

Media Content Access: Image-Based Filtering

Rehan Ullah Khan¹, Ali Alkhalifah²
Information Technology Department
Qassim University, Al-Qassim, KSA

Abstract—As the content on the internet contains sensitive adult material, filtering and blocking this content is essential for the social and ethical values of the many societies and organizations. In this paper, the content filtering is explored from still images' perspectives. Thus, this article investigates and analyses the content based filtering which can help in the flagging of the images as adult nature or safe images. As the proposed approach is based on the Chroma (colour) based skin segmentation and detection for detecting the objectionable content in images; therefore, the approach proceeds in the direction of the classical Machine Learning approaches and uses the two well-known classifiers: 1) The Random Forest; and 2) the Neural Network. Their fusion is also investigated. Skin colour is analyzed in the YCbCr colour space and in the form of blob analysis. With the "Adult vs. Safe" classification, an Accuracy of 0.88 and the low RMSE of 0.313 is achieved, indicating the usefulness of the detection model.

Keywords—Skin detection; content based filtering; content analysis; machine learning; random forest; neural network

I. INTRODUCTION

The most famous influential media platforms that allow the users to upload the recorded content (images and videos) are the Flickr, Facebook, Twitter, YouTube, and the DailyMotion. This is not limited to the social platforms. In fact, the internet itself is a gigantic and a general platform for digital media including images and videos resources. The content uploaded to platforms and the Internet itself is increasing rapidly as more and more people are finding access and enjoying the services provided by these providers. The negative side is that the media content; however, is becoming more and more liberal and one can easily access and see the partially/fully naked images on the internet. This opens risks in terms of many social factors. One of the major problems is the availability of these resources to the younger generation. The most feared element, however, nowadays is the availability of these media resources to the kids.

Therefore, in this paper, we analyze skin color based image content filter which can help in the flagging of the images as adult nature or safe images. As the proposed approach is based on color based skin segmentation and detection for detecting the objectionable content in images, therefore, we proceed in the direction of the classical Machine Learning approach and use the two well-known classifiers: The Random Forest and the Neural Network. For skin color analysis, we also investigate their fusion. We analyze skin color in the YCbCr color space and in the form of blob analysis. With the "Adult vs Safe" classification, we get an Accuracy of 0.88 and the low RMSE of 0.313, indicating the usefulness of our detection model. The success of this filter can have profound applications in media

filtering which will benefit not only the general society but also will be especially, useful for parental control over the media for the children and similar requirements. The basic filter developed can be further extended for videos and online streaming like YouTube and other resources.

Certain IP level porn and adult content blocking is possible, however, blocking the user's content on the IP level is always by-passable as there is certain Proxy bypassing tools available to bypass the restrictions and access the corresponding resources containing explicit content. Therefore, blocking and filtering the multimedia based on its content is of utmost importance.

There is interesting work available in the state-of-the-art regarding content based retrieval [1], [2], skin detection [3], [4], and content based filtering [5]. In [5], the authors propose a method combining evidence including video sequences, key shots, and key frames and evaluating performance with three the social networks. The work in [6] describes a sampling, based on the adaptive sampling analysis achieving an acceptable detection rates of 87%. The author in [7] use "bag of visual words approach" for filtering and blocking nude images using the associated voting scheme, analyzed and evaluated achieving 93.2 Accuracy. The work in [8] targets skin locus detection for content filtering using the 24 colors transformations in widely available images and videos. The framework of [9] produces an augmented classification model with independent of the access scenarios with the promising results. The authors in [10] combines the key-frame based approaches with a statistical MP4 motion vectors.

Generally, Image and Video Retrieval (IVR) paradigm is divided into two directions. One that uses media content directly by taking advantage of the visual information present in the images and videos. This type of approach is generally termed as the Content-Based Image and Video Retrieval (CBIVR). In CBIVR, the images and videos are retrieved and searched using the low-level features, for example, color, shape, texture and pixel relationships [11], [12]. Another approach that is nowadays advocated in conjunction with CBIVR is the Textual-Based Image and Video Retrieval (TBIVR). The textual analysis of images and videos uses mostly the tags assigned during the production of the media source. In order to overcome the limitations and drawbacks of the CBIVR, the TBIVR gracefully represent the visual information during the production or editing phase by users manually assigned keywords and or tags. The systems can also allow for later assignments of tags, however, this might introduce wrong tags and tags that are unnecessary. The TBIVR systems allow the users to type the information need as a text query.

In [13], the authors analyze content filtering for a Web-based P2P using the Machine learning to filter the explicit content and results show the feasibility of the approach. The work in [14] uses two visual features and are constructed from the video in question using decision variable based on the single frame and a group of frames. In [15], authors use the Hue-SIFT for nude and explicit content detection. In [16], authors propose visual motion analysis approach augmenting the visual motion features with the audio repetition in video. The [17] demonstrate a multimodal hierarchy based filtering for explicit content. The algorithm is made of 3 phases, i.e. Detecting initial video scene using the hashing signatures, real-time detection estimating the degree of explicit content, and finally, exploiting features from frame-group and thus achieving high detection rate. The authors in [18] uses optical flow for content filtering based on the selection of the frame as the key-frame. The work in [19] use similar approach of the motion estimation. In [20], authors present an objectionable (porn) video filtering approach using a fusion of audio features and video features. Training the Support Vector Machine (SVM) on the chromatic and texture cues of the SIFT key points for adult image frames integrated by the Bayes statistics for classification.

II. COLOR BASED SKIN DETECTION

For content filtering in still images, the basic feature for the detection starts with a reliable color based skin locus detection in images. We start with the experimental setup of the robust skin analysis using color features in the YCbCr color space. As the proposed approach is based on the Chroma (color) based skin segmentation and detection, therefore, we use the two well-known classifiers: The Random Forest and the Neural Network, selecting them due to the good classification performance in many related tasks.

III. CLASSIFICATION

As the proposed approach is based on the Chroma (colour) based skin segmentation and detection for detecting the objectionable content in images, therefore, we proceed in the direction of the classical Machine Learning approaches and use the two well-known classifiers: The Random Forest and the Neural Network. We also investigate their fusion in the Experiment section.

A. Random Forest

Recently, the tree based classifiers have gained considerable popularity. This popularity stems from the intuitive nature and the overall easy training paradigm. Classification trees, however, suffer from the classification and generalization accuracy. It is not possible to increase both the classification and generalization accuracy at the same time. Leo Breiman [21] introduced the Random Forest for addressing these issues. Random forest takes advantage of the combination of many trees from the same dataset. Random forest constructs a forest of trees such that each tree is

generated based on the random seed augmented on the data and assigns classes based on voting scheme from the trees.

B. Neural Network (ML)

From the neural network paradigm, we use the Multilayer Perceptron (MLP). A Multilayer perceptron is a Neural Network; a feed-forward Artificial Neural Network that functions by mapping the input variable of a dataset onto the output labels. Generally, it is significantly different than the linear perceptron in the way it takes advantage of the two or more layers of artificial neurons by integrating the non-linear activation functions. It can thus model linear and non-linear problems as well.

IV. FUSION

To achieve an effective and robust skin segmentation performance, the fusion of the two machine learning algorithms, i.e., the Random Forest and the Multilayer Perceptron is investigated. We believe that the fusion of the classifiers might report interesting results and increased classification performance. This fusion of the classification is investigated on the five parameters as follows:

- 1) *Average of probabilities*
- 2) *The product of probabilities*
- 3) *Majority voting*
- 4) *Minimum probability*
- 5) *Maximum probability*

V. SKIN BLOBS

By blob analysis, we mean the physical shapes the skin and non-skin regions/objects represent in an image. We believe that it may provide a good measure for prediction based on the shapes of skin regions. We use the 6 blob features for the skin non-skin regions analysis. These features are:

- Area
- Convex-Area
- Eccentricity
- Orientation
- Perimeter
- Solidity

VI. EXPERIMENTAL ANALYSIS

In this section, we discuss the datasets, the experimental setup and the results.

A. Dataset

The dataset for our experiments is a hybrid dataset containing images from [22] and our own additions. The total image patches present are 3242 skin and non-skin patches. This dataset contains two types of images, one is the original set of

images and the other is mask images as shown in Fig. 1. The dataset consists of images taken in different lighting conditions. It represents many types of skin ranging from white to black. Some images also have complex backgrounds similar to that of human skin color. We represent this dataset as DS_SKIN.

For adult content analysis, another dataset is also used for which contains an adult image, safe images and suspicious images (confusing the algorithm). This dataset contains 6000 images and most of them are extracted from the [23]. The dataset is represented as DS_ADULT and is used for the experimentation of differentiating between adult and non-adult content using skin features.



Fig. 1. Mask images from dataset.

B. Results and Evaluation

We discuss the experimental evaluation performed for different parameters discussed previously. First experiment consists of pixel based skin model generation in the YCbCr color space and analyzing its performance. The performance is then also analyzed in terms of the fusion of the two classifiers: The Random Forest and the MLP using different strategies.

Fig. 2 shows the evaluation of YCbCr color space using the Random Forest and the MLP and the five fusion strategies. The random forest reports an increased classification performance compared to the MLP.

The performance of the five fusion approaches of the Random Forest and the MLP shows that the average performance of the five fusion approaches is almost similar and has increased the performance of the MLP but has slightly decreased the performance of the Random Forest. In Fig. 2, the RMSE; however, shows variations with smallest RMSE for minimum probability and maximum for majority voting. This analysis shows that as of other fields, the Random Forest also shows increased classification performance in the pixel-based classification. The fusion though theoretically may increase performance; however, in practice, we did not find a big difference with the five strategies of fusion of classification. The minimum probability fusion reports a decreased RMSE in all the cases of the fusion strategies.

For blob analysis, a blob represents the physical shape of the skin and non-skin regions/objects present in an image. In blob analysis, we discuss the performance analysis using the six features of:

- 1) Area
- 2) Convex-Area
- 3) Eccentricity
- 4) Orientation
- 5) Perimeter
- 6) Solidity

Fig. 3 shows the performance analysis using blob analysis based on these six features. The random forest once again has comparatively higher performance than the MLP. The Random Forest has F-measure of 0.90, an accuracy of 0.90 and RMSE of 0.26. The MLP reports slightly decreased performance compared to the Random forest with an F-measure of 0.89, an accuracy of 0.89 and RMSE of 0.29. We get an approximately 1% of the increase in this case, which is not significant compared to the pixel analysis of Fig. 2.

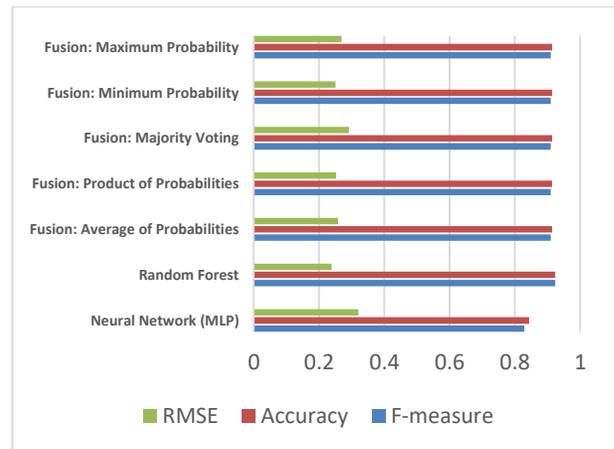


Fig. 2. Skin detection performance using YCbCr color space in the Random Forest and MLP setup and the fusion of classifiers.

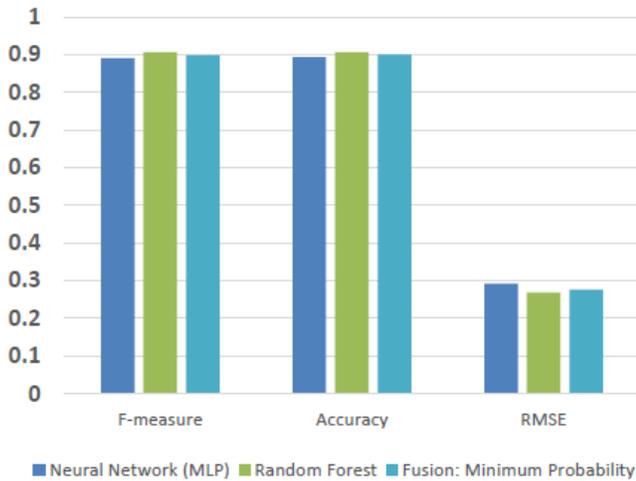


Fig. 3. Performance analysis of blobs of skin and non-skin regions.

For adult sensitive data classification, we use the dataset of images containing adult and non-adult content and is distributed into three classes:

- 1) Adult
- 2) Suspicious
- 3) Safe images

Adult images are sensitive images taken from porn movies. Suspicious images are purely not adult images but rather shots containing naked skin and naked people. It also consists of images with confusing backgrounds and having skin like color objects. Safe images are those with people and objects that are acceptable to most societies. This dataset contains almost 6000 images and most of them are extracted from [23]. The dataset is represented as DS_ADULT and is used for the experimentation of differentiating between adult and non-adult content using skin features. Fig. 4 shows the distribution of the images and their classes. The dataset consists of 2001 images of adult nature, 2001 suspicious images, and 2000 safe images.

Based on the over-all good performance in the previous experiments of skin and the state-of-the-art, compared to the MLP, we select the random forest for sensitive data classification. Fig. 5 shows the performance evaluation of the Random Forest within the four evaluation dimensions and the two parameters of the Accuracy and the F-measure. The four performance evaluation dimensions are:

- 1) Adult class vs. the suspicious class vs. the safe class
- 2) Suspicious vs. safe
- 3) Adult vs. suspicious
- 4) Adult vs. safe

From Fig. 5, it can be seen that the “Adult vs. suspicious vs. safe” gets an F-measure of 0.766, an accuracy of 0.766, and an RMSE of 0.356. Meaning that over-all, on average, out of 100 images, approximately, 76 are correctly identified as either “Adult”, “Suspicious” or “Safe”. The evaluation of “Suspicious vs. safe” images reports an increased classification compared to the previous class evaluation. With the “Suspicious vs. safe image”, we get an F-measure of 0.809, an Accuracy of 0.808, and an RMSE of 0.393. In Fig. 5, the “Adult vs. suspicious”

classification reports an increased classification of 0.859 (F-measure) and an Accuracy of 0.859 with low RMSE of 0.336. Our main interest is in the “Adult vs. the safe” classification. With the “Adult vs. safe” classification, we get an increased F-measure of 0.887, an increased Accuracy of 0.887 and low RMSE of 0.313. As the application of our research requires above 84% accurate model for adult vs. safe images, we get satisfactory results. Also since, suspicious images may be flagged as adult material; the latter can then manually be checked, thus satisfying our objectives.

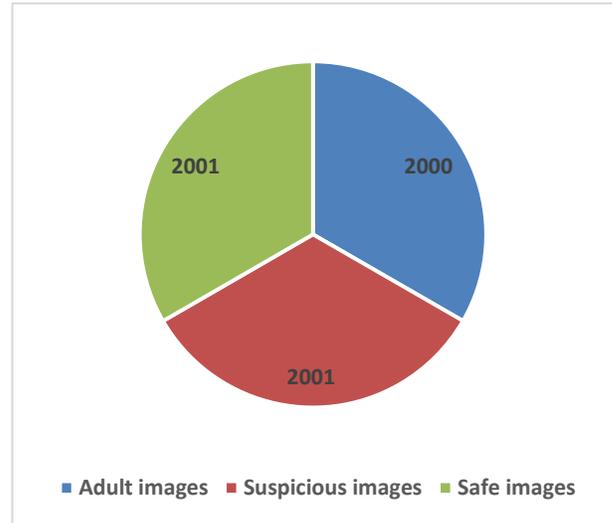


Fig. 4. Dataset distribution for image classification into three categories.

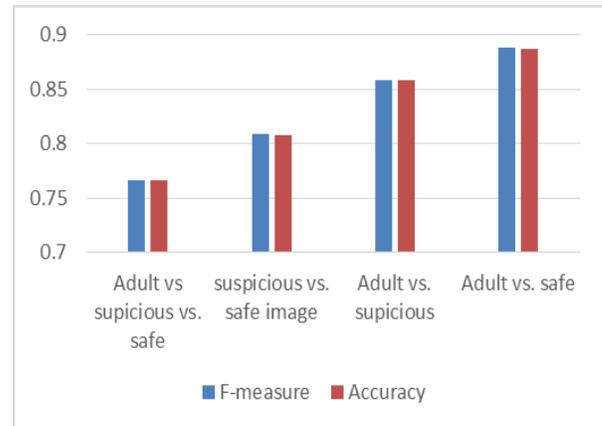


Fig. 5. Adult image classification performance.

VII. CONCLUSION

In this article, we explored the skin and content-based analysis of still images. As the proposed approach is based on the Chroma (color) based skin segmentation and detection for detecting the objectionable content in images; therefore, we walked in the direction of the classical Machine Learning approach and used the two well-known classifiers: The Random Forest and the Neural Network. This analysis showed that as of other fields, the Random Forest also shows increased classification performance in the pixel-based classification. Also, the fusion though theoretically may increase performance; however, in practice, we could not find a big

difference with the five strategies of fusion of classification. The Random Forest has increased classification performance in all the cases of three color spaces. The minimum probability fusion reports a decreased RMSE in all the cases of the fusion strategies. For the blob analysis, we got an increment of 1%, which is not significant compared to the pixel analysis. With the “Adult vs. safe” classification, we get an increased F-measure of 0.88, an increased Accuracy of 0.88 and low RMSE of 0.31. As the application of our research required above 84% accurate model for “adult vs. safe” images, we get satisfactory results. Also since, suspicious images may be flagged as adult material; the latter can then manually be checked.

ACKNOWLEDGMENT

The work in this article is funded in its entirety by the Dean of Scientific Research (SRD), Project number: 1335-coc-2016-1-12-S at the Qassim University, Kingdom of Saudi Arabia. We are thankful to Irfanullah of the Chongqing University for helping in experimental setup.

REFERENCES

- [1] S. H. Shirazi, A. I. Umar, S. Naz, N. A. Khan, M. I. Razzak, B. AlHaqbani, “Content-Based Image Retrieval Using Texture Color Shape and Region,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, 7(1), 2016.
- [2] M. Alkhwilani, M. Elmogy and H. Elbakry, “Content-Based Image Retrieval using Local Features Descriptors and Bag-of-Visual Words,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, 6(9), 2015.
- [3] S. A. S. Hani K. Al-Mohair, Junita Mohamad-Saleh, “Human skin color detection: A review on neural network perspective,” *Int. J. Innov. Comput. Inf. Control IJICIC*, vol. 8, no. 12, pp. 8115–8131, 2012.
- [4] K. Y. Md Foisal Hossain, Mousa Shamsi, Mohammad Reza Asharif, “Automatic facial skin detection using Gaussian Mixture Model under varying illumination,” *Int. J. Innov. Comput. Inf. Control IJICIC*, vol. 8, no. 2, pp. 1135–1144, 2012.
- [5] E. Valle, S. Avila, F. Souza, M. Coelho, and A. de A. Araujo, “Content-Based Filtering for Video Sharing Social Networks,” in *XII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais—SBSeg 2012*, 2011, p. 28.
- [6] P. Monteiro, S. Eleuterio, M. De, and C. Polastro, “An adaptive sampling strategy for automatic detection of child pornographic videos.”
- [7] A. P. B. Lopes, S. E. F. de Avila, A. N. A. Peixoto, R. S. Oliveira, M. de M. Coelho, and A. de A. Araújo, “Nude Detection in Video Using Bag-of-Visual-Features,” in *2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*, 2009, pp. 224–231.
- [8] A. Abadpour and S. Kasaei, “Pixel-Based Skin Detection for Pornography Filtering,” *Iran. J. Electr. Electron. Eng.*, vol. 1, no. 3, pp. 21–41, 2005.
- [9] N. Agarwal, H. Liu, and J. Zhang, “Blocking objectionable web content by leveraging multiple information sources,” *ACM SIGKDD Explor. Newsl.*, vol. 8, no. 1, pp. 17–26, Jun. 2006.
- [10] C. Jansohn, A. Ulges, and T. M. Breuel, “Detecting pornographic video content by combining image features with motion information,” in *Proceedings of the seventeen ACM international conference on Multimedia - MM '09*, 2009, p. 601.
- [11] C. K. B. Dinakaran, J. Annapurna, “Interactive Image Retrieval Using Text and Image Content,” *Cybern. Inf. Technol.*, vol. 10, no. 3, pp. 20–30, 2010.
- [12] M. Y. M. Latha, B. C. Jinaga, and V. S. K. Reddy, “Content Based Color Image Retrieval via Wavelet Transforms,” *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 7, no. 12, 2007.
- [13] J. H. Wang, H.-C. Chang, M.-J. Lee, and Y.-M. Shaw, “Classifying Peer-to-Peer File Transfers for Objectionable Content Filtering Using a Web-based Approach.”
- [14] H. Lee, S. Lee, and T. Nam, “Implementation of high performance objectionable video classification system,” in *2006 8th International Conference Advanced Communication Technology*, 2006, p. 4 pp.-pp.962.
- [15] A. P. B. Lopes, A. P. B. Lopes, R. E. F. De Avila, A. N. A. Peixoto, R. S. Oliveira, and A. De A. Araújo, “a bag-of-features approach based on hue-sift descriptor for nude detection.”
- [16] N. Rea, G. Lacey, R. Dahyot, and C. Lambe, “Multimodal periodicity analysis for illicit content detection in videos,” in *3rd European Conference on Visual Media Production (CVMP 2006)*. Part of the 2nd Multimedia Conference 2006, 2006, pp. 106–114.
- [17] S. Lee, W. Shim, and S. Kim, “Hierarchical system for objectionable video detection,” *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, pp. 677–684, May 2009.
- [18] Y. S. L. Li, “Objectionable videos detection algorithm based on optical flow,” *Comput. Eng.*, vol. 12, p. 77, 2007.
- [19] Y. Qu, ZY; Liu, YM; Liu, Y; Jiu, K; Chen, “a method for reciprocating motion detection in porn video based on motion features,” in *IEEE International Conference on Broadband Network and Multimedia Technology*, 2009, pp. 183–187.
- [20] Z. ZHAO, “Combining SVM and CHMM classifiers for porno video recognition,” *J. China Univ. Posts Telecommun.*, vol. 19, no. 3, pp. 100–106, Jun. 2012.
- [21] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [22] Q. Zhu, C.-T. Wu, K.-T. Cheng, and Y.-L. Wu, “An adaptive skin model and its application to objectionable image filtering,” in *Proceedings of the 12th annual ACM international conference on Multimedia - MULTIMEDIA '04*, 2004, p. 56.
- [23] S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo, “Pooling in image representation: The visual codeword point of view,” *Comput. Vis. Image Underst.*, vol. 117, no. 5, pp. 453–465, May 2013.