Challenges and Limitations Concerning Automatic Child Pornography Classification

Anton Moser, Marlies Rybnicek and Daniel Haslinger

Institute of IT Security Research, St. Poelten University of Applied Sciences,
Matthias Corvinus-Strasse 15, St. Poelten, Austria

Keywords: Computer Vision, Automated Child Pornography Detection, Content Classification.

Abstract: The huge volume of data to be analyzed in the course of child pornography investigations puts special demands on tools and methods for automated classification, often used by law enforcement and prosecution. The need for a clear distinction between pornographic material and inoffensive pictures with a large amount of skin, like people wearing bikinis or underwear, causes problems. Manual evaluation carried out by humans tends to be impossible due to the sheer number of assets to be sighted. The main contribution of this paper is an overview of challenges and limitations encountered in the course of automated classification of image data. An introduction of state-of-the-art methods, including face- and skin tone detection, face- and texture recognition as well as craniofacial growth evaluation is provided. Based on a prototypical implementation of feasible and promising approaches, the performance is evaluated, as well as their abilities and shortcomings.

1 INTRODUCTION

"Operation Spade" (Service, 2014) was a great success against child pornography for law enforcement agencies. In 3 years of investigation, 45 TB of child pornography as well as a customer database containing hundredths of datasets were seized. The dark figure of such illegal (digital) assets is assumed to be higher in the order of magnitudes. Companies like Google extended their efforts in detecting and removing such content. Teaming up with Microsoft, they are now successfully removing more illegal content than ever before by using human workpower for manual categorization of images and videos. (BBC, 2013) Since the number of digital content is rising and due to the limits of technological capacities in the field of automated categorization, further research is inevitable. Another application that turns up due to the growing popularity of social networks is that a substantial amount of interpersonal communication - especially of young people - takes place in social networks. New media brings along new perils and threats like Sexting or Posing. Monitoring tools (Rybnicek et al., 2013) can be equipped by such automated pornography detectors to raise awareness of adolescents. The main contribution of this paper is an introduction of previously done research in the field of automated child pornography classification, along with a prototypical implementation. The results are discussed and future work that needs to be done in order to overcome limitations is outlined.

In the course of this paper we start with the related work in Section 2. Section 3 describes our prototypical implementation. Furthermore, we give an overview of our experimental setup. In Section 4 and 5 we summarize our findings and give an outlook for future developments in this field.

2 RELATED WORK

In order to establish the state-of-the-art, we decided to search the digital libraries of ACM, IEEE and Springerlink. Additionally we investigated the bibliography of existing literature for more information. The following detection approaches are considered to be feasible for automated nudity, pornography and age detection. They can be roughly categorized as Face Detection, Skin Tone Detection, Texture Detection and Age Estimation.

2.1 Face Detection

In the process of sieving related literature (Talele and Kadam, 2009) or (Zakaria and Suandi, 2011),

the Viola-Jones algorithm as well as an extended version of Eigenfaces enhanced through neural networks turned out to be the most feasible methods. A new revolution in Face Detection is introduced by Facebook called 'DeepFace' (Taigman et al., 2014). Based on coupling a 3D model-based alignment with large feedforward models an accuracy of 97.35% is reached.

2.2 Skin Tone Detection

Skin Tone Detection is important to determine the skin tone of depicted persons in order to detect the percentage of skin visible on the digital asset. The resulting value can be used for sieving input data for nudity in order to minimize the number of images to be processed in the next steps. Selecting a color space suitable for skin tones is important. Usual color spaces are: RGB (Red, Green, Blue), YCbCr (Luminance, Chroma Blue, Chroma Red), HSV (Hue, Saturation, Value), YIQ (Luminance, Cyan-Orange Balance, Magenta-Green Balance) and YUV (Luminance, Chroma U, Chroma V). While all these color spaces are suitable for skin tone detection in first place, there are limitations to the process of applying manual thresholds in order to parameterize the detection algorithm. External influences like reflections, illumination and poor image quality lead to decreasing detection rates. (Yang et al., 2011)

Red/Green (R/G) ratio and Human Composition Matrix (HCM) are the two main processes of the hierarchical image filtering method, introduced by (Polpinij et al., 2008). R/G ratio is preferably used because it shows significant results for skin colors that are commonly found in African, Asian and Caucasian skins. This provides a feasible way of determining the thresholds for skin tone detection algorithms. If R/G ratio is not able to deliver reliable results, HCM is applied as a next processing step. The input image is sectored and compared against skin- and non-skin histogram models. Further, the probability of the color being a skin-tone is derived.

Another approach is the combination of 2-D histograms and the usage of Gaussian models (Tan et al., 2012). In our proof-of-concept, an eye-detector was used in order to refine the skin model. The major advantage of this algorithm is that it does not depend on training data and can cope with different ethnicities and varying illumination of the image.

To enhance the performance of skin tone detection mechanisms, local features and descriptors can be extracted as introduced in (Jiang et al., 2007) or (Ng and Pun, 2011).

2.3 Shape Detection

Shape detection is usually performed in succession to skin tone detection. Most shape detection algorithms follow the same approach: After areas of interest are determined, they are characterized based on the contour of the object. The decision between normal and pornographic images is made based on the extracted contour and a set of post processing steps as described in (Tan et al., 2012). Hu et al (Hu et al., 2006) proposed a method for torso detection in still images: "[..] the image is segmented into uniform areas. Then, dominant colors of the torso are adaptively selected using a color probability model. Finally, the torso candidates can be extracted based on the dominant colors"

2.4 Age Detection

To be able to automatically distinguish between pornography and child pornography, the age of every person depicted on the source material is vital. Throughout the last couple of years, age detection gained significance, as shown by the sheer number of research (Selvi and Vani, 2011) (Takimoto et al., 2006) (Li et al., 2012) (Fu et al., 2010) done in this field. The face is the only part of the human body that allows to visually determine the age of the person. Measuring the cranio-facial growth shows the most significant changes during the first 20 years of life. To detect the age, a set of features has to be extracted from the face, including eyes, nose and mouth. While research proposes different ways of detecting age, approaches based on distances, ratio and landmarks turn out to provide the best performance. Weda et al (Weda and Barbieri, 2007) show that the extraction of single facial features, e.g. the iris, also provide good results in age estimation. Since the human iris does not change in size in a persons lifetime while the head certainly does, the iris/head ratio can be used to determine the approximate age. The prerequisite for this approach is the availability of frontal images, something that is rare in the particular domain. In (Geng et al., 2013), the authors address that the main difficulty in facial age estimation is the lack of sufficient training data for many ages. Based on the fact that the growth of faces is a slow and smooth process, an algorithm named IIS-LLD is introduced which learns from labeled distributions. The basic idea behind their approach is that a face image contributes to not only the learning of its real age, but also the learning of its neighboring ages. Another approach is introduced by Guo et al (Guo et al., 2009) who use biologically inspired features for human age estimation.

3 EXPERIMENTAL SETUP

In this chapter the prototypical implementation of an automated child porn detection scheme based on the most promising approaches is described. The various methods of face detection, skin tone detection and shape detection as well as age estimation found in existing literature were compared and chosen based on accuracy and the number of samples used. Based on the prototype, we tried to verify if the following objectives can be achieved:

- child detection and

- nudity detection

The prototypical implementation has been done in Python. External libraries like OpenCV, Numpy, SciPy, Cython and Scikit were used. A modular architecture allows pluggable functionality and extensibility. The foundation builds the automated face detection which is needed to distinguish children, adolescents and adults. Features like eyes, nose and mouth are used for age estimation. Parallel skin detection is done based on dynamic thresholding and boosted pixel-based skin detection algorithm. Additional texture analysis enables the detection of explicit body regions. The prototype was evaluated with 100 images divided into two categories: age estimation and porn material. The reason for using two evaluation phases is that the possession of child pornography is illegal and therefore not justified. The accuracy rates of age estimation and nudity detection are separately accounted.

Preprocessing and Face Detection. After preprocessing an image based on Histogram Equalization, Face Detection (Viola and Jones, 2001) is executed. Face Detection is essential for further processing steps like age estimation and skin tone detection.

Age Estimation. Based on an already found and pruned face image, an additional eye detection algorithm is applied. After left- and right eye detection, an face orientation angle is computed to adjust varying postures of the head. A concluding analysis based on edge detection determines whether it is really a human face or not. To detect this, the following rules are used:

- Region I (area of the eyes) has more edges than region II (left cheek)

- Region I (area of the eyes) has more edges than region IV (right cheek)

- Region III (nose) has more edges than Region II (left cheek)

- Region III (nose) has more edges than region IV (right cheek)

For age estimation again eyes, nose and mouth detection algorithms are used. The final rule-based analysis (Tanner, 2011) tries to minimize recognition failures. Distances and ratios are calculated based on cranio-facial growth models (Izadpanahi and Toygar, 2012), see figure 1.
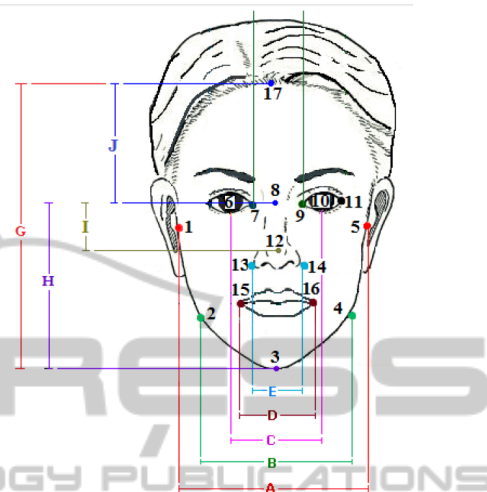

Figure 1: Landmarks for distance and ratio calculation (Izadpanahi and Toygar, 2012).

Skin Tone Detection. After face detection, dynamic skin color analysis takes place, resulting in dynamic thresholds for skin color analysis which are applied to the original image for nudity detection. Small modifications of the original algorithm (Yogarajah et al., 2012) were carried out. Dynamic thresholds were calculated based on the rotated face, before the extraction process took place. A squared small area of the center of the face is taken as a benchmark to guarantee that no external influences like hair or background affect the accuracy. Further the values of YCbCr, R/G ratio, 1D and HSV color spaces are calculated. All computed values are processed by means of histograms. 5% of the exceeding values are cutted to reduce noise. The results are upper and lower boundaries for the thresholds. The resulting thresholds are compared and a threshold scope is defined. The scopes are passed to a boosted pixel-based skin classifier (Sajedi et al., 2007) which produces a skin mask consisting of every skin pixel in the image.

These masks are used for extracting Regions of Interests (ROI) (Karavarsamis et al., 2013). Therefore the whole image is divided into squares, each about 2.5% of the total image dimensions. Each of these squares is analyzed and marked as ROI if the amount of skin-toned color exceeds 50%. The surface is further scanned for discontinuities, as this might indicate the presence of cloth. Additional texture analysis and Zoning (Santos et al., 2012) allow to further improve

the accuracy. The Canny edge detection algorithm for example is suitable for nipple detection. Zoning assumes that exposed body parts are usually positioned near to the center of the image. Therefore it divides the asset into three areas: Zone 0 defines the image as a whole, zone 2 removes 15% of the border area, zone 3 removes 25%. This enables the algorithm to analyze every zone individually to detect the distribution of skin throughout the picture. After each of the zones is processed, the final classification is performed.

In figure 2 some successful detection steps are displayed: out of the input image a skin-mask (b) and skin-ROIs (c) are extracted. For shape detection the contours are extracted in (d).
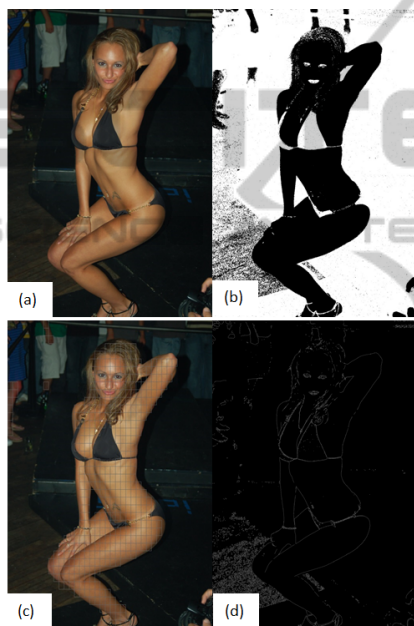

Figure 2: Successful detection: (a) input image (Colgan, 2011), (b) skin mask, (c) skin ROIs, (d) contours.

4 CHALLENGES AND LIMITATIONS

During setup and testing of the experimental setup, limitations and obstacles where encountered that had notable impact on the performance and accuracy of the final classification:

- **Background Color.** Even though dynamic thresholds where applied, areas that feature skin-like background colors impede the proper detection of ROIs. Detecting false-positive areas of skin has negative impacts on the skin masks created, which in the end interferes with the final classification process. The usage of multiple color

spaces and weights improves the classification, but still shows a certain amount of inaccuracy. Figure 3 shows an example where the skin-area could not be separated from the background.

Figure 3: Example result for skin detection with skin-like background color (MonsterMarketplace, 2015).

- **Combination of Algorithms.** None of the separate algorithms used for detecting skin-tone, eyes, face etc. result in 100% of accuracy. This leads to missing features, which would be needed as input for subsequent processes. While this can be mitigated by running through several stages of cascaded detections, the problem still persists to some extent.

- **Age estimation:** This is one of the most challenging analysis steps due to its requirements for images that come in high quality and resolution. The posture of the head is also important for feature extraction and age estimation. Figure 4 shows four examples, that could not be classified correctly. Although the faces and features are clearly visible and in a high resolution, no faces are found in (a) and (b). In (c) and (d) the faces were classified correctly. However, not all features were found, which results in an imprecise estimation.

- **Absence of Faces.** Images that do not feature faces are processed using default thresholds. This increases the error probability in form of false positives.

The evaluation process of our prototype was carried out using two different test sets. One set consisted of 40 images showing faces of adults and children to evaluate the performance of age estimation. The second set, consisting of 60 images, had its focus on determining the performance of pornography and nudity detection approaches. All the images were retrieved randomly from Google Image Search. The ratio of images showing children, adults, pornography and holiday scenes is evenly distributed.

84% of the images were correctly determined to include nudity or at least expose a certain amount of skin. A clear distinction between holiday pictures and
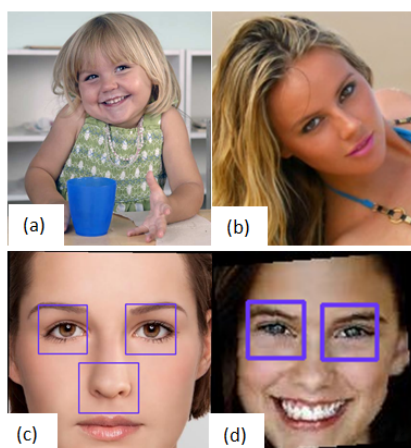
Figure 4: Example images (kinder.de, 2015) (Santa-Banta.com, 2015) (Neuss, 2015) (Bokelberg, 2015) with wrong detection: (a) and (b) no faces found, (c) mouth not found, (d) only eyes found.

pornographic content could not be achieved. Age estimation turned out to be even more challenging, as 2/3 of the images were not correctly classified into the categories "adolescent" and "adult".

5 CONCLUSION

In this paper, we highlighted open research areas which are required to develop automatic child pornography detection, in order to help law enforcement agencies to speed up investigations and implement or enhance (automated-) monitoring tools. Current approaches are not sufficient to guarantee a clear recognition of adolescents and nudity. Based on a prototypical implementation, we showed that the main challenges are dealing with source material that lacks of resolution and quality. Further, face mimics and positioning (e.g. angle and rotation) of depicted individuals lead to problems with age estimation. It has to be possible to detect the eyes and other important face features to compute distances. Different face mimics make research attempts even more difficult. Skin tone analysis of images leads to an acceptable detection rate, but provides no clear distinction between nudity and - for example - holiday pictures with lot of skin visible. Therefore it is necessary to implement texture analysis processes. We achieved an 84% accuracy to detect nudity which includes the exposition of a certain amount of skin. However, a clear distinction between holiday pictures and pornographic content could not be achieved. Age estimation turned out to be even more challenging as 2/3 of the images were not correctly classified. Therefore a 'non-face based

age estimation'-extension of our prototypical implementation is necessary.

We further suggest a modular implementation to enable an easy way for extending our proof of concept implementation with new processes. The use of separate modules also facilitates the exchange between research groups interested in the topic.

REFERENCES

BBC (2013). Google and microsoft agree steps to block abuse images. Google and Microsoft agree steps to block abuse images (last access: 30.10.2014).

Bokelberg (2015). Bokelberg. http://www.bokelberg.com/DE/search/gallery/12783/10/1/ (last access: 09.01.2015).

Colgan, P. (2011). Wikimedia commons. http://commons.wikimedia.org/wiki/File:Bikini_contest_-_black_bikini.jpg? uselang=de (last access: 29.05.2014).

Fu, Y., Guo, G., et al. (2010). Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1955–1976.

Geng, X., Yin, C., and Zhou, Z.-H. (2013). Facial age estimation by learning from label distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10):2401–2412.

Guo, G., Mu, G., Fu, Y., and Huang, T. S. (2009). Human age estimation using bio-inspired features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 112–119.

Hu, Z., Lin, X., and Yan, H. (2006). Torso detection in static images. In *8th International Conference on Signal Processing*, volume 3. IEEE.

Izadpanahi, S. and Toygar, O. (2012). Geometric feature based age classification using facial images. In *IET Conference on Image Processing (IPR)*, pages 1–5.

Jiang, Z., Yao, M., and Jiang, W. (2007). Skin detection using color, texture and space information. In *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, volume 3, pages 366–370. IEEE.

Karavarsamis, S., Ntarmos, N., Blekas, K., and Pitas, I. (2013). Detecting pornographic images by localizing skin rois. *International Journal of Digital Crime and Forensics (IJDCF)*, 5(1):39–53.

kinder.de (2015). Ihr Kind im 5. Lebensjahr. http://www.kinder.de/themen/kleinkind/entwicklung/artikel/ihr-kind-im-5-lebensjahr.html (last access: 09.01.2015).

Li, W., Wang, Y., and Zhang, Z. (2012). A hierarchical framework for image-based human age estimation by weighted and ohranked sparse representation-based classification. In *5th IAPR International Conference on Biometrics (ICB)*, pages 19–25.

MonsterMarketplace (2015). Picker-back bikini with metal ball studs. http://www.monstermarketplace.com/

brazilian-bikinis/ pucker-back-bikini-with-metal-ball-studs (last access: 09.01.2015).

Neuss, A. (2015). Perfekte gesichter. http://www.anikaneuss.de/fotografie_ perfektegesicht1.html (last access: 09.01.2015).

Ng, P. and Pun, C.-M. (2011). Skin color segmentation by texture feature extraction and k-mean clustering. In *Third International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN)*, pages 213–218. IEEE.

Polpinij, J., Sibunruang, C., Paungpronpitag, S., Chamchong, R., and Chotthanom, A. (2008). A web pornography patrol system by content-based analysis: In particular text and image. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 500–505.

Rybnicek, M., Poisel, R., and Tjoa, S. (2013). Facebook watchdog: A research agenda for detecting online grooming and bullying activities. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2854–2859.

Sajedi, H., Najafi, M., and Kasaei, S. (2007). A boosted skin detection method based on pixel and block information. In *5th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 146–151.

SantaBanta.com (2015). Bikini. http://www.santabanta.com/photos/bikini/14001216.htm?high=1 (last access: 09.01.2015).

Santos, C., dos Santos, E. M., and Souto, E. (2012). Nudity detection based on image zoning. In *11th International Conference onInformation Science, Signal Processing and their Applications (ISSPA)*, pages 1098–1103. IEEE.

Selvi, V. T. and Vani, K. (2011). Age estimation system using mpca. In *International Conference on Recent Trends in Information Technology (ICRTIT)*, pages 1055–1060. IEEE.

Service, T. P. (2014). Project spade saves children. http://www.torontopolice.on.ca/modules.php?op=modload& name=News& file=article& sid=7171 (last access 31.10.2014).

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1701–1708. IEEE.

Takimoto, H., Mitsukura, Y., Fukumi, M., and Akamatsu, N. (2006). A design of gender and age estimation system based on facial knowledge. In *International Joint Conference (SICE-ICASE)*, pages 3883–3886.

Talele, K. and Kadam, S. (2009). Face detection and geometric face normalization. In *TENCON 2009-2009 IEEE Region 10 Conference*, pages 1–6. IEEE.

Tan, W. R., Chan, C. S., Yogarajah, P., and Condell, J. (2012). A fusion approach for efficient human skin detection. *IEEE Transactions on Industrial Informatics*, 8(1):138–147.

Tanner, K. (2011). Modeling automated detection of children in images. Master's thesis, University of Rhode Island.

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

Weda, H. and Barbieri, M. (2007). Automatic children detection in digital images. In *IEEE International Conference on Multimedia and Expo*, pages 1687–1690. IEEE.

Yang, L., Li, H., Wu, X., Zhao, D., and Zhai, J. (2011). An algorithm of skin detection based on texture. In *4th International Congress on Image and Signal Processing (CISP)*, volume 4, pages 1822–1825.

Yogarajah, P., Condell, J., Curran, K., McKevitt, P., and Cheddad, A. (2012). A dynamic threshold approach for skin tone detection in colour images. *International Journal of Biometrics*, 4(1):38–55.

Zakaria, Z. and Suandi, S. A. (2011). Face detection using combination of neural network and adaboost. In *TENCON 2011-2011 IEEE Region 10 Conference*, pages 335–338. IEEE.