Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-10819-x

1169: INTERDISCIPLINARY FORENSICS: GOVERNMENT, ACADEMIA AND INDUSTRY INTERACTION


# An offline parallel architecture for forensic multimedia classification

**Luca Spalazzi**[1] · **Marina Paolanti**[1] · **Emanuele Frontoni**[1]

Received: 15 April 2020 / Revised: 1 January 2021 / Accepted: 10 March 2021 /
Published online: 02 April 2021


**Abstract**
Nowadays, the volume of the multimedia heterogeneous evidence presented for digital forensic analysis has significantly increased, thus requiring the application of big data technologies, cloud-based forensics services, as well as Machine Learning (ML) techniques. In digital forensics domain, ML algorithms have been applied for cybercrime investigation such as child abuse investigations, malware classification, and image forensics. This paper addresses this issues and deals with forensic analysis of digital images and videos. In particular, this work aims at proposing a multimedia classification tool with a parallel software architecture for a fast inspection, which is easy to use (to be used by officers during a search), requires limited hardware resources and it is built on an open-source software to limit its costs. Moreover, this tool must be able to quickly inspect multiple devices at a time. When positives are found in a device, such device will be seized for a deeper analysis later in the lab. It will not be seized otherwise, reducing the inconvenience for the suspect as well as the time required for the next analysis phase. As a case study, we focus on the identification of child pornography images. Experimental results show that the proposed architecture is capable of guaranteeing a high recall, a fast process and high performances in real scenarios.

**Keywords** Parallel architecture · Machine learning · Digital forensic

✉ Marina Paolanti
m.paolanti@univpm.it

Luca Spalazzi
l.spalazzi@univpm.it

Emanuele Frontoni
e.frontoni@univpm.it

[1] Department of Information Engineering (DII), Università Politecnica delle Marche, Via Brecce Bianche, Ancona, Italy

## 1 Introduction

Digital forensics is the part of forensic science dealing with the recovery and investigation of the material found in digital devices [16, 39, 40]. This discipline is employed in various application domains, notably, in this work, we focus on investigations carried out by the law enforcement on all those crimes where digital devices play a primary role. In this context, it should be noticed that, although several process models have been proposed by the literature for digital forensics, all of them comprise four phases: seizure, acquisition, analysis, and reporting.

**Seizure** First of all, one must identify which devices can contain useful information for the ongoing investigation and move on to their seizure. This activity, in the domain of interest for this work, is carried out by law enforcement who act under a specific search warrant.

**Acquisition or imaging** Once a device of interest has been identified, law enforcement, acting under a specific seizure warrant, must create an exact duplicate in order to guarantee: the analysis of the duplicate without compromising the integrity of the original device; the reproducibility of the analysis.

**Analysis** After the acquisition, the image of the device is analyzed in the laboratory to identify any evidence against the suspect or in its defense.

**Reporting** Once the analysis is complete, its results must be described in a official report.

From this brief overview, it boils up that access to the device contents occurs in two completely different scenarios in terms of time, place and characteristics: during the search and during the analysis.

- *During the search*, law enforcement officers are at the suspect's, the purpose is to define whether and which devices to seize. In this phase, it is important to have a high recall (low false negative rate) but unfortunately computational resources and time are limited. In other words, device inspection must be fast and based on poor computational resources.
- *During the analysis*, law enforcement officers act within their laboratories, the aim is to collect evidence for a formal examination before a court. In this phase, it is essential to have a high accuracy (both a low false positive rate and a low false negative rate).

Although numerous tools have been proposed for the analysis phase, so far little attention has been paid to the search phase. For this reason, this work focuses exclusively on this phase, in particular it focuses on the search for images related to the crimes of child pornography, violence, terrorism, drug trafficking, and other illicit trafficking. In fact, nowadays, by the advent of the social web and social networks, billions of individuals globally use digital technology daily.[1, 2] Among them, more and more people use such technologies for illegal trafficking: a market that earns more than a trillion dollars every year.[3] As a consequence, the amount of data that police forces have to inspect during a search is really impressive, tens and tens terabytes. Just to give an idea, let us suppose to be able to process

[1] https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates

[2] https://www.forbes.com/sites/michelleevans1/2018/12/17/5-stats-you-need-to-know-about-the-digital-consumer-in-2019/#6a59c6d636bd

[3] https://dataprot.net/statistics/cybercrime-statistics/

32 MB per second, the search would require 9 hours to inspect 1 TB of data. This means that in a realistic scenario, the search could go on several days. A search whose duration is so long usually has a significant impact on the life of the suspect and those around him. An even worse situation, both in terms of duration and impact, occurs when the search involves corporate devices. In cases like these, the activity of the whole company might be slowed down or stopped altogether by the search. The strategy usually used to reduce this impact is to seize all the devices without inspecting them and therefore without distinction, postponing a thorough inspection to the analysis phase. But even this strategy has a substantial impact, because the suspect (or worse the company) remains without devices for a long time. Impact that is even higher if the suspicions prove unfounded. For these reasons, the only acceptable strategy is to have on one hand a tool for a quick inspection with a high recall. By contrast, this tool will have limited computational resources, since it is unreasonable to assume that the police forces have with them, during the search, high-performance processing servers or rely on online services.

This work therefore aims at proposing an image and video classification tool with the following requirements:

- a *parallel software architecture* (for a fast inspection),
- *easy to use* (to be used by officers during a search),
- requiring *limited hardware resources* and based on *open-source software* (to limit its costs),
- however capable of guaranteeing a *high recall*.

In other words, this tool must be able to quickly inspect multiple devices at a time. When positives are found in a device, such device will be seized for a deeper analysis later in the lab. It will not be seized otherwise, reducing the inconvenience for the suspect as well as the time required for the next analysis phase. It should be noticed that some false positives are tolerated, as the analysis phase give the opportunity to enhance the accuracy of the classification (at least by means of a manual inspection).

As a case study for the experimental part, we concentrate on the identification of child pornography images. This is in consideration of the fact that child pornography is one of the the most relevant phenomenon in the area of illegal contents and whose social impact is very high.

In a first inspection, the device verifies the presence of pornographic images since it is difficult to distinguish an adult from a child. In literature, several works have been proposed to face these analysis. Solutions proposed range from nudity detection [30] and facial analytics [6, 29], as proxies for child pornography classification to bags of visual words [32, 36] and behavioral analytics [1] to network profiling [4, 8, 27, 34] and sensitive hashing techniques [11, 29]. In the lab, during the next analysis phase, forensic technicians have time for a deeper check, in order to discriminate illegal and legal contents. For this reason, the methods proposed in this paper are novel with reference to the state of the art, since their aim is to achieve an high recall without a particular focus on accuracy. The methods are available as open source and are novel with respect to the state of the art. The officer only need a sufficient number of evidences to be manually inspected and signed before ending the seizure) and a very fast processing time to ensure the main goal of the fast inspection.

The main purpose of the result section is to prove the effectiveness of the proposed pipeline with a particular focus on i) computational performances (all the system is designed to work with low cost hardware), ii) configurability (the solution is based on the idea that every end user con easily collect a dataset and train the AI algorithms to personalize the specific use case), iii) human based result verification (the final goal is to provide to end

users a short list of the most sensible contents for the final forensic verification, reducing the human verification work from thousands of multimedia contents to tens, saving time to work on more effective actions). All these features are tested on standard methods, given that main novelties are mainly on the whole pipeline and conceptual design, with respect to the specific machine learning and deep learning methods.

The paper is organized as follows. Section 2 provides a description of the approaches that were adopted for digital forensic. Section 3 describes the proposed architecture for image classification. In Section 4, an evaluation of our approach is offered, as well as a detailed analysis of of a specific case study. Finally, in Section 5, conclusions and discussion about future directions for this field of research are drawn.

## 2 Related work

Recently, digital images have become widespread in our daily life. The images, compared with textual content, are more spontaneous and can convey much more information [20]. Despite these advantages, the simple accessibility of digital images has emerged in important security problems, as a way to evaluate the authenticity of digital images and to detect illegal content. The new technologies allow to create, collect, and analyse the image contents. Manuscripts have been focused on video and image verification for evaluating if any manipulation exists [26, 38].

Kamenicky et al. [18] have introduced tools and method for images and videos analysis in context of criminal inquiry.

Several efforts have been devoted to video and images sources recognition [21, 25]. In [21], the authors have presented an algorithm which allows to extract photo response non-uniformity (PRNU) noise from video files obtained by the camera of mobile phone. The authors of [2] have presented a source identification approach for video files posted on social networks, such as Facebook,Twitter, Wechat, etc. Using PRNU, the method introduced by Amerini et al. can gathered fingerprint for camera phone.

Another approach proposed for social media data has been described in [10]. In this paper, the authors applied machine learning methods and a-priori knowledge gained through image processing. They have developed an approach that automatically understand which Social Network has handled an images as well as the software used for uploading it. This approach also consider as a feature if any adjustment has been introduced.

In [17], the authors have described an images analysis and processing by bringing together face recognition methods to detect covered facial information. Considering the devices quality, it is possible to apply this method in forensic vide/image processing for criminal identification, as already done in [14]. Maksymowicz et al. [23] proposed method for the crime reconstruction of an event or a scene using 3D analysis of video and image.

Recently, in [12, 15], it has been explored the automatic detection of risky circumstances for public security. Instead, for image enhancement techniques it has been employed histogram equalization (HE), in which image histogram is defined as statistic probabilistic distribution of gray levels [42].

The relevant approaches for anti-pornography systems are classified in two stages: skin detector and pornography classifier [41].

Regular and irregular patches have been explained as algorithms for skin colour detection [44]. The method applied by authors achieve 98.8% recall and a 96.5% precision. However, the results are good it is slow and it is not presented a test set. The resolution

and the quality and the brightness of images considerably influence the results, and hence considering another algorithms is of limited importance.

In [33], the authors have proposed a model based on skin detection which filters pornographic image. It merely detects static images which rely on a certain threshold. One issue is related to the fact that the skin region is too bright or too dark due to the variant illumination environment. Another reason is that there are several objects with skin-similar colours in the image background.

In [7], a hybrid approach is proposed. It aims at detecting pornographic contents in images. At any rate, as regards applications, knowledge modelling is complex and features are diverse.

A system based on neural networks for classifying images into pornography and non-pornography is proposed in [31]; the outcome shows that pornographic images can be classified by the system through an association of visual cues, which includes human and colour figure features. The shortcoming of the described pornography classifier arises for different reasons. In few images, the pornographic content is poor to identify, instead in others, skin zones are not-saturated, thus causing the fail of the skin classifier.

In [24], the authors have proposed a framework based on a multi-colour skin model for identifying pornographic images. This used RGB, normalised RGB, YCbCr, and HSI colour spaces. In [43], a SVM is employed by the authors. They trained a pixel-based skin classifier with the task of discriminating skin and non-skin pixels based on HSV colour space.

For pornographic image detection, promising results have been achieved. However, significant work still needs to be done to reach an automatic application that can detect pornographic images. Moreover, image identification is an image object recognition problem and it is demanding for different reasons [9]. Images are captured under different illumination degrees and are digitised in various resolutions. Another issue regards those images that may contain parts of human body in different poses or those ones' which may be partially dressed. Additionally, some art pictures are similar to pornographic images. Another basic question is linked to the prevalence of skin regions in pornographic images. In this paper, an anti-pornography architecture is described; it relies on the exposed skin detector module with the intention of overcoming problems mentioned.

Main contributions in the forensic fields are: i) an open source and low cost tool to speed up low enforcement officers inspection search phase; ii) a novel off line parallel hardware and software architecture for large scale multimedia data processing and classification; iii) a set of fast and high recall ML and DL algorithms to cover the most common inspection cases ; iv) an extensive test in real cases to demonstrate the effectiveness of the proposed process and the inspection time reduction.

## 3 Materials and methods

In this section, the parallel architecture for forensic multimedia classification is introduced as well as the case study used for evaluation. The solution provided is schematically outlined in Fig. 1. Further details of the suggested solution are delivered in the following subsections.

### 3.1 Digital inspection bag

The aim of this work is to elaborate an innovative "digital inspection bag" easily transportable and adaptable to any inspection purpose and able to detect on-site and off-line illegal multimedia contents. Moreover, it is easily usable by a non-expert user. The overall

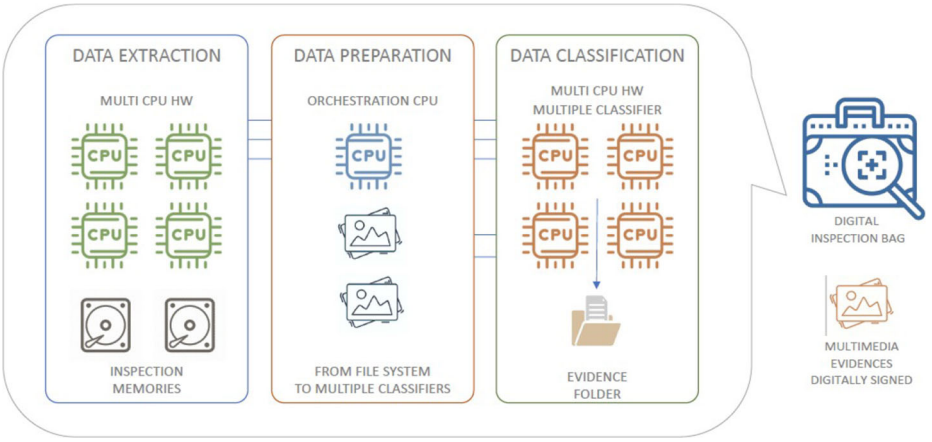


**Fig. 1** Overview of the implemented solution

architecture is base on open source software and low cost hardware (i.e. 8 Raspberry Pi 4 Model B) to ensure an easy introduction in real scenarios with very low investment.

A multi-CPU master-slave architecture is designed to extracts images and videos from file (directly from the file system and from deleted files with the method of searching for known multimedia file types signatures) to be sent to a slave classifier architecture that detects classes in every input files: (a) each slave uses N classifiers, one for each type of object to be detected; (b) the distribution is on the data preparation and load and not on the logic.

The preparation phase is also taking care of image resizing up to a max vertical & horizontal dimension of 1000 pixels and video fame extraction (only 1 frame by 50 is processed by default, with the possibility to change this parameters in special inspection cases) that are The proposed workflow works in series where master search for image files (header analysis) and randomizes the list of files, partitions and sends them to the slaves, which can stop, taking into account local heuristics. Each slave activates the classifiers from different points in the dataset.

The proposed method allow also the applications of different heuristics to reduce the inspection time and stop the classification process, asking for a human verification and confirmation of the evidence folder contents.

The randomization in file order deprives the dataset of the logical order imposed by the user. This allows to define the stop criteria (local heuristics) for slaves:

- Global: all the classifiers in the slave stop when
  - more than 60% of the list of file in the device has been analyzed;
  - more than 10% of the list of file in the device has been positively classified (illegal).
- For every classifier: only one by N stops
  - less than 10% after analyzing already 40% of the file list in the device.

Figure 2 represents the block diagram of the whole implemented architecture and shows the interfacing between master and slave. All described parameters can be personalized in the tool configuration files.

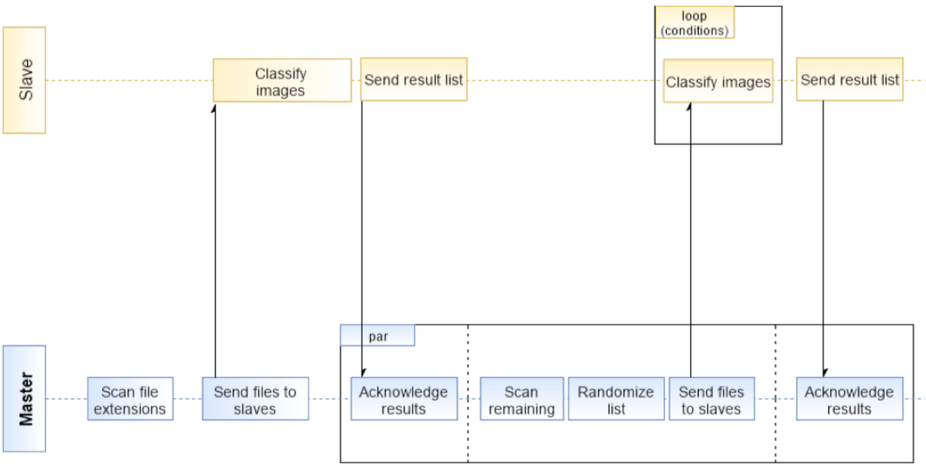


**Fig. 2** Block diagram of distributed architecture

### 3.2 Multimedia classification architecture

The classification architecture is based on different classifier that can be used by law enforcement officers, based on the seizure and acquisition purposes. All methods are completely open source and are novel with respect to the state of the art, mainly because the purpose of every method if to have a very high recall without a particular focus on accuracy (the officer only need a sufficient number of evidences to be manually inspected and signed before ending the seizure) and a very fast processing time to ensure the main goal of the fast inspection. All the special purpose classifiers are here following:

- Pornography and child pornography: is based on fast color and shape based feature and will be better described in next paragraph as the use case of this paper;
- Violence: is based on deep learning scene understanding using .
- Terrorism: is based on special symbol detection only using a SURF point feature detector
- Illicit trafficking: based on a gun identification algorithm based on shapes detection and classification

In the next section, only the Skin detection approach for pornography inspection application is explained as a Use Case of the contemplated solution. This use case is also used for results discussion.

### 3.3 Pornography use case

This use case, as previously discussed, is designed with the main purposes of high classification speed with high recall results. First, an initial phase of *Skin detection*, in which the portion of the skin in an input image is detected. This phase represents the initial step for the most applications that classify pornographic images: it is natural to think that an image, to be considered pornographic, should contain large portions of exposed skin. This procedure classifies each pixel of an image as belonging or not to human skin. The simplest approaches to model skin color are those that use explicit rules, described by logical expressions. Kovac

and others [19] used RGB space to define the regions of human skin. Others [5] have used $YC_bC_r$ space excluding luminance in their model. Moreover Hsieh and others [13] introduced thresholds in HSI space. Although these solutions appear very intuitive and perhaps with limited performance, allow a rather fast classification of the skin, an important aspect for this type of application. In this work, we compared two thresholding methods: $YC_bC_r$ and a composed RGB and HSV space of colors. This last combination guarantees a higher level of accuracy. The second step consists of the *Features extraction*, in which from the information on the skin contained in the image, we will define its discriminating characteristics. Then machine learning algorithms use features extracted in order to make recognizing and classification of the input images. The choice of the number of features extracted by an object is fundamental for the success of the following phases of training and testing of a classifier based on those characteristics, since the number of features affects the size of the feature vector, and therefore also of the feature space. It is important to choose a model that minimizes overfitting and underfitting problems. Finally, a *Classification* algorithm of pedo-pornographic images is developed, it must classify an image as legal or not legal.

### 3.3.1 Skin detection

For the first approach the skin detection method used is $YC_bC_r$ as the diagram in Fig. 3a shows. The method performs the thresholding of the input image after converting it into the

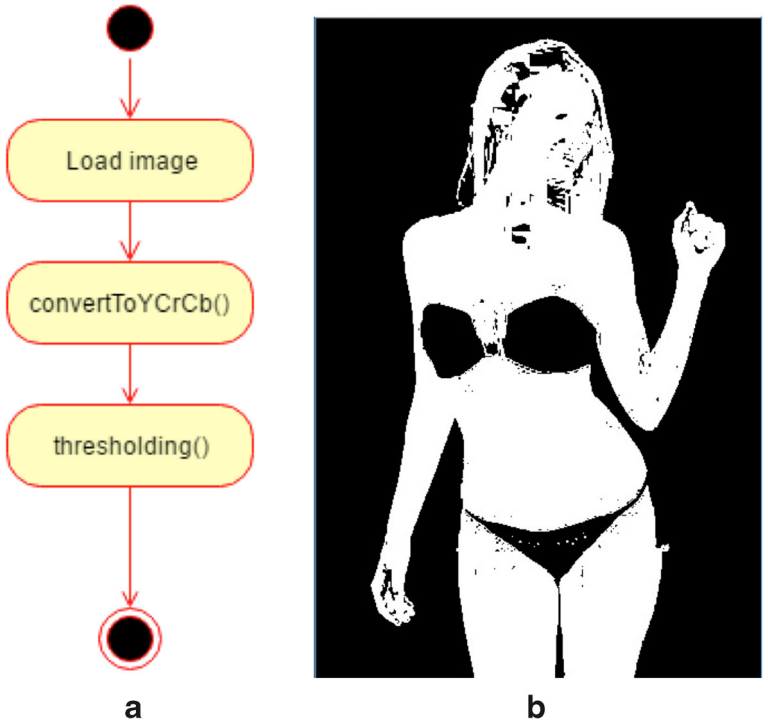


**Fig. 3** **a** Represents the diagram of skin detection using $YC_bC_r$ space. **b** is the result of the application of $YC_bC_r$ thresholding rules

luminance-chrominance space and returns the percentage value of the number of pixels in the skin of the entire image.

Figure 3b shows the resulting image after a skin detection procedure in $YC_bC_r$ space.

The flowchart in Fig. 4a shows the skin detection scheme using a combination of RGB and HSV (RGB+HSV) spaces of colors.

Given an input image, after being converted into HSV space, return a mask with all the pixels considered as skin in RGB+HSV space. This method runs morphological operations on the image, as for example the histogram equalization and closure (made with a cross kernel 6x6), with the purpose of improving the output quality. The choice to classify a pixel as skin occurs through a thresholding procedure, that consists to verify more or less complex logical expressions on the values of individual image channels. If a pixel verifies the conditions expressed through, it is a pixel of skin and so its colour is blank. The thresholding procedure uses rather complex logical rules, that examine all the components (R, G, B, H, S and V). The main expressions are in [19] and [35]. Figure 4b shows the resulting image after a skin detection procedure in RGB+HSV space. Comparing Fig. 3a and b we can see that RGB+HSV thresholding appears less noisy than $YC_rC_b$ even if the detection of the skin is very good.

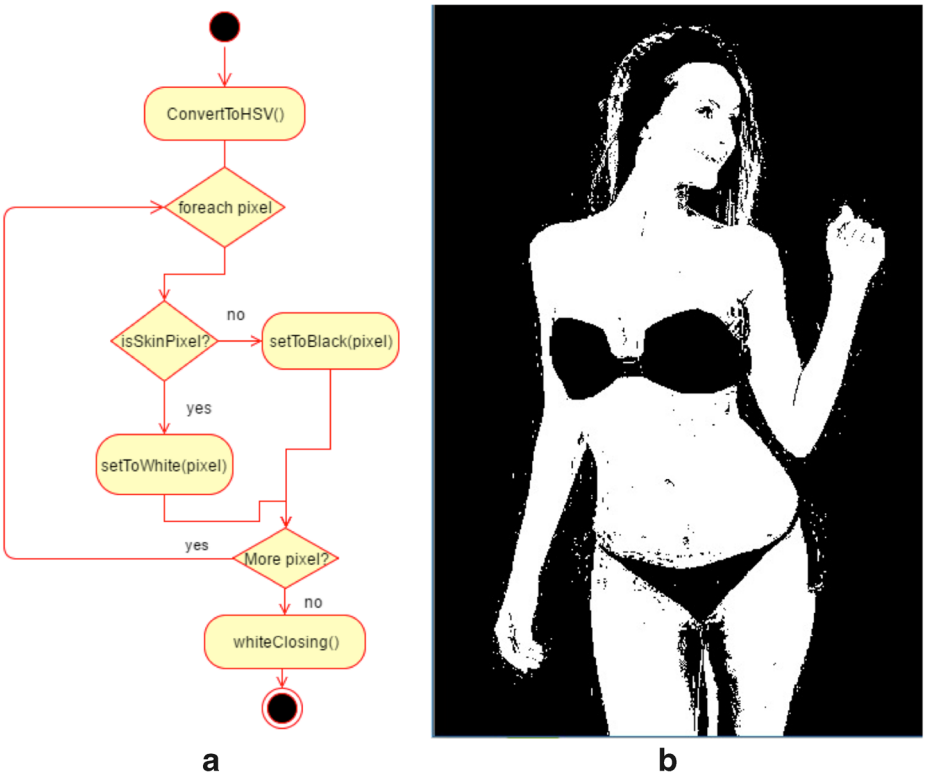


**Fig. 4** **a** Represents the diagram of skin detection using RGB+HSV space. **b** is the result of the application of RGB+HSV thresholding rules

### 3.3.2 Feature extraction

The next step is to find the contours (*findContours* in Fig. 5a) of the image through a chain-code algorithm. Then are determined the features, departing from the detected regions. After the calculation of features, a verify is made on the *isBad* attribute of the region. If *isBad* is true, all the pixels of the region become blacks. The attribute *isBad* is set to true if the area of the region does not exceed a threshold (i.e. the region is too small to be considered when the total area in pixel is less than the 2% of the total segmented area).

The output of the system is the region with the largest area and finally the percentage of skin pixels in relation to the whole image. Taking into account other papers concerning classification of pornographic images [3, 22, 28] finally overall ten features are extracted. Nine features are extracted from the maximum region and more the feature that describes the overall percentage of the skin in the image. Each feature is normalized between 0 and 1, to be processed by the next classification phase. The feature vector consists of the following features:

– Percentage of skin of the largest region.

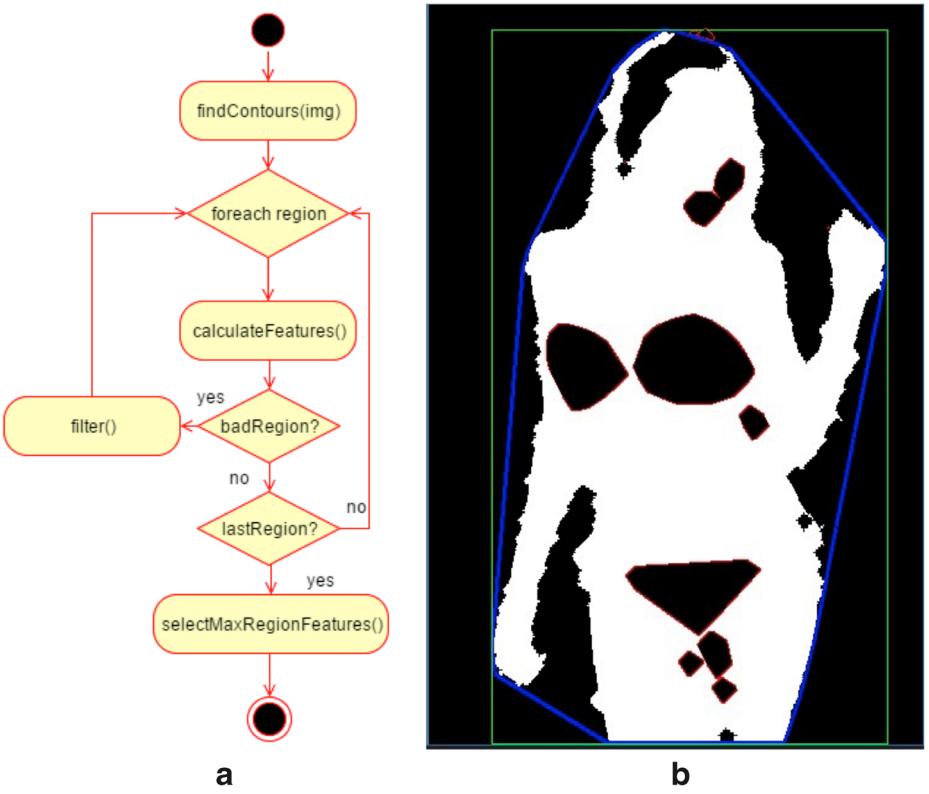


**Fig. 5** **a** Represents the diagram of of feature extraction. **b** is an example of application of geometric filtering

– Compactness of the largest region:

$$Compactness = \sqrt{\frac{4\pi(Regionarea)}{(RegionPerimeter)^2}}$$

– Rectangularity of the largest region

$$Rectangularity = \frac{(Regionarea)}{(RegionBoundingRectangleArea)}$$

– The average values of R, G and B of the largest region.
– The standard deviation of R, G and B of the largest region.
– The percentage of skin in the image.

Figure 5b shows the bounding boxes after the filtering in an image with large portions of skin.

#### 3.3.3 Classification

The last procedure concerns the classification phase that is based on Support Vector Machine (SVM) algorithm [37], also named at the maximum margin since the algorithm is designed to find the hyperplane that maximizes the separation margin between the classes. SVMs are binary supervised classification models (also extended to the multiclass case) aiming at identifying the geometric place of points in the space, the hyperplane, that separates between them examples belonging to different classes. In our work, the SVM classifier has in input three parameters: the first is the matrix of the actual data, the second one allows to choose the layout of the training data (that is, if the feature vector are represented as rows or columns in the data matrix) and finally the third represents the label matrix associated with the data. The parameters of classification are automatically chosen by the algorithm: a k-fold procedure is performed inside the method, designed to maximize the parameters of the classifier given the input training set. The classifier indicates if an image is or not pornographic.

## 4 Results and discussions

This section reports results of multimedia classification experiments. Along with the performance of the chosen classifier, tests on real scenario are discussed, done in cooperation with enforcement law officers.

The datasets used in the test are designed to prove that even with a relatively small dataset and standard method sufficient results for the specific forensic use case can be achieved with the proposed architecture and the overall design concept. In the day by day forensic activities the system can be trained on novel dataset for a specific use case in a fast and affordable way (i.e. from pornography detection to anti-terrorism investigations, from copyrighted materials to pedophilia).

### 4.1 Pornography classification results

This section shows the results of the performances of a SVM classifier on only one feature, extracted from an image on which a $YC_rC_b$ thresholding was executed. This approach

**Table 1** Experimental results with 1200 examples and $YC_rC_b$ space of colors

| N-th testing set | Accuracy | Precision | Recall | Exec time (s) |
|---|---|---|---|---|
| 1 | 0.695833 | 0.393162 | 0.958333 | 5.67 |
| 2 | 0.633333 | 0.350746 | 0.979167 | 4.97 |
| 3 | 0.579167 | 0.306569 | 0.875 | 5.13 |
| 4 | 0.604167 | 0.339035 | 0.958333 | 5.25 |
| 5 | 0.675 | 0.375 | 0.9375 | 6.79 |

is the one used also on real test, described in the next paragraph. The test has been executed on a machine with the subsequent configuration: CPU: Intel i7-4700MQ 2,4 GHz, RAM: 12 GB, OS: Windows 10 - 64bit. In this case, the SVM classifier, given the low dimensionality of the feature vector, is based on a INTER type kernel. The dataset used consists of 1200 examples, of which: 240 positives (pornographic images) and 960 negatives (non-pornographic images). The validation technique is Cross Validation, with k=5. Then, there are 240 examples for fold, of which, always using a layered approach: 48 positives and 192 negatives. As indicated in the table, the solution has a very high **Recall** (just over 94%), compared to a discrete mean value of **Accuracy** (63% average). Although the performance looks very good, this model has a serious defect: under fitting. A single feature is not sufficient to create a reasonably complex model that well generalizes not classified data. Numerous attempts were made to train the model with dataset of different sizes. Moreover, various cross-validation operations were attempted, trying to change the quantity of examples in the fold or the number of fold itself. By executing one of these modifications, the model is no longer able to achieve the behavior shown in the table, going to classify all the examples as negative. In conclusion, this solution must be discarded (Table 1).

The second experiment involves a SVM classifier on 10 features, extracted from an image on which a RGB + HSV thresholding was executed. In this case, the SVM classifier, given the high dimensionality of the feature vector, is based on a LINEAR type kernel, which is a function that does not perform any mapping in a higher-dimensional space. A good separability of data is guaranteed by the high number of features. The dataset used consists of 1000 examples, of which: 500 positives and 500 negatives. The validation technique is Cross Validation, with k=5. Then, there are 200 examples for fold, of which, always using a layered approach: 100 positives and 100 negatives.

Observing the results in Table 2, we obtain a mean value of **Recall** less than the previous approach (from 94% to 88% of this solution), but we have a better value concerning the mean value of **Accuracy** equals to 76%. Moreover, this solution does not suffer of underfitting problem as the previous $YC_rC_b$ solution since the model appears more discriminant

**Table 2** Experimental results with 1000 examples and RGB+HSV space of colors

| N-th testing set | Accuracy | Precision | Recall | Exec time (s) |
|---|---|---|---|---|
| 1 | 0.765 | 0.7086661 | 0.9 | 7.29 |
| 2 | 0.78 | 0.741379 | 0.86 | 7.37 |
| 3 | 0.765 | 0.730435 | 0.84 | 9.74 |
| 4 | 0.775 | 0.735043 | 0.86 | 8.78 |
| 5 | 0.725 | 0.661871 | 0.92 | 10.22 |

thanks to the number of selected features. The last column (**Exec time**) denotes that the application has a mean execution time less than 9 seconds. Then, just for the purpose and the last end of our application, we can state that it is an appealing tool for a real-time and in-loco check. Experimental results show that this application satisfies better the initial requirements (Table 2).

### 4.2 Real case results

The final results reported in this paper are those coming from real tests on real scenario and done in cooperation with enforcement law officers.

To determine the effectiveness of the exposed methods and architecture with respect of manual scenario a comparison of different inspections was performed: a manual analysis vs a automatic one was conducted in 3 different real cases with a final reduction of inspection time of -94% on average on a medium size memory of 10TB with a multimedia content rate between 22% and 28%. The average number of collected evidences in the 3 pornography inspection cases was 38 images or video frames that was sufficient for evidence reporting and to start the second phase of the digital forensic inspection process. All the process was conducted off line with no further parameters settings. False positive rate during the manual evidence check was, on average, the 22% of the total number of evidence images selected by the tools. Usability, fast inspection processing and the priority to recall versus accuracy were the main motivations of these promising results, together with a low cost and easy to manage parallel hardware architecture.

## 5 Conclusion and future works

Digital forensic investigations are often required to identify, process, and analyse a consistent amount of heterogeneous multimedia contents in order to achieve valuable information and insightful knowledge that can allow them to rapidly react to a crime. In this work, a multimedia classification tool is advocated together with a parallel software architecture for a fast inspection, which is easy to use (to be used by officers during a search). It requires limited hardware resources and it is based on an open-source software valuable to limit its costs. Furthermore, this tool must allow a quick inspection of multiple devices at a time. Specifically, a set of fast and high recall ML and DL algorithms are adopted to cover the most common inspection cases. The experiments in real cases indicate that our tool is well-suited for forensic purposes. In particular, the tests described have demonstrated the success and the suitability of the provided process and for the inspection time reduction.

Further investigation will involve the improvement of the approach's robustness and the conduction of controlled tests on real-world cases. Moreover, several functionalities will be added such as automatic image (and probability map) segmentation. A larger dataset will be built for a more robust and reliable prediction system. This will allow the researchers to evaluate which are the resources necessary for the Convolutional Neural Network (CNN) layers planning. The following studies could also concern cybersecurity and cryptography taking into account virtualization, standardization of technologies, and specific regulations for protecting personal data. A future development of this work will be also devoted to the development of a standard method for data collection with an available public dataset. Since the multimedia classification architecture is presented in multiple scenarios (not only pornography), several experiments in domains like terrorism (i.e. looking for special symbols), illegal trafficking of copyrighted materials (i.e. searching for specific contents like last

published films or games), cultural heritage goods (i.e. looking for particular masterpiece pictures), etc. All these use cases can leverage on the proposed architecture and methods by using fast training procedures and automatic and fast methods to effectively search for positives in an inspected device.

**Acknowledgments** Authors thank the "Compartimento Polizia Postale e delle Comunicazioni di Ancona" for inspiring this work and testing the proposed solution in real scenarios.

**Funding** Open access funding provided by Università Politecnica delle Marche within the CRUI-CARE Agreement.



## References

1. Al Mutawa N, Bryce J, Franqueira VN, Marrington A (2015) Behavioural evidence analysis applied to digital forensics: an empirical analysis of child pornography cases using p2p networks. In: 2015 10th international conference on availability, reliability and security. IEEE, pp 293–302
2. Amerini I, Caldelli R, Del Mastio A, Di Fuccia A, Molinari C, Rizzo AP (2017) Dealing with video source identification in social networks. Signal Process Image Commun 57:1–7
3. Basilio JAM, Torres GA, Pérez GS, Medina LKT, Meana HMP (2011) Explicit image detection using ycbcr space color model as skin detection. Appl Math Comput Eng: 123–128
4. Bissias G, Levine B, Liberatore M, Lynn B, Moore J, Wallach H, Wolak J (2016) Characterization of contact offenders and child exploitation material trafficking on five peer-to-peer networks. Child Abuse & Neglect 52:185–199
5. Chai D, Ngan KN (1999) Face segmentation using skin-color map in videophone applications. IEEE Trans Circuits Sys Vid Technol 9(4):551–564
6. Chatzis V, Panagiotopoulos F, Mardiris V (2016) Face to iris area ratio as a feature for children detection in digital forensics applications. In: 2016 Digital media industry & academic forum (DMIAF). IEEE, pp 121–124
7. Choi B, Chung B, Ryou J (2009) Adult image detection using bayesian decision rule weighted by svm probability. In: 2009 Fourth international conference on computer sciences and convergence information technology. IEEE, pp. 659–662
8. Chopra M, Martin MV, Rueda L, Hung PC (2006) Toward new paradigms to combating internet child pornography. In: 2006 Canadian conference on electrical and computer engineering. IEEE, pp 1012–1015
9. Fanchang H, Xu C, Gongping Y, Lu Y, Chengdong L, Chenglong L, Chuanliang X (2020) Local image quality measurement for multi-scale forensic palmprints. Multimed Tools Appl: 1–24
10. Giudice O, Paratore A, Moltisanti M, Battiato S (2017) A classification engine for image ballistics of social data. In: International conference on image analysis and processing. Springer, pp 625–636
11. Grega M, Bryk D, Napora M (2014) Inact—indect advanced image cataloguing tool. Multimed Tools Appl 68(1):95–110
12. Grega M, Matiolański A., Guzik P, Leszczuk M (2016) Automated detection of firearms and knives in a cctv image. Sensors 16(1):47
13. Hsieh S, Fan KC, Lin C (2002) A statistic approach to the detection of human faces in color nature scene. Pattern Recogn 35(7):1583–1596
14. Iuliani M, Shullani D, Fontani M, Meucci S, Piva A (2018) A video forensic framework for the unsupervised analysis of mp4-like file container. IEEE Trans Inf Forensics Secur 14(3):635–645

15. Jang DM, Turk M (2011) Car-rec: A real time car recognition system. In: 2011 IEEE workshop on applications of computer vision (WACV). IEEE, pp 599–605
16. Jang EG, Koh BS, Choi YR (2012) A study on block-based recovery of damaged digital forensic evidence image. Multimed Tools Appl 57(2):407–422
17. Jenkins R, Kerr C (2013) Identifiable images of bystanders extracted from corneal reflections. PloS one 8(12)
18. Kamenicky J, Bartos M, Flusser J, Mahdian B, Kotera J, Novozamsky A, Saic S, Sroubek F, Sorel M, Zita A, et al. (2016) Pizzaro: Forensic analysis and restoration of image and video data. Forensic Sci Int 264:153–166
19. Kovac J, Peer P, Solina F (2003) Human skin color clustering for face detection. In: EUROCON 2003. Computer as a Tool. The IEEE Region 8, vol 2. IEEE, pp 144–148
20. Kumar A, Kansal A, Singh K (2019) An improved anti-forensic technique for jpeg compression. Multimed Tools Appl 78(18):25427–25453
21. Li J, Ma B, Wang C (2018) Extraction of prnu noise from partly decoded video. J Vis Commun Image Represent 57:183–191
22. Lin YC, Tseng HW, Fuh CS (2003) Pornography detection using support vector machine. In: 16th IPPR conference on computer vision, graphics and image processing (CVGIP 2003), vol 19, pp 123–130
23. Maksymowicz K, Tunikowski W, Kościuk J (2014) Crime event 3d reconstruction based on incomplete or fragmentary evidence material–case report. Forensic Sci Int 242:e6–e11
24. Mofaddel MA, Sadek S (2010) Adult image content filtering: A statistical method based on multi-color skin modeling. In: 2010 2nd International conference on computer technology and development. IEEE, pp 682–686
25. More LG, Brizuela MA, Ayala HL, Pinto-Roa DP, Noguera JLV (2015) Parameter tuning of clahe based on multi-objective optimization to achieve different contrast levels in medical images. In: 2015 IEEE International conference on image processing (ICIP). IEEE, pp 4644–4648
26. Pandey RC, Singh SK, Shukla KK (2016) Passive forensics in image and video using noise features: A review. Digit Investig 19:1–28
27. Peersman C, Schulze C, Rashid A, Brennan M, Fischer C (2014) icop: Automatically identifying new child abuse media in p2p networks. In: 2014 IEEE security and privacy workshops. IEEE, pp 124–131
28. Platzer C, Stuetz M, Lindorfer M (2014) Skin sheriff: a machine learning solution for detecting explicit images. In: Proceedings of the 2nd international workshop on Security and forensics in communication systems. ACM, pp 45–56
29. Ricanek KJr, Boehnen C (2012) Facial analytics: from big data to law enforcement. Computer (9):95–97
30. Sae-Bae N, Sun X, Sencar HT, Memon ND (2014) Towards automatic detection of child pornography. In: 2014 IEEE International conference on image processing (ICIP). IEEE, pp 5332–5336
31. Sayed U, Sadek S, Michaelis B (2009) Two phases neural network-based system for pornographic image classification. In: Proceedings of 5th international conference of sciences of electronic, technologies of information and telecommunications (SETIT2009), pp 1–6
32. Schulze C, Henter D, Borth D, 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, pp 353–360
33. Shen X, Wei W, Qian Q (2010) A pornographic image filtering model based on erotic part. In: 2010 3rd International congress on image and signal processing, vol 5. IEEE, pp 2473–2477
34. Shupo A, Martin MV, Rueda L, Bulkan A, Chen Y, Hung PC (2006) Toward efficient detection of child pornography in the network infrastructure. IADIS Int J Comput Sci Inf Sys 1(2):15–31
35. Smolka B, Czubin K, Hardeberg JY, Plataniotis KN, Szczepanski M, Wojciechowski K (2003) Towards automatic redeye effect removal. Pattern Recogn Lett 24(11):1767–1785
36. Ulges A, Stahl A (2011) Automatic detection of child pornography using color visual words. In: 2011 IEEE international conference on multimedia and expo. IEEE, pp 1–6
37. Vapnik V (2013) The nature of statistical learning theory. Springer Science & Business Media
38. Villalba LJG, Orozco ALS, López RR, Castro JH (2016) Identification of smartphone brand and model via forensic video analysis. Expert Syst Appl 55:59–69
39. Wang D, Gao T, Yang F (2018) A forensic algorithm against median filtering based on coefficients of image blocks in frequency domain. Multimed Tools Appl 77(18):23411–23427
40. Wang F, Hu L, Hu J, Zhao K (2016) Computer forensic analysis model for the reconstruction of chain of evidence of volatile memory data. Multimed Tools Appl 75(16):10097–10107
41. Zaidan A, Ahmad NN, Karim HA, Larbani M, Zaidan B, Sali A (2014) On the multi-agent learning neural and bayesian methods in skin detector and pornography classifier: An automated anti-pornography system. Neurocomputing 131:397–418

42. Zeng L, Chen J, Tong L, Yan B, Ping X (2013) Image contrast enhancement based on histogram similarity. In: 2013 IEEE International conference on medical imaging physics and engineering. IEEE, pp 269–273
43. Zhao Z, Cai A (2010) Combining multiple svm classifiers for adult image recognition. In: 2010 2nd IEEE internationalconference on network infrastructure and digital content. IEEE, pp 149–153
44. Zuo H, Hu W, Wu O (2010) Patch-based skin color detection and its application to pornography image filtering. In: Proceedings of the 19th international conference on World wide web, pp 1227–1228