IREX IV: Part 2

Compression Profiles for Iris Image Compression

NIST Interagency Report 7978

G. W. Quinn, P. Grother, M. Ngan

Information Access Division

National Institute of Standards and Technology

January 29, 2014

Acknowledgements

The authors would like to thank the sponsors of this activity. These are the Federal Bureau of Investigation, the U.S. VISIT office in the Department of Homeland Security (DHS), and the Science and Technology Directorate, also in DHS. Vitally, the authors would also like to thank the United States Department of Defense's Biometrics Identity Management Agency for their support and active collaboration.

Disclaimer

Specific hardware and software products identified in this report were used in order to perform the evaluations described in this document. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

Executive Summary

Background: With the publication of ISO/IEC 19794-6:2011 and ANSI/NIST ITL 1-2011 as stable and tested iris image interchange standards, iris can be exploited as a powerful interoperable biometric in a range of one-to-one and one-to-many roles. Some applications require compression of the iris data because resource constraints exist. These are the limited size and communication speed of ISO/IEC 7816 based smartcard credentials, and the limited bandwidth that is sometimes available for transmission of biometric data to backend servers. While one solution to this has been transmission of relatively small biometric templates, this is problematic for iris recognition because there are no *standardized* templates for iris data. The current solution is the transmission of compressed standardized iris image data. Compression can be lossless, but sometimes only lossy compression can fulfil file size requirements. Iris recognition is advantaged by standardized formats that assist compression, and by the fact that the iris texture can sustain considerable compression damage and still remain viable for recognition. The published image interchange standards do not yet contain definitive detailed guidance on compression. This report addresses this need.

Approach: We seek a formal compression profile for the application of the ISO/IEC 15444-1 JPEG 2000 compression algorithm to iris image data. This establishes settings for the various JPEG 2000 parameters by empirically quantifying their effect on iris recognition accuracy. This is done for applications in which either or both of the enrollment and search samples are compressed. Accuracy is measured using state-of-the-art commercial algorithms applied to over 3 million iris images.

Results: Most iris cameras emit 8-bit grayscale images with pixel dimensions 640x480. These are standardized as Image Kind 2[1] in ISO/IEC 19794-6:2011. Their size is 307 kilobytes (KB), but they can be compressed to 150KB without any loss of pixel information, and to as little as 16KB with only small losses in accuracy. When the iris is centered and the periphery cropped and masked, as required for Image Kind 7 of the ISO standard, the resulting image can be compressed to as little as 2KB with only small losses in accuracy. Such sizes support inclusion of iris data on secure identity credentials, and fast network-based recognition. Efforts to reduce sizes substantially below 2KB produce elevated error rates that would not be tolerable for many applications.

For electronic passports, the International Civil Aviation Organization (ICAO) 9303 specification should be revised to note availability of iris images that are more than 10 times smaller than those conceived of in the first editions of that profile. Particularly, the Data Group 4 container in the Logical Data Structure (LDS) could now be populated with ISO/IEC 19794-6:2011 iris images of size 3KB or smaller, rather than the 30KB currently indicated. Facial images stored on e-Passports typically have sizes of 10-20KB, and standardized fingerprint templates have sizes around 0.5KB. Digital signatures associated with any of these elements can readily have sizes around 0.5KB.

Technical Summary

Further technical results are listed below. Each item roughly corresponds to a section or subsection from the main body of the report.

- *JPEG 2000 Compresson Profile:* The IREX I evaluation identified JPEG 2000 as more effective at compressing iris images than alternatives such as traditional JPEG. This study extends that research by recommending that iris images be compressed with
    - a single tile,
    - a block size of 64-by-64,
    - a base quantization step size of $1/256$, and
    - 3 decomposition levels,

  when compressing with the irreversible (CDF 9/7) wavelet transform. These parameter values ensure minimal loss in recognition accuracy. The values do not change depending on the file size.

  The file-size specification follows.

- *Compression Limits for Kind 2 Formats:* Kind 2 records are produced by most iris cameras and do not undergo special processing to assist with compression. These images can be compressed to 16KB with only small losses in accuracy.

[1] ISO/IEC 19794-6:2011 actually refers to these as "Type 2 images" rather than "Kind 2 images", but the terminology is changed in this report to avoid confusion with Type 2 records in ANSI/NIST ITL 1-2011.

At a fixed decision threshold, the False Negative Identification Rate (FNIR) increases by no more than a factor of $1/4$. When compressed to 8KB, FNIR doubles at a fixed decision threshold while the behavior of the False Positive Identification Rate (FPIR) is matching algorithm dependent. Of the four implementations, it does not change for two, increases by about a factor of 5 for one, and increases by about a factor of 20 for the last. Since iris Detection Error Trade-off (DET) curves tend to be low sloping, the increase in FPIR can often be corrected without a significant impact on FNIR by adjusting the decision threshold. When images are compressed to 6KB, the FNIR increases by a factor of 3 to 4 (at fixed FPIR), depending on the matching algorithm. Compressing Kind 2 records to sizes smaller than 16KB is not recommended.

- *Compression Limits for Kind 7 Formats:* Kind 7 records can be compressed to as little as 2KB with only small losses in accuracy (FNIR at fixed FPIR increases by no more than a factor of $1/4$). Note this is an order of magnitude lower than ICAO's recommendation of 30KB as the optimal compression size. Error rates increase more appreciably when images are compressed to sizes smaller than 2KB. When compressed to 1KB, FNIR increases by a factor of 2 to 3 at a fixed threshold, and FPIR does not change for 2 implementations, increases by a factor of 5 for one implementation, and by a factor of about 20 for another. When compressed to 768 bytes, FNIR increases by a factor of 3 to 5 (at fixed FPIR), depending on the implementation used to generate the Kind 7 records. The sclera must be masked and the eyelid boundaries blurred to keep error rates low. Compressing Kind 7 records to sizes smaller than 2KB is not recommended.

  In comparison to other biometric modalities, standardized fingerprint minutiae information can be stored in 500 bytes, and face images can be compressed to about 8KB, although ISO/IEC 19794-6 recommends 30KB, and ICAO recommends 10-20KB for e-passports.

- *Relevance to 1:1 Verificaton:* This study was conducted using 1:N algorithms where the enrolled population was typically 160 000. Given our focus on FPIR values around 10^{-3}, this study has approximate correspondence to 1:1 false match rates below 10^{-8}, a security level more stringent than would be used for most high security applications.

- *Two-Eye Matching:* When a fixed amount of storage space is available, sometimes greater accuracy can be achieved by storing images of both eyes rather than a more lightly compressed version of just one. This study found that a cross-over occurs for Kind 7 records, where one-eye matching is more accurate than two-eye matching at lower storage capacities, but less accurate at larger storage capacities. At sizes of 1.5KB or less, one-eye matching is consistently more accurate. At 4KB or more, two-eye matching is always more accurate. Note that matching with two eyes introduces an additional computation penalty (to both template generation and searching) that may offset possible accuracy benefits.

- *Resolution Downsampling:* Downsampling selectively discards the highest frequency information in the image. If this information is relatively unimportant to the matching algorithm, then ensuring that it is discarded during compression will ensure that a maximum of the encoding budget is dedicated to representing the more important features. Two-by-two pixel averaging as well as selective retention of only the lower-frequency resolution levels of an image's JPEG 2000 representation were tested. However, neither method consistently yielded better performance over the case when images are not downsampled. Care should be taken when downsampling images since excessive tuning of the compession process runs the risk of compromising interoperability.

- *Lossless Compression:* Lossless compression retains all of the information in an image so that it can be perfectly reconstructed. As such, it cannot achieve compression ratios as high as lossy compression. JPEG 2000 typically manages to compress Kind 7 records losslessly to sizes between 10KB and 40KB, with a mean size of 20KB. Kind 2 records typically compress to sizes between 100KB and 170KB, with a mean size of 135KB. Lossy compression should only be applied to images if there is a documented need for small image sizes.

  The ISO/IEC 19794-5 and ANSI/NIST-ITL 2011-1 standards also allow iris images to be stored as lossless PNGs. The IREX I evaluation compressed iris images with PNG and found that Kind 7 records compress to a median file size of 25KB, and Kind 2 images to a median size of 150KB, which are comparable to the sizes lossless JPEG 2000 compression achieves in this evaluation.

Contents

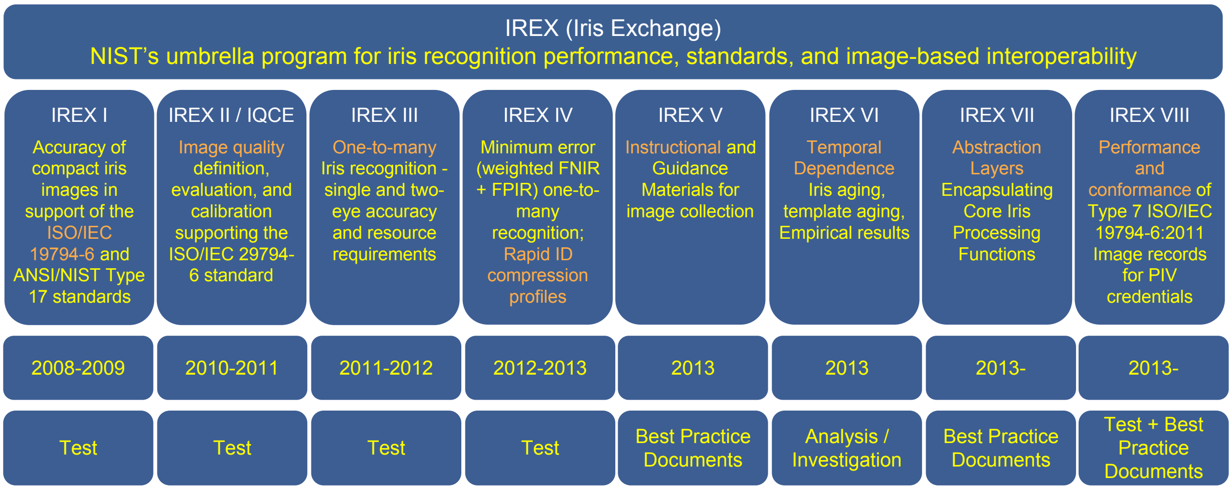


Figure 1: Timeline of the IREX program, including planned future installments.

1 The IREX Program

The Iris Exchange (IREX) Program was initiated by the National Institute of Standards and Technology (NIST) to support an expanded marketplace of iris-based applications. IREX provides quantitative support for iris recognition standardization, development, and deployment. To date, 5 activities have been completed and 3 more are tentatively planned (see Figure 1). Each is summarized below.

- **IREX I** [1] was a large-scale, independently administered, evaluation of one-to-many iris recognition. It was conducted in cooperation with the iris recognition industry to develop and test standard formats for storing iris images. Standard formats are important for maintaining interoperability and preventing vendor lock-in. The evaluation was conducted in support of the ISO/IEC 19794-6 and ANSI/NIST-ITL 1-2011 standards.
- **IREX II** [2] supported industry by establishing a standard set of quality metrics for iris samples. Although iris recognition has the potential to be extremely accurate, it is highly dependent on the quality of the samples. The evaluation tested the efficacy of 14 automated quality assessment algorithms in support of the ISO/IEC 29794-6 standard [3].
- **IREX III** [4] was a performance test of the latest iris recognition algorithms over operational data. Despite growing interest in iris-based technology, at the time there was a paucity of experimental data to support published theoretical considerations and accuracy claims. IREX III constituted the first public presentation of large-scale performance results using operational data.
- **IREX IV** builds upon IREX III as a performance test of one-to-many iris recognition. In addition to providing participants from previous evaluations an opportunity to further develop and test their recognition algorithms, this evaluation explores the potential for using a cost equation model for optimizing algorithms for specific applications.
- **IREX V** will provide best practice recommendations and guidelines for the proper collection and handling of iris images.
- **IREX VI** [5] explores a possible aging effect for iris recognition. The intrinsic features of the iris may naturally change over time in a way that affects recognition accuracy. Factors such as subject habituation and aging of the camera may also introduce a time dependency.
- **IREX VII** intends to define a framework for communication and interaction between components in an iris recognition system. By introducing layers of abstraction that isolate underlying vendor-specific implementation details, a system can become more flexible, extensible, and modifiable.
- **IREX VIII** will test the performance of ISO/IEC 19794-6:2011 Type 7 images and lay the groundwork for conformance testing of Type 7 record generators.

The latest information on the IREX Program can be found on the IREX website [6].

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

2 Introduction

2.1 Purpose

The IREX I evaluation determined JPEG 2000 is the best format for lossy compression of iris images in terms of minimizing the loss in recognition accuracy. However, JPEG 2000 contains many customizable parameters that were not explored in the evaluation. These parameters affect not only the pixel representation of an image, but also computation time, memory usage, and the ability ot perform computations in parallel. This study extends IREX I by identifying an optimal combination of parameter values for compressing standard iris images with JPEG 2000.

Secondarily, this report tests the ability of automated iris recognition algorithms to match highly compressed iris images. Unlike IREX I, the current study deals with the more difficult problem of matching in a one-to-many, rather than one-to-one, mode.

2.2 Market Drivers

Many biometric systems operate with restrictions on the size of their biometric samples. A system that reads samples from a smartcard is limited by the storage capacity of its digital chip. For compaison, the NIST Special Publication 800-73-4 [7] limits the container size on government Personal Identity Verification smartcards to no less than 12 710 bytes for a face image, and no less than 4 006 bytes for fingerprint minutiae information. FIPS 201-2 [8] recommends no less than 3 000 bytes for an iris image. The Registered Traveler Pilot Program [9], which operated from 2005 to 2009, allocated 4 000 bytes per iris image. The images were stored in a variation of the polar format later rejected for inclusion in ISO/IEC 19794-6. ICAO Doc 9303 [10] conservatively recommends 30 720 bytes for optimal storage of an iris image on e-passports even though subsequent studies have found that iris images can be stored at much smaller sizes without detrimentally affecting recognition accuracy. Even without a fixed upper limit on the container size, smaller samples transfer more quickly from the card to the reader. This can influence how many samples are selected for transfer (e.g. one iris or both), or which modality (e.g. face or fingerprint) is used for recognition.

(a) Reading samples from an Identity Credential **(b)** Transfering images across a limited-bandwidth network

Figure 2: Scenarios where the size of biometric samples impacts performance. Figure 2a depicts a system that reads biometric samples from a limited storage capacity smartcard. Figure 2b depicts a system that transfers samples across a bandwidth-limited network.

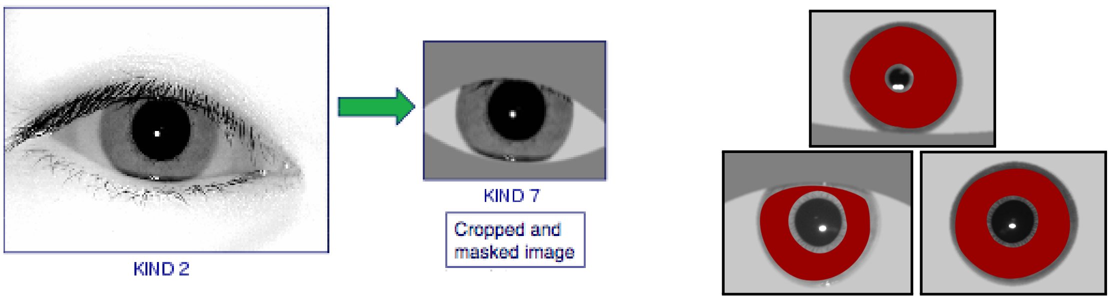
The size of biometric samples also affects the performance of systems that must transfer samples across bandwidth-limited networks (see Figure 2b). A prominent example is India's Unique Identity (UID) scheme. Private banks linked to the scheme will soon deploy thousands of "micro ATMs" across rural parts of the country to provide citizens with better access to their accounts [11]. These Micro ATMs will verify users' identities by capturing biometric samples locally and transferring them to a central facility for matching. The samples must be transferred using India's existing telecommunications infrastructure, but since coverage is weak in some parts of the country, the rate at which data can be transferred is sometimes severely limited. The iris would be viable only if the samples could be compressed to a few kilobytes.

The Department of Defense uses the iris for rapid identification in the field, and smaller samples facilitate faster response times. Some applications might benefit from transferring a highly compressed version of the iris sample for quick identification, followed by a better quality (non-compressed) version of the sample for retention as the authoritative sample.

2.3 Standard Iris Formats

Standard iris images are not iris templates. Rather, they are interoperable images designed for efficient storage and transmission. Templates are proprietary "black box" feature representations specific to a single provider's recognition algorithm. As such, their content is non-standard, non-interoperable, and not suitable for cross-agency or cross-vendor exchange of iris data. Although proprietary templates are sometimes smaller than raw iris images, discarding the original images locks the system into using a particular version of a provider's software. Not only does this undermine interoperability, but it prevents the system from exploiting future improvements in the feature extraction and matching procedures.

The ISO/IEC 19794-6:2011 [12] and ANSI/NIST-ITL 1-2011 [13] standards define Kind 2 and Kind 7 record formats for storing iris images. Kind 2 image records are usually output directly by iris cameras and do not undergo processing to facilitate compression. In contrast, Kind 7 image records provide a much more compact representation of the iris but require further processing to generate. In addition to cropping out much of the periocular region around the iris, the sclera and eyelids must each be masked with a solid color. Uniform regions of solid color require very little space to encode, thus ensuring that a maximum of the encoding budget is dedicated to representing the actual iris features. The first amendment to ISO/IEC 19794-6:2011 defines 4 cases for masking eyelids. The first is depicted in Figure 3a, and 3 alternatives are depicted in Figure 3b. When the upper or lower eyelids do not intersect the iris boundary, they do not need to be localized and masked with a color distinct from that of the sclera. This study focuses its attention solely on the Kind 2 and Kind 7 formats.



(a) **(b)**

Figure 3: Examples of (a) Kind 2 and Kind 7 image records defined in ISO/IEC 19794-6, and (b) alternative eyelid masking procedures supported by the standard. (Some iris textures are masked with red to prevent identification of the individuals).

3 Methodology

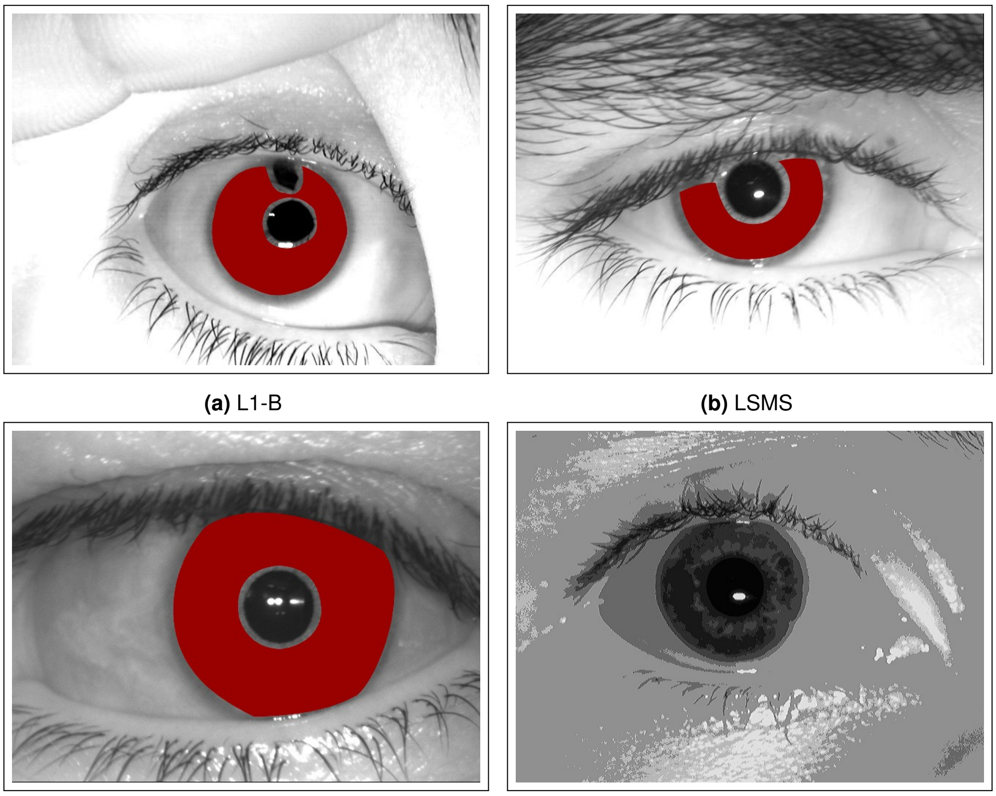
This section describes the test procedures, software, and performance metrics used in this study. Much of this information is covered in greater depth in the *IREX IV Concept of Operations (CONOPS)* [14] document. The reader is referred to this document for further details on the evaluation process.

A technology evaluation [15] such as this focuses on algorithm performance over factors that may be relevant to the deployment and operation of a biometric system (e.g. policy drivers, societal and economic considerations, availability of legacy data). Performance is assessed using metrics that give a general idea of the technology's capabilities. The relative importance of these metrics will depend on how the technology will be used.

3.1 Test Environment

The evaluation was conducted offline at a NIST facility. Testing was performed on high-end PC-class blades running the LINUX operating system, which is typical of central server applications. Most of the blades had 6 quad-core AMD Opteron processors running at 2.4 GHz with 192 GB of main memory. The test harness used concurrent processing to distribute workload across dozens of blades.

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate



(a) L1-B **(b)** LSMS

(c) COGT **(d)** Posterized

Figure 4: Examples of iris images from the test dataset. The subfigure captions for (a) through (c) refer to a four-letter software code that is an indicator of the type of camera used. The subfigure in (d) is an example of a highly posterized image.

3.2 Iris Dataset

The evaluation uses images from the Operational Set (OPS) II, which consists of approximately 7.5 million field-collected images from several commercial capture systems, predominantly the Securimetrics HIIDE and PIER, and the Crossmatch I SCAN and SEEK. The images occasionally suffer from poor sample quality (e.g. high amounts of occlusion, specular reflections) that are typical of an operational system. Many were captured outside and contain heavily constricted pupils. Figure 4 shows some examples from the set. The iris in 4a an anatomical defect that occurs rarely in the dataset (see the *IREX III Supplemental Report* [16]). All images have a pixel resolution of 640x480. The pathological 330x330 images discussed in IREX III and its supplement are excluded from this evaluation. Some of the subjects' irides were captured by more than one camera model on different days. Further details on the images can be found in the IREX III supplement.

Some images in the OPS-II dataset suffer from what appears to be a *posterization* effect, an artifact of color quantization that can lead to significant reductions in the amount of useful feature information in the images. Figure 4d shows an example of a highly posterized iris image. The IREX III supplement identifies posterization as one of the more common causes of failed identifications for the current dataset. The dithered texture introduced by posterization also introduces a lot of high frequency information that makes it more difficult to efficiently compress these images to small file sizes.

3.3 Matching Algorithms

Twelve commercial organizations and academic institutions submitted 66 iris recognition software libraries for evaluation. The participation window opened on May 16th, 2012 and closed on August 2nd, 2012. Participation was open worldwide to anyone with the ability to implement a large-scale one-to-many iris identification algorithm. There was no charge to participate.

Support for Kind 7 record generation was optional but encouraged. Five of the 12 participants provided software libraries

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

Participant	Letter Code	Class P Submissions	Kind 7 Support?
University of Bath	A	A00P, A01P, A02P	
Neurotechnology	B	B00,P B01P, B02P	✓
Smart Sensors	C	C00P, C01P, C02P	
3M Cogent	D	D00P, D01P, D02P	
IriTech	E	E00P, E01P, E02P	
MorphoTrust	F	F00P, F01P, F02P	✓
iSciLab	G	G00P, G01P, G02P	
Delta ID	H	H00P, H01P	
University of Cambridge	I	I00P, I01P, I02P	✓
Iris ID	J	J00P, J01P	
Morpho	K	K00P, K01P, K02P	✓
Nihon Systems	L	L00P	

Table 1: Participants of IREX IV along with their NIST-assigned letter codes, algorithm identifiers, and whether the submissions support Kind 7 record generation.

capable of generating Kind 7 records from raw Kind 2 images, although one provider's implementations were untestable because they produced runtime errors. All submissions were required to support the proper handling of Kind 7 records even if they could not generate the records themselves. However, this report focuses predominantly on the most recent algorithm submissions from the 4 participants who support Kind 7 record generation.

Table 1 lists the IREX IV participants along with the alpha-numeric codes assigned to their algorithms. Participants were allowed to submit up to 3 Class P algorithms. Briefly, Class P means the algorithms are intended for use in positive (as opposed to negative) identification systems. Positive identification systems verify the claim that the user is enrolled and typically grant special privileges or access to enrolled users. Four of the participants (University of Bath, iSciLab, Delta ID and Nihon Systems) are new to the IREX program while the other 8 have participated in previous IREX evaluations. For each participant, the algorithms are labeled by chronological order of submission.

Four participants (lettered B, F, I, and K) provided implementations that support for the generation of Kind 7 records. However, the implementations from participants and B and K are only partially conformant to the standard since they do not always mask the sclera or blur the sclera-eyelid boundaries.

3.4 Compression Algorithm

This evaluation uses version 7.0 of Kakadu Software's JPEG 2000 developer toolkit [17]. The software is proprietary and fully compliant with Part 1 of the JPEG 2000 standard [18]. Open-source alternatives to Kakadu include OpenJPEG [19], and JasPer [20]. Only part 1 of the JPEG 2000 standard is used to compress images in this evaluation. The more flexible second part of the standard is not widely supported as of this writing and is not tested.

The quality and fidelity of the compressed images are the most important performance characteristics, although compression time is also sometimes measured and reported.

3.5 Performance Metrics

3.5.1 Operational Scope

This evaluation measures iris recognition performance for *open-set* applications, meaning individuals are searched against a database of previously enrolled persons, but without any guarantee that searched individuals are enrolled. Most real-world applications of biometrics operate in this way. For example, a system that grants building access cannot assume that every user who attempts access has provided the system with an enrollment sample on a previous occasion. *Closed-set* applications, which assume every searched individual is enrolled (and thus only concern themselves with identifying

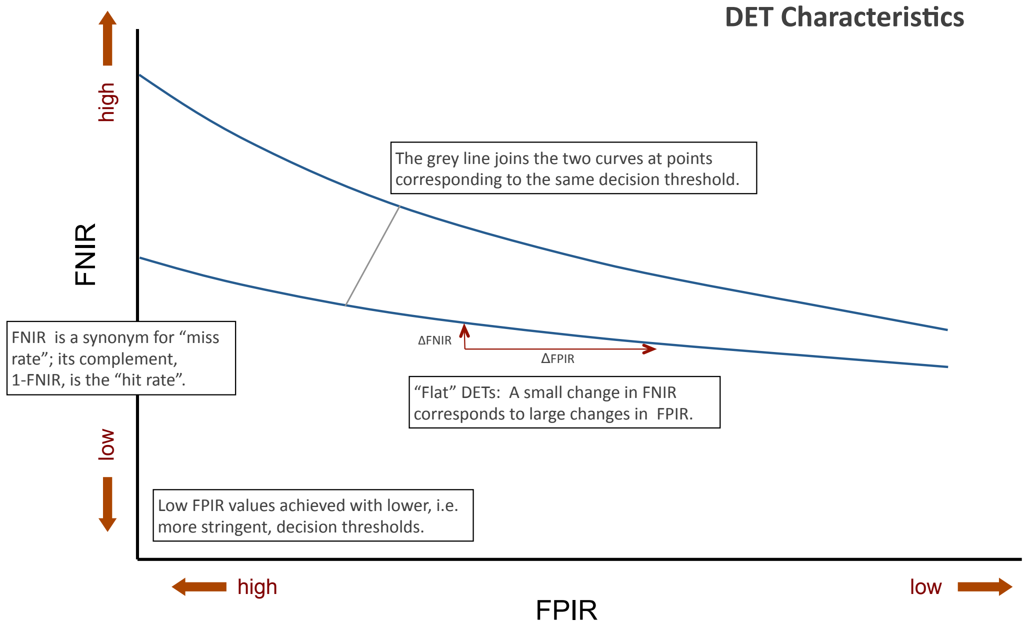


Figure 5: Notional DET curves. Each point along the curve corresponds to a particular decision threshold. Adjusting the decision threshold decreases the rate of one error type, but at the expense of the other.

the searched person from among the enrolled population) are operationally uncommon for iris and are not tested in this evaluation.

Consideration is further restricted to *positive* identification systems only, which verify the (often implicit) claim that the user is enrolled in the database. Such systems typically grant special privileges or access to enrolled users. For example, the NEXUS Program [21] uses iris recognition to positively identify registered travelers for expedited security screening at ports of entry in the US and Canada. Negative identification systems, which were not specifically tested in this evaluation, verify the claim that the person is *not* enrolled and often impose restrictions on enrolled individuals. An example is the United Arab Emirates (UAE)'s border-crossing control system, which uses iris recognition to prevent expelled individuals from re-entering the country.

3.5.2 Matching Accuracy

Matching accuracy is measured for open-set biometric systems, which are tasked with searching an individual against an enrollment database and returning zero or more candidates. A candidate is returned if the implementation determines that dissimilarity to the searched image is below a pre-determined decision threshold. A false positive occurs when a search returns a candidate for an individual that *is not* enrolled in the database. A false negative occurs when a search *does not* return the correct candidate for an individual that *is* enrolled in the database. Raising the decision threshold increases the rate of false positives but decreases the rate of false negatives.

Core matching accuracy is presented in the form of Detection Error Tradeoff (DET) plots [22], which show the trade-off between the False Positive Identification Rate (FPIR) and the False Negative Identification Rate (FNIR) as the decision threshold is adjusted. Figure 5 shows a notional DET plot. Low security applications (e.g. theme park access) might operate at high decision thresholds, toward the right end of the figure. High security applications (e.g. access to highly sensitive information) are more likely to operate at low decision thresholds, toward the left end of the figure. Iris recognition is known for having lightly sloping DETs compared to other biometric modalities.

The integrity of ground truth information is a matter of concern in any biometric evaluation. Identity mistakes are known to exist in OPS II. To negate their impact on the FPIR, we chose to horizontally flip search images prior to template generation when the searches were non-mated. Replacing a search image with its mirror image ensures that even if a mate is enrolled, the textures will still appear different (see IREX III Section 6.3 for a detailed explanation and analysis). Unfortunately, this does not solve the problem where two or more different people are assigned the same identifier. Although this type of error can inflate estimates of FNIR, the IREX III Supplemental Report found it to be a rare occurrence (of 17,017 mated searches,

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

only 28 failures were attributed to this type of ground truth error).

Due to the high frequency of erroneous (left/right) eye labelings in the OPS-II dataset, we chose to always enroll both eyes for an individual as separate entries and credit the algorithm with a match if either of the subject's eyes were matched. We suspect the mislabelings are due to ambiguity with respect to whether "left" is intended to refer to the subject's left eye (correct) or the eye on the left from the perspective of the camera operator (incorrect).

False positives are computed exclusively from non-mated searches (i.e. searches for which the searched individual is not enrolled in the database). This is more reflective of operation than if false positives had been computed from mated searches with the correct candidates removed from the list. Similarly, false negatives are computed exclusively from mated searches.

3.5.3 Computation Time

Timing statistics are presented for compression operations as the actual time elapsed according to the Bash shell's `time` command, which has a resolution of one millisecond on our platform. The command reports end-to-end runtime, which includes the time it takes to read an image from disk. To reduce the impact of I/O on timing statistics, the images are read from `/dev/sh` to ensure they are already cached in main memory. The alternative C function `clock()`, which measures the amount of processor time dedicated to the process, has insufficient resolution and would not report useful timing statistics for multithreaded runs.

Timing statistics were collected on an unloaded machine having the specifications described in Section 3.1 (a high-end PC-class blade with 6 quad-core AMD Operteron CPUs running at 2.4 GHz).

3.5.4 Uncertainty Estimation

Some figures and tables convey information about the uncertainty associated with a statistic in the form of confidence intervals or estimates of standard deviation. These estimates are intended to capture random variation in the observed value if one assumes repeated *iid* sampling from the same population. They are *not* intended to reflect how the statistic might change over different test data or even different sampling strategies over the same data.

Estimates of uncertainty are computed using the Wilson Score method [23] which overcomes certain problems associated with applying the Central Limit Theorem to a discretized estimator. We make several simplifying assumptions when applying the method to biometric identification. Most notably, separate searches against the same enrollment database are treated as independent samples, yet we know positive correlations exist due to Doddington's Zoo [24]. We also report estimates of the variability of FNIR at a fixed FPIR when in fact it is the decision threshold that is fixed. Uncertainty with respect to what decision threshold corresponds to the targeted FPIR results in increased uncertainty about the true value of FNIR. However, our estimates of FPIR are fairly tight due to the large number of non-mated searches performed, so they are not expected to have a large impact on the estimates.

4 Results

4.1 Toward a JPEG 2000 Compression Profile

JPEG 2000 includes a number of customizable parameters that affect the pixel representation of an image when it is compressed. The goal of this section is to identify the optimal combination of parameter values that minimize the loss in recognition accuracy. The effect that some of these parameters have on other performance metrics, such as computation time, is also investigated.

Lossy compression was always performed with the irreversible (9/7) wavelet transform. Section 4.4 explores lossless compression and uses the reversible (5/3) wavelet transform.

Overview of the Encoding Procedure

The basic steps of the compression process are depicted in Figure 6. The diagram is high-level and over-simplified in some places to only depict the steps most pertinent to the current study. After the image is divided into non-overlapping rectangular tiles, each tile is wavelet transformed. The coefficients are then quantized to reduce the number of bits required to represent them. The image data is then partitioned into code blocks that are passed directly to the entropy coder, which uses the Embedded Block Coding with Optimal Truncation (EBCOT) algorithm to perform the core optimization of JPEG 2000. The final result is a serialized bit-stream.

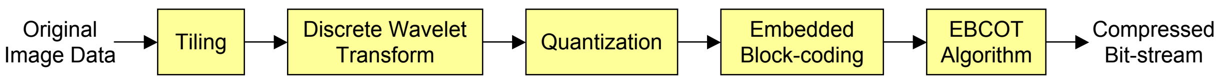


Figure 6: Basic steps of JPEG 2000 compression.

The bit-rate is the most important parameter because it is used to target a desired file size. The other parameters varied in this study are listed below.

- ***Tile Size:*** Before wavelet conversion, the image is separated into non-overlapping rectangular tiles that are each encoded separately. Setting the tile size to a small value can introduce blocking artifacts similar to traditional JPEG, since no attempt is made to smooth the borders between adjoining blocks. Partitioning in this manner allows operations to be performed on individual tiles without having to load the entire image into memory.
- ***Number of Decomposition Levels:*** After the wavelet transformation, tiles are broken into multiple decomposition levels such that higher levels describe finer details in the image. The process, known as Dyadic decomposition [25], is a "divide and conquer" strategy that has desirable mathematical properties.
- ***Quantization Step Size:*** The quantization step size determines the granularity of the wavelet coefficients. Quantization performs both rounding and truncation of the coefficients. Small step sizes correspond to finer granularity (i.e. greater precision and fidelity to the original image), but require more bits to represent. Quantization is performed prior to entropy coding.
- ***Block Size:*** Specifies the dimensions of the rectangular code blocks. After the wavelet transformation, the image is separated into frequency subbands. Each frequency subband is further separated into code blocks, and each block is independently coded as a bit stream. The entropy coder operates directly on these blocks, truncating each at a point that minimizes the overall squared error loss. The code block size is normally set to either 32-by-32 or 64-by-64.

4.1.1 Tile Size

Introduction

Before wavelet conversion, the image is divided into non-overlapping rectangular tiles. Tiling is intended to reduce memory usage when viewing or modifying high-resolution images since many operations only require some tiles to be loaded into memory. The downside to tiling is that it can introduce blocking artifacts along the borders between tiles (see Figure 7). Iris images have relatively small pixel dimensions and are not expected to benefit from using more than a single tile to cover the entire image. Generally speaking, tiling is intended for use with much higher resolution images, such as those produced by medical imaging devices [26]. Tile dimensions are usually powers of 2 and cannot vary within an image (with the exception of those running along the right and bottom image border that are sometimes truncated). Kakadu's default tile size is the smallest possible that encompasses the entire image.

Results and Recommendations

Figure 8 shows recognition accuracy as a function of file size when search images are compressed with different tile sizes. All other compression parameters are left at their default values (i.e. 3 decomposition levels, a quantization step size of $1/256$, and a block size of 64). Although the tile size has little effect at large file sizes, the benefit to using a single tile is apparent at sizes below 2 048. When compressing to a size of 1 024 bytes, a tile size of 128 increases FNIR by between 10 and 60 percent, depending on the matching algorithm. This report recommends that only a single tile be used to represent the image. For Kind 2 and Kind 7 images, a tile size of 1024x1024 pixels is sufficient. Setting the tile size to be greater than the dimensions of the image does not detrimentally affect compression.

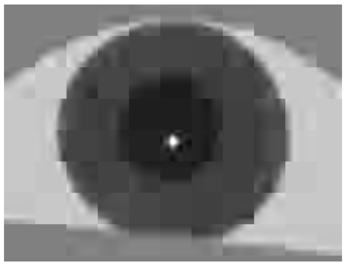


Figure 7: Demonstration of blocking artifacts in a highly compressed JP2 image.

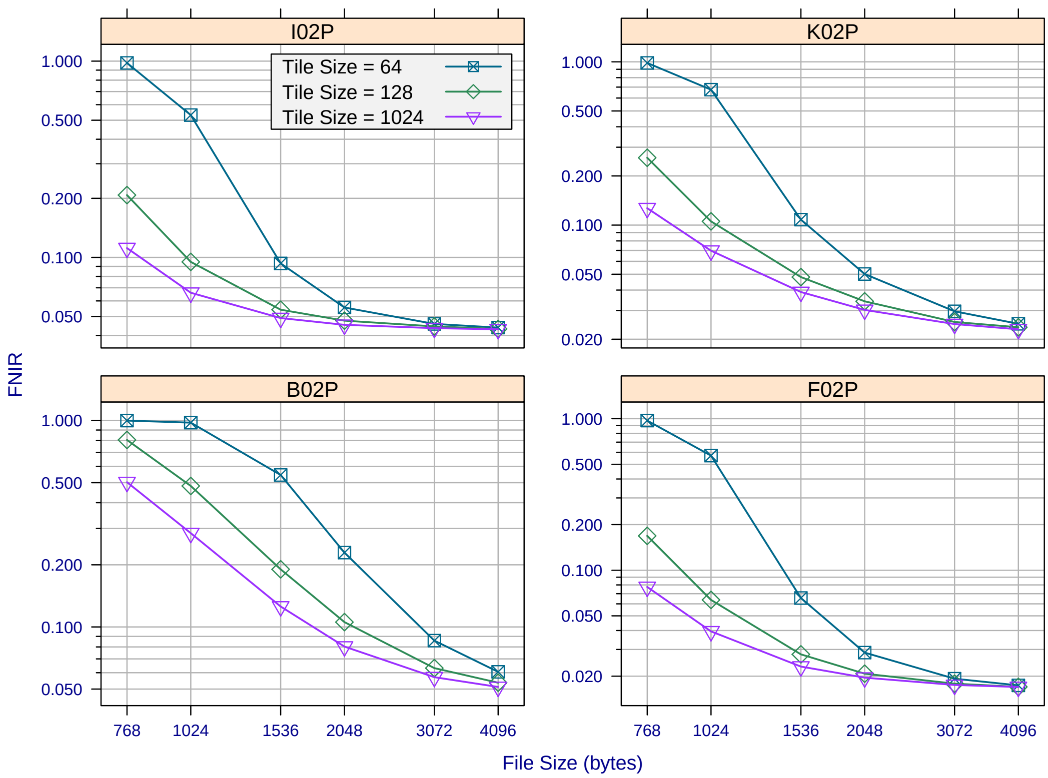


Figure 8: Comparison of FNIR (at FPIR=0.001) for 4 matching algorithm when search images are compressed with various tile sizes. Compressed Kind 7 records (created by the matching algorithms) are searched against an enrolled population of 160 000 non-compressed images. Each point is generated from 80 135 mated searches.

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

4.1.2 Quantization Step Size

Introduction

After the wavelet transform, the coefficients are quantized to reduce the number of bits required to represent them. Quantization can introduce rounding and truncation error. The amount of quantization is determined by a step size parameter, where larger values correspond to finer granularity (i.e. greater precision and fidelity to the original image), but require more bits to represent. A different step size can be specified for each tile and decomposition level, but since there are no obvious theoretical benefits to doing so, this study only measures the effect of varying a single global value. This value is appropriately scaled according to the resolution of the decomposition level. A small step size is often recommended since quantization is a lossy procedure, and selective retention of information should be handled primarily by the entropy coder. Kakadu's default value is $1/256$, which is quite small.

Results and Recommendations

Figure 6 shows recognition accuracy as a function of file size when search images are compressed with different step sizes. All other compression parameters are left at their default values. The figure demonstrates a clear benefit to using smaller step sizes when the file size is larger (≥ 2048 bytes), although the benefit diminishes at smaller file sizes. Step sizes $1/256$ and $1/64$ yield nearly identical results. We recommend using a step size of $1/256$ because small step sizes never seem to perform worse than larger ones, and because it is a commonly used value.

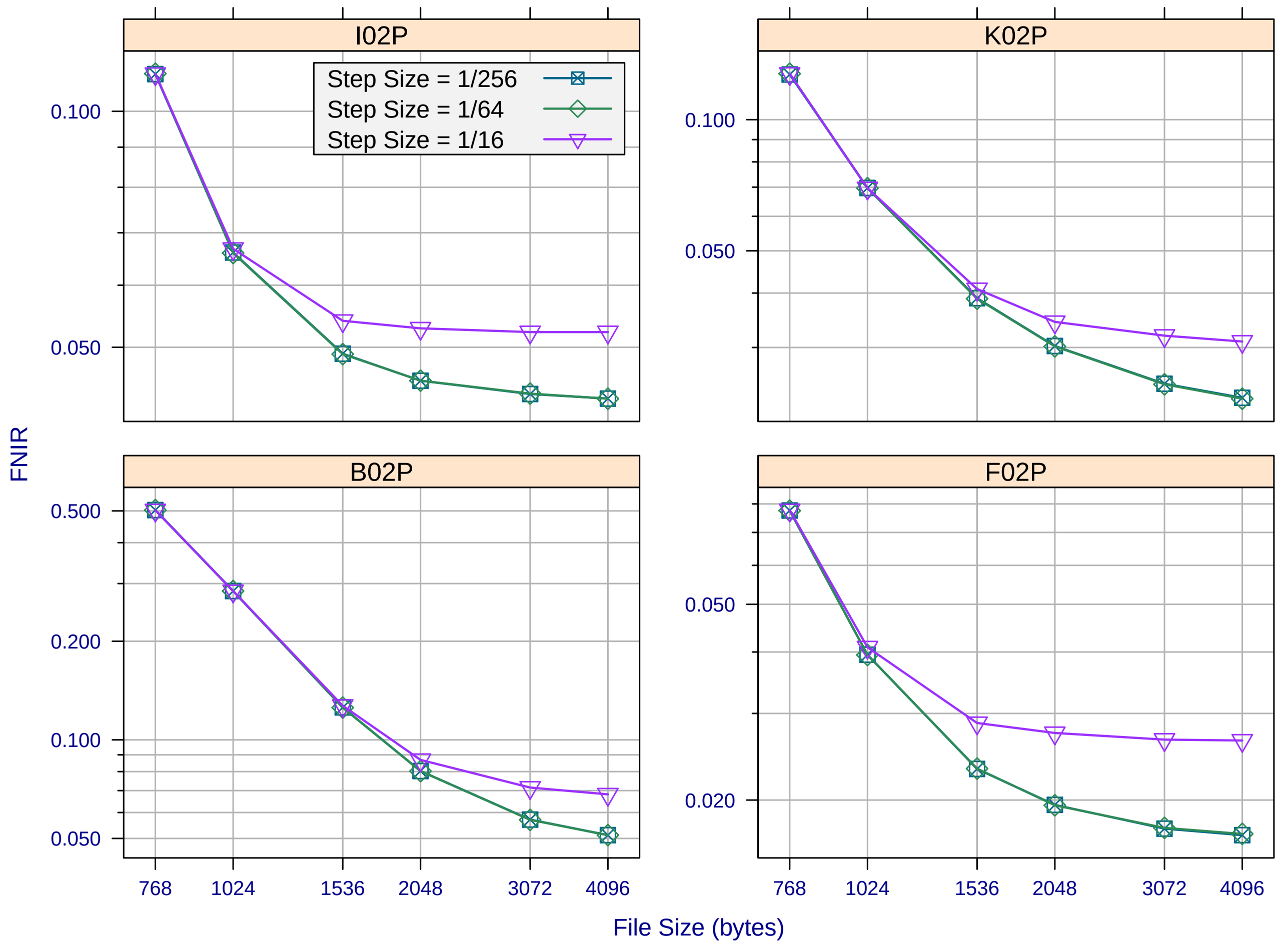


Figure 9: FNIR (at FPIR=0.001) as a function of file size for 4 matching algorithms when search images are compressed at different base step sizes. Compressed Kind 7 records (created by the matching algorithms) are searched against an enrolled population of 160 000 non-compressed images. Each point is generated from 80 135 mated searches.

4.1.3 Number of Decomposition Levels

Introduction

The wavelet transform decomposes the image into a number of different resolution levels. The process has been shown to work well for wavelet-based compression techniques. A further advantage of separating the image into multiple resolution levels is that the image can be re-constructed up to a certain resolution by only decompressing those levels that correspond to the lower frequencies. This can save time when rendering images on low-resolution embedded devices.

Results and Recommendations

Figure 10 shows recognition accuracy as a function of file size when search images are compressed with different numbers of decomposition levels. All other compression parameters are left at their default values (i.e. a single tile, a quantization step size of $1/256$, and a block size of 64). Performance tends to be poorest when only one decomposition level is used. At small file sizes (≤ 1024), 3 decomposition levels always yield the best performance for all recognition algorithms. Thus, we recommend setting the number of decomposition levels to 3 when compressing iris images.

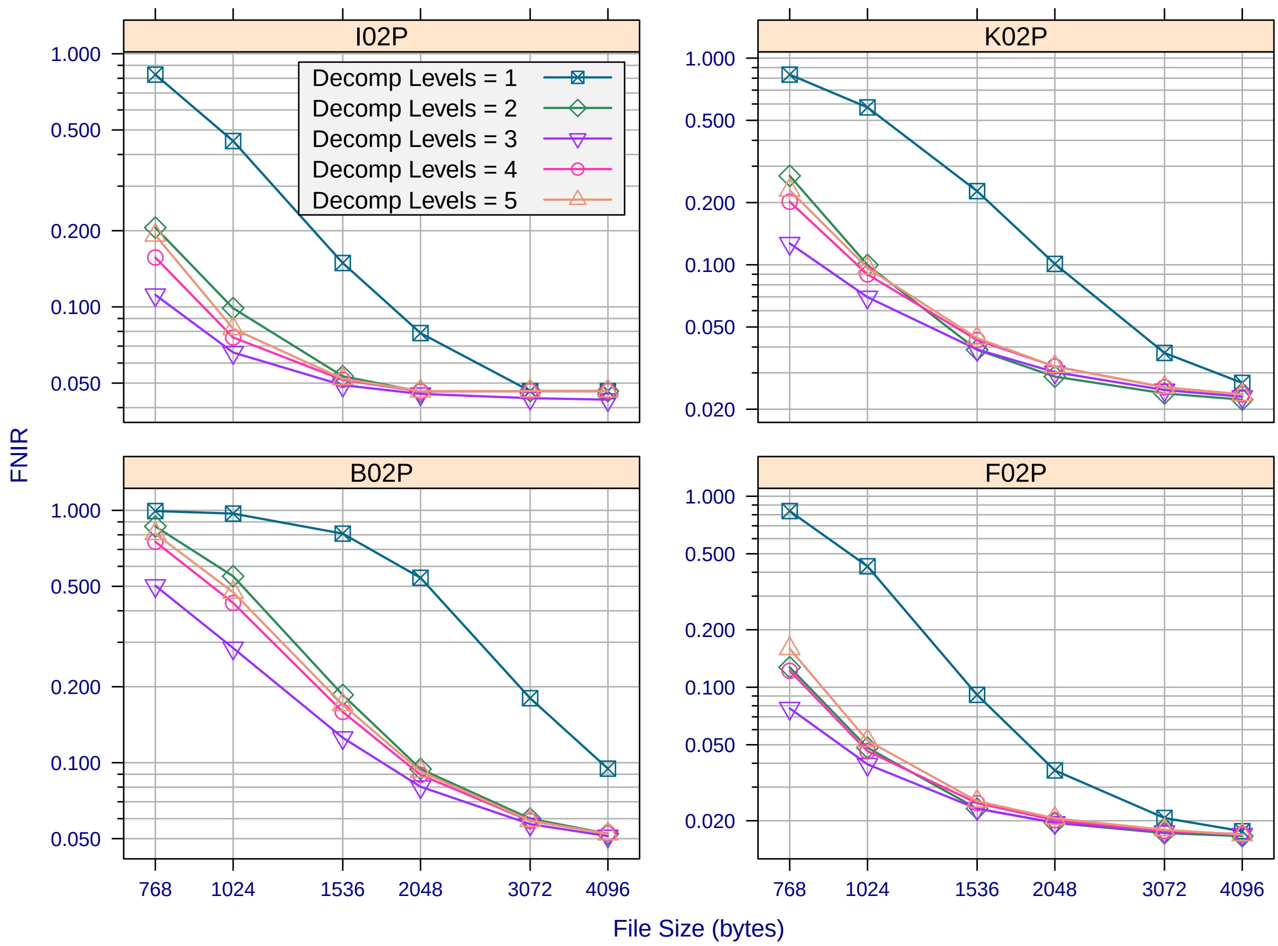


Figure 10: FNIR (at FPIR=0.001) as a function of file size for 4 matching algorithms when search images are compressed using different numbers of decomposition levels. Compressed Kind 7 records (created by the matching algorithms) are searched against an enrolled population of 160 000 non-compressed images. Each point is generated from 80 135 mated searches.

4.1.4 Block Size

Introduction

Each decomposition level is further divided into code blocks. Code blocks are encoded as bit-streams, with the most important bits located earlier in the stream. Since each code-block is coded independently of the others, a multi-threaded machine can encode blocks in parallel, saving computation time. The entropy coder operates directly on these blocks, truncating each at a point that minimizes the overall squared error loss. The code block size is normally set to either 32-by-32 or 64-by-64. These dimensions refer to the number of wavelet coefficients in the vertical and horizontal directions, both of which must be powers of 2. We see no obvious reason to test non-square dimensions given the properties of iris images. Since the total number of coefficients cannot exceed 4 096, this restricts the block size to no more than 64-by-64.

Results and Recommendation

Figure 11 shows recognition accuracy as a function of file size when search images are compressed using different code-block sizes. All other compression parameters are left at their default values. Performance is poorest for block sizes of only 8-by-8. The difference between code blocks of 32-by-32 and 64-by-64 is too small to establish statistical significance. For the sake of consistency, we recommend setting this value to 64-by-64.

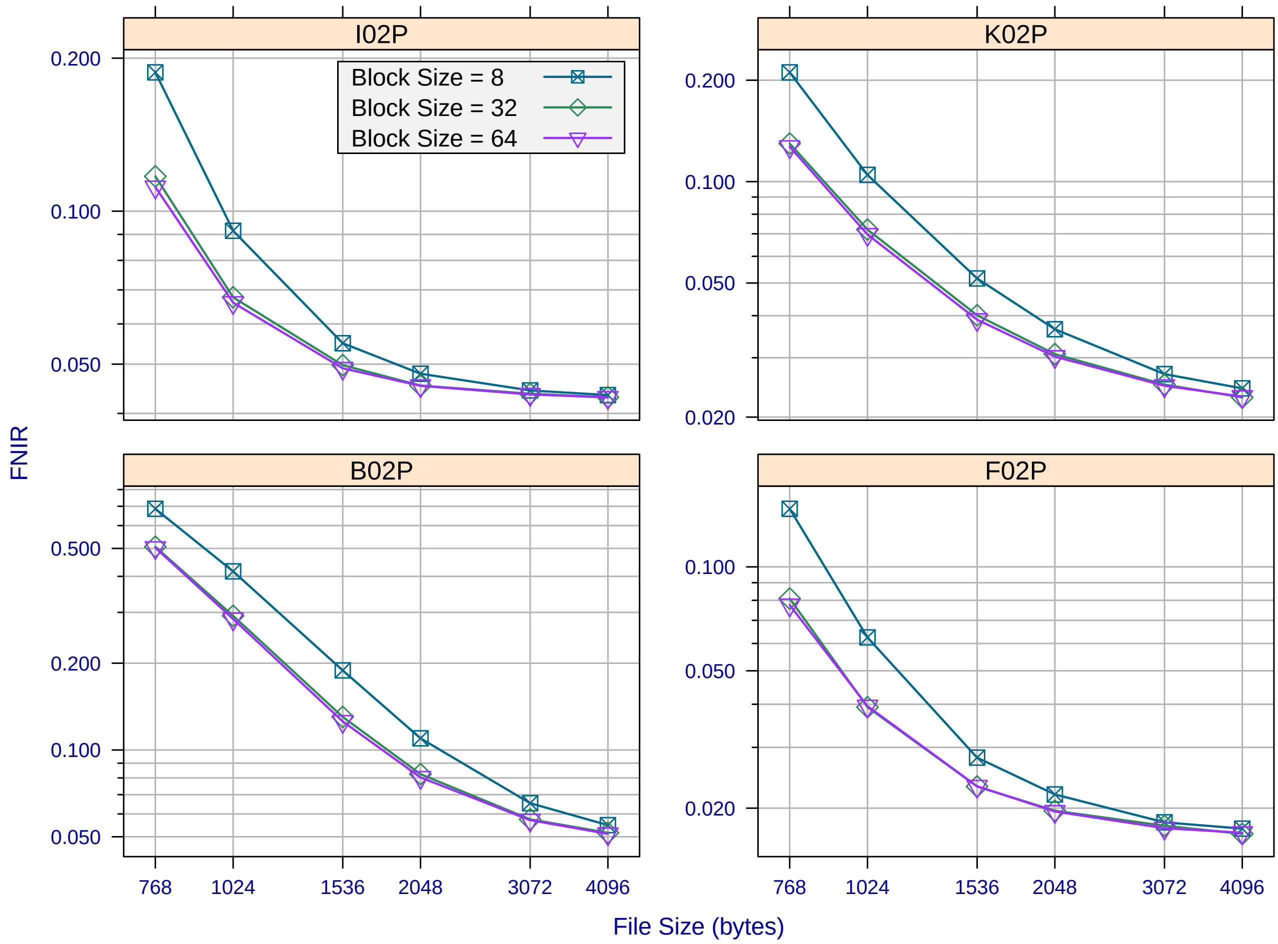


Figure 11: FNIR (at FPIR=0.001) as a function of file size for 4 matching algorithms when search images are compressed using different code-block sizes. Compressed Kind 7 record (created by the matching algorithms) are searched against an enrolled population of 160 000 non-compressed images. Each point is generated from 80 135 mated searches.

4.1.5 Timing Statistics

The time it takes to compress an iris image can be an important performance factor, especially on computationally limited devices. Figure 12 shows the distribution of compression times when Kind 2 and Kind 7 records are compressed to various file sizes. Parameters such as tile and block size are set to optimal values listed in Section 4.1.6. Section 3.5.3 describes the procedure for collecting timing statistics and Section 3.1 outlines the specifications of the timing machine, a high-end PC class blade with 6 quad-core AMD Opteron CPUs running at 2.4 GHz.

Although the targeted file size does not significantly affect compression time for Kind 7 records, Kind 2 records take longer to compress at larger file size. Generally speaking, Kind 2 records take longer to compress than Kind 7 records, possibly because the former have larger pixel dimensions and contain additional textures. Compression takes an average of 0.014 seconds for Kind 7 records, which is small compared to other steps performed during identification. Some recognition algorithms are capable of creating matching templates in as little as 3 hundredths of a second, but others require half a second or longer on a timing machine with identical specifications (see the IREX IV: Part 1 Final Report). Searching a template against an enrolled population of 1.6 million irides can take anywhere from half a second to half a minute, depending on the recognition algorithm.

Figure 13 shows the distribution of compress times when Kind 7 records are compressed with different numbers of processing threads and block sizes. Sometimes the median compression time equals the 10th or 90th percentile because the timer has only millisecond resolution. Dedicating more threads to compression does not improve end-to-end compression time, possibly because the overhead of loading software libraries, which can only be done with a single thread, dominates execution time. Median compression time is sometimes lowest when only one thread is used. Only certain steps of the compression process can utilize multiple threads. One such step is code blocking, which is performed immediately prior to EBCOT encoding. Since iris records are fairly low resolution images that compress quickly, it may be that multithreading incurs an overhead greater than the benefit of coding blocks concurrently. Compressing with a block size of only 8-by-8 appears to increase computation time. Otherwise, there is no pronounced difference in computation time for block sizes of 32-by-32 and 64-by-64.

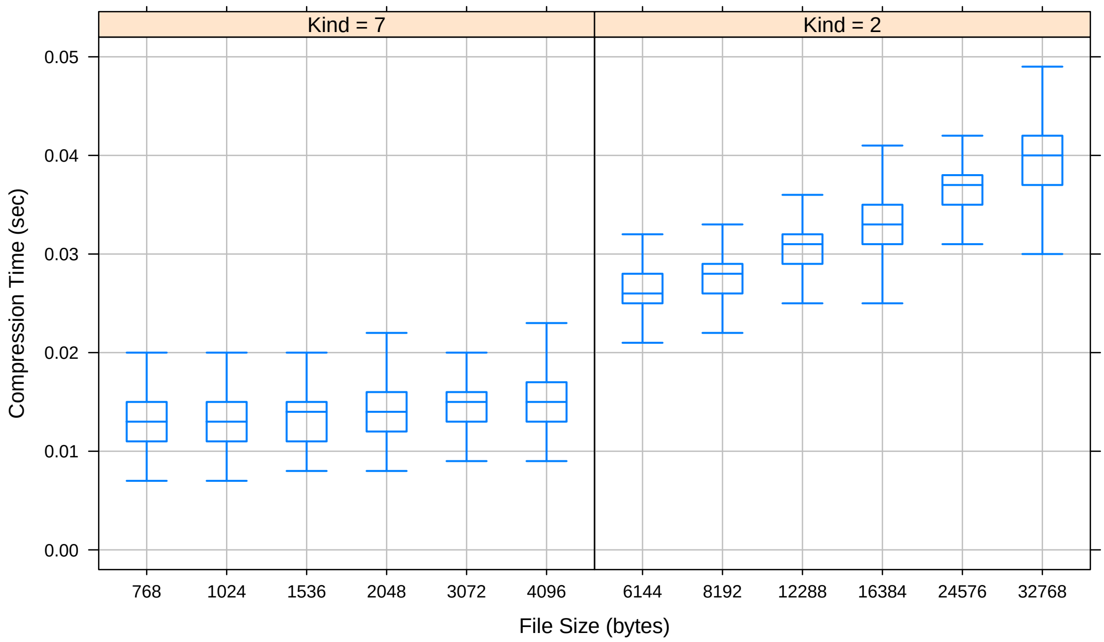


Figure 12: Distribution of compression times when Kind 2 and Kind 7 records are compressed to various file sizes on an otherwise unloaded high-end machine with 6 quad-core AMD Opteron processors operating at 2.4 GHz. Each plot is generated from compressing 5 000 images. The resolution of the timer is 0.001 seconds (rounded).

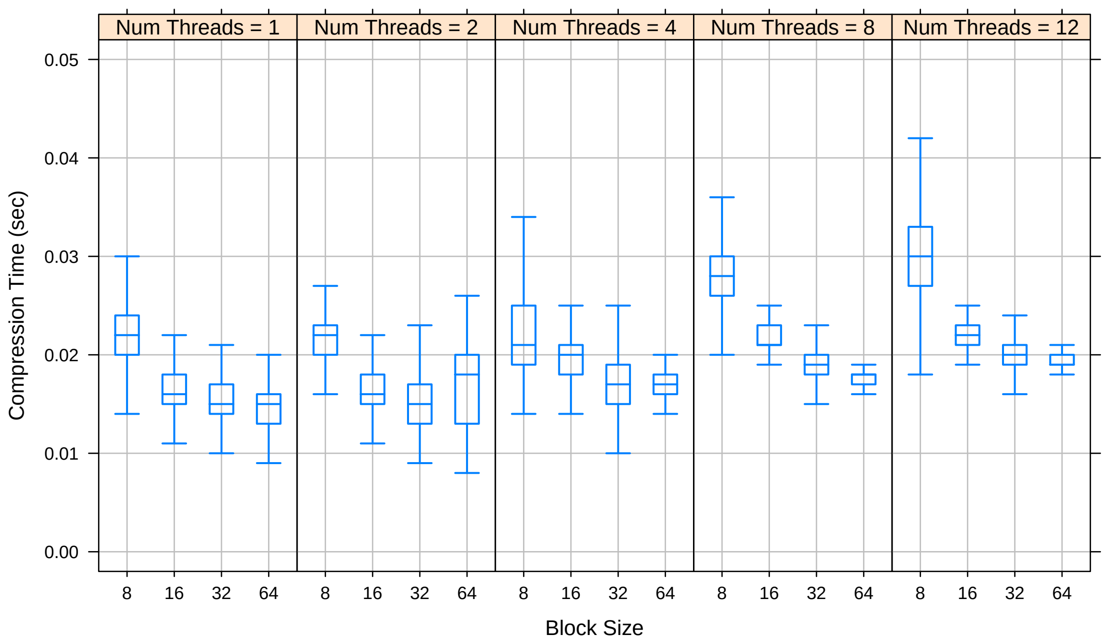


Figure 13: Distribution of compression times for Kind 7 records when different numbers of processing threads and block sizes are used. Block sizes are always square. Images are always compressed to 3 072 bytes. Each plot is generated from compressing 5 000 images. The resolution of the timer is 0.001 seconds (rounded).

4.1.6 Summary of Recommendations

The following parameter values were found to produce the smallest drop in recognition accuracy:

- a single tile,
- a block size of 64-by-64,
- a quantization step size of $1/256$.
- 3 decomposition levels.

At small file sizes (e.g. those below 2 048 bytes), recognition accuracy is especially sensitive to the number of tiles and the number of decomposition levels. At larger file sizes, recognition accuracy is more sensitive to the quantization step size. The best combination of parameters does not change depending on the file size.

4.2 Performance of Kind 7 Format

This section explores the effect of lossy compression on Kind 7 records. The Kind 7 format is the most compact format for storing iris images described in ISO/IEC 19794-6. The standard requires the iris to be centered in the image, and the superfluous area around the iris to be cropped and masked. The masking ensures that a maximum of the encoding budget is dedicated to representing the iris features rather than less relevant periocular textures.

Iris images are compressed to fixed file sizes to address applications that impose a hard upper-limit on the container size. Since Kind 7 records can vary in pixel dimensions, the bit rate (in bits-per-pixel) input to the compression software had to be adjusted for each image to correspond to the correct file size. Sometimes the image is a few bytes smaller than the targeted file size, but it is never greater.

4.2.1 Compressing Only Search Images

Applications

Some setups may only require compression of images on one side of the comparison process. Systems that compare compressed samples from digital smartcards to live captures often have no need to compress the live captures, especially when they only need to exist for the duration of the transaction. Systems that transfer compressed samples across bandwidth-limited networks may search these samples against previously enrolled samples that were never compressed. The results in this section are more closely related to the latter example since images are only compressed on the search side.

Results

Figure 14 shows DET accuracy when Kind 7 records are compressed to different file sizes and searched against an enrolled population of 160 000. Compression parameters such as block size and step size are set to optimal values identified in Section 4.1.6. Line segments connect *points of equal threshold* between curves, which show the specific effect that compression has on the mated and non-mated comparison score distributions. The following conclusions are drawn from the figure:

- Compressing search images down to 2 048 bytes results in only a small drop in accuracy ($\sim 1/3$ increase in FNIR at fixed FPIR) for algorithms I02P and F02P. Algorithm B02P experiences a greater decrease in accuracy (about a factor of 10 increase in FNIR at fixed FPIR) since it fails to mask the sclera.

- Accuracy drops much more appreciably when search images are compressed to 1 024 bytes. At fixed FPIR, FNIR increases by about a factor of 2 for I02P, a factor of 3 for F02P, a factor of 5 for K02P, and more than 10 for B02P.

- High amounts of compression decrease non-mated dissimilarity scores for algorithms I02P and B02P. In the case of I02P, the increase in FPIR at a fixed threshold is approximately 4 fold when search images are compressed to 1 024 bytes. The increase is much larger for B02P. Algorithms F02P and K02P do not exhibit appreciable increases in FPIR.

Operational Relevance

Algorithms that mask the sclera and blur the eyelid boundaries achieve noticeably superior accuracy when images are compressed to small file sizes ($\leq 3,072$ bytes). Kind 7 records can be compressed to sizes as small as 2 048 bytes with only minor degradation in recognition accuracy. Accuracy drops much more quickly when search images are compressed to sizes smaller than 2 048 bytes. Compared to other biometric modalities, standard fingerprint minutiae information can be

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

stored in as little as 400 bytes [8]. and face images can be compressed to about 8KB [27], although the ISO/IEC 19794-5 standard [28] recommends 30KB to be safe.

High amounts of compression tend to increase dissimilarity scores for mated comparisons. High amounts of compression sometimes decrease dissimilarity scores for non-mated comparisons, but to a lower extent. Sometimes it may be advantageous to adjust the decision threshold depending on the amount of compression applied to the images.

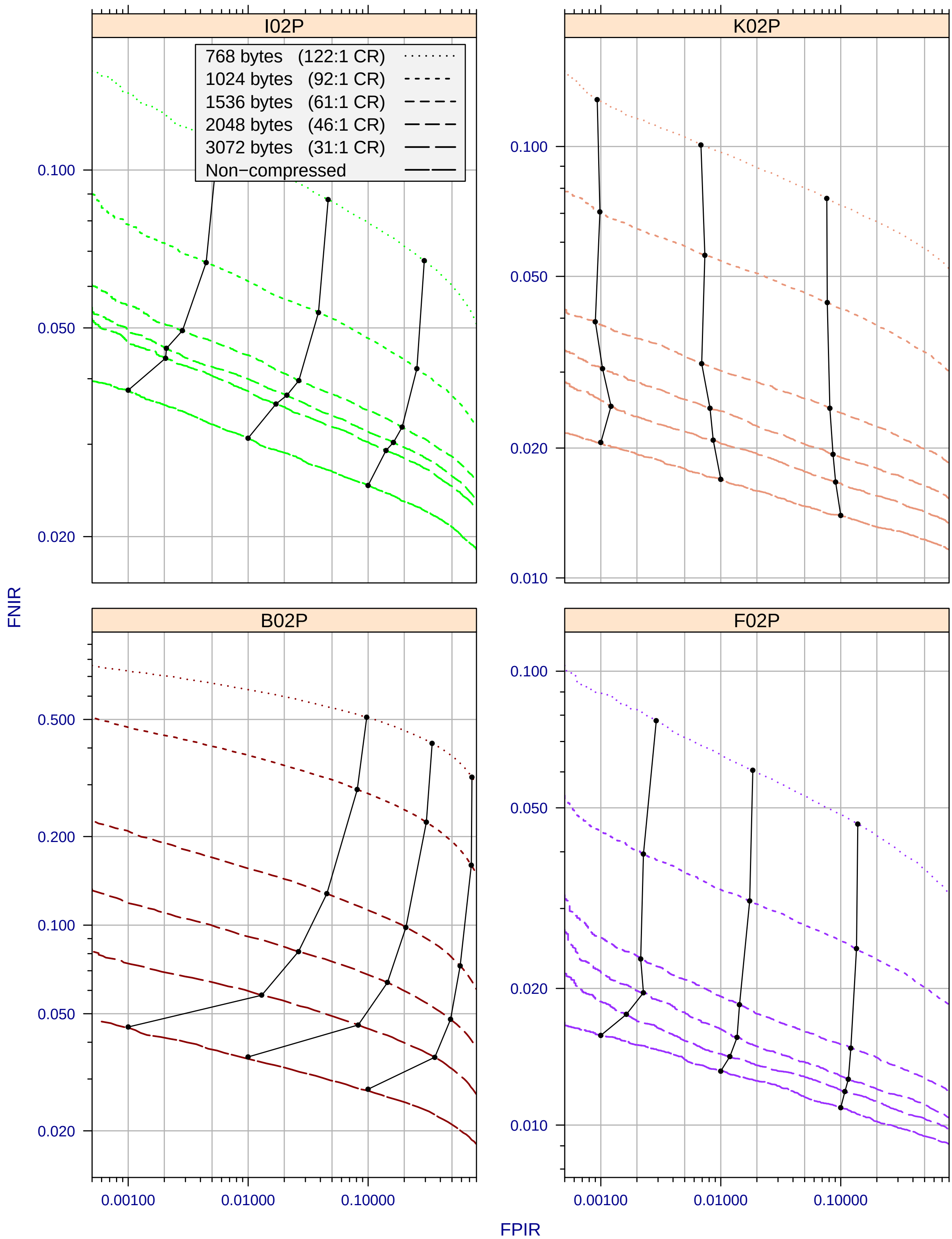


Figure 14: DET accuracy when Kind 7 records are compressed to different file sizes and searched against an enrolled population of 160 000. Enrolled images were never compressed. Line segments connect points of equal threshold. The mean compression ratio corresponding to each file size is included in the legend. Plots are generated using 80 135 mated and 60 000 non-mated searches.

4.2.2 Compressing Search and Enrollment Images

Applications

Some iris recognition systems may match compressed images against other compressed images. Systems that must transfer both the smartcard sample and the live capture across a network are likely to compress both. A setup was described in Section 4.2.1 where compressed iris samples are transferred and searched against a back-end database of enrolled samples. After a positive identification is made, the system may opt to replace the sample on the enrollment side with the newly acquired (and compressed) sample that it matched.

Results

Figure 15 shows DET accuracy when Kind 7 search and enrollment images are similarly compressed to targeted file sizes. The enrolled population is 160 000. Compression parameters such as block size and step size are set to optimal values identified in Section 4.1.6. Lines segments connect *points of equal threshold* between curves, which show the specific effect that compression has on the mated and non-mated comparison score distributions. The following conclusions are drawn from the figure:

- Accuracy is similar to when only the search images are compressed. Compression sizes of 2 048 bytes result in only a small drop in accuracy ($\sim 1/5$ factor increase in FNIR at fixed FPIR) for algorithms I02P and F02P. Algorithm B02P experiences a greater decrease in accuracy (about a factor of 10 increase in FNIR at fixed FPIR) since it fails to mask the sclera.

- Accuracy drops much more appreciably when images are compressed to sizes smaller than 2 048 bytes. At fixed FPIR and a compression size of 1 024 bytes, FNIR increases by about a factor of 2 for I02P, a factor of 2.5 for F02P, a factor of 4.5 for K02P, and more than 20 for B02P.

- High amounts of compression increase non-mated dissimilarity scores for algorithms I02P and B02P, but to a lesser extent than if only search images had been compressed (determined by comparing these results to Figure 14). In the case of I02P, the increase is minor (no more than a factor of 3 increase in FPIR at fixed threshold when images are compressed to 1 024 bytes). Algorithms F02P and K02P do not experience appreciable increases in FPIR.

Operational Relevance

Results are similar to when only search images are compressed. Algorithms that mask the sclera and blur the eyelid boundaries achieve noticeably superior accuracy when images are compressed to small file sizes ($\leq 3{,}072$ bytes). Kind 7 records can be compressed to sizes as small as 2 048 bytes with only minor degradation in recognition accuracy. Accuracy drops much more quickly when images are compressed to sizes smaller than 1 536 bytes. High amounts of compression tend to increase dissimilarity scores for mated comparison. Dissimilarity scores for nonmated comparisons decrease somewhat for some algorithms, but to a lesser extent than when both images are compressed.

Lossy compression discards potentially identifying information. While it is preferable to discard as little information as possible, compressing enrolled images by an amount comparable to search images does not lead to an appreciable drop in accuracy. This may be because compression tends to discard similar feature information in both images, and iris matchers benefit mostly from feature information only when it is present in both images.

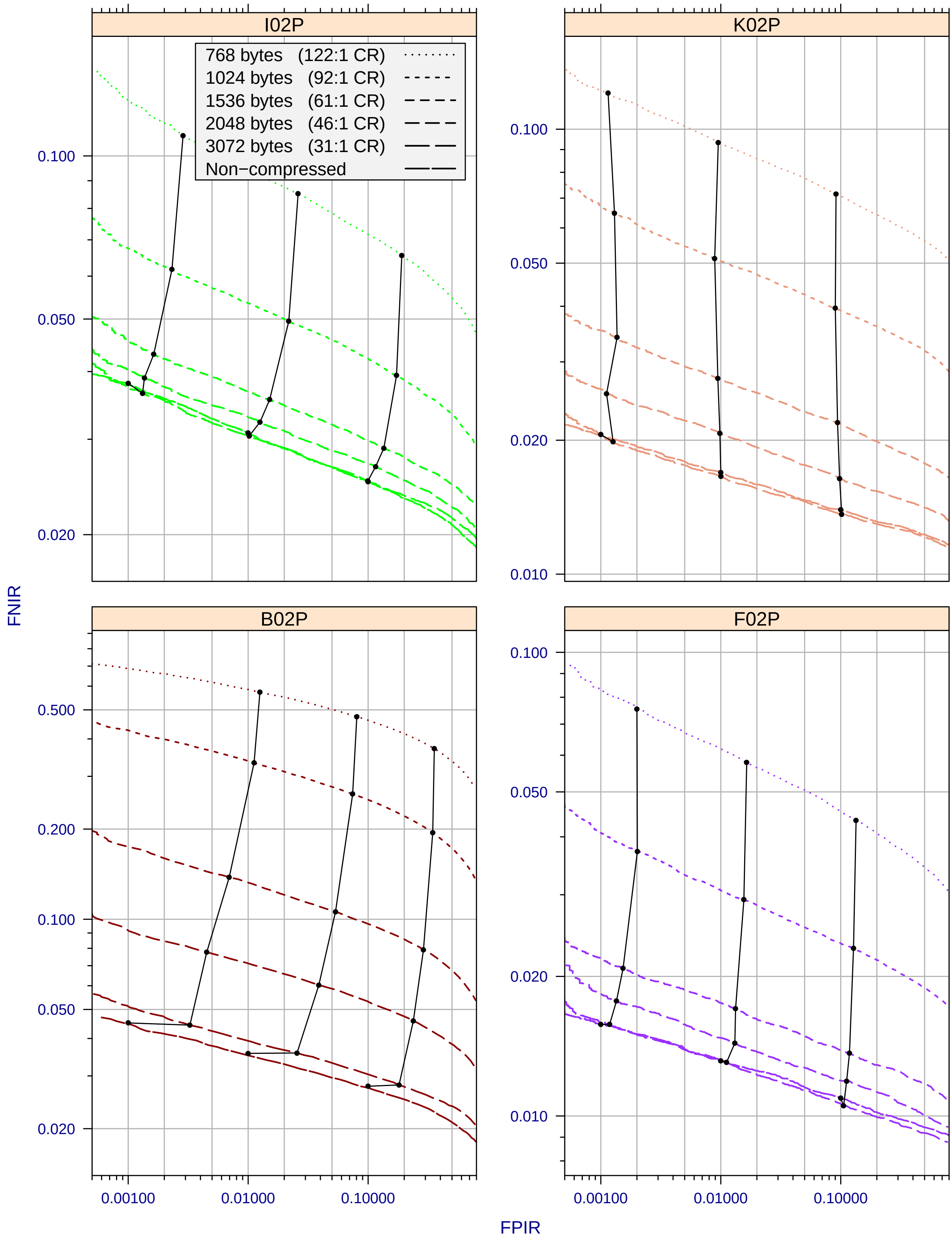


Figure 15: DET accuracy when Kind 7 records are compressed to different file sizes. The size of the enrollment population is 160 000. Both search and enrollment images are compressed. Line segments connect points of equal threshold. The mean compression ratio corresponding to each file size is included in the legend. Plots were generated using 80 135 mated and 60 000 non-mated searches.

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

4.2.3 When to use Both Eyes

Applications

If the goal is to maximize recognition accuracy, images of both eyes should be used for matching whenever available. However, if only a fixed amount of storage space is available, the question becomes whether better accuracy is achieved by storing images of both eyes, or a more lightly compressed version of just one. Part 1 of the IREX IV report found that using both eyes results in only about a factor of 3 to 4 reduction in FNIR at fixed FPIR, which is indicative of a high degree of positive correlation between left and right eyes captured during the same session. This makes sense since people tend to blink or look off to the side simultaneously with both eyes. Unfortunatley, it diminishes the benefit to using both eyes for matching.

Results and Recommendations

Figure 16 shows recognition accuracy as a function of file size when searches are performed with one eye, and with two eyes. Compression parameters such as tile and block size are set to optimal values identified in Section 4.1.6. Only search images are compressed. The figure demonstrates that there is a crossover point, where one-eye matching is more accurate at lower storage capacities, but less accurate than two-eye matching at larger storage capacities. The crossover tends to occur between 2 and 3KB. At sizes of 1.5KB, one-eye matching is consistently more accurate. At 4KB or more, two-eye matching is usually more accurate. Note that matching with two eyes introduces an additional computation penalty that may offset possible accuracy benefits.

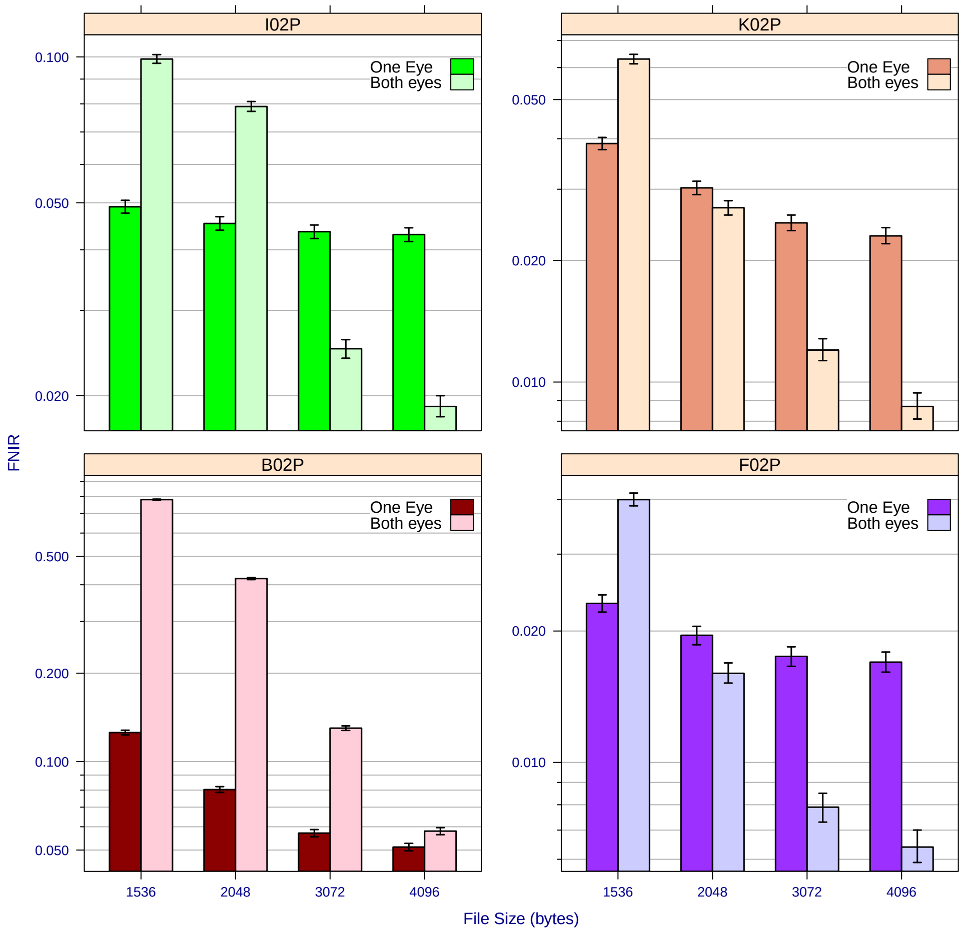


Figure 16: Comparison of FNIR (at an FPIR of 0.001) for one-eye and two-eye matching at different maximum storage capacities. Only search images are compressed. Each point is generated from 80 135 mated searches and 60 000 nonmated searches against an enrolled population of 160 000.

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

4.2.4 Should Images be Downsampled?

Introduction

Downsampling an image's pixel resolution prior to performing compression may improve recognition accuracy when the amount of compression is high. Downsampling selectively discards the highest frequency information in the image. If this information is less important for matching, then ensuring that it is discarded during the compression procedure may improve recognition accuracy. Extending this line of thinking a bit further, an optimal combination of resolution downsampling and JPEG 2000 compression could be identified that selectively retains the frequency ranges most important for matching. However, over-tuning of the compression procedure runs the risk of compromising interoperability, especially since not all recognition algorithms use precisely the same features for matching.

Two methods of downsampling are tested in this study. The first involved simple 2x2 pixel averaging. The downsampled image is then passed to the JPEG 2000 compression algorithm. The compressed image is then decompressed and upscaled to its original size before it is passed to the matcher. Upscaling is performed via bilinear interpolation. The second method of downsampling simply instructs the Kakadu implementation to allocate no space to representing the highest frequency information in the image. This has a roughly similar effect to 2x2 pixel averaging, but allows the step to be performed directly by the JPEG 2000 encoder. One would also expect the JPEG 2000 encoder to do a better job of minimizing the mean square error loss subject to the given constraint.

A similar study on automated face recognition [27] concluded that downsampling provided no perceivable accuracy benefit since JPEG 2000 already preferentially discards the higher frequency information during compression.

Results and Recommendations

Figure 17 compares recognition accuracy when different methods of compression are applied to search images. Compression parameters such as tile and block size are set to optimal values identified in Section 4.1.6. Results are inconsistent across algorithms, and no single method of compression works best in all cases. Reducing the image resolution through Kakadu achieves the best results for algorithms I02P and F02P, and at compression sizes ≥ 2048 for algorithm K02P. However, the improvement is sometimes so small that it may not be statistically significant. Furthermore, whenever downsampling appears to offer a benefit, the fractional drop in FNIR remains almost constant over the full range of file sizes. One would expect the performance disparity to be greater at smaller file sizes. More likely, downsampling is removing some type of noise (e.g. camera shot noise) that leads to the improvement.

Neither method of downsampling results in a consistent improvement in recognition accuracy. Many other methods of decimation and/or filtering are possible, and some may reap clear benefits. Although further investigation is warranted, downsampling cannot be recommended at this time.

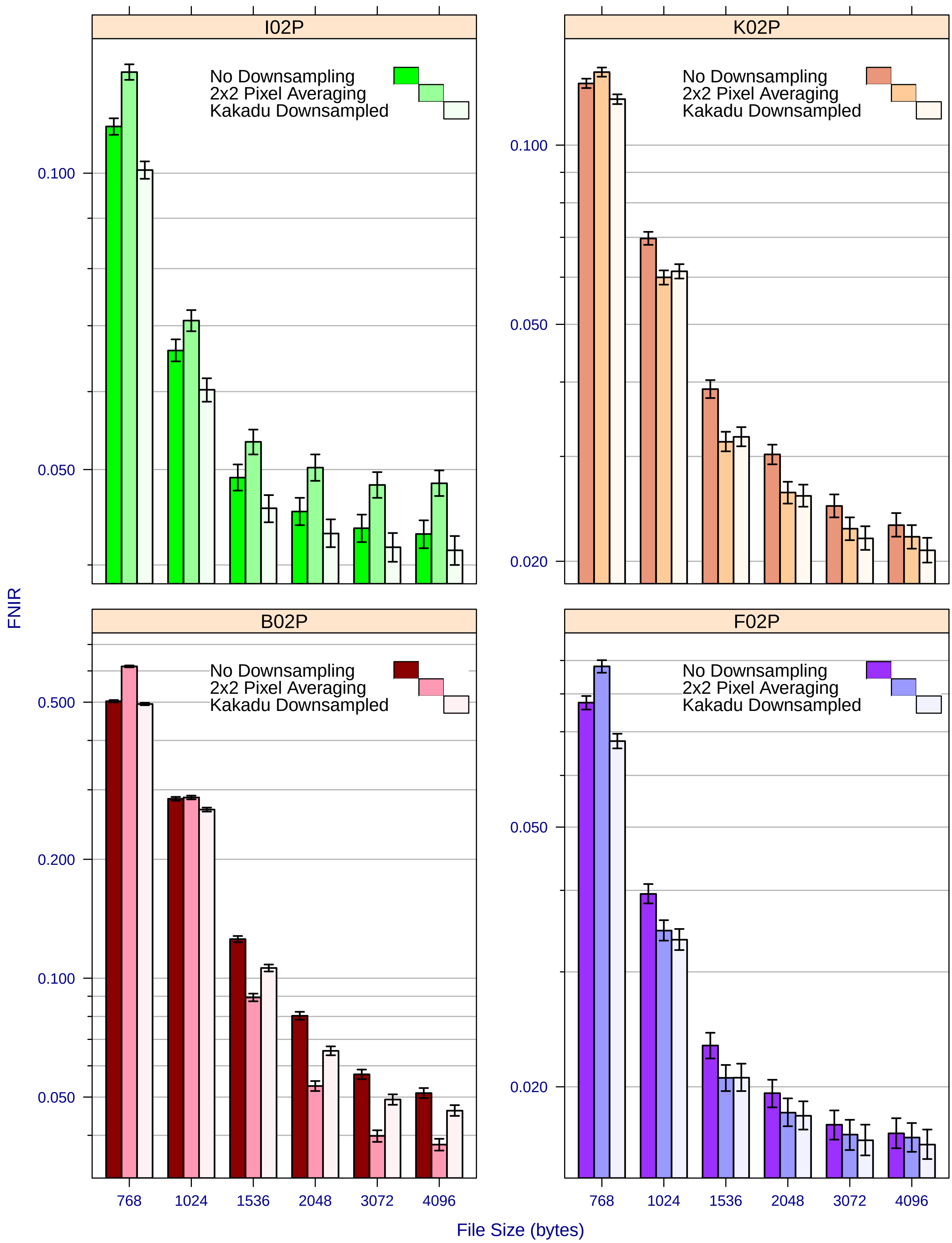


Figure 17: Comparison of FNIR (at FPIR=0.001) with and without downsampling applied to search images. 80 135 searches are performed against an enrolled population of 160 000 (non-compressed) iris images. Error bars show 95% confidence intervals.

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

4.3 Performance of Kind 2 Format

Kind 2 records must have pixel dimensions of 640x480 and a bit depth of 8, but otherwise do not require a masking or cropping of the area outside the iris. As such, they cannot be compressed to sizes as small as Kind 7 records without suffering much greater losses in accuracy. Iris cameras typically return Kind 2 images, so no additional processing is required to create them. Creating Kind 7 records, on the other hand, requires localization of the iris center as well as the limbus and eyelid boundaries, which is a non-trivial task.

4.3.1 Compressing Only Search Images

Introduction

When file size constraints are more relaxed, it may not be necessary to convert Kind 2 records into more compact Kind 7 representations. Doing so requires localization of the limbus and eyelid boundaries, often on the client side where computational resources may be limited.

Results and Recomendations

Figure 18 shows DET accuracy when Kind 2 are compressed to different file sizes and searched against an enrolled population of 160 000. Compression parameters such as block and step size are set to optimal values identified in section 4.1.6. Line segments connect *points of equal threshold* between curves, which shows the specific effect that compression has on the mated and non-mated comparison score distributions. The following conclusions are drawn from the figure:

- Compressing search images down to 16 384 bytes results in only a moderate to small drop in accuracy ($< 1/3$ increase in FNIR at fixed FPIR) for all algorithms.
- Accuracy drops more appreciably when search images are compressed to 8 192 bytes. At fixed FPIR, FNIR increases by about a factor of 2 for I02P and K02P, and a bit more than a factor of 2 for B02P and F02P. At 6 144 bytes, the factor increase in FNIR ranges from about 2.5 to 4 depending on the algorithm.
- High amounts of compression decrease non-mated dissimilarity scores for algorithms I02P and B02P. In the case of I02P, the increase in FPIR at a fixed threshold is no more than a factor of 5 when search images are compressed to 6 144. Algorithms F02P and K02P do not experience appreciable increases in FPIR.

Figure 19 compares the ability of iris recognition algorithms to match highly compressed Kind 2 and Kind 7 records. Kind 7 records typically perform better at file sizes under 16 384. At larger file sizes there appears to be little or no accuracy benefit to matching Kind 7 records over their Kind 2 counterparts. Kind 7 records achieve the same FNIR as Kind 2 records at only a fraction of the size. FNIR is comparable when file sizes are reduced by a factor of 5 for B02P and K02P, and a factor of 8 for I02P and F02P.

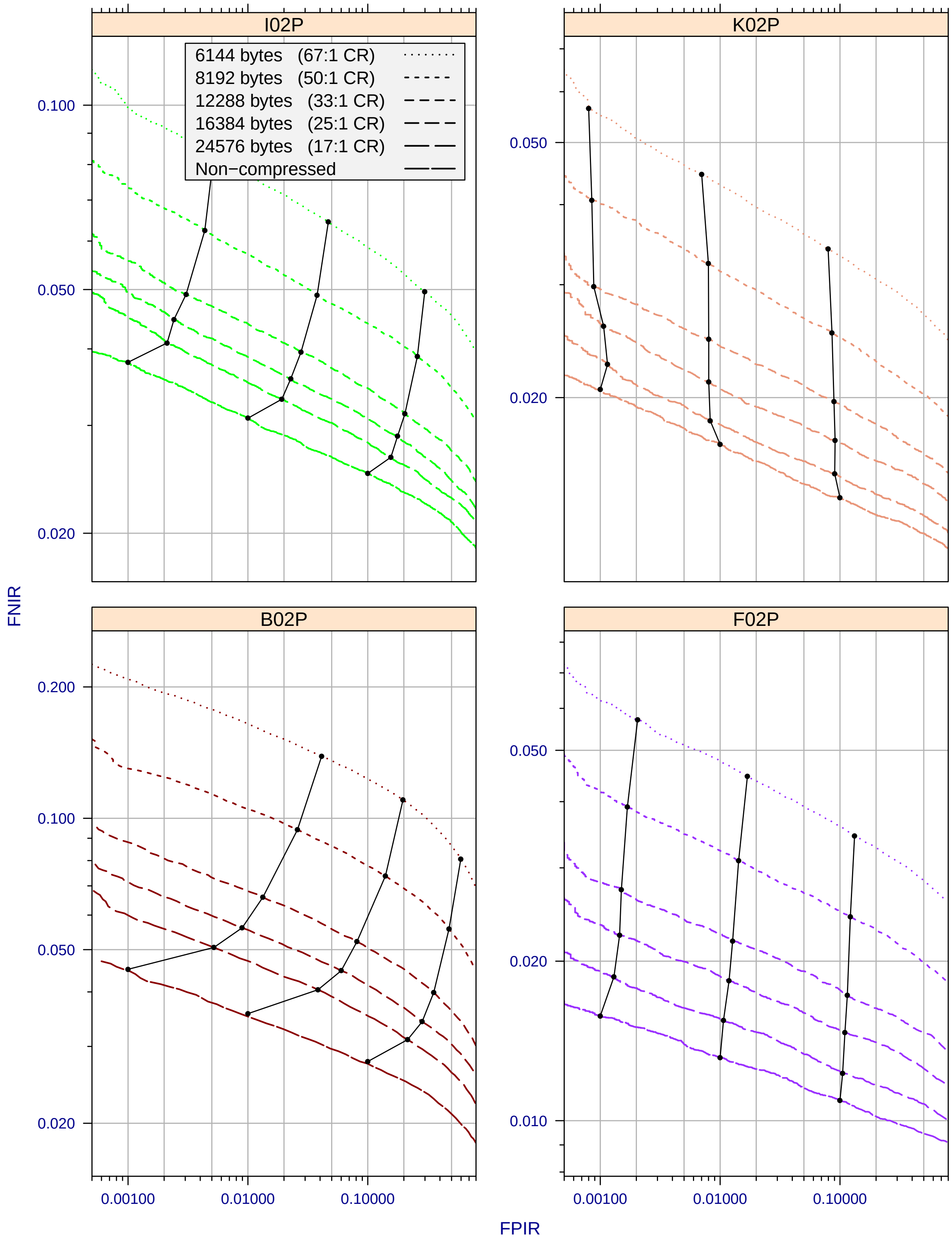


Figure 18: DET accuracy when Kind 2 records are compressed to different file sizes and searched against an enrolled population of 160 000. Enrolled images were never compressed. Line segments connect points of equal threshold. The compression ratio corresponding to each file size is included in the legend. Plots are generated using 80 135 mated and 60 000 non-mated searches.

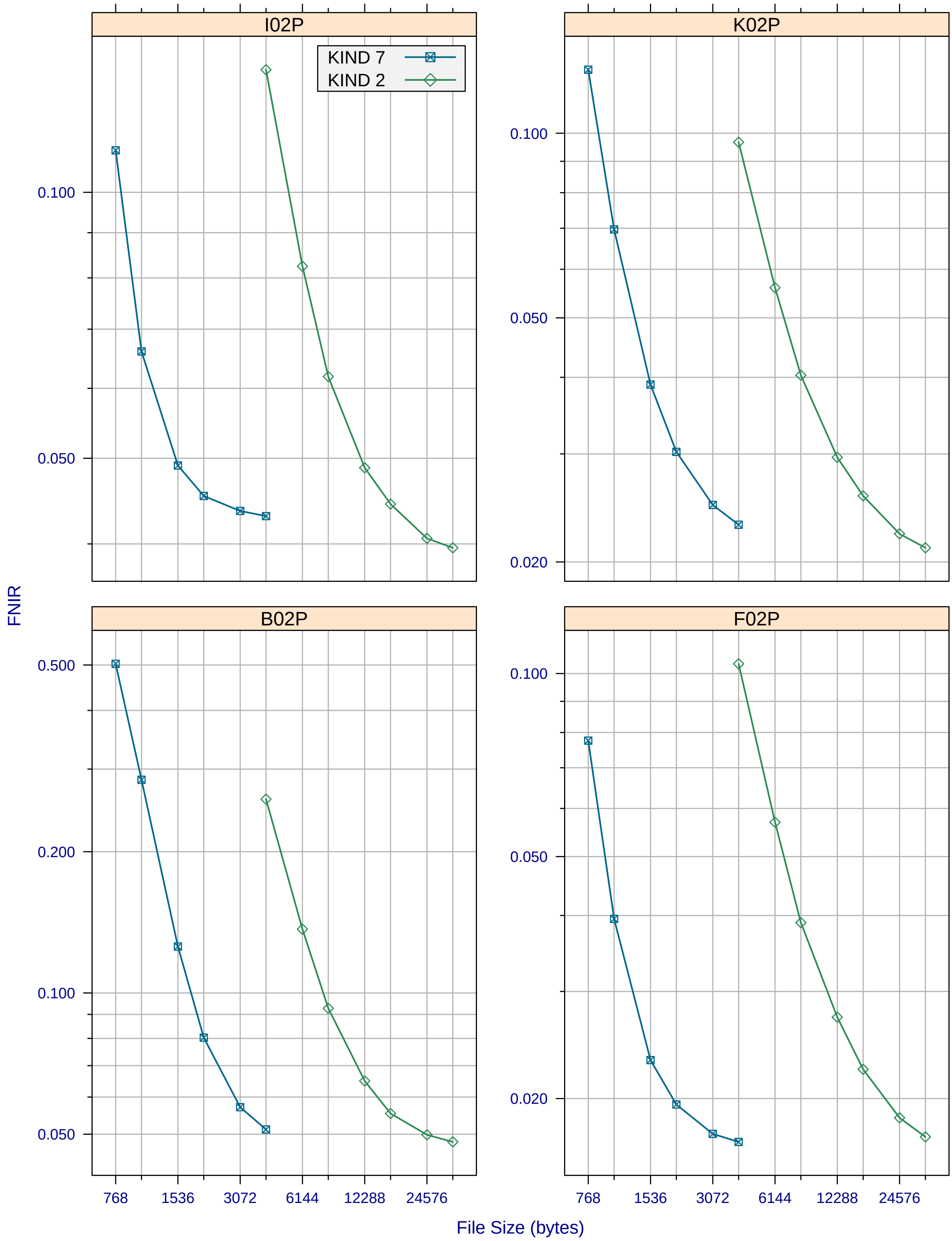


Figure 19: Comparison of FNIR (at FPIR=0.001) when Kind 2 and Kind 7 records are compressed to various file sizes and searched against an enrolled population of 160 000 non-compressed images. Each point is generated from 80 135 mated searches. Kind 7 records perform better at smaller file sizes because they undergo cropping and masking that allows them to be stored more compactly.

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

4.3.2 Compressing Search and Enrollment Images

Introduction

Some iris recognition systems may match compressed Kind 2 records against other compressed iris records. Possible setups are described in Section 4.2.2.

Results and Recomendations

Figure 20 shows DET accuracy when Kind 2 search and enrollment images are similarly compressed to targeted file sizes. The size of the enrolled population is 160 000. Compression parameters such as block and step size are set to optimal values identified in section 4.1.6. Line segments connect *points of equal threshold* between curves, which shows the specific effect that compression has on the mated and non-mated comparison score distributions. The following conclusions are drawn from the figure:

- Accuracy is similar to when only the search images are compressed. Compression sizes of 16 384 bytes result in moderate drops in accuracy ($< 1/4$ increase in FNIR at fixed FPIR) for all algorithms.
- Accuracy drops more appreciably when search images are compressed to 8 192 bytes. At fixed FPIR, FNIR increases by less than a factor of 2 for I02P, about a factor of 2 for F02P and K02P, and about a factor of 2.5 for B02P. At 6 144 bytes, the factor increase in FNIR ranges from about 2 to 4.5 depending on the algorithm.
- High amounts of compression increase non-mated dissimilarity scores for algorithms I02P and B02P, but to a lesser extent than if only search images had been compressed (determined by comparing these results to Figure 18). In the case of I02P, the increase is minor (no more than a factor of 3 increase in FPIR at fixed threshold when images are compressed to 6 144 bytes). Algorithms F02P and K02P do not experience appreciable increases in FPIR.
- Light compression appears to actually improve performance for all algorithms. For algorithms I02P and K02P, the drop in FNIR at fixed FPIR is almost $1/3$ when images are compressed to 24576 bytes. The compression may be removing some type of noise (e.g. shot noise) from the images. Lightly compressing iris images for the sole purpose of improving accuracy is not recommended since these results may not translate to other iris data. Furthermore, such processing should be handled internally by the recognition algorithm.

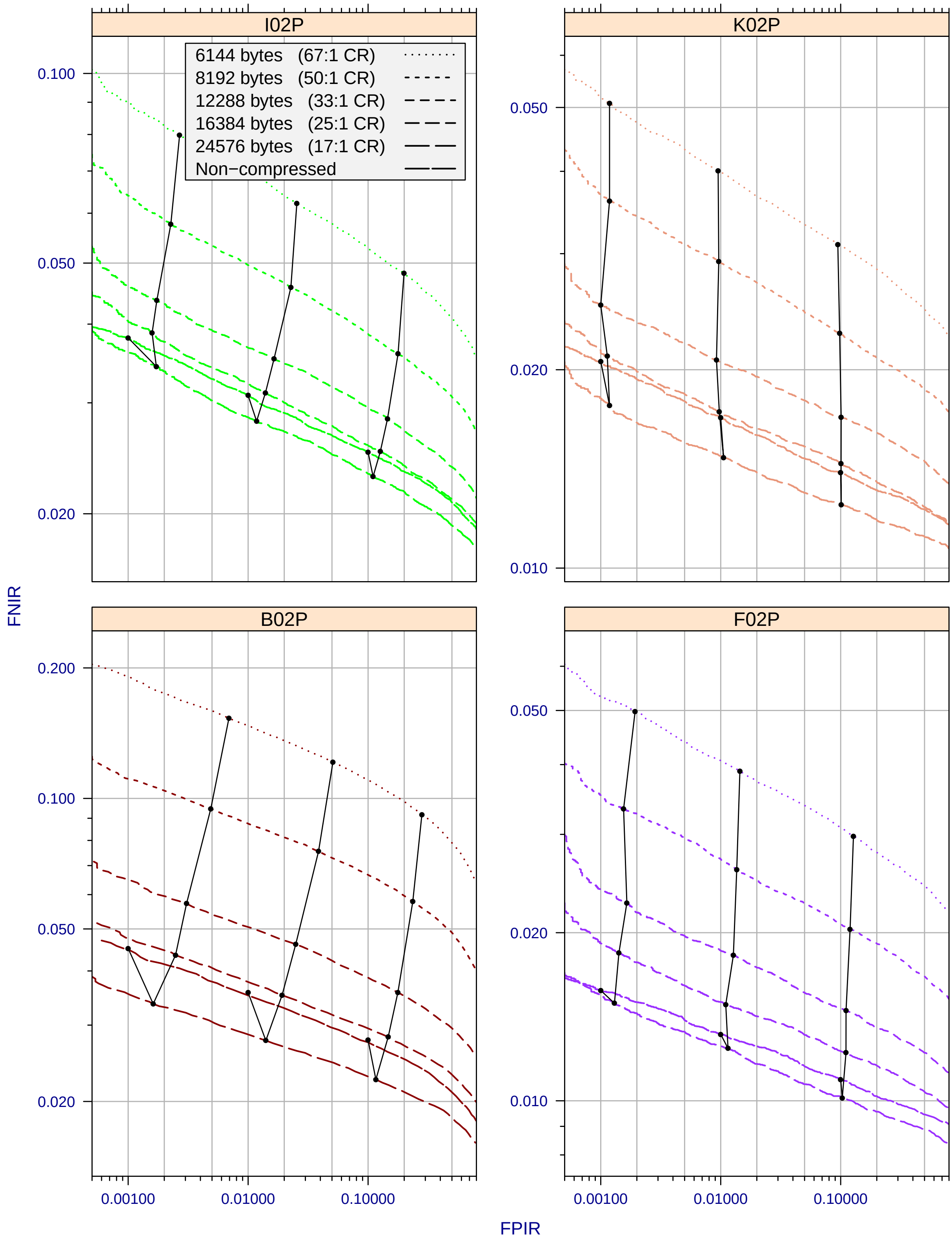


Figure 20: DET accuracy when Kind 2 search and enrollment images are similarly compressed to targeted file sizes. The size of the enrolment population is 160 000. Line segments connect points of equal threshold. The compression ratio corresponding to each file size is included in the legend. Plots are generated using 80 135 mated and 60 000 non-mated searches.

4.4 Lossless Compression

Introduction

JP2 can compress images losslessly. Lossless compression retains all of the information in the image so that it can be perfectly reconstructed to its original form. Since no information can be discarded, it cannot achieve file sizes as small as when lossy compression is used. Smaller sizes can still be achieved by converting the image to a Kind 7 before compressing it since they have smaller pixel dimensions and uniform areas of solid color that are easy to represent compactly. Lossless compression requires the use of the 5/3 CDF wavelet transform.

Results

Figure 21 shows the distribution of file sizes achieved when images are compressed losslessly in their original Kind 2 format, and when they are converted to Kind 7 formats by 4 different algorithms. Kind 2 images compress to a mean size of 135KB. Algorithms I02P and F02P achieve the lowest file sizes for Kind 7 images since they mask the sclera and blur the eyelid boundaries. The mean file sizes are 20KB and 21KB for algorithms I02P and F02P respectively. An alternative lossless compression format is PNG. The IREX I Final Report found that libpng [29] compresses Kind 7 records to a median size of 25KB, and Kind 2 images to a median size of 150KB.

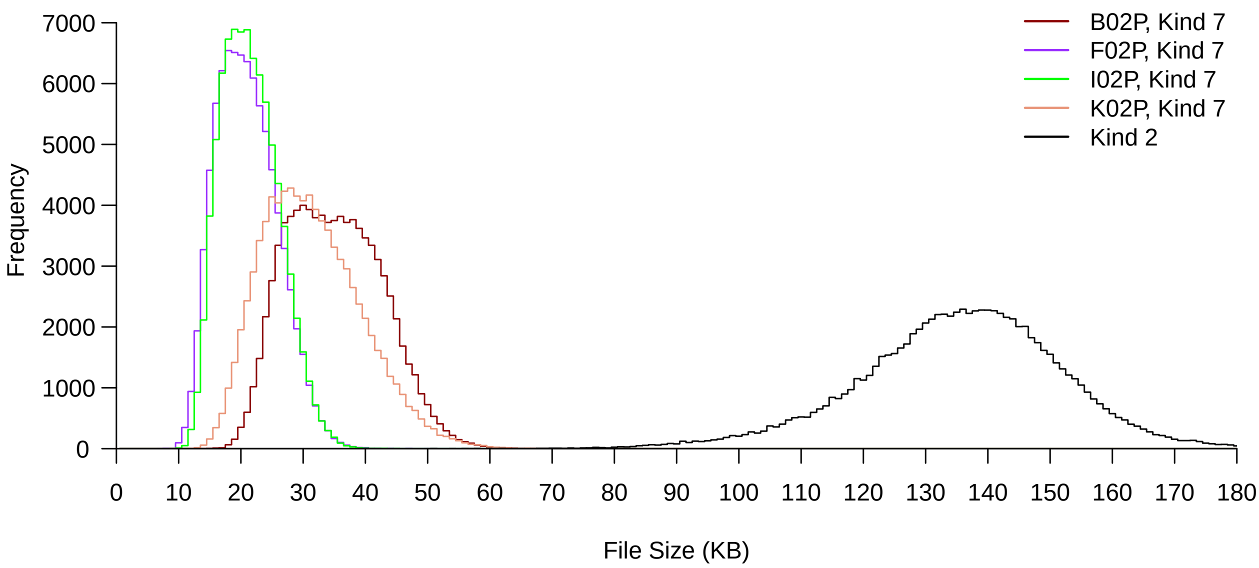


Figure 21: Distribution of file sizes when Kind 2 images, and Kind 7 images generated by different algorithms, are compressed losslessly. Each histogram is generated from 96 635 iris images.

5 References

[1] P. Grother, E. Tabassi, G. W. Quinn, and W. Salamon, “Performance of Iris Recognition Algorithms on Standard Images.” http://www.nist.gov/itl/iad/ig/irex.cfm, 2009. 1

[2] E. Tabassi, P. Grother, and W. Salamon, “IREX - IQCE Performance of Iris Image Quality Assessment Algorithms.” http://www.nist.gov/itl/iad/ig/irexii.cfm, 2011. 1

[3] *ISO/IEC 29794-6 - Biometric Sample Quality Standard- Part 6: Iris Image*. 2012. 1

[4] P. Grother, G. Quinn, J. Matey, M. Ngan, W. Salamon, G. Fiumara, and C. Watson, “IREX III: Performance of Iris Identification Algorithms.” http://www.nist.gov/itl/iad/ig/irexiii.cfm, 2011. 1

[5] P. Grother, J. R. Matey, E. Tabassi, G. W. Quinn, and M. Chumakov, “IREX VI: Temporal Stability of Iris Recognition Accuracy.” http://biometrics.nist.gov/cs_links/iris/irexVI/irex_report.pdf, 2013. 1

[6] “The IREX Program.” http://www.nist.gov/itl/iad/ig/irex.cfm. 1

[7] “Special Publication 800-73-4 - Interfaces for Personal Identity Verification.” http://csrc.nist.gov/publications/PubsDrafts.html, 2013. 2

[8] *FIPS PUB 201-2: Personal Identity Verification (PIV) of Federal Employees and Contractors*. 2013. 2, 16

[9] “Registered Traveler Interoperability Consortium - Technical Interoperability Specification Version 1.7.” http://www.rtconsortium.org/_docpost/RTICTIGSpec_v1.7.pdf, 2008. 2

[10] “ICAO Doc 9303 - Machine Readable Travel Documents.” http://www.icao.int/publications/pages/publication.aspx?docnum=9303, 2008. 2

[11] “Prime Minister launches Aadhaar Enabled Service Delivery.” Press Release, October 2012. http://uidai.gov.in/images/2nd_anniversary/uidai_press_release_for_oct_20.pdf. 2

[12] *ISO/IEC 19794-6 - Biometric Data Interchange Formats - Iris Image Data*. 2011. 3

[13] *ANSI/NIST-ITL 1-2011 Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information*. 2011. 3

[14] G. Quinn and P. Grother, “Irex iv: Evaluation of one-to-many iris recognition ,concept, evaluation plan, and api specification.” http://www.nist.gov/itl/iad/ig/irexiv.cfm, May 2012. 3

[15] P. J. Phillips, A. Martin, C. I. Wilson, and M. Przybocki, “An introduction to evaluating biometric systems,” *Computer*, vol. 33, pp. 56–63, Feb. 2000. 3

[16] G. Quinn and P. Grother, “IREX III Supplement I: Failure Analysis.” http://www.nist.gov/itl/iad/ig/irexiii.cfm, 2011. 4

[17] D. Taubman, “Kakadu software version 7.0.” http://www.kakadusoftware.com/index.php?option=com_content&task=view&id=56&Itemid=26. 5

[18] *ISO/IEC 15444-1:2004 - Information technology – JPEG 2000 image coding system: Core coding system*. 2004. 5

[19] H. Drolon, F. Devaux, A. Descampe, Y. Verschueren, D. Janssens, and B. Macq, “OpenJPEG.” http://www.openjpeg.org/. 5

[20] M. D. Adams, “JASPER.” http://www.jpeg.org/jpeg2000/testlinks.html/. 5

[21] “Nexus program description.” http://www.cbp.gov/xp/cgov/travel/trusted_traveler/nexus_prog/nexus.xml. 6

[22] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, “The DET curve in assessment of detection task performance,” in *Proc. Eurospeech*, pp. 1895–1898, 1997. 6

[23] L. D. Brown, T. T. Cai, and A. Dasgupta, “Interval estimation for a binomial proportion,” *Statistical Science*, vol. 16, pp. 101–133, 2001. 7

Kind 2 = Unprocessed from iris camera
Kind 7 = Cropped and masked to facilitate compression
FNIR = False Negative Identification Rate
FPIR = False Positive Identification Rate

[24] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, “Sheep, goats, lambs and wolves a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation,” in *INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING*, 1998. 7

[25] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, “Image Coding Using Wavelet Transform,” *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, 1992. 8

[26] “Wellcome digital library.” `http://wellcomelibrary.org/about-us/projects/digitisation/`. 9

[27] G. Quinn and P. Grother, “NIST Internal Report 7830: Performance of Face Recognition Algorithms on Compressed Images.” `http://www.nist.gov/manuscript-publication-search.cfm?pub_id=908515`, 2011. 16, 21

[28] *ISO/IEC 19794-5 - Biometric Data Interchange Formats - Face Image Data*. 2005. 16

[29] “Libpng open source reference library.” `http://www.libpng.org/pub/png/libpng.html`. 28