Comments on the CASIA Version 1.0 Iris Data Set

P. Jonathon Phillips, *Senior Member*, *IEEE*,
Kevin W. Bowyer, *Fellow*, *IEEE*, and
Patrick J. Flynn, *Senior Member*, *IEEE*

Abstract—We note that the images in the CASIA version 1.0 iris data set have been edited so that the pupil area is replaced by a circular region of uniform intensity. We recommend that this data set no longer be used in iris biometrics research unless there is a compelling reason that takes into account the nature of the images. In addition, based on our experience with the Iris Challenge Evaluation (ICE) 2005 technology development project, we make recommendations for reporting results of iris recognition experiments.

Index Terms—Iris recognition, biometrics.

THE paper by Ma et al. [1] made a number of contributions to iris biometrics, including a novel iris recognition algorithm, a benchmark of several approaches to iris recognition, and the popularization of the CASIA version 1.0 iris image data set. The Acknowledgments section of the paper states that "A public version of the CASIA Iris Database is available from http://www.sinobiometrics.com." The iris image data set known as CASIA version 1.0 [2] was made available and widely distributed.

The CASIA version 1.0 data set is actually not the same as the data set used in [1]. The CASIA version 1.0 data set contains 756 images, whereas the results in [1] are based on 2,255 images. Ma et al. [1] provide a detailed description of the iris image acquisition process. However, the description omits one key feature of the CASIA version 1.0 images: Each image has been edited so that the original pupil area is "replaced with a circular region of constant intensity to mask out the specular reflections from the near infrared illuminators." The authors state that the experimental results in [1] were obtained using unedited versions of the images [3]. However, all of the images appearing in the figures in [1] have been edited in the same manner as the images in CASIA version 1.0.

Fig. 1 shows an example of a CASIA version 1.0 image, a different display of the same image that highlights the edited region, and examples of unedited images from the more recent CASIA version 3.0 data set [4] and the Iris Challenge Evaluation (ICE) 2005 data set [5]. The particular intensity value for the edited pupil regions in the CASIA version 1.0 data set varies between images, but for each image, the original pupil area has been overwritten with a dark, circular region of constant intensity value.

Recognition results on the CASIA version 1.0 data set have been extensively reported in the literature [6]. Many of these algorithms are presented by the authors as fully automatic in that they process iris images without manual intervention. In one sense, these claims

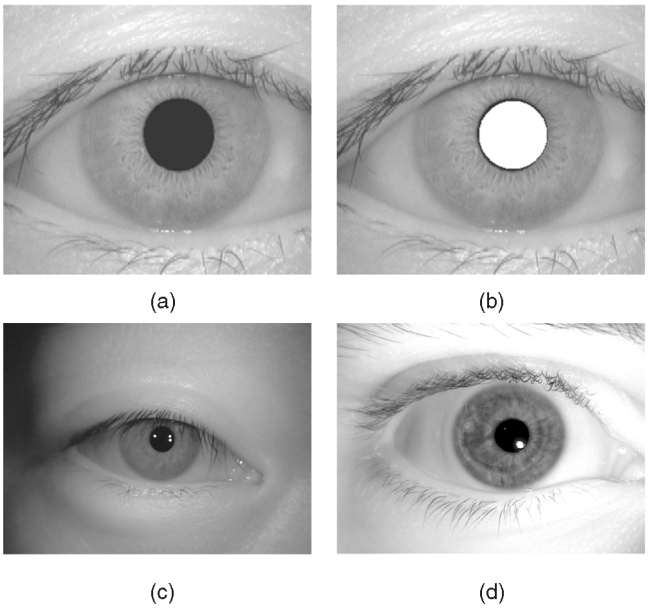

Fig. 1. Example iris images showing editing and specular reflections. Image (a) is image "001_1_1.bmp" from the CASIA version 1.0 data set, displayed as distributed by CASIA. Image (b) is the same image, but with the circular region of constant intensity in the pupil area displayed in white so as to make the editing obvious. Image (c) is from the more recent CASIA data set and shows the specular highlights from the illuminator. Image (d) is from the ICE 2005 data set, also showing specular highlights of the type normally visible.

are correct since the editing of the images was performed prior to them being distributed. However, since the pupil regions of the CASIA version 1.0 images were transformed into constant-intensity circular regions prior to distribution of the data, automatic detection of the pupil-iris boundary was made artificially simple. Original iris images have a vastly more complex distribution of intensity values in the pupil area, typically including specular highlights from the illumination source, as seen in Fig. 1. Because the editing of the CASIA version 1.0 images is not documented in [1] or [2], there is the potential for confusion in interpreting the nature of the data set and performance reported on the data set.

Table 1 provides a list of publicly available iris data sets for use in iris recognition experiments.[1] Because of the availability of these data sets, we recommend that reporting experimental results on the CASIA version 1.0 data set be discontinued unless there is a compelling scientific reason to use it. In any case, publications using the data set should disclose the fact that the images have been edited to have artificially simple pupil regions.

Recent results in iris recognition suggest that additional sophistication should be adopted in designing and reporting experiments. Initial results from the ICE 2005 suggested there might be a correlation between the left and right irises [5]. Also, various researchers have commented that race and eye color have the potential to influence performance. However, currently, there is no definitive scientific evidence supporting any of these claims. To put iris performance on a more solid scientific footing, it is recommended that performance be broken out by left and right

- *P.J. Phillips is with the National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8940. E-mail: jonathon@nist.gov.*
- *K.W. Bowyer and P.J. Flynn are with the Department of Computer Science and Engineering, University of Notre Dame, 384 Fitzpatrick Hall, Notre Dame, IN 46556-5637. E-mail: kwb@cse.nd.edu, flynn@nd.edu.*

Manuscript received 26 July 2006; revised 22 Mar. 2007; accepted 22 Mar. 2007; published online 28 Mar. 2007.
Recommended for acceptance by D. Fleet.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0559-0706.
Digital Object Identifier no. 10.1109/TPAMI.2007.1137.

1. Specific data sets identified in this paper were used in order to perform the analysis described in this document. In no case does such identification imply recommendation or endorsement by the US National Institute of Standards and Technology or the University of Notre Dame nor does it imply that the data sets identified are necessarily the best available for the purpose. Authors should examine the data set documentation and images to verify that they are appropriate for their experiments.

TABLE 1
Iris Data Sets Available to the Iris Research Community as of July 2006

Dataset	Contact point	Number of subjects	Number of images
Bath	http://www.bath.ac.uk/elec-eng/research/sipg/irisweb/index.htm	400 eyes	8000
CASIA 3	http://nlpr-web.ia.ac.cn/english/irds/irisdatabase.htm	60	2400
ICE 2005	http://iris.nist.gov/ice	132	2953
UBIRIS	http://iris.di.ubi.pt	241	1877
UPOL	http://phoenix.inf.upol.cz/iris	64	384
WVU	arun.ross@mail.wvu.edu	~400	~3000

irises, race, eye color, and whether or not the person is wearing contact lenses or eyeglasses. To be able to compare performance among different approaches, we strongly recommend that papers include experimental results on publicly available data sets and deposit the training and testing sets and source code for scoring experiments as online supplemental material. These recommendations will be a significant step forward by allowing repeatable experiments and systematic accumulation of scientific understanding in iris recognition.

REFERENCES

[1] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Personal Identification Based on Iris Texture Analysis" *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 25, no. 12, pp. 1519-1533, Dec. 2003.
[2] "Specification of CASIA Iris Image Database(ver 1.0)," *Chinese Academy of Sciences*, http://www.nlpr.ia.ac.cn/english/irds/irisdatabase.htm, Mar. 2007.
[3] T. Tan, personal communication, 2006.
[4] "Note on CASIA-Iris v.3," *Chinese Academy of Sciences*, http://www.cbsr.ia.ac.cn/IrisDatabase.htm, Mar. 2007.
[5] P.J. Phillips, "Test Director's ICE 2005 Presentation," http://iris.nist.gov/ice/, Mar. 2006.
[6] K.W Bowyer, K. Hollingsworth, and P.J. Flynn, "A Survey of Iris Image Analysis for Biometrics" CSE technical report, Univ. of Notre Dame, 2006.