ISSN 2395-8618

Sports Video Anonymisation and Accuracy Prediction

César Marrades Cortés, Keith Quille, Jelena Vasić, and Seán McHugh

***Abstract*—The anonymisation of people featuring in videos is required in many contexts. One of these is the physical education state exam in Ireland, where secondary school students are assessed as prescribed by the National Council for Curriculum and Assessment (NCCA), with the use of video recordings among other tools. For reasons of GDPR, all the video material presented for grading must not reveal the identity of the students. The work presented here was undertaken on consultation with the NCCA regarding their needs in this area and comprised two distinct tasks: (1) the implementation and testing of an anonymiser program in C#, supported by libraries Accord and EmguCV, each defining a different variant of the program; (2) the use of machine learning predictive models in Weka to investigate which of various factors (such as camera quality, camera angle, sport) affect the anonymisation program's performance on sports videos. One hundred video inputs, resulting in 200 outputs (one for each of the two libraries, per input), were used and the best anonymisation success prediction model had an accuracy of 94% and a specificity of 95.2%. This work forms a base upon which a full automated video anonymisation system could be built, most importantly generating knowledge on what measures can be taken towards the optimisation of video anonymisation performance.**

***Index Terms*—Video anonymisation, machine learning.**

I. Introduction

THE WORK presented here addresses a specific image processing problem presented by the Irish National Council for Curriculum and Assessment (NCCA). Physical education exams are often recorded by students in video format and then evaluated by the NCCA. The project aimed to develop a face recognition system that demonstrates the possibility of automatically anonymising the videos prior to evaluation, by means of using face recognition libraries to detect students' faces and of blurring them. This is an important concern for the NCCA as identity preservation is required by the current European General Data Protection Rules (GDPR). It is also a way of avoiding bias during content evaluation.

The paper presents findings and recommendations in two distinct, if related, areas: (1) the development and integration of a software system prototype, with emphasis on facial recognition libraries used and (2) a data-analytical investigation of the capabilities and limitations of those libraries, including what features of video material make it most amenable to processing for anonymisation.

The libraries used in the work at hand have been employed by other researchers. In her master's thesis [1], Suad Hajr Ahmed Omar discusses the general limitations of the Emugcv library in terms of an existing optimal camera-to-subject distance of 1 to 3 feet, required illumination and problems with recognition when the face is partially covered, e.g. with sunglasses. Team TBD [2] found that in a situation with time constraints and related performance requirements the Emgucv library presented as the most stable and complete solution. Mohammed [3] also used the Emugcv library, while recognising that several factors can limit its usability, for example unusual poses and facial expressions or occlusion by sunglasses or make-up.

II. Methodology

The prototype development and performance analysis phases of the work are logically and operationally separate, each self-contained with respect to methodology.

A. Prototype Development

The main functional requirement placed on the prototype was that it accept a video clip of a sporting activity as input and that it output a video based on the original clip, in which the faces of any people are rendered unrecognisable.

The prototype was created as a Microsoft Windows executable file, developed using C# with the Emgucv and Accord face recognition libraries. These libraries were chosen based on favourable reports in literature on similar uses and availability of documentation (Accord), in combination with their availability at no monetary cost. As the focus of the project was to generically investigate automated anonymisation for a particular type of video rather than provide an exhaustive review of implementation possibilities, it was sufficient to use these two libraries. Other libraries considered were Microsoft Cognitive Services, Censor Face and OpenCV.

The prototype was designed to generate two outputs, one produced by each of the face recognition libraries, for every received input video. It transforms all the video files found in a specified input folder.

Manuscript received on June 12, 2019, accepted for publication on September 7, 2019, published on December 30, 2019.

The authors are with Technological University Dublin, Ireland (e-mail: cmarrades@hotmail.com, {keith.quille, jelena.vasic, sean.mchugh}@tudublin.ie).

ISSN 2395-8618

TABLE I
SPORTS: DETAILS OF SPORTS THAT FEATURE IN THE SET OF INPUT VIDEOS, TOGETHER WITH CO-OCCURRING CONDITIONS (OTHER VARIABLE VALUES).

Sport	**Conditions**
Basketball	indoors and outdoors, dynamic cameras, medium and far distances
Karate, Judo	indoors, multiple backgrounds, multiple angles, tendency to close to medium distances, head protectors
Boxing	indoors, multiple angles, head protectors, referee, lose, medium and far ranges
Hurling	tendency towards far distances, high angles, with multiple people
Gaelic Football	tendency towards far distances, medium and high angles, with multiple people
Tennis	multiple angles, mixed cam movement types, high angles
Table Tennis	medium and far distances
Soccer, Soccer Interior	multiple angles, mixed cam movement types, medium, high angles
High Jump, High Jump Pole	medium angles from close and medium distances different camera types
Weight Lifting	static cameras from close and medium distances and medium angles

B. Analysis

The data used in the analysis phase comprised input and output videos and information about the videos, which was extracted in a manual process.

Videos of various sports activities, in mp4 format, were downloaded from YouTube to serve as inputs to the anonymiser prototype. They were chosen so as to provide a variety of different scenarios covering the most common cases handled by the NCCA when evaluating students in exams. Longer videos require longer processing time, therefore shorter videos had a preference over longer ones.

The attributes of both the input and the output videos were extracted 'manually', by observation and noting of properties on the part of a researcher. This manual processing of the 300 videos took about 7 days. The attributes extracted were the following:

– Sport (sports defined in Table I)
– Indoors/outdoors
– Light conditions
– Whether the video was recorded using a phone or a dedicated camera
– Video quality
– Camera movement e.g. static or dynamic
– Camera recording angle e.g. low, medium or high
– Camera distance e.g. medium or far
– Number of people in the video (ranging from 1 to more than 10)
– Whether people were wearing head protectors
– Field type e.g. court, grass
– Field colour
– Background colour
– Whether there was crowd on the background of the video
– Whether there was a referee
– Ground position: whether people in the video spend a considerable amount of time on the floor e.g. with wrestling or combat sports in general

Table I lists the sports that feature in the input video set. It also shows co-ocurring conditions i.e. other attribute values that tend to pertain to videos of the listed sport. The ten sports have been chosen to cover a wide variety of scenarios i.e. attribute combinations and are not based on a definite list from the NCCA.

The output videos were labelled according to two different scales for measuring the quality of the blurring: a 1-10 scale that rates outputs from worst to best and a binary one, with values of *full* and *partial* for the level of facial blurring. The different values (classes) on these scales are shown in Table II.

The following rules were used when grading outputs using the classifications in Table II:

– *no accuracy*: no faces were recognised or blurred

ISSN 2395-8618

TABLE II
OUTPUT CLASSIFICATION: VALUES AND THEIR DESCRIPTIONS FOR THE TWO TARGET VARIABLES.

10-Value Scale Class	10-Class Description	2-Value Scale Class
1	no accuracy, high noise	Partial
2	no accuracy, noise	Partial
3	extremely low accuracy, noise (any)	Partial
4	low accuracy, high noise	Partial
5	low accuracy, noise	Partial
6	good accuracy, high noise	Partial
7	good accuracy, low noise	Full
8	almost no failures, noise (any)	Full
9	no failures, noise	Full
10	no failures, no noise	Full

– *extremely low accuracy*: almost no faces were recognised or blurred
– *low accuracy*: some of the faces were recognised in some parts of the video
– *good accuracy*: the majority of faces were recognised and blurred
– *almost no failures*: very few faces not recognised or blurred
– *no failures*: all faces were recognised and fully blurred

The Weka tool was used for all the steps required to develop the prediction model i.e. for data pre-processing, attribute selection and the selection of a final model that would be used to predict the quality of the outputs produced by the anonymiser.

III. RESULTS

Explicit measurements were not taken of the anonymiser prototype's performance but the following account provides insight into the relevant orders of magnitude. The other two subsections respectively contain a detailed discussion of the videos and their attributes and of the predictive models that were built in Weka.

A. Running the Prototype

The anonymiser prototype was used to create 200 output videos from 100 videos collected to serve as input. A micro Amazon Web Services (AWS) instance was initially set up to process the 100 inputs. This instance was later scaled to a large instance as the process was using the CPU at maximum capacity. After a few days, a second Intel Core i7 2.60GHz machine was added to speed up the process. Processing all the 100 inputs and producing 200 outputs took around 8 days following these steps.

B. Video Attributes

The input videos were viewed and attribute values, as listed above, recorded for each of them. The output videos were viewed and the blurring quality assessed using the scales presented in Table II. Thus the data set for analysis was completed.

The initial prediction model defined in Weka used the 10-value target variable shown in Table II and performed poorly, with a maximum accuracy of 56%. For this reason a second, binary target variable with values *Full* and *Partial*, denoting 'fully blurred' and 'partially blurred', was defined as shown in Table II. The best accuracy achieved when predicting this new target was 94%. Although no further tests were done, final results indicate that increasing the *Full* threshold to rank 8 would still have high probability of positive results.

Data inspection and observations revealed that videos are more or less evenly divided between indoors and outdoors. Only 5% of the videos are recorded under bad light conditions and there is a high tendency to record these videos with cameras rather than with mobile phones, as only 5% of them are recorded using a phone. There are very few videos where persons are wearing head protections and there is a tendency to keep the same cam distance for the duration of the video. Recordings are evenly divided between having and not having a referee and similarly for the presence of a crowd.

Very few output videos scored highly for successful facial recognition and blurring. No obvious correlations were identified by visual correlation inspection. The distributions show that *fully* blurred outputs have very specific characteristics, while the *partially* blurred ones constitute most of the data set, with only 10% of the set scoring above 6. The plots indicate that all the *fully* blurred outputs are taken from close distances and tend to be recorded from static or 'steady' cameras, from a medium angle, with no crowd. These attributes will be important in determining whether a video will be successfully anonymised.

Figure 1 shows the class distributions for the 10-class and 2-class target labelling.

TABLE III
MOST RELEVANT ATTRIBUTES.

Attribute	Rank
Crowd	0.3242
CamDistance	0.3203
In-Out	0.3203
CamMovement	0.2958
CamAngle	0.2695
Referee	0.261

ISSN 2395-8618

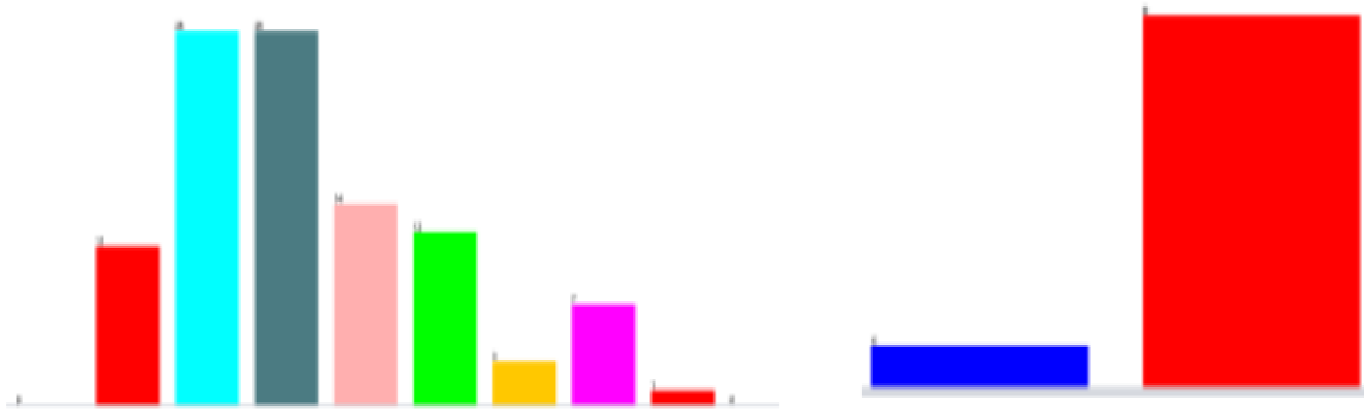

Fig. 1. Output video class distribution with 10-class labelling (left, classes 1 to 10 shown from left to right) and 2-class labelling (right, blue = fully blurred, red = partially blurred).

TABLE IV
PREDICTIVE MODEL PERFORMANCE.

Attribute Set	Algorithm	Accuracy	Sensitivity	Specificity
1	Naïve Bayes	91%	100%	90%
1	Trees.J48	89%	30%	95.5%
2	Naïve Bayes	94%	90%	94.4%
2	Trees.J48	89%	30%	95.5%

Weka correlation attribute evaluation found *Crowd*, *Camera distance*, *Indoors/outdoors* and *Camera movement* as the most relevant attributes affecting video blurring accuracy. A longer list of relevant attributes and their ranks is given in Table III.

C. Predictive Models

The attribute set in Table III, chosen using Weka correlation attribute evaluation on the dataset with the 2-class target, was used in two variants, with *Referee* (attribute set 1) and without *Referee* (attribute set 2), each to build two different predictive model types. The performance of the 4 resulting models is presented in Table IV, in terms of accuracy, sensitivity and specificity. Accuracy represents overall predictive power, while sensitivity and specificity respectively measure the power of predicting the *fully* blurred and *partially* blurred classes.

IV. Evaluation and Conclusion

The work described in this paper investigates solutions to two different interrelated problems: the automated anonymisation of sports videos and the evaluation of the anonymisation process. The anonymisation prototype was successfully built and run on a batch of a 100 sports videos downloaded from the Web to produce anonymised output videos. The properties of the input and output videos were summarised and the video attributes most predictive of successful anonymisation identified.

A. Anonymiser Prototype

The prototype is a functioning piece of software, however, the many parameters that informed its development, in particular the face recognition libraries, and its eventual performance characteristics both in terms of speed and anonymisation success, demonstrate that automating the removal of person-identifying visual information from videos is not a trivial task.

For example, the processing of 100 videos took 8 days on two high-end processors. This could be improved upon through code parallelisation.

The face recognition libraries used have particular limitations, including difficulty with identifying faces obstructed with arms or head-wear such as hats and sunglasses or poor performance with unfocussed or blurred footage. In this sense the prototype is a very specific implementation of the generic functionality of interest. However, its main value is in the proof of concept, not only in the functional sense but also as it serves as a kind of a 'wet-lab' for the analytic work presented here, which heavily relies on it.

B. Anonymisation Performance Prediction

A large number of video attributes, some specifically pertinent to sports videos, were investigated and tested for relevance to the success of face detection in a video. For example, the presence of a crowd, camera distance and indoor vs. outdoor filming affect the success of anonymisation. This kind of information is helpful when recommendations can

ISSN 2395-8618

be made in the production of video material for a particular purpose such as examination.

The idea of automated extraction of attribute values, which in this instance had to be identified 'manually' from the videos, while secondary to the work at hand, would have wide application in video-related research. Closer to the central topic of this paper, further work could continue on to solidify and generalise the conclusions through investigation across a greater number of face identification libraries and video types.

References

[1] H. Suad and O. Ahmed, "Biometric system based on face recognition system," Master's dissertation, Firat University, 2016.

[2] P. Bergeron, K. Kuchta, M. Olenik, B. Stern, and C. Tosswill, "Kodak Facebook collage project," Rochester IoT, Tech. Rep., 2016.

[3] S. Abdo and A. Mohammed, "Recognition of face detection system based on vide," Universiti Teknikal Malaysia Melaka, Tech. Rep., 2015.