DISCRIMINATION BETWEEN CHILDREN AND ADULT FACES USING BODY AND HEAD RATIO AND GEOMETRIC FEATURES

C.A. Reyes-García[1], E. Morales-Vargas[1], H. Peregrina-Barreto[1], C. Manfredi[2]

[1]División de ciencias computacionales, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México
[2]Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
{emoralesv, kargaxxi, hperegrina} @ccc.inaoep.mx, claudia.manfredi@unifi.it

Abstract: **The classification between a child and an adult is an important task in the medical and security field. It consists in the discrimination between children and adult faces in image sequences. Image processing techniques were used to identify regions of interest and computational intelligence was used for classification. First, in the input image, a body detector was used to measure the size of the human body in the image presented. Later, a face detector projects a bounding box around the face. Then, a facial model is fitted in the face of the person to improve the measure of the heads size. Finally, a normalization step is used to ensure size invariance in the image. Using the body size, head size and facial landmarks, geometric features were extracted to perform the classification. Experimental results show accuracy rates of 86%.**
Keywords: **Classification, image processing**

I. INTRODUCTION

Object detection is a widely used technique of image processing, it consists in defining the boundaries of certain objects with a semantic meaning (a car in a street or a tumor in a medical image for example). Basically, the main purpose of object detection is to put a marker or bounding box in the location of a specific object when it is detected. It's suggested that the object recognition be made in real time to expand the applications of the methods. Then, to detect an object in an image, the features used to feed a classifier need to be robust and easy to calculate [1]. In Computer Science, discrimination between children and adult faces in images is a problem not commonly tackled, it consists in the identification and boundary defining of children and adult faces using computational algorithms. Classification between children and adults is useful for safety, medical or surveillance purposes [2] and is often performed in a very controlled environment (insignificant camera rotations or illumination changes, absence of other objects in the background or absence of facial occlusions) [3]. On the other hand, the problem being a variant of pedestrian detection, initial steps to tackle this problem are performed using algorithms for human body detection in images, which can be useful for surveillance purposes. Although it's a different problem, some approaches aim to detect children from adults using age classification problems which is probably not the best suitable option using closed circuit television cameras [4]. Although, some works obtained good results in closed environments the problem here relies in the environment where the human face to be discriminated is at, not in the whole image, to allow the algorithms to extract facial features as wrinkles in an easy way. To attack the problem in this kind of environments where no subtle facial features can be extracted, we proposed a methodology for differentiating between children and adults in outdoor scenes.

The proposed method aims to extract features based on body and head sizes avoiding subtle features, such as the relation between the body and face or geometric features, to perform the differentiation between children and adult faces. Also, the effect of fitting a facial model to the initial head recognition was studied to improve the measures and their impact in the recognition rates. The results suggest that it's possible to improve actual methods based on body and head ratio with acceptable accuracy rates when the human body and face are present in an image.

II. METHODS

A methodology for children and adult face discrimination that is divided into three steps was proposed: (A) identification of the region of interest where the human body and their related face are detected. Next (B) the feature extraction step is performed, and finally (C) a classification model is applied. The method initiates in the identification of the regions of interest, which are the face and body of a human. For this purpose, a detector using Aggregate Channel Features [5], [6]trained with the INRIA person dataset was used [6]. For face recognition Viola-Jones algorithm was used [7]. Later geometrical features of detected face and body were extracted, the geometrical features are representing the distance between facial region of interest. Geometrical features were extracted obtaining the distance between the region of interest and normalizing them using the distance between the eyes.

Claudia Manfredi (edited by), *Models and Analysis of Vocal Emissions for Biomedical Applications : 11th International Workshop, December, 17-19, 2019*, © 2019 FUP, CC BY 4.0 International, published by Firenze University Press (www.fupress.com), ISSN 2704-5846 (online), ISBN (online PDF) 978-88-6453-961-4

A normalization step was studied to know it's effects in the classification rates. Finally, an SVM was used to perform the classification step. Fig. 1 depicts the proposed methodology for children and adult faces discrimination.

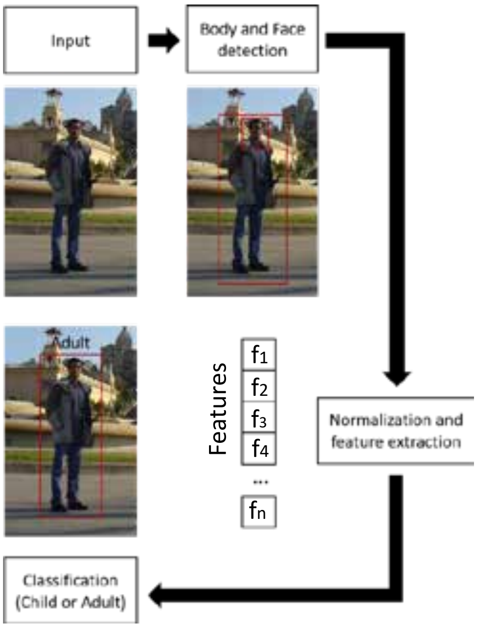

Fig. 1. Proposed methodology for children and adult discrimination using image processing and computational intelligence.

A. Body and head detection

The proposed method focuses on discrimination between children and adult faces. Some studies show that not only the proportion of the face can provide enough information for this purpose. The proportion of the face can provide enough information to discriminate between children and adults because the human head reaches its full size until a person becomes a teenager. For this reason, it's focused first on the identification of the body in the image because it occupies the largest region in the picture. The body identification also serves to reduce the area of search of the face to decrease the number of false positives. In other words, first it's focus on the identification of the body and within the detection area a face that is known to be present is looked for, if there were several detections, we select the best rated or the one that is most likely to be a face.

For the identification of human bodies in an image we use a detector trained using the INRA database with Aggregate Chanel Features. The detection algorithm provides the area where a human body is in the image, using this data, the Viola-Jones algorithm was used to detect all the faces present in the image. Viola-Jones extracts Haar-like characteristics to identify patterns in the image using a set of templates to match them into distinguishable facial features. An example of the templates used to compute Haar-like features are depicted in Fig. 2. Extracted characteristics pass through a series of weak classifiers templates trained with the Adaboost algorithm. The main characteristic of the decision algorithm is that all the weak classifiers must agree that a face is being detected to further reduce the detection of false positives.

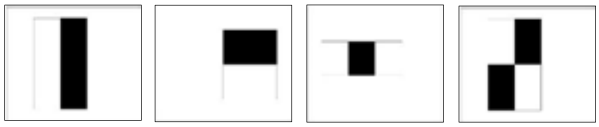

Fig. 2. Templates used to extract the Haar-like features used to detect faces in the image.

B. Feature extraction

Second step to perform the classification of children and adult faces consists in the extraction of features to feed a model of classification. Many features to represent the ratio between the size of the body and the size of the head were explored, the rationale behind this decision is due to the fact that studies remark that head size is bigger at younger ages; extraction of head size as it cannot be a standalone indicator because size in terms of pixels is a relative measure of sizes, it does not take into account the closeness of the object. For this reason, the relation of the body and the head is calculated using Eq 1. where $hsize$ represents the size of the detected bounding head in pixels and $bsize$ is the measured size of the body.

$$r = \frac{hsize}{bsize} \tag{1}$$

After obtaining the bounding box surrounding a face, using the area of where a face is located as an initialization step, a set of 68 facial landmarks were fitted into the face using Intraface models [5], [8]. Fig. 3a depicts an initial facial model before its fitting and 3b show the facial model fitted.

(a) (b)

Fig. 3. Matching between a facial model and an image. (a) shows the initialization and (b) the fitted model.

Considering that algorithms for children and adult faces only use the size of the detected bounding box provided by a face detection algorithm to obtain the features and aiming to obtain a better measure, the fitted facial model was used to improve the measure of not only the head but also the facial regions of interest such as: eyebrows, mouth, or nose, allowing the extraction of distance between the different facial regions of interest to improve classification rates. Then, in addition to the ration of head and body, 31 geometric features describing sizes of facial distinctive areas were extracted. The features were obtained computing either the distance or the angle between two points of the obtained facial landmarks. All the features were normalized respective to the distance obtained with points 40 and 43 of Fig. 4. Fig. 4 shows an enumeration of the 68 facial landmark points and Table 1 displays the coordinate pairs to compute the distance or angle.

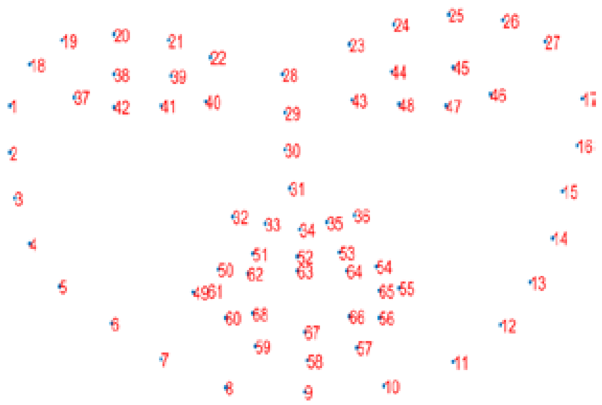

Fig. 4. Enumeration of the facial landmarks obtained using Intraface [5].

Table 1. Coordinate pairs to features computing.

Description	Pairs
Eyebrow (angles)	{20,28,42}, {25,28,47}
Eyebrow stretching	{19,37}, {26,46}
Eyes stretching	{37,38}, {38,39}, {39,40} {40,41}, {41,42}, {42,47} {43,44}, {44,45}, {45,46} {46,47}, {47,48}, {48,43}
Eyes opening	{38,42}, {39,41}, {44,48} {45,47}
Mouth stretching	{49,52}, {52,55}, {55,58}, {58,49}
Mouth openness	{63,67}
Facial sizes	{40,9}, {43,9}, {43,34}, {40,34}, {58,9}, {34,52}

C. Classification

The final stage of the methodology consists in the classification model, for this purpose, an SVM classifier with a linear kernel was used. For validation purposes a k-fold cross validation with 10 folds was used.

III. EXPERIMENTS AND RESULTS

For training and validation purposes, images taken from the INRIA database were used [7]. The INRIA database contains 1,805 64x128 images of humans cropped from a large set of photos. People in the images are usually standing but in any orientation. From the INRIA database a set of 145 images were taken. Images were a human or a child is standing with their face are visible were taken. 74 images of adults and 71 images of children were used for training and validation of the model.

Here the question of how the results of differentiating between children and adult faces can be improved is addressed. Two experiments were performed to analyze if the results can be improved: first, it was explore if adding geometrical features to the model could lead to a better representation and second, it was explore if fitting facial landmarks could improve the measure of head size, leading to better discrimination (BLHR). Figure 5 depicts how facial landmarks can lead to a better understanding of facial sizes and relations rather than the Viola-Jones algorithm output.

Fig.5. Facial landmarks fitted into an image using Intraface [8]. The initial adjustment was made using the located face with the Viola-Jones algorithm [7].

We performed experiments with (BNLHR) and without facial landmarks alignment aiming to mitigate the rotation and occlusions that can cause head sizes to be measured incorrectly. The normalization method consists in a heuristic based on affine transformations, first, a rotation aligns the face using the canthus of the eyes and the facial landmarks are normalized in the range [0,1] [9]. Fig. 6 depicts the normalization process. Finally, experiments adding the geometric features were performed (+GF). Results are shown in Table 2.

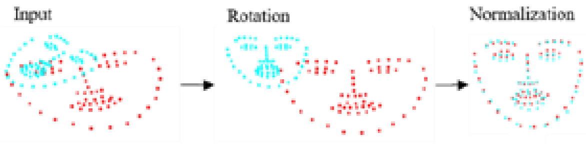

Fig. 6. Graphical representation of the normalization of the facial landmarks to reduce rotation and size noise.

Table 2. Results obtained with the proposed methodology.

Method	Accuracy
BHR [10]	61%
BHR+GF	78%
BLHR	64%
BLHR+GF	**86.4%**
BNLHR	74.6%
BNLHR+GF	84.7%

IV. DISCUSSION

Although there are not many works tackling the problem of the discrimination between children and adult faces, current solutions are focused in controlled environments which decreases the number of applications to use this kind of models. The proposed model archived a higher accuracy with the reference models [2], [4] fitting a set of 68 facial landmarks to the output of the recognition model to improving the measurement of the face and the classification rates. On the other hand, a limitation of the model is that the body and the face must be present in the image, a possible way to tackle this limitation may be to use an algorithm to recognize heads instead of a face recognition algorithm.

V. CONCLUSIONS

In this work a methodology for the classification of children and adult faces based on facial geometric features was presented. The proposed model not only use the relation between the head and body, but it also uses geometric features describing the normalized sizes of different facial regions of interest. For this approach the obtained accuracy rates increased from 61% to 78% which is seen as an improvement. Trying to alleviate the noise introduced by image size, a normalization was included to the methodology but without getting a significant improvement including or excluding the geometrical features. Improving the measure of the head was using a model fitted that uses the output of a facial recognition algorithm got the best improvement in terms of accuracy of our model and with the addition of the geometrical features betters results were obtained.

In conclusion, the results suggest that it's possible to improve discrimination between children and adult faces measuring the head by using facial landmarks instead of the bounding boxes provided by a facial recognition algorithm.

VI. ACKNOWLEDGMENTS

This research has been funded by project "Analysis and classification techniques of voice and facial expressions: application to neurological diseases in newborns and adults" from AMEXCID, supported by México and Italy governments.

REFERENCES

[1] K. Lee, C. Lee, S. A. Kim, and Y. H. Kim, "Fast object detection based on color histograms and local binary patterns," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, pp. 1–4, 2012.

[2] M. B. Hans Weda, "AUTOMATIC CHILDREN DETECTION IN DIGITAL IMAGES," pp. 1687–1690, 2007.

[3] W.-B. Horng, C.-P. Lee, and C.-W. Chen, "Classification of Age Groups Based on Facial Features," Tamkang J. Sci. Eng., vol. 4, no. 3, pp. 183–191, 2001.

[4] O. F. Ince, I. F. Ince, J. S. Park, J. K. Song, and B. W. Yoon, "Child and adult classification using biometric features based on video analytics," ICIC Express Lett. Part B Appl., vol. 8, no. 5, pp. 819–825, 2017.

[5] B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Aggregate channel features for multi-view face detection," IJCB 2014 - 2014 IEEE/IAPR Int. Jt. Conf. Biometrics, 2014.

[6] T. Surasak, I. Takahiro, C. H. Cheng, C. E. Wang, and P. Y. Sheng, "Histogram of oriented gradients for human detection in video," Proc. 2018 5th Int. Conf. Bus. Ind. Res. Smart Technol. Next Gener. Information, Eng. Bus. Soc. Sci. ICBIR 2018, pp. 172–176, 2018.

[7] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Surf. Sci., vol. 394, no. 1–3, 1997.

[8] F. De La Torre, W. S. Chu, X. Xiong, F. Vicente, X. Ding, and J. Cohn, "IntraFace," 2015 11th IEEE Int. Conf. Work. Autom. Face Gesture Recognition, FG 2015, no. July, 2015.

[9] O.-E. F. Morales-Vargas E., Reyes-García C.A., Peregrina-Barreto H., "Facial expression recognition with fuzzy explainable models," Model. Anal. Vocal Emiss. Biomed. Appl., 2017.

[10] O. F. Ince, J. S. Park, J. Song, and B. W. Yoon, "Child and Adult Classification Using Ratio of Head and Body Heights in Images," Int. J. Comput. Commun. Eng., vol. 3, no. 2, pp. 120–122, 2014.