Examining the Performance for Forensic Detection of Rare Videos Under Time Constraints

Johan Garcia
Department of Mathematics and Computer Science, Karlstad University, Karlstad, Sweden

Keywords: Digital Forensics, Video Classification, Monte-Carlo Simulations.

Abstract: In many digital forensic investigations large amounts of material needs to be examined. Investigations involving video files are one instance where the amounts of material can be very large. To aid in examinations involving video, automated tools for video content classification can be employed. In this work we examine the performance of several different video classifiers in the context of forensic detection of a small number of relevant videos among a large number of irrelevant videos. The higher level task performance that is of interest is thus the ability to detect a relevant video in a limited amount of time. The performance on this higher level task is a combination of the classification performance, but also the run-time performance of the classifiers. A variety of video classification techniques are available in the literature. This work examines task performance for 6 video classification approaches from literature using Monte-Carlo simulations. The results illustrate the interdependence between run-time and classification performance, and show that high classification performance in terms of true positive and false positive rates not necessarily lead to high task performance.

1 INTRODUCTION

Video content classification has been a topic of interest to the research community for a considerable amount of time. Many different approaches exist which differs in classification and runtime performance. One area where video content classification may be beneficial is in digital forensics. When performing forensics examinations, the amounts of data that needs to be handled are often substantial. By employing various automated classification tools, examinations can be performed more efficiently.

In this work we focus on examinations that are performed under time constraints. One instance when this could happen is when there is a suspicion of inappropriate material, such as adult videos, on computers or storage devices belonging to a school or an organization where such material is not allowed on the organizations equipment. Another case where time constraints may be present is during the initial phase of investigations related to child sexual abuse (CSA) material. Such examinations can take place when there is a degree of suspicion but not sufficiently to warrant the seizure of equipment. The equipment might be examined in place for a period of time and if data of relevance is located then this would be grounds for collecting the equipment and making a thorough forensic analysis.

The particular task that the video classification should support in these contexts is the detection of a small number of relevant videos among a large number of non-relevant videos. This particular use case presents particular challenges to the video classifier with regards to classification and runtime performance. In many cases video classifiers were designed and evaluated on classification performance and with no, or little, consideration of runtime characteristics. However, since the runtime performance of video classification approaches can vary several orders of magnitude it can be of considerable importance for task performance in a time constrained setting. In this paper we examine six different video classification approaches in the context of detecting rare relevant videos that occur with low frequency among a large number of irrelevant videos. In this examination the classifiers consider the relevant videos to be detected adult videos among a set of regular videos. However, the same trade-off between classification and runtime performance occurs for the case where a small number of CSA videos should be detected among a large number of non-CSA videos. To perform the evaluation of the video detection perfor-

Garcia J..
Examining the Performance for Forensic Detection of Rare Videos Under Time Constraints.
DOI: 10.5220/0005574204190426
In *Proceedings of the 12th International Conference on Security and Cryptography* (SECRYPT-2015), pages 419-426
ISBN: 978-989-758-117-5

mance Monte Carlo simulations are employed. The results show that for the considered classification approaches the difference in runtime characteristics are so large that they dominate over the classification performance in the majority of settings. Only when the number of rare videos that can be detected is very low, a slower but better-classifying approach is more appropriate for a subset of the considered time constraint times.

The paper is structured as follows. In the next section the background and related works are discussed. Section 3 covers the evaluation setup, followed by the results in Section 4. Finally, Section 5 provides the conclusions.

2 BACKGROUND

A problem when comparing different video classification approaches is that the papers describing the different approaches use different data sets. A valid comparison of the classification and run-time performance of the different approaches can not be achieved when different data sets and different hardware are employed by the different authors. As the video detection task is dependent on data for both the video classification performance and the runtime performance under comparable conditions the amount of usable input from the literature is limited. In some cases authors also re-implement approaches described earlier in the literature and evaluate this collection of approaches on the same data sets and hardware. This is the case for the work described in (Jung et al., 2014). Our present study is based on the classification and performance results reported by them. In their paper Jung et al. consider eight different video classification approaches involving various characteristics of the videos with some focus on the presence of periodic motion. Out of these eight, two are proposed by the authors.

The Jung1 approach proposed by the authors are based on classification performed by a 70-dimensional support vector machine (SVM). The features used are based on the magnitude and frequency of periodic motion along with a skin color feature vector based on hue and saturation histograms. They also propose the Jung2 approach which is similar but does not consider the skin color information. In addition to their approach, Jung et al. have also implemented an approach suggested in (Behrad et al., 2012). The Behrad approach is based on locating the largest section that is a skin colored, and compute the correlation between frames, and subsequently analyze the discrete Fourier transform of the correlation. Another examined approach suggested in (Ulges et al., 2012) is based on fusing several features including skin color, the discrete cosine transform coefficients of image patches using a bag of visual words model, motion histograms and also audio information based on cepstral coefficients. This multitude of employed features comes with a corresponding impact on classification speed. Another approach is suggested in (Endeshaw et al., 2008) which has low computational requirements as it works only on the motion vectors present in the compressed video stream. While such approach has considerable runtime benefits, it was also found to yield worse classification performance than the previously discussed approaches. In addition to these approaches, Jung et al. also evaluates three schemes designed to detect periodic motions. These approaches (Briassouli and Ahuja, 2007; Niyogi and Adelson, 1994; Liu and Picard, 1998) were not originally designed with any consideration of using them for classifying adult video material.

While the current study focuses on the detection of adult videos, the same consideration regarding the trade-off between runtime performance and classification performance exists in the domain of CSA investigations. As long as the rare video detection task can be achieved using different approaches with varying classification and runtime characteristics the same methodology as used in this study is applicable. The paper (de Castro Polastro and da Silva Eleuterio, 2012) highlights the importance of being able to quickly detect videos in the context of CSA investigations. The automated classification of CSA video material is discussed in (Schulze et al., 2014), which examines a multi-modal approach. Hence the possibility exists to examine the resulting task performance for the different modalities based on their varying run-time and classification characteristics in a manner similar to what is done here.

With regards to Monte Carlo simulations as employed here, one example where it has been used previously in the domain of digital forensics is (Garcia, 2014). There, the implications of including side information for file hashing has been explored using approaches similar to what is used in this study but with more focus on storage device characteristics.

3 EVALUATION SETUP

3.1 Performance of Classifiers

The video classification performance for the eight classifiers discussed in the background section are shown in Figure 1. The figure shows the Receiver Op-

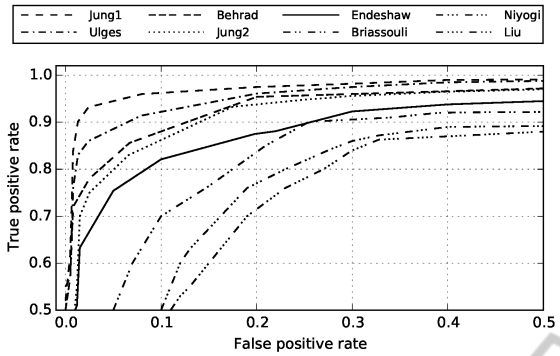

Figure 1: Classification performance (from (Jung et al., 2014)).

erating Curve (ROC) which presents a curve of all the possible operating points of a binary classifier. The x-axis shows the false alarm rate, or false positive rate (FP), which in this case corresponds to the fraction of videos classified as being relevant although they actually are irrelevant. The y-axis shows the true positive rate (TP), which is the fraction of the relevant videos that are actually detected as such. Ideally, FP should be low and TP high, leading to the upper left corner representing the optimal classifier. The curves in the graphs represents the possible trade-offs between FP and TP that results from setting the detection threshold for an individual classifier to different values. By setting the threshold to a specific value the classifier will work at a particular operating point with specific FP and TP.

In addition to the classification performance shown above, runtime aspects are also of importance in this evaluation. The run time characteristics of the eight classifiers as reported in (Jung et al., 2014) are shown in Table 1. The table shows the amount of

Table 1: Run time performance (times are from (Jung et al., 2014)).

Method	Processing time for 6 hours video	Relative speed
Jung1	9 min 30 sec	37.90x
Ulges	3 h 17 min	1.83x
Behrad	18 min 20 sec	19.64x
Jung2	9 min 9 sec	39.34x
Endeshaw	1 min 21 sec	266.67x
Briassouli	5 h 33 min	1.08x
Niyogi	2 h 32 min	2.37x
Liu	2 h 26 min	2.47x

time necessary for processing video files with the total duration of six hours, and the resulting speed of the classifier in relation to normal real-time. These results were obtained by Jung et al. using a PC with a 2.8GHz CPU and 4GB of RAM. The videos had a resolution of 640×480 pixels. As can be seen in the table, there is a large variation between the different classification approaches. These differences relate to how the classification is performed, as discussed in the background section.

3.2 Chosen Parameter Setup

Based on their performance characteristics, six classification approaches were selected for a further Monte Carlo simulation-based performance evaluation. As the Niyogi and Liu approaches had relatively poor classification performance for this task, and had no particular runtime advantages they were not considered in the further evaluation. Based on the ROC classification performance characteristics provided above, suitable operating points for the different remaining classification approaches were decided using heuristics. For this particular task operating with a low false positive rate is considered more important than having the highest possible true positive rate. The chosen operating points are shown in Table 2.

Table 2: Selected operating points.

Method	True positive rate	False positive rate
Jung1	0.90	0.013
Ulges	0.83	0.013
Behrad	0.76	0.008
Jung2	0.70	0.015
Endeshaw	0.63	0.015
Briassouli	0.60	0.070

In addition to classifier characteristics, the simulator also needs to model the characteristics of the video material to be evaluated. This study focuses on the case where a small number of relevant videos is to be detected among a large number of non-relevant videos. The videos are modeled as coming from three different classes with different characteristics. The first class represents material obtained via video sharing sites such as YouTube. The YouTube statistics is also considered to represent amateur-generated content in general. Statistics for the run time, compression rate, and byte length of videos present on YouTube are reported in (Ameigeiras et al., 2012). The T3 data set provided by that study is also used here, although it was filtered to only consider videos with a width of 480 pixels. The two other classes considered are downloaded TV shows and movies. For these classes the byte length distribution reported in (Carlsson et al., 2012) is used in a simplified way, using the reported medians, but with a uniform distribu-

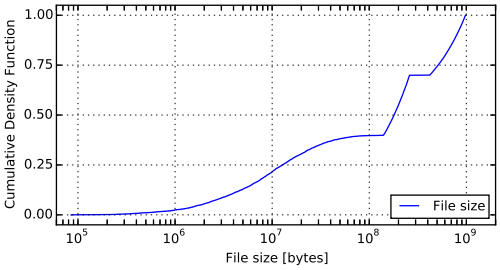

Figure 2: Video file size distribution.

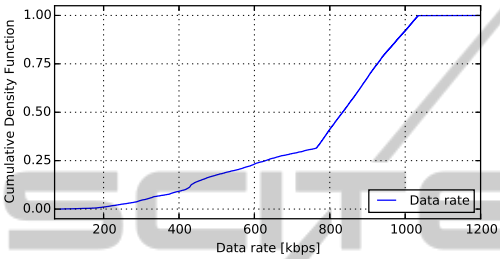

Figure 3: Compression distribution.

tion of ±30% for TV shows and ±40% for movies. As no empirical data was found for the compression rate of these classes, it was set to vary uniformly within the range 765-1035 kbps. The relative proportion of video files were set to 0.4 for YouTube/amateur videos, 0.3 for TV shows, and 0.3 for movies.

For these proportions, the resulting cumulative density function (CDF) for the variation of byte length file sizes in the modeled video material is shown in Figure 2. The leftmost hump in the figure corresponds to the YouTube/amateur material, and the two rightmost slopes correspond to TV shows and movies. Here, the x-axis is logarithmic. Similarly, the resulting compression rate CDF is shown in Figure 3. Finally, the modeled video duration CDF is shown in Figure 4. Here, the part under 1000 seconds comes mainly from the empirical time distribution found in the YouTube data set, whereas the part above 1000 seconds mainly model the durations of the TV series and movie classes. As can be seen in the figure the

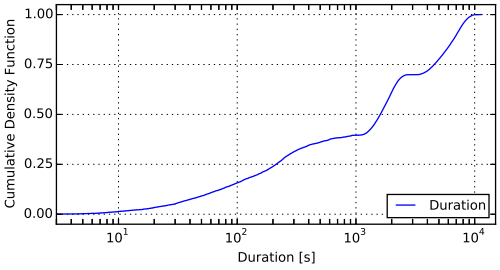

Figure 4: Video duration distribution.

median time for modeled TV show is around 30 min., and for the movie class around 1 hour 40 min.

3.3 Simulation Configuration

The simulation configuration models 500 GB of video material stored on a storage device. As the bottleneck in this study is classifier runtime performance rather than storage device read characteristics, no storage device related delays were included in the simulation model. One Monte-Carlo simulation round consists of filling the space on the storage device with video files with the characteristics given by the file size, compression rate, and duration distributions discussed in the previous subsection. Then, a particular number or fraction of files are randomly marked as relevant, or *rare*. The different evaluation metrics are then computed using the aggregated values after considering the operating point and run-time characteristics for each of the classifiers for each file. The simulations used 10000 Monte-Carlo runs. Each simulation run generates an average of around 1800 video files.

3.4 Validity Aspects

Although data from literature was employed for several parameters and model aspects for the simulation, it should be recognized that some parametrization simplifications were made. Some simplifications were performed by Jung et al when implementing some of the classification approaches as discussed in their paper.

A sensitivity analysis using different video distributions and operating points showed that the relative performance of the examined approaches is robust to such variations in the parameterization. Thus, these results are considered to be sufficiently general to strongly support that run-time characteristics needs more attention when designing classifiers intended for the forensic domain.

4 RESULTS

The results section is divided into three parts. In the first part the focus is on the amount of time until the first detection of a rare video. The performance here depends mainly on runtime characteristics, but when the number of rare videos is very small the true positive rate is also of importance. The second part considers the evolution of the number of detected videos over time, and the amount of manual verification work

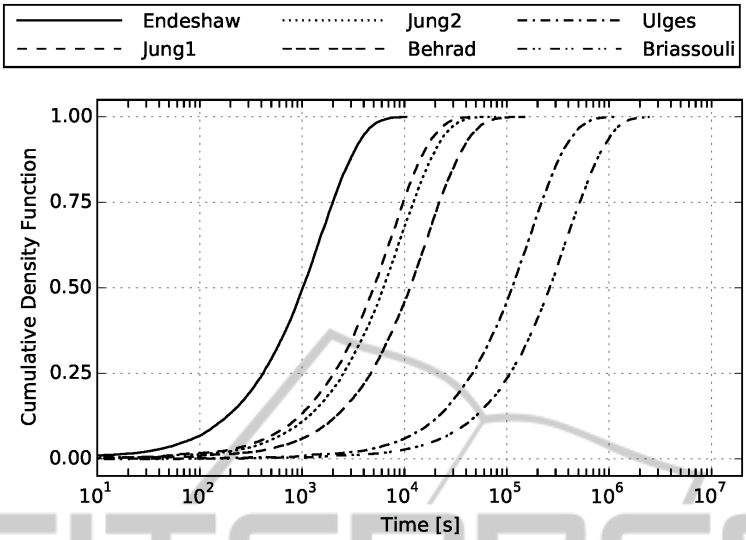

Figure 5: CDF of time to first detection of a rare video, with a 0.01 fraction of rare videos.

that is necessary. The third part shows how a combination of classifiers can be employed to create improved performance for the metrics discussed in the first two parts.

4.1 Time to First Detect

The time required to detect the first rare video depends on where in the sequence of examined videos the first rare video happens to be placed. In reality this varies from case to case, and this variation is captured by the Monte-Carlo simulations which randomizes the placement between each simulation run. The first examination considers how the time to detect the first rare video varies as a result of the randomness in the placement of the rare files and other random factors. The results for the different classifiers when the fraction of rare files is 0.01 are shown as a CDF in Figure 5. Note the logarithmic x-axis. As can be seen in the figure, the amount of clock time to detect the first video varies considerably. The median time varies from ca 20 min to ca 83 hours. In a practical setting such difference can have a considerable impact.

The CDF can also be used to infer the probability of detecting one rare video in a particular time. If a time constraint of 1 hour ($=3.6 \times 10^3$ s) is present, the probability of detection varies between 92% and 1%. This difference is mainly dependent on the run-time characteristics of the classifier, rather than its classification performance. As one focus with this examination is how video classification approaches function under time constraints, this graph can give an indication of how appropriate different classification approaches would be for particular time limits.

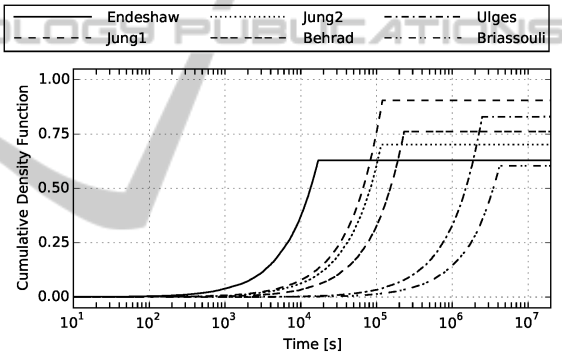

Figure 6: CDF of time to detection of a single rare video, when only one single rare video is present.

Figure 5 shows the conditions for a rare video fraction of 0.01. If this value is changed, this affects the time to detect the first video. As one extreme point, it is possible to examine the characteristics when there is only one rare video stored among the large (ca 1800) number of non-relevant videos. The case of a single rare video is shown in Figure 6, which shows the corresponding CDF. Here it is noticeable that the detection times have moved considerably to the left in the figure, making the average time it takes to detect the video considerably longer. Also visible in the figure is the fact that for a single rare video the true positive rate of the classification approach has a marked impact on the results. As can be seen in the figure the approaches tops out at a CDF probability that is equal to their true positive rate. For the fastest Endeshaw approach this means that among

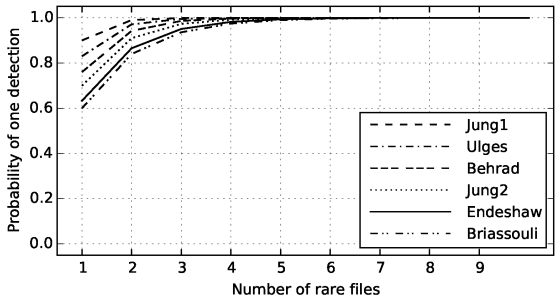

Figure 7: Probability to detect at least one rare video.

the 10,000 simulation runs, for a fraction of 0.37 of these runs the single rare file was not detected by the classifier. While this is a weakness of the Endeshaw classifier, it can nevertheless be observed that for a large fraction of the considered time periods it is the best choice. For all time periods below 8.5×10^4 s ($\approx$ 24 hours) the Endeshaw classifier has the highest probability of detecting the video.

In most cases it is reasonable to expect that the number of rare videos to be detected is larger than one. Based on the true positive rates it is possible to calculate the probability of detecting at least one rare video as a function of the number of rare videos stored. The results are illustrated in Figure 7, which shows that when there are 4-5 rare videos present, the difference in probability of detecting at least one rare video becomes very small.

4.2 Detections Over Time

It is possible to examine how the number of detected videos vary as a function of time for the different classifiers. Such data makes it possible to consider the performance implications of using another threshold than one video. The median number of detected videos over time is shown in Figure 8. The time to detect the first video in this figure corresponds to the median value of the CDF in Figure 5. While the Endeshaw classifier will be able to detect the most videos for a large fraction of the time period, it cannot detect as many videos as other classifiers due to its lower TP rate. In addition to examining the median, it is also possible to consider the 95th percentile of the 10000 MC runs. This is shown in Figure 9 which shows that both the time to detection as well as the maximum number of detected videos is lower, as can be expected. It can be noted that the transition point where Jung1 performs better than Endeshaw is set to similar point in time for both the median and 95th percentile.

Another factor to be considered is the amount of time that has to be spent validating the videos classified as relevant. As the false positive rate at the chosen operating points vary, the number of false positive videos also vary. The relative classification speed also plays a role, as the faster the classification is performed, the more videos can potentially be misclassified. The resulting validation amount is shown in Table 3. After one clock hour the Jung1 approach has scanned circa 38 hours of video and generated an average of 29 minutes of video which needs to be manually verified, in addition to any true positives detected. The faster and less precise Endeshaw approach has during one hour scanned around 267 hours of video

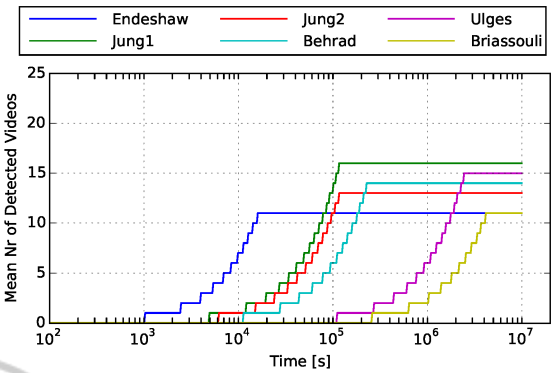

Figure 8: Number of detected videos over time (50th percentile).

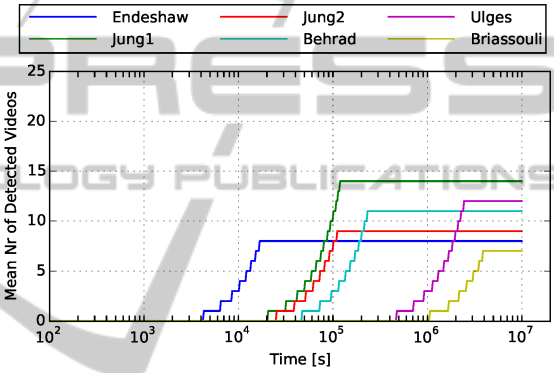

Figure 9: Number of detected videos over time (95th percentile).

Table 3: Generated validation amounts.

Method	FP rate	Relative speed	Validation amount
Endeshaw	0.015	266.67x	4x
Jung1	0.013	37.89x	0.493x
Jung2	0.015	39.34x	0.590x
Behrad	0.008	19.64x	0.157x
Ulges	0.013	1.83x	0.024x
Briassouli	0.070	1.08x	0.076x

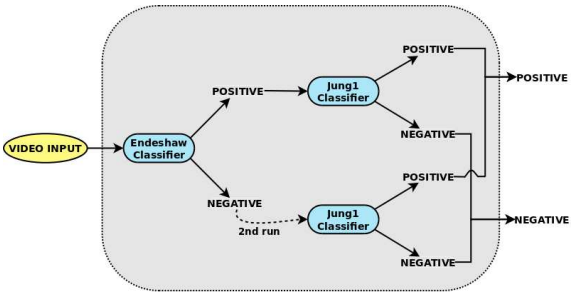

Figure 10: Endeshaw-Jung1 cascading classifier approach.

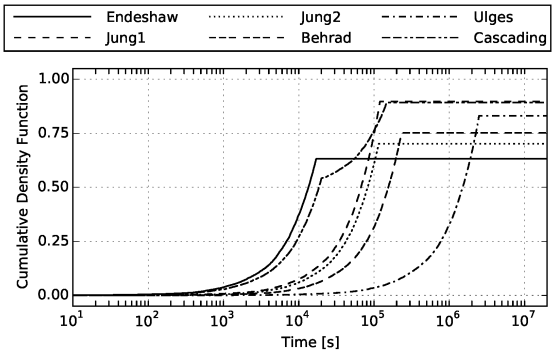

Figure 11: CDF of time to detect a single rare video, including an Endeshaw-Jung1 cascading classifier.

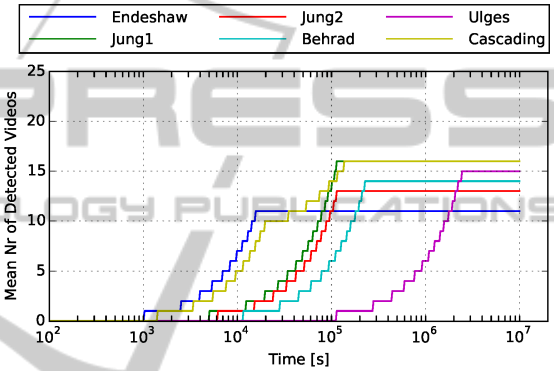

Figure 12: Median number of detected videos over time, including an Endeshaw-Jung1 cascading classifier.

and generated 4 hours of video material that needs verification. Out of all detected material some are actual true positives containing relevant material, and others are irrelevant material that the classifier misclassified. To separate these true positives from the false positives manual validation is necessary. The manual validation can be performed in different ways as deemed appropriate, but viewing the video with a moderate amount of fast forward is one simple approach.

4.3 Cascading Classifiers

In addition to using a single classifier it is also possible to cascade several classifiers after each other. This is done to improve some aspect of the classification system, although at the expense of having a more complex system. An early example of using cascading classifiers is the face detection approach described by (Viola and Jones, 2001). This subsection examines the performance effects of using a cascading approach for rare video detection, with the results derived under the assumption of classifier independence. The results in this subsection are thus indicative of a best case cascading scenario. As the modes of operation of the cascaded classifiers are significantly different they are expected to be quite independent. The examined cascading approach is shown in Figure 10. The fast Endeshaw classifier is first employed, and for every video classified as positive the Jung1 classifier is then directly employed as shown in the upper part of Figure 10. This cascading reduces the amount of false positives and the corresponding need for manual verification. Thus, the amount of video per clock hour needing manual verification is reduced from 4 hours to less than 10 minutes. The downside is a slight reduction in relative speed, which still is above 200x.

Furthermore, the approach of Figure 10 employs a second run after the first run of all videos has finished. This second run uses the Jung1 classifier with the goal of detecting the rare videos among the negatives returned by the Endeshaw classifier. This way a larger fraction of the total number of rare videos can be detected, and it will also increase the probability of finding a single rare video to the same level as for the Jung1 classifier.

The resulting performance of the cascading approach for the case of one single rare video is shown in Figure 11. In the figure the increased probability of detecting the video is evident, as well as a slight time increase in relation to the Endeshaw approach. The knee visible for the cascading classifier corresponds to the changeover point between videos classified as positive in the first or second run.

Figure 12 shows the median number of detected videos for a 0.01 fraction of rare videos. The results show a similar trend as was shown in Figure 11, with a slight increase in time and a higher number of detected videos. If the decrease in manual validation amount provided by the upper classifier of Figure 10 is not required, the upper classifier cascade can be removed. Such removal would lead to run-times below the knee being similar to those of the Endeshaw classifier, while still being able to detect the same amount of videos as the Jung1 approach.

5 CONCLUSIONS

Various classification approaches are used in a variety of contexts. Some of these contexts involve time-constraints where the run-time characteristics of the classification approach becomes especially relevant. In this work we have studied the higher level task of detecting a rare video with relevant characteristics among a large set of irrelevant videos. For such tasks the run time characteristics can be expected to have high importance in addition to the classification performance.

In this study Monte-Carlo simulations were performed to evaluate the task completion time as related to the run-time and classification performances of six different video classification approaches. Performance using cascading classifiers were also examined. The presented study reflects only a subset of the available approaches and has some simplifications. It nevertheless provides strong support for the statement that classification performance cannot be the only factor in consideration when designing classification systems used in time constrained environments. The results show that high classification performance in terms of true positive and false positive rates not necessarily lead to high task performance.

ACKNOWLEDGEMENTS

The authors wish to thank Soonhung Jung for sharing data on the receiver operating conditions from their study.

REFERENCES

Ameigeiras, P., Ramos-Munoz, J. J., Navarro-Ortiz, J., and Lopez-Soler, J. M. (2012). Analysis and modelling of youtube traffic. *Transactions on Emerging Telecommunications Technologies*, 23(4):360–377.

Behrad, A., Salehpour, M., Ghaderian, M., Saiedi, M., and Barati, M. N. (2012). Content-based obscene video recognition by combining 3d spatiotemporal and motion-based features. *EURASIP Journal on Image and Video Processing*, 2012(1):1–17.

Briassouli, A. and Ahuja, N. (2007). Extraction and analysis of multiple periodic motions in video sequences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(7):1244–1261.

Carlsson, N., Dán, G., Mahanti, A., and Arlitt, M. (2012). A longitudinal characterization of local and global bittorrent workload dynamics. In *Passive and Active Measurement*, pages 252–262. Springer.

de Castro Polastro, M. and da Silva Eleuterio, P. M. (2012). A statistical approach for identifying videos of child pornography at crime scenes. In *Availability, Reliability and Security (ARES), 2012 Seventh International Conference on*, pages 604–612. IEEE.

Endeshaw, T., Garcia, J., and Jakobsson, A. (2008). Classification of indecent videos by low complexity repetitive motion detection. In *Applied Imagery Pattern Recognition Workshop, 2008. AIPR '08. 37th IEEE*, pages 1–7.

Garcia, J. (2014). An evaluation of side-information assisted forensic hash matching. In *Computer Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International*, pages 331–336.

Jung, S., Youn, J., and Sull, S. (2014). A real-time system for detecting indecent videos based on spatiotemporal patterns. *Consumer Electronics, IEEE Transactions on*, 60(4):696–701.

Liu, F. and Picard, R. W. (1998). Finding periodicity in space and time. In *Computer Vision, 1998. Sixth International Conference on*, pages 376–383. IEEE.

Niyogi, S. A. and Adelson, E. H. (1994). Analyzing gait with spatiotemporal surfaces. In *Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on*, pages 64–69. IEEE.

Schulze, C., Henter, D., Borth, D., and Dengel, A. (2014). Automatic detection of csa media by multi-modal feature fusion for law enforcement support. In *Proceedings of International Conference on Multimedia Retrieval*, ICMR '14, pages 353:353–353:360, New York, NY, USA. ACM.

Ulges, A., Schulze, C., Borth, D., and Stahl, A. (2012). Pornography detection in video benefits (a lot) from a multi-modal approach. In *Proceedings of the 2012 ACM international workshop on Audio and multimedia methods for large-scale video analysis*, pages 21–26. ACM.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511–I–518 vol.1.