

A New Color Image Database for Benchmarking of Automatic Face Detection and Human Skin Segmentation Techniques

Abdallah S. Abdallah, Mohamad Abou El-Nasr, and A. Lynn Abbott

Abstract—This paper presents a new color face image database for benchmarking of automatic face detection algorithms and human skin segmentation techniques. It is named the VT-AAST image database, and is divided into four parts. Part one is a set of 286 color photographs that include a total of 1027 faces in the original format given by our digital cameras, offering a wide range of difference in orientation, pose, environment, illumination, facial expression and race. Part two contains the same set in a different file format. The third part is a set of corresponding image files that contain human colored skin regions resulting from a manual segmentation procedure. The fourth part of the database has the same regions converted into grayscale. The database is available on-line for noncommercial use. In this paper, descriptions of the database development, organization, format as well as information needed for benchmarking of algorithms are depicted in detail.

Keywords—Image database, color image analysis, face detection, skin segmentation.

I. INTRODUCTION

THE detection and recognition of human faces is an important topic in computational image analysis. The last decade has shown dramatic progress in this area, with emphasis on such applications as human-computer interaction (HCI), biometric analysis, content-based coding of images and videos, and surveillance.

The availability of large sets of images is essential in these fields for the development and testing of new analysis techniques. For human face *recognition*, several large image databases are widely available (e.g., [11, 13, and 30]). The images in such databases typically show human subjects conveniently posed and properly illuminated, and with each face centered in the image and appropriately scaled.

For face detection, on the other hand, relatively few image databases are available to the research community. This might seem surprising, because the detection step ultimately must be performed in advance of further processing that involves the face. Unfortunately, the lack of a standard data set represents a substantial burden for researchers. Without reference test

cases, it is difficult to develop evaluation criteria for benchmarking new face detection algorithms and implementations.

Face detection algorithms have been categorized by Yang and Kriegman [1] into four classes: knowledge-based, feature invariant, template matching, and appearance-based. More details on face detection algorithms are presented in the survey by Hjelmås and Low [2]. No matter what class of algorithm is under consideration, a benchmarking data set has to satisfy many criteria. For example, the number of images must be adequate for both training and testing, and the images should present many different poses, backgrounds, lighting conditions, facial expressions, and variations of skin color. Databases with a wide variety of these attributes will help in assessing particular algorithms, and in making comparisons between different techniques.

This paper introduces a new image database that has been developed to address these needs. Known as the VT-AAST image database, it was developed jointly by researchers at Virginia Tech and the Arab Academy for Science, Technology, and Marine Transport. The database is unique in that every color image is accompanied by a corresponding image file that has been manually segmented to identify skin regions. Such “ground truth” information is expected to be extremely valuable in the development of automatic skin-detection systems. The high-level goal of this database is to fill the void that exists for face-detection research, by providing a large set of images that can serve as a standard for development and comparison of face-detection and skin-segmentation algorithms.

The rest of this paper is organized as follows: Section 2 presents a brief overview of the face-detection databases that are currently available, and describes the motivation for the creation of a new database. Section 3 depicts in detail the process of development, filing, formatting and documenting the VT-AAST database. Section 4 presents concluding remarks.

II. PREVIOUS IMAGE DATABASES FOR FACE DETECTION

This section discusses existing databases, describing some advantages and disadvantages of each. Several grayscale face image databases have been compiled. Among them are the well-known CMU test set (Rowley, et al. [3, 4]) and CMU profile test set (Schneiderman and Kanade [5]), which have

A. S. Abdallah is Direct PhD Student, Bradley Dept. of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia, USA.

M. A. El-Nasr is Assistant Professor, Computer Engineering Dept., Arab Academy for Science, Technology, and Marine Transport, Alexandria, Egypt.

A. L. Abbott Associate Professor, Bradley Dept. of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia, USA.

been used for benchmarking purposes (e.g., [6, 7, 8]).

There have been very few attempts, on the other hand, to build databases of color face images. One of those is the Kodak face image database (Loui, et al. [9]), with 80 images and with 90% of the faces shown in a frontal pose. Another is the UCD color face image database compiled by Sharma and Reilly [10]. These databases, and several others including our VT-AAST database, are listed in Table I.

TABLE I
 FACE DETECTION DATABASES

| Database | Quantity |
|-------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| FERET (Philips, et al.[11]) | 14,051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles. |
| MIT (Sung and Poggio [12]) | Set one includes 301 grayscale frontal and almost frontal face mugshots for 71 persons. Set 2 (also called MIT-23) has 23 images including 149 faces. |
| CMU test set (Rowley, et al. [3][4]) | 507 frontal faces from 130 grayscale images (also called CMU-130). In addition, there are 50 grayscale images containing 223 faces. |
| CMU profile test set (Schneiderman and Kanade [5]) | 208 grayscale images with profile views. |
| Kodak (Loui, et al.[9]) | 80 color images in TIFF format. Approximately 10% of the faces are not frontal. The face size ranges from 13×13 pixels to 300×300 pixels. |
| AR (Martinez and Benavente [13]) | This set includes 4,000 color images with faces of 126 individuals (70 men and 56 women). It uses a single person per image on a simplified uniform background. Format: TIFF, 768×576, 24 bits/pixel. |
| CVL Face Database[30] | 798 color images with faces of 114 persons, 7 images for each person. All are in JPEG format, and in size of 640*480 pixels. |
| UCD (Sharma and Reilly [10]) | 94 color images in GIF format. Contains 299 faces with 182 in frontal pose, 91 intermediate pose, and 26 in profile. Wide variety of face size, 3D pose, orientation, and occlusion. |
| VT-AAST | 286 color images in both JPEG and GIF formats, including 1027 faces. The poses are as follows: 515 frontal, 390 intermediate, 91 profile, and 29 over-profile. All GIF images are 300×225 pixels in size. For every color image, 2 manually segmented "skin" images are also provided in color and grayscale. |

Each is described in terms of the number of images and file formats, as well as attributes of image content such as pose and number of faces present.

Most of the existing databases have some drawbacks, for the purpose of face detection. For example, a traditional approach for human faces is to show one person per image against a simple, uniform background. (Most of the FERET, Kodak, AR and CVL images follow this convention.) This approach is helpful for face recognition work, but many face detection applications require an ability to accommodate more complex situations. Another aspect of many face databases is that they provide grayscale images only. This is quite limiting

for applications that require skin detection, which often relies heavily on color-based analysis (e.g., [14, 15, 16, and 17]).

To the best of our knowledge, the UCD color database [10] is the only publicly available database that satisfies our requirements of color imagery plus a variety of face sizes, poses, and background scenery. As shown in the table, the VT-AAST database is somewhat larger than the UCD database. The VT-AAST database also provides some additional advantages over existing databases, and these advantages will be described in the next section.

III. THE VT-AAST COLOR FACE IMAGE DATABASE

This section describes the VT-AAST database in more detail. The images were captured using several consumer-grade digital cameras from different vendors. Image-array sizes ranged from 3 to 5.2 megapixels. Part of the reason for using these "point-and-shoot" cameras instead of professional-grade imagers was to mimic expected applications of face-detection systems. We assume that most of these applications will expect to rely on relatively inexpensive cameras, such as those mounted on laptop computers, or those typically used for surveillance in public places.

Our cameras provided images in JPEG format. (Most of today's consumer-grade cameras do not provide images in uncompressed form.) These original images are given in part one of the database. These images were further compressed to the Graphics Interchange Format (GIF), with 300×225 pixels per image to form part two of the database. Therefore, each photograph is provided in both JPEG and GIF formats. In addition, a manual segmentation procedure has been used to identify skin regions for each case. These segmented images are available in two forms: with color information retained for pixels with the skin regions, and with only grayscale information retained for the skin regions. All of the segmented images are stored as GIF files of size 300×225. These segmented cases represent the third and fourth parts of the VT-AAST database, respectively.

An example from the database is shown in Fig. 1. The choice of image size matches those commonly used for video analysis (e.g., [8, 23, and 29]). The figure illustrates the results of the manual segmentation process.

The current size of the database is 286×4 images, where the factor of 4 represents the different image formats provided for each original photograph. These were obtained in both indoor and outdoor environments, with widely varying backgrounds. Several examples from the database are shown in Fig. 2. The goal was to provide a wide variety in the following:

- 3-dimensional pose (frontal, intermediate, profile, and "over-profile").
- 2-dimensional orientation (upright and rotated).
- Facial sizes in the images.
- Facial expressions.
- Luminance conditions.
- Occlusion.
- Structural components (hair, beards, mustaches,

- glasses).
- Gender.
- Human race (White, Black, Asian, etc., as specified by the U.S Census Bureau [18]).
- Scene and background (outdoor and indoor, simple and cluttered).



Fig. 1 Example images from the VT-AAST database. (a) Color image, available in both JPEG and GIF forms. (b) Corresponding segmented image in which the skin regions have been retained, but all other pixels have been changed to a dark background value. The segmented image is also available with the skin regions represented using grayscale pixel values.

Table II provides a breakdown of these categories, as represented in the database. For 3D pose, we follow the definitions given by Reilly [10]. Essentially, for a frontal view the subject faces the camera, and for a profile view a plane parallel to the image plane divides the face equally. The intermediate pose is sometimes called a three-quarter view, and indicates that the head has turned slightly away from the camera. Finally, the over-profile view refers to the case that the subject is looking away from the camera, so that the ear and cheek are visible but the tip of the nose is not.

Structural components typically refer to obstructions that cover the face partially. Typical examples include facial hair and eyeglasses. Racial information is categorized in the database according to the latest definitions given by U.S Census Bureau [18].

The rest of this section is a summary of the tools and formatting processes that were used to build the VT-AAST database. To store image details, we used both Microsoft Excel [26] and the Microsoft Access database engine [27]. Our database consists of a single table whose public key is the image ID; each image entry has specific fields that correspond to one of the attributes described above. Fig. 3 presents an example record, with values shown for the different fields.

The Excel spreadsheet consists of three columns. The first column contains a unique identifier for each set of 4 images, corresponding to 1 photograph. The second column states the total number of faces in each image. In some cases, a few face regions of an image are not included in this count. This occurred when a face was not noticed during the manual segmentation step because of color and brightness patterns that caused it to blend into the background. To indicate these cases, a “+” symbols appears in the 3rd column of the

spreadsheet. The manual extraction was done using the color range selection tool in Adobe Photoshop CS2 [22].

A further reason for using the Microsoft tools was to facilitate integration with the Matlab computing environment [28]. Matlab’s Visual Query Builder provides a capability for database querying, and Matlab’s Excel Link tool allows the incorporation of Matlab commands within Excel spreadsheets.

IV. CONCLUSION

This paper has introduced a new image database that fills an important need for researchers who are interested in human face and skin detection. Created as a joint effort by Virginia Tech and the Arab Academy for Science, Technology, and Marine Transport, the VT-AAST color image database is large enough to be used for both training and testing. The images are sufficiently varied in content to provide a wide range of poses, skin colors, illumination conditions, and background complexity. This is distinguished from previous databases by the extensive tabulation of image content (number of faces for each image, etc.), and by providing manually segmented “ground truth” images of skin regions for each case.

The number of faces in the images is relatively large, and therefore facilitates the identification of separate training and testing sets. Example approaches include support vector machines (e.g., [20, 21]), artificial neural networks (e.g., [3, 4, 16]), AdaBoost (e.g., [19]) and FloatBoost ([24, 25]). A wide range of skin colors appear in the images, and this presents a challenge to skin-detection algorithms, as well as an opportunity for further research.

The VT-AAST database can serve as a standard for comparison for future face- and skin-detection research. Access to the images is provided through the Internet. They are currently available at <http://filebox.vt.edu/users/yarab/VT-AAST Database>, although a password is needed to access the site. Access will be provided to users who sign an agreement for noncommercial use of the images.

REFERENCES

- [1] M. H. Yang, and D. J. Kriegman, “Detecting Faces in Images: A Survey”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, Jan. 2002, pp. 34 - 58.
- [2] E. Hjeltnæs, and B. K. Low, “Face Detection: A Survey”, *Computer Vision and Image Understanding*, Vol. 83, No. 3, Sept. 2001, pp. 236-274.
- [3] H. Rowley, S. Baluja, and T. Kanade, “Neural Network-Based Face Detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, Jan. 1998, pp. 23-38.
- [4] H. Rowley, S. Baluja, and T. Kanade, “Rotation Invariant Neural Network-Based Face Detection”, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1998, pp. 38-44.
- [5] H. Schneiderman, and T. Kanade, “A Statistical Method for 3D Object Detection Applied to Faces and Cars”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, 2000, pp. 746-751.
- [6] H. Kruppa, M. A. Bauer, and B. Schiele, “Skin Patch Detection in Real-World Images”, *Proc of the DAGM-Symposium*, 2002, pp. 109-116.
- [7] P. Viola and M. Jones., “Robust real-time face detection”, *International Journal of Computer Vision (IJCV)* 57(2), 2004, pp. 137-154.

- [8] T. Theoharides, G. Link, N. Vijaykrishnan, M. J. Irwin, and W. Wolf, "Embedded Hardware Face Detection", Proc. 17th International Conference on VLSI Design, Jan. 2004, p.133.
- [9] A. C. Loui, C. N. Judice, and S. Liu, "An Image Database for Benchmarking of Automatic Face Detection and Recognition Algorithms", Proc. IEEE Conference on Image Processing, Vol. 1, 1998, pp.146-150.
- [10] P. Sharma, and R.B. Reilly, "A Color Face Image Database for Benchmarking of Automatic Facial Detection Algorithms", Proc. 4th European Conference of Video/Image Processing and Multimedia Communications, July 2003, pp, 423 - 428.
- [11] P. J. Phillips, P. Rauss, and S. Der, "FERET (Face Recognition Technology) Recognition Algorithm Development and Test Report", Technical Report ARL-TR 995, U.S. Army Research Laboratory, October 1996.
- [12] K-K. Sung, and T. Poggio, "Example-Based Learning for View-based Human Face Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, Jan. 1998, pp. 39-51.
- [13] A. Martinez, and R. Benavente, "The AR Face Database," Technical Report CVC 24, Purdue Univ., 1998.
- [14] R-L. Hsu, and M. Abdel-Mottaleb, "Face Detection in Colour Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, May 2002, pp. 696 - 706.
- [15] C. Garcia, and G. Tziritas, "Face Detection Using Quantized Skin Colour Regions Merging and Wavelet Packet Analysis", IEEE Transactions on Multimedia, Vol. 1, No. 3, September 1999, pp. 264 - 277.
- [16] C. Lin, "Face Detection by Colour and Multilayer Feedforward Neural Network", Proc. IEEE International Conference on Information Acquisition, June-July 2005, pp. 518-523.
- [17] K. Sandeep and A.N. Rajagopalan, "Human Face Detection in Cluttered Color Images using Skin Color and Edge Information", Proc. Indian Conference on Computer Vision, Graphics and Image Processing, Dec. 2002.
- [18] U.S. Census Bureau, 2000 Census of Population, Public Law 94-171 <http://factfinder.census.gov/>.
- [19] Y. Wei, X. Bing, and C. Chareonsak, "FPGA Implementation of ADABOOST Algorithm for Detection of Face Biometrics", Proc. IEEE International Workshop on Biomedical Circuits and Systems, Dec. 2004, pp. S1/6.17-20.
- [20] E. Osuna, R. Freund, and F. Girosit, "Training Support Vector Machines: An Application to Face Detection", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 1997, pp. 130-136.
- [21] Yongmen Li, S. Gong, and H. Liddell, "Support Vector Regression and Classification-based Multi-view Face Detection and Recognition", Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition, March 2000, pp. 300-305.
- [22] Adobe Photoshop CS2, <http://www.adobe.com/products/photoshop/>.
- [23] O. Bernier, M. Collobert, R. Feraud, V. Lemaire, J. E. Viallet, and D. Collobert, "MULTRAK: A System for Automatic Multiperson Localization and Tracking in Real-time", Proc. International Conference on Image Processing, ICIP 98. Vol. 1, Oct. 1998, pp. 136-140.
- [24] S.Z. Li and Z. Zhang, "FloatBoost Learning and Statistical Face Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, Issue 9, Sept. 2004, pp. 1112-1123.
- [25] S.Z. Li, L. Zhu, Z. Zhang, and H.-J. Zhang, "Learning to Detect Multi-view Faces in Real-time", Proc. 2nd International Conference on Development and Learning, June 2002, pp. 172-177.
- [26] Microsoft Office Excel, <http://office.microsoft.com/en-us/excel/default.aspx>.
- [27] Microsoft Office Access, <http://office.microsoft.com/en-us/access/default.aspx>.
- [28] Matlab Online, <http://www.mathworks.com/products/matlab/>.
- [29] T. Theoharides, N. Vijaykrishnan, and M. J. Irwin, "A Parallel Architecture for Hardware Face Detection", Proc. IEEE Computer Society Annual Symposium on Emerging VLSI Technologies and Architectures, March 2006.
- [30] Computer Vision laboratory, Faculty of Computer and Information Science, University of Ljubljana, Slovenia, <http://www.lrv.fri.uni-lj.si/facedb.html>.



Fig. 2 More example images from the VT-AAST database. (a-c) Many are group shots, providing several faces to detect, in a variety of indoor and outdoor environments. (d) Unusual orientations are present in some of the images, along with a mixture of structural components such as sunglasses and facial hair. (e) Different 3-D orientations of the individuals' heads present are shown. (f) Three-dimensional perspective may cause faces to appear at dramatically different sizes in the images.

| image name | total faces | frontal | intermediate |
|------------|-----------------|---------|--------------|
| 214 | 11 | 11 | |
| profile | over profile | upright | |
| | | 2 | |
| rotated | beard | glasses | |
| 9 | 4 | 4 | |
| female | male | white | black |
| 4 | 7 | 8 | 2 |
| asian | native hawaiian | other | |
| | | 1 | |
| occluded | | | |
| 5 | | | |

Record: 214 of 286

Fig. 3 Example record from the VT-AAST database. This type of tabulated information is provided for every image in the database. Blank fields represent values of zero.

TABLE II
 STATISTICAL DATA FOR VT-AAST IMAGE DATABASE

| Total Number of Images = 286 | | | | | | | | | | | | | | | | |
|------------------------------|--------------|---------|--------------|-------------|---------|--------|--------|-------|-------|-----------------|-------|----------|-------|-----------------------|-------------------|-----------|
| Total Number of Faces = 1027 | | | | | | | | | | | | | | | | |
| 3D Pose | | | | Orientation | | Gender | | | | Race | | | | Structural Components | | |
| Frontal | Intermediate | Profile | Over profile | Upright | Rotated | Male | Female | White | Black | American Indian | Asian | Hawaiian | Other | Sunglasses or Glasses | Mustache or Beard | Occlusion |
| 515 | 390 | 91 | 29 | 885 | 142 | 532 | 495 | 828 | 102 | 1 | 0 | 0 | 93 | 334 | 142 | 106 |

The 286 photographs contain 1027 instances of human faces in a variety of situations and environments.