Temporal Stability of Visual Search-Driven Biometrics

Hong-Jun Yoon[a], Tandy R. Carmichael[b], Georgia Tourassi[a]

[a] Health Data Sciences Institute, Biomedical Science and Engineering Center, Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, TN, USA 37831;
[b] Department of Electrical and Computer Engineering, Tennessee Technological University, One William L. Jones Drive, Cookeville, TN, USA 38505

ABSTRACT

Previously, we have shown the potential of using an individual's visual search pattern as a possible biometric. That study focused on viewing images displaying dot-patterns with different spatial relationships to determine which pattern can be more effective in establishing the identity of an individual. In this follow-up study we investigated the temporal stability of this biometric. We performed an experiment with 16 individuals asked to search for a predetermined feature of a random-dot pattern as we tracked their eye movements. Each participant completed four testing sessions consisting of two dot patterns repeated twice. One dot pattern displayed concentric circles shifted to the left or right side of the screen overlaid with visual noise, and participants were asked which side the circles were centered on. The second dot-pattern displayed a number of circles (between 0 and 4) scattered on the screen overlaid with visual noise, and participants were asked how many circles they could identify. Each session contained 5 untracked tutorial questions and 50 tracked test questions (200 total tracked questions per participant). To create each participant's "fingerprint", we constructed a Hidden Markov Model (HMM) from the gaze data representing the underlying visual search and cognitive process. The accuracy of the derived HMM models was evaluated using cross-validation for various time-dependent train-test conditions. Subject identification accuracy ranged from 17.6% to 41.8% for all conditions, which is significantly higher than random guessing (1/16 = 6.25%). The results suggest that visual search pattern is a promising, temporally stable personalized fingerprint of perceptual organization.

Keywords: eye tracking, perceptual organization, user modeling

1. INTRODUCTION

Physical and behavioral characteristics have been studied extensively as the basis for biometric systems for identification or authentication of individuals [1,2]. With the continuous advancement and accessibility of eye-tracking technology, there has been an increased interest in exploring gaze as the source of biometric information by studying different aspects such gaze direction, velocity, etc. [3-10]. By observing how people pursue and achieve various visual tasks, signal processing scientists can derive new biometrics. The quality of the gaze-based biometrics however is highly dependent on several external and internal factors, such as the quality of the eye-tracker, the quality of the calibration process, and the whether an individual's behavioral trait changes over time and under different external conditions or stressors. It is also unclear whether gaze is behavioral or physiological trait.

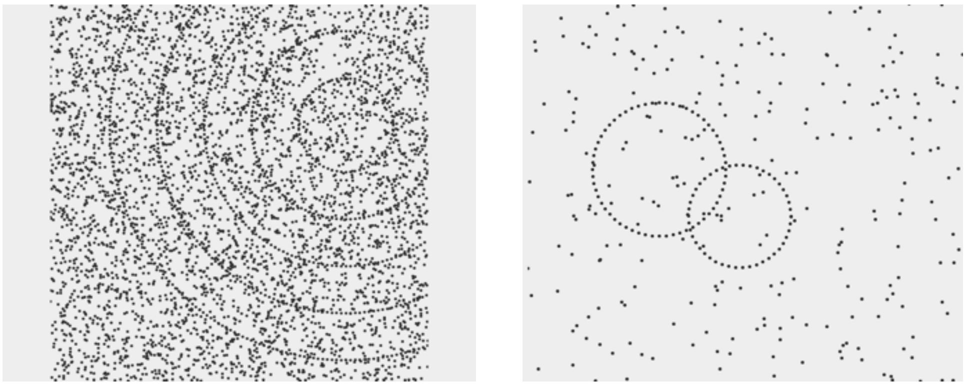
In our previous study we explored gaze velocity during visual search as a possible biometric. That study focused on visual search tasks involving different spatial arrangements [11] to determine i) which task captures best the individual's visual search behavior, and ii) whether the behavior is consistent across tasks. The visual search pattern related to the law of closure (i.e., "the mind completes shapes that are not whole") was found to derive a particularly effective biometric. However the previous study did not address the temporal stability of the biometric. To address this gap, we performed a new study with more subjects, strict calibration protocols, and a temporally variable study design. Understanding better the temporal stability of a gaze-based biometric and the factors that may influence its robustness are critical steps for effective implementation and utilization of the biometric.

2. METHODS

2.1 Perceptual Organization Tasks

Perceptual organization tests based on the Gestalt grouping principles of similarity, continuation, proximity, and closure [12] have been used before to study their potential as possible biomarkers for neurological disorders such as Alzheimer's disease [13]. Intrinsically, these tests encourage frequent visual processing for detection of various spatial cues, followed by memorization and recognition of visual components by grouping and separation. In our previous study, we investigated gaze tracking while individuals performed five perceptual organization tests to capture the uniqueness of the detection and recognition process of these individuals. That study suggested that one test was particularly helpful in deriving a fairly accurate biometric, producing significantly highest reader identification accuracy among the other five tests.

This study focuses on this specific perceptual organization test to evaluate the temporal stability of the derived biometrics. The test uses a spatial organization pattern known as the "the glass pattern test". The glass pattern test displayed a group of concentric circles made up of dots, illustrated in Figure 1(a). The circles were shifted to the left or right side of the image. Noise was superimposed over the image to introduce distractors. The test asked the participant which side of the image the circle was centered on. We also created a variant of the glass pattern test. Specifically, we scattered between 0-4 circles of varying size and position across an image. Noise was superimposed over the image as with the previous test. The participants were asked how many circles were displayed on the screen. Example of the Test 2 is shown in Figure 1(b). For each pattern we varied the task difficulty by adding or removing distractors. We determined the test difficulties from a staircase-type pilot study we performed before the actual experiment.



(a) (b)

Figure 1. Examples of stimulus patterns used in each of the two tests, (a) Test 1: glass patterns, (b) Test 2: a variant of the glass patterns.

2.2 Eye-Tracking Data Collection

For eye-tracking data collection, we used the Mirametrix S2 remote eye-tracker [14] based on its accuracy, availability, and ease of use. The Mirametrix S2 eye-tracker refreshed at 60 frames per second, which was considered reasonable for our study. The eye-tracker was placed directly under the monitor and positioned such that it was able to capture the subject's gaze to any part of the monitor. The eye-tracker was placed about two feet from the subject, which was the manufacturer's recommended distance. To minimize any interference with gaze tracking, the overhead lights were always turned off during tracking. We also carefully considered the eye-tracking process to maximize the accuracy of our results. Before we began the testing, we calibrated the eye-tracker using the typical 9-point calibration pattern until a precise calibration was achieved. Calibration was considered precise if the cursor correctly followed the subject's gaze to every section of the monitor. During the test, we constantly monitored the feedback from the eye-tracking system for any possible errors and adjusted the eye-tracker accordingly [15]. Recruited participants repeated the calibration until three consecutive acceptable calibrations were obtained. Calibration quality acceptance was determined by the following two conditions: a) there exists more than seven out of nine valid calibration points, and b) the average calibration error was less than 80 pixels. We dismissed participants whose calibrations could not satisfy the eligibility criteria within 10 calibration attempts.

2.3 Study Design

We scheduled four data collection sessions split into two morning and two afternoon sessions--a morning and afternoon session for each test described in section 2.1. The sessions were scheduled as close to the beginning or end of the participants' workday as possible to help us investigate whether work overload is a confounding factor. We ordered the sessions specifically so no participant would view the same test twice in a row. Only one morning session and one afternoon session could be scheduled on the same day. This separated the participants into four testing groups depending on which session order they desired:

- morning, afternoon, afternoon, morning
- morning, morning, afternoon, afternoon
- afternoon, afternoon, morning, morning
- afternoon, morning, morning, afternoon

We divided the study participants randomly following this counter-balanced design. Participants were guided to report their tiredness before starting the session as scale level from 1 (not tired at all) to 5 (very tired).

For each one of the 2 tests, 50 different images of various difficulty levels were generated. Before starting each test, the subjects reviewed 5-guided introductory questions consisting of obvious cases, mid-level, and difficult cases to familiarize themselves with the tests as well as the graphical user interface. The study software was designed to collect a good amount of valid eye gaze samples per each test pattern. In this paper, the software shows patterns to collect up to 200 valid eye gaze samples.

Sixteen subjects (8 males and 8 females) ranging from 19 yrs old to 47 yrs old were recruited for the study. The average age of the female subjects was 27.1 (±11.1) and of male subjects was 21.4 (±2.3). Six out of the sixteen subjects had corrected vision (glasses). Human subject recruitment and data collection were done according to a protocol approved by the Oak Ridge Site-Wide Internal Review Board. All participants signed an informed consent form.

2.4 Data Analysis

Following the same data analytics methodology used in the previous study, we applied Hidden Markov Models (HMMs) to analyze gaze velocity information. However, unlike the previous study where the number of hidden states of HMMs was five for all subjects, we applied an adaptation rule to find the optimal number of hidden states for each human subject to ensure better identification performance as well as avoid convergence failure. The number of optimal states was determined empirically by monitoring the log-likelihood values of the HMMs. Gaze velocities from 50 test patterns were used to train HMMs using 2, 3, 4, and 5 states for each subject, to choose the HMM showing the highest Log-likelihood.

We trained 16 HMMs, one per each subject, and computed the log-likelihood from each HMM for the test data. We applied leave-one-case-out cross-validation sampling. That is, each case (i.e., gaze sequence) was excluded once to serve as test case. The remaining cases were used to train 16 HMMs thus producing 16 user profiles, one for each subject. The 16 user profiles were applied to the test case. The test case was assigned to the subject whose HMM user profile gave the most likely output. The RHmm software package [16] was used in this study.

3. RESULTS

Participants' information and their eye tracker calibration errors as well as tiredness are reported in Table 1. Note that all participants fulfilled 9 valid calibration points. It is observed that there is no difference on average calibration error between morning sessions and afternoon sessions. In addition, the Pearson's correlation between subjects' tiredness and average calibration error is 0.157, therefore, very weak evidence of correlation between gaze tracking calibration performance and subjects' tiredness. Likewise, the average calibration error for participants with (=32.63) and without (=35.77) corrected vision was not significantly different. These findings suggest the gaze calibration process was done very thoroughly in this study.

Table 1. Eye tracker calibration errors and participants' tiredness per each session of the study.

			Test 1 AM		Test 1 PM		Test 2 AM		Test 2 PM	
Gender	**Age**	**Vision**	**Tired**	**Error**	**Tired**	**Error**	**Tired**	**Error**	**Tired**	**Error**
Female	16	None	3	56.51	3	45.66	2	28.12	1	34.66
Female	18	Glasses	3	55.67	1	57.53	5	26.61	1	41.13
Female	20	None	2	23.83	2	48.82	4	61.08	1	34.18
Female	22	None	2	25.85	3	42.76	5	38.44	4	42.76
Female	22	None	2	27.88	1	36.30	1	33.24	5	65.87
Female	35	Glasses	1	22.97	1	30.42	3	28.76	1	24.40
Female	37	None	2	64.61	2	35.92	3	34.15	3	34.40
Female	47	Glasses	3	31.58	1	30.03	3	25.33	2	28.80
Male	18	Glasses	3	17.87	2	51.51	3	24.67	3	28.40
Male	20	None	3	25.63	4	40.03	2	30.58	3	38.53
Male	21	None	1	26.70	2	51.51	2	41.78	1	28.45
Male	21	None	3	25.62	3	43.67	2	26.98	3	28.64
Male	21	None	2	21.98	3	24.15	3	34.29	4	25.31
Male	22	Glasses	1	19.85	1	46.64	1	48.43	1	39.12
Male	22	None	3	21.11	1	39.68	2	16.90	2	24.13
Male	26	Glasses	3	29.49	1	23.18	2	25.24	1	25.40
Average			2.3	31.07 ±14.37	1.9	40.49 ±10.00	2.7	32.79 ±10.67	2.3	34.01 ±10.48

Table 2 summarizes the identification accuracy for each test pattern and each cross-validation scheme investigated. We explored two cross-validation scenarios. The first scenario, considered all gaze sequences collected for an individual as one universal set to derive its biometric. The second scenario treats morning and evening data as separate tests for training and testing. The second scenario aims to determine if biometric signatures change with time, even within the same day. Please note that with 16 subjects, random guessing for subject identification would be 6.25% (=1/16).

In all experiments reported in Table 2, identification accuracy was statistically significantly higher than random guessing (=0.0625). However, compared to our previously reported study, the identification accuracy for Test 1 was lower. This difference most likely can be attributed to different subjects and more precise collection of gaze data. Also, identification accuracy from the Leave One Out (LOO) experiment was significantly higher than that for the time-based data split experiment. These findings suggest that we need further investigation of eye-gaze features to achieve robustness as well as time-invariance.

Table 2. Identification accuracy with 95% confidence intervals

Test Pattern	Cross-Validation Scheme		Accuracy
Test 1	Leave One Out	AM Only	41.8% (38.3%~45.1%)
		PM Only	36.6% (33.8%~39.8%)
		AM and PM	33.6% (31.4%~35.9%)
	Data Split	Train AM – Test PM	17.6% (15.1%~20.4%)
		Train PM – Test AM	21.5% (18.8%~24.4%)
Test 2	Leave One Out	AM Only	38.1% (34.6%~41.6%)
		PM Only	37.6% (34.3%~40.8%)
		AM and PM	25.9% (23.8%~28.1%)
	Data Split	Train AM – Test PM	17.6% (15.1%~20.1%)
		Train PM – Test AM	23.8% (20.9%~26.8%)
Test 1 + Test 2	Leave One Out	AM Only	39.9% (37.6%~42.3%)
		PM Only	37.1% (34.8%~39.4%)
		AM and PM	29.8% (28.2%~31.3%)
	Data Split	Train AM – Test PM	17.6% (15.1%~20.1%)
		Train PM – Test AM	22.6% (19.9%~25.6%)

Table 2 shows the identification accuracy based on Test 2 was often inferior to that for Test 1 in LOO tests. Also, the subjects' average gaze velocity for test pattern 1 was consistently higher than that for test pattern 2, for morning and afternoon reading sessions (Test_1_AM=49.4 pixels, Test_1_PM=39.4 pixels, Test_2_AM=53.3 pixels, Test_2_PM=50.1 pixels). This result supports our previous study findings that higher identification accuracy correlates with a lower average gaze velocity. Gaze data observed in the morning resulted in better identification accuracy suggesting that visual fatigue may compromise somewhat the quality of the gaze-based biometric. However, the decline in identification accuracy observed when using the data collected in late afternoon was not statistically significant.

In accordance to Table 1, Table 3 lists identification accuracy of LOO tests by sessions, along with average accuracy values between female and male subjects, young and old subjects, and subjects with glasses and no glasses. Differences between subgroups were not significant. Pearson's correlation coefficient between the tiredness and identification accuracy was 0.109. Likewise, correlation coefficient between the average calibration error and identification accuracy was -0.005, resulting in very weak correlation or no correlation observed in between.

4. CONCLUSION

The follow up study performed in this paper consists mainly of repetition of tests with different subjects and different time intervals between the eye tracking data collection sessions. Consistency of results regardless of gender, age and vision condition demonstrated good potential that an individual's gaze during perceptual organization visual search tasks can be utilized as a biometric. However, lower performance in time-based data split pointed out the major challenge in time-invariance to achieve a temporally robust user authentication system. The result that identification accuracy in morning session is moderately higher than that in afternoon session gives some clue between fatigue and consistency of eye gaze patterns. These findings will lead to more accurate personalized modeling and identification based on eye gaze, as well as have implications in Radiology training and the development of personalized e-learning environments.

Table 3. Leave One Out test accuracy of identification of individual readers and their summations of groups by gender, age, and vision conditions.

Gender	Age	Vision	Test 1 AM	Test 1 PM	Test 2 AM	Test 2 PM	Average
Female	16	None	0.080	0.160	0.120	0.320	0.170
Female	18	Glasses	0.700	0.580	0.580	0.340	0.550
Female	20	None	0.100	0.240	0.260	0.240	0.210
Female	22	None	0.460	0.080	0.380	0.660	0.395
Female	22	None	0.260	0.260	0.440	0.160	0.280
Female	35	Glasses	0.120	0.360	0.140	0.320	0.235
Female	37	None	0.820	0.580	0.640	0.760	0.700
Female	47	Glasses	0.640	0.840	0.660	0.580	0.680
Male	18	Glasses	0.700	0.140	0.500	0.140	0.370
Male	20	None	0.100	0.060	0.060	0.180	0.100
Male	21	None	0.040	0.400	0.500	0.140	0.270
Male	21	None	0.940	0.740	0.280	0.660	0.655
Male	21	None	0.800	0.880	0.720	0.740	0.785
Male	22	Glasses	0.640	0.260	0.200	0.360	0.365
Male	22	None	0.100	0.040	0.360	0.040	0.135
Male	26	Glasses	0.180	0.240	0.260	0.380	0.265
Female Subjects			0.398	0.388	0.403	0.423	0.403 (0.355, 0.448)
Male Subjects			0.438	0.345	0.360	0.330	0.368 (0.323, 0.413)
Young (<21.5) Subjects			0.433	0.400	0.378	0.345	0.389 (0.338, 0.435)
Old (>21.5) Subjects			0.403	0.333	0.385	0.408	0.382 (0.330, 0.430)
Subjects with glasses			0.497	0.403	0.390	0.353	0.411 (0.357, 0.467)
Subjects without glasses			0.370	0.344	0.376	0.390	0.370 (0.328, 0.412)

ACKNOWLEDGMENTS

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

[1] Ortega-Garcia, J., Bigun, J., Reynolds, D., and Gonzalez-Rodriguez, J., "Authentication gets personal with biometrics." Signal Processing Magazine, *IEEE* 21(2), 50-62 (2004).

[2] Goudelis, G., Tefas, A., and Pitas, I., "Emerging biometric modalities: a survey." Journal on Multimodal User Interfaces 2(3-4), 217-235 (2008).

[3] Kasprowski, P., and Ober, J., "Eye movements in biometrics." In *Biometric Authentication*, Springer Berlin Heidelberg, 248-258 (2004).

[4] Deravi, F., and Guness, S.P., "Gaze Trajectory as a Biometric Modality." In *BIOSIGNALS*, 335-341 (2011).

[5] Bednarik, R., Kinnunen, T., Mihaila, A., and Fränti, P., "Eye-movements as a biometric." In *Image analysis*, Springer Berlin Heidelberg, 780-789 (2005).

[6] Galdi, C., Nappi M., Riccio D., Cantoni V., and Porta, M., "A New Gaze Analysis Based Soft-Biometric." In *Pattern Recognition*, Springer Berlin Heidelberg, 136-144 (2013).

[7] Cantoni, V., Porta, M., De Maio, L., Distasi, R., and Nappi, M., "Towards a novel technique for identification based on eye tracking." 2012 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 1-4 (2012).

[8] Rigas, I., Economou, G., and Fotopoulos, S., "Biometric identification based on the eye movements and graph matching techniques." Pattern Recognition Letters 33(6), 786-792 (2012).

[9] Holland, C. D., and Komogortsev, O. V., "Complex eye movement pattern biometrics: Analyzing fixations and saccades." 2013 International Conference on Biometrics (ICB), 1-8 (2013).

[10] Rigas, Ioannis, George Economou, and Spiros Fotopoulos. "Human eye movements as a trait for biometrical identification." In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, IEEE, 217-222 (2012).

[11] Yoon, H, Carmichael, T. R., and Tourassi, G., "Gaze as a biometric" *Proc. SPIE* 9037, Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment, 903707 (2014); doi:10.1117/12.2044303.

[12] Kurylo, D. D., Corkin, S., and Growdon, J. H., "Perceptual Organization in Alzheimer's Disease." Psychology and Aging, 562 (1994).

[13] Wertheimer, M., "Laws of Organization in Perceptual Forms", A Sourcebook of Gestalt Psychology, 71-88 (1938).

[14] "S2 Eye Tracker," Mirametrix. Available: http://mirametrix.com/products/eye-tracker/.

[15] Duchowski, A. T., "Eye Tracking Methodology". London: Springer (2007).

[16] Taramasco, O. and Bauer, S., "RHmm: Hidden Markov Models Simulations and Estimations," Available: http://CRAN.R-project.org/package=RHmm.