

Fusión de Sistemas de Reconocimiento Biométrico de Pisadas y Cara en Escenarios No Controlados y Sin Supervisión

Rubén Vera-Rodríguez, Pedro Tomé, Julián Fierrez y Javier Ortega-García

Biometric Recognition Group - ATVS, Universidad Autónoma de Madrid,
Avda. Francisco Tomás y Valiente, 11 - 28049 Madrid, Spain
{ruben.vera, pedro.tome, julian.fierrez, javier.ortega}@uam.es

Abstract. Este artículo presenta un análisis experimental de la fusión de sistemas de reconocimiento biométrico de pisadas y cara para la autenticación de personas en escenarios no controlados y sin supervisión. El reconocimiento biométrico usando pisadas es una técnica relativamente nueva que se basa en la captura de señales de personas andando sobre sensores de presión. La idea de la fusión entre señales de pisadas e imágenes de la cara surge de la premisa de que en un área donde hay instalados sensores de captura de pisadas es muy sencillo posicionar una cámara que capture la cara de la persona que anda sobre los sensores. Este sistema puede encontrar aplicación en escenarios de edificios inteligentes, asistencia en ambientes autónomos (AAL), cuidado de personas mayores, o accesos controlados. Este artículo presenta un análisis comparativo de ambas técnicas biométricas usando la misma base de datos y protocolos experimentales. En el trabajo experimental se consideran dos tipos de aplicaciones: casas inteligentes (grupo pequeño de usuarios con muchos datos de entrenamiento) y accesos controlados (grupo mayor de usuarios con pocos datos de entrenamiento), obteniendo resultados del 0.9% y 5.8% EER respectivamente para la fusión de ambas modalidades. Esto supone una mejora significativa del rendimiento en comparación con los resultados obtenidos por los sistemas individuales.

1 Introduction

Los sistemas biométricos no intrusivos están recibiendo una gran atención de la comunidad investigadora recientemente dado el alto grado de aceptabilidad por parte de los usuarios en distintas aplicaciones. Las señales que se pueden capturar de las pisadas y de la cara son buenos ejemplos de modos biométricos no intrusivos que podrían ser fusionados. El reconocimiento biométrico usando pisadas es una técnica relativamente nueva para discriminar entre personas usando características de la forma de andar extraídas de sensores instalados en el suelo. Uno de los beneficios de las señales de pisadas son la robustez de las señales frente a distintas condiciones ambientales, con mínimas fuentes externas de ruido que puedan corromper las señales [1]. Por otro lado, los sistemas de cara tienen mejores tasas de reconocimiento comparadas con las pisadas, pero las señales de

cara están más afectadas por factores externos como la iluminación, la pose o la distancia entre el sujeto y la cámara entre otros [2].

Este artículo se centra en la fusión entre señales de pisadas y cara en ambientes no controlados y sin supervisión. Trabajos previos relacionados han llevado a cabo la fusión entre señales de cara y de videos de personas andando [3,4], obteniendo muy buenos resultados debido principalmente a la poca correlación entre ambas modalidades biométricas. Este es un caso similar al presentado aquí, aunque las señales de pisadas son más controladas que los videos de las personas andando, pero son más robustas. En nuestro caso, las señales de pisadas y cara son muy fáciles de capturar conjuntamente instalando una cámara que captura la cara de la persona que camina sobre los sensores que capturan las pisadas. Este es un buen ejemplo de un sistema biométrico multimodal no obstrusivo y transparente a los usuarios, ya que la persona camina libremente sobre un área sin tener que interactuar con ningún dispositivo.

La base de datos considerada en este artículo (para más detalles ver [5]) fue capturada de forma no controlada y sin supervisión en su mayoría, es decir, los factores que producen variabilidad en cada modo biométrico como son la iluminación, pose, etc., para la cara, y el tipo de calzado o la velocidad al andar no fueron controlados, lo cual hace este problema más complejo pero a la vez los resultados obtenidos son más realistas en términos de la amplitud de condiciones consideradas.

La fusión de las pisadas y la cara se lleva a cabo a nivel de las puntuaciones (score-level). Estas puntuaciones han sido normalizadas de varias formas para hacerlas comparables en ambos sistemas. Dos arquitecturas de fusión han sido consideradas, una correspondería con el caso ideal de tener el mismo número de señales de pisadas que de imágenes de la cara, y otra se correspondería con un caso más realista donde se lleva a cabo una fusión adaptativa en el caso de no tener el mismo número de señales de ambas modalidades a fusionar. Por otro lado, se presentan resultados experimentales simulando dos posibles escenarios de aplicación para este sistema multibiométrico: el caso de casas inteligentes, donde se tendría un grupo pequeño de usuarios pero con muchos datos de entrenamiento, y el caso de accesos de seguridad, donde habría un grupo grande de usuarios pero con pocos datos disponibles de entrenamiento. Los mejores resultados obtenidos son del 0.9% y 5.8% EER para los casos de la fusión de pisadas y cara para cada aplicación respectivamente.

Este artículo está organizado del siguiente modo. La Sección 2 describe las señales de pisadas consideradas y el sistema de reconocimiento biométrico usado para este modo. La Sección 3 es similar a la anterior pero para el caso de la cara. La Sección 4 describe el protocolo experimental seguido para llevar a cabo la fusión. La Sección 5 presenta los resultados experimentales obtenidos, y finalmente las conclusiones se esbozan en la Sección 6.

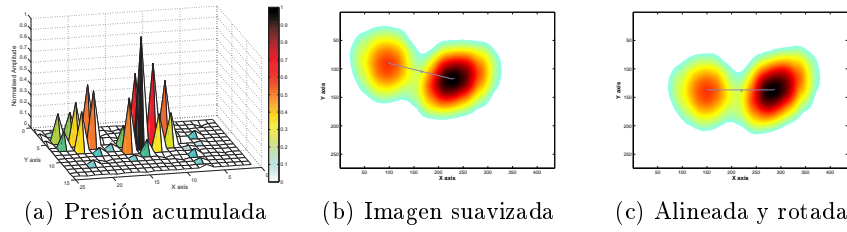


Fig. 1. (a) Ejemplo de la señal de una pisada con la presión acumulada en los ejes X e Y. (b) Igual que (a) después de aplicar un filtro Gaussiano. (c) Imagen de la pisada después de rotar y alinear a un centro común.

2 Sistema Reconocimiento de Pisadas

Esta sección describe las características de las señales de pisadas así como el sistema de reconocimiento desarrollado. La principal característica de las señales de pisadas consideradas aquí es que contienen información tanto en el dominio del tiempo como en el del espacio, a diferencia de trabajos previos [6,7,8]. En este caso se usa una matriz con alta densidad de sensores piezoeléctricos (650 sensores por m^2) distribuidos homogéneamente, los cuales capturan la presión transitoria. La tasa de muestreo es de 1.6 kHz. El área donde los sensores están instalados es lo suficientemente grande como para capturar la señal procedente de una zancada (pisada de pie derecho a pie izquierdo).

En este artículo, las características extraídas de las señales de pisadas para llevar a cabo el reconocimiento de personas están basadas en la presión acumulada por cada sensor piezoeléctrico durante una pisada, al igual que en [9]. La Figura 1(a) muestra un ejemplo de una señal de pisada con la presión acumulada por cada sensor en los ejes X e Y. Sobre este tipo de imágenes se lleva a cabo alineamiento y rotación para poder situar la pisada en posiciones fijas, pero antes de esto, las imágenes se suavizan usando un filtro Gaussiano para obtener una imagen continua. La Figura 1(b) muestra la imagen resultado para el ejemplo dado después de usar el filtro Gaussiano, vista desde arriba.

Para realizar el alineamiento, en primer lugar se detectan los dos picos de la señal, que corresponden con el talón y la parte frontal, es decir, las posiciones de mayor presión. Después se calcula el centro de esas posiciones y se alinea y rota como se observa en la Figura 1(c). De este modo todas las señales de pisadas se encuentran centradas y alineadas en rotación y se puede llevar a cabo el reconocimiento biométrico. Para reducir la dimensionalidad de las imágenes se aplica PCA (análisis de componentes principales) antes de ser introducidas en el clasificador, reteniendo un 96% de la información original usando las primeras 200 componentes. En cuanto al clasificador, SVM (máquina de vectores soporte) es usado en este caso con un kernel RBF (radial basis function), dados los buenos resultados obtenidos en estudios previos en este área [8,10].

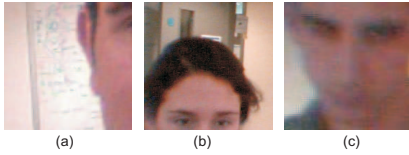


Fig. 2. Ejemplos de imágenes de la cara eliminadas.



Fig. 3. Ejemplos de imágenes de la cara consideradas en los experimentos. Cabe destacar que es un escenario no controlado.

3 Sistema de Reconocimiento de Cara

Esta sección describe las imágenes de cara consideradas y el sistema de reconocimiento biométrico desarrollado. Las imágenes de la cara se obtienen usando una cámara de video comercial de baja calidad, con una frecuencia de muestreo de 30 imágenes por segundo a una resolución de 640×480 píxeles. Cada señal de zancada tiene enlazado un video de la cara de la persona que camina hacia la cámara.

En la fase de preprocesado, VeriLook SDK v2.0 (sistema comercial) es aplicado sobre todas las imágenes capturadas para ubicar y segmentar la cara en la imagen. Las imágenes de cara resultantes son normalizadas a 64×80 píxeles (ancho \times alto). VeriLook proporciona un índice de calidad a cada imagen de la cara en el rango de 0 a 100. De este modo, la imagen con mayor índice de calidad de cada secuencia de imágenes es seleccionada para llevar a cabo el reconocimiento de las personas. Además, para descartar imágenes de baja calidad se fijó un umbral al valor de 55 basado en este índice de calidad.

La Figura 2 muestra algunos ejemplos de imágenes de la cara descartadas debido, por ejemplo, a que sólo contenían parte de la cara (a), (b) o son imágenes borrosas (c). Es importante notar que estas son las imágenes con mayor índice de calidad dentro de su secuencia. La Figura 3 muestra algunos ejemplos de imágenes de cara consideradas en los experimentos. La Figura 3(a) muestra un ejemplo de una cara de buena calidad, (b) está un poco borrosa, y (c) y (d) muestran dos imágenes de la misma persona con diferentes condiciones de iluminación. La Figura 3 muestra la dificultad del problema considerado aquí, con imágenes obtenidas en escenarios no controlados y sin supervisión. Los dos únicos parámetros controlados fueron la edad de los usuarios (ya que la base de datos se obtuvo en un periodo de 16 meses) y la cámara, que fue la misma durante toda la captura. De las señales seleccionadas, la pose es compensada respecto

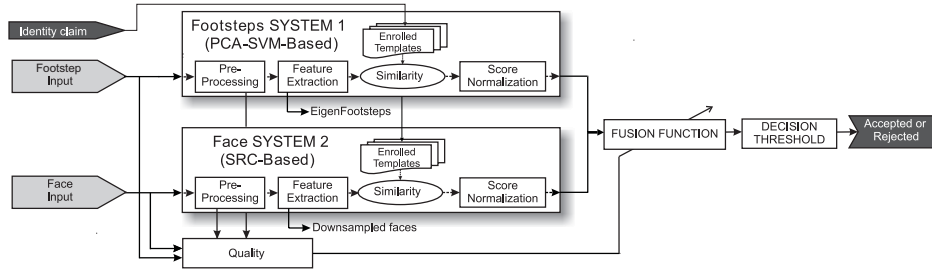


Fig. 4. Arquitectura general para la fusión de los sistemas de pisadas y cara. La arquitectura de la fusión depende de los casos ideal o realista definidos.

a la posición de los ojos, y la iluminación con compensación fotométrica para obtener mejores resultados en el reconocimiento biométrico.

En cuanto al sistema de clasificación empleado, es un sistema de verificación en el estado del arte basado en “sparse representation classification” (SRC) [11]. Las características extraídas son submuestras de las imágenes con un ratio de 1/4, obteniendo 320 dimensiones de las 5120 originales, lo cual forma el vector de características.

4 Protocolo Experimental

4.1 Fusión

En esta sección se describe la fusión de las modalidades de pisadas y cara. En general, la fusión biométrica multimodal puede ser llevada a cabo en diferentes niveles. En nuestro caso la fusión se lleva a cabo en el nivel de las puntuaciones, que es el método más común debido a la facilidad de la combinación de las puntuaciones generadas por distintos clasificadores [12,13]. La Figura 4 muestra un diagrama general de la fusión de los dos sistemas. Se han considerado dos arquitecturas diferentes de fusión:

1. Caso ideal. En este caso, siempre hay una imagen de la cara que se fusiona con una señal de la zancada (pisada de derecha a izquierda). Para la fusión se hace la suma de las puntuaciones de las dos modalidades. Como se muestra en la Figura 4. En la Sección 5.1 se muestran los resultados obtenidos para este tipo de fusión.

2. Fusión adaptativa (caso realista). En este caso, debido a la eliminación de algunas imágenes de la cara de mala calidad, existen más señales de pisadas que imágenes de la cara para realizar la fusión. Por lo tanto, la fusión adaptativa se lleva a cabo haciendo la suma de las puntuaciones de las pisadas y la cara (como en el caso ideal) cuando hay señales para las dos modalidades, y en el caso de que una señal de la pisada no tenga una imagen de la cara correspondiente se da todo el peso a la señal de la pisada. La Sección 5.2 muestra los resultados para este caso.

		Security Access (SA)			Smart Home (SH)		
		Train	Test		Train	Test	
Subjects		P1-P54	P1-P54	P55-P78	P1-P15	P1-P15	P16-P78
Foot	#Signals	1080	7725	250	3000	3113	630
Face	#Signals	809	5932	179	2148	2561	484
	Discarded	25,1%	23,2%	25,4%	28,4%	17,7%	23,2%

Table 1. Configuración de la base de datos para las dos aplicaciones consideradas: acceso de seguridad (SA) y casas inteligentes (SH).

Antes de llevar a cabo la fusión de las dos modalidades biométricas, las puntuaciones son normalizadas para poder hacer una transformación de las mismas para hacerlas comparables, es decir, para ponerlas en un dominio común. Siguiendo los resultados de Jain *et al.* [13], las reglas de normalización de las puntuaciones que se han seguido son: max-min, Z-norm y tanh-norm, las cuales son comparadas en la Sección 5. En los casos de Z-norm y tanh-norm, antes de aplicar la normalización las puntuaciones son transformadas usando una función logit [14] dada por:

$$s' = \text{logit}(s) = \log \frac{s}{1-s} \quad (1)$$

donde s son las puntuaciones y s' denota la puntuación transformada con la función “logit”.

4.2 Escenario de Aplicación

En cuanto al protocolo experimental, se han diseñado dos escenarios de aplicación distintos: casas inteligentes (*smart homes*) y accesos seguros (*security access*). Una de las características de la base de datos considerada aquí es que contiene una gran cantidad de datos para un grupo pequeño de usuarios (>200 señales por usuario, para 15 usuarios), lo cual podría servir para simular un escenario de casa inteligente; y contiene una cantidad pequeña de datos para un grupo más grande de usuarios (>20 señales por usuario, para 54 usuarios), lo cual podría servir para simular un escenario de acceso de seguridad. Esto refleja el modo de captura, que fue voluntario y sin supervisión.

La Tabla 1 muestra las divisiones de la base de datos en sets de entrenamiento y test para cada aplicación considerada. El número de señales de pisadas y de imágenes de la cara son distintos debido a la eliminación de algunas imágenes de cara, como ya se ha descrito. En particular, el 24,5% de imágenes de la cara son eliminadas en el caso de la aplicación de acceso de seguridad (SA), y un 23,1% son eliminadas en el caso de la aplicación de casas inteligentes (SH).

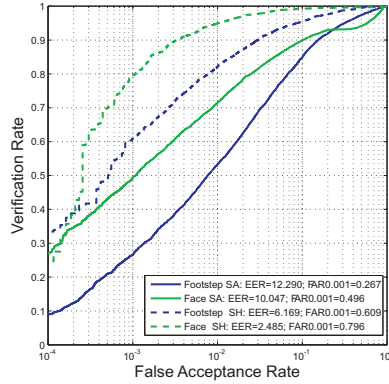


Fig. 5. Curvas ROC para los sistemas de pisadas y cara para las aplicaciones de acceso de seguridad (SA) y casas inteligentes (SH).

5 Resultados Experimentales

Esta sección describe los resultados experimentales de la fusión de los modos de pisadas y cara para un escenario no controlado y sin supervisión. Los resultados se muestran mediante el uso de curvas ROC con tasas de verificación en EERs (*equal error rates*) y tasas de verificación (VR) trabajando con FAR (tasa de falsa aceptación) de 0,001.

5.1 Caso Ideal

Esta sección presenta los resultados obtenidos para el caso de la fusión ideal (descrita en la Sección 4.1). La Figura 5 muestra las curvas ROC para los sistemas de reconocimiento de pisadas y cara de forma independiente. En este caso hay el mismo número de señales de pisadas y de imágenes de cara, que corresponde con el número de imágenes de cara mostrado en la Tabla 1. Las curvas ROC de las dos aplicaciones consideradas muestran resultados significativamente mejores para el caso de casas inteligentes (SH) comparado con el acceso de seguridad (SA), debido principalmente a la mayor cantidad de datos de entrenamiento usados por persona y a que el grupo de usuarios considerados es menor. Además, los resultados obtenidos para la cara son mejores que los obtenidos para el caso de las pisadas.

El rendimiento del sistema biométrico multimodal ha sido estudiado bajo varias técnicas de normalización de las puntuaciones, que son las siguientes: max-min, transformación logit con Z-norm, tanh-norm, y transformación logit con tanh-norm. Las curvas ROC para las aplicaciones SA y SH se muestran en la Figura 6. Como se puede observar, entre las distintas técnicas de normalización la tanh-norm con y sin transformación logit son las que mejores resultados obtienen tanto para valores bajos como altos de FAR. En particular, los mejores

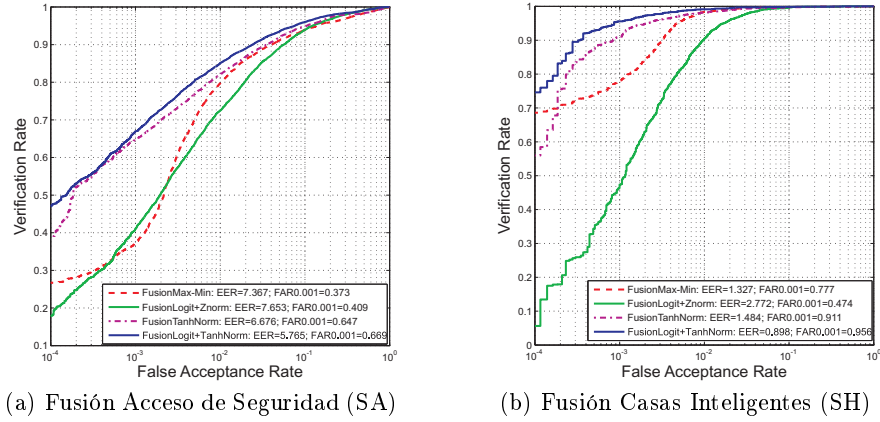


Fig. 6. Curvas ROC para la fusión de los sistemas de pisadas y cara para el caso ideal.

resultados obtenidos para el caso de acceso de seguridad (SA) don del 66.9% de tasa de verificación (VR) a una FAR=0.001 y 5.7% EER. Para el caso de casas inteligentes (SH) los mejores resultados obtenidos son del 95.6% de VR y 0.9% EER, en ambos casos usando la transformación logit con la normalización tanh-norm, que es la técnica usada también en el caso de la fusión realista (Sección 5.2). La Tabla 2 muestra los resultados obtenidos para los distintos casos.

5.2 Caso Realista

Esta sección muestra los resultados para el caso de la fusión realista, la cual es una fusión adaptativa tal y como se describe en la Sección 4.1. La Figura 7 muestra las curvas ROC para los casos de las dos aplicaciones comparando el rendimiento del sistema de pisadas, la fusión en el caso ideal y el realista. Como se puede observar, el rendimiento para la fusión realista no es tan bueno como para el caso ideal, pero muestra una mejora significativa respecto al caso de sólo usar las pisadas. Además, como se ha comentado, cabe destacar que alrededor de un 24% de los datos disponibles no han sido considerados en el caso ideal, cosa que no sería demasiado sensata en un escenario real. En este caso los resultados obtenidos son del 7.7% EER y 41.5% VR para la aplicación SA y del 2.3% EER y 80.7% VR para la aplicación SH.

6 Conclusiones

Este artículo presenta por primera vez resultados experimentales para la fusión de las modalidades biométricas de pisadas y cara en un entorno no controlado y sin supervisión. Las señales de las pisadas tienen el beneficio sobre otros modos biométricos de poder ser capturadas de forma encubierta, lo cual es muy conveniente para los usuarios, además este tipo de señales son muy difíciles de

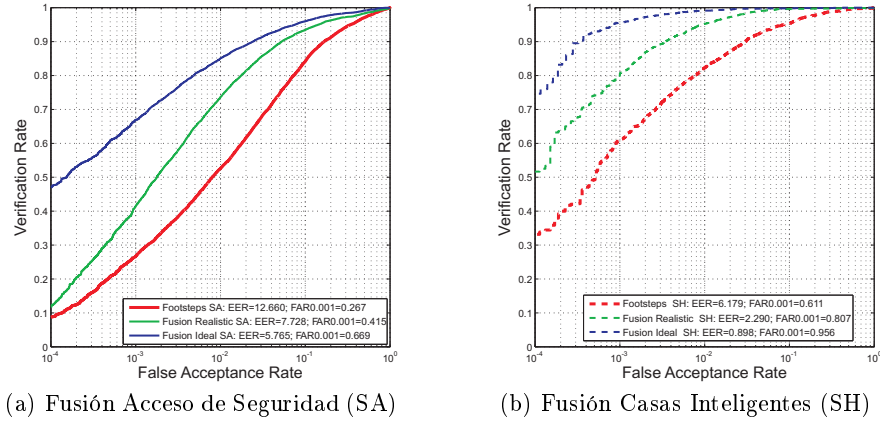


Fig. 7. Curvas ROC para la fusión de los sistemas de pisadas y cara para el caso realista comparado con el caso ideal y con el caso del sistema de pisadas de forma individual.

	Security Access (SA)		Smart Home (SH)	
	VR	EER	VR	EER
Footstep	26.7	12.7	61.1	6.2
Face	49.6	10.0	79.6	2.5
Fusion Ideal	66.9	5.8	95.6	0.9
Fusion Realistic	41.5	7.7	80.7	2.3

Table 2. Resultados en tasa de verificación (VR) para FAR=0.001 y EER.

ser imitadas. La fusión de señales de pisadas con otros modos biométricos no intrusivos como la cara o la forma de andar (*gait*) puede ser llevada a cabo fácilmente. En el protocolo experimental se han considerado dos aplicaciones: acceso de seguridad y casas inteligentes. La fusión se lleva a cabo en el nivel de las puntuaciones y siguiendo dos arquitecturas: (i) un caso ideal donde se obtienen resultados del 5.8% y 0.9% EER para cada aplicación respectivamente; y (ii) un caso realista donde no siempre hay una imagen de la cara enlazado a cada señal de la pisada. En este caso se lleva a cabo una fusión adaptativa obteniendo resultados del 7.7% y 2.3% EER para cada aplicación respectivamente, lo cual supone un 39.4% y 62.9% de relativa mejora del EER comparado con el caso de las pisadas de forma individual.

Agradecimientos

Este trabajo ha sido financiado parcialmente por los proyectos Contexts (S2009/TIC-1485), Bio-Challenge (TEC2009-11186) y la Cátedra UAM-Telefónica. R.V.R es recipiente de un contrato Juan de la Cierva del MICINN.

References

1. Vera-Rodríguez, R., Lewis, R., Mason, J., Evans, N.: Footstep recognition for a smart home environment. *International Journal of Smart Home. Special Issue on Future Generation Smart Space (FGSS)* **2** (2008) 95–110
2. Tome, P., Fierrez, J., Alonso-Fernandez, F., Ortega-Garcia, J.: Scenario-based score fusion for face recognition at a distance. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. (2010) 67 – 73
3. Kale, A., Roychowdhury, A., Chellappa, R.: Fusion of gait and face for human identification. In: *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)*. *IEEE International Conference on*. Volume 5. (2004) V – 901–4 vol.5
4. Geng, X., Smith-Miles, K., Wang, L., Li, M., Wu, Q.: Context-aware fusion: A case study on fusion of gait and face for human identification in video. *Pattern Recognition* **43** (2010) 3660 – 3673
5. Vera-Rodríguez, R., Mason, J., Evans, N.: Automatic cross-biometric footstep database labelling using speaker recognition. In: *Proceedings of the IAPR/IEEE International Conference on Biometrics (ICB)*. (2009)
6. Yun, J.S., Lee, S.H., Woo, W.T., Ryu, J.H.: The User Identification System Using Walking Pattern over the ubiFloor. In: *Proceedings of International Conference on Control, Automation, and Systems*. (2003) 1046–1050
7. Middleton, L., Buss, A.A., Bazin, A.I., Nixon, M.S.: A floor sensor system for gait recognition. In: *Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*. (2005) 171–176
8. Suutala, J., Roning, J.: Methods for person identification on a pressure-sensitive floor: Experiments with multiple classifiers and reject option. *Information Fusion. Special Issue on Applications of Ensemble Methods* **9** (2008) 21 – 40
9. Vera-Rodríguez, R., Mason, J.S., Fierrez, J., Ortega-Garcia, J.: Analysis of spatial domain information for footstep recognition. *IET Computer Vision, Special Issue on Future Trends in Biometrics Processing* (2011)
10. Vera-Rodríguez, R., Mason, J., Fierrez, J., Ortega-Garcia, J.: Analysis of time domain information for footstep recognition. In: *Proc. 6th International Symposium on Visual Computing (ISVC'2010)*. *Lecture Notes in Computer Science*, Springer (2010)
11. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31** (2009) 210–227
12. Alonso-Fernandez, F., Fierrez, J., Ramos, D., Gonzalez-Rodriguez, J.: Quality-based conditional processing in multi-biometrics: application to sensor interoperability. *IEEE Trans. Sys. Man Cyber. Part A* **40** (2010) 1168–1179
13. Jain, A., Karthik, N., Ross, A.: Score normalization in multimodal biometric systems. *Pattern recognition* **38** (2005) 2270–2285
14. Poh, N., Kittler, J., Marcel, S., Matrouf, D., Bonastre, J.F.: Model and score adaptation for biometric systems: Coping with device interoperability and changing acquisition conditions. In: *Pattern Recognition (ICPR), 2010 20th International Conference on*. (2010) 1229 –1232