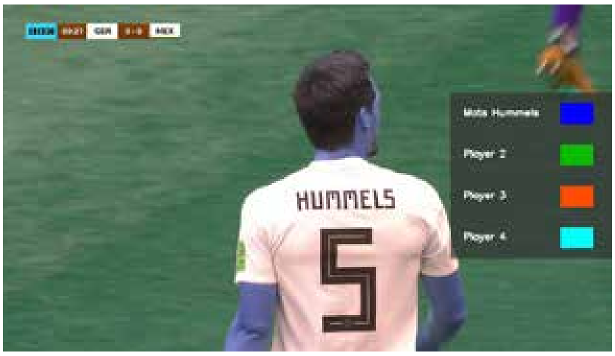


Figure 5: An example of a video frame with additional data

A convolutional neural network consists typically of an input layer, which holds the raw pixel values of the image, in our case an image of width 28, height 28, and one grayscale channel. The convolutional layer computes the outputs of neurons that are connected to local regions in the input layer. Each neuron will compute the dot product between small input regions and their weights. The max pooling layer performs a down sampling operation along the spatial dimensions of width and height. The flatten layer flattens the feature map from the previous layer. The vector encodes the features, which are distilled throughout the previous steps. With this approach, we prepare input data for the dense layer. Neurons in the dense layer have full connections to all activations in the previous layer. They classify input features into a predefined number of classes.

In our approach, a dropout layer is used to eliminate co-dependency among neurons during training, which curbs the individual power of each neuron, leading to over-fitting of training data. During the runtime phase, we use all the activations, but reduce them to account for the missing activations during training.

The architecture of the convolutional neural network used in the system is presented in Figure 4.

3.5. Video annotation

An example of the data that are presented to the user is shown in Figure 5. Video frames of close-up shots are annotated using an original video track, and graphical elements are added to the video track consequently. For the implementation of user interface the [17] is used, which enables creation of a highly adjustable user interface. The user interface adapts itself based on the number of recognised players. It shows the information of player names for the recognised numbers. The information about player names is displayed for a predefined time, to enable users to interact with the content. The interaction depends on the use case in which the video annotation service is used. In our example, users are able to interact with the content by adding emotional signs to the players to express their opinions about player action.

4. Evaluation

In this section, we describe the evaluation of the previously described processing steps, being part of our system. Shot segmentation and classification processes are described and evaluated well in reference literature. The precision reaches up to 95%. Because of the good precision of video shot segmentation and classification methods, these two processes could not have a negative impact on the overall performance of the system.

4.1 Test data preparation

Teams in the match	Analysed team	Number of frames
Atletico Madrid - Real Madrid	Atletico Madrid	3
Colombia - Japan	Colombia	12
Colombia - Japan	Japan	8
Germany - Mexico	Germany	3
Poland - Senegal	Poland	5
Real Madrid - Liverpool	Liverpool	1
Real Madrid - Liverpool	Real Madrid	2
Sweden - South Korea	Sweden	4
Estonia - Slovenia	Estonia	8
Estonia - Slovenia	Slovenia	4
Lithuania - Slovenia	Lithuania	4
Lithuania - Slovenia	Slovenia	6
Slovenia – San Marino	Slovenia	8
Slovenia – San Marino	San Marino	6
Slovenia - Switzerland	Slovenia	10
Slovenia - Switzerland	Switzerland	5

Table 1: The structure of the evaluation dataset

For evaluation purposes, a small test dataset was collected of close-up shots. Figure 6 presents some examples of video frames from the evaluation dataset. Data consist of 14 categories. In each category, there are frames from one football match. In each category, there are 10 to 15 video frames of close-up shots. Frames of close-up video shots were chosen randomly from videos, which were classified automatically as closeup shots. From each shot, one frame was chosen randomly. Table 1 presents the structure of the evaluation dataset used for evaluation of number extraction and number recognition processes.

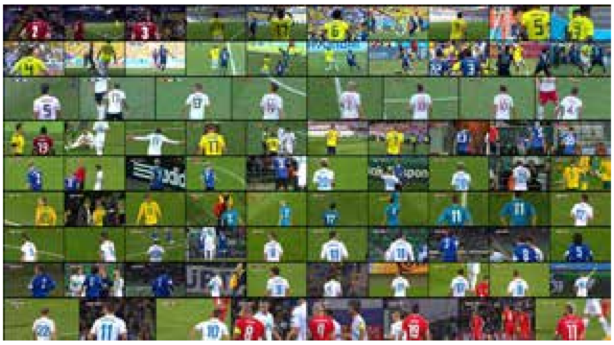


Figure 6: Randomly selected frames from the evaluation dataset

4.2 Evaluation of shirt localization and number extraction

We used the prepared dataset for the purpose of shirt localization and number extraction evaluation. As input we used close-up shots, and annotated them manually with information about numbers located in the frame. We ran processes of shirt localization and number extraction on those frames and calculated accuracy results for each category in the dataset. Table 2 shows the results of the evaluation processes. Using the previously described techniques for frame binarization, connected neighbours` segmentation and additional criteria for number extraction, our approach performed well, and did not add significant error to the overall performance of the system. We noted some number extraction problems on jerseys with stains. We also noted that using the HSV colour model instead of RGB, which was used in our previous work, the accuracy of both processes is significantly better.

Teams in the match	Analyzed team	Number of extracted digits	Number of digits in the reference	Accuracy [%]
Atletico Madrid - Real Madrid	Atletico Madrid	3	3	100.00
Colombia - Japan	Colombia	11	17	64.71
Colombia - Japan	Japan	8	13	64.54
Germany - Mexico	Germany	5	5	100.00
Poland - Senegal	Poland	8	8	100.00
Real Madrid - Liverpool	Liverpool	2	2	100.00
Real Madrid - Liverpool	Real Madrid	3	4	75.00
Sweden - South Korea	Sweden	5	5	100.00
Estonia - Slovenia	Estonia	8	10	80.00
Estonia - Slovenia	Slovenia	7	8	87.50
Lithuania - Slovenia	Lithuania	5	6	83.33
Lithuania - Slovenia	Slovenia	9	9	100.00
Slovenia – San Marino	Slovenia	11	13	84.62
Slovenia – San Marino	San Marino	6	6	100.00
Slovenia - Switzerland	Slovenia	15	16	93.75
Slovenia - Switzerland	Switzerland	7	7	100.00
Overall		113	132	85.60

Table 2: Evaluation results of the shirt localization and number extraction processes.

4.3 Evaluation of number recognition

Digit	Number of successful recognitions	Number of failed recognitions	Accuracy[%]
0	1	4	20.00
1	20	6	76.92
2	15	2	88.23
3	8	1	88.88
4	5	1	83.33
5	17	0	100.00
6	1	2	33.33
7	10	0	100.00
8	6	1	85.71
9	1	3	25.00
Overall	84	20	80.77

Table 3: Evaluation results of the number recognition process.

For the evaluation of number recognition process, we used an annotated evaluation dataset. We ran the recognition process on all images in the dataset and compared results with reference numbers. The accuracy of the number recognition process is given in Table 3.

4.4 Runtime analysis

For the runtime analysis, we ran each individual processes five times on a virtualised server with 2 CPU cores with frequency of 2 GHz. Table 4 shows the average time of execution for each process, and the overall time needed to execute all the processes. There is a need to run two processes of shirt localization, number extraction and number recognition in parallel, one for each team. During evaluation, we evaluated the execution time of a video frame with two players that belong to different teams, each having a two digit number on the shirt. Close-up shots with two players belonging to the same team are, in real-world videos, very rare. If we prosume, that the framerate of a broadcast video is 25 frames per second, the process will add a time delay of 0,825 to video stream. To achieve real time operation, 20 sequential frames of input video need to be processed in parallel.

Process	Time in seconds
Shirt localization	0.091
Number extraction	0.216
Number recognition	0.545
Overall	0,825

Table 4: Runtime analysis of all processes

5. Conclusion

In future work we plan to include the described annotation service in the shown example use case. Then we will evaluate the service as a whole from the usability aspect.

We plan to merge the implemented service to an existing IMS network, described in [16], and to expose it to the stress test. Based on the results of the stress test, we can note the key bottlenecks and improve the current implementation of the annotation service.

Conflicts of interests

Described work was not published in any other paper. Publication of this article is confirmed by all authors.

Literature

1. Zhang, Y. ; Lu, H. ; Xu, C. Collaborate ball and player trajectory extraction in broadcast soccer video. International Conference on Pattern Recognition 2008, 1-4.
2. Kang, M.S. ; Lee, J.W. ; Kang, S.J. ; Lim, Y.C. Semi-automatic player detection for real-time broadcasting systems. IEEE International Conference on Consumer Electronics (ICCE) 2017, 262-264.
3. Heydari, M. ; Moghadam A:M.E. An MLP-based player detection and tracking in broadcast soccer video. International Conference of Robotics and Artificial Intelligence 2012, 195-199.
4. Nunez, J.R. ; Facon, J. ; Souza Brito, A. Soccer video segmentation: Referee and player detection. International Conference on Systems, Signals and Image Processing 2008, 279-282
5. Andrade, E.L. ; Khan, E. ; Woods, J.C. ; Ghanbari, M. Player identification in interactive sport scenes using region space analysis prior information and number recognition. International Conference on Visual Information Engineering VIE 2003, 57-60
6. Ballan, L. ; Bertini, M. ; Del Bimbo, A. ; Nunziati, W. Soccer Players Identification Based on Visual Local Features. Proc. of ACM International Conference on Image and Video Retrieval (CIVR) 2007.
7. Mackowiak, S. ; Konieczny, J. ; Kurc, M. ; Maćkowiak, P. ; Maćkowiak, P. Football Player Detection in Video Broadcast. Computer Vision and Graphics - International Conference 2010.
8. Gerke, S. ; Müller, K. ; Schäfer, R. Soccer Jersey Number Recognition Using Convolutional Neural Networks. IEEE International Conference on Computer Vision Workshop 2015, 734-741.
9. Frejlichowski, D. Identification of Football Players based on Generic Fourier Descriptor Applied for the Recognition of Numbers. Image Processing & Communications, 21, 13-18.
10. Frejlichowski, D. A Method for Data Extraction from Video Sequences for Automatic Identification of Football Players Based on Their Numbers. International Conference on Image Analysis and Processing 2011, 356-364.
11. Bagheri-Khaligh, A. ; Raziperchikolaei R. ; Ebrahimi Moghaddam, M. A new method for shot classification in soccer sports video based on SVM classifier. Southwest Symposium on Image Analysis and Interpretation 2012, 109-112.
12. Shen, R.K. ; Lin, Y.N. ; Tony Tong-Ying Juang ; Victor R. L. Shen ; Soo Yong Lim. Automatic Detection of Video Shot Boundary in Social Media Using a Hybrid Approach of HLFPN and Keypoint Matching. IEEE Transactions on Computational Social Systems 2017, 5, 210-219.
13. Lu, Z.M. ; Shi, Y. Fast Video Shot Boundary Detection Based on SVD and Pattern Matching. IEEE Transactions on Image Processing 2013, 22, 5136-5145.
14. Li, Z. ; Liu, X. ; Zhang, S. Shot Boundary Detection based on Multilevel Difference of Colour Histograms. International Conference on Multimedia and Image Processing 2018, 15-22.
15. Tsamoura, E. ; Vasileios Mezaris ; Ioannis Kompatsiaris. Gradual transition detection using color coherence and other criteria in a video shot meta-segmentation framework. IEEE International Conference on Image Processing 2008, 45-48.
16. Mlakar, I. ; Zimšek, D. ; Kacic, Z. ; Rojc, M. A Novel IMS based UMB-SmartTV system for Integrating Multimodal Technologies. International Journal of Computers and Communications 2014, 8, 7-16.
17. Bradski, G. The OpenCV Library, Dr. Dobb's Journal of Software Tools, 2008.
18. LeCun, Y. ; C. Cortes. MNIST handwritten digit database. http://yann.lecun.com/exdb/mnist/, 2010