

NIST Technical Note 2154

**Iris Recognition on
Noah Kalina's *Everyday***

James R. Matey

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.TN.2154>

NIST
**National Institute of
Standards and Technology**
U.S. Department of Commerce

NIST Technical Note 2154

Iris Recognition on Noah Kalina's *Everyday*

James R. Matey
*Information Access Division/Image Group
Information Technology Laboratory*

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.TN.2154>

April 2021



U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
*James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce
for Standards and Technology & Director, National Institute of Standards and Technology*

Certain commercial entities, equipment, software or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology (NIST), nor is it intended to imply that the entities, materials, software or equipment are necessarily the best available for the purpose.

NIST policy requires that all research is reviewed and approved by the human subjects and/or vertebrate animal review process before research involving human or animal subjects is conducted. The National Institute of Standards and Technology Research Protections Office reviewed the protocols for this project and determined they are either (1) “not human subjects research (NHSR)” or (2) they meet the criteria for “exempt human subjects research (EHSR)” – as defined in 15 CFR 27, the Common Rule for the Protection of Human Subjects. The NIST protocols relevant for this paper are:

- ITL-17-0004 Evaluation of Biometric Algorithms on Public Images from Internet (EHSR)
- ITL-2019-0156 IREX X: Ongoing Evaluations of Iris Recognition (NHSR)

All images shown in this paper have been checked for conformance to protocols/licensing regarding publication.

The high resolution data from Noah Kalina’s *Everyday* project that is used in this analysis is copyrighted by Noah Kalina, credit Noah Kalina and used by permission and under license. NIST has no license to redistribute the high resolution images. Images from the *Everyday* project that are shown in this paper are lower resolution images available online under Creative Commons licenses.

National Institute of Standards and Technology Technical Note 2154
Natl. Inst. Stand. Technol. Tech. Note 2154, 17 pages (April 2021)
CODEN: NTNOEF

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.TN.2154>

Abstract

In January 2000, Noah Kalina began taking a photograph of himself every day – and has continued for more than 20 years. A time-lapse video of his images can be seen online: [Everyday](#) . These images, and images from similar projects, provide us with interesting opportunities to explore the effects of time lapse on iris recognition employed on images that were not originally intended for iris recognition.

NIST obtained a license from Kalina to use a subset of original, high resolution, digital images in biometric studies. This paper is our first published analysis of those images. Our license does not permit redistribution of the high resolution images.

Iris images were extracted from the visible light, digital, face images and compared using iris recognition algorithms originally designed for near infrared images from purpose built iris cameras. Although the majority of the extracted iris images in this dataset did not provide solid matches, a significant fraction did. This demonstrates that iris recognition is *possible* on iris image pairs from visible light images that were not collected for the purpose of iris recognition. Such capability may be useful in forensic applications.

Key words

iris recognition; biometrics; match score variation; comparison score variation; visible light iris recognition; forensic iris.

Table of Contents

1	Introduction	1
2	Description of Data and Analysis	1
3	Results: Kalina Everyday	5
4	Conclusions and Future Plans	11

List of Tables

Table 1	Description of dataset	1
---------	------------------------	---

List of Figures

Fig. 1	Example of an image from the <i>Everyday</i> project by Noah Kalina.	2
Fig. 2	Example of an image from a commercial, off-the-shelf iris camera.	3
Fig. 3	Histogram of interpupillary distances for the Kalina <i>Everyday</i> images	4
Fig. 4	Match quality histogram for high quality image pairs.	5
Fig. 5	Comparison quality for all mated and un-unmated image pairs.	6
Fig. 6	Comparison score histograms for mated and non-mated image pairs.	7
Fig. 7	Match score as a function of time delta for mated and non-unmated image pairs.	8
Fig. 8	Match scores as a function of elapsed time for mated and unmated image pairs.	9
Fig. 9	ROC curves.	10

1. Introduction

In January 2000, Noah Kalina began taking a photograph of himself every day – and has continued for more than 20 years. A video of the images can be seen online: [EveryDay](#)¹; an example image can be seen in figure 1. These images, and images from similar projects, provide opportunities to explore the effects of elapsed time on iris recognition employed on images that were not originally intended for iris recognition.

NIST obtained a license from Kalina to use a subset² of the original, high resolution, digital images in biometric studies.

It is important to note that the algorithms used in the current analysis were not designed to work with the iris images that we extracted from the EveryDay project: the algorithms were designed to work with near-infrared images collected by iris cameras that are built for compliance with the ISO/IEC 29794-6 standard[1].

2. Description of Data and Analysis

Table 1 summarizes some important characteristics of the dataset used in this paper; in this context FTE means *failure to enroll*: the algorithm could not process the image. It is important to note that the numbers in this table represent a snapshot of the dataset at a particular point in time. Kalina continues to collect images for his project.

Table 1. Description of dataset

Number of images	1317
Number of iris images	2634
Number FTE, iris2pi	10
Number FTE, VeriEye	48
Time span	7 years
Image type	face
Image size	2816x2112
Illumination	Visible
Camera	Nikon Coolpix S10
Notes	2009-2015, half years

The high resolution, color, *Everyday* images required pre-processing to convert them to a form useful for commercial, off-the-shelf (COTS) iris recognition algorithms:

1. the iris images, including the immediate periocular region, were extracted from the face images;
2. the iris images were scaled to 640x480;
3. the iris images were converted from RGB to 8 bit gray-scale to approximate the image format provided by standard commercial iris cameras as illustrated in figure 2. Our license does not permit us to publish the extracted iris images from the original, high resolution, EveryDay images.

¹<https://petapixel.com/2020/01/13/photographer-noah-kalina-has-shot-a-selfie-every-day-for-20-years-now>

²Seven half years (January to June) comprising a six year span, 2009-2015.



Fig. 1. Low resolution example of an image from the *Everyday* project by Noah Kalina. Screen capture from <http://noahkalina.com/582/597>. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License, see <http://noahkalina.com/566>.

The preprocessing can be affected by the camera used for collection. Over the years Kalina made use of different cameras. For the images from 2009-2015, he used a Nikon CoolPix S10³, as recorded in the meta-data of the images. The images from the S10 have dimensions of 2816x2112 pixels, approximately 6 mega-pixels.

To carry out the preprocessing we used a python script based on an example, `face_landmark_detection.py`, in the `dlib` library⁴ to locate facial landmarks for each of the *EveryDay* images. We then input the images and landmarks into a program described in NIST-TN-2098[3] to extract and scale the iris images to a standard 640x480 format; convert to gray scale by extracting the red channel⁵; and compute the interpupillary distances in the original images.

For the example image in figure 1, the interpupillary distance is about 0.15 of the image width; the iris diameter is about 0.25 of the interpupillary distance. Hence in images with a width of 2816 pixels, and similar pose, we expect interpupillary distances of approximately $0.15 * 2816 \sim 420$ pixels and iris diameters in the original image of approximately $0.15 * 0.25 * 2816 \approx 100$ pixels. This is about half the resolution seen in a high quality image from a purpose built iris camera, e.g. figure 2. It is also above the approximately 50 pixel diameter at which Quinn et al. saw strong false non match rate (FNMR) degradation in their study of the effects of reduced resolution[4] and comparable to the 100 pixels/cm resolution that Matey et al. reported in the *Iris On the Move* systems[5].

Figure 3 shows the actual distribution of interpupillary distances as determined from the landmark detection algorithm. There is a secondary peak below 200 pixels. Visual inspection of those images showed that for an interval, the faces were captured at a pose that resulted in fewer pixels across the face. This is only apparent in the high resolution images. The images with the low interpupillary distances were excluded from subsequent analysis because they were taken under different conditions than the bulk of the images.

³See manufacturer's information at <https://imaging.nikon.com/lineup/coolpix/s/s10/index.htm>

⁴See <https://www.pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/>

⁵Using the red channel is a better approximation to the near infrared used in purpose built iris cameras

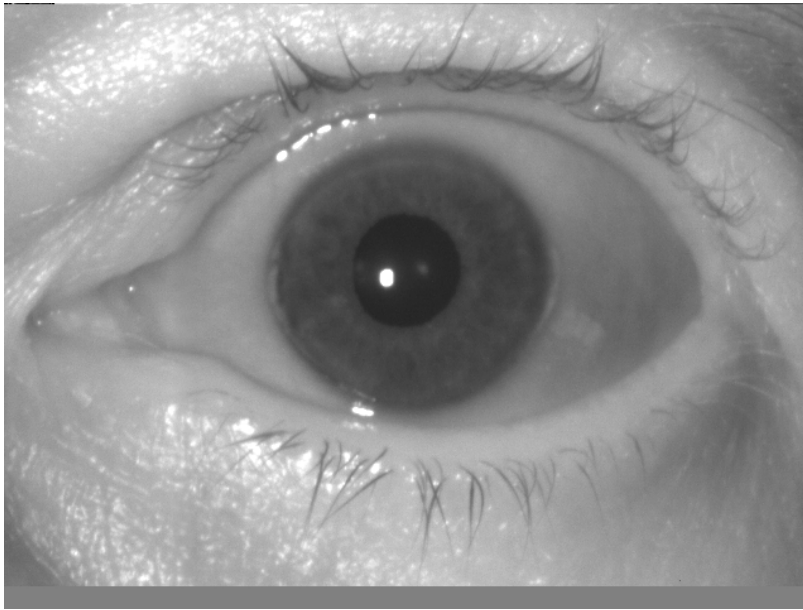


Fig. 2. Example of an iris image generated by a commercial off-the-shelf iris camera. The process described in the text generates images of this format, though of lower resolution and lower overall quality. The image in this case has approximately 200 pixels across the iris. We use this image as an illustration because our license does not permit publication of the iris images extracted from the high resolutions EveryDay images. This image obtained from a consented subject who provided a release for the use of the image in publications; it is from the NIST IrisDaily dataset[2].

The extracted images were scaled to 640x480 using a region of interest about the eye that resulted in iris diameters in the scaled images of 225 ± 4 pixels, as determined by the iris2pi algorithm; we obtained similar results for VeriEye.

The iris images were then processed in pairs to generate comparison scores using a commercial version of the iris2pi iris recognition algorithm whose internals have been described in detail by Daugman[6] [7] [8] [9] and variants of which have been implemented by academics, see for example Masek[10].

We also used another commercial iris recognition algorithm, Neurotechnology/VeriEye Version 10, for which the internals are not well known.

We used R[11] to produce the plots and statistical tests presented below.

In the plots below, the comparison scores are raw scores as reported by the algorithms for image pairs where both images satisfied the interpupillary criterion illustrated in figure 3. The iris2pi scores are dissimilarity scores⁶; they are fractional Hamming distances that have not been normalized using the Daugman correction⁷ for number of bits compared. Any iris2pi comparisons with less than 500 bits compared were discarded⁸; the nominal number of bits compared for the Daugman correction is 911. Though the Daugman correction is valuable in large scale applications to ensure the stability of the FMR statistics, for cases where the (dis)similarity between a small number of image pairs is of primary interest, it can give rise to confusion. As an example, for the case of three images which are essentially identical to

⁶Lower dissimilarity scores mean a better match between the images being compared.

⁷Equation 6 in [8]

⁸The total number of comparisons was 3441376 for iris2pi; after the application of the interpupillary and comparison quality criteria, we were left with 3441262 comparisons.

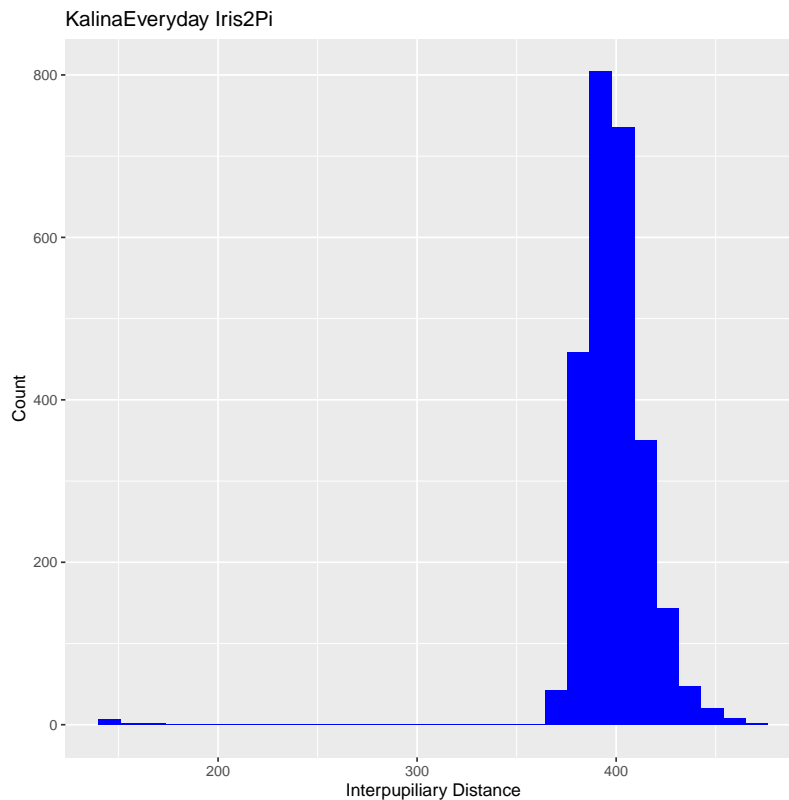


Fig. 3. Histogram of interpupillary distances for the Kalina *Everyday* images, as measured on the 2816x2112 original images. Visual inspection of the original images showed that the small peak at low interpupillary distance was due to a change in pose that occurred for a short period. The images with the low interpupillary distance were filtered out of the subsequent analysis.

within some noise level and for which two have no occlusion and one is partly occluded, the Daugman correction can introduce a significant difference between the score for the two non-occluded images and the scores between the occluded image and either non-occluded image that varies with occlusion, is not directly related to the similarity of the images and can be difficult to interpret.

The VeriEye scores are proprietary similarity scores for which we lack information to make detailed analysis.

For figures 6, 8 & 9, we selected the best left and right images as the reference images and made all comparisons against those images. The selection was made by rank ordering all of the mated comparison scores and selecting as reference images the one left and one right images that most often contributed to the 100 best matches. The images were

- iris2pi: 2015-05-13 (left) 2015-03-10 (right)
- VeriEye: 2015-05-19 (left) 2014-06-25 (right)

This simplifies the interpretation of variation in comparison score vs time delta and reduces the clutter on the point plots.

Some algorithms can provide a comparison quality metric that provides insight into how reliable the comparison score is expected to be. For iris2pi one such metric is the count of the number of usable

bits in the comparison score calculation. Figure 4 shows the comparison quality that can be routinely obtained with images of the quality of figure 2. There are 2048 bits (8 radial bands, 128 angular locations per band, 2 bits per location) in an iris2pi template; this leads to a theoretical maximum in usable bits of 2048. However, in practice, procedures in the comparison code that discard less reliable bits reduce the experimental maximum to about 1500. Hollingsworth, et al. called these bits *fragile bits* in their 2008 paper[12]. A careful read of the Daugman papers and patents including [6, 7, 13] suggest that a major reason for the reduction is the introduction of guard bands that mask noise in the digitization of phase near the horizontal and vertical axes, where a small variation in the phase can flip a bit from 0 to 1 by moving the phasor across the quadrant boundary. We do not have room here to present a full discussion; we refer the reader to the cited references.

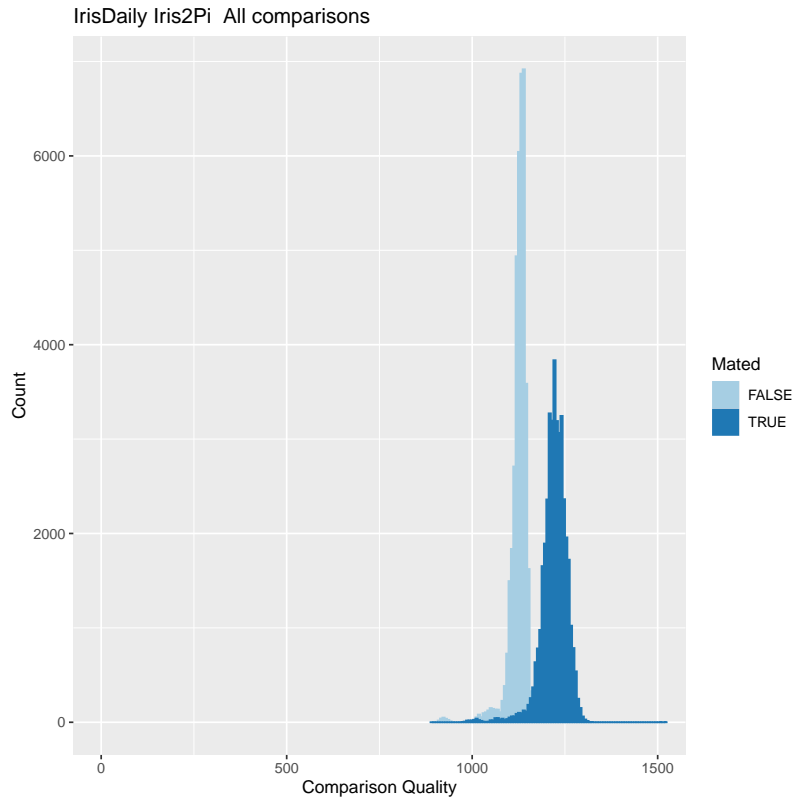


Fig. 4. Match quality quality histogram for high quality image pairs from the NIST IrisDaily dataset[2]. Match quality is *not* comparison score. It is a measure of the quality of the images/templates used in the comparison.

Daugman used that metric to adjust comparison scores in his iris2pi algorithm[8]. As noted above, we have not made such adjustments in this paper.

3. Results: Kalina Everyday

Figure 5 shows the comparison quality for mated and non-mated image pairs as reported by the iris2pi algorithm we used; the version of the VeriEye algorithm used here does not provide an equivalent metric. This plot represents comparisons of all images against all images. We see that matches based on iris images extracted from the Kalina Everyday images have lower quality than those seen in figure 4. It is important to understand that the number of bits used in a comparison are critical to reliability of the comparison. To illustrate: at the limit where only one bit was being compared, the binomial statistics

which are the basis of iris2pi[8] would give a 50% chance of all the bits matching by random chance for two un-related irides; for the case of N compared bits, the chance of all the bits matching by chance for two un-related irides is 2^{-N} . Hence, in the absence of the previously discussed Daugman correction, we need to impose a comparison quality cutoff. We imposed a minimum comparison quality criterion of 500 bits; comparisons with lower quality were discarded in this analysis.

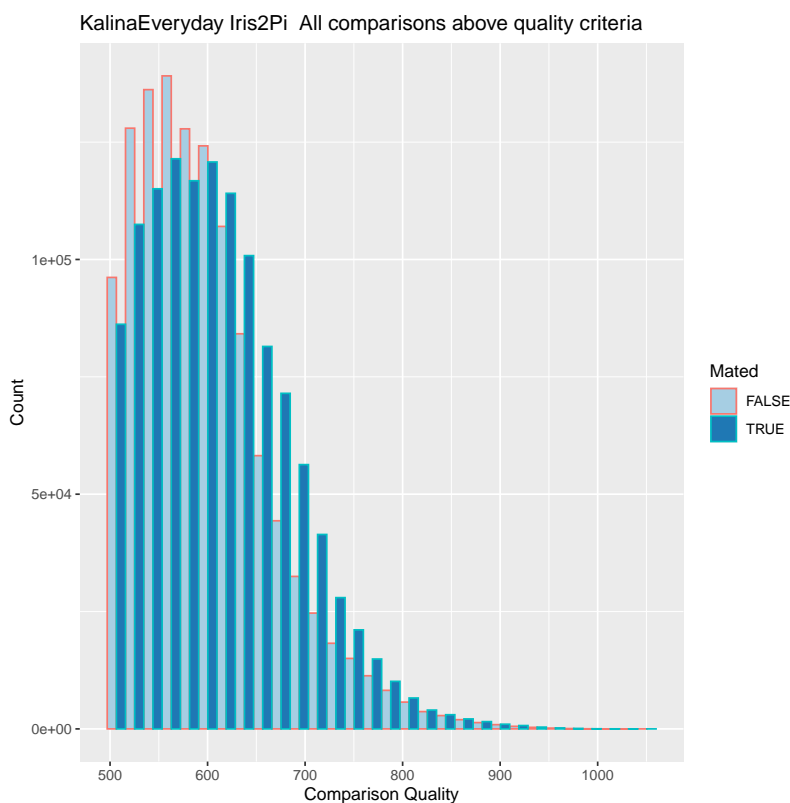


Fig. 5. Comparison quality for all mated and un-mated image pairs satisfying the interpupillary criterion discussed in the text. In this study we only had comparison quality from the iris2pi algorithm – a count of the number of bits compared. The cutoff at 500 was imposed, as discussed in the text. Note the suppressed zero to highlight details of differences between the two distributions.

Figure 6 presents the comparison score distributions for both algorithms for mated and non-mated pairs. Though there is substantial overlap between the mated and non-mated distributions, there are a significant number of mated comparisons beyond the tail of the non-mated comparisons. This figure presents all possible comparisons that met the image quality and comparison quality criteria.

Figure 7 presents violin plots[14] of the mated and non-mated scores by the time delta in years between the images. This figure presents all possible comparisons that met the image quality and comparison quality criteria. The mated comparisons extend to much better match scores than the non-mated. Shorter time deltas appear to correlate with better mated comparisons; the underlying reasons remain to be determined.

The temporal effect may be the result of changes in the ambient environment that can be seen in the online video: [EveryDay](#) and that we have not yet investigated.

The ROC plots in figure 9 are also based on matches from the single best left/right reference images. The ROC plots show that, for this set of iris images, there is a significant difference in performance between the two algorithms. This is not surprising because these algorithms were not optimized for the images in this dataset. This offers the hope that optimization for this application might yield improved results.

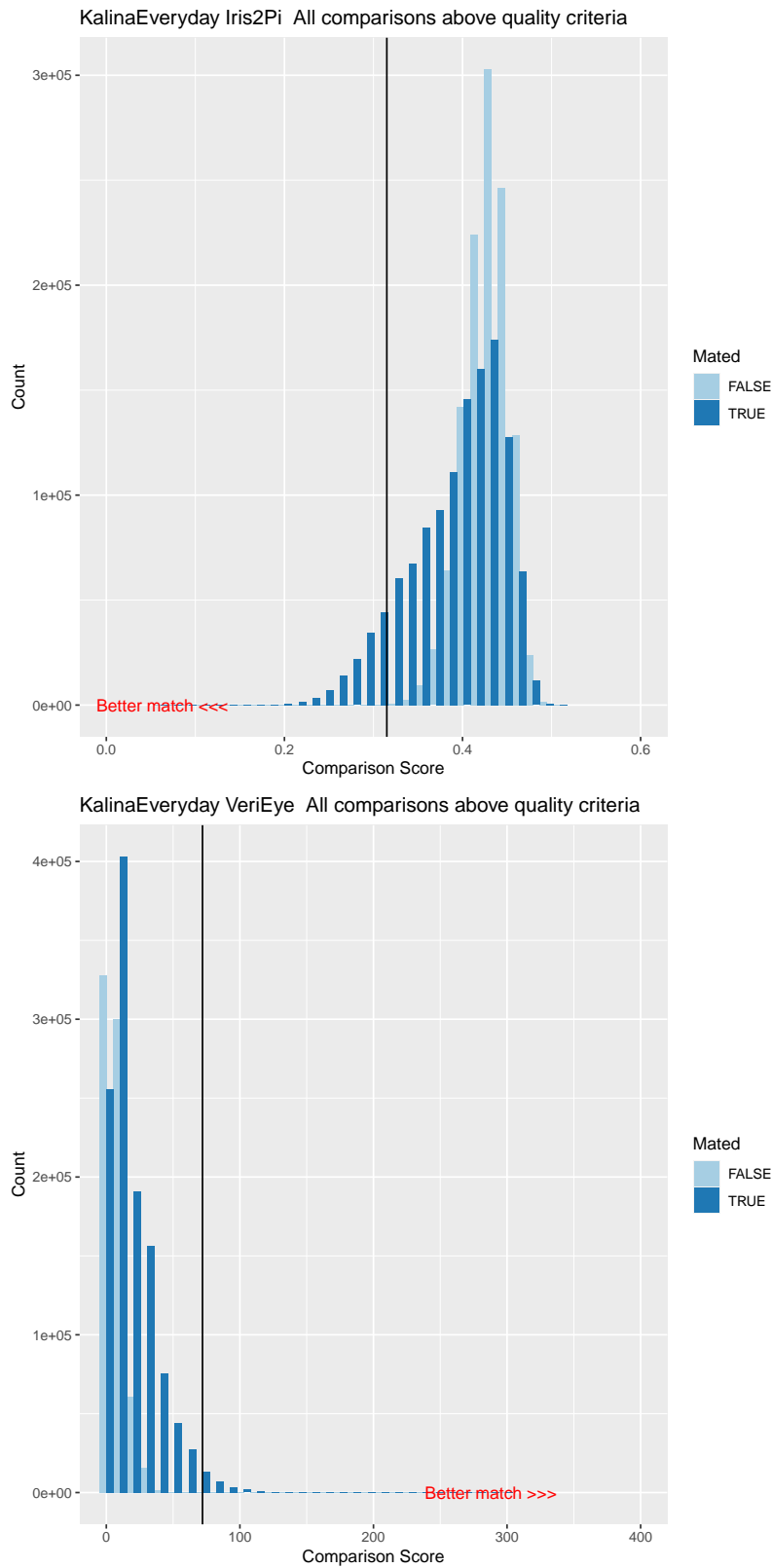


Fig. 6. Comparison score histograms for mated and non-mated image pairs. The comparisons are all against the single best left/right images as discussed in the text. The vertical line indicates the comparison score at which one might expect an FMR of 10^{-6} for NIR images from a purpose built iris camera that corresponds to a false match rate of approximately 1:1M on the basis of previously published results[8].

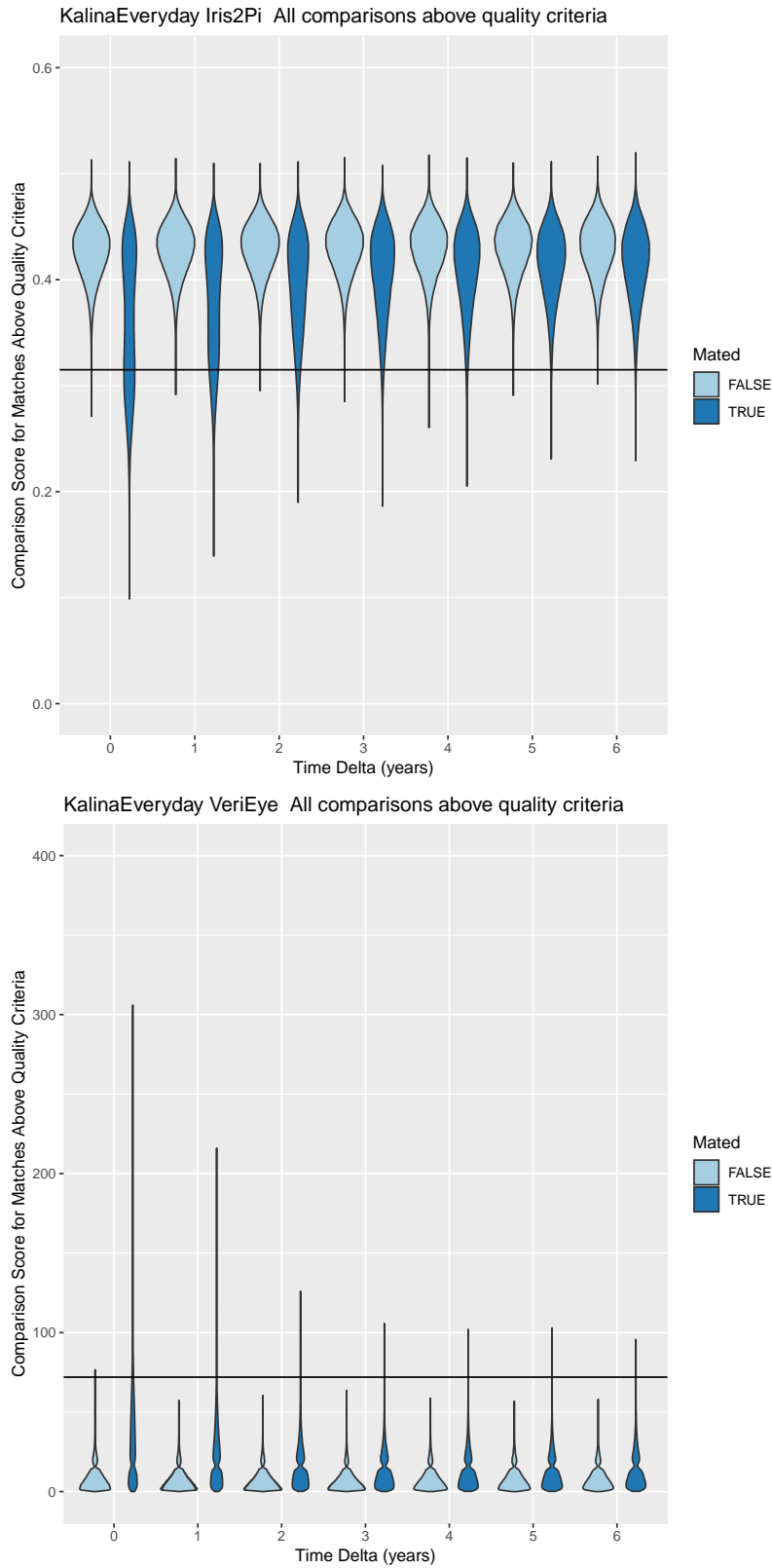


Fig. 7. Violin plots[14] of comparison scores as a function of time delta for mated and non-unmated image pairs. The horizontal line indicates the match score at which one might expect an FMR of 10^{-6} for NIR images from a purpose built iris camera on the basis of previously published results[8].

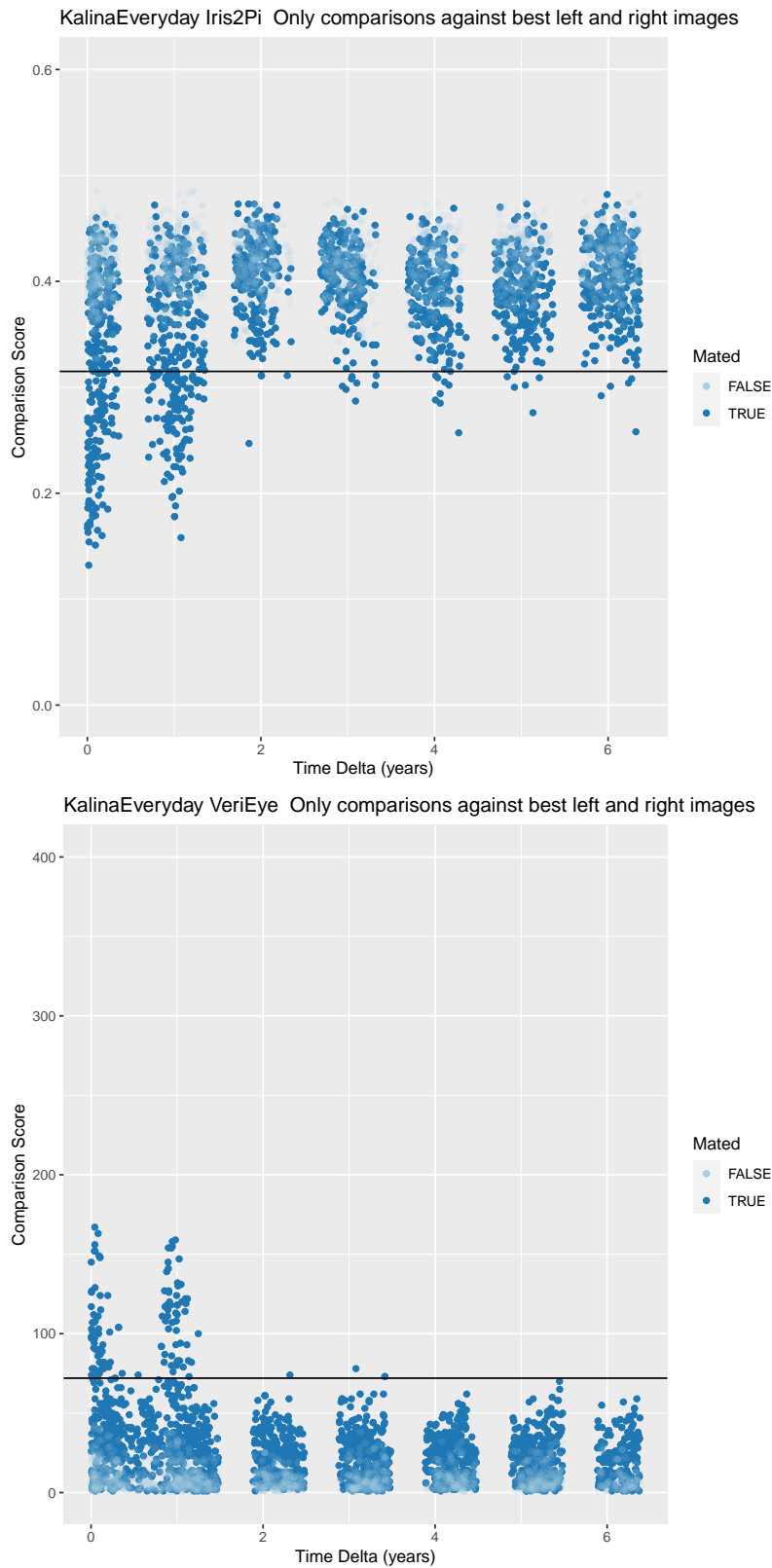


Fig. 8. Match scores as a function of elapsed time for mated and unmated image pairs. There is a single reference image for the left and for the right eye as described in the text. The horizontal line indicates the comparison score at which one might expect an FMR of 10^{-6} for NIR images from a purpose built iris camera on the basis of previously published results[8].

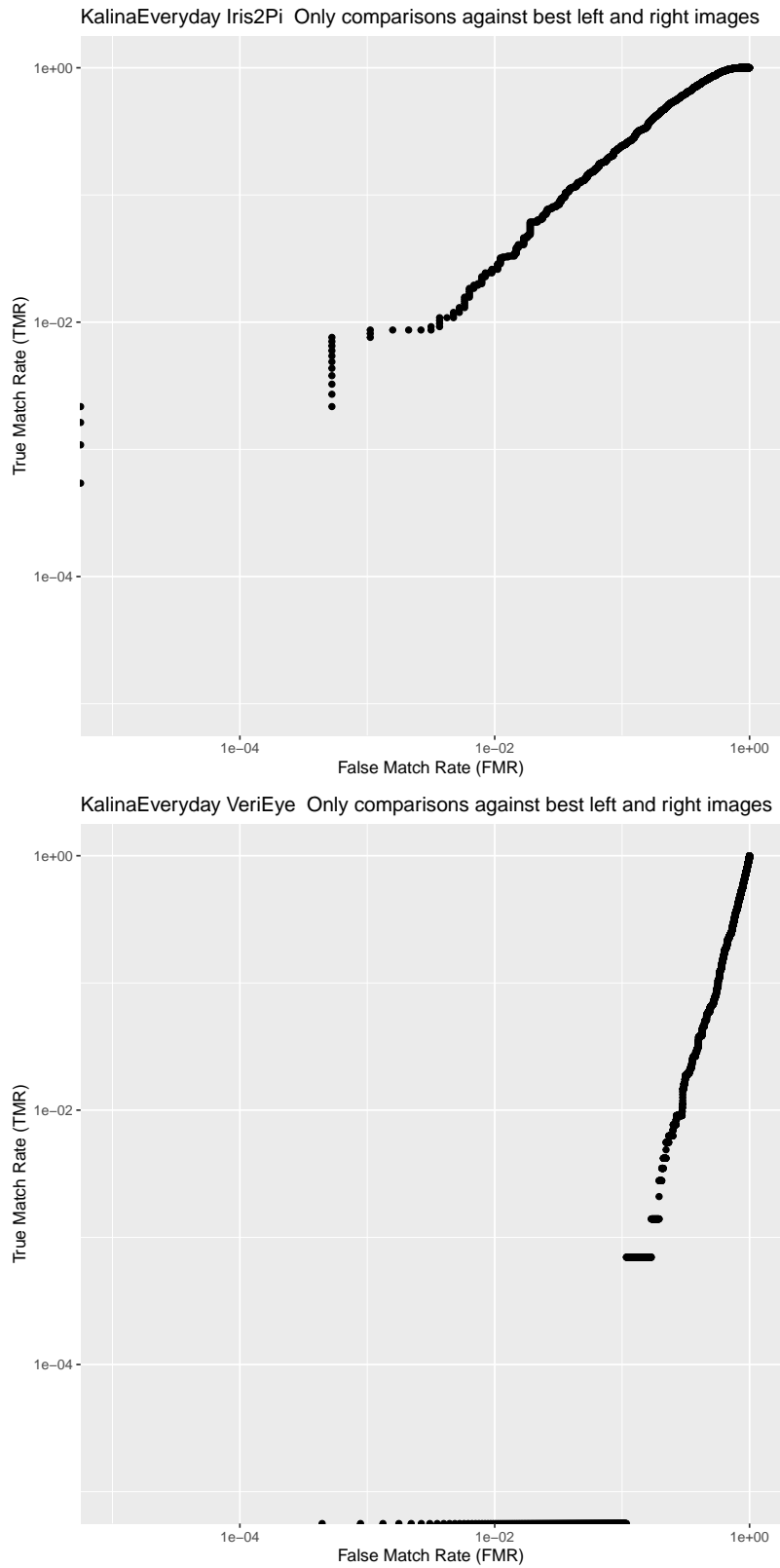


Fig. 9. ROC curves. There is a single reference image for the left and for the right eye as described in the text.

4. Conclusions and Future Plans

This analysis was conducted on visible light images captured with a commercial DSLR camera rather than an iris camera. For both mated and non-mated cases, the images and matches were subjected to prior quality metric filters described in the text.

Although the images were collected outside the infrared band normally used for iris recognition and under conditions that were not designed for iris recognition, figure 8 demonstrates that over a period of at least 6 years⁹, there are a significant number of instances where mated iris image pairs can be distinguished from non-mated pairs. This is reflected in the incomplete overlap between the mated and non-mated distributions in figure 6.

The ROC plot for iris2pi, figure 9, contains a region with comparison scores corresponding to false match rates of 0.1% at a true accept rate of 1%. Based on those ROC plots, the process laid out here is clearly inadequate for tasks such as access control or time & attendance. However, for comparison of high value image pairs, the process may still deliver valuable information. An example of a high value pair might be the case of a missing person, where the only imagery available has properties similar to the Kalina images.

Our future plans include:

- Expand the dataset by engaging with Mr. Kalina to obtain the latest images from his project.
- Evaluate the efficacy of other algorithms.
- Accept suggestions from the community for additional studies related to this dataset
- Conduct a more detailed analysis of iris2pi comparison quality and options for incorporating it into decisions involving visible light images from conventional, rather than iris, cameras.

Acknowledgments

Our thanks to Noah Kalina for making his images available to us, and to Patrick Grother, Greg Fiumara, George W. Quinn, Bethany Retton and Craig Watson for helpful comments on the manuscript.

⁹The time span of this dataset.

References

- [1] ISO 29794-6:2015 (2015) *Information technology – Biometric sample quality – Part 6: Iris image data* (ISO, Geneva, Switzerland), . URL <https://www.iso.org/standard/54066.html>.
- [2] Matey JR (2021) The NIST IrisDaily Dataset: Description and Analysis. NIST, Technical report. To be published March, 2021 URL doi.org/10.6028/NIST.IR.????
- [3] Matey JR, Grother PJ, Quinn GW (2020) Analysis of Iris Images in Nicholas Nixon: The Brown Sisters; NIST-TN-2098. NIST, Technical report. URL nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.2098.pdf.
- [4] Quinn GW, Quinn GW, Grother P, Matey J (2019) IREX IX Part Two: Multispectral Iris Recognition. NIST, Technical report. URL www.nist.gov/itl/iad/image-group/irex-ix.
- [5] Matey JR, et al. (2006) Iris on the Move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE* 94(11):1936–1947.
- [6] Daugman J (2001) Iris recognition. *American scientist* 89(4):326–333.
- [7] Daugman J (2004) How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on* 14(1):21–30. <https://doi.org/10.1109/TCSVT.2003.818350>
- [8] Daugman J (2006) Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE* 94(11):1927–1935. <https://doi.org/10.1109/JPROC.2006.884092>
- [9] Daugman J (2016) Evolving methods in iris recognition, . URL http://www.cse.nd.edu/BTAS_07/John_Daugman_BTAS.pdf.
- [10] Masek L (2003) *Recognition of human iris patterns for biometric identification*. Ph.D. thesis. Master’s Thesis, Univesity of Western Australia, .
- [11] R Core Team (2019) *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing Vienna, Austria, . URL <https://www.R-project.org/>.
- [12] Hollingsworth KP, Bowyer KW, Flynn PJ (2008) The best bits in an iris code. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(6):964–973.
- [13] Daugman J (1994) Biometric personal identification system based on iris analysis, . US Patent 5,291,560 URL <https://www.google.com/patents/US5291560>.
- [14] Hintze JL, Nelson RD (1998) Violin plots: a box plot-density trace synergism. *The American Statistician* 52(2):181–184.