Ataques directos usando imágenes falsas en verificación de iris

Virginia Ruiz-Albacete, Pedro Tome-Gonzalez, Fernando Alonso-Fernandez, Javier Galbally, Julian Fierrez, and Javier Ortega-Garcia

Biometric Recognition Group - ATVS
Escuela Politecnica Superior - Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - 28049 Madrid, España
http://atvs.ii.uam.es
{virginia.ruiz, pedro.tome, fernando.alonso, javier.galbally, julian.fierrez, javier.ortega}@uam.es

Resumen En este artículo se estudian las vulnerabilidades de un sistema de reconocimiento de iris frente ataques directos. Para ello se ha creado una base de datos con iris falsos, partiendo de los iris reales de la base de datos BioSec. Usando una impresora comercial, las imágenes de iris han sido impresas y posteriormente capturadas con nuestro sensor de iris. Para los experimentos usamos un sistema de reconocimiento de libre distribución. Basándonos en los resultados obtenidos tras las pruebas en distintos modos de operación, demostramos que el sistema es vulnerable frente ataques directos, resaltando la importancia de desarrollar contramedidas para este tipo de acciones fraudulentas.

Palabras Clave: Biometría, reconocimiento de iris, ataques directos, iris falso

1. Introducción

El importante aumento del número de aplicaciones que requieren una correcta identificación de individuos ha provocado un creciente interés en la biometría. El término *biometría* hace referencia al reconocimiento de forma automática de un individuo, basándose en sus rasgos físicos (por ejemplo, huellas dactilares, iris, geometría de la mano, orejas, huella palmar) o a características en su comportamiento (como la firma, forma de andar y forma de teclear) [1]. Los sistemas biométricos presentan varias ventajas frente a los métodos de seguridad tradicionales basados en algo que sabes (password, PIN), o algo que tienes (tarjeta, llave). En ellos no se necesita conocer una clave (que puede ser olvidada) ni requiere un instrumento (véase una llave que puede ser perdida o robada), sino que realiza un reconocimiento basado en lo que uno es. Entre todas las técnicas biométricas, el reconocimiento de iris ha sido tradicionalmente considerado como uno de los métodos más fiables y precisos de los disponibles [2]. Además, cuenta con la ventaja de ser un rasgo bastante estable a lo largo de la vida de una persona y permite una identificación no invasiva, ya que se trata de un órgano visible de forma externa [3].

Sin embargo, a pesar de todas estas ventajas, los sistemas biométricos presentan algunos inconvenientes, tales como [4]: *i*) la falta de privacidad (por ejemplo, todo el mundo conoce nuestra cara y puede obtener nuestra huella), o *ii*) el hecho de que un rasgo biométrico no pueda ser reemplazado (mientras que una clave olvidada puede ser fácilmente redefinida, una huella dactilar no puede ser regenerada si ha sido robada). Además, los sistemas biométricos son vulnerables a ataques externos, lo que puede reducer su nivel de seguridad. En [5] se identifican 8 puntos posibles de ataque a un sistema de reconocimiento biométrico. Estos puntos de vulnerabilidad son descritos en la Figura 1, y pueden ser agrupados, de forma general, en dos grupos principales:

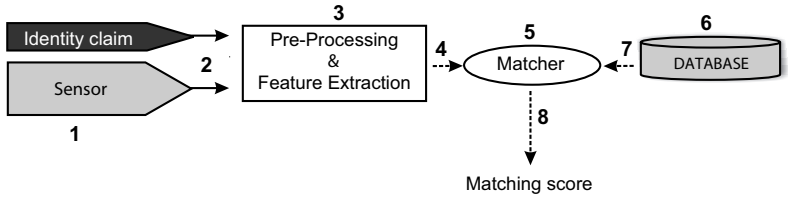


Figura 1. Arquitectura de un sistema automático de verificación biométrica. Los posibles puntos de ataque se han numerado del 1 al 8.

- **Ataques directos**. Este tipo de ataques se basa en el uso de un rasgo biométrico sintético, como por ejemplo un dedo de goma, con el que se intenta acceder al sensor (punto 1 de la Figura 1). Es importante destacar que para este tipo de ataques no es necesario tener ningún conocimiento especial sobre el sistema. De hecho, tiene lugar en el plano analógico, fuera de los límites digitales del sistema, por lo que mecanismos de protección digitales (como la firma digital, el watermarking, etc.) no pueden ser usados.

- **Ataques indirectos**. Este grupo incluye los 7 puntos de ataque restantes identificados en la Figura 1. Los ataques 3 y 5 se pueden llevar a cabo mediante un *troyano* que anule los módulos del sistema. En el ataque 6, se manipula la base de datos del sistema. El resto de ataques (2, 4, 7 y 8) representan posibles puntos débiles en los canales de comunicación del sistema. Al contrario que los ataques directos, en este caso el intruso debe conocer alguna información adicional sobre el funcionamiento interno del sistema, y en muchos casos tener un acceso físico a algún componente de la aplicación. La mayoría de los trabajos referentes a ataques indirectos usan algún tipo de técnica basada en hill-climbing, introducida en [6].

En este trabajo nos centramos en estudiar los ataques directos en sistemas basados en reconocimiento de iris. Para ello hemos creado una base de datos con imágenes sintéticas de iris generadas a partir de los 50 usuarios de la base de datos multimodal BioSec [7]. Este artículo se estructura de la siguiente manera. En la

Sec. 2 se detalla el proceso seguido para la creación de los iris falsos y presentamos la base de datos usada en los experimentos. El protocolo experimental, algunos resultados y comentarios se exponen en la Secc. 3. Finalmente exponemos las conclusiones en la Secc. 4.

IMPRESORA	PAPEL	PRE-PROCESADO [8]
Chorro de tinta	Papel blanco	Equalización de histograma
Láser	Papel reciclado	Filtrado de ruido
	Papel Fotográfico	Apertura/cierre
	Papel de alta calidad	Top hat
	Papel cebolla	
	Cartulina	

Cuadro 1. Pruebas realizadas para la creación de iris falsos.

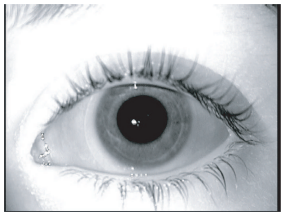
2. Base de Datos de Iris Falsos

Se ha creado para este trabajo una nueva base de datos a partir de imágenes de iris de 50 usuarios extraidos de la base de datos de referencia BioSec [7]. Se ha dividido el proceso en 3 pasos: *i*) Primero se han preprocesado las imágenes originales para obtener una mejor calidad en pasos posteriores, después *ii*) han sido impresas en un papel usando una impresora comercial y por último, *iii*) las imágenes impresas se han presentado al sensor de iris, obteniendo así las imágenes falsas.

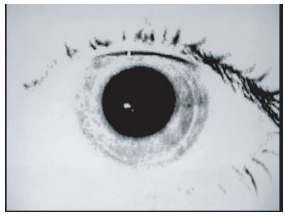
2.1. Método de generación automática de iris

Para lograr una nueva base de datos de forma correcta, es necesario tener en cuenta diversos factores que afectan a la calidad de las imágenes falsas adquiridas. Se han encontrado como principales variables, con una importancia significativa para la calidad de iris: el preprocesado de las imágenes originales, el tipo de impresora y el tipo de papel.

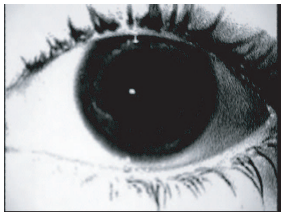
Hemos probado dos impresoras diferentes: una HP Deskjet 970cxi (de chorro de tinta) y una HP LaserJet 4200L (láser). Ambas proporcionan una calidad bastante buena. Por otra parte, hemos observado que la calidad de las imágenes adquiridas depende del tipo de papel usado. En este punto aparece la mayor variedad de opciones. Los tipos de papel probados se indican en la Tabla 1. En nuestros experimentos, el pre-procesado adquiere especial importancia ya que hemos observado que la cámara de iris no captura la mayoría de las imágenes originales impresas que no han sido previamente modificadas. Por ello hemos llevado a cabo diferentes métodos de mejora y refuerzo de las imágenes antes de ser impresas, de forma que sea posible adquirir imágenes falsas de buena calidad. Las opciones probadas se encuentran también resumidas en la Tabla 1. Probando todas las posibilidades con un pequeño grupo de imágenes, la combinación que



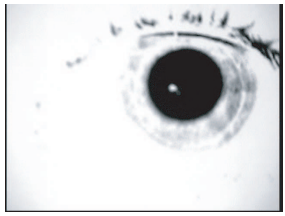
(a) Imagen original - sin mejora



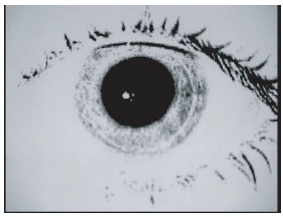
(b) Imagen falsa - sin mejora



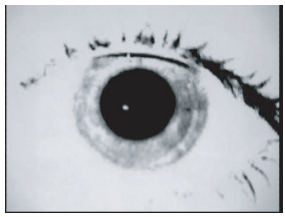
(c) Imagen falsa - Equalización de histograma



(d) Imagen falsa - filtrado de ruido



(e) Imagen falsa - TopHat



(f) Imagen falsa - Apertura+TopHat

Figura 2. Imágenes capturadas con las distintas modificaciones, usando papel de alta calidad e impresora de chorro de tinta.

ha resultado en una mejor segmentación y por tanto en la mejor calidad para la posterior comparación ha sido la impresora a chorro, con papel de alta calidad y un preprocesado de "Apertura-TopHat". En la Figura 2, se muestran ejemplos de diferentes técnicas de preprocesado con este tipo de papel y dicha impresora.

2.2. Base de Datos

La base de datos de iris falsos creada sigue la misma estructura que la base de datos original Biosec. Por tanto, los datos de los experimentos consisten en 50 usuarios $\times$ 2 ojos $\times$ 4 imágenes $\times$ 2 sesiones = 800 imágenes falsas, cada una de las cuales tiene su correspondiente imagen real. La adquisición de las imágenes falsas se ha realizado con la misma cámara que capturó las imágenes de BioSec, una LG IrisAccess EOU3000.

3. Experimentos

3.1. Sistema de reconocimiento

Para nuestros experimentos hemos usado el sistema desarrollado por Libor Masek[1] [9]. Consiste en la secuencia de pasos descritos a continuación: segmentación, normalización, codificación y comparación de plantillas.

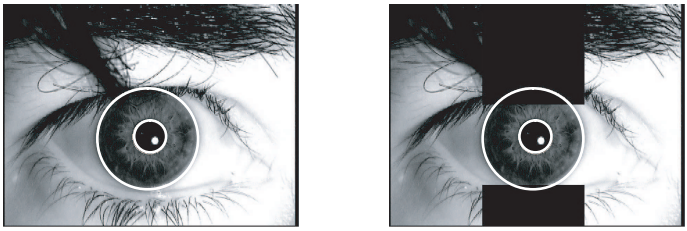
Para la segmentación del iris, el sistema utiliza la transformada circular de Hough para detectar las fronteras de la pupila y del iris. Las fronteras del iris se modelan como dos círculos. El sistema también realiza un paso de detección de párpados. Los párpados son detectados mediante una línea en la parte superior e inferior haciendo uso de la transformada lineal de Hough (ver la Figura 3(a) derecha, en la que la línea de los párpados corresponde al borde del bloque negro). La detección de pestañas se basa en una umbralización del histograma, y está implementado en el código, pero nosotros no lo utilizaremos para nuestros experimentos. A persa de que las pestañas son bastante oscuras comparadas con la región del iris, existen otras zonas del iris con el mismo tono oscuro debido a las condiciones de la imagen. Por ello, un corte basado en este umbral para aislar las pestañas resultaría también en la eliminación de otras partes importantes del iris. Sin embargo, para nuestra base de datos, la oculsión por pestañas no es demasiado prominente.

Para mejorar el funcionamiento de la segmentación, primero pre-estimamos el centroide de la pupila mediante umbralización del histograma, ya que se ha comprobado que la región de la pupila es la de menores niveles de gris de una imagen de iris. Esta pre-estimación nos permite reducir el área de búsqueda de la transformada circular de Hough. Además imponemos tres condiciones a los dos círculos que van a modelar las fronteras de pupila e iris: *i*) a pesar de que estos dos círculos no tienen porqué ser concéntricos, se impone un valor máximo a la distancia entre sus centros; *ii*) no se permite que ninguno de los dos círculos puede tener partes fuera de la imagen de iris; y *iii*) los radios de los círculos no pueden ser similares.

Para la normalización de la región del iris, se utiliza una técnica basada en el "modelo de goma"desarrollado por Daugman [10]. El centro de la pupila se considera el punto de referencia, y basándonos en él, generamos un vector de 2D según el mapeo del radio angular de la región de iris segmentada. En la Figura 3 se muestra un ejemplo de los pasos que se llevan a cabo esta normalización.

La extracción de características se implementa mediante convolución de la imagen de iris normalizada con un filtro Log-Gabor 1D. Las filas del patrón normalizado en 2D se toman como señales en 1D, cada fila correspondiente a un anillo circular de la región del iris. Usa la dirección angular puesto que la independencia máxima tiene lugar en esta dirección. La salida del filtrado se cuantifica en fase en cuatro niveles siguiendo el método de Daugman [10], con cada filtro, produciendo dos bits de datos. La salida de la cuantificación en fase es un código de grises, de modo que al desplazarse de un cuadrante a otro, sólo

[1] El código puede descargarse de forma gratuita en `www.csse.uwa.edu.au/~pk/studentprojects/libor/sourcecode.html`



(a) Imagen original e imagen de ruido

(b) Plantilla normalizada de iris

(c) Máscara de ruido

Figura 3. Ejemplos de los pasos de la normalización.

hay 1 bit cambio. Esto reducirá al mínimo el número de bits en desacuerdo para imágenes del mismo iris ligeramente desalineadas [9]. El proceso de codificación produce una plantilla binaria con un número de bits de información, y la correspondiente máscara de ruido que representa las zonas corruptas (párpados) dentro de los patrones de iris (ver Figura 3 (c)).

Para la comparación entre plantillas (matching), la métrica elegida para el reconocimiento es la distancia de Hamming (HD). La distancia de Hamming empleada incorpora el enmascaramiento de ruido, de modo que sólo los bits sin ruido se utilizan en el cálculo de la misma. La fórmula de la distancia de Hamming modificada viene dada por:

$$HD = \frac{1}{N - \sum_{k=1}^{N} Xn_k(OR)Yn_k} \cdot \sum_{j=1}^{N} X_j(XOR)Y_j(AND)Xn'_j(AND)Yn'_j$$

donde X_j y Y_j son los dos bits a comparar de la plantilla, Xn_j y Yn_j son las correspondientes mascaras de ruido para X_j y Y_j, y N es el numero de bits representado por cada plantilla.

Con el fin de dar cuenta de las incoherencias rotacionales, cuando se calcula la distancia de Hamming de dos modelos, una plantilla se desplaza a nivel de bit a izquierda y derecha, calculándose una serie de valores de la distancia de Hamming a partir de sucesivos cambios [10]. Este método corrige desajustes rotacionales en el modelo normalizado del iris causados por diferencias en rotación de imágenes. De entre los valores de distancia calculada se adopta el menor valor calculado.

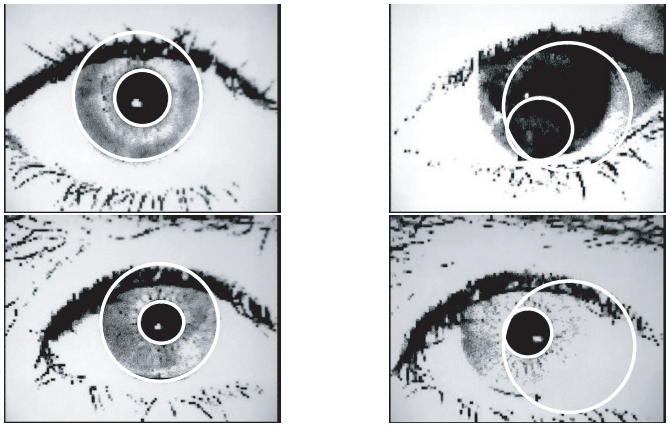
3.2. Protocolo Experimental

Para los experimentos, cada ojo de la base de datos se considera un usuario diferente. De esta forma, tenemos dos sesiones con 4 imágenes cada una de los 100 usuarios (50 participantes $\times$ 2 ojos por participante).

En los experimentos se consideran dos diferentes escenarios de ataque y se comparan al modo de funcionamiento normal:

- **Modo normal de funcionamiento (NOM)**: en este modo, el registro y las pruebas se llevan a cabo con iris reales. Este será el escenario de referencia. En este contexto, la Tasa de Falsa Aceptación (FAR) del sistema se define como el número de veces que un impostor, usando su propio iris, consigue acceder al sistema como un usuario original, por lo que se puede interpretar como la robustez del sistema frente a un ataque sin esfuerzo. Del mismo modo, la Tasa de Falso Rechazo (FRR) denota el número de veces que un usuario original es rechazado por el sistema.
- **Ataque 1**: aquí, tanto el registro como las pruebas se llevan a cabo con iris falsos. En este caso, el atacante se registra en el sistema con un iris falso correspondiente a un usuario genuino y luego intenta entrar a la aplicación usando también un iris falso del mismo usuario. En este escenario, un ataque no exitoso (es decir, el sistema rechaza al atacante) será cuando el impostor no sea capaz de acceder al sistema usando el iris falso. Por tanto, la tasa de éxito del ataque (SR) de este esceracio se puede calcular como: SR = 1−FRR.
- **Ataque 2**: el registro se lleva a cabo con un iris real y las pruebas se realizan con un iris falso. En este caso el usuario genuino se registra con su iris y el atacante intenta acceder a la aplicación con el iris falso correspondiente al usuario legal. Un ataque exitoso tendrá lugar si el sistema confunde un iris falso con su correspondiente genuino, es decir: SR = FAR.

Para calcular el rendimiento del sistema en un modo normal de funcionamiento, el protocolo experimental ha sido el siguiente. Para un usuario dado, todas las imágenes de la primera sesión se consideran como muestras de registro. Las comparaciones genuinas se obtienen comparando las muestras de registro con las imágenes correspondientes de la segunda sesión del mismo usuario. Las comparaciones de impostores se obtienen comparando una imagen de registro con una muestra aleatoria de la segunda sesión de cada uno de los usuarios restantes. De forma similar, para calcular el FRR en el ataque 1, todas las imágenes falsas de la primera sesión de cada usuario son comparadas con las correspondientes imágenes falsas de la segunda sesión. En el segundo ataque, solo se calculan los resultados de los impostores, comparando las 4 muestras originales de registro de cada usuario con sus 4 muestras falsas de la segunda sesión. En nuestros experimentos, no todas las imágenes fueron segmentadas correctamente por el sistema de reconocimiento Por ello no fue posible usar todas las imágenes de ojos para los experimentos de prueba.



(a) Detección correcta de iris (b) Detección incorrecta de iris

Figura 4. Ejemplos de las imágenes falsas correctamente segmentadas (izquierda) y con una detección incorrecta de irir (derecha).

3.3. Resultados

En la Figura 4, se muestran varios ejemplos de imágenes falsas con detección correcta e incorrecta de iris. El número de imágenes correctamente segmentadas para la base de datos original es de 792 (99 % de las 800 disponibles) y 574 para la base de datos falsa (71.75 % de las 800). Es importante resaltar que más del 70 % de las imágenes falsas pasan con éxito las fases de segmentación y normalización. Gracias a las modificaciones incluidas en el paso de segmentación (ver Sección 3.1), hemos mejorado el porcentaje de segmentación del sistema original, el cual se situaba en anteriores experimentos en un 80.56 % y un 38.43 % para la base de datos original y falsa respectivamente. Es importante destacar también que al intentar mejorar el porcentaje de correcta segmentación de las imágenes reales, estamos también mejorando el de las imágenes falsas. En la Tabla 2 se muestra el Porcentaje de éxito (SR) de los ataques directos al sistema de reconocimiento en cuatro puntos de funcionamiento distintos, considerando únicamente la comparación entre imágenes correctamente segmentadas. El umbral de decisión está fijado para alcanzar un FAR=$\{0.1, 1, 2, 5\}$ % en el modo normal de funcionamiento, y después se calcula el porcentaje de éxito de los dos ataques propuestos. Como podemos observar, para cualquier punto de funcionamiento el sistema es vulnerable para ambos ataques (de hecho se observa una tasa de éxito mayor ó igual al 35 %). Esto se hace especialmente evidente según aumentamos la FAR del modo normal de funcionamiento, consiguiendo un éxito de ataque de más de la mitad de las pruebas. También es importante resaltar que el porcentaje de éxito del ataque 1 es similar al del ataque 2. En el ataque 1, un intruso sería registrado correctamente en el sistema usando un iris falso de otra persona y posteriormente se le permitiría el acceso al sistema usando dicha imagen falsa.

NOM	Ataque 1	Ataque 2
FAR - FRR (%)	SR (%)	SR (%)
0.1 - 16.84	33.57	36.89
1 - 12.37	48.02	52.44
2 - 10.78	53.03	56.96
5 - 8.87	61.19	64.56

Cuadro 2. Evaluación del sistema de verificación frente ataques directos. NOM se refiere al modo normal de funcionamiento del sistema y SR al porcentaje de éxito del ataque.

4. Conclusiones

Se ha presentado un estudio de las vulnerabilidades de un sistema basado en reconocimiento de iris. Los ataques se han realizado usando imágenes falsas creadas a partir de imágenes reales de la base de datos de referencia BioSec. Después de imprimir las imágenes con una impresora comercial, estas fueron presentadas al sensor de iris. Se han estudiado diferentes factores que afectan a la calidad de las imágenes falsas adquiridas, incluyendo el pre-procesado de las imágenes originales, el tipo de impresora y el tipo de papel. Hemos elegido la combinación que nos da la mejor calidad y después hemos construido una base de datos falsa con las imágenes de 100 ojos, con 8 imágenes de iris por ojo. La adquisición de las imágenes falsas se ha llevado a cabo con las misma cámara usada en BioSec.

Se han comparado dos escenarios de ataque distintos con el modo normal de funcionamiento del sistema, usando un sistema de reconocimiento de disponibilidad pública. El primer ataque consiste en registrarse y acceder al sistema con un iris falso. El segundo simula el registro con un iris original y el acceso con un iris falso. Los resultados mostraron que el sistema es vulnerable para ambos ataques. También se ha observado que alrededor del 72 % de las imágenes falsas fueron segmentadas correctamente por el sistema. Cuando esto ocurre, al intruso se le garantiza la entrada con una probabilidad bastante alta, alcanzando un porcentaje de éxito en ambos ataques del 50 % o más.

Una posible contramedida para prevenir estos ataques es usar procedimientos de detección de vida. Para el caso de sistemas de reconocimiento de iris, se propone la detección de reflexiones de luz, la detección de movimiento del iris, o la respuesta del iris a modificaciones repentinas de luz [11,12]. Estas líneas de investigación serán seguidas en nuestros trabajos futuros. También se explorará el uso de otro tipo de sensores, como el de soporte manual de OKI usado en la base de datos de Casia[2].

Agradecimientos. Este trabajo ha sido financiado por el proyecto TEC2006-13141-C03-03 del Ministerio de Educación y Ciencia y por la Red de Excelencia Europea

[2] http://www.cbsr.ia.ac.cn/databases.htm

BioSecure IST-2002-507634. El autor F. A.-F. agradece a la Consejeria de Educacion de la Comunidad de Madrid y al Fondo Social Europeo por financiar sus estudios de Doctorado. El autor J. G. está siendo fianciado por una beca FPU del Ministerio de Educación y Ciencia. El autor J. F. está siendo fianciado por una Marie Curie Fellowship de la Comisión Europea.

Referencias

1. Jain, A., Ross, A., Pankanti, S.: Biometrics: A tool for information security. IEEE Trans. on Information Forensics and Security **1** (2006) 125–143
2. Jain, A., Bolle, R., Pankanti, S., eds.: Biometrics - Personal Identification in Networked Society. Kluwer Academic Publishers (1999)
3. Monro, D., Rakshit, S., Zhang, D.: DCT-Based iris recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence **29**(4) (April 2007) 586–595
4. Schneier, B.: The uses and abuses of biometrics. Communications of the ACM **48** (1999) 136
5. Ratha, N., Connell, J., Bolle, R.: An analysis of minutiae matching strength. Proc. International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA **Springer LNCS-2091** (2001) 223–228
6. Soutar, C., Gilroy, R., Stoianov, A.: Biometric system performance and security. Proc IEEE Workshop on Automatic Identification Advanced Technologies, AIAT (1999)
7. Fierrez, J., Ortega-Garcia, J., Torre-Toledano, D., Gonzalez-Rodriguez, J.: BioSec baseline corpus: A multimodal biometric database. Pattern Recognition **40**(4) (April 2007) 1389–1392
8. Gonzalez, R., Woods, R.: Digital Image Processing. Addison-Wesley (2002)
9. Masek, L., Kovesi, P.: Matlab source code for a biometric identification system based on iris patterns. The School of Computer Science and Software Engineering, The University of Western Australia (2003)
10. Daugman, J.: How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology **14** (2004) 21–30
11. Daugman, J.: Anti spoofing liveness detection. available on line at http://www.cl.cam.ac.uk/users/jgd1000/countermeasures.pdf
12. Pacut, A., Czajka, A.: Aliveness detection for iris biometrics. Proc. IEEE Intl. Carnahan Conf. on Security Technology, ICCST (2006) 122–129