Sensitivity Analysis for Biometric Systems: A Methodology Based on Orthogonal Experiment Designs

Yooyoung Lee
James J. Filliben
Ross J. Micheals
P. Jonathon Phillips

1

Sensitivity Analysis for Biometric Systems: A Methodology Based on Orthogonal Experiment Designs

Yooyoung Lee
Ross J. Micheals
P. Jonathon Phillips
Information Access Division
Information Technology Laboratory

James J. Filliben
Statistical Engineering Division
Information Technology Laboratory

National Institute of Standards and Technology
Gaithersburg, MD 20899

May 2012

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 the insight and understanding of the key factors—algorithmic, subject-based, camera-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 allows 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, DNA, etc.

Keywords: sensitivity analysis, iris recognition, biometrics, experiment design, VASIR, video-based biometric system, orthogonal, fractional factorial, characterization, uncertainty, factor effects, interaction effects, optimization, main effects plots, interaction plots

1. Introduction

Biometrics is the study of identifying an individual based on human physiological or behavioral characteristics. 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 attractive 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 [1].

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 (and therefore confidence), 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 [2]. Hornberger and Spear [3] 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 in which they can be made to produce virtually any desired behavior, often with both plausible structure and "reasonable" parameter values. We believe that the above statement applies equally well for biometric systems.

Sensitivity analysis is the study of how the output of a system is affected by different inputs to the system [4], while optimization analysis is the determination of system parameter settings that yield the best system performance. 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 held constant while one factor is focused on and varied to examine its effect. This design, though popular [5,6], 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.

This is the motivation to introduce to the biometrics community an alternative method for conducting a sensitivity analysis. We demonstrate the method on the VASIR iris recognition system to yield more accurate and reliable results with the advantage that:

1) The system is better understood;
2) The system is better characterized;
3) The system is better optimized—with the 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 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 [7].

The general structure of this paper is three-fold:

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 [8,9].

2) *Sensitivity 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 <u>ranked list</u> 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 estimable, and such estimates should have as small as possible bias and sampling uncertainty.

As it turns out, such minimal 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 <u>capacity</u> 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).

Both parts are important, and when optimally used in concert, they yield copious insight into the relative importance of the various computational components (and interactions) that affect biometric system performance. 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 (see the details in Section 3).

2.1 Experiment Design

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

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 [10];

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 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 <u>randomly</u> samples a user-specified number from that population space (see details in Kleijnen [11]). 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 in succession is changed to an alternative value. 1FAT designs are simple, intuitive, and inexpensive (costing as little as $n = 1+k$ runs). On the other hand, we will demonstrate below that such designs routinely produce effect estimates (and hence the subsequent ranked list of factors) which can be significantly biased and imprecise (see Box et al [12] and Saltelli et al [4]). Further, 1FAT designs have no ability whatsoever to estimate interactions—which are quite common in biometric systems. Of course, such interactions will make themselves evident by contaminating (biasing) the one thing that 1FAT designs do produce—estimates of the k main factors.

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 [10]). The advantage of a full factorial design is that it is information-rich—such a design will provide 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. In terms of information content, it is difficult to go wrong with a full factorial design. On the other hand, the significant disadvantage of full factorial designs (which prevents its widespread use for large-k systems) 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 [10]). Designs with both 1- and 2-dimensional balance are referred to as "orthogonal designs", 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 in 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 in the n runs. Note that for a k-factor design to be orthogonal, the design must have 1-dimensional balance for each (and every) one of the k factors under study, and must have 2-dimensional balance for each (and every) one of the $\binom{k}{2}$ combinatorial pairs of factors in the design.

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 (e.g., -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. To retain the estimation virtues of orthogonality while avoiding the run-expense of full factorial designs, leads one to consider the use of orthogonal <u>fractional</u> 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 is a fractional design—but a decidedly <u>non-orthogonal</u> fractional design and so has poor estimation properties.

As alluded to above, estimates from an orthogonal factorial design are far superior (unbiased and precise) to estimates from the 1FAT design. To demonstrate this, we draw on an example from the chemical engineering industry where our response variable of interest is percentage reacted (0 % to 100 %) for an industrial chemical process [13]. The k=5 factors for this example are feed rate, catalyst, agitation rate, temperature, and concentration. For simplicity these five factors have only two levels. The sensitivity analysis goal is the usual: determine which of the five factors are most important in affecting percentage reacted. We carry out four separate sensitivity analyses of this k=5 chemical reaction process—using four different experiment designs. Figure 1 shows these four designs—three orthogonal designs and the 1FAT design:

1) Two-level orthogonal full factorial design (2^5, $n = 32$)—expensive,
2) Two-level orthogonal fractional factorial design (2^{5-1}, $n = 16$)—half expensive,
3) Two-level orthogonal fractional factorial design (2^{5-2}, $n = 8$)—inexpensive, and
4) Two-level 1-Factor-at-A-Time (1FAT) design $(n = 6)$—inexpensive.

Note that one might argue that 1FAT designs are poor due to the smaller number of runs. This, however, is not the case as may be seen by the 2^{5-2} orthogonal design—consisting of a similar small number of runs (8)—and yet having vastly superior estimation properties over the 1FAT. The left-most column in Figure 1 has the name of the design and the number of runs (n = 32, 16, 8, and 6, respectively), The second-left panel has the corresponding experimental design matrix (details omitted) along with a response column "Y" resulting from a behind-the-scenes simulation model with known factor and interaction coefficients (see details in the book written by Box et al [13]). Given the generic sensitivity analysis question: determine the most important factors in this k=5 factor experiment", the rightmost panel gives the "answer"—the ranked list that results after the design, data collection, and data analysis. Note that the behind-the-scenes simulation model had the property that the most important factor was X2 (with an effect size of 19.5), followed by the next most important factor (a 2-term X2*X4 interaction with value 13.25), etc. We assess the relative merit of the four designs by comparing their resulting ranked list estimates with this known ground truth ($B_2 = 19.5, B_{24} = 13.25, \dots$).

Design Name	Design Tableau	Effect Estimators	
2^5 Full Factorial Design $n=32$	X1 X2 X3 X4 X5 (two blocks of + / − rows) with Y values	$\hat{\beta}_2 = 19.50$ $\hat{\beta}_{24} = 13.25$ $\hat{\beta}_{45} = -11.00$ $\hat{\beta}_4 = 10.75$ $\hat{\beta}_5 = -6.25$ $\hat{\beta}_1 = -1.38$ $\hat{\beta}_3 = -0.63$	$\hat{\beta}_2 = 19.5$
2^{5-1} Fractional Factorial Design $n=16$	X1 X2 X3 X4 X5 with Y values	$\hat{\beta}_2 = 20.50$ $\hat{\beta}_4 = 12.25$ $\hat{\beta}_{24} = 10.75$ $\hat{\beta}_{45} = -9.50$ $\hat{\beta}_5 = -6.25$ $\hat{\beta}_1 = -2.00$ $\hat{\beta}_3 = 0.00$	$\hat{\beta}_2 = 20.5$
2^{5-2} Fractional Factorial Design $n=8$	X1 X2 X3 X4 X5 — Y: 44, 53, 70, 93, 66, 55, 54, 82	$\hat{\beta}_2, \hat{\beta}_{14} = 20.25$ $\hat{\beta}_4, \hat{\beta}_{12} = 13.25$ $\hat{\beta}_{23}, \hat{\beta}_{45} = -12.75$ $\hat{\beta}_1, \hat{\beta}_{24}, \hat{\beta}_{35} = 12.25$ $\hat{\beta}_{25}, \hat{\beta}_{34} = 6.25$ $\hat{\beta}_5, \hat{\beta}_{13} = -3.75$ $\hat{\beta}_3, \hat{\beta}_{15} = -0.75$	$\hat{\beta}_2 = 20.25$
1FAT 1-Factor-at-a-Time Design $n=6$	X1 X2 X3 X4 X5 — Y: 61, 53, 63, 53, 69, 56	$\hat{\beta}_1 = -8.00$ $\hat{\beta}_3 = -8.00$ $\hat{\beta}_4 = 8.00$ $\hat{\beta}_5 = -5.00$ $\hat{\beta}_2 = 2.00$	$\hat{\beta}_2 = 2.0$

Figure 1 Comparison of three orthogonal designs and the 1FAT design determining the most important factors out of five factors

In particular, we can go to the top of the four ranked lists and extract the "most important factor" as estimated by each of the four particular experiment designs and compare it with the known truth. In doing so we see that:

1) The expensive ($n=32$) 2^5 full factorial design identified factor X2 as its most important factor with an estimated effect size of $\hat{B}_2 = 19.5$; this was identical to the $B_2 = 19.5$ coefficient in the underlying simulation function.

2) The half-as-expensive ($n=16$) 2^{5-1} orthogonal fractional factorial design also determined X2 as the most important factor and estimated its effect as $\hat{B}_2 = 20.5$—close to 19.5 but with only half the experimental expense/effort.

3) The inexpensive ($n=8$) 2^{5-2} orthogonal fractional factorial design also chose X2 as the most important factor and estimated its effect as $\hat{B}_2 = 20.25$—also close (but with an issue of confounding with the X1*X4 effect which must be examined).

4) On the other hand, the least expensive ($n=6$) 1FAT fractional factorial design determined (incorrectly) that X1 was the most important factor and—even worse—found that X2 was the least important factor with its estimated effect being only $\hat{B}_2 = 2.0$—incorrect (biased) by an order of magnitude from the truth.

The two-level orthogonal fractional factorial designs were remarkably accurate, insightful, and extremely efficient, while the 1FAT design yielded grossly biased estimates and incorrect sensitivity analysis conclusions. The above illustration is not necessarily the exception. In a scientific/biometric experimental situation, if the performance response function has any interactions whatsoever, then orthogonal fractional factorial designs typically yield excellent, trusted sensitivity analysis conclusions, while 1FAT designs typically yield incorrect and skewed conclusions. Our proposed sensitivity analysis approach acknowledges this superiority, and thus uses orthogonal fractional factorial designs as the centerpiece for its 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 [12]), 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: 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 loss of feeling as to whether the resulting conclusions are consistent with the data.

For this reason, we find that a superior (in regard to maximizing insight) approach to the analysis of data from 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 (judiciously constructed and augmented) plots. In this regard, we have constructed 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 [8,14,15] and is an integral part of the NIST-developed analysis tool Dataplot [14,16].

The battery consists of the following ten graphical procedures: 1) Ordered Data plots; 2) Dex (design of experiment) Scatter plots; 3) Main Effects plots; 4) Dex Interaction Effects Matrix; 5) Block plots; 6) Dex Youden plots; 7) Effects plots; 8) Half-normal Probability plots; 9) Cumulative Residual Standard plots; and 10) Contour plots. A full application of all ten of these graphical tools for this biometrics example is unneeded here and is beyond the scope of this paper—it will suffice for us to demonstrate the power of the Sensitivity Analysis methodology to utilize three of the ten tools: 1) the Ordered Data plot; 2) the Main Effects 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) [17] algorithm—was implemented in the C programming language from Masek's Matlab code [5]. 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 [7,18] 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.

Although VASIR was developed with less-constrained video-based iris recognition in mind, VASIR also robustly accommodates the constrained still-based iris recognition. VASIR supports multiple matching scenarios:

1) Video (at a distance) to Video (at a distance) (VV),
2) Video (at a distance) to Still (classical) (VS), and
3) Still (classical) to Still (classical) (SS) iris recognition.

In VV matching, the extracted iris region from video frames is matched to other frames from the same video, or to frames from a different video sequence of the same person. VS means that the video captured at a distance is compared to classical still images, captured by a different device. SS matching is a traditional iris recognition scenario in which a classical still-image is matched against other classical still-images of the same person that were captured by the same (or similar) device.

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 Figure 2, the VASIR system can principally be categorized into three modules: 1) image acquisition, 2) video processing, and 3) iris recognition.
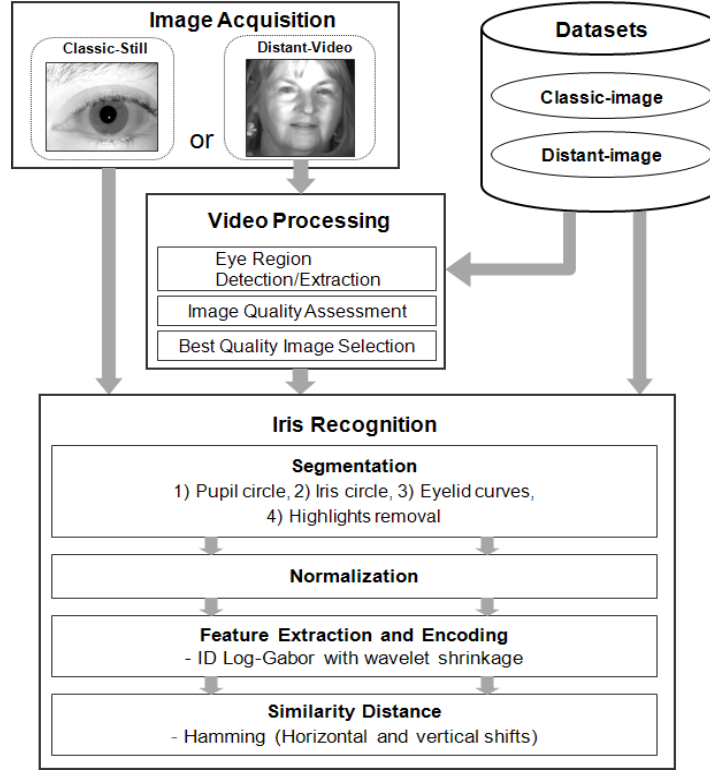
Figure 2 VASIR system architecture

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 quality score, 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 features 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 [7] 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.

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 <u>multiple</u> 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 33 factors from the first three categories, which could be examined unto themselves to assess the robustness of our algorithmic conclusions.

Table 1 Possible robustness factors that affect a biometric system

Environmental conditions	Image conditions	Subject characteristics
1) indoor/outdoor 2) lighting 3) background 4) weather 5) camera 6) day/night	7) focus or sharpness 8) contrast 9) brightness 10) motion blur 11) resolution 12) noise 13) color 14) distortion 15) artifacts 16) reflections 17) video /still (scenarios)	18) contact lenses 19) glasses 20) left/right 21) gender 22) age 23) cosmetics (e.g. mascara) 24) hair color 25) race 26) movement 27) distance from camera 28) skin color 29) eye color 30) eyelashes 31) eyebrows 32) pupil dilation 33) usability (e.g., behavior, training, perception)

Table 2 illustrates a list of 38 algorithmic factors (VASIR-specific) which potentially affect VASIR performance.

Table 2 VASIR possible algorithmic factors 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
Eye region detection and extraction (FX)	Algorithms	D	Haar,PCA, …	10
	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	X2: Algorithms	D	Sobel, LoG, CON,…	15
	Image enhancement/restoration	D	Super-resolutions,…	5
Best Quality Image Selection	Algorithms	D	ABIS,…	3
Segmentation (SEG)	* Pupil circle			
	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 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)

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 Table 1 and Table 2, k would equal 71 (=33+38). In general, the number of runs n must exceed the number of factors k to obtain estimates

for the k=71 factors and 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 factors of VASIR, we thus choose to focus primarily on only the eight specific algorithmic factors.

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

To accommodate affordability, the second step was to reduce the number of factors k. We hence choose to limit ourselves to only k=8 algorithmic factors (one or two factors from each of the eight key components in Table 2); we highlighted these eight algorithmic factors in Table 2 with gray. The reason why we have chosen these eight algorithmic factors is so that at least one factor was varied for each VASIR key component (excepting the Best Image Quality Selection component whose setting is dictated by the choice utilized for preceeding 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 setting levels of the eight algorithmic factors. Table 3 summarizes the eight chosen algorithmic factors and their setting levels.

Table 3 The eight algorithmic factors and their setting levels (see detailed procedures in the paper [7])

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 (SMAlg): Similarity metrics	-2 (HD)	-1 (COR)		+1 (COS)	+2 (WED)
X8 (SMSh): Horizontal shifting		-1 (10)		+1 (5)	

The detailed methods for each factor are described 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 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 [7] were introduced to measure one or another aspect of image quality, we focused on the following three 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 [19,20,21]. The gradient at each point and the orientation of that gradient (gradient magnitude) can be measured by:

   $$SOB = \sqrt{(Dx(x,y))^2 + (Dy(x,y))^2}$$

   The Sobel operator consists of a pair of 3×3 filters defined as Dx (left) and Dy (right). These filters are designed to respond to edges running vertically and horizontally relative to the pixel grid.

   B. Laplacian of Gaussian (LoG) filter

   LoG is an important filter with much attention given to it [22,23]. 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 9x9 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 [24,25].

14

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

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 by Albregtsen [24]). A higher contrast is considered as an indicator of better quality iris image in our study.

3) X3 (SegEye):Eyelids segmentation

In the IrisBEE algorithm [5], the eyelids were removed by inserting horizontal <u>flat lines</u> 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 <u>different curves</u> to segment the actual upper and lower eyelid shape. We examined how the two 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 [26]) 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:

$$I(x(r,\theta),y(r,\theta)) \rightarrow I(r,\theta)$$
$$x(r,\theta) = (1-r)x_p(\theta) + rx_s(\theta)$$
$$y(r,\theta) = (1-r)y_p(\theta) + ry_s(\theta)$$

(1)

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 (θ =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 two settings for radial size: 20 and 32 pixels.

5) X5 (FXWL): Wavelength

The IrisBEE algorithm employed a 1D Log-Gabor filter—introduced by Yao et al. [27]—to process the feature extraction from normalized iris images. The frequency response of a Log-Gabor filter is given as:

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

where w_0 represents the filter's center frequency (wavelength) and σ gives the filter bandwidth. We used different wavelengths (18px, 16px) 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 [26]. 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 [28]. The paper written by Lee [7] suggested that VASIR's approach based on the distance (magnitude) of the coefficient from the origin is superior than the Hallingsworth's fragile bit approach [28] 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 80 % (vs. 90 %) of the bits before computing the similarity distance.

7) X7 (SMAlg): 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 [29,30,31]. This section compares and examines four similarity distance metrics (HD, COR, COS, WED):

A. Hamming Distance (HD)

The Hamming Distance counts the number of positions, given two templates of the same size, that binary digits (0 and 1) are different [32]. It 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 [29] and later re-implemented by Masek [5]. 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:

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

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 $[-1, 1]$.

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

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:

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

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 [5,29]; 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 two 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 similarity distance values for matching and non-matching 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)—the detailed definition is described in the papers [7,33]. Symbolically the three responses are

        Y1 = VR|FAR = .01 (or TAR|FAR = .01)

        Y2 = VR|FAR = .10 (or TAR|FAR = .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 to 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, 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 [12] (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 Figure 3—note that since some of the eight factors are more time-consuming to change than others, the run order of the above design was optimized to minimize execution time (a 4x speed-up was achieved).

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	-1	-2	-1	63	-1	0	1	1	1	1	-2	-1	125	-1	-1	-1	1	1	1	-1	1
2	1	-1	-1	-1	-1	-1	-2	1	64	1	0	1	1	1	1	-2	1	126	1	-1	-1	1	1	1	-1	-1
3	-1	-1	1	-1	-1	-1	-2	1	65	-1	1	-1	-1	-1	-1	-2	-1	127	-1	-1	1	1	1	1	-1	-1
4	1	-1	1	-1	-1	-1	-2	-1	66	1	1	-1	-1	-1	-1	-2	1	128	1	-1	1	1	1	1	-1	1
5	-1	-1	-1	1	-1	-1	-2	1	67	-1	1	1	-1	-1	-1	-2	1	129	-1	0	-1	-1	-1	-1	-1	-1
6	1	-1	-1	1	-1	-1	-2	-1	68	1	1	1	-1	-1	-1	-2	-1	130	1	0	-1	-1	-1	-1	-1	1
7	-1	-1	1	1	-1	-1	-2	-1	69	-1	1	-1	1	-1	-1	-2	1	131	-1	0	1	-1	-1	-1	-1	1
8	1	-1	1	1	-1	-1	-2	1	70	1	1	-1	1	-1	-1	-2	-1	132	1	0	1	-1	-1	-1	-1	-1
⋮									⋮									⋮								
245	-1	0	-1	1	-1	1	1	-1	311	-1	-1	1	1	-1	1	2	1	377	-1	1	-1	-1	1	1	2	-1
246	1	0	-1	1	-1	1	1	1	312	1	-1	1	1	-1	1	2	-1	378	1	1	-1	-1	1	1	2	1
247	-1	0	1	1	-1	1	1	1	313	-1	-1	-1	-1	1	1	2	-1	379	-1	1	1	-1	1	1	2	1
248	1	0	1	1	-1	1	1	-1	314	1	-1	-1	-1	1	1	2	1	380	1	1	1	-1	1	1	2	-1
249	-1	0	-1	-1	1	1	1	-1	315	-1	-1	1	-1	1	1	2	1	381	-1	1	-1	1	1	1	2	1
250	1	0	-1	-1	1	1	1	1	316	1	-1	1	-1	1	1	2	-1	382	1	1	-1	1	1	1	2	-1
251	-1	0	1	-1	1	1	1	1	317	-1	-1	-1	1	1	1	2	1	383	-1	1	1	1	1	1	2	-1
252	1	0	1	-1	1	1	1	-1	318	1	-1	-1	1	1	1	2	-1	384	1	1	1	1	1	1	2	1

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

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, 32/4=8; this is referred to as 2-dimensional balance.

For the $(2^{6-1}) \times (3^1) \times (4^1)$ design, we should obtain near-precise estimates of main effects and two-term interactions—rivaling the quality of a $(2^6) \times (3^1) \times (4^1)$ full factorial design—but with only half (384) the number of runs.

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) [34]. 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 challenges

for the MBGC dataset is to recognize a person from the NIR and high definition video as the person walks through a portal. In this experiment, we there 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 2048x2048 resolutions; with face/hair/neck visible in the screen.

For the MBGC distant-video dataset, Table 4 shows the number of video sequences captured by IOM system.

Table 4 Distant-video (face-visible) number of videos captured by IOM

Distant-Video	MBGC
Total # of videos	628
# of used videos	204

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 ten 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.

Figure 4 shows an example of distant-video within the MBGC dataset.

(a) Person A taken by IOM   (b) Person B taken by IOM
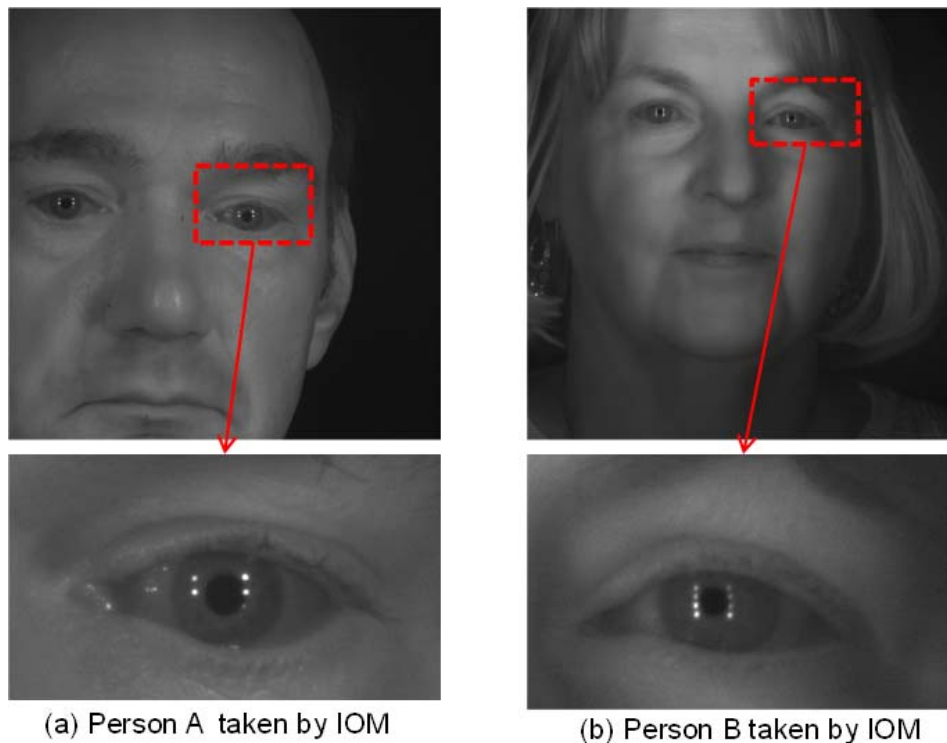
Figure 4 Sub-images extracted from MBGC distant-video taken by IOM

Figure 4 (a) 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).

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 pair-wise 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 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 ten graphical procedures: 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 [15].

Figure 5 shows the Main Effects plot of VASIR's k=8 factors on the VV Left case with the VR|FAR=.01 performance response.
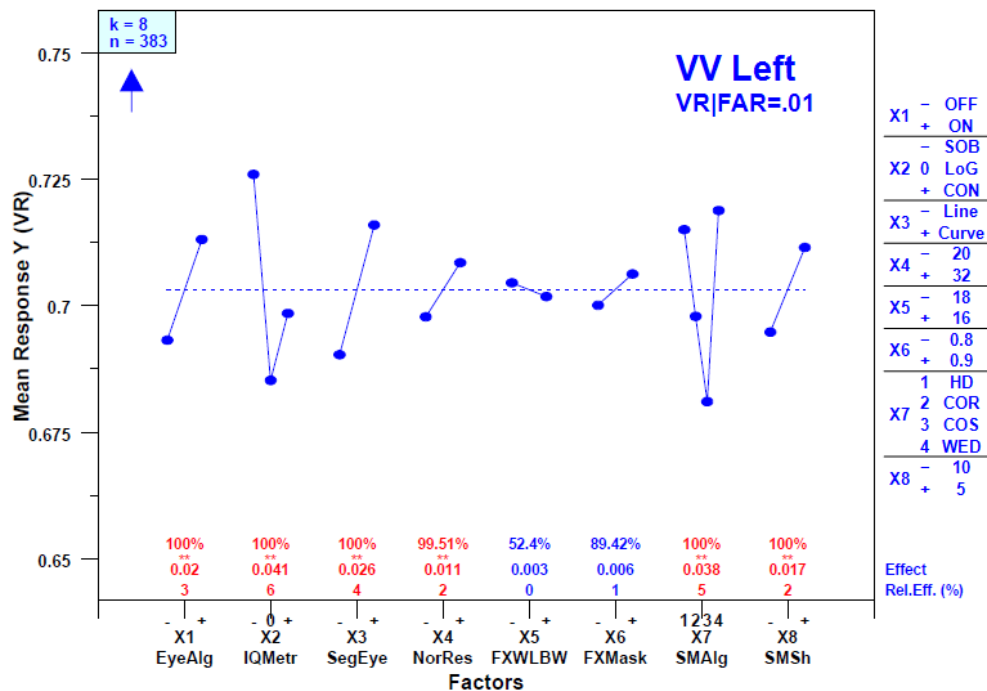
Figure 5 Main Effects plot with the VR|FAR=.01 response for VV Left—note that the four coded settings (1,2,3,4) for X7 are equal to (-2,-1,+1,+2). The importance of factors X2, X7, X3, X1, X8, and X4 in order

The horizontal axis illustrates the eight factors (X1 to X8) and the coded factor settings (e.g., "−1", "0", or "+1") for 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 presents 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 Figure 5 plot has n=383 because one value out of the 384 doesn't 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 ≤ 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" = $100 \times$ 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 by:

X2 (IQMetr): Image quality metrics (effect = .041), followed closely by

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; 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 Figure 5, 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 5.

Table 5 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.

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.

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 *n* 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 [10]. Ordered Data plots can reconfirm the results from main effects.

Figure 6 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 Figure 6 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 the data properly. So we illustrated only the 60 most important 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.
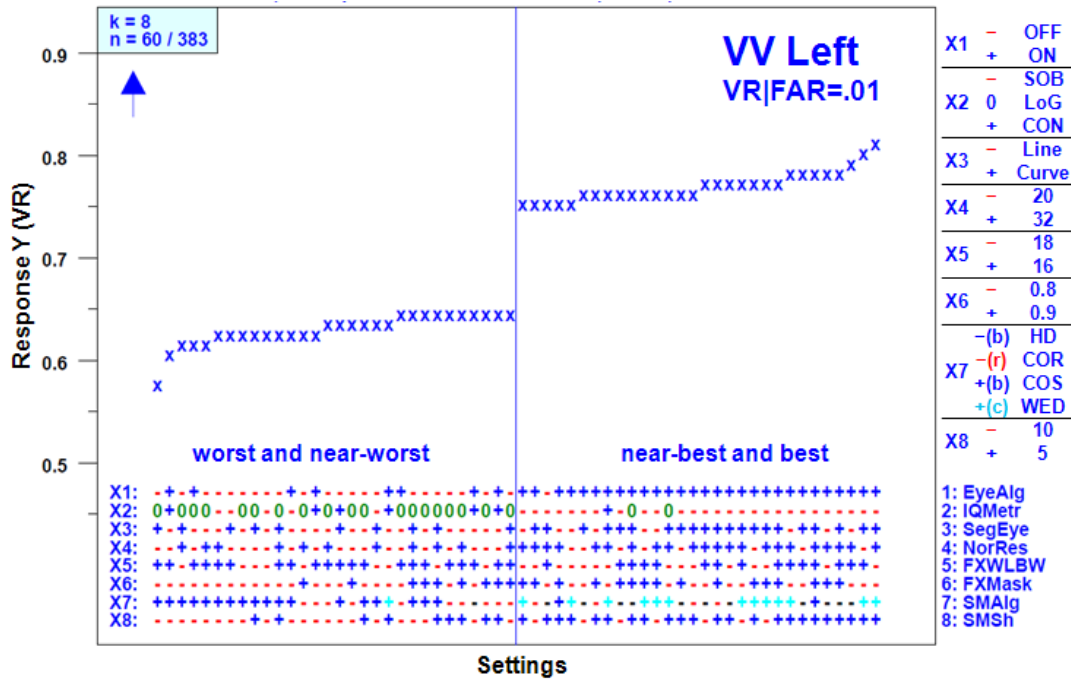
Figure 6 Ordered Data plot with VR|FAR=.01 response for VV Left. We illustrated only the 60 most important extreme points—30 (best) + 30 (worst)

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, 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. For the 30 worst responses, 22 came from X1=-1. Thus, X1 (EyeAlg) is an important factor.

Table 6 summarizes the counting levels from the plot in Figure 6. Highlighted in orange are the counts of two levels ≥ 20 for the best response and ≤ 10 for the worst response—note that since factor X7 has four levels, the criteria for this count is ≥ 12 and ≤ 2, respectively, and for the three levels of factor X2, the criteria is ≥ 15 and ≤ 5. Based on the above table, we thus conclude that the important factors for VR|FAR=.01 are as follows: X1 (EyeAlg), X2 (IQMetr), X3 (SegEye), X7 (SMAlg), and X8 (SMSh)—where both worst and best responses were highlighted in orange. From Figure 6, the best settings for the VR|FAR=.01 response are (+1,−1, +1, +1, −1, −1, +2, +1)—which is exactly the same as the results from the Main Effects plot in Figure 5.

25

Table 6 Summary of counting levels for the VR|FAR=.01 worst/best response

Factor	Worst response					Best response				
	-2	-1	0	+1	+2	-2	-1	0	+1	+2
X1 (EyeAlg)		22		8			1		29	
X2 (IQMetr)		5	19	6			27	2	1	
X3 (SegEye)		20		10			9		21	
X4 (NorRes)		19		11			8		22	
X5 (FXWL)		9		21			15		15	
X6 (FXMask)		21		9			14		16	
X7 (SMAlg)	1	10		8	1	11	4		2	13
X8 (SMSh)		21		9			7		23	

Table 7 shows the summary of important factors and the best settings –yielded from the Ordered Data plot—for the VV Left case.

Table 7 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
VR|FAR=.01	X1, X2, X3, X7, X8	+1 (ON)	−1 (SOB)	+1 (Curves)	+1 (32)	−1 (18)	−1 (.8)	+2 (WED)	+1 (5)
VR|FAR=.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)

From the analysis of all responses for this VV Left case in Ordered Data plots, factors X1 (EyeAlg), X2 (IQMetr), X3 (EyeSeg), X7 (SMAlg), and X8 (SMSh) are the important factors and the best settings are (+1,−1, +1, +1, −1, −1, −2, +1)—which is relatively similar to the results derived from the Main Effects plots.

6.3 Interaction Effects Matrix

Knowledge of the main effects is an incomplete description of the influence of factors because interactions exist among factors—complex systems such as biometrics are frequently more complicated than simple (linear/additive) models. When a factor effect changes by depending on the setting of another factor, the two factors are said to interact, and those *interaction effects* must be considered. For factors A and B with two levels (+1 and -1), the interaction effect is estimated by calculating the factor A effect (when factor B is at the + level), calculating the factor A effect (when factor B is at the – level), and then differencing the two factor A effects.   If there is no interaction between the two factors A and

B, then the factor A effect will be the same for both and hence the difference will be 0. We now address the problem as to whether two-term interaction effects exist in the VASIR system and determine whether those interaction effects are important. VASIR has $\binom{8}{2} = 28$ such two-term interactions among the eight algorithmic factors.

There are two methods to represent the interaction effects between the two factors:

1) interaction effects with two-levels and
2) interaction effects with 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:

$$(-1) \times (-1) = +1$$
$$(-1) \times (+1) = -1$$
$$(+1) \times (-1) = -1$$
$$(+1) \times (+1) = +1$$

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, the eight algorithmic factors have different levels; 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 Figure 7 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. Figure 7 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.
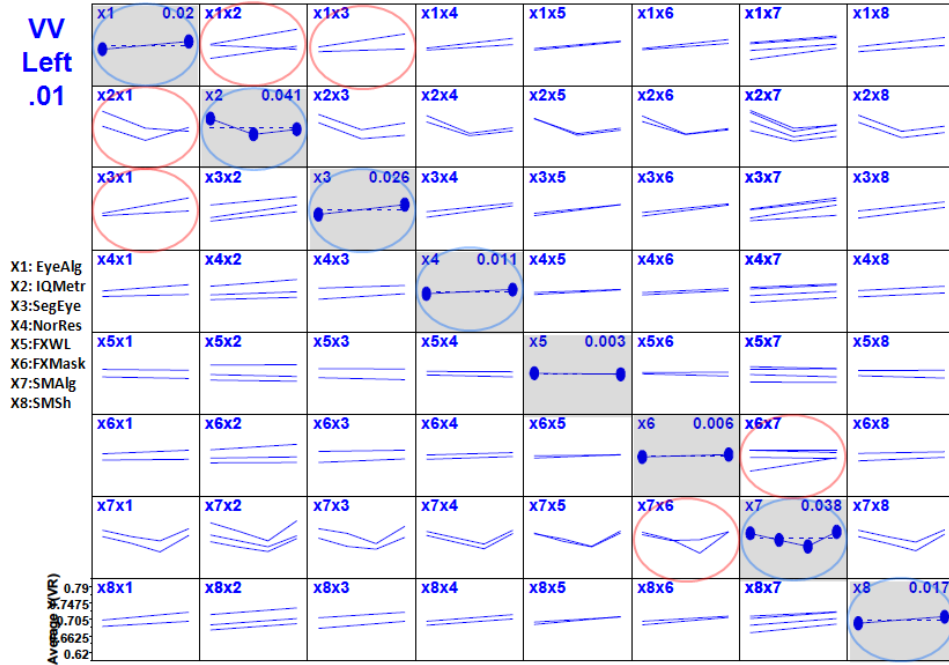
Figure 7 Interaction Effects Matrix plot with the VR|FAR=.01 response for VV Left. The blue circles indicate significant factors and the red circles indicate two-term interactions where effects appear significant.

The plots on the diagonal are identical to those seen on the Main Effects plot; the horizontal axis is the factor levels and the 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 (=2x3) 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 <u>not</u> 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 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 is no interaction 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, Figure 7 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 SMAlg).

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

Table 8 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•SMAlg)
VR|FAR=.10	X1X2 (EyeAlg•IQMetr)	X1X7 (EyeAlg•SMAlg)	
EER	X1X2 (EyeAlg•IQMetr)	X1X7 (EyeAlg•SMAlg)	

For this VV Left eye case, the X1X2 (EyeAlg•IQMetr) interaction occurred for all three responses and the X1X7 (EyeAlg•SMAlg) 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 (SMAlg). The results show that X3 (SegEye), X4 (NorRes), X5 (FXWLBW), X6 (FXMask), and X8 (SMsh) barely have any interaction in the VASIR system.

The results indicated that almost all of the 28 interaction plots have parallel traces—only 2 or 3 interaction effects out of 28. 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 Curve

The *ROC (Receiver Operating Characteristic)* curve is a plot of FAR (False Accept Rate) against VR (Verification Rate) which presents the tradeoff between the estimates of the verification rate and false accept rate. In an ideal system, the FAR should be low and the VR should be high.
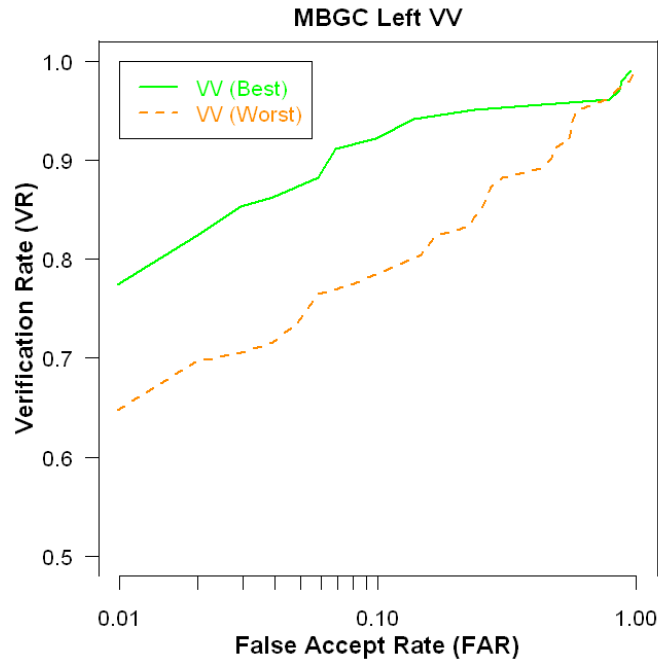
Figure 8 Comparisons of the VASIR's best and worst settings for VV Left

We now address the question as to how good the optimal settings are as compared to other possible settings for the VASIR system. Figure 8 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, −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 X2 (IQMetr) is image quality metrics and factor X7 (SMAlg) is similarity metrics—these are the two <u>most important algorithmic factors</u> affecting VASIR performance—this is robustly true over all three responses. Factors X5 (FXWL: wavelength) and X6 (FXMask: wavelet magnitude) are universally the <u>least important factors</u> 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 <u>VASIR's performance increased 12.8 % and EER decreased 9.2 %</u>. We demonstrated that the experiment design we employed was very <u>efficient</u> and <u>effective</u> 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 [19], CON was first introduced for general quality assessment by Albregtsen [24]; LoG was specifically used for the iris image quality assessment by Wan et al [35]. 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 (SMAlg): Similarity metrics

The second most important factor affecting VASIR performance was SMAlg (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 SMAlg 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. SMAlg 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 [7] has a significant improvement over IrisBEE's approach [5]. 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 (SMAlg: Similarity metrics). As expected, the best setting for the (EyeAlg, IQMetr, EyeSeg, SMAlg) 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 [5] used 3 bits shifting for left and right for the LEI [36] dataset and 8 bits for the CASIA [37] dataset. IrisBEE [17] used 10 bits shifting for the ICE2005 [17] dataset. For VASIR, we examined the use of 5 and 10 bits shifting for the MBGC video dataset. The two settings for SMSh are therefore 5 and 10. 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 factors, 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 20px

and 32px. We found that the larger radius resolution (32px) had a better matching performance than the smaller resolution (20px) 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 (SMAlg: 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 16px and 18px (the IrisBEE default). Our study found that 18px was slightly (but not significantly) better than 16px—the difference is negligible. Masek [5] found that 18px 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 between16px and 18px 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; 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 (SMAlg: 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 (SMAlg):** 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 (SMAlg)—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 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; and 3) 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).

References

[1] A. K. Jain, P. J. Flynn, and A. A. Ross, "Introduction to Biometrics," *Handbook of Biometrics, Springer*, 2008.

[2] G. Stirling Andrew, "On Science and Precaution in the Management of Technological Risk," *European Science and Technology Observatory*, 2002.

[3] G. M. Hornberger and S. R.C., "An approach to the preliminary analysis of environmental systems," *Journal of Environmental Management*, vol. 12, pp. 7–18, 1981.

[4] A. Saltelli, "Sensitivity analysis for importance assessment," *Risk Analysis*, vol. 22, no. 3, pp. 579–590, 2002.

[5] L. Masek and P. Kovesi, "MATLAB Source Code for a Biometric Identification System Based on Iris Patterns," *School of Computer Science and Software Engineering, University of Western Australia*, 2003.

[6] K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostyn, S. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, N. Poh, and others, "Performance Characterization of Face Recognition Algorithms and Their Sensitivity to Severe Illumination Changes," *Biometric Authentication*, pp. 8–19, 2004.

[7] Y. Lee, "VASIR: Video-based Automatic System for Iris Recognition," Dissertations, School of Computer Science and Engineering in Chung-Ang University, South Korea, 2011.

[8] J. J. Filliben, S. Cetinkunt, W. L. Yu, and A. Donmez, "Exploratory Data Analysis Techniques as Applied to a High-Precision Turning Machine," *Elsevier Science Publishers B.V.*, p. 199, 1993.

[9] J. J. Filliben, "Experiment Design for Scientists and Engineers," Short Course, National Institute of Standards and Technology, 2002.

[10] National Institute of Standards and Technology, "Engineering Statistics HandBook (5.5.9. An EDA approach to experimental design)," *National Institute of Standards and Technology*, 2003. Available: http://www.itl.nist.gov/div898/handbook/pri/section5/pri59.htm. [Accessed: 09-Mar-2011].

[11] J. P. . Kleijnen, "Design and analysis of Monte Carlo experiments," *Handbook of Computational Statistics*, vol. 1, 2004.

[12] G. E. P. Box, W. G. Hunter, and J. S. Hunter, "Statistics for experimenters: an introduction to design, and model building." New York Singapore: John Wiley, 1978.

[13] G. E.P. Box, J. S. Hunter, and W. G. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery, 2nd Edition*. Wiley, 2005.

[14] J. J. Filliben and A. Hecket, "Dataplot Homepage." Available: http://www.itl.nist.gov/div898/software/dataplot/. [Accessed: 11-Jul-2011].

[15] K. L. Mills, J. J. Filliben, C. Dong Yeon, S. Edward, and G. Daniel, "Chapter 4: Sensitivity Analysis of MesoNet," *Study of Proposed Internet Congestion-Control Mechanisms*, 2010. Available: http://www.nist.gov/itl/antd/Congestion_Control_Study.cfm. [Accessed: 09-Mar-2011].

[16] J. J. Filliben, "DATAPLOT—An Interactive High-level Language for Graphics, Non-linear Fitting, Data analysis, and Mathematics," *ACM SIGGRAPH Computer Graphics*, vol. 15, no. 3, 1981.

[17] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, and W. T. Scruggs, "The iris challenge evaluation 2005," *The 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2008)*, pp. 1–8, 2008.

[18] Y. Lee, P. J. Phillips, and R. Micheals, "An automated video-based system for iris recognition," *Advances in Biometrics*, pp. 1160–1169, 2009.

[19] J. R. Beveridge, G. H. Givens, P. J. Phillips, B. A. Draper, and Y. M. Lui, "Focus on quality, predicting FRVT 2006 performance," in *8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1–8, 2009.

[20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Prentice Hall, 2002.

[21] N. Kanopoulos, N. Vasanthavada, and R. L. BAKER, "Design of an Image Edge Detection Filter Using the Sobel," *IEEE JOURNAL OF SOLID-STATE CIRCUITS*, vol. 23, no. 2, 1988.

[22] R. Wang, "Lecture Notes: Computer Image Processing and Analysis," 2011. Available: http://fourier.eng.hmc.edu/e161/lectures/gradient/. [Accessed: 21-Oct-2011].

[23] S. R. Gunn, "On the discrete representation of the Laplacian of Gaussian," *The Journal of the Pattern Recognition*, vol. 32, pp. 1463–1472, 1999.

[24] F. Albregtsen, "Statistical Texture Measures Computed from Gray Level Coocurrence Matrices," *Image Processing Laboratory, Department of Informatics, University of Oslo*, 1995.

[25] M. Hall-Beyer, "GLCM Texture Tutorial." Available: http://www.fp.ucalgary.ca/mhallbey/tutorial.htm. [Accessed: 02-Jul-2011].

[26] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 2002.

[27] P. Yao, J. Li, X. Ye, Z. Zhuang, and B. Li, "Iris recognition algorithm using modified log-gabor filters," *Pattern Recognition*, vol. 4, pp. 461–464, 2006.

[28] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "The best bits in an iris code," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 964–973, 2008.

[29] J. Daugman, "How Iris Recognition Works," *IEEE Transaction on Circuits System and Video Technology*, vol. 14, no. 1, pp. 21–30, Jan. 2004.

[30] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 2002.

[31] S. Lim, K. Lee, O. Byeon, and T. Kim, "Efficient iris recognition through improvement of feature vector and classifier," *ETRI journal*, vol. 23, no. 2, pp. 61–70, 2001.

[32] Hamming R.W., "Error Detecting and Error Correcting Codes," *The Bell System Technical Journal*, vol. XXIX, p. 2, 1950.

[33] P. Grother, R. Micheals, and P. J. Phillips, "Face Recognition Vendor Test 2002 Performance Metrics," *Fourth International Conference on Audio-Visual Based Person Authentication*, 2003.

[34] P. J. Phillips, P. J. Flynn, J. R. Beveridge, K. W. Bowyer, W. T. Scruggs, A. O'Toole, and et al., "Overview of the Multiple Biometrics Grand Challenge," The third IAPR/IEEE International Conference of Biometrics (ICB2009), pp.705-714, 2009.

[35] J. Wan, X. He, and P. Shi, "An Iris Image Quality Assessment Method Based on Laplacian of Gaussian Operation," in *Proc. IAPR IAPR Conf. Machine Vision Applications*, pp. 248–251, 2007.

[36] C. Barry and N. Ritter, "Database of 120 Greyscale Eye Images.," Lions Eye Institute,, Perth Western Australia.

[37] Center for Biometrics and Security Research (CBSR), "CASIA Iris Image Database," *Service Note on CASIA Iris Image Databases*. Available: http://www.cbsr.ia.ac.cn/IrisDatabase.htm. [Accessed: 31-Aug-2011].

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	
SEG	*** Pupil circle**		
	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
1	-1	-1	-1	-1	-1	-1	-2	-1
2	1	-1	-1	-1	-1	-1	-2	1
3	-1	-1	1	-1	-1	-1	-2	1
4	1	-1	1	-1	-1	-1	-2	-1
5	-1	-1	-1	1	-1	-1	-2	1
6	1	-1	-1	1	-1	-1	-2	-1
7	-1	-1	1	1	-1	-1	-2	-1
8	1	-1	1	1	-1	-1	-2	1
9	-1	-1	-1	-1	1	-1	-2	1
10	1	-1	-1	-1	1	-1	-2	-1
11	-1	-1	1	-1	1	-1	-2	-1
12	1	-1	1	-1	1	-1	-2	1
13	-1	-1	-1	1	1	-1	-2	-1
14	1	-1	-1	1	1	-1	-2	1
15	-1	-1	1	1	1	-1	-2	1
16	1	-1	1	1	1	-1	-2	-1
17	-1	-1	-1	-1	-1	1	-2	1
18	1	-1	-1	-1	-1	1	-2	-1
19	-1	-1	1	-1	-1	1	-2	-1
20	1	-1	1	-1	-1	1	-2	1
21	-1	-1	-1	1	-1	1	-2	-1
22	1	-1	-1	1	-1	1	-2	1
23	-1	-1	1	1	-1	1	-2	1
24	1	-1	1	1	-1	1	-2	-1
25	-1	-1	-1	-1	1	1	-2	-1
26	1	-1	-1	-1	1	1	-2	1
27	-1	-1	1	-1	1	1	-2	1
28	1	-1	1	-1	1	1	-2	-1
29	-1	-1	-1	1	1	1	-2	1
30	1	-1	-1	1	1	1	-2	-1
31	-1	-1	1	1	1	1	-2	-1
32	1	-1	1	1	1	1	-2	1
33	-1	0	-1	-1	-1	-1	-2	-1
34	1	0	-1	-1	-1	-1	-2	1
35	-1	0	1	-1	-1	-1	-2	1
36	1	0	1	-1	-1	-1	-2	-1
37	-1	0	-1	1	-1	-1	-2	1
38	1	0	-1	1	-1	-1	-2	-1
39	-1	0	1	1	-1	-1	-2	-1
40	1	0	1	1	-1	-1	-2	1
41	-1	0	-1	-1	1	-1	-2	1
42	1	0	-1	-1	1	-1	-2	-1
43	-1	0	1	-1	1	-1	-2	-1
44	1	0	1	-1	1	-1	-2	1
45	-1	0	-1	1	1	-1	-2	-1
46	1	0	-1	1	1	-1	-2	1
47	-1	0	1	1	1	-1	-2	1
48	1	0	1	1	1	-1	-2	-1
49	-1	0	-1	-1	-1	1	-2	1
50	1	0	-1	-1	-1	1	-2	-1
51	-1	0	1	-1	-1	1	-2	-1
52	1	0	1	-1	-1	1	-2	1
53	-1	0	-1	1	-1	1	-2	-1
54	1	0	-1	1	-1	1	-2	1
55	-1	0	1	1	-1	1	-2	1
56	1	0	1	1	-1	1	-2	-1
57	-1	0	-1	-1	1	1	-2	-1
58	1	0	-1	-1	1	1	-2	1
59	-1	0	1	-1	1	1	-2	1
60	1	0	1	-1	1	1	-2	-1
61	-1	0	-1	1	1	1	-2	1
62	1	0	-1	1	1	1	-2	-1

n/k	X1	X2	X3	X4	X5	X6	X7	X8
63	-1	0	1	1	1	1	-2	-1
64	1	0	1	1	1	1	-2	1
65	-1	1	-1	-1	-1	-1	-2	-1
66	1	1	-1	-1	-1	-1	-2	1
67	-1	1	1	-1	-1	-1	-2	1
68	1	1	1	-1	-1	-1	-2	-1
69	-1	1	-1	1	-1	-1	-2	1
70	1	1	-1	1	-1	-1	-2	-1
71	-1	1	1	1	-1	-1	-2	-1
72	1	1	1	1	-1	-1	-2	1
73	-1	1	-1	-1	1	-1	-2	1
74	1	1	-1	-1	1	-1	-2	-1
75	-1	1	1	-1	1	-1	-2	-1
76	1	1	1	-1	1	-1	-2	1
77	-1	1	-1	1	1	-1	-2	-1
78	1	1	-1	1	1	-1	-2	1
79	-1	1	1	1	1	-1	-2	1
80	1	1	1	1	1	-1	-2	-1
81	-1	1	-1	-1	-1	1	-2	1
82	1	1	-1	-1	-1	1	-2	-1
83	-1	1	1	-1	-1	1	-2	-1
84	1	1	1	-1	-1	1	-2	1
85	-1	1	-1	1	-1	1	-2	-1
86	1	1	-1	1	-1	1	-2	1
87	-1	1	1	1	-1	1	-2	1
88	1	1	1	1	-1	1	-2	-1
89	-1	1	-1	-1	1	1	-2	-1
90	1	1	-1	-1	1	1	-2	1
91	-1	1	1	-1	1	1	-2	1
92	1	1	1	-1	1	1	-2	-1
93	-1	1	-1	1	1	1	-2	1
94	1	1	-1	1	1	1	-2	-1
95	-1	1	1	1	1	1	-2	-1
96	1	1	1	1	1	1	-2	1
97	-1	-1	-1	-1	-1	-1	-1	-1
98	1	-1	-1	-1	-1	-1	-1	1
99	-1	-1	1	-1	-1	-1	-1	1
100	1	-1	1	-1	-1	-1	-1	-1
101	-1	-1	-1	1	-1	-1	-1	1
102	1	-1	-1	1	-1	-1	-1	-1
103	-1	-1	1	1	-1	-1	-1	-1
104	1	-1	1	1	-1	-1	-1	1
105	-1	-1	-1	-1	1	-1	-1	1
106	1	-1	-1	-1	1	-1	-1	-1
107	-1	-1	1	-1	1	-1	-1	-1
108	1	-1	1	-1	1	-1	-1	1
109	-1	-1	-1	1	1	-1	-1	-1
110	1	-1	-1	1	1	-1	-1	1
111	-1	-1	1	1	1	-1	-1	1
112	1	-1	1	1	1	-1	-1	-1
113	-1	-1	-1	-1	-1	1	-1	1
114	1	-1	-1	-1	-1	1	-1	-1
115	-1	-1	1	-1	-1	1	-1	-1
116	1	-1	1	-1	-1	1	-1	1
117	-1	-1	-1	1	-1	1	-1	-1
118	1	-1	-1	1	-1	1	-1	1
119	-1	-1	1	1	-1	1	-1	1
120	1	-1	1	1	-1	1	-1	-1
121	-1	-1	-1	-1	1	1	-1	-1
122	1	-1	-1	-1	1	1	-1	1
123	-1	-1	1	-1	1	1	-1	1
124	1	-1	1	-1	1	1	-1	-1

n/k	X1	X2	X3	X4	X5	X6	X7	X8
125	-1	-1	-1	1	1	1	-1	1
126	1	-1	-1	1	1	1	-1	-1
127	-1	-1	1	1	1	1	-1	-1
128	1	-1	1	1	1	1	-1	1
129	-1	0	-1	-1	-1	-1	-1	-1
130	1	0	-1	-1	-1	-1	-1	1
131	-1	0	1	-1	-1	-1	-1	1
132	1	0	1	-1	-1	-1	-1	-1
133	-1	0	-1	1	-1	-1	-1	1
134	1	0	-1	1	-1	-1	-1	-1
135	-1	0	1	1	-1	-1	-1	-1
136	1	0	1	1	-1	-1	-1	1
137	-1	0	-1	-1	1	-1	-1	1
138	1	0	-1	-1	1	-1	-1	-1
139	-1	0	1	-1	1	-1	-1	-1
140	1	0	1	-1	1	-1	-1	1
141	-1	0	-1	1	1	-1	-1	-1
142	1	0	-1	1	1	-1	-1	1
143	-1	0	1	1	1	-1	-1	1
144	1	0	1	1	1	-1	-1	-1
145	-1	0	-1	-1	-1	1	-1	-1
146	1	0	-1	-1	-1	1	-1	1
147	-1	0	1	-1	-1	1	-1	1
148	1	0	1	-1	-1	1	-1	-1
149	-1	0	-1	1	-1	1	-1	-1
150	1	0	-1	1	-1	1	-1	1
151	-1	0	1	1	-1	1	-1	1
152	1	0	1	1	-1	1	-1	-1
153	-1	0	-1	-1	1	1	-1	-1
154	1	0	-1	-1	1	1	-1	1
155	-1	0	1	-1	1	1	-1	1
156	1	0	1	-1	1	1	-1	-1
157	-1	0	-1	1	1	1	-1	1
158	1	0	-1	1	1	1	-1	-1
159	-1	0	1	1	1	1	-1	-1
160	1	0	1	1	1	1	-1	1
161	-1	1	-1	-1	-1	-1	-1	-1
162	1	1	-1	-1	-1	-1	-1	1
163	-1	1	1	-1	-1	-1	-1	1
164	1	1	1	-1	-1	-1	-1	-1
165	-1	1	-1	1	-1	-1	-1	1
166	1	1	-1	1	-1	-1	-1	-1
167	-1	1	1	1	-1	-1	-1	-1
168	1	1	1	1	-1	-1	-1	1
169	-1	1	-1	-1	1	-1	-1	1
170	1	1	-1	-1	1	-1	-1	-1
171	-1	1	1	-1	1	-1	-1	-1
172	1	1	1	-1	1	-1	-1	1
173	-1	1	-1	1	1	-1	-1	-1
174	1	1	-1	1	1	-1	-1	1
175	-1	1	1	1	1	-1	-1	1
176	1	1	1	1	1	-1	-1	-1
177	-1	1	-1	-1	-1	1	-1	1
178	1	1	-1	-1	-1	1	-1	-1
179	-1	1	1	-1	-1	1	-1	-1
180	1	1	1	-1	-1	1	-1	1
181	-1	1	-1	1	-1	1	-1	-1
182	1	1	-1	1	-1	1	-1	1
183	-1	1	1	1	-1	1	-1	1
184	1	1	1	1	-1	1	-1	-1
185	-1	1	-1	-1	1	1	-1	1
186	1	1	-1	-1	1	1	-1	1

n/k	X1	X2	X3	X4	X5	X6	X7	X8
187	-1	1	1	-1	1	1	-1	1
188	1	1	1	-1	1	1	-1	-1
189	-1	1	-1	1	1	1	-1	1
190	1	1	-1	1	1	1	-1	-1
191	-1	1	1	1	1	1	-1	1
192	1	1	1	1	1	1	-1	1
193	-1	-1	-1	-1	-1	-1	1	-1
194	1	-1	-1	-1	-1	-1	1	1
195	-1	-1	1	-1	-1	-1	1	1
196	1	-1	1	-1	-1	-1	1	1
197	-1	-1	-1	1	-1	-1	1	1
198	1	-1	-1	1	-1	-1	1	1
199	-1	-1	1	1	-1	-1	1	1
200	1	-1	1	1	-1	-1	1	1
201	-1	-1	-1	-1	1	-1	1	1
202	1	-1	-1	-1	1	-1	1	1
203	-1	-1	1	-1	1	-1	1	1
204	1	-1	1	-1	1	-1	1	1
205	-1	-1	-1	1	1	-1	1	1
206	1	-1	-1	1	1	-1	1	1
207	-1	-1	1	1	1	-1	1	1
208	1	-1	1	1	1	-1	1	1
209	-1	-1	-1	-1	-1	1	1	1
210	1	-1	-1	-1	-1	1	1	1
211	-1	-1	1	-1	-1	1	1	1
212	1	-1	1	-1	-1	1	1	1
213	-1	-1	-1	1	-1	1	1	1
214	1	-1	-1	1	-1	1	1	1
215	-1	-1	1	1	-1	1	1	1
216	1	-1	1	1	-1	1	1	-1
217	-1	-1	-1	-1	1	1	1	1
218	1	-1	-1	-1	1	1	1	1
219	-1	-1	1	-1	1	1	1	1
220	1	-1	1	-1	1	1	1	1
221	-1	-1	-1	1	1	1	1	1
222	1	-1	-1	1	1	1	1	1
223	-1	-1	1	1	1	1	1	1
224	1	-1	1	1	1	1	1	1
225	-1	0	-1	-1	-1	-1	1	1
226	1	0	-1	-1	-1	-1	1	1
227	-1	0	1	-1	-1	-1	1	1
228	1	0	1	-1	-1	-1	1	1
229	-1	0	-1	1	-1	-1	1	1
230	1	0	-1	1	-1	-1	1	1
231	-1	0	1	1	-1	-1	1	-1
232	1	0	1	1	-1	-1	1	1
233	-1	0	-1	-1	1	-1	1	1
234	1	0	-1	-1	1	-1	1	-1
235	-1	0	-1	1	1	-1	1	-1
236	1	0	1	-1	1	-1	1	1
237	-1	0	-1	1	1	-1	1	1
238	1	0	-1	1	1	-1	1	1
239	-1	0	1	1	1	-1	1	1
240	1	0	1	1	1	-1	1	-1
241	-1	0	-1	-1	-1	1	1	1
242	1	0	-1	-1	-1	1	1	1
243	-1	0	1	-1	-1	1	1	1
244	1	0	1	-1	-1	1	1	1
245	-1	0	-1	1	-1	1	1	1
246	1	0	-1	1	-1	1	1	1
247	-1	0	1	1	-1	1	1	1
248	1	0	1	1	-1	1	1	-1
249	-1	0	-1	-1	1	1	1	-1
250	1	0	-1	-1	1	1	1	1
251	-1	0	1	-1	1	1	1	1
252	1	0	1	-1	1	1	1	-1
253	-1	0	-1	1	1	1	1	1
254	1	0	-1	1	1	1	1	-1
255	-1	0	1	1	1	1	1	-1
256	1	0	1	1	1	1	1	1
257	-1	1	-1	-1	-1	-1	1	1
258	1	1	-1	-1	-1	-1	1	1
259	-1	1	1	-1	-1	-1	1	1
260	1	1	1	-1	-1	-1	1	1
261	-1	1	-1	1	-1	-1	1	1
262	1	1	-1	1	-1	-1	1	-1
263	-1	1	1	1	-1	-1	1	-1
264	1	1	1	1	-1	-1	1	1
265	-1	1	-1	-1	1	-1	1	1
266	1	1	-1	-1	1	-1	1	-1
267	-1	1	1	-1	1	-1	1	-1
268	1	1	1	-1	1	-1	1	1
269	-1	1	-1	1	1	-1	1	-1
270	1	1	-1	1	1	-1	1	1
271	-1	1	1	1	1	-1	1	1
272	1	1	1	1	1	-1	1	-1
273	-1	1	-1	-1	-1	1	1	1
274	1	1	-1	-1	-1	1	1	-1
275	-1	1	1	-1	-1	1	1	-1
276	1	1	1	-1	-1	1	1	1
277	-1	1	-1	1	-1	1	1	1
278	1	1	-1	1	-1	1	1	1
279	-1	1	1	1	-1	1	1	1
280	1	1	1	1	-1	1	1	-1
281	-1	1	-1	-1	1	1	1	-1
282	1	1	-1	-1	1	1	1	1
283	-1	1	1	-1	1	1	1	1
284	1	1	1	-1	1	1	1	-1
285	-1	1	-1	1	1	1	1	1
286	1	1	-1	1	1	1	1	-1
287	-1	1	1	1	1	1	1	-1
288	1	1	1	1	1	1	1	1
289	-1	-1	-1	-1	-1	-1	2	-1
290	1	-1	-1	-1	-1	-1	2	1
291	-1	-1	1	-1	-1	-1	2	-1
292	1	-1	1	-1	-1	-1	2	1
293	-1	-1	-1	1	-1	-1	2	-1
294	1	-1	-1	1	-1	-1	2	-1
295	-1	-1	1	1	-1	-1	2	1
296	1	-1	1	1	-1	-1	2	1
297	-1	-1	-1	-1	1	-1	2	1
298	1	-1	-1	-1	1	-1	2	-1
299	-1	-1	1	-1	1	-1	2	1
300	1	-1	1	-1	1	-1	2	1
301	-1	-1	-1	1	1	-1	2	-1
302	1	-1	-1	1	1	-1	2	1
303	-1	-1	1	1	1	-1	2	1
304	1	-1	1	1	1	-1	2	-1
305	-1	-1	-1	-1	-1	1	2	-1
306	1	-1	-1	-1	-1	1	2	-1
307	-1	-1	1	-1	-1	1	2	-1
308	1	-1	1	-1	-1	1	2	1
309	-1	-1	-1	1	-1	1	2	-1
310	1	-1	-1	1	-1	1	2	1
311	-1	-1	1	1	-1	1	2	1
312	1	-1	1	1	-1	1	2	1
313	-1	-1	-1	-1	1	1	2	-1
314	1	-1	-1	-1	1	1	2	1
315	-1	-1	1	-1	1	1	2	1
316	1	-1	1	-1	1	1	2	-1
317	-1	-1	-1	1	1	1	2	1
318	1	-1	-1	1	1	1	2	-1
319	-1	-1	1	1	1	1	2	-1
320	1	-1	1	1	1	1	2	1
321	-1	0	-1	-1	-1	-1	2	-1
322	1	0	-1	-1	-1	-1	2	1
323	-1	0	1	-1	-1	-1	2	1
324	1	0	1	-1	-1	-1	2	-1
325	-1	0	-1	1	-1	-1	2	1
326	1	0	-1	1	-1	-1	2	-1
327	-1	0	1	1	-1	-1	2	1
328	1	0	1	1	-1	-1	2	1
329	-1	0	-1	-1	1	-1	2	-1
330	1	0	-1	-1	1	-1	2	1
331	-1	0	1	-1	1	-1	2	1
332	1	0	1	-1	1	-1	2	1
333	-1	0	-1	1	1	-1	2	-1
334	1	0	-1	1	1	-1	2	1
335	-1	0	1	1	1	-1	2	1
336	1	0	1	1	1	-1	2	-1
337	-1	0	-1	-1	-1	1	2	1
338	1	0	-1	-1	-1	1	2	-1
339	-1	0	1	-1	-1	1	2	1
340	1	0	1	-1	-1	1	2	1
341	-1	0	-1	1	-1	1	2	-1
342	1	0	-1	1	-1	1	2	1
343	-1	0	1	1	-1	1	2	1
344	1	0	1	1	-1	1	2	1
345	-1	0	-1	-1	1	1	2	-1
346	1	0	-1	-1	1	1	2	1
347	-1	0	1	-1	1	1	2	1
348	1	0	1	-1	1	1	2	1
349	-1	0	-1	1	1	1	2	-1
350	1	0	-1	1	1	1	2	-1
351	-1	0	1	1	1	1	2	-1
352	1	0	1	1	1	1	2	1
353	-1	1	-1	-1	-1	-1	2	1
354	1	1	-1	-1	-1	-1	2	1
355	-1	1	1	-1	-1	-1	2	-1
356	1	1	1	-1	-1	-1	2	1
357	-1	1	-1	1	-1	-1	2	1
358	1	1	-1	1	-1	-1	2	1
359	-1	1	1	1	-1	-1	2	1
360	1	1	1	1	-1	-1	2	1
361	-1	1	-1	-1	1	-1	2	1
362	1	1	-1	-1	1	-1	2	1
363	-1	1	1	-1	1	-1	2	-1
364	1	1	1	-1	1	-1	2	1
365	-1	1	-1	1	1	-1	2	-1
366	1	1	-1	1	1	-1	2	1
367	-1	1	1	1	1	-1	2	1
368	1	1	1	1	1	-1	2	-1
369	-1	1	-1	-1	-1	1	2	1
370	1	1	-1	-1	-1	1	2	1
371	-1	1	1	-1	-1	1	2	1
372	1	1	1	-1	-1	1	2	1
373	-1	1	-1	1	-1	1	2	-1
374	1	1	-1	1	-1	1	2	1
375	-1	1	1	1	-1	1	2	1
376	1	1	1	1	-1	1	2	-1
377	-1	1	-1	-1	1	1	2	-1
378	1	1	-1	-1	1	1	2	1
379	-1	1	1	-1	1	1	2	1
380	1	1	1	-1	1	1	2	-1
381	-1	1	-1	1	1	1	2	1
382	1	1	-1	1	1	1	2	-1
383	-1	1	1	1	1	1	2	1
384	1	1	1	1	1	1	2	1