



Sensitivity analysis for biometric systems: A methodology based on orthogonal experiment designs[☆]

Yooyoung Lee^{*}, James J. Filliben, Ross J. Micheals, P. Jonathon Phillips

National Institute of Standards and Technology, Information Technology Laboratory, 100 Bureau Drive, Gaithersburg, MD 20899, USA

ARTICLE INFO

Article history:

Received 22 March 2012

Accepted 9 January 2013

Available online 16 January 2013

Keywords:

Sensitivity analysis

Iris recognition

Biometrics

Experiment design

VASIR

Video-based biometric system

Orthogonal

Fractional factorial

Characterization

Uncertainty

ABSTRACT

The purpose of this paper is to introduce an effective and structured methodology for carrying out a biometric system sensitivity analysis. The goal of sensitivity analysis is to provide the researcher/developer with insight and understanding of the key factors—algorithmic, subject-based, procedural, image quality, environmental, among others—that affect the matching performance of the biometric system under study. This proposed methodology consists of two steps: (1) the design and execution of orthogonal fractional factorial experiment designs which allow the scientist to *efficiently* investigate the effect of a large number of factors—and interactions—simultaneously, and (2) the use of a select set of statistical data analysis graphical procedures which are fine-tuned to unambiguously highlight important factors, important interactions, and locally-optimal settings. We illustrate this methodology by application to a study of VASIR (Video-based Automated System for Iris Recognition)—NIST iris-based biometric system. In particular, we investigated $k = 8$ algorithmic factors from the VASIR system by constructing a $(2^{6-1} \times 3^1 \times 4^1)$ orthogonal fractional factorial design, generating the corresponding performance data, and applying an appropriate set of analysis graphics to determine the relative importance of the eight factors, the relative importance of the 28 two-term interactions, and the local best settings of the eight algorithms. The results showed that VASIR's performance was primarily driven by six factors out of the eight, along with four two-term interactions. A virtue of our two-step methodology is that it is systematic and general, and hence may be applied with equal rigor and effectiveness to other biometric systems, such as fingerprints, face, voice, and DNA.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Biometrics is the automated recognition of individuals based on their biological and behavioral characteristics [1]. The characteristics can include fingerprints, face, iris, ocular area, retina, ear, voice, DNA, signature, gait, and hand geometry among others. The use of biometrics has many advantages—especially as an alternative to keys, passwords, smartcards, and other artifacts for physical entry. In this regard, biometric-based technologies are increasingly being incorporated into specific security fields and applications, such as industrial access control, law-enforcement, military, border control, and forensics [2].

A significant problem in biometric studies is that researchers/developers often present results that lack an assessment of intrinsic system uncertainty. A high degree of input and output numerical precision often gives the impression of great accuracy, but

neglects to give attention to the critical questions of the sensitivity of the final results to different algorithms, environments, subject characteristics, and biometric sample conditions [3]. Hornberger and Spear [4] made the following paraphrased statement about simulation models: Most such models are complex, with many parameters, state-variables and underlying non-linear relations; under optimal circumstances, such systems have many degrees of freedom and—with judicious adjustments—are susceptible to over-fitting with both plausible structure and “reasonable” parameter values. We believe that the above statement applies equally well for biometric systems, especially for iris recognition system.

Sensitivity analysis has been successfully conducted in areas such as computer vision and computer network [5–7]. Sensitivity analysis is the study of how the output of a system is affected by different inputs to the system [8]. In essence, a biometric system is a data monitoring and decision-making “machine.” A good biometric system has a high proportion of correct decisions. All biometric systems are susceptible to incorrect decisions—especially in the presence of less-than-optimal conditions.

In practice, the performance of many biometric systems is frequently examined and optimized via a series of one-factor-at-a-time experiment designs in which most factors in the system are

[☆] This paper has been recommended for acceptance by Rudolf M. Bolle.

^{*} Corresponding author. Tel.: +1 301 975 8487/3234; fax: +1 301 975 5287.

E-mail addresses: yooyoung@nist.gov (Y. Lee), filliben@nist.gov (J.J. Filliben), rossm@nist.gov (R.J. Micheals), jonathon@nist.gov (P. Jonathon Phillips).

held constant while one factor is focused on and varied to examine its effect. This design, though popular [9,10] has the disadvantage that it can yield grossly biased (incorrect) estimates of factor effects. Further, this design has no capacity to estimate factor interactions—which are intrinsic to many biometric systems.

The motivation for this paper is thus to introduce to the biometrics community an alternative method for conducting a sensitivity analysis with the advantage that:

- (1) The system will be better understood.
- (2) The system will be better characterized.
- (3) The system will be better optimized—with the net effect that system performance is significantly improved in a computationally efficient fashion.

Thus, in short, the objective of this paper is to introduce and apply a structured “Sensitivity Analysis” approach for gaining insight and understanding about the system’s key components—those which most affect the quality and performance of a biometric system—and to optimize the settings of these key components.

Sensitivity analysis as we describe it has two separate and distinct steps:

- (1) *Experiment Design* (the structured plan for collecting the data), and
- (2) *Statistical Analysis* (the structured methodology for analyzing the data).

Both parts are critical, and when optimally used in concert yield enhanced insight into the relative importance and effect of the various computational components (and interactions) affecting biometric system performance. The experiment design and data analysis are demonstrated by a particular iris-based biometric system, VASIR (Video-based Automated System for Iris Recognition), which has verification capability for both traditional still iris images and video sequences captured at a distance while a person walks through a portal [11].

The general structure of this paper is threefold:

- (1) *Orthogonal fractional factorial design*: Introduce to the biometrics community a structured orthogonal fractional factorial experiment design methodology to efficiently gain insight and understanding (“sensitivity analysis”) of critical system parameters, interactions, and their optimal settings—this introduces and applies an established method within the statistical community [12,13].
- (2) *Statistical analysis*: Present effective and insightful statistical analysis methodologies for carrying out sensitivity studies.
- (3) *Demonstration with VASIR*: Demonstrate our experiment design and analysis methodologies for VASIR, with potential application to the broader biometrics field.

This sensitivity analysis approach provides a tool for understanding the computational components affecting the overall performance of a biometric system. Based on such understanding, the logical follow-up is to carry out an optimization analysis (identifying the optimal global settings of the components), and a robustness analysis (assessing the range of validity for our sensitivity and optimization conclusions). Our current paper focuses on the details of the sensitivity analysis only. To demonstrate the elements of the sensitivity analysis approach, we restrict our focus to a fixed setting for two robustness factors: (1) eye position (left eye only), and (2) image type (video matching: non-ideal to non-ideal image only).

2. Sensitivity analysis methodology

Sensitivity analysis is the experimental process by which we determine the relative importance of the various factors of a system. Suppose a system has k factors (input parameters) which potentially affect its performance. The minimal deliverable of a sensitivity analysis is to produce a ranked list of those k factors—ordered most to least important. For complicated systems (e.g., biometrics), a more desirable deliverable is to produce a ranked list which contains not only the k main factors, but also includes the various interactions of a system. To generate such a list implicitly means that each factor effect must be estimative, and such estimates should have as minimal bias and uncertainty.

As it turns out, such bias and uncertainty is driven primarily by the choice of experiment design that the analyst employs—some designs yield noisy effect estimates, while others yield very accurate estimates.

A good experiment design is important—it assures that the resulting data from the design has the capacity to answer the scientific question at hand—in particular, the data must have the capacity to yield a valid and rigorous ranking of the factors under study. Important as the experiment design component is, a complementary component is also important, namely, the statistical analysis methodology employed to analyze the data resulting from the design—what techniques must be brought to bear on the data so as to optimally estimate, order, and highlight the various factor effects. Hence sensitivity analysis consists of two separate and distinct steps:

- (1) *Experiment Design* (the structured plan for collecting the data), and
- (2) *Statistical Analysis* (the structured methodology for analyzing the data).

The detailed elements of the two components for our sensitivity analysis are illustrated by application to a particular iris-based biometric system: VASIR (Video-based Automated System for Iris Recognition), which has verification capabilities not only for traditional still iris images but also for video sequences taken at a distance with moving subject (see the details in Section 3).

2.1. Experiment design

Experiment design as a discipline is a systematic and rigorous approach for scientific and engineering problem-solving. The general goal of experiment design is threefold:

- (1) To produce insight and understanding into the factorial dependencies of a system.
- (2) To produce unambiguous, valid, and defensible conclusions.
- (3) To achieve both of the above with as small a sample size (time and cost) as possible [14].

Sensitivity Analysis offers to the biometric scientist the understanding and insight as to what is important and what is not in a system—and where the scientist should focus near-term and long-term research efforts. In this regard we shall briefly review and compare four commonly used experiment designs for sensitivity analysis: (1) Randomization Designs (Monte Carlo), (2) One-Factor-at-a-time (1FAT) Designs, (3) Full Factorial Designs, and (4) Orthogonal Fractional Factorial Designs.

- (1) *Randomization Designs (Monte Carlo)*.

Monte Carlo is a common methodology for many scientific sensitivity analysis studies. In essence, it considers the entire population space of factors and settings and then randomly samples a

user-specified number from that population space (see details in Kleijnen [15]). The advantage of Monte Carlo is that “on the average” the estimated main effects and the interactions will be unbiased. The primary disadvantage with Monte Carlo is that it tends to be expensive; the required sample size to estimate factor effect of the desired precision frequently exceeds an affordable sample size.

(2) One-Factor-at-a-Time (1FAT) Designs.

1FAT is an extremely common and popular sensitivity analysis methodology in both the biometric community and the larger scientific community. In the 1FAT design, an initial run is made where all k factors are set at a baseline value, and k subsequent runs are made whereby each factor is changed in succession to an alternative value. 1FAT designs are simple, intuitive, and inexpensive (costing as little as $n = 1 + k$ runs). On the other hand, such designs routinely produce effect estimates (and hence the subsequent ranked list of factors) which are frequently significantly biased and imprecise (see Box et al. [16] and Saltelli et al. [8]). Further, 1FAT designs have no ability whatsoever to estimate interactions—which are commonly existent in biometric systems.

(3) Full Factorial Designs.

Full factorial designs consist of running all possible combinations of all levels of all factors (see details in the NIST website [14]). The advantage of a full factorial design is that it provides rigorous information (estimates) about the relative importance of all k main factors, and all interactions (2-term, 3-term, . . . , k -term). Given these estimates, it is thus an easy step to generate the ideal ranked list, which includes all main effects and all interactions of all orders. On the other hand, the significant disadvantage of full factorial designs is that such designs are frequently too expensive. For example, if all k factors have $l_i = 2$ levels, the total cost is $n = 2^k$; if all k factors have $l_i = 3$ levels, the total cost is $n = 3^k$, and hence even for modest values of k , the resulting n can quickly become unaffordable/expensive.

(4) Orthogonal Fractional Factorial Designs.

Orthogonal Fractional Factorial Designs are a viable alternative to full factorial designs because they use only a fraction of the runs needed for a full factorial design, while still yielding good effect estimates (see details on the NIST website [14]). Orthogonal designs in this context refer to designs with both 1- and 2-dimensional balance, which is balanced for every single factor in the design and balanced for every pair of factors in the design, respectively. Balanced for a single factor means that every setting in that factor occurs the same number of times over the n runs. Balanced for a pair of factors means that every possible pair of settings across the two factors occurs the same number of times over the n runs.

As an example, for the simplest case in which each of k factors has only two levels, then for an n -run experiment, a factor is singly-balanced if the two levels (here coded as -1 and $+1$) occur the same number of times ($= n/2$) across the n runs. Similarly, two factors with two levels are doubly-balanced if the four combinations: $(-1, -1)$, $(+1, -1)$, $(-1, +1)$, and $(+1, +1)$ occur the same number of times ($= n/4$) across the n runs. The primary virtue of orthogonal designs is that their balance yields excellent (small bias and high precision) statistical estimates for the main effects and interactions.

We note in passing that full factorial designs—though typically expensive—are inherently orthogonal. This is because if the factor X_i of a full factorial design has l_i levels, then each level of X_i will occur the same number of times, namely, n/l_i . Similarly, if two factors X_i and X_j have l_i and l_j levels, respectively, then each pair of levels of X_i and X_j will occur the same number of times, namely, $n/(l_i * l_j)$. To

retain the estimation virtues of orthogonality while avoiding the run-expense of full factorial designs, leads one to consider the use of these orthogonal fractional factorial designs.

Note that since a fractional factorial design is any design whose total number of runs n is less than the corresponding full factorial design, then by definition the 1FAT design qualifies as a fractional design—but a decidedly non-orthogonal fractional design and so has poor estimation properties.

In a scientific/biometric experimental situation, if the performance response function has any interactions, then orthogonal fractional factorial designs typically yield excellent, trusted sensitivity analysis conclusions. Our proposed sensitivity analysis approach acknowledges this superiority, and thus uses orthogonal fractional factorial designs as the centerpiece for our experiment design component.

2.2. Data analysis

For the analysis of data drawn from orthogonal fractional factorial designs, the available statistical methods fall into two broad categories: (1) quantitative and (2) graphical. For quantitative methods, there are a variety of tools that could be employed (see Box et al. [16]), with the classical ANOVA (Analysis of Variance) method being the most common quantitative tool for sensitivity analysis. On the other hand, these quantitative methods are not always the best practical choice for the analysis of sensitivity experiments for a variety of reasons, namely, too much of a “black box” of statistical procedures, too much of a removal from the raw data, too many assumptions which must be adhered to (and tested for), and too much of a disconnect as to whether the resulting conclusions are consistent with the data.

For this reason, the approach that we apply to the analysis of data from our sensitivity experiments is via graphical data analysis methods—in particular EDA (Exploratory Data Analysis) graphical methods. Such methods take the approach of keeping “close to” the data, and judiciously displaying the data in such a fashion that the relative importance of the factors (and interactions) becomes evident from the (carefully constructed and augmented) plots. In this regard, we have assembled a battery of graphical procedures fine-tuned for sensitivity studies conducted via orthogonal (full or fractional) experiment designs. This battery of procedures was developed at NIST [12–14] and is an integral part of the NIST-developed analysis software tool Dataplot [17,18].

We will demonstrate the power of the sensitivity analysis methodology by using three graphical tools, in particular: (1) the Main Effects plot; (2) the Ordered Data plot; and (3) the Interaction Effects Matrix. These graphical data analysis methods serve as an important post-data component which complements the pre-data orthogonal experiment design component.

3. VASIR (Video-based Automatic System for Iris Recognition)

Iris recognition is a popular biometric system approach whose effectiveness is due to the highly distinctive features of the human iris. Most commercial systems for iris recognition are relatively expensive and are computational black boxes that run proprietary algorithms. In this light, to advance iris-based biometrics technology—IrisBEE (Iris Biometric Evaluation Environment) [19] algorithm—was implemented in the C programming language from Masek’s Matlab code [9]. IrisBEE was developed as a baseline for traditional still-based iris recognition, and hence there is still a need to overcome a number of challenges for images taken under more flexible acquisition and environmental condition (e.g., video taken at a distance).

In contrast to IrisBEE, VASIR [11,20] was developed with Near-Infrared (NIR) face-visible video-based iris recognition as part of its

domain scope. VASIR is a fully automated system for video-based iris recognition, capable of handling videos that were captured under less-constrained environmental conditions, such as a person walking through a portal at a distance. The VASIR system was designed, developed, and optimized to be robust—to address the challenge of recognizing a person in less-than-optimal environments, while coping with both high and low image/video quality. Unlike commercial iris recognition software (expensive and black box), VASIR provides an opportunity for the biometrics research community to examine the effect of algorithmic component changes, to extract and re-use its freely available source code, and to otherwise advance the state-of-the art of iris recognition technology.

VASIR has the capacity to automatically detect and extract the eye region, and subsequently to automatically assess and select the best quality iris image from NIR face-visible video. After this process, VASIR carries out a comprehensive image comparison analysis that in turn yields a verification result. As shown in Fig. 1, the VASIR system can principally be categorized into three modules: (1) image acquisition, (2) video processing, and (3) iris recognition.

Each module consists of several components that have all been designed, developed, and optimized to achieve high verification performance. In the Image Acquisition module, VASIR loads a still image or video sequence. In the Video Processing module, VASIR automatically detects the eye region from face/hair/shoulder visible frames in a video sequence and extracts the left/right iris images. VASIR then calculates automatically the image quality score of the extracted iris image from each frame. Based on the score calculated by the automatic image quality measurement (AIQM) method—a metric based on Skewness and Laplacian of

Gaussian (LoG) as developed in [11], the best quality iris images—one for left and one for right—are automatically chosen from all available frames. The Iris Recognition module is fed either the resulting iris images from the Video Processing module or the still iris image. For both video and still iris images, VASIR localizes the iris region based on the results of the segmentation algorithm. The segmented iris regions are then extracted and normalized based on polar coordinates and interpolation. Next, VASIR extracts the iris features—defined as a unique texture within the annular region between the sclera and the pupil—from the normalized iris images and encodes the extracted features as binary strings along with a noise-mask. In the matching stage, VASIR matches the extracted biometric template to existing templates. Note that all procedures are fully automatic—see Lee's paper [11] for detailed methods and procedures.

4. Experiment design: VASIR case

The purpose of a Sensitivity Analysis is first and foremost to gain insight into the important factors (and interactions) which drive the biometrics system. In this regard, the primary Sensitivity Analysis output is a ranked list of factors and interactions along with estimates of the magnitude of their effects. To achieve that, the biometrics researcher needs to provide information about the following:

- (1) model;
- (2) factors;
- (3) responses;
- (4) max affordable number of runs; and
- (5) choice of design.

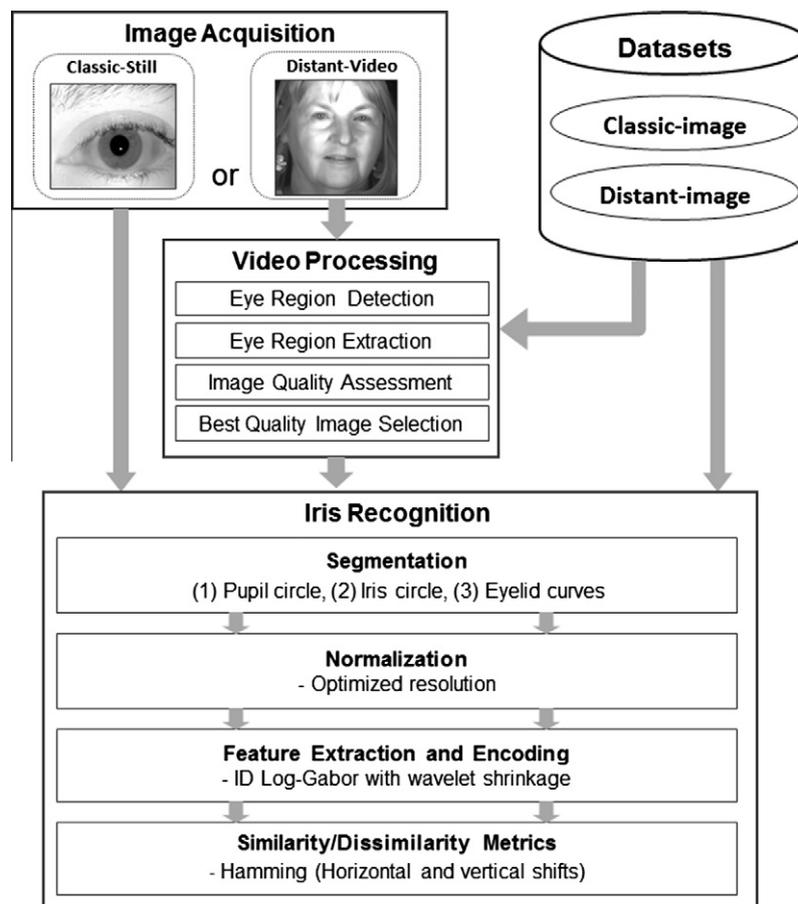


Fig. 1. VASIR system architecture.

4.1. Model

The starting point for a formal experiment design is to represent a model for the system response. For many scientific applications, a generic response model can be expressed as:

$$Y = f(X_1, X_2, \dots, X_k)$$

where Y represents a general response and k represents the number of factors that affect the response.

In the case where we have multiple responses, as in VASIR, the model can be generalized to

$$Y_i = f_i(X_1, X_2, \dots, X_k)$$

to indicate that the different factors may affect each individual response Y_i in its own and separate way.

4.2. Factors

For a biometric system in general, there are many factors (parameters or variables) that can affect a system's response (or performance)—the VASIR system is no exception. The VASIR factors naturally fall into four categories:

- (1) environmental conditions;
- (2) image conditions;
- (3) subject characteristics; and
- (4) algorithmic components.

Factors described in Table 1 are a representative—but partial—list of such potential factors from the three categories, which could be further examined unto themselves to assess the robustness of our conclusions about our fourth category (algorithmic components).

At a minimum, an experiment design is characterized by two numbers (k, n) where k is the number of factors in our study and n is the affordable number of runs. If we chose to examine and vary all of the factors listed in Tables 1 and 2, k would equal 71 (=33 + 38). In general, the number of runs n in an experiment must exceed the number of factors k ; hence to obtain estimates for all $k = 71$ factors would require $n > 71$. Further, to estimate the $\binom{k}{2}$ two-term interactions, this would require $n > 2500$. This is too expensive for our time and cost constraints. Inasmuch as the purpose of our study is to demonstrate the efficiency of our experiment design technique and to understand and optimize the algorithmic

Table 1
Potential robustness factors (33) that can affect a biometric system.

Environmental conditions	Image conditions	Subject characteristics
(1) Indoor/outdoor	(7) Focus or sharpness	(18) Contact lenses
(2) Lighting	(8) Contrast	(19) Glasses
(3) Background	(9) Brightness	(20) Left/right
(4) Weather	(10) Motion blur	(21) Gender
(5) Camera	(11) Resolution	(22) Age
(6) Day/night	(12) Noise	(23) Cosmetics (e.g. mascara)
	(13) Color	(24) Hair color
	(14) Distortion	(25) Race
	(15) Artifacts	(26) Movement
	(16) Reflections	(27) Distance from camera
	(17) Video/still (scenarios)	(28) Skin color
		(29) Eye color
		(30) Eyelashes
		(31) Eyebrows
		(32) Pupil dilation
		(33) Usability (e.g., behavior, training, perception)

factors of VASIR, we thus choose to focus primarily on the algorithmic category only.

An important early step in the structured experiment design process was to collect and construct the superset of possible variables (“factors”) that may affect the quality and performance of the VASIR algorithm. For this study, if we analyzed all 38 algorithmic factors, this would require >750 observations—still too expensive.

To accommodate affordability, the next step is to reduce the number of factors k . We hence choose to limit ourselves to only $k = 8$ algorithmic factors; we highlighted these eight algorithmic factors in Table 2 with gray. The reason why we have chosen these particular eight algorithmic factors is so that at least one factor was drawn from each of VASIR's eight key components (excepting the Best Image Quality Selection component whose setting is dictated by the choice utilized for preceding Image Quality Assessment component). Other reasons for our choice of these eight algorithmic factors were to analyze the sensitivity of new methods being considered for VASIR, to focus on those methods of greatest research interest, and to concentrate on those specific algorithms for recent upgrades for VASIR.

As a next step, we reduced the number of settings (levels) of the eight algorithmic factors. In particular, if each factor could (judiciously) be represented by $l = 2$ levels, then n will be reduced markedly. Table 3 summarizes the eight chosen algorithmic factors and their setting levels.

The detailed description for each factor and its levels are as follows:

- (1) X1 (EyeAlg): Eye position alignment.

A tilted head or subject movement results in a larger angular difference between the target and the query iris image—causing rotational inconsistency; i.e., the matching point within the two iris templates is different. In VASIR, to compensate for the angular difference, the positions of the left and right eyes were automatically and angularly aligned according to the estimated degree of the distance difference of the left and right pupil center. We analyzed whether the eye position alignment approach (ON) was actually better than without eye position alignment (OFF).

- (2) X2 (IQMetr): Image quality metrics.

The iris image quality assessment can help to predict whether an image is usable or recognizable, and it can also help to determine which image out of a set of frames in a video sequence has the best quality. VASIR developed multiple quality metrics for measuring iris image quality automatically for images (or videos) captured in different environments. VASIR automatically selects the best quality iris frame out of a video on the quality assessment metric. Although 16 metrics [11] were introduced to measure one or another aspect of image quality, we focused on the following $l = 3$ metrics (SOB, LoG, CON) for our sensitivity analysis:

A. Sobel (SOB) filter.

The Sobel operator has been used extensively for image edge detection and for measuring the focus level of an image [21–23]. The gradient at each point and the orientation of that gradient magnitude $G(x, y)$ can be measured by:

$$G(x, y) = \sqrt{(D_h(x, y))^2 + (D_v(x, y))^2}$$

$$SOB = \frac{\sum_1^n G(x, y)}{N}$$

The Sobel operator consists of a pair of 3×3 filters defined as D_h and D_v and these filters are designed to respond to edges run-

Table 2
VASIR possible algorithmic factors (38) and their popular settings that can be controlled for optimization.

Key components	Possible parameters/methods/factors	Type	Popular settings	#Settings
Image acquisition (IA)	Image quality control	D	Quality control system, ...	10
	Algorithms	D	Haar,PCA, ...	10
Eye region detection and extraction (FX)	Classifier cascade types	D	eye-pair, left-only, right...	10
	Minimum resolution for eye region	C	0-200	200
	Width of nose bridge	C	1-100	100
	X1: Pupil position alignment left/right	D	(on, off)	2
Image quality Assessment (AIQM)	X2: Algorithms	D	Sobel, LoG, CON,...	15
	Image enhancement/restoration	D	Super-resolutions,...	5
Best Quality Image Selection (ABIS)	Algorithms	D	ABIS,...	3
	* Pupil circle			
Segmentation (SEG)	Algorithms	D	Hough, active contours, ...	10
	Datatype	D	Video, still iris, webcam,...	5
	Ratio between width and height	C	1-12	12
	Thresholds	C	0-255	256
	Scale	C	1-4	5
	Radius limits	C	1-200	200
	Closing iterations	C	0-4	5
	Opening iterations	C	0-4	5
	* Iris circle			
	Algorithms	D	Hough, active contours, ...	10
	Datatype	D	Video, still iris, webcam..	5
	Thresholds	C	0-1 (0.1,0.2,...)	100
	Scale	C	1-4	5
	Radius limits	C	1-200	200
	Distance between pupil c and iris c	C	1-100	100
	* Eyelids			
	X3: Algorithms	D	line/curve fitting, ...	10
Adjustment	C	0-9	10	
* Noise removal				
Thresholding for eyelashes	C	0-19	20	
Thresholding for reflections	C	235-255	20	
Normalization (NORM)	Algorithms	D	Polar, interpolations, ...	4
	X4: Radial resolution	C	1-100	100
	Angular resolution	C	1-500	500
Feature extraction and encoding (FX)	Algorithms	D	Gabor, SIFT, DAISY, ...	10
	* Gabor-wavelet parameters			
	X5: Wavelength (pixel unit)	C	1-30	30
	Bandwidth	C	0-1	100
	X6: Masking a level of magnitudes	C	(0-5), (80-100)	25
	Encoding scale	C	1-10	10
Similarity/dissimilarity metrics (SM)	X7: Algorithms	D	Hamming, L2, COS,...	10
	* Directional bit-wise shifting method			
	X8: Horizontal (left and right)	C	0-29	30
	Vertical (Up and down)	C	0-9	10

Type: C = Continuous/quantitative/ordinal, D = Discrete/qualitative/non-ordinal

ning horizontally and vertically relative to the pixel grid. N is the total number of input image pixels and the SOB value is calculated by averaging the gradient value.

B. Laplacian of Gaussian (LoG) filter.

LoG is an important filter with much attention given to it [24]. LoG is defined as;

$$LoG = \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} e^{-(x^2+y^2/2\sigma^2)}$$

We used a 9×9 filter with $\sigma = 1.4$ in our experiment. To measure the quality score, the LoGED (Edge Density) is computed as;

$$LoGED = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} |LoG(x,y)|$$

where $M \times N$ is the number of pixels in the search area, and $LoG(x,y)$ is the calculated value at location (x,y) .

C. Contrast (CON).

Contrast is a measure of the intensity differences between a pixel and its neighbor over the whole image [25,26].

$$CON = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j)^2 P(i,j|d, \theta)$$

where G is the number of gray levels. The matrix element $P(i,j|d, \theta)$ contains the second order statistical probability values for changes between gray levels i and j at a particular displacement distance d and at a particular angle θ (see details written

Table 3

The eight algorithmic factors and their setting levels (see detailed procedures in the paper [11]).

Eight algorithmic factors	Number of setting levels				
X1 (EyeAlg): Eye position alignment		-1 (OFF)		+1 (ON)	
X2 (IQMetr): Image quality metrics		-1 (SOB)	0 (LoG)	+1 (CON)	
X3 (SegEye):Eyelids segmentation		-1 (Lines)		+1 (Curves)	
X4 (NorRes): Radial resolution		-1 (20px)		+1 (32px)	
X5 (FXWL): Wavelength		-1 (18px)		+1 (16px)	
X6 (FXMask): Wavelet magnitude		-1 (0.8)		+1 (0.9)	
X7 (SMAIlg): Similarity metrics	-2 (HD)	-1 (COR)		+1 (COS)	+2 (WED)
X8 (SMSH): Horizontal shifting		-1 (10)		+1 (5)	

by Albreghsen [25]). A higher contrast is considered as an indicator of better quality iris image in our study.

(3) X3 (SegEye): Eyelids segmentation.

Although many methods for iris segmentation has been suggested in the literature [27–29], we first compare IrisBEE segmentation approach with VASIR's approach. In the IrisBEE algorithm [9], the eyelids were removed by inserting horizontal flat lines to delimit the upper and lower eyelids. It is important to understand that human eyes are known to have different curvatures for the upper and lower eyelids. In addition, the shape of the eye can be significantly different depending on the person; e.g., race: caucasian, asian, etc. VASIR therefore developed two different curves to segment the actual upper and lower eyelid shape. We examined how the $l=2$ eyelids segmentation approaches (Lines, Curves) influence VASIR performance.

(4) X4 (NorRes): Radial resolution for normalization.

VASIR involves the comparison of two biometric iris samples. Even for multiple images of the same subject, a complication arises in such a comparison due to pupil dilation, non-concentric pupil displacement, or varying eye distance to the capture device. To facilitate the comparison, the multiple images must be stretched or compressed to a standardized scale (normalization).

For the normalization step, a standardized 2D image of the iris region is generated by a polar coordinate-based method (proposed by Daugman [30]) based on two circular constraints (pupil and iris). The Daugman's rubber sheet model assigns to each point within the iris region a pair of real coordinates (r, θ) where the radius r lies on the unit interval $[0, 1]$ and θ is the angle over $[0, 2\pi]$. The remapping of the iris image $I(x, y)$ from Cartesian coordinates (x, y) to polar coordinates (r, θ) is classically represented as:

$$I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta)$$

$$x(r, \theta) = (1 - r)x_p(\theta) + rx_s(\theta)$$

$$y(r, \theta) = (1 - r)y_p(\theta) + ry_s(\theta)$$

where $(x_p(\theta), y_p(\theta))$ and $(x_s(\theta), y_s(\theta))$ are the coordinates of the pupil and iris boundaries respectively along the θ direction. As in IrisBEE's algorithm, an iris pattern image was generated by normalizing the angular size ($\theta = 240$ pixels) and the radial size ($r = 20$ pixels).

For our experiment, to determine the relative importance of r , its potential interactions, and its optimal values, we examined the effect of $l = 2$ settings for radial size: 20 and 32 pixels.

(5) X5 (FXWL): Wavelength (Gabor filter parameter).

Some different approaches [31,32] have been suggested to represent the patterns from the iris region. The IrisBEE algorithm employed a 1D Log-Gabor filter—introduced by Yao et al. [33]—to process the feature extraction from normalized iris images. The frequency response of a Log-Gabor filter is given as:

$$G(w) = e^{-\log\left(\frac{w}{w_0}\right)^2 / 2 \log\left(\frac{\sigma}{w_0}\right)^2}$$

where w_0 represents the filter's center frequency (wavelength) and σ gives the filter bandwidth. We used $l = 2$ different wavelengths (18 px, 16 px) to determine the relative importance of w_0 , its interactions, and its optimal settings.

(6) X6 (FXMask): Wavelet magnitude.

To encode coefficients (complex value) with the binary iris code $[0, 1]$, VASIR employed the phase information with four quadrants proposed by Daugman [30]. To determine each bit of the iris code for the coefficient, if the real part of the coefficient is negative, the iris code is mapped to "0", otherwise it is mapped to "1". If the imaginary part is negative, the iris code is mapped to "0", otherwise it is mapped to "1". This is to assure that unimportant bits would not be included when measuring the similarity between two biometric templates [34]. The paper written by Lee [11] suggested that VASIR's approach based on the distance (magnitude) of the coefficient from the origin is superior than the Hollingsworth's fragile bit approach [34] based on filter responses near the axes of the real or imaginary part. We then examined what levels of small/large magnitude values affect VASIR performance. For this, we varied the masking of significant bits based on larger magnitudes (0.8, 0.9)—that is the effect of keeping $l = 2$ 80% (vs. 90%) of the bits before computing the similarity distance.

(7) X7 (SMAIlg): Similarity metrics.

Similarity metrics provide a quantitative measure for the degree that the two templates match. A wide variety of similarity metrics have been proposed in the iris-based biometrics community [35–37]. For this case when we are comparing different iris images (from the same or from different person), this section examines $l = 4$ metrics for comparison scores (HD, COR, COS, WED):

A. Hamming Distance (HD).

HD can be used to decide whether two iris templates are of the same person. In VASIR, a fractional HD was applied to its iris recognition system—initially proposed by Daugman [35] and later re-implemented by Masek [9]. A noise mask helps to

exclude insignificant bits (e.g., eyelashes, eyelids, reflections) of a template. The fractional HD is given by:

$$HD(T, Q) = \frac{\sum_{i=1}^N (T_i \oplus Q_i) \cap (Tm_i \cap Qm_i)}{N - \sum_{k=1}^N (Tm_k \cup Qm_k)}$$

where target (T) and query (Q) are two bit-wise templates and Tm and Qm are the corresponding noise masks. N is the total bits of a template.

B. Normalized Cross-Correlation (COR).

COR is a measure of similarity between two templates for image processing. The images are normalized by subtracting the mean value and dividing by the standard deviation as follows:

$$COR(T, Q) = \frac{\sum_{i \in \{Tm_i \cap Qm_i\}} (T_i - \mu T) \cdot (Q_i - \mu Q)}{\sqrt{\sum_{i \in \{Tm_i \cap Qm_i\}} (T_i - \mu T)^2 \cdot \sum_{i=1}^N (Q_i - \mu Q)^2}}$$

where μT and μQ are the means of T and Q .

C. Cosine similarity (COS).

COS is a measure of similarity between two vectors by measuring the cosine of the angle; the angular separation is $[-1, 1]$.

$$COS(T, Q) = \frac{\sum_{i \in \{Tm_i \cap Qm_i\}} T_i \cdot Q_i}{\sqrt{\sum_{i \in \{Tm_i \cap Qm_i\}} T_i^2 \cdot \sum_{i \in \{Tm_i \cap Qm_i\}} Q_i^2}}$$

D. Weighted Euclidean Distance (WED).

WED can be used to determine similarity of an unknown sample set to a known one and is given as:

$$WED(T, Q) = \sum_{i \in \{Tm_i \cap Qm_i\}} \frac{1}{(\sigma_i^Q)^2} (T_i - Q_i)^2$$

where σ_i is the standard deviation of the i th feature of the template Q .

(8) X8 (SMSh): Horizontal shifting.

The starting point for normalizing the iris region of an iris image varies due to the subject's head tilt, movement, and when the subject looks in a different direction—we call this rotational inconsistency. To overcome this rotational inconsistency between two iris templates, one template is two bit-wise shifted left or right, and the similarity score is selected from successive shifts [9,35]; e.g., the smallest value is a successive shift value for the Hamming Distance case.

VASIR developed a new shifting method in which the template is shifted not only left and right (horizontal) bit-wise, but also upward and downward (vertical); the values for these horizontal and vertical direction shifts are indicated by X and Y , respectively. We were interested in the effect of rotational inconsistency on VASIR performance; such inconsistency is linked to horizontal (as opposed to vertical) bitwise shifting and so we varied such horizontal bitwise shifting (by $l = 2$ levels: 5 and 10 bits) to determine the relative importance of a number of shifted bits, their potential interaction, and their optimal settings.

All remaining factors were fixed. We should note that our conclusions about the chosen eight algorithmic factors may be dependent on the settings of the remaining 30 (=38–8) algorithmic factors and the 33 robustness factors (see Appendix A and Table 1, respectively).

It is a sobering reality that “nature” will propagate the effect of these remaining algorithmic and robustness factors onto the performance of our VASIR system and onto our conclusions about the relative importance of these eight factors—regardless of

whether we identify such factors or not. For a good experiment design, it is critically important to pre-identify, control, or at least record the settings of these robustness factors during the entire course of the experiment.

4.3. Responses (Y_i)

Y refers to the response of interest for which we wish to evaluate the effect of the various factors. In biometric systems, it is quite common to have multiple responses of interest and these responses are frequently identical to various performance metrics.

Our study had three responses that were based on matching and non-matching comparison scores: (1) the VR (Verification Rate) when the FAR (False Accept Rate) was 0.01; (2) the VR when the FAR was 0.10; and (3) the EER (Equal Error Rate)—detailed definition is described in the papers [11,38] and note that VR is equal to True Positives (TP) and FAR is equal to False Positives (FP). Symbolically the three responses are.

$$Y1 = VR|FAR = .01 \text{ (or TP|FP = .01).}$$

$$Y2 = VR|FAR = .10 \text{ (or TP|FP = .10).}$$

$$Y3 = EER.$$

For a biometric system, higher values of VR and smaller values of EER indicate superior system performance. Common alternatives for VR|FAR are FAR = .001 and FAR = .0001. We chose not to use these as our performance metrics because they would not have been meaningful due to the relatively small number of different subjects (~ 70 – 100) in the chosen MBGC dataset selected for our study.

4.4. Max affordable number of runs

Our computing platform consisted of a dual core CPU (3.33 GHz) and RAM (64 GB) with Windows Server 2008 as the operating system. The practical constraint under which our study was operating was that the total amount of CPU time for the experiment in total would not exceed two weeks. This maximum two-week time constraint was chosen to allow for design re-execution due to possibility of “real world” negative events that many times do occur in large scientific investigations; e.g., crashes (hardware/software), memory leaks, debugging problems, dataset problems, design access problems, anomalous (unusual) looking results, and data analysis problems. In reality, in light of all of the above possibilities, it did in fact take approximately six months to design, collect data, and carry out a sensitivity analysis for the two-week run. Hence, even in a parallel computing environment, this ideal two-week time constraint translated into an upper limit of $n = 500$ runs to examine the eight algorithmic factors.

4.5. Choice of design

Given the ($k = 8$ factor, $n < 500$ run) constraint with six factors at two levels, factor X2 at three levels, and factor X7 at four levels, full factorial is an excellent design, but too expensive—far exceeding our $n = 500$ limit ($n = 2^6 \times 3^1 \times 4^1 = 768$ for all possible combinations of all $k = 8$ factors). On the other hand, orthogonal fractional factorial design is excellent with good main effects and interaction estimation properties, and is highly efficient.

The question then arises as to how to fractionate and what factors to fractionate on. Since fractionating three- and four-level factors is more difficult and subtle, we choose to fractionate on the six factors at two levels: this is referred to as a 2^{6-1} design and is described in Box et al. [16] (see p.276). In combination with the other levels of the two factors, the design that we chose is a

$(2^{6-1}) \times (3^1) \times (4^1)$ orthogonal fractional factorial. This design examines the eight algorithmic factors with $n = 384$ run—half the runs of a full factorial and well below our 500-run limit. This design utilizes two levels for each of the six algorithmic factors (X1, X3, X4, X5, X6, X8), three levels for X2, and four levels for X7; these two levels will be coded as $(-1, +1)$, three levels $(-1, 0, +1)$, and four levels $(-2, -1, +1, +2)$.

The ($k = 8$ factor, $n = 384$ run) $(2^{6-1}) \times (3^1) \times (4^1)$ orthogonal design matrix that we employed is shown in Appendix B—note that since some of the eight factors are more time-consuming to change than others, the run order of the design was optimized to minimize execution time (a $4\times$ speed-up was achieved).

Note that this design matrix has 8 columns (factors) and 384 rows (runs); since the design is orthogonal, each of six two-level factors (X1, X3, X4, X5, X6, X8) has the same number ($384/2 = 192$) of -1 's and $+1$'s. X2 factor has three settings ($384/3 = 128$) of -1 's, 0 's, and $+1$'s while X7 has four settings (each occurring $384/4 = 96$) of -2 's, -1 's, $+1$'s, and $+2$'s—this equality property is referred to as 1-dimensional balance. Due to orthogonality, each and every one of the $\binom{6}{2} = 15$ pairs of the six factors at two levels has the same number of $(-1, -1)$, $(-1, +1)$, $(+1, -1)$, and $(+1, +1)$ combinations—namely, $384/4 = 96$; this is referred to as 2-dimensional balance.

5. Data

5.1. Dataset

For the purpose of this sensitivity analysis study, we evaluated the VASIR system performance using datasets collected by MBGC (Multiple Biometric Grand Challenge) [39]. These MBGC datasets include iris images of varying illumination conditions, low quality, and off-angle or occluded images in both still and video imagery. One of the challenges for the MBGC dataset is to recognize a person using an iris from the NIR and high definition video as the person walks through a portal. In this experiment, we therefore use the NIR face-visible video dataset, which we will call “distant-video”;

the distant-video samples were captured with a video camera by the Sarnoff IOM system in 2048×2048 resolutions; with face/hair/neck visible in the screen.

For the MBGC distant-video dataset, the number of video sequences captured by IOM system is 628.

A small number (<50) of the subjects appeared in only one video sequence—and so these subjects were excluded because we wanted to have replication over at least two videos at a different time. Other subjects existed in multiple video sequences, some appeared in as many as 10 sequences. For parsimony, if a subject happened to appear in three or more video sequences, we extracted for our study two such video sequences—each sequence representing a different session. In summary, out of the 628 videos, we thus used 204 videos involving 102 subjects.

Fig. 2 shows an example of distant-video within the MBGC dataset.

Fig. 2a and b show the face-visible frame from a video and the sub-image (eye region) extracted from that frame. Distant-video data is normally considered to be a poor-quality image source since the video was taken with moving subjects, having motion blur, poor contrast, off-angle, poor illumination, and other deficiencies.

5.2. Data collection procedure

This section describes the procedure of collecting data for our focus in this paper: the left eye position and the face-visible distant-video to distant video (VV) matching scenario. For this VV matching scenario, we used 204 videos (102 subjects) whereby each subject appeared in two different video sequences and each taken in different sessions.

The VASIR system then proceeds as follows:

- (1) VASIR automatically adjudges, identifies, and extracts all admissible/visible left iris images (=565 in this case) out of these 204 video sequences (face-visible) based on criteria related to factor X1 (Eye region detection/extraction with pupil position alignment).

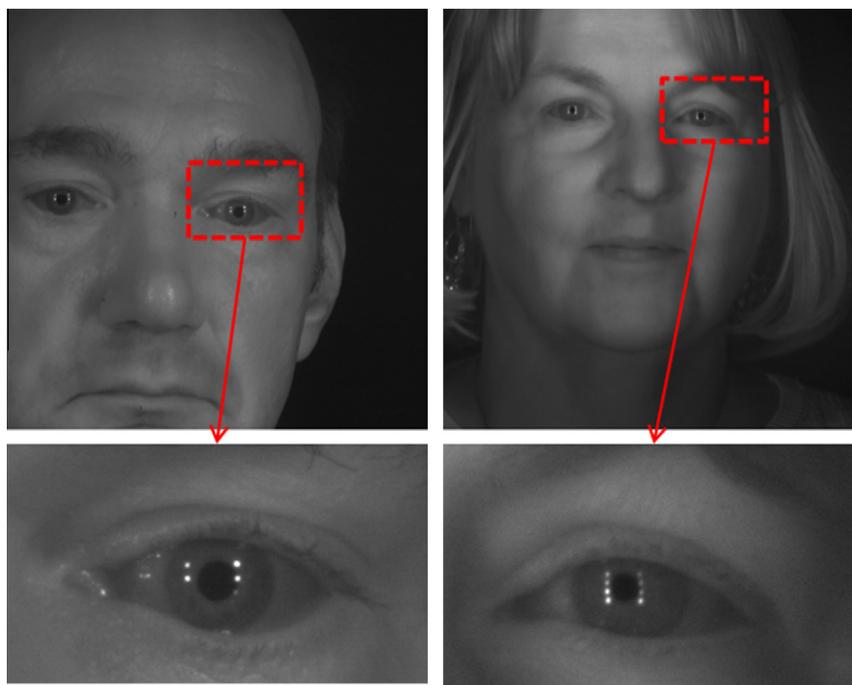


Fig. 2. Sub-images extracted from MBGC distant-video taken by IOM.

- (2) VASIR then selects—automatically—the best 204 iris images based on quality score criteria associated with factor X2 (Image quality metrics for quality assessment and the best image selection).
- (3) VASIR then takes the resulting 204 images and segments the iris region based on segmentation algorithms that included factor X3 (eyelid segmentation algorithm).
- (4) The segmented iris regions are extracted using polar-coordinates and then normalized based on the resolution related to factor X4 (Radial resolution).
- (5) VASIR extracts features from the normalized iris images based on the feature extraction algorithm associated with factor X5 (wavelength).
- (6) Then VASIR encodes the extracted features and masks out noise based on factor X6 (masking with wavelet magnitude).
- (7) The encoded results are then used for carrying out the pairwise biometric templates matching based on factor X7 (similarity metrics).
- (8) Finally to correct rotational inconsistencies between two biometric templates, VASIR applies factor X8 (horizontal shifting).

The above VASIR procedure is executed for all $n = 384$ runs of the $k = 8$ factors as specified by the $2^{6-1} \times (3^1) \times (4^1)$ experiment design. For a given run, we obtain a set of match scores and a set of non-match scores. Based on the set of match and non-match scores, VASIR automatically produces similarity scores which in turn yields the three performance responses (VR|FAR = .01, VR|FAR = .10, and EER). The sensitivity analysis was then initiated to determine the effect of the eight algorithmic factors on the above three responses.

6. Data analysis

This section describes the details of the sensitivity analysis carried out on the $k = 8$ algorithmic factors for the VASIR system. This analysis is carried out and presented for the fixed settings of the remaining algorithmic factors (see Appendix A) and robustness factors (see Table 1). In particular, the results presented in the remainder of this section are for the VV matching scenario and the left eye position.

The specific deliverables from the sensitivity analysis of the VASIR system are as follows:

- (1) A ranked list of all eight algorithmic factors—ordered by relative importance.
- (2) Inclusion of the relevant two-term interactions within that ranked list.
- (3) Estimates of the magnitude and the direction of factor effects and interactions.
- (4) Determination of the most important factor(s) and interaction(s).
- (5) Specification of the optimal (local) settings for the eight factors.

As we discussed in Section 2.2, to fulfill the goals in our study, we demonstrate three especially important plots of the 10 graphical procedures from [14]: (1) Main Effects plot, (2) Interaction Effects Matrix plot, and (3) Ordered Data plot.

6.1. Main Effects plot

The *Main Effects* plot is the most important graphical data analysis technique to identify the influential and statistically significant factors affecting performance responses. This plot provides the mean response for each setting of each factor and highlights their difference to show the effect of changes on the response(s) due to that factor [7].

Fig. 3 shows the Main Effects plot of VASIR's $k = 8$ factors on the VV Left case with the VR|FAR = .01 performance response.

The horizontal axis is the eight factors (X1 to X8) and the coded factor settings (e.g., “−1”, “0”, or “+1”) within each factor—the four coded factor settings for X7 (1, 2, 3, and 4) are equivalently referred to as “−2”, “−1”, “+1”, and “+2”. The vertical axis is the mean response for each setting of each factor. For each factor, a line connects the mean values for that factor. The magnitude of the line is the factor effect; longer lines indicate the factor has effects while shorter lines indicate the factor has no effect. The slope of the line indicates whether the factor has an increasing or decreasing effect on VASIR's responses.

The number of runs (n) in the design was originally $384 = 2^{6-1} \times 3^1 \times 4^1$. Some of the runs are ignored if the response value does not exist (due to a small number of subjects) for the relevant run, therefore, n may be less than 384. For example, the legend box of the Fig. 3 plot has $n = 383$ because one value out of the 384 does not exist for this VR|FAR = .01 response.

On the inside of the plot above the horizontal axis, the top number gives the percentage from a one-way ANOVA f -test; two asterisks (**) signifies that a factor effect is statistically significant at the 1% level (>0.99) and one asterisk (*) signifies significance at the 5% level ($>0.95 \leq 0.99$). The second number is the estimated factor effect in raw response units. In case of factors with two-level settings (coded as − and +, and with corresponding mean values \bar{y}_- and \bar{y}_+), the effect is uniquely defined as $\bar{y}_+ - \bar{y}_-$. For three-level settings (−, 0, and +, and with corresponding mean values \bar{y}_-, \bar{y}_0 , and \bar{y}_+), we define the factor effect as the largest in magnitude out of given three differences: $\bar{y}_+ - \bar{y}_-, \bar{y}_+ - \bar{y}_0$, and $\bar{y}_0 - \bar{y}_-$. For four-levels, we also define the factor effect as the largest out of $\binom{4}{2} = 6$ possible differences. The bottom number is the estimated percentage change (“relative effect” = $100 \times \text{effect/global mean}$). Note that since the design is orthogonal, such effect estimates are identically the least squares estimates that would result from a multi-linear regression.

For the response VR|FAR = .01 of the VV Left eye case, the important factors influencing VASIR's performance are ordered (most important to least important) as follows:

- X2 (IQMetr): Image quality metrics (effect = .041),
- X7 (SMAlg): Similarity metrics (effect = .038),
- X3 (SegEye): Eyelids segmentation (effect = .026),
- X1 (EyeAlg): Eye position alignment (effect = .020),
- X8 (SMSh): Horizontal shifting (effect = .017), and then
- X4 (NorRes): Radial resolution for normalization (effect = .011),
- X6 (FXMask): Wavelet magnitude (effect = .006), and
- X5 (FXWLBW): Wavelength (effect = .003).

Six factors (X2, X7, X3, X1, X8, and X4) are statistically significant (** or *) and are highlighted in red in Fig. 3; two factors (X5 and X6) are not statistically significant.

The Main Effects plot is also useful for determining optimal settings on the average—i.e., based on actual settings utilized in the experiment design. From Fig. 3, those settings for each factor which yield a large value (closer to 1.0) of VR would be the preferred optimal setting. The optimal settings are (+1, −1, +1, +1, +2, +1), where “•” indicates both settings are equivalent or the mean difference is not statistically significant.

From the Main Effects plots, a summary of the factor effects and optimal settings for all three responses is given in Table 4.

For all three responses, the most important factor out of these eight algorithmic factors in VASIR is X2 (IQMetr) and the second important factor X3 (SegEye) and X7 (SMAlg). The statistically significant factors having an influence on the VASIR system are (X2, X7, X3, X1, X8, and X4) in order from most to least important. The least important factors are X5 (FXWL) and X6 (FXMask) for all responses.

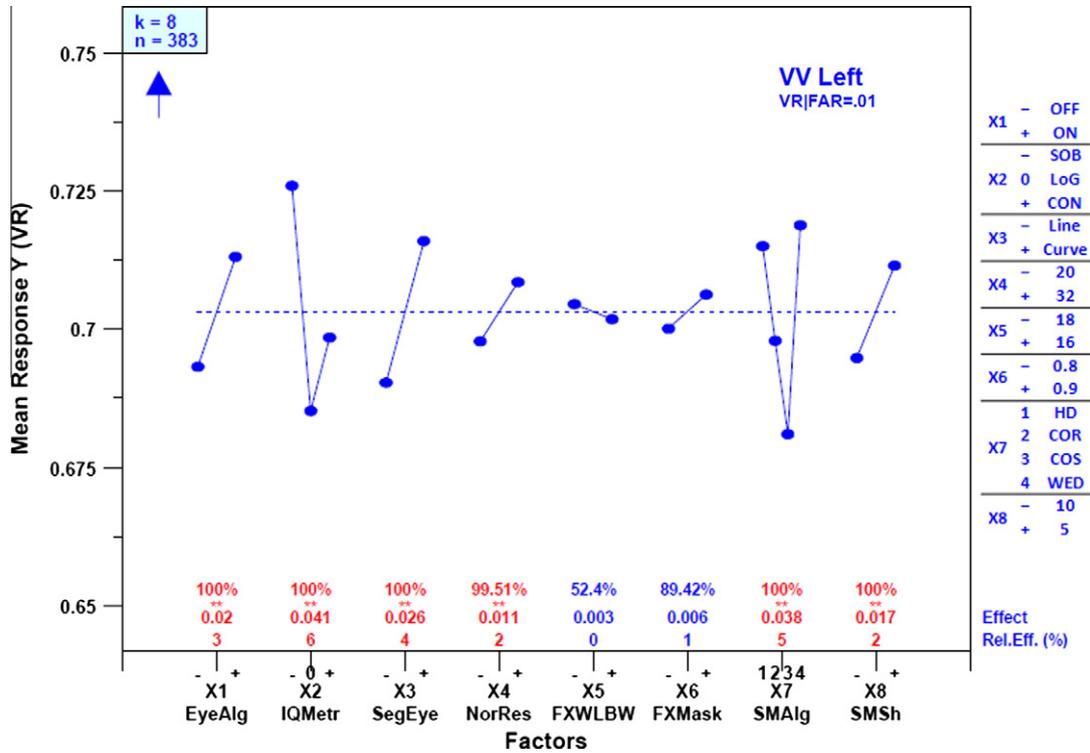


Fig. 3. Main Effects plot with the VR|FAR = .01 response for VV Left—note that the four coded settings (1,2,3,4) for X7 are equal to (−2,−1,+1,+2). The importance of factors X2, X7, X3, X1, X8, and X4 in order.

6.2. Ordered Data plot

The Ordered Data plot yields the settings that correspond to each response value. The plot is generated by ordering all responses—worst to best—and carrying along the corresponding settings for the $k = 8$ factors. If a factor has one setting for the best and near-best response values and the opposite setting for the worst and near-worst response values, then that factor is usually a relatively important factor [14]. Ordered Data plots can reconfirm the results from Main Effects plots.

Fig. 4 illustrates the Ordered Data plot for the VR|FAR = .01 response for the VV Left case. In our sensitivity analysis, we used a total of 384 runs from the eight algorithmic factors and so the plot potentially contains a maximum of 384 data points; for example, the Ordered Data plot in Fig. 4 has only 383 runs for the VR|FAR = .01 response.

Presenting all data points and their settings make the plot too crowded and difficult to interpret. So we illustrated only the 60 most extreme points—the best and near-best response values (30 points) on the right-side and the worst and near-worst response

values (30 points) on the left-side; the inserted blue vertical line distinguishes between the best-side and worst-side points.

The upper left corner box of the plot shows the number of factors ($k = 8$) and the number of data points (60) followed by the total number of runs (383) that have the corresponding VR|FAR = .01 response. The horizontal axis shows the eight factor labels and their settings for each of the 60 runs ordered from the smallest to the largest response values. The vertical axis is the value of the VR|FAR = .01 response.

The Ordered Data plot not only provides the best settings but also reaffirms important factors. On the average, if a factor has no effect then there should be a near even split of 15 + 1's and 15 − 1's, the more the divergence from 15/15, the greater the significance of the factor. For 30 trials, based on the binomial distribution, it is statistically significant whenever the count is ≥ 20 or ≤ 10 . For instance, we may pose the question as to whether factor X1 (EyeAlg) is important? Out of the 30 best VR|FAR = .01 responses, 29 of them came with X1 = +1 (statistically significant). For the 30 worst responses, 22 came from X1 = −1 (also statistically significant). Thus, X1 (EyeAlg) is an important factor.

Table 4 VASIR's most important (IMP) factors and optimal settings with three responses for VV Left—from Main Effects plots.

Responses	1st Most IMP	2nd Most IMP	Least IMP	Optimal Setting							
				X1	X2	X3	X4	X5	X6	X7	X8
VR FAR=.01	X2**	X7**	X5, X6	+1 (ON)	−1 (SOB)	+1 (Curves)	+1 (32)	•(−1) (18)	•(+1) (.9)	+2 (WED)	+1 (5)
VR FAR=.10	X2**	X3**	X5, X6	+1 (ON)	−1 (SOB)	+1 (Curves)	+1 (32)	•(−1) (18)	•(−1) (.8)	−2 (HD)	+1 (5)
EER	X2**	X3**	X5, X6	+1 (ON)	−1 (SOB)	+1 (Curves)	+1 (32)	•(−1) (18)	•(−1) (.8)	−2 (HD)	+1 (5)

** represent statistical significance at the 1 % levels, respectively.

• represents that the mean difference between settings is not statistically significant.

Table 6
VASIR's relative important factors and the best settings with three responses for VV Left—from Ordered Data plots.

Responses	Important Factors	Best Setting							
		X1	X2	X3	X4	X5	X6	X7	X8
VRIFAR=.01	X1, X2, X3, X7, X8	+1 (ON)	-1 (SOB)	+1 (Curves)	+1 (32)	-1 (18)	-1 (.8)	+2 (WED)	+1 (5)
VRIFAR=.10	X1, X2, X3, X8	+1 (ON)	-1 (SOB)	+1 (Curves)	-1 (24)	+1 (16)	-1 (.8)	-2 (HD)	+1 (5)
EER	X2, X7, X8	+1 (ON)	-1 (SOB)	+1 (Curves)	+1 (32)	-1 (18)	-1 (.8)	-2 (HD)	+1 (5)

- (1) interaction effects when the factors have two-levels and
- (2) interaction effects when the factors have three or more levels.

For the first method, where the two factors have only two levels, the representation becomes simpler. In such cases, the factor cross products serve as a reasonable surrogate for the interactions. In particular, if factor X1 takes on the coded values -1 and +1, and if factor X2 takes on the coded values -1 and +1, then the cross product X1 * X2 also takes on the coded values -1 and +1:

$$\begin{aligned}
 (-1) \times (-1) &= +1 \\
 (-1) \times (+1) &= -1 \\
 (+1) \times (-1) &= -1 \\
 (+1) \times (+1) &= +1
 \end{aligned}$$

The advantage of this method is that the cross product X1 * X2 (and for that matter any cross product) becomes just another -1 and +1 factor in the orthogonal system and so the interaction effects may be directly estimated and compared with one another and with the algorithmic factors.

On the other hand, if the two factors have three or more levels then the representation of such interactions is more complicated and the cross product representation is of little help. For our case with eight algorithmic factors, six factors have two levels, the factor X2 has three levels, and X7 four levels.

With this in mind, we make use of the Interaction Effects Matrix plot (see Fig. 5 below) which is a multi-plot per page display of the original main effects and all of the two-term interactions. In practice for an experiment with *k* factors, the total number of possible two-term interactions is:

$$\binom{k}{2} = \frac{k!}{2!(k-2)!} = \frac{k(k-1)}{2}$$

In our experiment, we have 28 two-term interactions from a ($2^{6-1} \times 3^1 \times 4^1$) experiment (*k* = 8) for each response. Fig. 5 presents the interaction effects matrix—it consists of 64 plots (8 Main Effects plots + 2 × 28 Interaction Effects plots)—two plots for each two-term interaction.

The plots on the diagonal are identical to those seen on the Main Effects plot; the horizontal axis is the factor levels and the

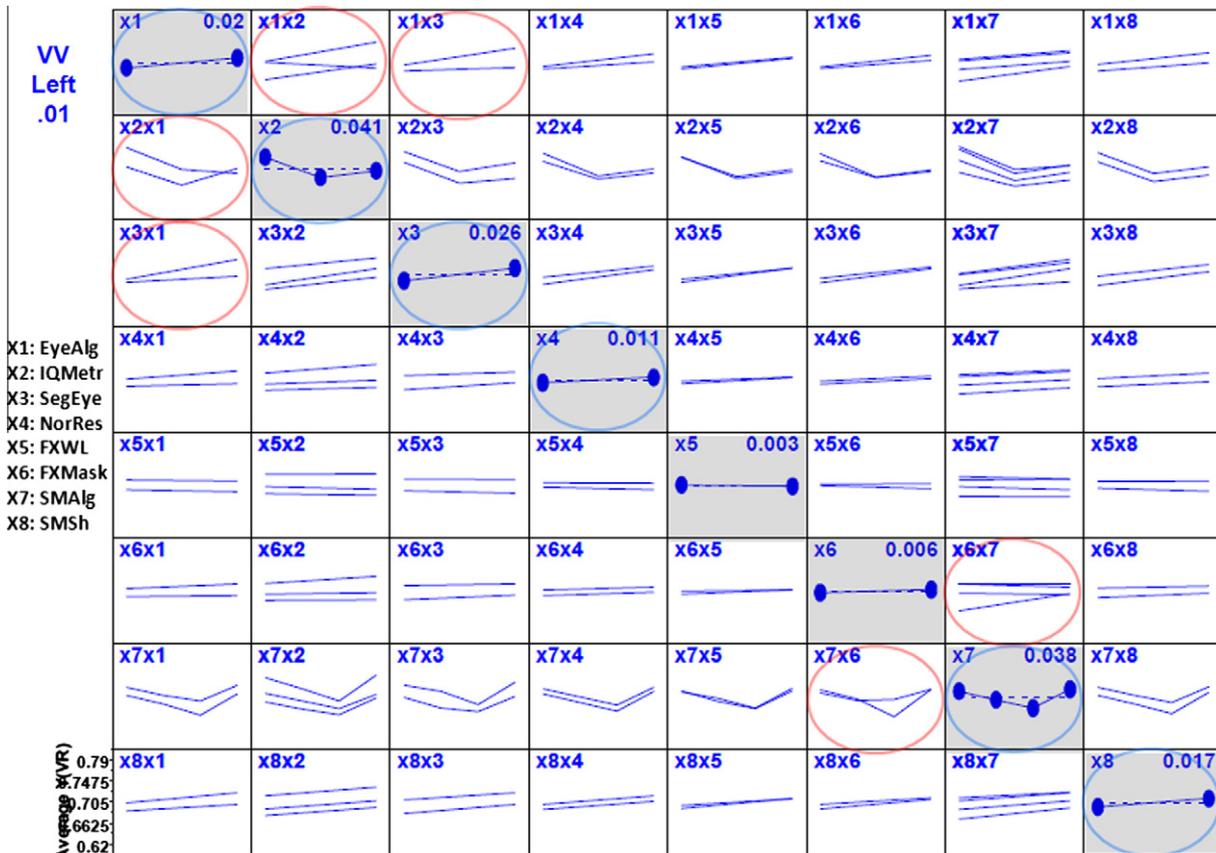


Fig. 5. Interaction Effects Matrix plot with the VR|FAR = .01 response for VV Left. The blue circles indicate significant factors and the red circles indicate two-term interactions where effects appear significant. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7
Summary of VASIR's interaction effects with three responses for VV left.

Response	Interaction Effects		
VR FAR=.01	X1X2 (EyeAlg•IQMetr)	X1X3 (EyeAlg•SegEye)	X6X7 (FXMask•SMAIlg)
VR FAR=.10	X1X2 (EyeAlg•IQMetr)	X1X7 (EyeAlg•SMAIlg)	
EER	X1X2 (EyeAlg•IQMetr)	X1X7 (EyeAlg•SMAIlg)	

vertical axis is the mean response at each of those levels; the index in the upper left corner identifies factors (e.g., X1) and the number following the index is the (least squares) estimated effect due to that factor (e.g., X1 (+1) effect $\beta_1 = .02$). The plots off-diagonal are the two-term interaction plots; the index in the upper left corner of each plot identifies the two-term interaction (e.g., X1X2, X1X3). For the X1X2 interaction (X1 with two levels and X2 with three levels), the horizontal axis is the two levels of X1 and the vertical axis is the mean response at each of those two levels; it is computed three times—once for each level of X2—thus yielding six ($= 2 \times 3$) points and three traces. From each of the three traces, we can compute an X1 effect—the steeper the trace, the larger the X1 effect. When the three traces are parallel, then the resulting three estimates of the X1 effect are identical—and so the factor X1 effect is not dependent on the level of X2 (no interaction). When the three traces are not parallel, then the X1 effect depends on the setting of X2 (interaction).

In summary, the Interaction Effects plots are to be interpreted as follows:

- (1) If the lines within the plot are parallel, then this implies no interaction.
- (2) If the lines within the plot are not parallel, this implies an interaction—the more divergent, the stronger the interaction.

If there are no interactions among factors, we can conclude that the system response is driven primarily by main effects. Otherwise, the interactions can lead to unexpected effects on the responses.

For the VR|FAR = .01 response in the VV Left case, Fig. 5 shows the relative importance of factors X2, X7, X3, X1, X8, and X4 in order—see the blue circles on the diagonal. For the most important factor X2, it is seen that there exists an interaction effect between X1 (EyeAlg) and X2 (IQMetr)—see the non-parallel lines in the X1X2 and X2X1. Similar cases are seen between X1 (EyeAlg) and X3 (SegEye), and between X6 (FXMask) and X7 (SMsh)—see the red circles on the off-diagonal.

The Interaction Effects Matrix plot for VR|FAR = .01 response shows that almost all of the 2×28 plots have near-parallel traces and hence do not interact—only three two-term interactions exist with non-parallel lines out of 28:

- (1) X1 and X2 (EyeAlg and IQMetr).
- (2) X1 and X3 (EyeAlg and EyeSeg).
- (3) X6 and X7 (FXMask and SMAIlg).

Thus, only a few of the eight algorithmic factors are interacting with one another to affect the VR|FAR = .01 VASIR matching performance response.

Table 7 summarizes the results from all three responses interaction effects analysis, one per response.

For this VV Left eye case, the X1X2 (EyeAlgIQMetr) interaction occurred for all three responses and the X1X7 (EyeAlgSMAIlg) interaction for only the VR|FAR = .10 and EER responses.

The most important interaction effects are between factor X1 (Eye position alignment) and X2 (Image quality metrics) for the VASIR VV Left case. The interaction effects with factor X1 (EyeAlg) appeared the most frequently, followed by X2 (IQMetr) and X7 (SMAIlg). The results show that X3 (SegEye), X4 (NorRes), X5 (FXWLBW), X6 (FXMask), and X8 (SMsh) barely have any interaction in the VASIR system.

Thus, we conclude that the VASIR system is a near-linear system—driven by main effects with virtually little effect from interactions. Such near-linearity suggests that algorithmic changes to optimize a particular factor of interest are unlikely to influence the effects of other algorithmic factors in the VASIR system (quasi-independence)—note that this conclusion may change depending on other remaining algorithmic factors and robustness conditions.

6.4. ROC (Receiver Operating Characteristic) curve

There are two common graphical procedures for assessing performance of biometric systems; DET (Detection Error Tradeoff) and ROC (Receiver Operating Characteristic). Because of its preponderance in the biometrics literature, we choose to use the alternative ROC representation for analyzing VASIR performance. ROC curve is a plot of FAR (or FP) on the x-axis against VR (or TP) on the y-axis represented parametrically as a function of the decision threshold [38]. In an ideal system, the FAR should be low and the VR should be high.

We now address the question as to how good the optimal settings are as compared to other possible settings for the VASIR system. Fig. 6 illustrates the comparison between the best setting (+1, -1, +1, +1, -1, -1, -2, +1) and the worst setting (-1, 0, -1, -1, +1, +1, -1) (as obtained from the Main Effects plot analysis) by plotting the corresponding ROC curves.

Based on the ROC curve, when the settings were changed from worst (orange line) to best (green line), VR at FAR = .01 in VASIR's performance increased approximately 12.8%. EER decreased approximately 9.2%.

These results show that factor settings do make a difference. This also demonstrates the efficiency and the power of the orthogonal experiment design approach for setting, characterizing, and optimizing multiple factors simultaneously.

7. Results

This section is a summary of the results for the factor effects, most important to least important factors, optimal settings, and performance improvement.

For this VV Left case, the ranked list of factors is (X2, X7, X3, X1, X8, X4, X6, X5) of which six factors: X2 through X4 are statistically significant; factors X6 and X5 are not statistically significant. Factor

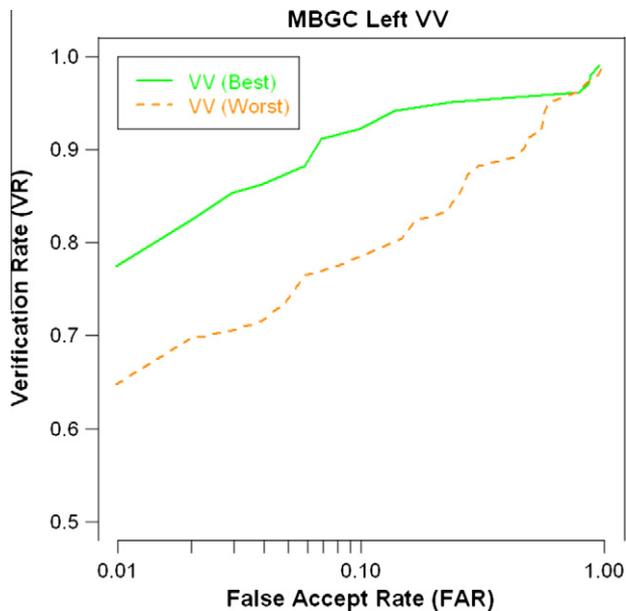


Fig. 6. Comparisons of the VASIR's best and worst settings for VV Left.

X2 (IQMetr) is image quality metrics and factor X7 (SMAIlg) is similarity metrics—these are the two most important algorithmic factors affecting VASIR performance—this is robustly true over all three responses. Factors X5 (FXWL: wavelength) and X6 (FXMask: wavelet magnitude) are universally the least important factors for the three responses. The overall improvement in the VR|FAR = .01 change from the worst to best settings of the eight factors was significant—VR at FAR = .01 in VASIR's performance increased 12.8% and EER decreased 9.2%. We demonstrated that the experiment design we employed was very efficient and effective in determining which factors are important, interaction effects, and optimal settings.

The eight algorithmic factors and our detailed findings that are new and unexpected are as follows:

(1) X2 (IQMetr): Image quality metrics.

Out of the eight examined factors, factor X2 (IQMetr) was the most important factor affecting VASIR performance for all three responses. Measuring image quality and then selecting the best quality image using quality metrics had a strong factor effect (e.g., effect = .041 at VR|FAR = .01) on VASIR's performance responses, and the effect was statistically significant. The three settings of IQMetr consist of SOB (Sobel), CON (Contrast) and LoG (Laplacian of Gaussian). SOB was introduced as a tool for face image quality evaluation by Beveridge et al. [21,18], CON was first introduced for general quality assessment by Albregtsen [25]; LoG was specifically used for the iris image quality assessment by Wan et al. [40]. For this VV Left case, we found that the SOB setting led to the best matching performance and was statistically different (better) than the other two IQMetr settings: CON (second best) and LoG (worst). The discrepancy between SOB and LoG is a bit surprising since both the SOB and LoG approaches measure the focus based on edge-based spatial domain metrics, while the CON approach measures the contrast based on a GLCM (Gray Level Co-occurrence Matrix). Hence, SOB gave the best results in our iris verification experiments. Unexpectedly, contrast (CON) was (slightly) better than the LoG operator. We also found that X2 (IQMetr) had significant interactions with factor X1 (EyeAlg: Eye Position Alignment). As expected, the best settings for the (EyeAlg, IQMetr) combination was (ON, SOB) and the worst setting was (OFF, LoG).

(2) X7 (SMAIlg): Similarity metrics.

The second most important factor affecting VASIR performance was SMAIlg (Similarity metric). The similarity metrics are important (effect = .038 at VR|FAR = .01) and statistically significant. The motivation for this factor is to identify one metric as best for VASIR performance.

The four settings for SMAIlg are HD (Hamming Distance), COR (Correlation), COS (Cosine) and WED (Weighted Euclidean Distance). Interestingly, for the VR|FAR = .01 response, WED had slightly better matching results than the other three metrics. On the other hand, for both VR|FAR = .10 and EER responses, HD led to much better matching performance than the other metrics. In summary, HD had the best matching results, followed by COR, WED, and COS. SMAIlg had minor interactions with factor X1 (EyeAlg: Eye position alignment) and factor X6 (FXWL: Wavelet magnitude).

(3) X3 (SegEye): Eyelids segmentation.

We found that the eyelids segmentation is an important factor (effect = .026 at VR|FAR = .01) and statistically significant for all three responses. This factor examines the effect of different algorithms for eyelids segmentation. The two settings for SegEye are "Lines" and "Curves." For all three responses, the use of two different curves for segmenting the upper and lower eyelids led to a better matching performance than the use of horizontal flat lines—this was expected. This conclusion reaffirmed that VASIR's eyelid segmentation approach [11] has a significant improvement over IrisBEE's approach [9,19]. The presence of parallel lines in the Interaction Effect Matrix for factor SegEye with the other seven factors (apart from minor interaction with factor X1 [EyeAlg: Eye position alignment]) indicates no interaction effect; hence SegEye is primarily driven by the main effect.

(4) X1 (EyeAlg): Eye position alignment.

Eye position alignment is also an important factor (effect = .020 at VR|FAR = .01) and statistically significant for all three responses. This factor addresses the difficulty in iris recognition from rotational differences between the target and query iris templates caused by tilted/rotated head with various movements. The two settings for EyeAlg are ON (correcting the inconsistency using an eye position alignment algorithm) and OFF (not correcting the inconsistency). VASIR had a better matching performance when the system used the ON setting to correct the rotational inconsistency between the two templates. EyeAlg had significant interactions with factor X2 (IQMetr: Image quality metrics), X3 (EyeSeg: Eyelids segmentation), and X7 (SMAIlg: Similarity metrics). As expected, the best setting for the (EyeAlg, IQMetr, EyeSeg, SMAIlg) combination was (ON, SOB, Curves, HD) and the worst setting was (OFF, LoG, Lines, COS).

(5) X8 (SMSh): Horizontal bit shifting for the HD_XorY approach.

Horizontal bit shifting is important (effect = .017 at VR|FAR = .01) and statistically significant. The motivation of this factor is to correct rotational inconsistencies after normalizing the target and query template. Masek [9] used 3 bits shifting for left and right for the LEI [41] dataset and 8 bits for the CASIA [42] dataset. IrisBEE used 10 bits shifting for the ICE2005 [19] dataset. For VASIR, we examined the use of 5 and 10 bits shifting for the MBGC video dataset. We found that 5 bits shifting had better matching performance than 10 bits shifting across all three responses. Note that factor SMSh had no interaction with other fac-

tors, hence this factor could be optimized independently without adversely affecting other factors.

(6) X4 (NorRes): Radial resolution for normalization.

The last statistically significant factor is radial resolution (effect = .011 at VR|FAR = .01). This factor addresses the question as to whether template size affects VASIR performance; the two settings for NorRes are 20 px and 32 px. We found that the larger radius resolution (32 px) had a better matching performance than the smaller resolution (20 px) across all three responses. NorRes does not interact with other factors.

(7) X6 (FXMask): Wavelet magnitude.

Wavelet magnitude is unimportant (effect = .006 at VR|FAR = .01) and statistically insignificant. This factor examines the effect of declaring different levels (%) of important bits in the iris template. The two settings of FXMask are 0.8 (80%) and 0.9 (90%). For the VV Left case, the 80% setting is slightly (but not significantly) better than the setting 90% for all three VASIR's responses. Note that FXMask has a minor interaction with X7 (SMAIlg: Similarity metrics).

(8) X5 (FXWL): Wavelength.

The least important factor in our study is wavelength (effect = .003 at VR|FAR = .01)—it is statistically insignificant. This factor examines the effect of varying the center frequency of the Log-Gabor filter (for feature extraction) on VASIR performance. The two settings of FXWL are 16 px and 18 px (the IrisBEE default). Our study found that 18px was slightly (but not significantly) better than 16 px—the difference is negligible. Masek [9] found that 18 px was also optimal for the CASIA dataset—although the author stated that the optimal value can be attributed to the different imaging conditions of each dataset. On the other hand, our orthogonal experiments showed that difference in effect between 16 px and 18 px for MBGC video is insignificant. In comparing our results to Masek's results, two principal differences emerge: (1) the Masek experiment varied a few (one or two) factors; the VASIR experiment varied $k = 8$ factors; and (2) the VASIR experiment design allowed for the existence of interactions. Both of these reasons support the contention that the VASIR conclusions are more global and robust. FXWL had no interaction effect with any other factors.

8. Conclusions

We introduced to the biometrics community a structured methodology for sensitivity analysis to foster an understanding of the key factors (parameters) in biometric systems.

This Sensitivity Analysis methodology consists of two components:

- (1) Experiment design in which we utilize efficient orthogonal fractional factorial designs to estimate not only the k main effects but also the $\binom{k}{2}$ two-term interactions of a biometric system.
- (2) Graphical data analysis in which we utilize three procedures: Main Effects plot, Ordered Data plot, and Interaction Effects matrix to determine important factors, two-term interactions, and optimal (local) settings.

To demonstrate the utility of this methodology, using the MBGC distant-video dataset, we chose a ($k = 8$ factor, $n = 384$

run) $(2^{6-1}) \times (3^1) \times (4^1)$ orthogonal fractional factorial experiment design for our VASIR system—investigating eight algorithmic factors (X1 to X8) to determine the most important, their optimal settings, and the relative importance of the 28 two-term interactions.

For this Video to Video (VV) Left eye case (the focus of this paper), our experiments showed that the three most important (see bold below) out of these eight algorithmic factors that we studied in VASIR were X2 (IQMetr: Image quality metrics) with factors X3 (SegEye: Eyelids segmentation) and X7 (SMAIlg: Similarity metrics) being next in importance. The least important factors were X5 (FXWL: FX wavelength) and X6 (FXMask: FX masking with magnitude). We found that the optimal settings were (+1, -1, +1, +1, -1, +1, -2, +1) with details as follows:

- X1 (EyeAlg): Eye position alignment (+1: ON),
- **X2 (IQMetr):** Image quality metrics (-1: Sobel operator [SOB]),
- **X3 (SegEye):** Eyelids segmentation (+1: Curves),
- X4 (NorRes): Radial resolution for normalization (+1: 32),
- X5 (FXWL): FX wavelength (-1: 18),
- X6 (FXMask): FX masking with magnitude (+1: 0.9),
- **X7 (SMAIlg):** Similarity metrics (-2: Hamming distance [HD]),
- X8 (SMsh): Horizontal shifting number (+1: 5).

In order of decreasing importance, the statistically significant factors that had an influence on VASIR performance were the six factors (X2, X7, X3, X1, X8, and X4). On the other hand, factors X5 and X6 were found to have barely any effect on VASIR's overall performance.

We found that some two-term interactions did in fact exist—they involved factors X1 (EyeAlg) (primarily), X2 (IQMetr), and X7 (SMAIlg)—in particular, the $X1 * X2$ and $X1 * X7$ interactions were found to be important. On the other hand, for our VV Left eye case, we found that most of the interactions had minor effect on VASIR performance—hence the effect on performance of VASIR's algorithmic component was mostly additive and independent. It is noteworthy that when the VASIR settings were changed from worst to best, VASIR's verification rate at FAR = .01 increased significantly (12.8%) and EER decreased significantly (9.2%).

In summary, the choice of the image quality metric for selecting the best quality image in video had the strongest effect on VASIR performance, followed by the choice of similarity metric. Our data analysis also reaffirmed that eyelid segmentation was important and that VASIR's approach (Curves) had a significant improvement over IrisBEE's approach (Lines). Further, VASIR's two factors for correcting rotational difference due to head tilt or subject movement (eye position alignment, Horizontal shifting) were both important. It is of interest to note that comparing across studies, we found that optimal value of bit shifting for correcting inconsistency depended on the dataset (or the different imaging conditions of the dataset). Finally, we found that the larger radius size of the iris template had a better matching performance than the smaller size. Given our eight algorithmic factors, we found that VASIR is a near-linear system; thus, optimization of a particular factor is unlikely to influence the effects of the other algorithmic factors. We believe that the sensitivity analysis methodology demonstrated herein can be applied to other biometric systems.

Based on our study, opportunities for future research would include as follows: (1) carrying out a follow-up experiment to ascertain the robustness of our conclusions over other scenarios (eye position (left/right), matching scenarios (VV: Video to Video, VS: Video to Still, and SS: Still to Still)); (2) replacement of the two unimportant factors with other VASIR key algorithm

factors; (3) reaffirming our conclusions by using larger datasets (e.g., more videos and subjects); and (4) applying the same sensitivity analysis methodology to simultaneously examine a considerably larger (e.g., $k=20$ factor) set of VASIR algorithmic factors.

Disclaimer

The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology (NIST).

Appendix A. The chosen eight algorithmic factors with their multiple levels (marked with gray) and the remaining 30 factors with their fixed levels

	Possible factors	-1 setting	+1 setting
EX	Eye det./ext. (EX) algorithms	Haar-wavelet	
	Classifier cascade types	Eye-pair	
	Minimum resolution for eye region	200x45	
	Width of nose bridge	Width/5	
	X1: EX Pupil position alignment	OFF	ON
IQA	X2: Image quality metrics algorithms	-1: SOB, 0: LoG1, +1: CON	
	Image enhancement/restoration	None	
BIS	Best image selection algorithms	ABIS	
		* Pupil circle	
SEG	Pupil segmentation algorithms	Contour processing	
	Data type	Distant-Video, Classic-still	
	Ratio between width and height	8, 12	
	Thresholds	Minimum +(25,35)	
	Scale	1,2	
	Radius limits	42*Scale	
	Closing iteration	0,2	
	Opening iterations	3	
		* Iris circle	
	Iris segmentation algorithms	Hough transform (circle)	
	Data type	Distant-Video,Classic-still	
	Thresholds	Low:0.11, High:0.19	
	Scale	1,2	
	Radius limits	82*Scale	
	Distance between the pupil center and the iris center	(82*Scale)/4	
	* Eyelids		
X3: Eyelids segmentation algorithms	Hough transform (lines)	Hough transform (curves)	
Adjustment	2px		
	* Noise removal		
Thresholding for eyelashes	3		
Thresholding for reflections	250		
NOR	Normalization algorithms	Polar coordinates with bilinear interpolation	
	X4: Radial resolution	20	32
	Angular resolution	240	
FX	Feature Extraction (FX) algorithms	1D-Gabor wavelet	
	X5: FX Wavelength (WL)-pixel unit	18px	16px
	FX Bandwidth (BW)	0.5	
	X6: FX Masking w/h Mag (below,above)	(0.01, 0.8)	(0.01, 0.9)
	FX Encoding	ON	
FX Encoding scale	1		
SM	X7: Similarity metrics (SM) algorithms	-2: HD, -1: COR, +1: COS, +2: WED	
	X8: SM Horizontal shifting (left/right)	10	5
	SM Vertical shifting (Up and down)	2	

Appendix B. $(2^{6-1}) \times (3^1) \times (4^1)$ orthogonal fractional factorial design matrix for eight algorithmic factors in VASIR

n/k	X1	X2	X3	X4	X5	X6	X7	X8	n/k	X1	X2	X3	X4	X5	X6	X7	X8	n/k	X1	X2	X3	X4	X5	X6	X7	X8
1	-1	-1	-1	-1	-1	-2	-1	-1	63	-1	0	1	1	1	-2	-1	-1	125	-1	-1	1	1	1	-1	-1	
2	-1	-1	-1	-1	-1	-2	1	-1	64	1	0	1	1	1	-2	1	-1	126	1	-1	1	1	1	1	-1	
3	-1	-1	-1	-1	-1	-2	1	1	65	-1	0	1	1	1	-2	1	1	127	-1	-1	1	1	1	1	-1	
4	-1	-1	-1	-1	-1	-2	1	1	66	1	0	1	1	1	-2	1	1	128	1	-1	1	1	1	1	-1	
5	-1	-1	-1	-1	-1	-2	1	1	67	-1	1	1	1	1	-2	1	-1	129	-1	0	1	1	1	1	-1	
6	-1	-1	-1	-1	-1	-2	1	1	68	1	1	1	1	1	-2	1	-1	130	1	0	1	1	1	1	-1	
7	-1	-1	-1	-1	-1	-2	1	1	69	-1	1	1	1	1	-2	1	1	131	-1	0	1	1	1	1	-1	
8	-1	-1	-1	-1	-1	-2	1	1	70	1	1	1	1	1	-2	1	-1	132	1	0	1	1	1	1	-1	
9	-1	-1	-1	-1	-1	-2	1	1	71	-1	1	1	1	1	-2	1	1	133	-1	0	1	1	1	1	-1	
10	-1	-1	-1	-1	-1	-2	1	1	72	1	1	1	1	1	-2	1	-1	134	1	0	1	1	1	1	-1	
11	-1	-1	-1	-1	-1	-2	1	1	73	-1	1	1	1	1	-2	1	1	135	-1	0	1	1	1	1	-1	
12	-1	-1	-1	-1	-1	-2	1	1	74	1	1	1	1	1	-2	1	-1	136	1	0	1	1	1	1	-1	
13	-1	-1	-1	-1	-1	-2	1	1	75	-1	1	1	1	1	-2	1	1	137	-1	0	1	1	1	1	-1	
14	-1	-1	-1	-1	-1	-2	1	1	76	1	1	1	1	1	-2	1	-1	138	1	0	1	1	1	1	-1	
15	-1	-1	-1	-1	-1	-2	1	1	77	-1	1	1	1	1	-2	1	1	139	-1	0	1	1	1	1	-1	
16	-1	-1	-1	-1	-1	-2	1	1	78	1	1	1	1	1	-2	1	-1	140	1	0	1	1	1	1	-1	
17	-1	-1	-1	-1	-1	-2	1	1	79	-1	1	1	1	1	-2	1	1	141	-1	0	1	1	1	1	-1	
18	-1	-1	-1	-1	-1	-2	1	1	80	1	1	1	1	1	-2	1	-1	142	1	0	1	1	1	1	-1	
19	-1	-1	-1	-1	-1	-2	1	1	81	-1	1	1	1	1	-2	1	1	143	-1	0	1	1	1	1	-1	
20	-1	-1	-1	-1	-1	-2	1	1	82	1	1	1	1	1	-2	1	-1	144	1	0	1	1	1	1	-1	
21	-1	-1	-1	-1	-1	-2	1	1	83	-1	1	1	1	1	-2	1	1	145	-1	0	1	1	1	1	-1	
22	-1	-1	-1	-1	-1	-2	1	1	84	1	1	1	1	1	-2	1	-1	146	1	0	1	1	1	1	-1	
23	-1	-1	-1	-1	-1	-2	1	1	85	-1	1	1	1	1	-2	1	1	147	-1	0	1	1	1	1	-1	
24	-1	-1	-1	-1	-1	-2	1	1	86	1	1	1	1	1	-2	1	-1	148	1	0	1	1	1	1	-1	
25	-1	-1	-1	-1	-1	-2	1	1	87	-1	1	1	1	1	-2	1	1	149	-1	0	1	1	1	1	-1	
26	-1	-1	-1	-1	-1	-2	1	1	88	1	1	1	1	1	-2	1	-1	150	1	0	1	1	1	1	-1	
27	-1	-1	-1	-1	-1	-2	1	1	89	-1	1	1	1	1	-2	1	1	151	-1	0	1	1	1	1	-1	
28	-1	-1	-1	-1	-1	-2	1	1	90	1	1	1	1	1	-2	1	-1	152	1	0	1	1	1	1	-1	
29	-1	-1	-1	-1	-1	-2	1	1	91	-1	1	1	1	1	-2	1	1	153	-1	0	1	1	1	1	-1	
30	-1	-1	-1	-1	-1	-2	1	1	92	1	1	1	1	1	-2	1	-1	154	1	0	1	1	1	1	-1	
31	-1	-1	-1	-1	-1	-2	1	1	93	-1	1	1	1	1	-2	1	1	155	-1	0	1	1	1	1	-1	
32	-1	-1	-1	-1	-1	-2	1	1	94	1	1	1	1	1	-2	1	-1	156	1	0	1	1	1	1	-1	
33	-1	0	-1	-1	-1	-2	-1	-1	95	-1	1	1	1	1	-2	-1	-1	157	-1	0	-1	1	1	1	-1	
34	-1	0	-1	-1	-1	-2	-1	-1	96	1	1	1	1	1	-2	-1	-1	158	1	0	-1	1	1	1	-1	
35	-1	0	-1	-1	-1	-2	-1	-1	97	-1	1	1	1	1	-2	-1	-1	159	-1	0	-1	1	1	1	-1	
36	-1	0	-1	-1	-1	-2	-1	-1	98	1	1	1	1	1	-2	-1	-1	160	1	0	-1	1	1	1	-1	
37	-1	0	-1	-1	-1	-2	-1	-1	99	-1	1	1	1	1	-2	-1	-1	161	-1	0	-1	1	1	1	-1	
38	-1	0	-1	-1	-1	-2	-1	-1	100	1	1	1	1	1	-2	-1	-1	162	1	0	-1	1	1	1	-1	
39	-1	0	-1	-1	-1	-2	-1	-1	101	-1	1	1	1	1	-2	-1	-1	163	-1	0	-1	1	1	1	-1	
40	-1	0	-1	-1	-1	-2	-1	-1	102	1	1	1	1	1	-2	-1	-1	164	1	0	-1	1	1	1	-1	
41	-1	0	-1	-1	-1	-2	-1	-1	103	-1	1	1	1	1	-2	-1	-1	165	-1	0	-1	1	1	1	-1	
42	-1	0	-1	-1	-1	-2	-1	-1	104	1	1	1	1	1	-2	-1	-1	166	1	0	-1	1	1	1	-1	
43	-1	0	-1	-1	-1	-2	-1	-1	105	-1	1	1	1	1	-2	-1	-1	167	-1	0	-1	1	1	1	-1	
44	-1	0	-1	-1	-1	-2	-1	-1	106	1	1	1	1	1	-2	-1	-1	168	1	0	-1	1	1	1	-1	
45	-1	0	-1	-1	-1	-2	-1	-1	107	-1	1	1	1	1	-2	-1	-1	169	-1	0	-1	1	1	1	-1	
46	-1	0	-1	-1	-1	-2	-1	-1	108	1	1	1	1	1	-2	-1	-1	170	1	0	-1	1	1	1	-1	
47	-1	0	-1	-1	-1	-2	-1	-1	109	-1	1	1	1	1	-2	-1	-1	171	-1	0	-1	1	1	1	-1	
48	-1	0	-1	-1	-1	-2	-1	-1	110	1	1	1	1	1	-2	-1	-1	172	1	0	-1	1	1	1	-1	
49	-1	0	-1	-1	-1	-2	-1	-1	111	-1	1	1	1	1	-2	-1	-1	173	-1	0	-1	1	1	1	-1	
50	-1	0	-1	-1	-1	-2	-1	-1	112	1	1	1	1	1	-2	-1	-1	174	1	0	-1	1	1	1	-1	
51	-1	0	-1	-1	-1	-2	-1	-1	113	-1	1	1	1	1	-2	-1	-1	175	-1	0	-1	1	1	1	-1	
52	-1	0	-1	-1	-1	-2	-1	-1	114	1	1	1	1	1	-2	-1	-1	176	1	0	-1	1	1	1	-1	
53	-1	0	-1	-1	-1	-2	-1	-1	115	-1	1	1	1	1	-2	-1	-1	177	-1	0	-1	1	1	1	-1	
54	-1	0	-1	-1	-1	-2	-1	-1	116	1	1	1	1	1	-2	-1	-1	178	1	0	-1	1	1	1	-1	
55	-1	0	-1	-1	-1	-2	-1	-1	117	-1	1	1	1	1	-2	-1	-1	179	-1	0	-1	1	1	1	-1	
56	-1	0	-1	-1	-1	-2	-1	-1	118	1	1	1	1	1	-2	-1	-1	180	1	0	-1	1	1	1	-1	
57	-1	0	-1	-1	-1	-2	-1	-1	119	-1	1	1	1	1	-2	-1	-1	181	-1	0	-1	1	1	1	-1	
58	-1	0	-1	-1	-1	-2	-1	-1	120	1	1	1	1	1	-2	-1	-1	182	1	0	-1	1	1	1	-1	
59	-1	0	-1	-1	-1	-2	-1	-1	121	-1	1	1	1	1	-2	-1	-1	183	-1	0	-1	1	1	1	-1	
60	-1	0	-1	-1	-1	-2	-1	-1	122	1	1	1	1	1	-2	-1	-1	184	1	0	-1	1	1	1	-1	
61	-1	0	-1	-1	-1	-2	-1	-1	123	-1	1	1	1	1	-2	-1	-1	185	-1	0	-1	1	1	1	-1	
62	-1	0	-1	-1	-1	-2	-1	-1	124	1	1	1	1	1	-2											

Appendix B (continued)

n/k	x1	x2	x3	x4	x5	x6	x7	x8	n/k	x1	x2	x3	x4	x5	x6	x7	x8	n/k	x1	x2	x3	x4	x5	x6	x7	x8
187	-1	-1	-1	-1	-1	-1	-1	-1	253	-1	0	-1	1	1	1	1	1	319	-1	-1	1	1	1	1	2	-1
188	1	1	1	1	1	1	1	1	254	1	0	-1	1	1	1	1	1	320	-1	-1	1	1	1	1	2	1
189	-1	-1	-1	-1	-1	-1	-1	-1	255	-1	0	1	1	1	1	1	1	321	-1	0	-1	-1	-1	-1	2	-1
190	1	1	1	1	1	1	1	1	256	1	0	1	1	1	1	1	1	322	1	0	-1	-1	-1	-1	2	1
191	-1	-1	-1	-1	-1	-1	-1	-1	257	-1	-1	-1	-1	-1	-1	-1	-1	323	-1	0	-1	-1	-1	-1	2	1
192	1	1	1	1	1	1	1	1	258	1	-1	-1	-1	-1	-1	-1	-1	324	1	0	-1	-1	-1	-1	2	-1
193	-1	-1	-1	-1	-1	-1	-1	-1	259	-1	1	1	1	1	1	1	1	325	-1	0	1	1	1	1	2	1
194	1	1	1	1	1	1	1	1	260	1	1	-1	-1	-1	-1	-1	-1	326	1	0	1	-1	-1	-1	2	-1
195	-1	-1	-1	-1	-1	-1	-1	-1	261	-1	1	-1	-1	-1	-1	-1	-1	327	-1	0	1	-1	-1	-1	2	-1
196	1	1	1	1	1	1	1	1	262	1	-1	-1	-1	-1	-1	-1	-1	328	1	0	1	-1	-1	-1	2	1
197	-1	-1	-1	-1	-1	-1	-1	-1	263	-1	1	1	1	1	1	1	1	329	-1	0	-1	-1	-1	-1	2	1
198	1	1	1	1	1	1	1	1	264	1	1	1	-1	-1	-1	-1	-1	330	1	0	-1	-1	-1	-1	2	-1
199	-1	-1	-1	-1	-1	-1	-1	-1	265	-1	-1	-1	-1	-1	-1	-1	-1	331	-1	0	-1	-1	-1	-1	2	1
200	1	1	1	1	1	1	1	1	266	1	-1	-1	-1	-1	-1	-1	-1	332	1	0	-1	-1	-1	-1	2	-1
201	-1	-1	-1	-1	-1	-1	-1	-1	267	-1	1	1	1	1	1	1	1	333	-1	0	1	1	1	1	2	1
202	1	1	1	1	1	1	1	1	268	1	-1	-1	-1	-1	-1	-1	-1	334	1	0	-1	-1	-1	-1	2	-1
203	-1	-1	-1	-1	-1	-1	-1	-1	269	-1	1	1	1	1	1	1	1	335	-1	0	1	1	1	1	2	1
204	1	1	1	1	1	1	1	1	270	1	-1	-1	-1	-1	-1	-1	-1	336	1	0	1	-1	-1	-1	2	-1
205	-1	-1	-1	-1	-1	-1	-1	-1	271	-1	1	1	1	1	1	1	1	337	-1	0	-1	-1	-1	-1	2	1
206	1	1	1	1	1	1	1	1	272	1	-1	-1	-1	-1	-1	-1	-1	338	1	0	-1	-1	-1	-1	2	-1
207	-1	-1	-1	-1	-1	-1	-1	-1	273	-1	1	1	1	1	1	1	1	339	-1	0	1	1	1	1	2	1
208	1	1	1	1	1	1	1	1	274	1	-1	-1	-1	-1	-1	-1	-1	340	1	0	-1	-1	-1	-1	2	-1
209	-1	-1	-1	-1	-1	-1	-1	-1	275	-1	1	1	1	1	1	1	1	341	-1	0	-1	-1	-1	-1	2	1
210	1	1	1	1	1	1	1	1	276	1	-1	-1	-1	-1	-1	-1	-1	342	1	0	-1	-1	-1	-1	2	-1
211	-1	-1	-1	-1	-1	-1	-1	-1	277	-1	1	1	1	1	1	1	1	343	-1	0	1	1	1	1	2	1
212	1	1	1	1	1	1	1	1	278	1	-1	-1	-1	-1	-1	-1	-1	344	1	0	-1	-1	-1	-1	2	-1
213	-1	-1	-1	-1	-1	-1	-1	-1	279	-1	1	1	1	1	1	1	1	345	-1	0	-1	-1	-1	-1	2	1
214	1	1	1	1	1	1	1	1	280	1	-1	-1	-1	-1	-1	-1	-1	346	1	0	-1	-1	-1	-1	2	-1
215	-1	-1	-1	-1	-1	-1	-1	-1	281	-1	1	1	1	1	1	1	1	347	-1	0	1	1	1	1	2	1
216	1	1	1	1	1	1	1	1	282	1	-1	-1	-1	-1	-1	-1	-1	348	1	0	-1	-1	-1	-1	2	-1
217	-1	-1	-1	-1	-1	-1	-1	-1	283	-1	1	1	1	1	1	1	1	349	-1	0	-1	-1	-1	-1	2	1
218	1	1	1	1	1	1	1	1	284	1	-1	-1	-1	-1	-1	-1	-1	350	1	0	-1	-1	-1	-1	2	-1
219	-1	-1	-1	-1	-1	-1	-1	-1	285	-1	1	1	1	1	1	1	1	351	-1	0	1	1	1	1	2	1
220	1	1	1	1	1	1	1	1	286	1	-1	-1	-1	-1	-1	-1	-1	352	1	0	-1	-1	-1	-1	2	-1
221	-1	-1	-1	-1	-1	-1	-1	-1	287	-1	1	1	1	1	1	1	1	353	-1	0	1	1	1	1	2	1
222	1	1	1	1	1	1	1	1	288	1	-1	-1	-1	-1	-1	-1	-1	354	1	0	-1	-1	-1	-1	2	-1
223	-1	-1	-1	-1	-1	-1	-1	-1	289	-1	1	1	1	1	1	1	1	355	-1	0	1	1	1	1	2	1
224	1	1	1	1	1	1	1	1	290	1	-1	-1	-1	-1	-1	-1	-1	356	1	0	-1	-1	-1	-1	2	-1
225	-1	-1	-1	-1	-1	-1	-1	-1	291	-1	1	1	1	1	1	1	1	357	-1	0	1	1	1	1	2	1
226	1	1	1	1	1	1	1	1	292	1	-1	-1	-1	-1	-1	-1	-1	358	1	0	-1	-1	-1	-1	2	-1
227	-1	-1	-1	-1	-1	-1	-1	-1	293	-1	1	1	1	1	1	1	1	359	-1	0	1	1	1	1	2	1
228	1	1	1	1	1	1	1	1	294	1	-1	-1	-1	-1	-1	-1	-1	360	1	0	-1	-1	-1	-1	2	-1
229	-1	-1	-1	-1	-1	-1	-1	-1	295	-1	1	1	1	1	1	1	1	361	-1	0	1	1	1	1	2	1
230	1	1	1	1	1	1	1	1	296	1	-1	-1	-1	-1	-1	-1	-1	362	1	0	-1	-1	-1	-1	2	-1
231	-1	-1	-1	-1	-1	-1	-1	-1	297	-1	1	1	1	1	1	1	1	363	-1	0	1	1	1	1	2	1
232	1	1	1	1	1	1	1	1	298	1	-1	-1	-1	-1	-1	-1	-1	364	1	0	-1	-1	-1	-1	2	-1
233	-1	-1	-1	-1	-1	-1	-1	-1	299	-1	1	1	1	1	1	1	1	365	-1	0	1	1	1	1	2	1
234	1	1	1	1	1	1	1	1	300	1	-1	-1	-1	-1	-1	-1	-1	366	1	0	-1	-1	-1	-1	2	-1
235	-1	-1	-1	-1	-1	-1	-1	-1	301	-1	1	1	1	1	1	1	1	367	-1	0	1	1	1	1	2	1
236	1	1	1	1	1	1	1	1	302	1	-1	-1	-1	-1	-1	-1	-1	368	1	0	-1	-1	-1	-1	2	-1
237	-1	-1	-1	-1	-1	-1	-1	-1	303	-1	1	1	1	1	1	1	1	369	-1	0	1	1	1	1	2	1
238	1	1	1	1	1	1	1	1	304	1	-1	-1	-1	-1	-1	-1	-1	370	1	0	-1	-1	-1	-1	2	-1
239	-1	-1	-1	-1	-1	-1	-1	-1	305	-1	1	1	1	1	1	1	1	371	-1	0	1	1	1	1	2	1
240	1	1	1	1	1	1	1	1	306	1	-1	-1	-1	-1	-1	-1	-1	372	1	0	-1	-1	-1	-1	2	-1
241	-1	-1	-1	-1	-1	-1	-1	-1	307	-1	1	1	1	1	1	1	1	373	-1	0	1	1	1	1	2	1
242	1	1	1	1	1	1	1	1	308	1	-1	-1	-1	-1	-1	-1	-1	374	1	0	-1	-1	-1	-1	2	-1
243	-1	-1	-1	-1	-1	-1	-1	-1	309	-1	1	1	1	1	1	1	1	375	-1	0	1	1	1	1	2	1
244	1	1	1	1	1	1	1	1	310	1	-1	-1	-1	-1	-1	-1	-1	376	1	0	-1	-1	-1	-1	2	-1
245	-1	-1	-1	-1	-1	-1	-1	-1	311	-1	1	1	1	1	1	1	1	377	-1	0	1	1	1	1	2	1
246	1	1	1	1	1	1	1	1	312	1	-1	-1	-1	-1	-1	-1	-1	378	1	0	-1	-1	-1	-1	2	-1
247	-1	-1	-1	-1	-1	-1	-1	-1	313	-1	1	1	1	1	1	1	1	379	-1	0	1	1	1	1	2	1
248	1	1	1	1	1	1	1	1	314	1	-1	-1	-1	-1	-1	-1	-1	380	1	0	-1	-1	-1	-1	2	-1
249	-1	-1	-1	-1	-1	-1	-1	-1	315	-1	1	1	1	1	1	1	1	381	-1	0	1	1	1	1	2	1
250	1	1	1	1	1	1	1	1	316	1	-1	-1	-1	-1	-1	-1	-1	382	1	0	-1	-1	-1	-1	2	-1
251	-1	-1	-1	-1	-1	-1	-1	-1	317	-1	1	1	1	1	1	1	1	383	-1	0	1	1	1	1	2	1
252	1	1	1	1	1	1	1	1	318	1	-1	-1	-1	-1	-1	-1	-1	384	1	0	-1	-1	-1	-1	2	-1

References

[1] ISO/IEC 2382-37:2012(E) Information technology—Vocabulary—Part 37: Biometrics, International Organization for Standardization and International Electrotechnical Committee, December 2012.

[2] A.K. Jain, P.J. Flynn, A.A. Ross, Introduction to biometrics, in: Handbook of Biometrics, 2008.

[3] G. Stirling Andrew, On science and precaution in the management of technological risk, Eur. Sci. Technol. Observatory (2002).

[4] G.M. Hornberger, S.R.C., An approach to the preliminary analysis of environmental systems, J. Environ. Manage. 12 (1981) 7–18.

[5] Y. Zhang, D.B. Goldgof, S. Sarkar, L.V. Tsap, A sensitivity analysis method and its application in physics-based nonrigid motion modeling, Image Vision Comput. 25 (3) (2007) 262–273.

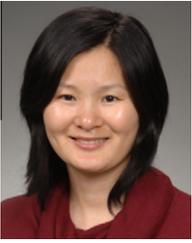
[6] Y. Lu, S. Payandeh, On the sensitivity analysis of camera calibration from images of spheres, Comput. Vision Image Understand. 114 (1) (2010) 8–20.

[7] K.L. Mills, J.J. Filliben, C. Dong Yeon, S. Edward, G. Daniel, Sensitivity Analysis of MesoNet, Study of Proposed Internet Congestion-Control Mechanisms, 2010, <http://www.nist.gov/itl/antd/Congestion_Control_Study.cfm> (accessed 09.03.11). Chapter 4.

[8] A. Saltelli, Sensitivity analysis for importance assessment, Risk Anal. 22 (3) (2002) 579–590.

[9] L. Masek, P. Kovetski, MATLAB Source Code for a Biometric Identification System based on Iris Patterns, School of Computer Science and Software Engineering, University of Western Australia, 2003.

[10] K. Messer, J. Kittler, M. Sadeghi,



Yooyoung Lee is a guest researcher at NIST and primary focuses on video-based biometrics and evaluations. She received her Ph.D. in Computer Engineering from Chung-Ang University, Seoul, Korea. She received the Associate of the Year (2012) Award from the Information Technology Laboratory (ITL) at NIST for analyzing and evaluating the performance of biometric algorithms. She developed Video-based Automatic System for Iris Recognition (VASIR) as a benchmark baseline. Her research interests are video-based human recognition, biometrics, performance evaluations, computer vision, pattern recognition, image processing, and sensitivity analysis.



James J. Filliben is a mathematical statistician at NIST. He received his Ph.D. in statistics from Princeton University (1969). He is the author of more than 100 papers, the developer of a widely-cited test statistic for normality, a Fellow of the American Statistical Society, and the 2003 recipient of the ASA Youden Award (Interlab Testing). He is the author of Dataplot: an extensive software system for statistical graphics and modeling. His NIST and inter-agency contributions have been recognized with Department of Commerce Gold (3), Silver (1), and Bronze (3) medals. His research interests include exploratory data analysis, statistical graphics, distributional modeling, and experiment design.



Ross J. Micheals leads an effort researching biometric client technologies with a particular focus on web services & usable next-generation interfaces for acquisition. Ross also pioneered NIST's research into the usability of biometric systems, an initiative in which he is still involved with today. Most recently, he served as the acting deputy director for the National Strategy for Trusted Identities in Cyberspace (NSTIC) Program Office. He earned his Ph.D. in computer science from Lehigh University, and has also worked in the field of computer vision for both Texas Instruments and Carnegie Mellon's Robotics Institute. Since 1998, he has collaborated with



30 co-authors and has been cited by nearly 600 researchers. Along with his colleagues, he was awarded a Department of Commerce Gold Medal in 2003 for his work on biometric system performance assessment.

Jonathon Phillips is a leading technologist in the fields of computer vision, biometrics, and face recognition. He is at NIST, where he runs challenge problems and evaluations to advance biometric technology. His previous efforts include the Iris Challenge Evaluations (ICE), the Face Recognition Vendor Test (FRVT) 2006 and the Face Recognition Grand Challenge and FERET. From 2000–2004, Dr. Phillips was assigned to DARPA. For his work on the FRVT 2002 he was awarded the Department of Commerce Gold Medal. His work has been reported in the New York Times and the Economist. He has appeared on NPR's Science Friday show. He is a fellow of the IEEE and IAPR.