

An overview of NuDetective Forensic Tool and its usage to combat child pornography in Brazil

Pedro Monteiro da Silva Eleuterio and Mateus de Castro Polastro

Brazilian Federal Police

Abstract. In many countries, the possession of files containing child and teen pornography is a heinous crime and is desirable for law enforcement be able to detect such files in a timely manner at crime scenes. However, mainly at crime scenes, it is impossible to manually examine all files that can be stored in digital storage devices. The NuDetective Forensic Tool was developed to assist forensic examiners to identify child pornography at crime scenes. NuDetective uses automatic nudity detection in images and videos, file name comparison and also uses hash values to reduce the files to be analyzed by forensic examiners. Despite the high detection rates achieved in past experiments, the authors did not get any formal feedback of NuDetective users about its performance in real forensic cases. So, this work presents a detailed review of the four main features of NuDetective Forensic Tool, including all techniques and methods implemented, and also the results of an unpublished survey conducted to evaluate the real effectiveness of NuDetective by its Brazilian users. The results obtained showed that NuDetective is helping to arrest pedophiles and to combat the child sexual exploitation in the digital era.

Keywords: child pornography, nudity detection, digital crime scene

1 Introduction

The crimes of sexual exploitation of children increase every day, especially with the use of the Internet to exchange child pornographic files. A recent report by the Brazilian Federal Police [1] shows an increase of 127% in “red-handed” arrests in 2013 in cases of child pornography in Brazil, compared with the previous year, as shown in Figure 1. The report also reported the increasing number of cases investigated on the subject, which grows every year. Data extracted from SaferNet Brazil [2] show that in the last eight years, were received over 1 million complaints regarding child pornography in websites, which were hosted in 90 countries spread across the five continents on earth. Therefore, the sexual exploitation of child is a global crime and, unfortunately, is growing every year.

To combat this crime, the possession of these files is already a crime in many countries [3]. Countries such as Canada, United States, United Kingdom, Australia, and Brazil [4], have legislation to criminalize people who store these files on any digital storage device. Consequently, requests for forensic examination to identify illegal content are becoming increasingly common, including at crime scenes.

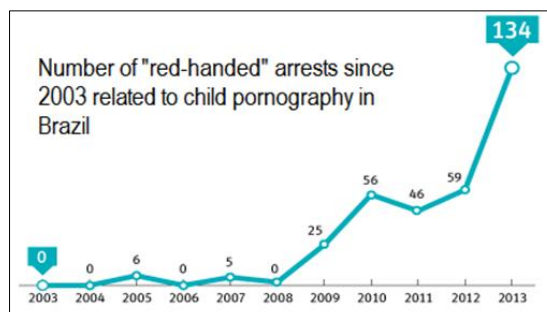


Fig.1.Number of “red-handed” arrests in crimes related to child pornography in Brazil in the last 11 years.

With this prior motivation, the first version of NuDetective Forensic Tool was developed in 2010 to help the forensic examiners to find child and teen pornographic files in a timely manner at crime scenes [5]. Designed to be used mainly at crime scenes, NuDetective has the main objective to be quick enough to scan digital storage devices in a few minutes, identifying suspicious files related to child pornography, which must be analyzed by the forensic examiners.

The tool, which is free, was sent upon request to more than 40 entities in Brazil and seven around the world, including law-enforcement and public forensic laboratories, public ministry, public universities, among others. Despite the experiments had proved the effectiveness of the tool, the authors did not have any formal user feedback about the real usage of NuDetective in real forensic cases. Therefore, a study with Brazilian users of the tool was conducted in order to verify in practice its use and results obtained in real forensic cases. In addition, the user opinions are of great importance to the choice of the new features and improvements that could be developed in the near future.

This paper presents the unpublished results of the conducted survey, beyond a detailed review of the main features of NuDetective, including the background concepts and all the techniques used to identify suspicious files related to child pornography.

The paper is organized as follows: Section 2 presents some background concepts, like automatic human nudity detection in images and also pornography detection in videos. Section 3 presents an overview of NuDetective Forensic Tool, its main features and the techniques used to automatic detect child pornography. Section 4 shows the conducted survey with Brazilian users about the usage of NuDetective and its results in practice. Sections 5 and 6 present the conclusions and future work, respectively.

2 Background

In this section, key concepts related to automatic human nudity and pornography detection are reviewed, including the skin color detection in images, which is a prior step for human nudity detection. This section also reviewed some particular aspects of video files and the state-of-art in pornography detection in videos and a quick review

about file names and peer-to-peer networks. All these ideas are important to understand the techniques implemented in NuDetective, which is shown in details in Section 3.

2.1 Skin Color Classifier

The pixel classification of digital images to determine human skin regions tries to identify each pixel as skin or not. There are many difficulties to classify skin pixels, because many factors affects its color, including the use of different digital cameras, scanners and camcorders, which can assign different pixel colors of the same person in the same conditions of light. Other difficulties are the variable ambient light, such as sunlight, artificial light, use of flashlights, and the fact that skin color naturally varies from person to person [9]. To complicate a bit more, wood, leather, hair, and sand, among many other things, have similar digital color to the human skin, which can confuse automatic skin detectors in relation to the image background [10].

The human skin color mainly depends on the concentration of melanin, hemoglobin, and the lighting condition. It is not located randomly within a color space [11], is located within small areas of the color space, and it depends on what color space used by the classifier [10]. Thus, the state-of-art presents many conducted studies to determine ways of classifying skin color pixels among existing color spaces, such as RGB [12], HSI / HSV / HSL [13], and also orthogonal color spaces (YCbCr, YIQ, YUV, and YES) [14].

A study [15] has proposed a method for detecting skin color using the RGB color space. It consists of a process with simple steps and rules that allows constructing classifiers quickly [13]. A basic arithmetic equation can describe the relationship among the Red, Green, and Blue, as shown in (1) below:

$$R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and } \max R, G, B - \min R, G, B > 15 \text{ and} \quad (1)$$

$$R - G > 15 \text{ and } R > G \text{ and } R > B$$

Other researchers used different color spaces to define a skin pixel classifier. In [16], areas of color space belonging to the skin color were obtained using the normalized RGB and HSV color spaces from (2) and (3) below:

$$0.36 \leq r \leq 0.465, \quad 0.28 \leq g \leq 0.363 \quad (2)$$

$$0 \leq H \leq 50, 0.20 \leq S \leq 0.68, 0.35 \leq V \leq 1.0 \quad (3)$$

In [17], the YCbCr color space was used to divide the regions of skin color through the chrominance information. They achieved high rates of detection in images with good lighting conditions.

2.2 Nudity Detection in Images

The first phase in the nudity detection in images is the identification of skin pixels, as discussed before. Once identified these pixels, the next phase is to determine if the

regions bounded by skin color match to shapes of the human body. There are known methods in the literature that handle image features to detect human nudity. Also, there are techniques that compare the segmented images by skin color with trained database of pre-classified images.

One study [18] presents various ways to accomplish the nudity detection by extracting image features, like comparing the size of the largest skin region found with the image total size. The number of identified skin areas is another important attribute. Small human skin regions in large quantities are not common in nudity images. The percentage of skin pixels is also an indicative of the presence of human nudity [18].

Another research [19] also used extraction of image features to determine human nudity. It considers that nudity contains a several different skin tones and these regions are relatively close to each other. First, the suggested approach calculates the percentage of skin pixels among the total image pixels. Through empirical experiments, a threshold higher than 15% was set, so the image with more than 15% of skin pixels continues to be evaluated by the process [19]. Otherwise, it defines that the image has no nudity. Then, using computational geometry, the three major regions of skin color is used to perform amount calculus, determining if the image contains human nudity or not, based on the comparison of thresholds empirically obtained.

Other techniques use an image database to train the nudity classifier. One method [20] uses Support Vector Machine (SVM), which can learn about image detection of human nudity. In the model, is first verified the correlation among regions containing skin color or not, to train the SVM. Then, the nudity classifier uses this information to determine if a new image contains nudity or not, based on information acquired during the previously training phase. Another technique [21] proposed to compare the suspicious images with a set of features obtained from training database analysis.

2.3 Nudity and Pornography Detection in Videos

There are many ways to perform automatic nudity and pornography detection in videos. The literature shows many techniques to achieve this aim. There are some techniques based in video frame extraction. Motion segmentation and motion analysis are also reported in the literature, among other techniques depending on the objective of the research.

Poison and Tjoa [22] presented a review of the state-of-art of forensic investigations of multimedia data. For automatic classification of videos, it emphasizes the usage of motion analysis and keyframes as the main methods. Other methods, like periodicity detection (PER), the sliding window periodicity (PERWIN), motion histograms, and the bag-of-visual-words (BOVW) are also reported as methods for pornography detection in video files.

Nudity and pornography detection in videos using frame extraction was used by various studies [23] [24] [25]. The algorithm developed in [24] tries to identify global motions, identifying and isolating the foreground objects, classifying the frames through shape, skin color, and texture features. It also compares the identified objects with a previously trained database, by using the weighted Euclidean distance method.

Through methods known as bag-of-visual-features (BOVF), which was previously developed for nudity detection in images [26], Lopes et al. [23] used descriptors (a gray-level SIFT and hueSIFT), creating a point descriptor considering the color hue. Histograms were built with amount training images. A voting system determines if the video has nudity or not.

Other research [25] chooses keyframes when changes occur in the video scene. It tries to identify standards in the scene, such as dissolve, wipe, fade in, and fade out. The algorithms try to identify skin color, removing textures and unwanted parts of the image, before determining if the video has nudity or not. It reported that the developed method took about one tenth of the video length to classify it [25].

Unlike initiatives that use video frame extraction and subsequent detection in images, other studies used motion to determine pornography in videos [27] [28] [29]. One study [27] identified repetitive movements, calculating what they called Periodicity Detection (PER), naturally present in videos of sex, and generated motion histograms. The next step is to combine the frame static analysis, including bag-of-visual-words (BOVW), with the results of motion analysis to determine if video has pornography or not. Other work [28] divided the video file into small fixed-length segments and created motion vectors for each extracted frame. The algorithms detect repetitive motion during 16 second intervals, determining if videos have pornography or not. Another study [29] also used motion segmentation and built motion vectors to identify what they called "homogeneous motion" in videos, which was used to determine if a video has pornography [29].

Audio streams were also used to measure audio periodicity. A study [30] used audio to recognize sounds of pornography. The developed framework combines the results from audio and video classifiers to determine if a video file has pornography or not. That framework also uses extracted frames from videos to identify pornography using an image recognition algorithm, as previously shown.

2.4 File Names and Peer-to-peer Networks

There are many programs to share files on peer-to-peer networks [32], like eMule, Kazaa and Ares, among others. Although they are different, the way to conduct searches for files is very similar in all of them: the user must use keywords to search for related files. The important point is that all the results files often contain the desired keywords in their names. Thus, after downloading the interest files, they will be stored on the storage device of the used computer with the same names that were available on the network.

The authors' experience shows that the vast majority of child pornographic files downloaded through peer-to-peer programs contain regular expressions of pedophilia. A study [33] showed 26 explicit expressions commonly used by pedophiles to look for such files on the Internet, like "hussyfan", "kdquality", and "kidzilla". Another important point is that words often used by pedophiles to seek this file type on P2P networks change over time [34]. Thus, the comparison of file names, searching for typical keywords of pedophilia, can be a very interesting and efficient way to find the illegal files. The search list must be always updated. However, in a hard disk may be

many filenames that include some expressions related to pedophilia that are not related to the subject (false positives). That can occur because some pedophilia keywords may include few characters, like "pthc" or "pedo", which may be part of a name of a common file.

3 The NuDetective Forensic Tool

This Section provides a review of the main features of the NuDetective Forensic Tool [5] [6] [7] [8] to illustrate its basic operation, including its pixel analysis and automatic human nudity detection algorithms in images [5] [6]. It also presents the techniques used to detect videos of child pornography [7] [8], beyond other important features and developed techniques.

3.1 Basic Features

The NuDetective Forensic Tool was developed to help forensic examiners to identify files of child pornography in a timely manner at crime scenes. To develop NuDetective, the authors considered two main principles: quick processing and reduced number of false negatives, which is important in cases evolving child pornography.

Developed with the Java Standard Edition (JSE), NuDetective has four main features (Image Analysis, FileName Analysis, Hash Analysis, and Video Analysis) and some other features that can be configured by the forensic examiner. NuDetective can be run also in a single thumb drive, needing only the JRE (Java Runtime Environment) and some libraries (jar files). The Tool can be easily configured to display texts in Brazilian Portuguese, English and Spanish languages. New languages can be added in the future without the need of new implementation.

The basics steps to use NuDetective are: the forensic examiner chooses folders and/or entire logical partitions to search, configures search options (which is an optional task because NuDetective starts with all recommended options for using at crime scenes), and starts the scan. Thus, the NuDetective searches for suspicious files related to child pornography, submitting them to the internal detection algorithms. The main features are Image Analysis, which detects human nudity in images; Video Analysis, which identifies videos with child pornography content; FileName Analysis, which verifies the file and path names for common phrases of pedophilia, and; Hash Analysis, which calculates the hash value of suspicious files and compare them with a user list of hash values. All these four features are described in the next subsections. It is possible to use the main four features together, which the forensic examiners could achieve the best result in detection of child pornography.

The authors used Java threads in the development of NuDetective, allowing parallel execution of some tasks. This construction allows the forensic examiner to start the analysis of suspicious files before the tool finishes searching all files of the selected paths and/or logical partitions.

3.2 Image Analysis

The first main idea of developing NuDetective starts with the Image Analysis feature. Using human nudity detection in the search for suspicious files of child pornography was an innovation in forensic tools at that time. The NuDetective Forensic Tool can automatically detect nudity in images [5], using the RGB color space [12] for representing digital color space. According to a survey conducted in [14], RGB is the most common color space representation used for storing the digital image.

The nudity detection algorithms implemented in NuDetective use three simple steps. The first one is to classify each pixel of the image as human skin color or not. NuDetective use the method developed by Kovac et al [15], as shown in (1) described in Section 2.1.

Once detected the skin pixels, the second step is to verify if the image has human nudity or not. The authors implemented their own detection algorithms based on the Ap-apid [19] technique for nudity detection. These algorithms do not require one previous trained image database to classify them, because the algorithms use percentage of skin color and computational geometry to determine if the image has human nudity or not. One new implementation of these algorithms does not imply that the results will be similar to the original one, due to the natural differences in the authors' implementation. The Ap-apid [19] algorithms were chosen because they are computationally fast and work as a binary classifier, i.e., after analysis the results will be yes or no for human nudity.

After the analysis of each image, the third step is to immediately feed the results GUI with suspicious image, allowing the forensic examiner to analyze it.

The authors conducted several experiments to verify the effectiveness of the Image Analysis provided by the tool, using a controlled database and a real forensic case [5]. The results showed high automatic nudity detection rates in images, low false negatives, and quick processing. In a real forensic experiment, an entire 80GB hard disc drive was scanned in less than only 13 minutes [5]. The obtained values for recall and precision were 95.4% and 93.0%, respectively, which is a satisfactory result, considering the execution time was as minimal as possible.

The NuDetective has some optional specific settings for Image Analysis, including the option of Maximum Optimization, and minimum/maximum sizes of images to be submitted to the nudity detection algorithms. Maximum Optimization was a result of the work previously published by the authors [7], which could reduce up to 90% the runtime of nudity detection in high-resolution images. The strategy used reduces the high-resolution image to 0.01 Megapixel before submitting them to the nudity detection algorithms. The experiments proved that the variation of nudity detection in results is minimal and the runtime is up to 90% lower. The Maximum Optimization is enabled by default and its use is always recommended, especially at crime scenes. The other parameters of image size (minimum and maximum) filter the images to be analyzed. When the maximum value is set to zero (default value), there is no maximum size limit.

3.3 Video Analysis

In the literature, as described in subsection 2.3, exists many studies related to pornography detection in video, using different computational methods. All of them try to achieve a higher detection rate as possible and require much processing time. So, besides being excellent techniques to identify pornography in videos, they were not specifically developed to be used at crime scenes. Also, they are not specific for child pornography, but they try to identify videos of nudity or adult pornography. NuDetective Video Analysis uses the algorithms provided by its Image Analysis. However, all the empirically experiments to determine the parameters of the algorithms were conducted with real child pornographic videos. That means that the method described in this subsection is optimized to detect child pornographic videos and may not work properly to identify adult pornography.

Thus, the Video Analysis feature provided by NuDetective tries to quickly identify videos of child pornography [8] [9]. First, many experiments with 181 videos of child pornography were carried out to understand the behavior of these files, including the nudity distribution in the video frames versus the video length. The objective of the authors were to calculate a Cutoff Point (CP), which indicates the minimum percentage of nudity frames a video must have to be classified as suspicious. Thus, using the internal algorithms of Image Analysis for each video frame, the authors calculated the mean of nudity frames for each video and the mean of nudity frames of all 181 video means (X). After that, the authors determined the standard error (E), and the Cutoff Point (CP) of 65.3% ($CP = X - E$).

After the definition of the Cutoff Point of 65.3%, the authors needed to create a strategy to sample video frames in runtime, since processing all the video frames requires a long time, being unfeasible for NuDetective at crime scenes. The first strategy [8] used a fixed number of frames for each video: 10 video frames per each video and if more than six were classified as nudity, the video was classified as suspect. However, besides achieved high detection rates and low runtime, the first approach was not the best, because it delayed the detection process, especially in shorter videos, which is the majority in videos of child pornography.

So, the second and actual strategy provided by NuDetective to sample video frames was presented in [9]. It used different frame sample sizes for different video lengths and achieved better detection rates and runtimes. The authors performed several experiments with different sampling strategies in two sets of videos, allowing the definition of a new formula based on the logarithmic equation, as shown in (4).

$$y = 3 \log_{10}\left(\frac{x}{fps}\right) \quad (4)$$

In equation (4), x is the total number of video frames, fps the video temporal resolution, and y the rounded up number of frames that NuDetective will sample for the video. With that new formula, the Video Analysis provided by NuDetective achieved values for precision and recall of 85.7% and 87.3%, respectively. Also, the performed experiments showed that Video Analysis needed only an average of 1.14 seconds per video to classify as child pornography or not [9].

Therefore, for each video to be analyzed by NuDetective, the Video Analysis first calculates and extracts the number of video frames. These extracted frames are submitted to the Image Analysis algorithms to determine how many frames contain human nudity. It calculates the percentage of nudity frames of the video, comparing the result to the Cutoff Point (CP) of 65.3%. If the percentage of nudity frames of the video is greater than the CP, the Video Analysis classifies the video as suspect and immediately feeds the results GUI of NuDetective.

3.4 FileName Analysis

The FileName Analysis provided by NuDetective allows comparison of file names with an internal list of known keywords used by pedophiles for file sharing over the Internet. The internal list of keywords includes some expressions in Portuguese, English and Spanish too, and the 26 expressions listed in a study [33]. The NuDetective also allows an external keyword list to be used for this comparison. This list should contain one name per line in a text file, without the use of commas and periods. The internal list of names is the default option.

It is possible to configure NuDetective to compare the names of all files found in selected folders and/or logical partitions, or just images and videos, or only images/videos considered suspicious by the Image Analysis (if enabled) and Video Analysis (if enabled). FileName Analysis is quick because does not require reading the content of the files. Another possible configuration is to include the complete path as the filename or only the file name, without the path (default value). The authors' experience shows that folders name rarely contains such expressions.

3.5 Hash Analysis

Another technique to identify illegal files widely used in forensic computing is the use of hash functions. The hash is a value calculated from a one-way mathematical function [31] and serves as a unique identifier of the file. Commercial forensic tools such Access Data Forensic Toolkit and Guidance Encase use the hash functions.

The Hash Analysis provided by NuDetective [7] allows the forensic examiner to import a text file, which contains the list of previously known hash values, that will be compared with the files present in the selected folders and/or disk logical partitions. The text file must contain one hash value per line. The NuDetective also offers the option to calculate the hash values of only image/videos files, which is much more interesting, considering the main purpose of the tool. The forensic examiner can choose what hash type to use: MD5, SHA-1, SHA-256 and SHA-512.

The success of the Hash Analysis depends on how complete is the list of hash values of known files. Therefore, new and unknown files, which could be produced by pedophiles, for example, would never be identified by this technique. However, the usage of Hash Analysis in conjunction with the other analysis provided by NuDetective can make the results even more reliable and efficient.

The authors conducted experiments to measure the performance of Hash Analysis, with the main objective to check the additional delay of Hash Analysis. Using high

resolution images, the delay introduced by Hash Analysis was minimal, due to the use of Java Threads. However, it requires reading the entire content of a file, which is an I/O operation, to calculate its hash value. If there are too many big files in the storage device, enabling the Hash Analysis in all files stored may slow down the forensic analysis at all.

4 The usage of NuDetective: a Survey

This section reports a conducted survey with 28 NuDetectiveBrazilian users, showing the results obtained, as well as their opinions about the usage of the tool.

4.1 Methodology

The Brazilian users of NuDetective who have requested and received the tool were contacted by email to respond to a simple questionnaire of nine questions about the usage of NuDetective tool and achievements obtained. The contributions were spontaneous, users agreed to contribute anonymously to the research, and were warned about the further dissemination of the results in a scientific manner.

The survey was constructed with the use of specific and free website for conducting surveys. Users who wanted to contribute have answered the questions directly on the website. The results of the conducted study are in this section.

4.2 Survey Questions

The survey contains nine questions: four objective questions with only one option to choose (Type I), four multiple choice questions (Type II) and only a discursive question about suggestions, praise and/or criticism (Type III). Table 1 shows the questions in the conducted survey.

Table 1. Question about the survey answered by 28 NuDetective users.

ID	Question Title	Type
1	About the installation and usage of NuDetective, check one of the following:	I
2	In how many real forensic cases did you use the NuDetective?	I
3	If you NEVER used the tool in real forensic cases, check the options that justify that (check as many options as necessary):	II
4	If you use or have used the NuDetective in real forensic cases, check the options that led you and/or your institution to use the tool (check as many options as necessary):	II
5	What are the negatives / problems encountered during the use of NuDetective? (check as many options as necessary):	II
6	In how many real forensic cases the tool helped you to IDENTIFY FILES of child pornography?	I
7	Check what you suggest about new features and improvements of the	II

	tool (check as many options as necessary):	
8	Where do you work?	I
9	Send your reviews, comments, compliments and / or suggestions about the NuDetective tool.	III
<i>I-Objective mandatory (one answer) / II-Multiple choice / III-Discursive (optional)</i>		

4.3 Survey Answers

This subsection presents the complete results of the most important questions of the survey (questions 2, 4, 5, 6, and 7). The ninth question was optional and discursive, and its answers often included personal information. About the other questions, only the main answers will be described. The results are presented in Table 2.

Table 2. Answers for the questions 2, 4, 5, 6, and 7.

ReplyOptionstoQuestion2	Answers	%
I havenotused	6	21
Only1 time	5	18
2to 5 times	9	32
6to 10 times	4	14
More than 10 times	4	14
ReplyOptionstoQuestion4	Answers	%
The tool have a great performance and is quick to use in real cases	16	73
The tool is reliable and identifies the files that contain pornography stored in digital medias	10	45
The tool is free, so I can use without costs	16	73
I have not found another tool (free or not) that is better than NuDetective in the task of identifying child pornography	7	32
The articles published about the tool made me more confident about the results NuDetective can get	5	23
The NuDetective is easy to install	12	55
The NuDetective is easy to use	15	68
The tool can run on multiple platforms, like Windows and Linux	8	36
The tool can be run with JRE, so there is no need to install any program	11	50
The NuDetective can be executed from a common thumb drive	13	59
Another reason: specify	3	14
ReplyOptionstoQuestion5	Answers	%
I did not found any problem. Everything is ok!	8	33
The tool is slow	0	0
The tool does not find files of child pornography	0	0
NuDetective was unable to run on multiple platforms (Linux, iOS)	0	0
The main configuration interface is bad / confusing	0	0
The search for ZIP files is slow or inaccurate	1	4
The results interface is bad / confusing	0	0
The tool has many false positives	12	50
The tool documentation is poor and I cannot find answers to my ques-	0	0

tions		
The statistics shown in the main interface are confusing	0	0
The tool crashes or closes unexpectedly	0	0
Unable to run the tool via a thumb drive	2	8
Another reason: specify	4	17
ReplyOptionstoQuestion6	Answers	%
None. I never used the tool in real cases	5	18
None. I use the tool, but in the situations that I used NuDetective it didn't find child pornography files	6	21
One case	5	18
Two to five cases	10	36
Over five cases	2	7
ReplyOptionstoQuestion7	Answers	%
The NuDetective should be faster	6	23
The NuDetective should be more precise and use others techniques to identify nudity	10	38
The NuDetective should analyze video content to identify human nudity	12	46
The NuDetective should reduce the number of false positives, combining other techniques for reducing files unrelated to child pornography	14	54
The NuDetective should search for P2P programs installed in the machine in order to warn the user about the existence of these programs.	13	50
The NuDetective should have functionality of save and open projects files in order to facilitate its use	8	31
The NuDetective should have configurable results interface	7	27
The NuDetective GUI should be reviewed as it is confusing	0	0
The default settings of the tool should change	0	0
The NuDetective should search inside RAR and GZ files, not only ZIP	16	62
The NuDetective should more types of Hash Algorithms	3	12
The user manual should be more detailed	4	15
Another reason: specify	4	15

4.4 Survey analysis

A conducted survey about NuDetective showed that the tool is reliable and it is being used to assist the identification of child pornographic files in real forensic cases. About 79% of NuDetective users already used the tool in at least one real forensic case. The main factors for users do not use the NuDetective in real cases (remaining 21%) were: they did not work in child pornography cases (38%) and they did not have time to test/evaluate the tool (25%).

Another interesting part is about the effectiveness of the tool and its contribution to combat sexual violence against children. About 74% of users who have used the tool in real forensic cases have effectively found files of child pornography stored in the computers. That means if the tool was used in a crime scene, the suspect was immediately arrested after the confirmation of the presence of the illegal files stored in the suspect's computer, as provided in Brazilian law [4].

Among the strengths of the tool, users have highlighted the analysis speed of the tool (73%), the fact that the tool is free and have no cost to the institutions/users (73%), the ease of use of the tool (68%), the possibility of use from a simple thumb drive (59%), and the ease of installation (55%). By the other hand, the main complaint was the existence of a high number of false positives (50%), which users also recommended that NuDetective should reduce the number of false positives combining other techniques for reducing files unrelated to child pornography (54%). One third of the users reported they did not find any problem using the tool.

About new improvements, users suggested that NuDetective should also search inside RAR and GZ archives (62%), should verify the existence of programs for file sharing and peer-to-peer networks (50%), and should analyze the content of videos to search for human nudity (46%). This study was done before the release of the third version of the tool, which released the Video Analysis feature, as explained in subsection 3.2.

Furthermore, this research served to guide the authors in selecting new functionalities and improvements, as the recently implemented Video Analysis feature. The discussion of future work is carried out in Section 6.

5 Conclusion

This paper presented an overview of the main functionalities of the NuDetective Forensic Tool. The strategies and algorithm used by the tool and also some past experiment results were showed.

Despite the high detection rates achieved in past experiments and the intense use of the tool in real forensic cases, the authors had not obtained any formal feedback of NuDetective users about its performance. Therefore, in this work the authors exhibit an unpublished survey conducted to evaluate the real effectiveness of NuDetective by Brazilian users.

The survey, answered by 28 users, showed that about 79% of them used the tool in at least one real forensic case. Moreover, 74% of the users who have used the tool in real forensic cases found files of child pornography in digital media using the tool. So the tool proved to be efficient by most users in real cases.

In general, the results indicated that NuDetective is being used in several Brazilian law-enforcement (72%) and other public institutes (21%). The NuDetective users are confident with its performance, demonstrated by the high number of child pornographic files identified in real Brazilian forensic cases. Once the crimes of production, possession, and transmission of child pornographic files are global and countries are constantly fighting against the sexual exploitation of children, the survey is also important to disseminate the results obtained from the usage of NuDetective by Brazilian law-enforcement, so other countries can take advantage of the developed work.

The chart illustrated in the Figure 1 shows that the number of “red-handed” arrests after the year NuDetective was launched (2010) is bigger than previous years. This suggests that the tool achieved its mainly objective, that is help to combat child pornography.

The NuDetective Forensic Tool is free and available only to law enforcement and public entities. For instructions on how to get it, please contact the authors or send an e-mail to nudetective@gmail.com.

6 Future Work

One of the goals of the conducted survey is to identify new functionalities and improvements demanded by the real users. The main complaint obtained by the survey was the existence of a high number of false positives and the users recommended that NuDetective should combine other techniques for reducing files unrelated to child pornography. To achieve this objective, the authors will implement soon a face detection algorithm in order to eliminate pictures that have many skin pixels in faces, avoiding bringing for the forensic examiner unrelated files. Although face detection algorithms are time consuming, this technique will be applied only to the selected suspicious files and should not affect significantly in the runtime.

Another implementation requested in the survey is to search inside RAR and GZ archives, to amplify the files reached by the Tool.

Finally, another suggestion was to search for P2P programs installed in the machine in order to warn the user about the existence of these software, since they can be used to exchange child pornography files with other users around the world.

Acknowledgment

The authors thank to the Brazilian Federal Police for financial and logistic support.

References

- [1] Prisões em flagrante por pornografia infantil mais que duplicam em 2013 no Brasil - Jornal folha de Sao Paulo. Available from <http://www1.folha.uol.com.br/tec/2014/02/1417794-numero-de-prisoas-em-flagrante-por-pornografia-infantil-mais-que-duplica-em-2013-no-brasil.shtml>; [Visited Apr, 2014]; [in portuguese].
- [2] Central nacional de denúncias de crimes cibernéticos – SaferNet Brasil. Available from <http://indicadores.safernet.org.br/>; [Visited Apr, 2014]; [in portuguese].
- [3] Choo, K-K. R. Online child grooming: a literature review on the misuse of social networking sites for grooming children for sexual offences. Australian Institute of Criminology, pp 48-66. Available from <http://www.aic.gov.au/publications/current%20series/rpp/100-120/rpp103.aspx>; [Visited Apr, 2014].
- [4] BRAZIL. Law 11,829, November 25th, 2008. Brazil, 2008; [in portuguese].
- [5] Polastro, M. C.; Eleuterio, P. M. S. NuDetective: a Forensic Tool to Help Combat Child Pornography through Automatic Nudity Detection. Proceedings of the Twenty-First International Workshop on Database and Expert Systems Applications (DEXA'10 – VISM'10) - pgs. 349-353. IEEE Computer Society. Bilbao, Spain. Aug, 2010.
- [6] Eleuterio, P. M. S.; Polastro, M. C. Identification of High-Resolution Images of Child and Adolescent Pornography at Crime Scenes. THE INTERNATIONAL JOURNAL OF FORENSIC COMPUTER SCIENCE - IJoFCS, v. 5, p. 49-59, Apr, 2011 (published in prior year 2010).
- [7] Polastro, M.C.; Eleuterio, P.M.S. A statistical approach for identifying videos of child pornography at crime scenes. Proceedings of the Fifth Workshop in Digital Forensics (WSDF'12 – ARES'12). IEEE Computer Society. Prague, Czech Republic. Aug, 2012.
- [8] Polastro, M.C.; Eleuterio, P.M.S. - Quick Identification of Child Pornography in Digital Videos. The International Journal of Forensic Computer Science - V. 7, N. 2 (2012), p. 21-32 - DOI: 10.5769/IJ201202002 - Brazilian Association of High Technology Experts (ABEAT) - Brasilia, Brazil. Dez/2012.

- [9] Albiol, A.; Torres, L.; Delp, E. J. Optimum color spaces for skin detection. *International Conference on Image Processing*, volume 1, pages 122–124, 2001.
- [10] Elgammal, A.; Muang, C.; Hu, D. *Skin Detection - a Short Tutorial*. Encyclopedia of Biometrics, Verlag, 2009.
- [11] Jones, M. J., Rehg, J. M. Statistical color models with application to skin detection. *International Journal of Computer Vision (IJCV)*, pages 81–96, 2002.
- [12] Pascale, D. A Review of RGB Color Spaces ... from xyY to R'G'B'. The BabelColor Company. Montreal, Canada, 2003.
- [13] Vezhnevets, V.; Sazonov, V.; Andreeva, A. A Survey on Pixel-Based Skin Color Detection Techniques. *Proceedings Graphicon*, 2003.
- [14] Kakumanu, P.; Makrogiannis, P.; Bourbakis, N. A survey of skin-color modeling and detection methods. *Pattern Recognition* 40, 2007.
- [15] Kovac, J.; Peer, P.; Solina, F. Human skin colour clustering for face detection. *International Conference on Computer as a Tool EUROCON 2003*, Eds. B. Zajc, Ljubljana, Slovenia, September 2003.
- [16] Wang, Y.; Yuan, B. A novel approach for human face detection from color images under complex background. *Pattern Recognition* 34, pages 1983–1992, 2001.
- [17] Wang, H.; Chang, S. A highly efficient system for automatic face region detection in MPEG video, *IEEE Trans. Circuits Systems Video Technol.*, pages 615–628, 1997.
- [18] Carlsson, A.; Eriksson, A.; Isik, M. Automatic detection of images containing nudity. Master thesis in intelligent systems Design, University of Gothenburg, Sweden, 2008.
- [19] Ap-apid, R. An Algorithm for Nudity Detection. *Proceedings of the 5th Philippine Computing Science Congress*, Cebu City, 2005.
- [20] Lin, Y.; Tseng, H.; Fuh, C. Pornography Detection Using Support Vector Machine. *16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP)*. Kinmen, ROC, 2003.
- [21] Wang, J.; Li, J.; Wiederhold, G.; Firschein, O. System for Screening Objectionable Images Using Daubechies' Wavelets and Color Histograms. *Proc. of the International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services*, pages 10–30, 1997.
- [22] Poisel, R.; and Tjoa, S. Forensics investigations of multimedia data: A review of the state-of-the-art. *Proc. of the 6th International Conference on IT Security Incident Management & IT Forensics (IMF'11)*, Stuttgart, Germany. IEEE, May 2011, pp. 48–61.
- [23] Lopes, A.P.B., Avila, S.E.F., Peixoto, A.N.A., Oliveira, R.S., Coelho, M. de M. & Araújo, A. de A. Nude detection in video using bag-of-visual-features. *Proceedings of the XXII Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI*, IEEE Computer Society Press, Rio de Janeiro, RJ, Brazil, 2009, pp 224-231.
- [24] C.-Y. Kim, O.-J.Kwon, W.-G.Kim, and S.-R. Choi. Automatic System for Filtering Obscene Video. In *ICACT*, volume 2, pages 1435--1438, 2008.
- [25] Wang, D.; Zhu, M.; Yuan, X.; Qian, H. Identification and annotation of erotic film based on content analysis. *Proc. SPIE* 5637, 88 (2005).
- [26] Lopes, A.P.B., Avila, S.E.F., Peixoto, A.N.A., Oliveira, R.S. & Araújo, A. de A. A bag-of-features approach based on Hue-SIFT descriptor for nude detection (in press). *Proceedings of the 17th European Signal Processing Conference, EUSIPCO*, Glasgow, Scotland, 2009, pp. 1552-1556.
- [27] Jansohn, C.; Ulges, A. & Breuel, T. M. Gao, W.; Rui, Y.; Hanjalic, A.; Xu, C.; Steinbach, E. G.; El-Saddik, A. & Zhou, M. X. (ed.) Detecting pornographic video content by combining image features with motion information. *ACM*, 2009, 601-604.
- [28] Endeshaw, T.; Garcia, J.; Jakobsson, A. 2008. Classification of indecent videos by low complexity repetitive motion detection. In *Proceedings of the 2008 37th IEEE Applied Imagery Pattern Recognition Workshop (AIPR '08)*. IEEE Computer Society, Washington, DC, USA, 1-7.
- [29] Rea, N.; Lacey, G.; Lambe, C.; Dahyot, R. Multimodal Periodicity Analysis for Illicit Content Detection in Videos. *Visual Media Production, 2006.CVMP 2006.3rd European Conference on In Visual Media Production, 2006.CVMP 2006. 3rd European Conference on (2006)*, pp. 106-114.
- [30] Zuo, H.; Wu, O.; Hu, W.; Xu, B. Recognition of blue movies by fusion of audio and video. *Proceedings of the IEEE International Conference on Multimedia and Expo(ICME)*, pp. 37–40, April 2008.
- [31] Menezes, A. et al. "Handbook of Applied Cryptography". CRC Press, 1996.
- [32] Saroiu, S. et al. "A Measurement Study of Peer-to-Peer File Sharing Systems". *Proceedings of the Multimedia Computing and Networking (MMCN)*. San Jose, Jan, 2002.
- [33] Latapy, M.; Magnien C.; and Fournier, R. Technical report on the Automatic Detection of Paedophile Queries. Measurement and Analysis of P2P Activity Against Paedophile Content project. Released in 2009. Available at <http://antipaedo.lip6.fr/> [Visited Apr, 2014].
- [34] Belbèze, C. et al. Technical report on Automatic Identification of Paedophile Keywords. Measurement and Analysis of P2P Activity Against Paedophile Content project. Available at <http://antipaedo.lip6.fr/> [Visited Apr, 2014].