
IREX I

Performance of Iris Recognition Algorithms on Standard Images

NIST Interagency Report 7629

Supplement One

Patrick Grother

Information Access Division
National Institute of Standards and Technology



January 12, 2010

RELEASE NOTES

- ▷ This document supplements NIST Interagency Report 7629 and is intended to be read in conjunction with it. This document, the parent report and its executive summary are available online at <http://iris.nist.gov/irex>. The parent report establishes the IREX goals, methods, experimental context and caveats for this supplement, and is an indispensable companion to it.
- ▷ The IREX evaluation was conducted in accordance with the IREX API and CONOPS which was assembled over the period October 2007 through September 2008. It was developed in consultation with the public and received comments from seven different organizations. It imported content on the IREX image formats from documents submitted to SC 37 in support of ISO/IEC 19794-6. It is archived at <http://iris.nist.gov/irex>.
- ▷ Throughout this report the submitted iris recognition algorithms are identified by a letter and a numeral of the form Nx . The letter identifies the company or university that submitted the algorithm. The numeral x takes the values 1 and 2 for the algorithms submitted to the original IREX I evaluation, and 3 for those submitted after the main IREX results were published in NIST IR 7629. The use of these codes is intended to conserve space in its many tables. For reference, the letters are associated with the providers' names in a running footnote.
- ▷ A glossary of terms and definitions is given in the parent report.
- ▷ Much of the tabulated content in this report was produced automatically. This involved the use of scripting tools to generate directly typesettable \LaTeX content. This improves timeliness, flexibility and maintainability, and reduces transcription errors.
- ▷ Readers are asked to direct any correspondence regarding this report to the IREX@NIST.GOV.

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.

1. INTRODUCTION

The purpose of this document is twofold:

- ▷ To report on the performance of revised iris recognition algorithms submitted by three of the IREX participants. These are Cambridge University, L1 Identity Solutions, and Neurotechnology. The algorithms are identified here by the codes D3, E3, and J3 respectively. These continue the nomenclature adopted in the parent document, NIST Interagency Report 7629. The original IREX I study allowed each of the ten participants to submit two algorithms, denoted 1 and 2. In August 2009, NIST invited revised implementations to be submitted for evaluation. These are identified here by the numeral 3.

Methods and results are given in section 2.

Template sizes are reported in section 3.

- ▷ To extend the analysis on the effect of compression on iris recognition accuracy, and to thereby guide practitioners on the attainable filesizes for storage of standardized iris images. While NIST IR 7629 reported the accuracy effects of JPEG and JPEG2000 compression after images were compressed to 2000 and 3000 bytes, it did not give the functional dependencies FNMR (C, t, K) and FMR (C, t, K) for compression parameter C , threshold t and image kind K . This supplemental adds this information for SDKs D3 and E3. These algorithms represent recent implementations from the most prolific suppliers of recognition algorithms. The analysis has not been repeated for the January 2009 implementations documented in the main IREX report because of computational expense.

Methods and results are given in section 5.

2. ACCURACY

Figures 1, 2 and 3 give DET characteristics for uncropped and uncompressed KIND 1 images of the OPS, ICE, and BATH datasets respectively. These add the accuracy results for the D3, E3 and J3 SDKs to the DETs that appear in NIST Interagency Report 7629. Thus the graphs give performance for 22 iris recognition algorithms operating on generic iris images.

The notable results, and discussion thereof, are as follows:

- ▷ Accuracy of the E3 and D3 implementations is markedly higher than prior algorithms from these providers. At a fixed FMR of 0.00001 on OPS images, the E3 implementation gives FNMR = 0.0028 versus 0.0081 for the earlier E1 SDK. This approximately factor-of-three reduction in FNMR is also achieved for the ICE dataset (0.0341 \rightarrow 0.0092). The corresponding figures for the D2 to D3 evolution are 0.0157 \rightarrow 0.0048 for the OPS dataset and 0.0286 \rightarrow 0.0228 for ICE.
- ▷ These improvements are achieved by unspecified algorithmic changes. One notable observation, however, is that the incidence of template-generation failures are very much reduced here versus in the original IREX trial. Table 1 shows that the D3 and E3 template-generation failure rates are massively reduced. On ICE images, the E1 and E2 SDKs had failed to produce templates on 1.2% of the ICE images. These rates are essentially zero for E3. The IREX report (see Figure 8 there) identified a tendency of the E1 and E2, and especially the D1 and D2 SDKs to give false rejections on large iris radius images. This aspect was identified by Cambridge as one that could be easily fixed, and this assertion was key in initiating NIST's invitation for submission of revised SDKs.

SDK	OPS			ICE			BATH		
	k1	k3	k7	k1	k3	k7	k1	k3	k7
D3	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
	15	0	0	112	0	0	0	0	0
	32640	32640	32640	56666	56666	56666	23035	23035	23035
E3	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.016	0.017
	0	4	4	0	3	3	171	366	388
	32640	32640	32640	56666	56666	56666	23035	23035	23035
J3	0.001	0.000	0.000	0.002	0.000	0.000	0.016	0.020	0.019
	28	13	10	128	19	18	364	469	448
	32640	32640	32640	56666	56666	56666	23035	23035	23035

Table 1: The numbers and proportions of failed template and IREX record creation attempts. In each cell there are three quantities: First is the fraction of templates missing, second is the number of templates missing; third is the total number of images used. A template may be missing if the SDK failed to convert a raw image into an IREX record, or if it failed to convert an IREX record into a template. All rates refer to native operation, and are aggregated over the enrolment and verification image sets. **Colors:** Cells are shaded dark red when the topline FTE is above 1.0%, light red above 0.5%, light green below 0.1% and dark green with exactly zero errors.

- ▷ Significantly the accuracy improvements have been realized without sacrificing speed. Indeed the implementations are generally as fast, or faster, on the same hardware and image datasets as their precursors.
- ▷ The J3 implementation did not offer improved accuracy.

Discussion: The fact that iris recognition accuracy can be improved on the basis of the quantitative testing results given in NIST Interagency Report 7629 is evidence that tuning of algorithms to the intended image corpus can be leveraged operationally. Simultaneously, it is evidence that tuning is sometimes necessary. In a perfectly interoperable world, tuning with prior knowledge of the image properties would not be necessary. The gap between current reality and global interoperability is the gap that NIST 's IREX program and the ISO/IEC 19794-6 and 29794-6 iris image interchange and quality standards are designed to fill.

3. TEMPLATE SIZES

For the D3 implementation all templates were of size 1024 bytes.

For the E3 implementation all enrollment and verification templates were of sizes 1280 and 4320 bytes respectively.

For the J3 implementation all templates were of size 2328 bytes.

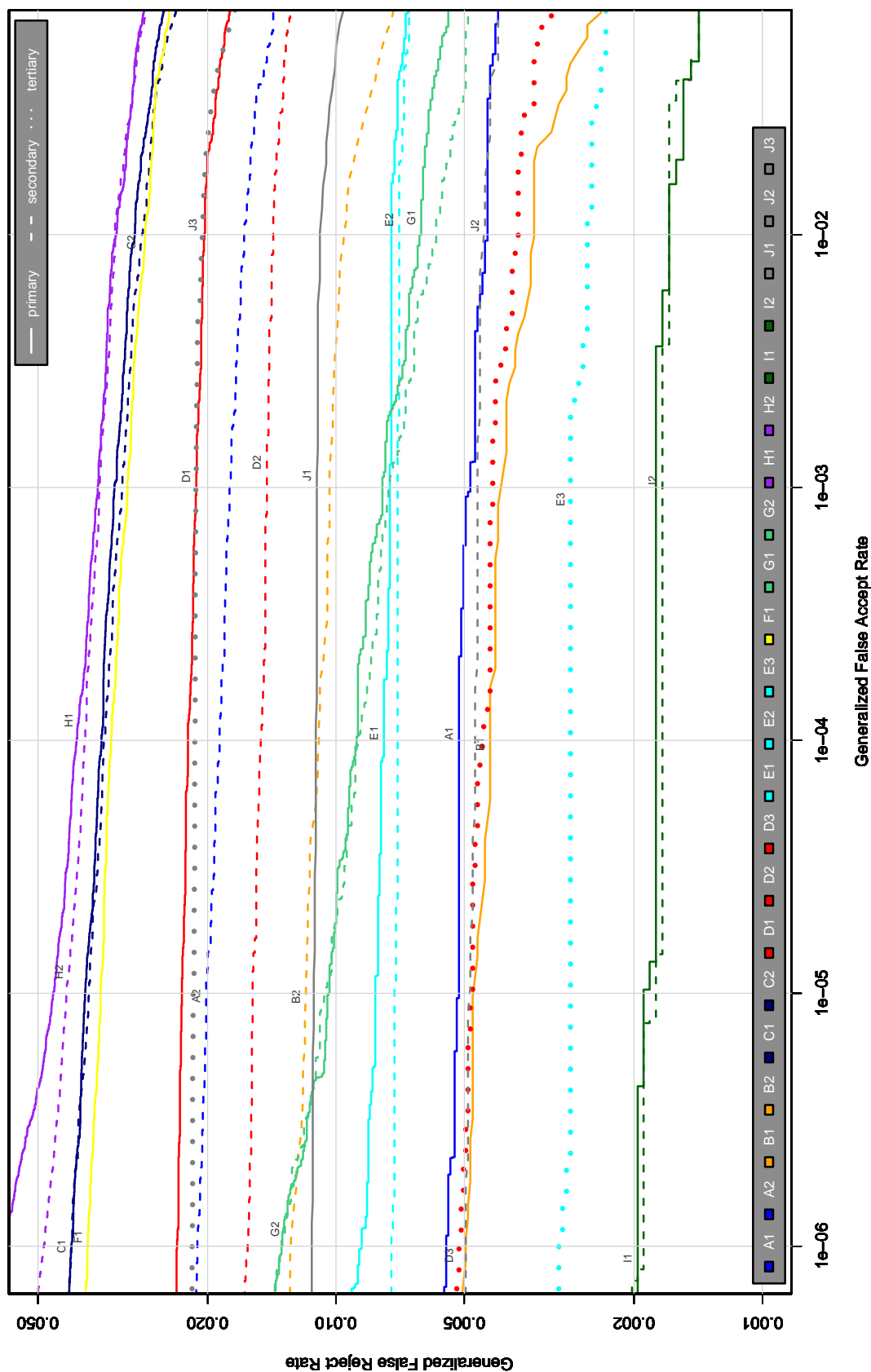


Figure 1: For each SDK the figure shows the DET for comparison of uncompressed KIND 1 images from the OPS dataset. The accuracy of the new (tertiary) SDKs are shown with dotted lines. The figures plot generalized false rejection and acceptance rates. As discussed in section 6.1 of the main IREX report, these include the effect of failed template creations, and are equivalent to FNMR and FMR when all templates are produced.

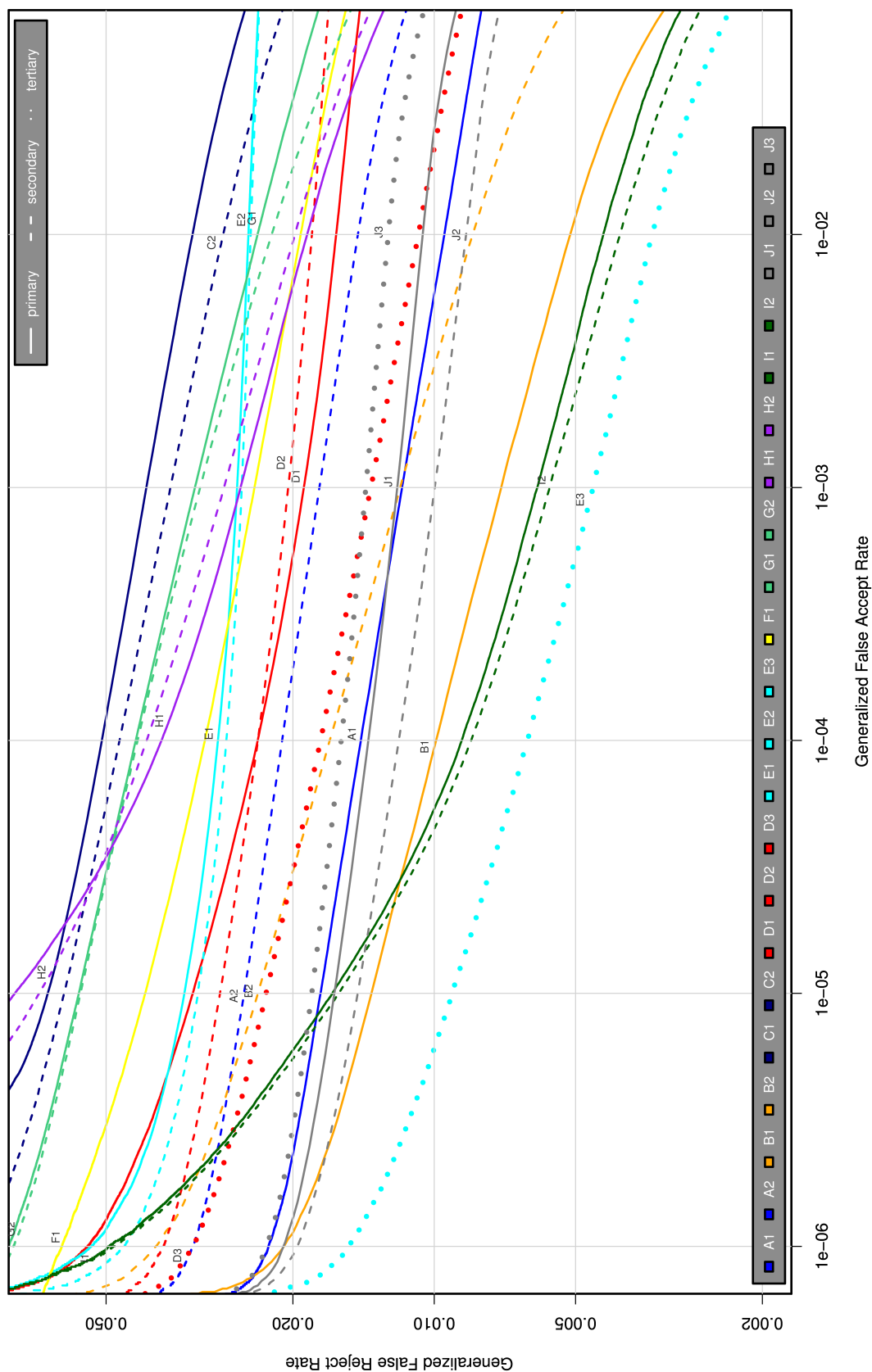


Figure 2: For each SDK the figure shows the DET for comparison of uncompressed KIND 1 images from the ICE dataset. The accuracy of the new (tertiary) SDKs are shown with dotted lines. The figures plot generalized false rejection and acceptance rates. As discussed in section 6.1 of the main IREX report, these include the effect of failed template creations, and are equivalent to FNMR and FMR when all templates are produced.

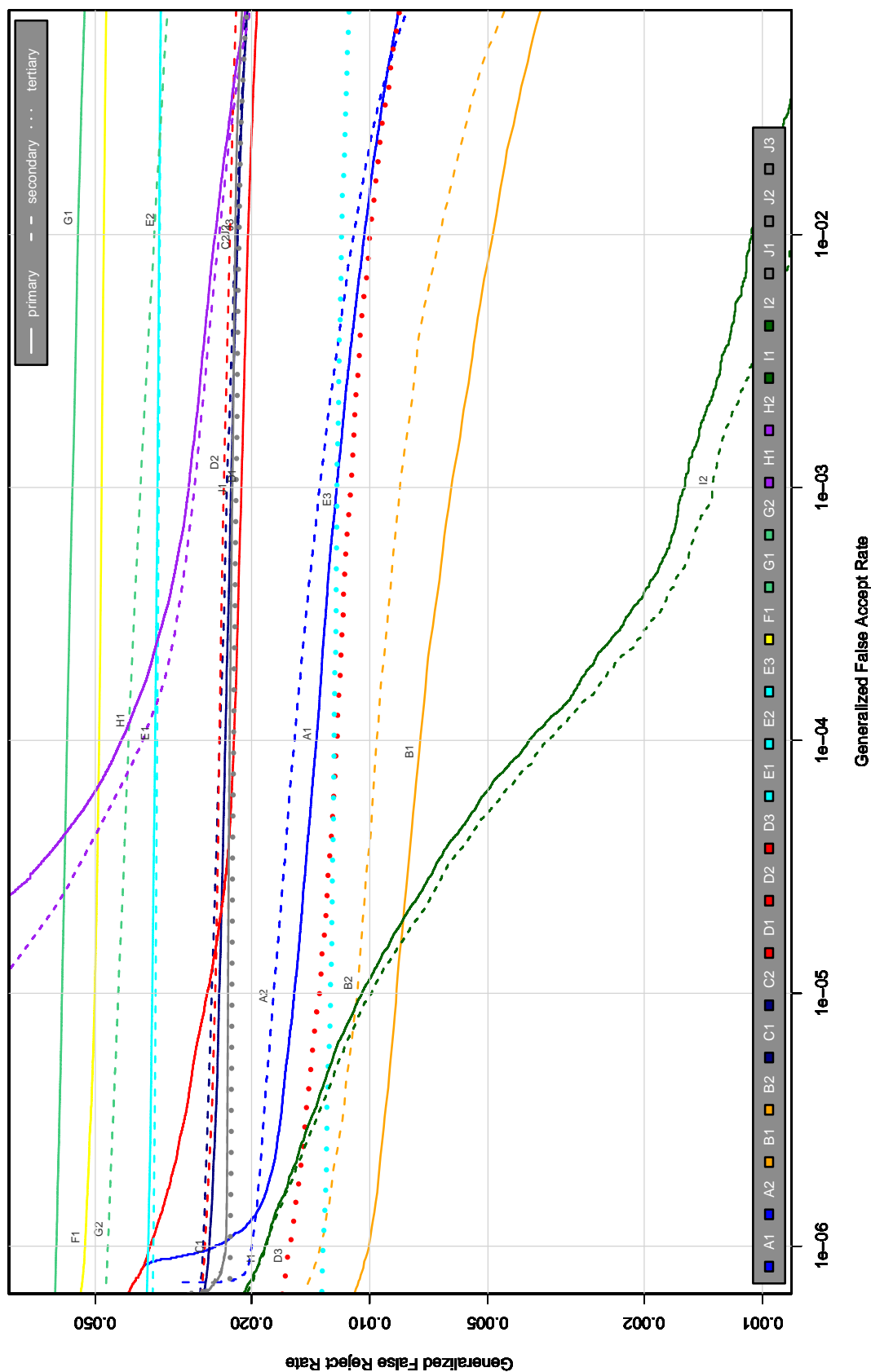


Figure 3: For each SDK the figure shows the DET for comparison of uncompressed KIND 1 images from the BATH dataset. The accuracy of the new (tertiary) SDKs are shown with dotted lines. The figures plot generalized false rejection and acceptance rates. As discussed in section 6.1 of the main IREX report, these include the effect of failed template creations, and are equivalent to FNMR and FMR when all templates are produced.

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY
KIND 1 = RAW 640x480	KIND 3 = CROP	KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR	

4. SPEED VS. ACCURACY TRADEOFFS

Given the wide ranges in processing times observed in the IREX report, this section includes the three new implementations in the speed-accuracy tradeoff analysis originally given in section 7.6 of the main IREX report, NIST Interagency Report 7629 (linked from <http://iris.nist.gov/irex>).

The four plots of Figure 4 extend the first report's analysis of speed-accuracy tradeoffs by including the three new algorithms in the tradespace plots. These plot a fixed statement of accuracy on the y-axis against the duration of the elemental processing operations: IREX record preparation (i.e. iris detection); template generation; matching; and a composite 1:N identification application. Within each plot, each point corresponds to one SDK. In each case the accuracy is quantified by FNMR at a fixed global threshold giving FMR = 0.00001 computed over uncompressed and uncropped ICE and OPS images. The y-axis false rejection rate value is the simple average of those observed over the OPS and ICE corpora.

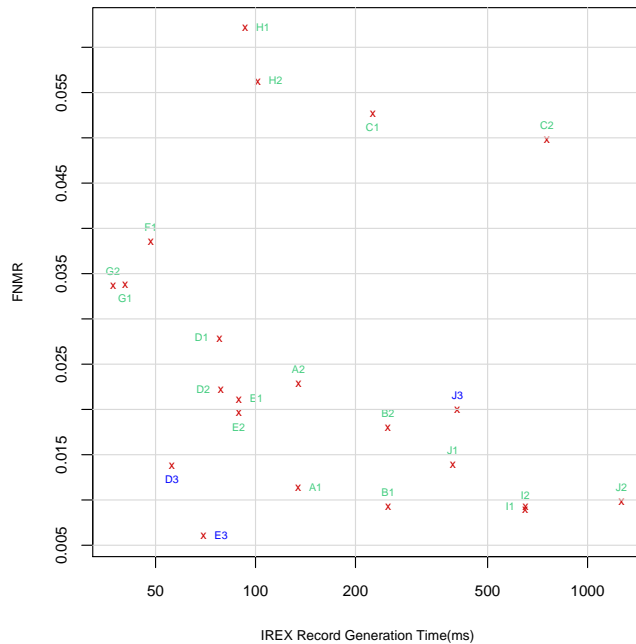
$$0.5(\text{FNMR}_{\text{ICE}}(\tau_{\text{ICE}}) + \text{FNMR}_{\text{OPS}}(\tau_{\text{OPS}})) \quad (1)$$

where the thresholds are set independently to give FMR = 0.00001. This method deviates from the main IREX report which used the simple average of FNMR over the OPS and ICE datasets at the threshold that gave FMR = 0.00001 on the combined dataset.

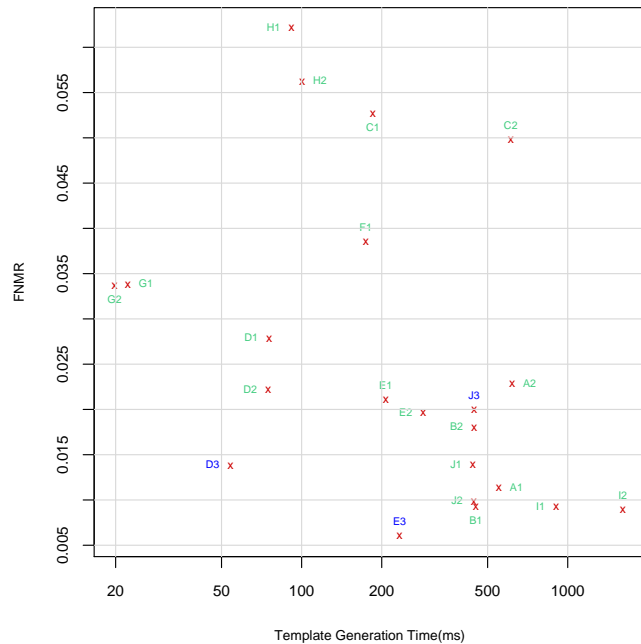
The notable observations are:

- ▷ The D3 SDK is faster in image processing than either of the D1 and D2 Cambridge precursors. The matching times are essentially unchanged.
- ▷ The E3 SDK is faster at the initial localization than the earlier E1 entry from L1. While the template generation time is slightly slower, the E3 implementation is considerably faster for matching.
- ▷ The J3 SDK exhibits closely the same speed as its earlier siblings J1 and J2.
- ▷ Both D3 and E3 give improved recognition accuracy without incurring a loss of computational efficiency.

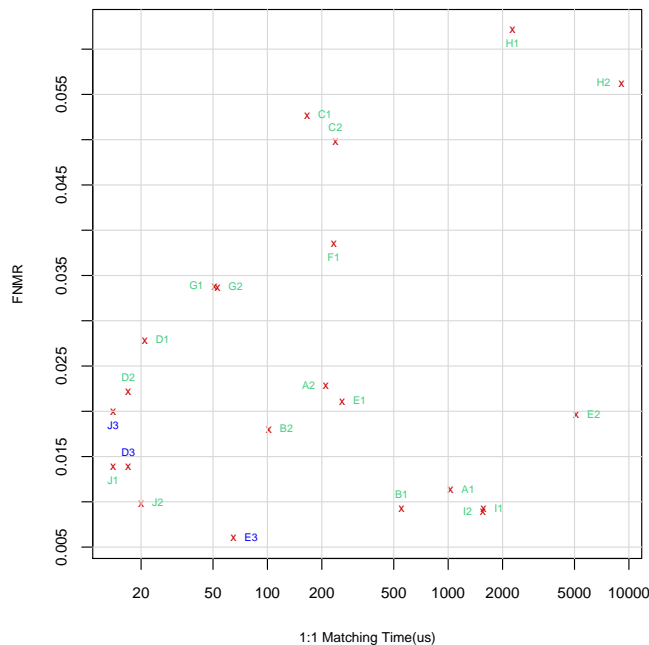
The D3 SDK ran under the linux operating system. The E3 and J3 implementations ran under Windows. The computers used have somewhat different hardware specifications (see NISTIR 7629) - the Windows machines are slower than the Linux machines. The times reported here, and in the original report, have *not* been adjusted to account for this. Users should understand that exact timing comparisons are difficult without an appropriate correction. Operationally computational time is strongly dependent on the architecture, input-output performance, the degree of tuning and optimization, and on other factors.



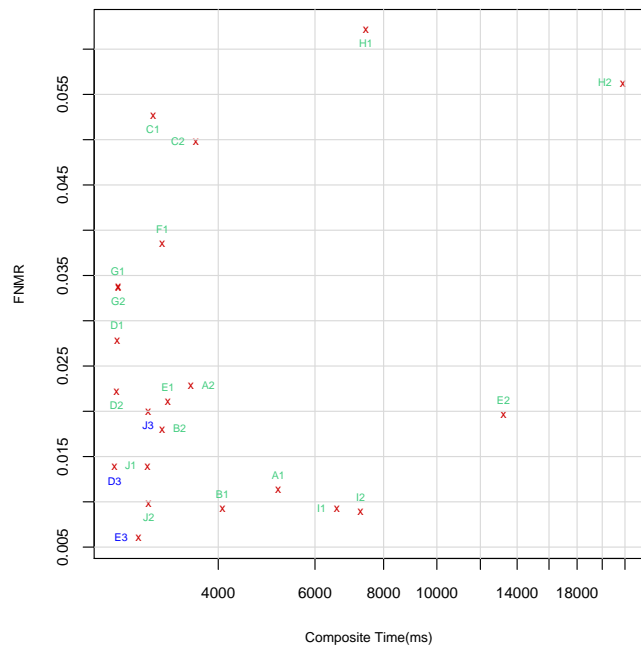
(a) FNMR vs. Time to make IREX record (mean KINDS 3, 7)



(b) FNMR vs. Time to make template from KIND 1 image



(c) FNMR vs. Time to compare templates



(d) FNMR vs. Composite 1:2000 search time

Figure 4: Duration of the IREX functions versus FNMR . At **top left** is the IREX record creation time measured in milliseconds; at **top right** is the template generation time also in milliseconds; at **bottom left** is the one-to-one comparison time in microseconds (μs); and finally at **bottom right** is the time for the hypothetical access control system documented in the IREX report. All times are on log-scales. The y-axis values are $0.5(FNMR_{ICE}(\tau_{ICE}) + FNMR_{OPS}(\tau_{OPS}))$ when the thresholds are set independently to give $FMR = 0.00001$. This deviates from the main IREX report which used the simple average of FNMR over the OPS and ICE datasets at the threshold that gave $FMR = 0.00001$ on the combined dataset.

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY
KIND 1 = RAW 640x480	KIND 3 = CROP	KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR	

5. EFFECT OF COMPRESSION

This section extends the analyses given in the main IREX report¹ by including full DET plots for many levels of JPEG2000 compression. The figures 8 through 15 show how both the false match and false non-match error rates change as the target file size is reduced. This is done for various combinations of the OPS, ICE, and BATH, datasets, and the dedicated compression formats, crop-only KIND 3 and crop-and-mask KIND 7.

A summary graph, Figure 5 shows the false rejection rates vs. compressed file size at a fixed false match rate of 0.0001.

This analysis is reported for D3 and E3. These implementations are believed to embed the core iris recognition algorithms from Cambridge University[1]. The results should be widely applicable because the the Cambridge algorithms are widely used. Note, however, that users will usually benefit from application of this kind of analysis for newer algorithms operating on there specific datasets.

The notable results are below.

- ▷ As stated in the main IREX report, the application of lossy compression algorithms to iris images causes incremental damage to the iris texture. The result of this is that the false rejection rate usually increases smoothly and monotonically with the amount of compression. Figures 6 and 7 show this dependency for the ICE dataset for SDKs D3 and E3 respectively.
- ▷ The smooth dependency is not observed for the OPS set. Here the error rates are absolutely low such that the number of false rejection errors is measured in tens. The effect of compression is to cause step changes in error rates corresponding to just a single image comparison (i.e. 1 out of 16320). Ultimately, below 2000 bytes, compression does elevate error rates. Inspection of the raw OPS images does shows no evidence that they have been previously compressed (e.g. by the camera or client).
- ▷ As shown in Figure 10, the functional dependency for the D3 SDK running on the BATH dataset is similar to that for ICE. For E3 (Figure 11) the situation is more complicated. The false non-match rates are higher than for D3, but the DET is flatter than that for D3. As shown in Table 1, the E3 SDK failed or elected not to produce templates from a larger fraction of BATH images than D3. The different forms of the DETs may be a result of different failure-to-enroll strategies.
- ▷ Regarding SDKs D3 and E3, while their absolute error rates differ, the functional form of the error rate dependency for the D3 and E3 SDKs is the same.
- ▷ The crop-only format, KIND 3, suffers elevated error rates at lighter compression rates than the crop-and-masked KIND 7 format. This arises because for KIND 3 images the compressor must dedicate many bits to the compression of the eyelids, eyelashes and periphery. The main IREX report recommended that KIND 7 is preferred to KIND 3 images for small target file sizes.

REFERENCES

- [1] J. Daugman. The importance of being random. *Pattern Recognition*, 36(2):279–291, February 2003.

¹NIST Interagency Report 7629 linked from <http://iris.nist.gov/irex>

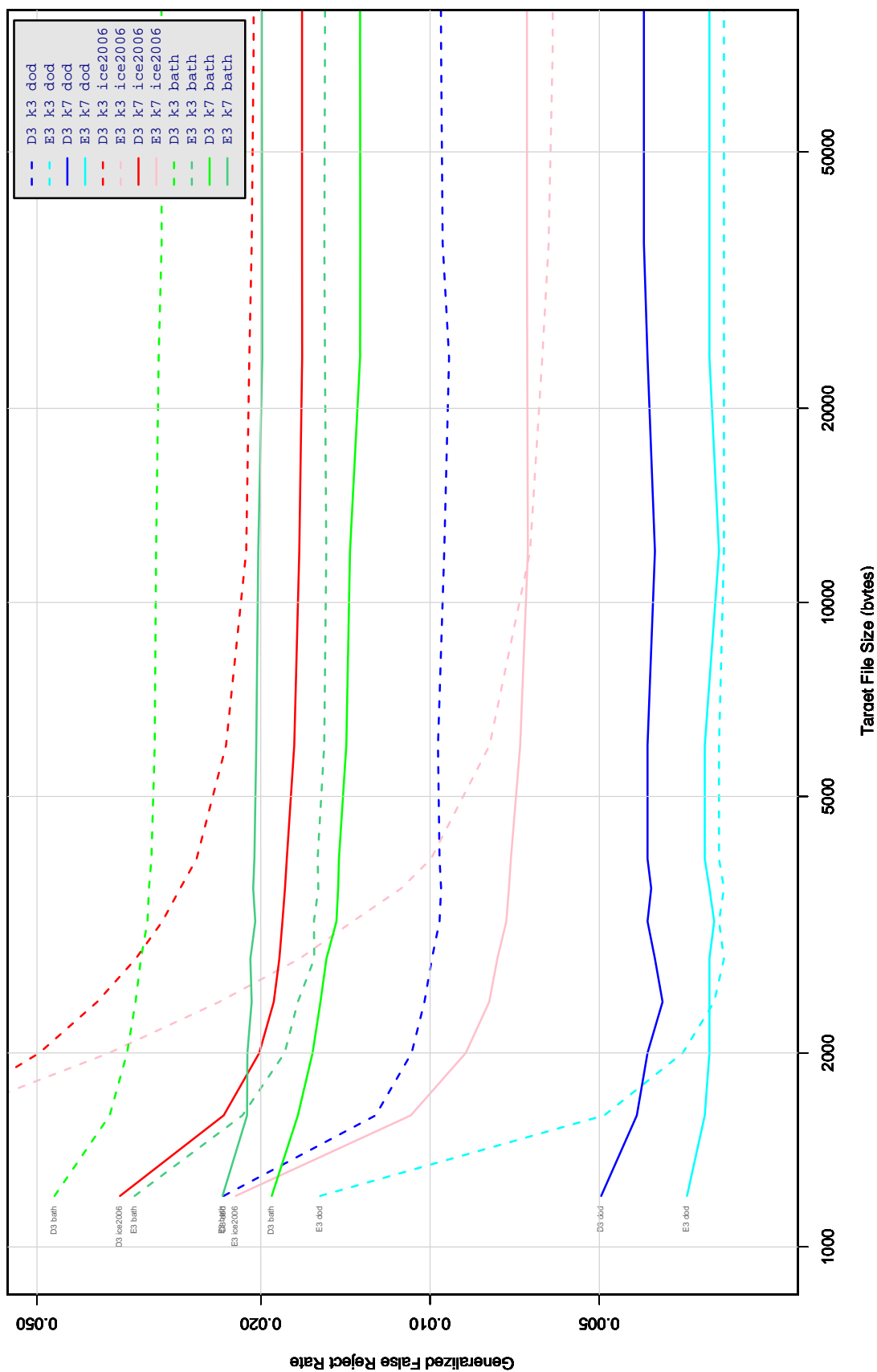


Figure 5: The variation of generalized false reject rate with target compressed file size for the D3 and E3 SDKs applied to the three IREX image sets in KIND 3 and KIND 7 formats. Accuracy is stated at a fixed generalized false accept rate of 0.0001.

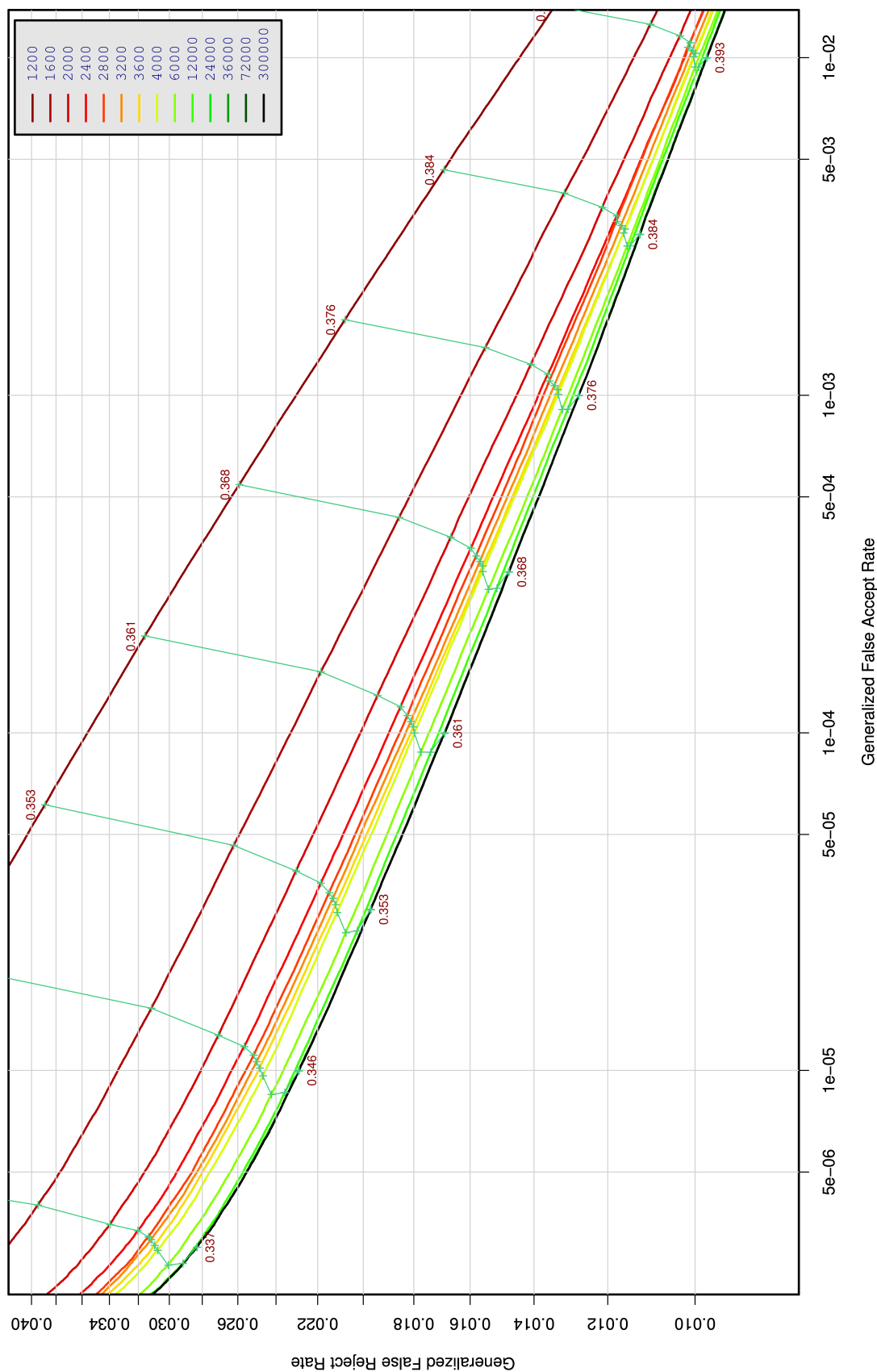


Figure 6: For the D3 SDK the figure shows DETs for comparison of JPEG2000-compressed KIND 7 images against uncompressed KIND 1 images of the ICE dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FNMR on compression, and deviation from vertical shows the dependence of FMR.

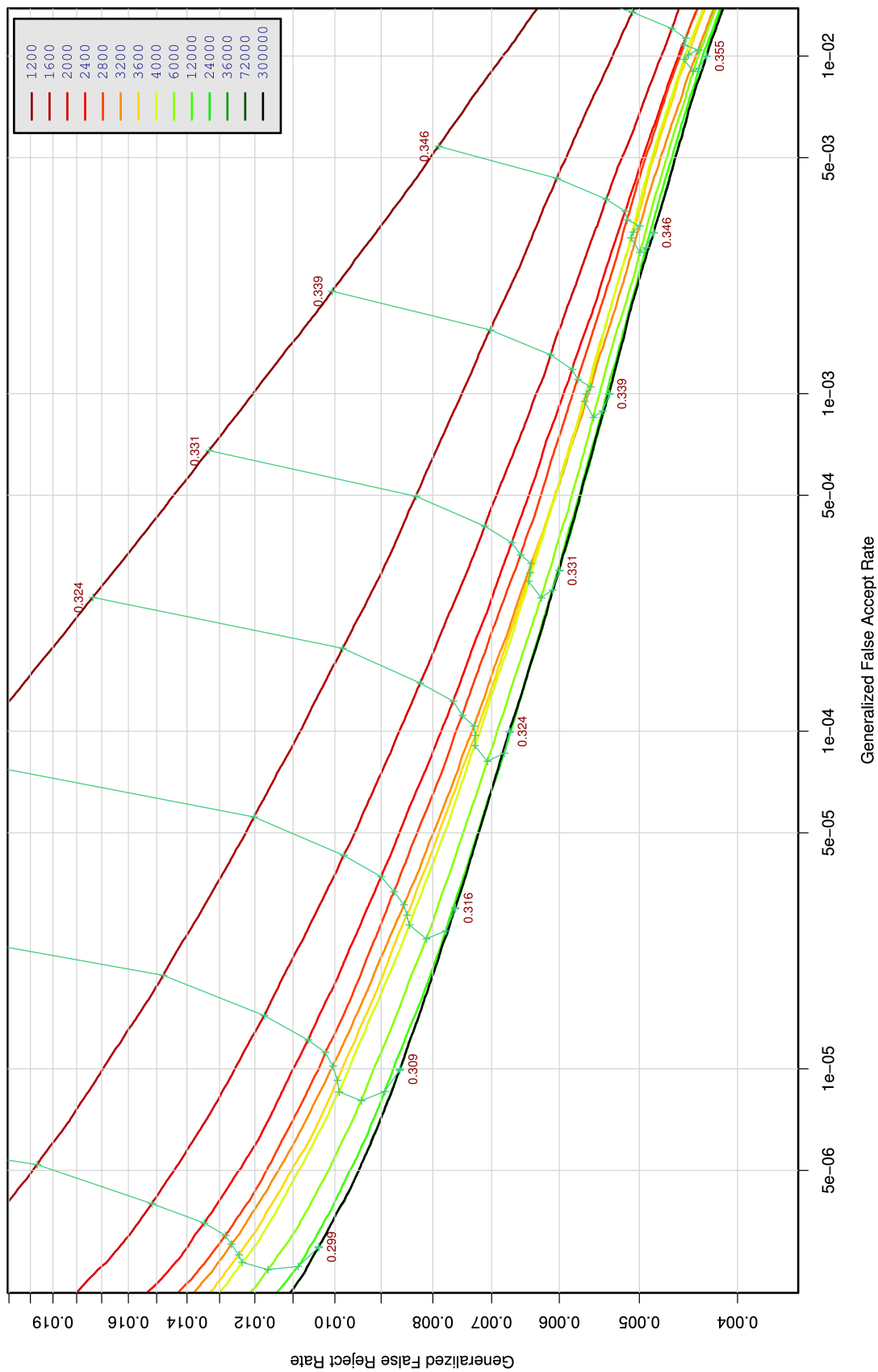


Figure 7: For the E3 SDK the figure shows DETs for comparison of JPEG2000-compressed KIND 7 images against uncompressed KIND 1 images of the ICE dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FNMR on compression, and deviation from vertical shows the dependence of FMR.

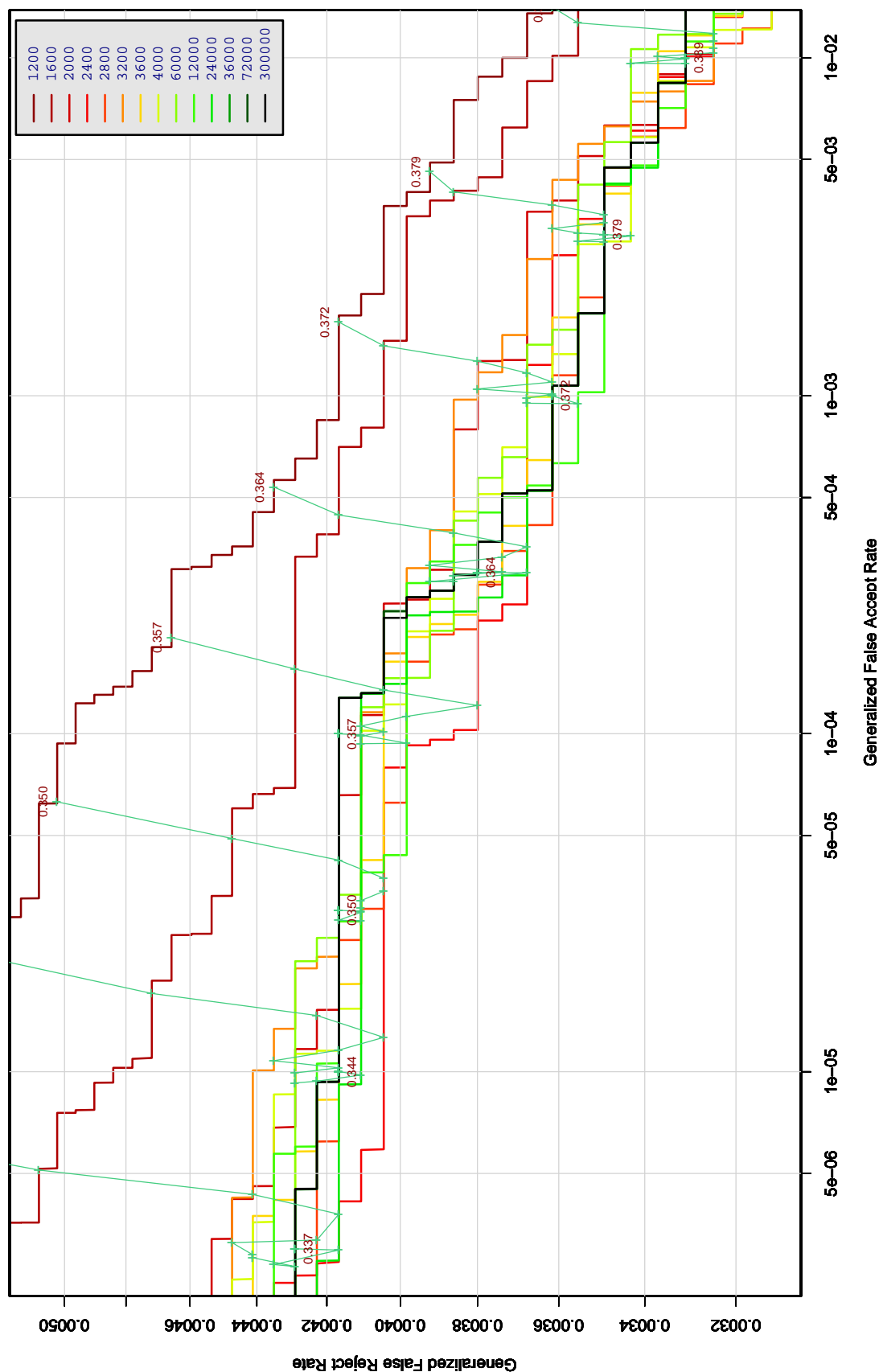


Figure 8: For the D3 SDK the figure shows DETs for comparison of JPEG2000-compressed KIND 7 images against uncompressed KIND 1 images of the OPS dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FMR on compression, and deviation from vertical shows the dependence of FMR.

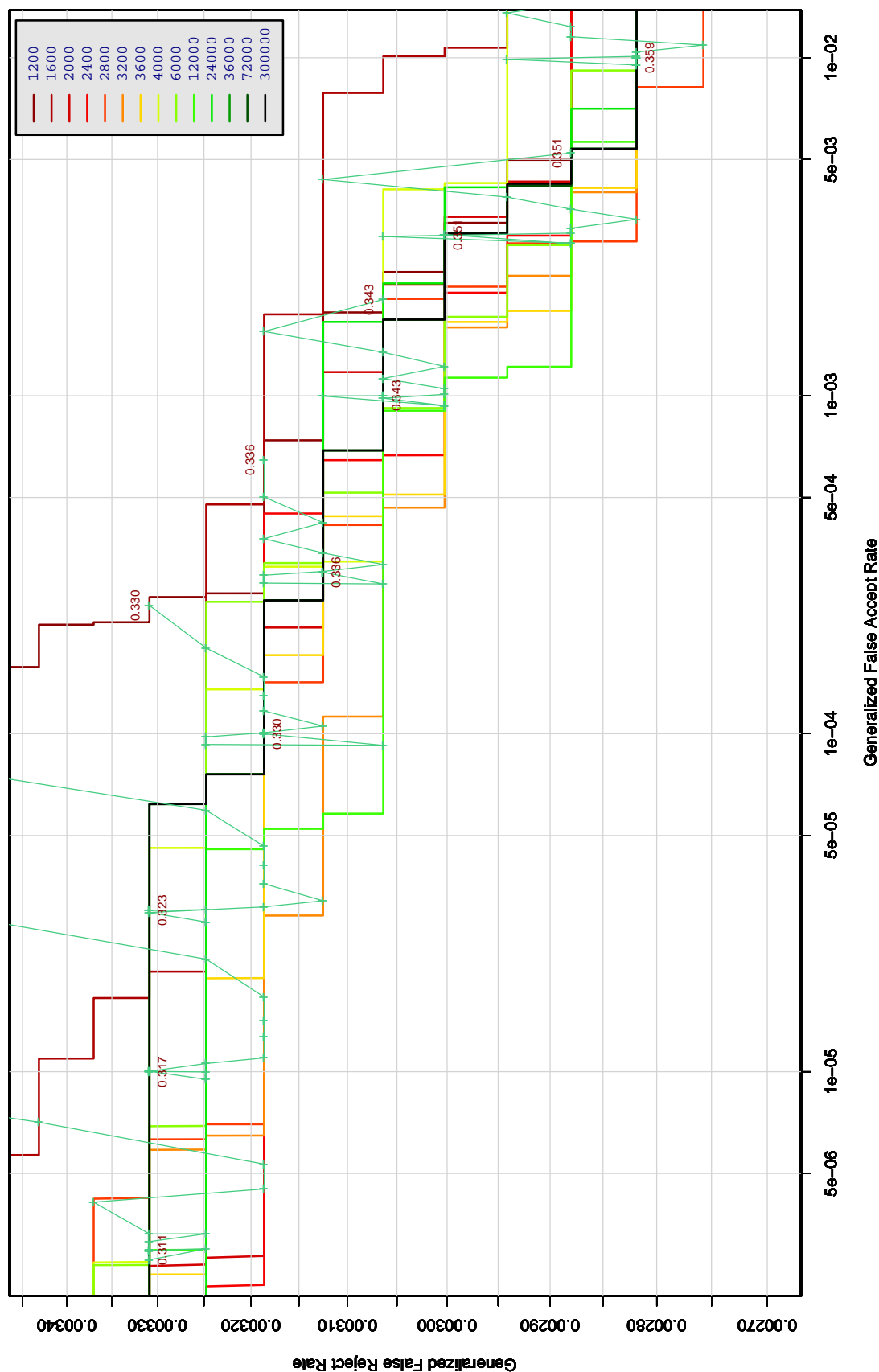


Figure 9: For the E3 SDK the figure shows DETs for comparison of JPEG2000 -compressed KIND 7 images against uncompressed KIND 1 images of the OPS dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FNMR on compression, and deviation from vertical shows the dependence of FMR.

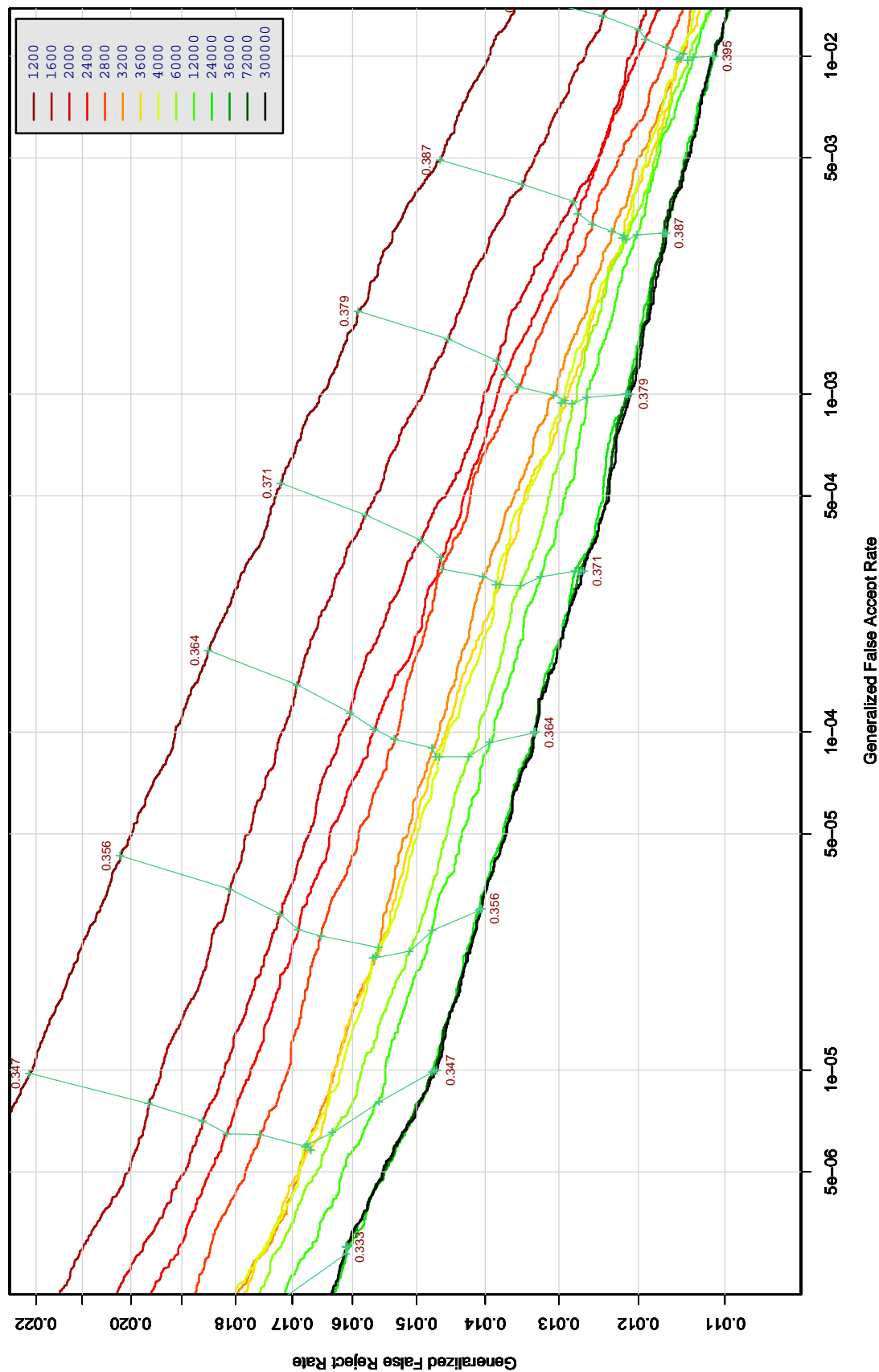


Figure 10: For the D3 SDK the figure shows DETs for comparison of JPEG2000-compressed KIND 7 images against uncompressed KIND 1 images of the BATH dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FNMR on compression, and deviation from vertical shows the dependence of FMR.

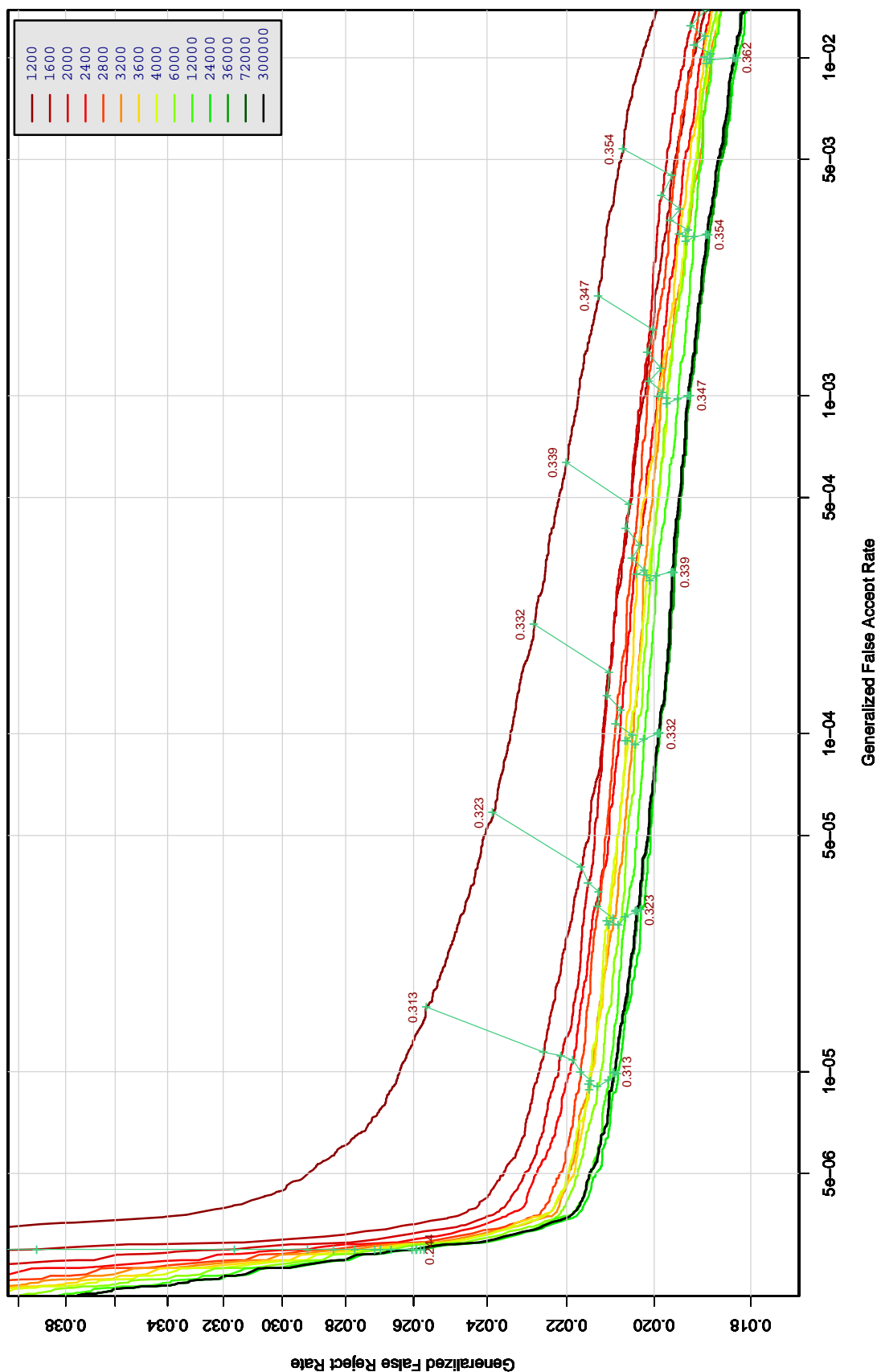


Figure 11: For the E3 SDK the figure shows DETs for comparison of JPEG2000 -compressed KIND 7 images against uncompressed KIND 1 images of the BATH dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FMR on compression, and deviation from vertical shows the dependence of FMR.

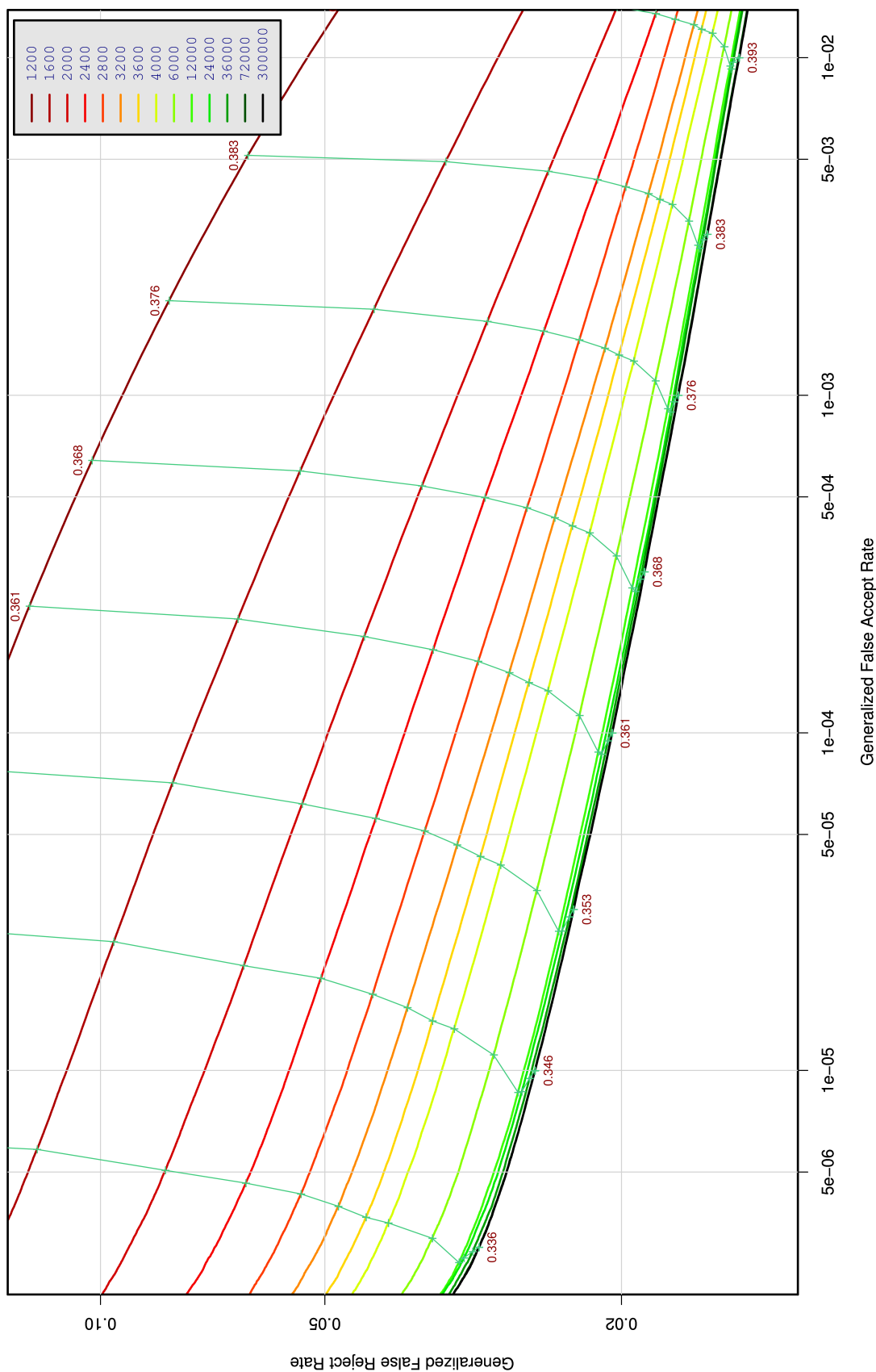


Figure 12: For the D3 SDK the figure shows DETs for comparison of JPEG2000 -compressed KIND 3 images against uncompressed KIND 1 images of the ICE dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FNR on compression, and deviation from vertical shows the dependence of FMR.

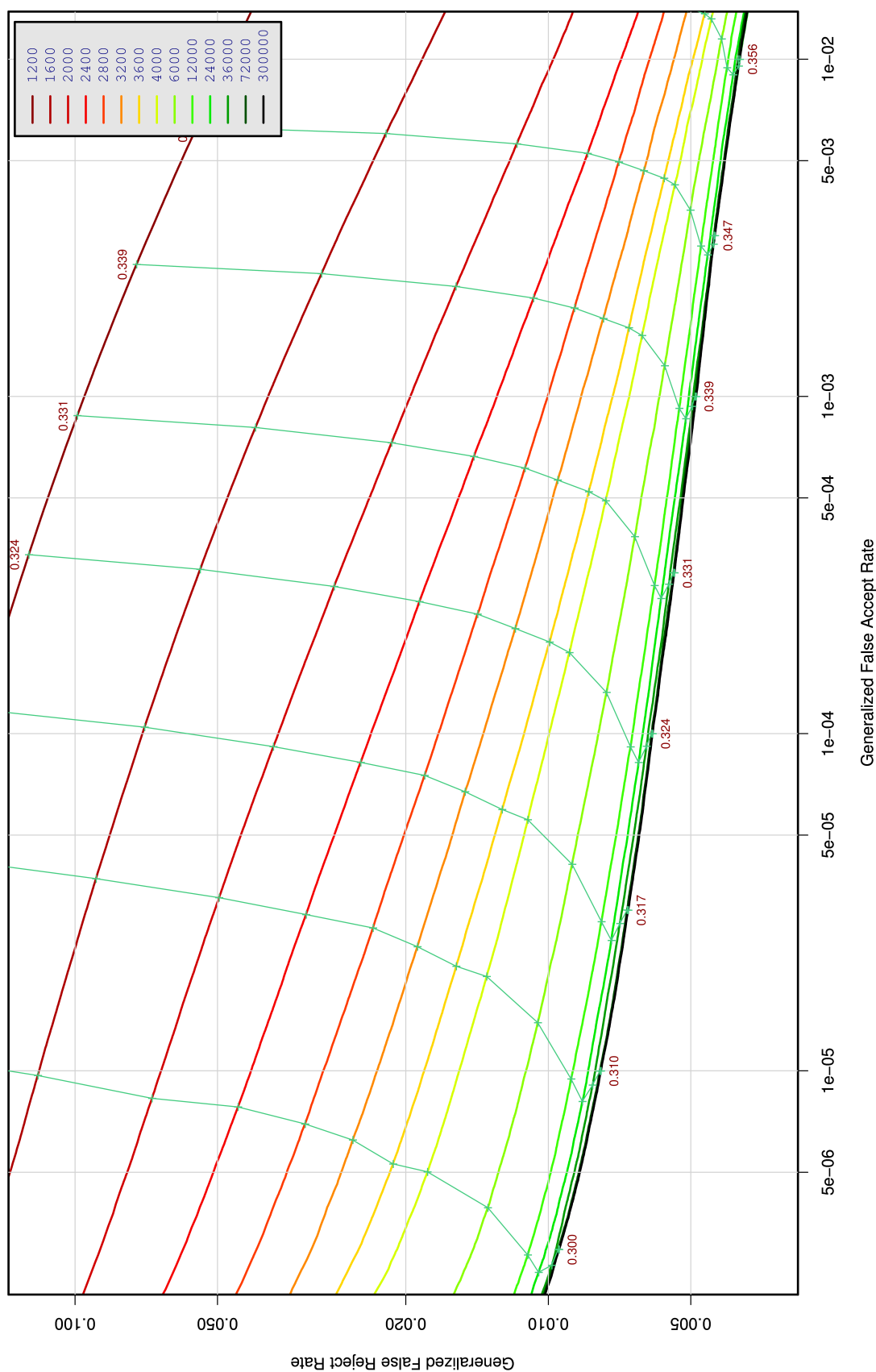


Figure 13: For the E3 SDK the figure shows DETs for comparison of JPEG2000-compressed KIND 3 images against uncompressed KIND 1 images of the ICE dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FNMR on compression, and deviation from vertical shows the dependence of FMR.

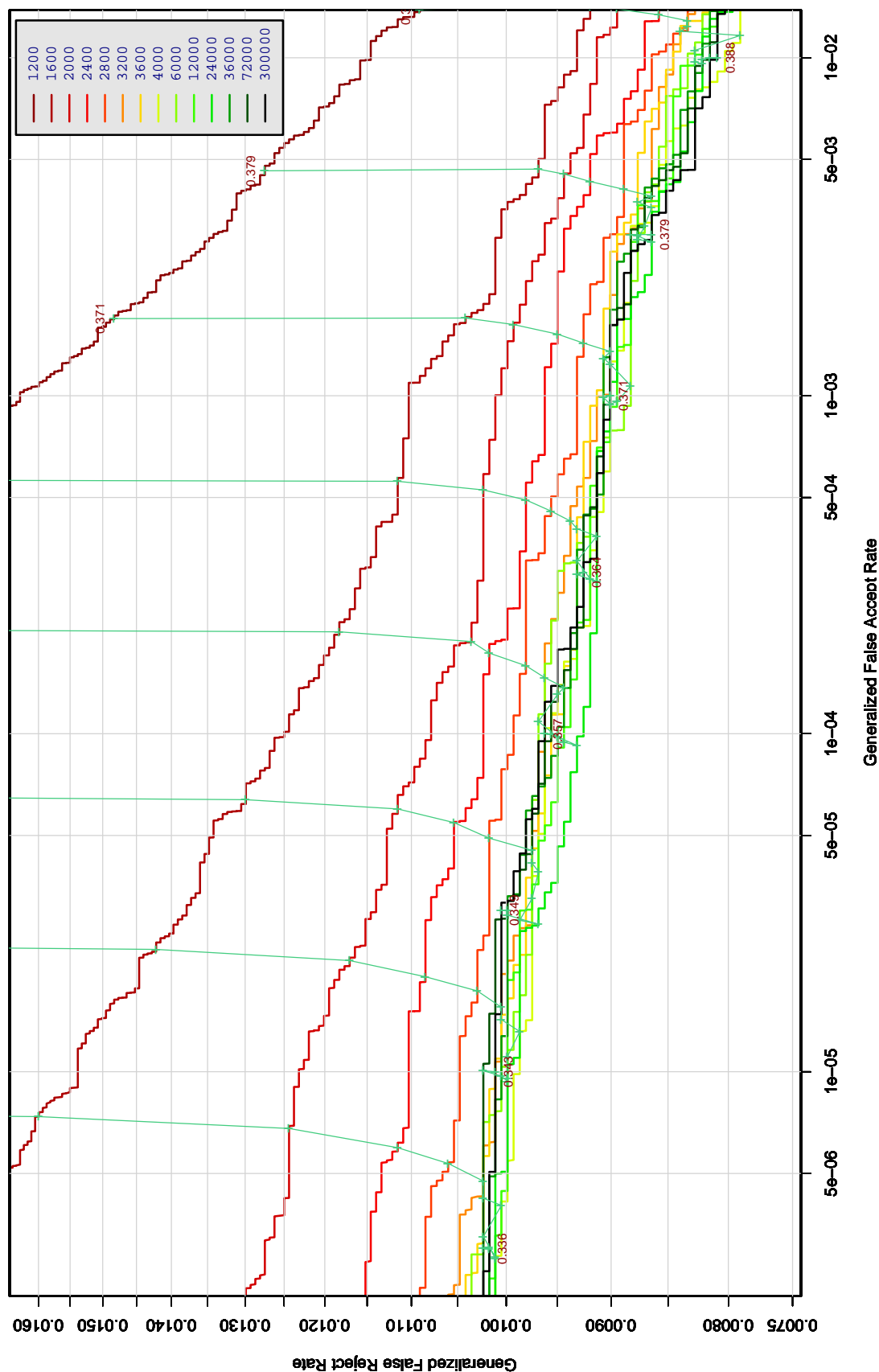


Figure 14: For the D3 SDK the figure shows DETs for comparison of JPEG2000 -compressed KIND 3 images against uncompressed KIND 1 images of the OPS dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FNMR on compression, and deviation from vertical shows the dependence of FMR.

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY
KIND 1 = RAW 640x480	KIND 3 = CROP	KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR	

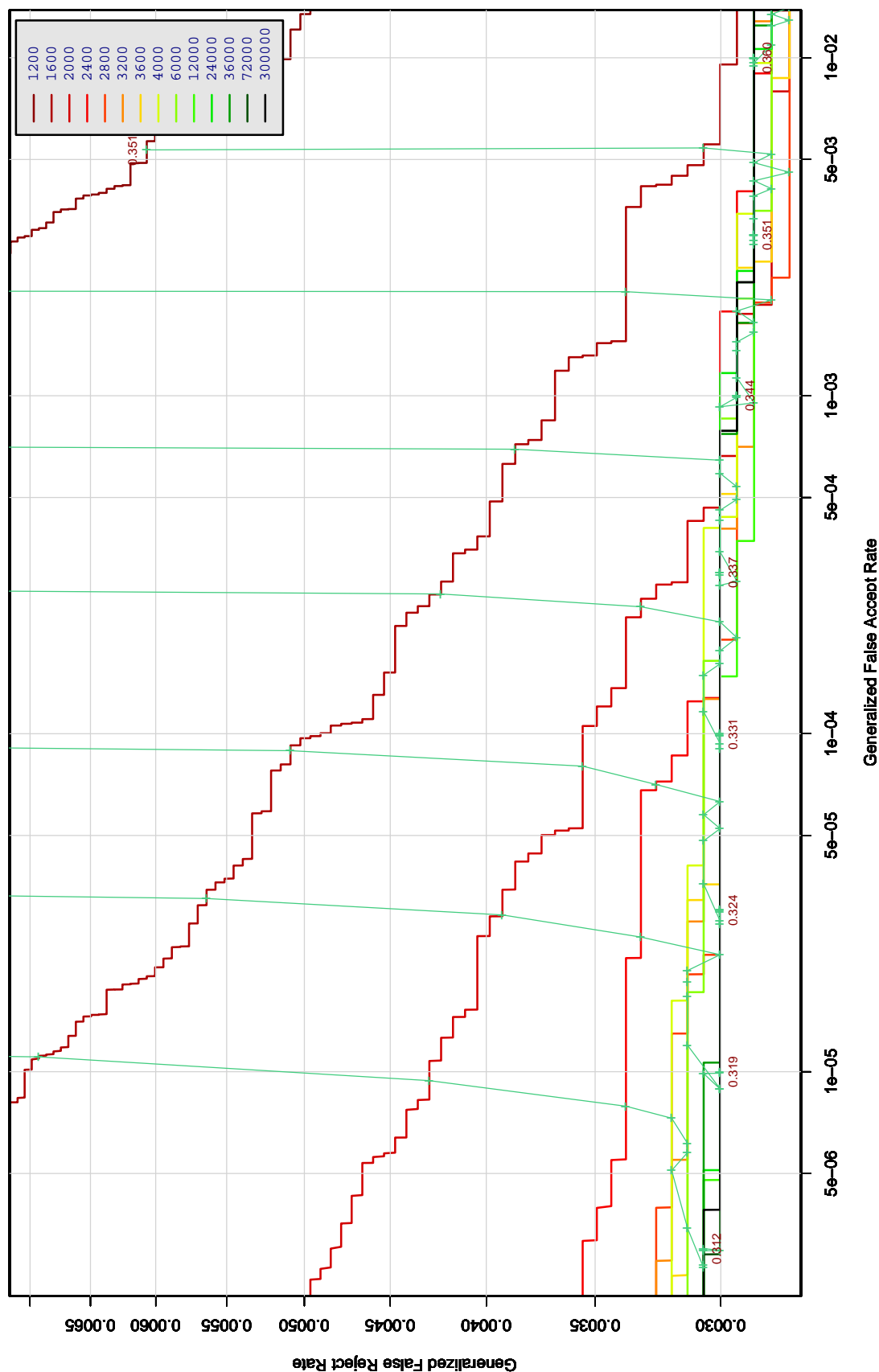


Figure 15: For the E3 SDK the figure shows DETs for comparison of JPEG2000-compressed KIND 3 images against uncompressed KIND 1 images of the OPS dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FMR on compression, and deviation from vertical shows the dependence of FMR.

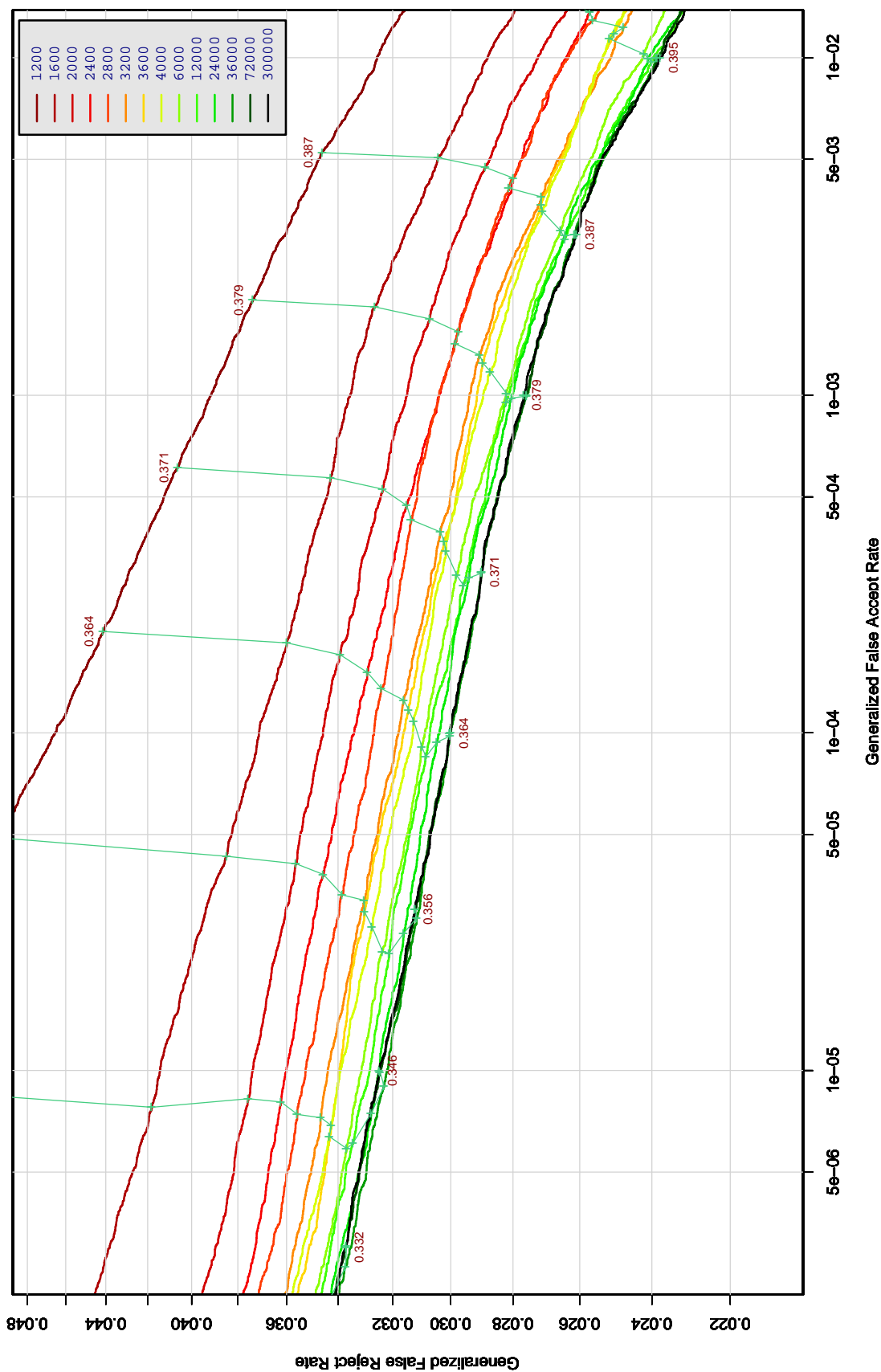


Figure 16: For the D3 SDK the figure shows DETs for comparison of JPEG2000-compressed KIND 3 images against uncompressed KIND 1 images of the BATH dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FMR on compression, and deviation from vertical shows the dependence of FMR.

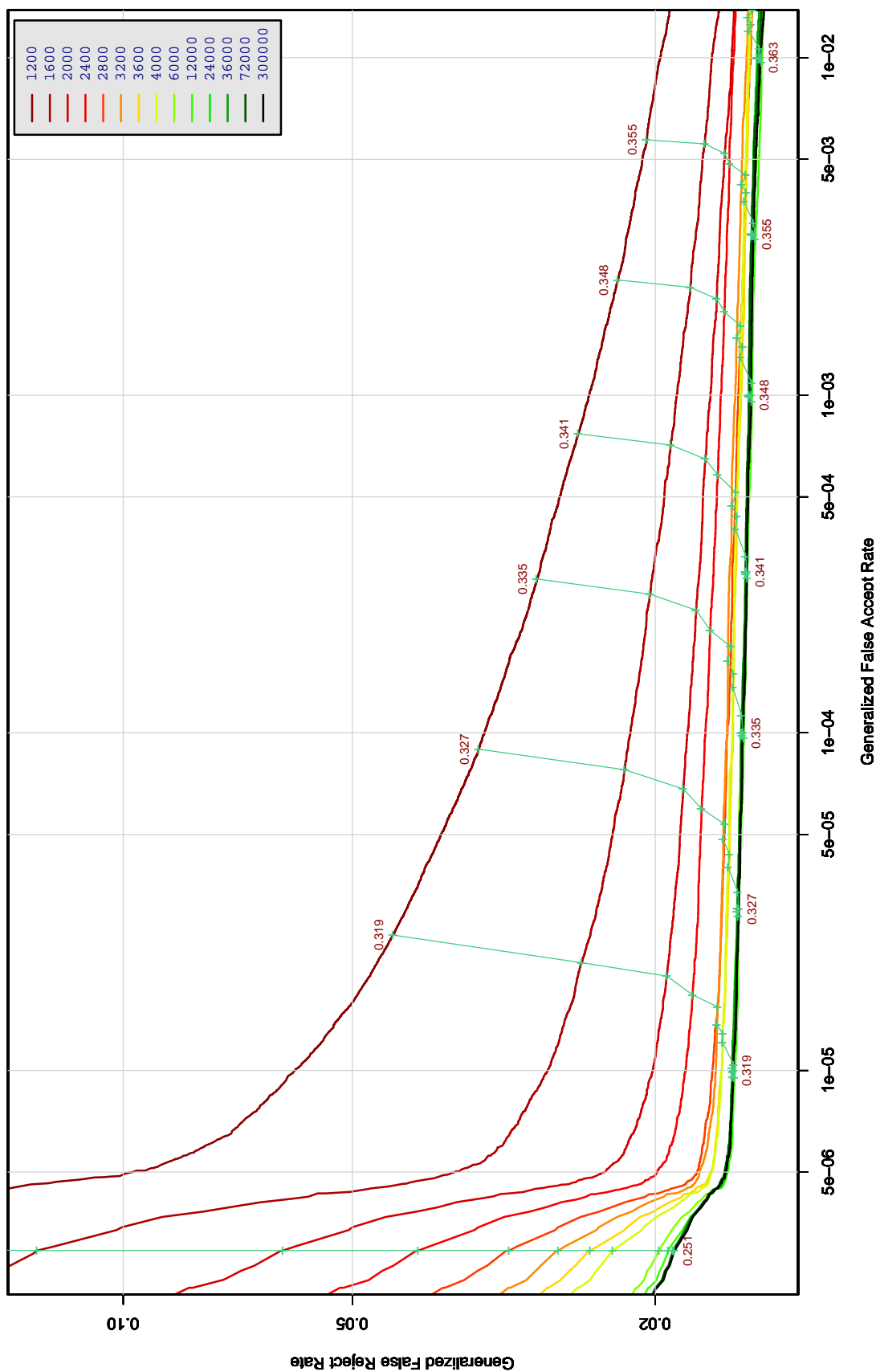


Figure 17: For the E3 SDK the figure shows DETs for comparison of JPEG2000 -compressed KIND 3 images against uncompressed KIND 1 images of the BATH dataset. Each DET corresponds to a particular target file size, in bytes. The green lines connect points of equal operating threshold - the deviation from horizontal shows the dependence of FNMR on compression, and deviation from vertical shows the dependence of FMR.