A Visual Attention Approach to Personal Identification

Anthony Maeder and Clinton Fookes
School of Electrical & Electronic Systems Engineering
Queensland University of Technology
GPO Box 2434 Brisbane, 4001 QLD Australia
{a.maeder,c.fookes}@qut.edu.au

Abstract

This paper describes the use of visual attention characteristics, monitored by gaze tracking during presentation of a known visual scene to a viewer, as a biometric for distinguishing between individual viewers. The positions and sequences of gaze locations during viewing may be determined by overt (conscious) or covert (sub-conscious) viewing behaviour. Methods to quantify the spatial and temporal patterns established by the viewer for a particular image are proposed, and distance measures between these are established. Experimental results suggest that both types of gaze behaviours can provide simple and effective biometrics for this application.

1. Introduction

Establishing the identity of a person when validating a request for access to a secure environment is an important and challenging application in the world of today. To complement the widespread practice of data-based authentication using private information (such as passwords) which is prone to attack or theft, some physical aspects of the human individual known as "biometrics" may also be used (such as fingerprints) [2]. While these are harder to violate than purely data-based methods, it is not impossible. Safer forms of biometrics would be based on non-visible and non-physiological information hidden within the person, such as behaviour or thought processes. Here we propose to exploit the personal aspects of visual attention processes determined by means of monitoring eye movements, in an attempt to develop such a biometric.

2. Visual Attention and Eye Movement

It has been estimated that approximately 80% of the information a human receives comes from visual inputs [13]. Visual information therefore plays a major role in our everyday life activities and also in our ability to make decisions based on this information. Visual attention is a complex and expanding research field which investigates aspects of human vision and how it relates to higher cognitive, psychological and neurological processes. The concept of attention, or conscious selecting and directing of perceptual information intake, arises because finite physical human limitations prevent us from perceiving all things at once. Rather, attention is used to focus our mental capacities on small portions of the sensory input gamut so that we can successfully assimilate the stimulus of interest [6].

The human visual system relies on positioning of the eyes to bring a particular component of the visible field of view into high resolution. This permits the person to view an object or region of interest near the centre of the field in much finer detail. In this respect, visual attention acts as a "spotlight" effect [9]. The region viewed at high resolution is known as the foveal region and is much smaller than the entire field of view contained in the periphery. Viewing of a visual scene consists of a sequence of brief gaze periods (typically 100-500ms) of visual concentration (fixations) at specific locations in the field of view, interspersed with sudden movements of the eyes (saccades) to reposition the foveal region at the next point of attention. This process provides the brain with detailed visual information over a succession of these fixation-saccade events covering a few comparatively small areas in the field of view, from which a "conceptual" image of the visual scene is constructed by combining these with the large area of low resolution information gained from the periphery. The fixation-saccade events may be consciously directed by the viewer to visit a sequence of specific points in the scene (overt), or else may be allowed to be directed sub-consciously by the brain according to its choice of points of interest (covert) [8].

In order to understand visual attention processes better, methods have been devised to track gaze location through eye movements: a simple approach uses a video camera image to recover the 3D orientation of the eye. By observing where and when a person's gaze is directed, it is possible to establish the fixation-saccade path followed by the viewer. This provides insights about what the viewer found interesting (i.e. what captured their attention) and perhaps reveal how that person perceived the visual scene they were viewing [6]. If the viewer is undertaking a defined task, such

as following a prescribed sequence of gaze locations, this is equivalent to providing a password. If a task is not specified, the path denotes the pattern of visual interest for that individual, corresponding to the scene being viewed. Either of these situations can provide a suitable foundation for a biometric, and both are explored here.

3. Biometric Gaze Measurements

Biometric research is a rapidly evolving field due to the increased demand on modern society to identify or authenticate an individual [11]. Whether for the purpose of entering a restricted room or building, or for credit card payments, there are numerous needs to validate identity. Biometric identification has adopted a wide range of human features and characteristics which may be used to identify or authenticate an individual with much stronger certainty. Typical examples include the fingerprint, iris and retina scan, voice print [12], face geometry [4, 10], DNA, handwriting, or even a person's typing style which can be used as a keyboard behavioral signature [1]. An identification system is also often formed as a combination of different traditional identification and biometric measures [5]. This paper describes another such biometric measure based on distinguishing between visual attention patterns for individual viewers.

As described earlier, a viewer will build up a perception or "conceptual image" of the visual scene by a sequence of fixation-saccade events. The spatial and temporal pattern of these events for a given visual scene varies widely between different people and accordingly can be used to compute a visual attention "signature" of each individual. These signatures can be formulated by quantitative analysis of the individual's gaze data [7]. Two possible ways that such signatures could be constructed are as follows.

1. The viewer could be presented with a known picture for which they had decided upon a personal sequence of fixation points, already informed to the authentication system in a training phase. The viewer would consciously (overtly) direct their gaze to these points in the order established, while the authentication system tracked the sequence of points.

2. The viewer could be presented with a known picture for which their unconscious (covert) pattern of gaze points when inspecting the picture had previously been captured by the authentication system in a training phase. The viewer would simply view the picture passively and allow the authentication system to collect the pattern of gaze points occurring naturally.

Both methods require some assumptions about the properties of eye movement and gaze tracking, as follows:

1. The operating characteristics of the tracking device need to be sufficient to allow the necessary detail in spatial and temporal resolution. A sampling rate of approximately 50ms or less will achieve the temporal requirement for detecting fixations generally. Spatial localisation to around 10% of the overall scene linear dimensions is judged sufficient to establish fixation points appropriate to this application.

2. A sufficient number of successive gaze points is required to allow unique characterisation of an individual within the population, and to override the effect of involuntary errors in gaze location and sequence, without requiring too many points or too long a gaze duration for practical purposes. A comfortable maximum viewing time for an image is approximately 20 secs, after which some fatigue/boredom effects typically occur. This allows for at least 40 fixations to be measured.

3. The gaze pattern followed by a viewer needs to be reasonably similar on different occasions. Substantial evidence from experiments with numbers of viewers indicates this expectation is realistic [14].

4. The covert viewing gaze patterns for different viewers of the same scene need to differ significantly and consistently from each other to allow effective detection. Evidence in the literature [6] suggests that this is the case.

5. An efficient and unbiased technique is needed to establish the distance between two gaze patterns, to allow easy decision-making for establishing similarity. A signature of relatively few numerical values representing a compacted form of the gaze pattern would be appropriate, and can be compared using ranking or matching type procedures.

6. A distance measure between gaze patterns also needs to make allowance for involuntary errors, such as sequence or duration variations. The signature should therefore be constructed to constrain the effects of same viewer variations.

Our approach to developing the biometrics described above, and establishing their viability subject to the above assumptions, is described in the sections below.

4. Methodology & Outcomes

The experimental methodology adopted consisted of recording gaze data of three different viewers for a particular image of an outdoor scene. This image is shown in Figure 1. For each image, the viewer was directed to examine the scene both consciously (overtly) and unconsciously

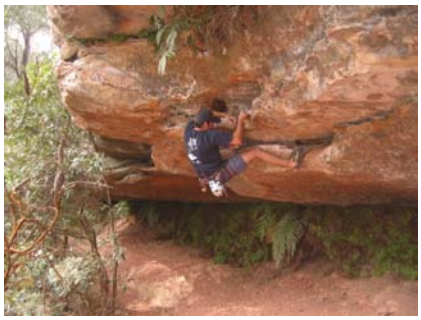


Figure 1. Rockclimb image used in the gaze-tracking experiments.

(covertly). The former approach relied on the observer to gaze at a select number of points and in a particular sequence known only to them. In the latter approach, the viewer was free to examine the image in their natural manner, i.e. as directed by their personal visual attention processes. For each case, the gaze-tracking experiment was repeated three times, each occasion being separated from the others by some other visual tasks to reduce the influence repetition.

4.1. Gaze-Tracking Device

The device used to record eye movements during these experiments was an EyeTech video-based corneal reflection eye tracker. This device is normally used for point-of-regard measurements, i.e. those that measure the position of the eye relative to 3D space rather than relative to the head [15]. The method of operation relies on tracking the corneal reflection from an infra-red light source, as it is invisible to the human eye and non-distracting. Although four separate reflections are formed during the reflection of a light source, referred to as Purkinje reflections [3], only the first Purkinje reflection is utilized in the video-based tracker and is located relative to the location of the pupil center using image processing techniques.

The gaze-tracker utilized operated at 15 frames per second, resulting in a sample of the observer's gaze direction approximately every 67ms. The experiments were conducted using an image and screen resolution of 1024×768 pixels.

4.2. Overt Experiment

For the overt experiments, each viewer was directed to look at six specific points in the Rockclimb image. As only a small number of people were utilised, the same six points were used in each case, however, the viewing sequence was re-arranged substantially for each person. This essentially provided a unique pattern which could be used to distinguish very clearly between the different individuals.

For both the overt and covert approaches to developing a biometric measure, the system may be considered to have two operating modes. The first mode consists of the off-line training wherein the library of gaze "signatures" is compiled for each individual. The second mode is the actual online operation of the authentication system. In this mode, an observer's gaze data is recorded online and compared against the database of signatures to identify or authenticate an individual. The off-line training mode for the overt experiment consisted of building the database of points to be viewed and their sequences manually. The recorded gaze data gathered during the online mode was then compared against the database to authenticate the individual.

For the given Rockclimb image, the observer was directed to gaze at each of the six specific points for approximately one second each and in their prescribed sequence. The gaze data (sampled at 15fps) was then passed through a spatial clustering algorithm to extract any fixations with an approximate viewing time of 1.0 secs and a tolerance of 0.3 secs, i.e. containing roughly 10 to 19 samples of gaze data. The locations of these fixations were then compared against the database, along with their viewing sequence, to authenticate the individual.

Figure 2 shows the original gaze data and fixations extracted for all subjects viewing the Rockclimb image, for comparison purposes. After clustering the original gaze samples with a constraint on the time of fixation, only six fixations were obtained and are plotted against the Rockclimb image. These fixations correspond to the six points the observer was directed to examine. Similar fixation plots were also obtained across the three repeat experiments from the same people.

For the overt experiment, a successful authentication is the case when the location of the fixations extracted from the clustering algorithm (subject to a time constraint of approx one second) is within a small threshold of the locations stored in the database. The sequence of viewing these fixations also had to correspond to those stored for each individual. The authentication process was successful in all cases.

4.3. Covert Experiment

Comparing the recorded gaze signatures of a person with those in a database is a straight forward procedure for the overt experiments. A valid authentication is simply the case when the location of the clustered fixations, and their viewing sequence, matches those stored in the database. However, the comparison of the gaze patterns for the covert experiments is a much more complicated and problematic endeavour. This is due to the inherent variations that exist not only between viewing patterns of different people, but more significantly, between different scans of the same person. This inter- and intra-variability makes the development of a simple authentication process extremely difficult. Consequently, more sophisticated measures need to be developed in order to truly identify commonalities, if any, that exist

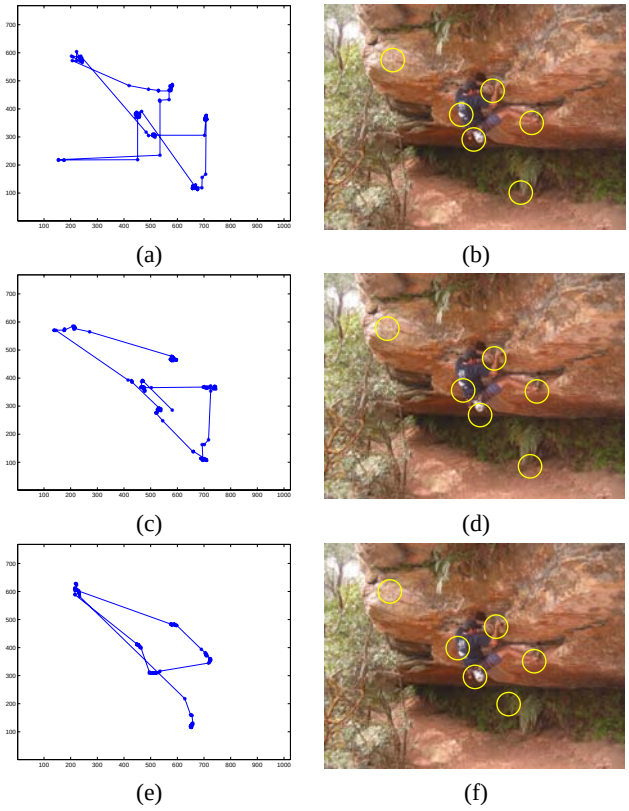


Figure 2. Overt Experiment: (a,c,e) show the original gaze samples while (b,d,f) show the extracted fixations plotted against the Rockclimb image for all three viewers.

between different scans of the same observer and between observers. The measures and data that will be discussed in this section are presented to ascertain if any patterns exist across the same observers and/or different observers. The existence of any such patterns will determine if a person's covert visual attention processes can in fact be used as a successful biometric. At this stage, no authentication procedure has been implemented on these measures.

Figure 3 presents some sample data for the covert experiments. Plots (a), (b) and (c) show fixations of Scan 1 for Person 1, Person 2, and Person 3 respectively, plotted against the Rockclimb image. The variations between observers in these cases are quite apparent. Figures (d), (e) and (f) represent the three repeated scans for Person 1 on the Rockclimb image. These three scans do appear to have many similarities after a visual comparison, however, there is still obviously some diversity between the three repeated scans, i.e. intra-viewer variation. Moreover, the plots in Figure 3 do not contain any information about the sequence these points were viewed in.

The measures employed to determine the similarities between different scans in this experiment are outlined below.

- $\mathbf{D}$: is used to measure or count the number of common fixations between any two scans. A common fixation is one which coincides spatially (within a given threshold) with another fixation in the second scan.

- $\mathbf{D}'$: encompasses the order information between two scans. It is a measure or count of the number of position changes in order for the common fixations viewed in both scans. This count is also normalised by the square of the number of common fixations, i.e. $\mathbf{D}^2$.

- $\mathbf{D}^*$: is used quantify the difference in visit periods of the first five fixations in two scans. This is achieved by the computation of a SAD score between the number of gaze samples in each of the first five fixations, $\mathbf{D}^* = \sum_{i=1}^{N} \left| p(F_i^1) - p(F_i^2) \right|$, where $N = 5$, F_i^1 and F_i^2 are the ith fixations of scan 1 and scan 2, and $p(.)$ is the number of gaze samples in each fixation.

- $\mathbf{D}^+$: quantifies the difference between the number of revisits of each of the first five fixations. The revisits is represented as a count of the number of times a viewer "revisits" one of the first five fixations during the entire viewing duration. This measure is implemented with another SAD score, $\mathbf{D}^+ = \sum_{i=1}^{N} \left| r(F_i^1) - r(F_i^2) \right|$ where $N = 5$, and $r(.)$ is the number of revisits for each fixation.

The first two distance measures are used to assess the commonality of the fixations and their viewed sequence between any two scans. $\mathbf{D}$ simply counts the number of common fixations while $\mathbf{D}'$ counts the number of position changes in the order of the fixation viewing sequence. This count is normalised by $\mathbf{D}^2$ to penalize those counts with a small number of common fixations. Thus, viewing sequences with a certain number of sequence changes and a large number of common fixations will be more significant than those with the same count for sequence changes and a small number of fixations.

The second two distance measures described above are used to quantify slightly different aspects of the visual attention process. $\mathbf{D}^*$ measures the difference between the period (or the number of sample points) of the first five fixations via a SAD score. Empirical evidence has shown that the most prominent of all fixations generally occur within the first five viewed. Later fixations as a general rule contain a much smaller period. These is the reason for only comparing the first five fixations. Note however, that these first five fixations can be revisited during the entire viewing duration (10 secs). Thus, $\mathbf{D}^*$ computes the SAD score between the total number of sample points between the first five fixations obtained over the entire viewing sequence. The last distance measure $\mathbf{D}^+$ quantifies the difference between the revisiting habits of the viewers. Different viewers will have different

underlying psychological and cognitive process which direct them in to revisit points or regions of interest to them in various manners. Thus, $\mathbf{D}^+$ measures the differences in the number of revisits of the first five prominent fixations via another SAD score.

Table 1 presents the $\mathbf{D}$ distance measures calculated between all possible scan combinations. The larger the value, the more similar the scans as they share a larger number of common fixations. The diagonal values in this table signifies comparison between a scan and itself, which simply yields the total number of fixations in that person's scan. From a simple visual inspection of the table, there is evidence that the $\mathbf{D}$ scores of intra-viewer comparisons (not including the diagonals) are generally larger than inter-viewer comparisons, except for a few cases. This measure however, is by no means sufficient enough to adequately distinguish between different viewers.

		P1			P2			P3		
		S1	S2	S3	S1	S2	S3	S1	S2	S3
P1	S1	14	9	9	6	8	7	5	5	3
	S2		10	8	3	7	5	3	3	2
	S3			11	3	5	5	4	4	3
P2	S1				15	7	9	4	3	3
	S2					13	6	2	2	2
	S3						13	4	5	3
P3	S1							9	7	6
	S2								10	4
	S3									7

Table 1. Quantitative measures: $\mathbf{D}$ calculated between scans of all viewers.

Table 2 presents the $\mathbf{D}'$ scores which signify the differences in order of the fixation viewing sequence, normalised by the square of the number of common fixations. In this case, the smaller the $\mathbf{D}'$ score, the more similar the scan. The diagonal values in this table are all zeros as a scan compared with itself will obviously have the exact same viewing sequence, so the changes between them will be zero. From a visual inspection of the table, it can be scene that the intra-viewer comparisons (not including the diagonals) are almost all smaller than inter-viewer comparisons, except for five distinct cases which the score is zero. These are cases when the viewing sequence of the common fixations are the same. However, from a cross-comparison with Table 1, it is clear that these instances generally only contain two or three common fixations between the two scans, so the possibility of having the same sequence is significantly higher. Future work will look at penalizing these instances rather than assigning it a zero value which would suggest similar scans, when in actual fact they are far from similar due to the small number of common fixations.

Table 3 presents the $\mathbf{D}^*$ SAD scores which measures the difference in the visit period (or number of gaze samples) between the first five fixations of any two scans, whatever location those fixations may be. This measure essentially compares how long a person views each of the first five fixations, which have been shown empirically to be the most prominent fixations. Similarly Table 4 presents the SAD scores between the number of revisits of the first five fixations. For the values in both Table 3 and 4, the smaller the score, the more similar the scans. Once again, from a visual inspection of both of these tables, there is an obvious trend where the intra-viewer comparisons are generally smaller than inter-viewer comparisons, except for a few cases.

		P1			P2			P3		
		S1	S2	S3	S1	S2	S3	S1	S2	S3
P1	S1	0	37	37	83	0	61	80	80	111
	S2		0	16	111	41	80	0	0	0
	S3			0	111	40	80	63	63	111
P2	S1				0	82	37	63	111	111
	S2					0	28	0	0	0
	S3						0	125	80	222
P3	S1							0	25	41
	S2								0	40
	S3									0

Table 2. Quantitative measures: $\mathbf{D}' \times 10^3$ calculated between scans of all viewers.

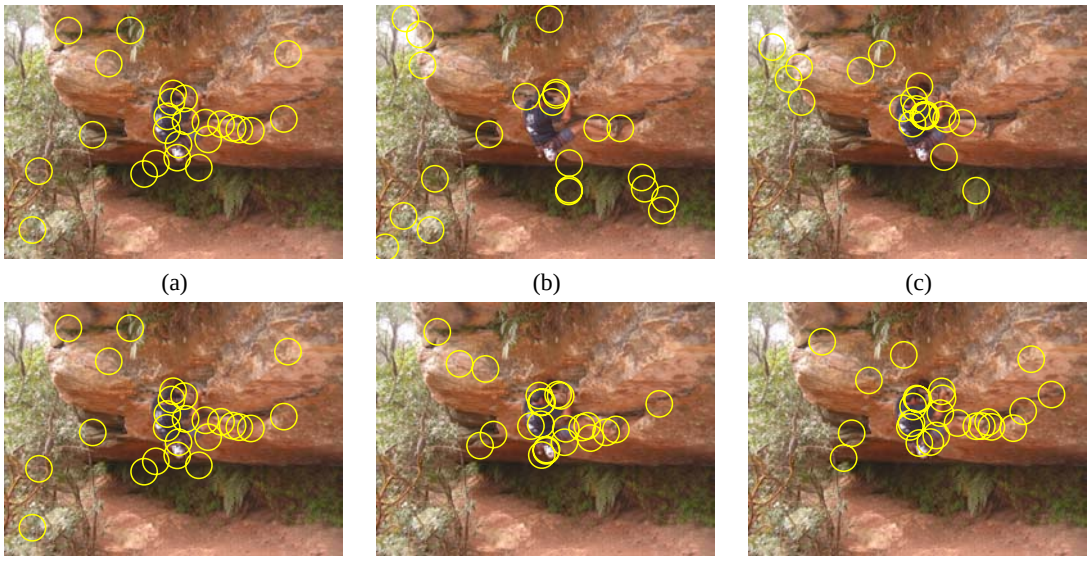
		P1			P2			P3		
		S1	S2	S3	S1	S2	S3	S1	S2	S3
P1	S1	0	27	56	65	45	72	87	85	59
	S2		0	53	80	54	81	80	78	48
	S3			0	63	69	82	69	85	49
P2	S1				0	38	31	76	108	82
	S2					0	51	78	72	68
	S3						0	95	115	83
P3	S1							0	36	44
	S2								0	54
	S3									0

Table 3. Quantitative measures: $\mathbf{D}^*$ calculated between scans of all viewers.

		P1			P2			P3		
		S1	S2	S3	S1	S2	S3	S1	S2	S3
P1	S1	0	3	8	8	7	9	8	8	4
	S2		0	5	9	10	12	7	5	3
	S3			0	10	11	13	8	6	4
P2	S1				0	3	3	8	8	8
	S2					0	4	7	9	7
	S3						0	9	11	9
P3	S1							0	6	4
	S2								0	4
	S3									0

Table 4. Quantitative measures: $\mathbf{D}^+$ calculated between scans of all viewers.

From the preliminary results presented thus far, there are some obvious trends in the data to suggest that a scan from one person is in actual fact more similar to other scans from that same person than to scans from other people. Future work will repeat these experiments on a greater range of images and subjects to provide a richer set of base data. Investigation will also be carried out on various statistical approaches to extract more concrete conclusions from these results, as well as examining more advanced comparative measures.



(a) (b) (c)

(d) (e) (f)

Figure 3. Covert Gaze Data: Fixations for Scan 1 of (a) Person 1, (b) Person 2, and (c) Person 3, for the Rockclimb image. Figures (d), (e) and (f) represent the three repeated scans for Person 1, i.e. intra-viewer variation.

5. Conclusion

The above experimental results would be enhanced by increasing the number of subjects tested, and refining the resolution of performance parameters associated with the complexity of the gaze tracking data and the corresponding compacted signature information. The simple visual attention biometrics described here could be extended in a number of ways, to increase the sophistication of the signatures extracted, or to introduce sufficient variability to make attacks harder: for example, expanding the number of different images available to a single viewer for the covert gaze method. The applicability of the method under highly constrained circumstances such as exist for present PC or PDA systems, using typical cheap camera technology for eye tracking and allowing no choice of images, is currently being investigated.

References

[1] T. Alexandre. Biometrics on smart cards: An approach to keyboard behavioural signature. *Future Generation Computer Systems*, 13:19–26, 1997.

[2] J. Ashbourn. *Biometrics: Advanced Identity Verification: The Complete Guide*. Springer, London, 2000.

[3] H. Crane. The purkinje image eyetracker, image stabilization, and related forms of stimulus manipulation. In D. Kelly, editor, *Visual Science and Engineering: Models and Applications*, pages 13–89, New York, NY, 1994. Marcel Dekker, Inc.

[4] J. Crowley. Vision for man-machine interaction. *Robotics and Autonomous Systems*, 19:347–358, 1997.

[5] U. Dieckmann, P. Plankensteiner, and I. Wagner. Sesam: A biometric person identification system using sensor fusion. *Pattern Recognition Letters*, 18:827–833, 1997.

[6] A. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer, London, 2003.

[7] J. Goldberg and X. Kotval. Computer interface evaluation using eye movements:methods and constructs. *International Journal of Industrial Ergonomics*, 24:631–645, 1999.

[8] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506, 2000.

[9] L. Itti and C. Koch. Feature combination strategies for saliency-based visual attenion systems. *Journal of Electronic Imaging*, 10(1):161–169, January 2001.

[10] A. Nikolaidis and I. Pitas. Facial feature extraction and pose determination. *Pattern Recognition*, 33:1783–1791, 2000.

[11] N. Ratha, J. Connell, and R. Bolle. Biometrics break-ins and band-aids. *Pattern Recognition Letters*, 24:2105–2113, 2003.

[12] C. Rebman, M. Aiken, and C. Cegielski. Speech recognition in the human-computer interface. *Information and Management*, 40:509–519, 2003.

[13] C. Roux and J.-L. Coatrieux, editors. *Contemporary Perspectives in Three-Dimensional Biomedical Imaging*, volume 30 of *Studies in Health Technology and Informatics*. IOS Press, Netherlands, 1997.

[14] A. Yarbus. *Eye Movements and Vision*. Plenum Press, New York, NY, 1967.

[15] L. Young and D. Sheena. Survey of eye movement recording methods. *Behavior research methods and instrumentation*, 7(5):397–439, 1975.