

# Robust Multi-modal and Multi-unit Feature Level Fusion of Face and Iris Biometrics

Ajita Rattani and Massimo Tistarelli

Computer Vision Laboratory,  
University of Sassari,  
Alghero, Italy  
ajita.rattani@diee.unica.it, tista@uniss.it

**Abstract.** Multi-biometrics has recently emerged as a mean of more robust and efficient personal verification and identification. Exploiting information from multiple sources at various levels i.e., feature, score, rank or decision, the false acceptance and rejection rates can be considerably reduced. Among all, feature level fusion is relatively an understudied problem. This paper addresses the feature level fusion of multi-modal and multi-unit sources of information. For multi-modal fusion the face and iris biometric traits are considered, while the multi-unit fusion is applied to merge the data from the left and right iris images. The proposed approach computes the *SIFT* features from both biometric sources, either multi-modal or multi-unit. For each source, feature selection on the extracted *SIFT* features is performed via spatial sampling. Then these selected features are finally concatenated together into a single feature *super-vector* using serial fusion. This concatenated super feature vector is used to perform classification.

Experimental results from face and iris standard biometric databases are presented. The reported results clearly show the performance improvements in classification obtained by applying feature level fusion for both multi-modal and multi-unit biometrics in comparison to uni-modal classification and score level fusion.

## 1 Introduction

Biometrics refers to the use of physiological, biological or behavioural characteristics to establish the identity of an individual. These characteristics are unique to each individual and remain partially un-altered during the individual's lifetime [1].

In the recent years biometric authentication has gained a considerable improvement in both reliability and accuracy. Nonetheless, the best biometric systems to date present several drawbacks, such as limited applicability, vulnerability to spoofing attacks, less discriminant features, performance degradation due to noisy data, and others. Some of these limitations are inherent in the sensor technology or in the nature of the biometric trait itself. Consequently, a mono-modal biometric system rarely can cope with the variety of requirements in real applications. This is especially true in non-ideal scenarios, like outdoor environments, or highly demanding applications, such as in large-scale systems.

By grouping evidence from multiple sources of information, multi-biometric systems [2] allow to overcome some of the drawbacks of the mono-modal biometric systems. The combination of multiple traits provide a better population coverage and provide a platform with wider usability. Multi-biometrics also offer an efficient counter-measure to spoofing, because it would be difficult for an impostor to simultaneously spoof multiple biometric traits of a genuine user [2].

Multi-biometric systems can rely on multi-modalities like face and fingerprint, multiple units like two or more fingerprints, multiple sensors like optical and capacitive sensors, multiple representations like LDA and PCA features for face [2]. The resulting multiple sources of information can be integrated at various levels. Ross and Jain [3] presented a wide overview of multi-modal biometric solutions describing different levels at which fusion can be performed, i.e. sensor level, feature extraction level, matching score level and decision level.

As reported in the literature [3] a biometric system that integrates information at an earlier stage of processing is expected to provide better performances than systems that integrate information at a later stage, because of the availability of more and richer information. Therefore, fusion at sensor and feature level is expected to be more efficient than fusion at matching score, rank and decision levels. Despite of the abundance of research papers related to multi-biometrics [4–6], fusion at feature level is a relatively understudied problem [7-9]. One possible reason is that fusion at feature level is relatively difficult to achieve in practice because different modalities may have incompatible feature sets and the correspondence among different feature spaces may be unknown. Moreover, fusing feature vectors may lead to the problem of the *curse of dimensionality*: due to the large dimension of the fused feature vector, a very complex matcher may be required to correctly classify the fused data. Nevertheless, the fused feature vector may contain noisy or redundant data thus leading to degradation in the performance of the classifier [7].

Among many biometric traits, face and iris possess a number of useful characteristics for recognition. For example, face sensing is non-invasive and friendly. Moreover, it is relatively easy to acquire face images of good quality. On the other hand, iris-based identification systems has proved to be among the best performing biometrics in many application domains. Nonetheless, both face and iris have their own limitations too. The accuracy of face recognition is often affected by illumination, pose, shelter and facial expression, while current iris recognition systems require the active cooperation of the user. Thus in a sense the two modalities are highly complementary. As face data is easy to acquire but delivers low to medium performances, iris data is more difficult to acquire but delivers very high performances. However, as they both belong to the same biological trait (the head) they can be easily sampled at the same time and with the same sensing technology. Therefore, it is expected that fusion of face and iris may overcome the intrinsic limitations of each modality, while delivering the following advantages [10-11]:

- 1) The total error rate (the combination of the false accept rate and the false reject rate) is decreased with respect to the best of the two modalities [11].
- 2) The probability of successful spoofing attacks is reduced, because of the increased difficulty in manufacturing (or reproducing in a simulated video) an artificial head with both the face and irises faithfully reproducing the genuine user's appearance. On the contrary, it also makes more difficult for an individual, in a surveillance scenario, to hide his/her identity.

- 3) The population coverage is larger in comparison to uni-modal biometric system. The adoption of the face modality can allow people with imperfect iris images to enroll, reducing the enrolment failure rate.
- 4) Conversely to other multi-biometric systems, which require multiple sensors, the combination of face and iris allows for simultaneous acquisition of all the data with a single device. Therefore, multi-modal and multi-unit fusion can be performed without introducing any additional cost for the hardware.

Few work related to face and iris fusion have been reported in the literature. The combination of iris and face at score level have been reported in [10][11]. References [12] [13] are the only one in literature, in our knowledge, describing a methodology for feature level fusion of face and iris. Both of the referred work [12-13] have applied projection to lower dimensional space for feature fusion and reduction, using prior training of the classifier and have used nearest neighbour distance for the classification of fused feature vector.

This paper proposes a new approach for feature level fusion of face and both left and right iris for an individual. In the proposed system, SIFT features are extracted from the face image and both irises. A spatial sampling technique is applied to select a subset of SIFT features individually from features extracted from each of the three biometric traits. The selected features are then combined into a single, higher dimensional fused feature set. Both the multi-modal (face and iris) and multi-unit (left and right irises) sources of information are combined to obtain a single super SIFT feature vector set which is used for matching and classification.

The main novelty of the work stems from the introduction of SIFT [14-16] features to process iris images and the combination of the features in a global SIFT feature vector set. The proposed approach has the advantage over [12][13] of adopting of common feature extractor which delivers more stable features. The applied feature reduction technique is very simple and does not require a preliminary training. Also for classification, a direct matching technique of the fused feature vectors is adopted. Moreover, the work also presented the fusion of both multi-modal and multi-unit sources of information at feature level which further enhanced the results.

Experimental results on the database, composed of a subset of iris images from the CASIA version 3 database [18] and its chimerical combination with the Equinox database [19] are reported. The results obtained demonstrate the performance enhancement of feature level fusion in comparison to uni-modal systems and score level fusion. Section 2, describes the Scale Invariant Features Transform and its application for face and iris. Section 3, describes the feature level fusion of face with iris, both left and right irises. Experimental results are presented in section 4. Conclusions are drawn in Section 5.

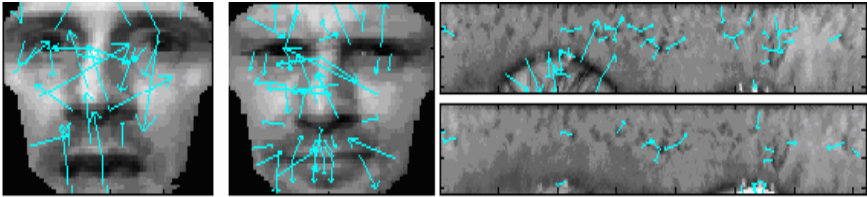
## 2 Scale Invariant Feature Transform (*SIFT*)

David Lowe proposed a method to extract distinctive and invariant features from images applied to general 3D object recognition [14]. These features known as *Scale Invariant Feature Transform* or *SIFT* features are invariant to image scale and rotation. Thus they provide a basis for object representation to perform a robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of

noise, and change in illumination. SIFT features are well localized in both the spatial and frequency domain, reducing the probability of disruption due to occlusion, clutter, or noise. In addition, SIFT features are highly distinctive. This property allows a single feature to be correctly matched with high probability against a large database of features, providing a basis for robust recognition. The cost of extracting these features is minimized by adopting a cascade filtering approach in which the more time consuming operations are applied only at image locations which have been selected at the initial phase [14].

Due to the stability and robustness of these features, they have been recently been applied for face and fingerprint biometrics [15][16]. Each extracted SIFT features can be defined as  $S = (s_1, s_2, \dots, s_m)$ , where each feature  $s_i = (x, y, \theta, Keydesc)$  includes the spatial location  $(x, y)$ , the local orientation  $\theta$  and the key descriptor of size  $1 \times 128$ . Key descriptor i.e., *Keydesc* part of each SIFT feature is invariant to affine transformations and is adopted for experiments in [15][16].

This work investigates, the applicability of *SIFT* features for iris biometrics and its fusion with the face biometric at feature level fusion. The extracted SIFT features from face and iris images as shown in Fig. 1. In this work also, keypoint descriptors (*Keydesc*) are considered for image representation and experiments.



**Fig. 1.** Shows extracted SIFT Features from the face and iris images. Even with the strong intra-class variations many common SIFT Features can be easily noticed for both the biometrics.

### 3 Feature Level Fusion

Feature level fusion is performed by the simple concatenation of the feature sets obtained from different sources. Let  $X = (x_1, x_2, \dots, x_m)$  and  $Y = (y_1, y_2, \dots, y_m)$  denote the feature vectors extracted from two different sources. The vector  $Z = (z_1, z_2, \dots, z_m)$  is formed by the concatenation of the two feature sets  $X$  and  $Y$ . Generally, before performing the concatenation, the two vectors  $X$  and  $Y$  are normalized to ensure the same range and scales of values. In order to reduce the dimensionality of the fused feature space, generally a feature selection technique is also applied to the individual or fused features. The similarity between two individuals is determined by matching the instances of the fused vector  $Z$  obtained from database and query sources of information.

By applying SIFT features to represent both the face and iris traits, the entire process of feature level fusion is greatly simplified without compromising recognition ability. In fact, the feature normalization process is not required, because the features from both sets are already commensurable and the matching step is reduced to the computation of the Euclidean distance metric and pairing of the nearest neighboring features between two concatenated database and query feature vectors.

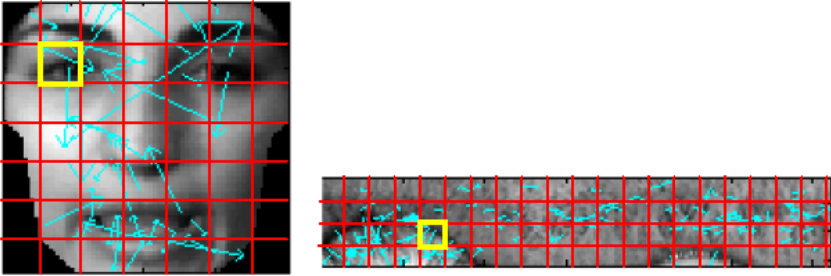
The entire matching process is described by the following steps:

- a) **Image Pre-processing and Feature Extraction:** Both face and iris images are pre-processed to extract only the relevant part of the image containing useful information. For face, a template mask is applied to the image to crop the central region of the face. This mask is registered and scaled to each face image on the basis of the position of the eyes and then the face portion is cropped. In the reported experiments, the eyes position are manually extracted but can be equally well extracted using an eye position extractor. It is worth noting that no further preprocessing is applied to the face images to compensate for illumination changes between samples. We deliberately chose not to perform an intensity normalization of the images to fully validate the robustness of the SIFT-based representation and the fused classifier. The iris images are segmented by using the technique proposed by Ross et al. [17] which is based on the implementation of Geodesic Curves. SIFT features [14] are extracted from the segmented and normalized iris images. As a result of which, a set of SIFT features is obtained from the face image and from left and right irises for each subject (Fig. 1).
- b) **Feature Selection:** The extracted SIFT features are then selected from both biometric traits by (Fig. 2)
  - i. dividing each image (face, left and right iris) into small windows of size  $3 \times 2$  for the iris and  $5 \times 5$  for the face. The total number of windows are 256 for face and 3780 for each iris image. The window size has been determined after several trials and taking the size which best captures the best discriminative information content in the data set. After performing several tests on different data, it was established experimentally that the optimal size only depend on the size of the input image.
  - ii. For each window the *SIFT* descriptor with average minimum Euclidean distance from all other descriptors within the same window, is selected. In case there is only one descriptor in the window, it is selected for inclusion into the fused feature vector.
- c) **Feature Concatenation:** The selected *SIFT* features from the face and the two irises are combined into a single super feature vector set, of dimension  $N \times 128$ , where  $N$  is equal to sum of all the features selected from different sources ,as a result of spatial sampling, for inclusion in the fused vector.

$$\begin{aligned}
 \text{Fused feature set} = & \text{face SIFT} \begin{pmatrix} 1 \times 128 \\ \dots\dots\dots \\ \text{left iris SIFT} \quad 1 \times 128 \\ \dots\dots\dots \\ \text{right iris SIFT} \quad 1 \times 128 \end{pmatrix}
 \end{aligned}$$

- d) **Feature Matching:** In order to match two fused feature sets  $f_1$  and  $f_2$ . Given a feature element  $p_{11}$  in  $f_1$ , the distances  $d_1$  and  $d_2$  between the first closest element  $p_{11}$  and the second closest element  $p_{21}$  in  $f_2$  are computed. If the ratio

$\frac{d_1}{d_2}$  is below 60%, then the vector element  $p_{11}$  is tagged to match with element  $p_{21}$ . The matching score between the two fused feature vectors is equal to the number of matching elements [16]. It is worth noting that this procedure maximizes the separation between matching and non-matching vector elements, thus reducing the probability of false matches between vector pairs.



**Fig. 2.** Example of feature selection by partitioning the face and iris images into a set of small windows and selecting one SIFT feature per window. The yellow boxes show an example of more than one SIFT feature in a single partition. Windows with no features are ignored.

## 4 Experimental Evaluation

CASIA iris database version 3 [18] and the Equinox face database [19] have been used to evaluate the performance of the proposed system. A total of 57 subjects with ten instances per each client have been taken from the Equinox database. The left and right normalized iris images of 125 clients from the CASIA version 3 database have been used for experiments.

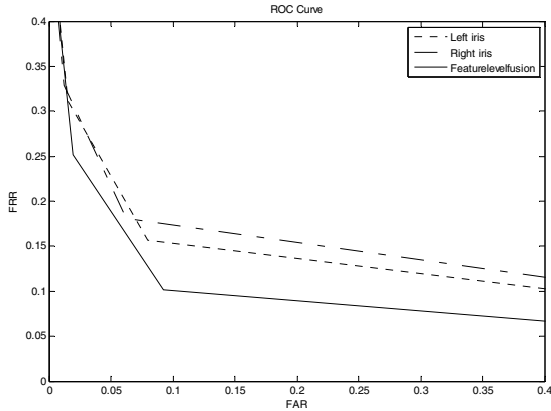
### 4.1 Protocol for Performance Evaluation

The following evaluation procedure has been applied for mono-modal and multi-biometric matching, with  $N$  subjects and processing 10 samples for each biometric trait per subject:

**Training:** one template is built for each subject using one image for each modality, i.e. one face image and one image for each of the two irises. The SIFT features are extracted from each image and both uni-modal and multi-modal feature vectors are built for the face, irises and the combination of the two irises and the face with one or two irises. In the performed experiments, the matching scores were computed to determine the discrimination capability of the single and fused representation.

**Testing:** Nine samples per person are used for testing and to generate the client matching scores. For the mono-modal systems, the impostor scores are generated by

matching the representation of each client against the first sample of all the other individuals in the dataset. In case of multimodal testing the client is tested against the first face and iris samples of the remaining of the chimerical users. Therefore, in total  $9xN$  client scores and  $9xNx(N-1)$  impostor scores for each of the uni-modal and multimodal representations are generated.



**Fig. 3.** ROC curves representing the matching performance of individuals based on the SIFT representation extracted from the left and right iris and the feature level fusion of the two irises

## 4.2 Multi-unit Fusion: Left and Right Irises

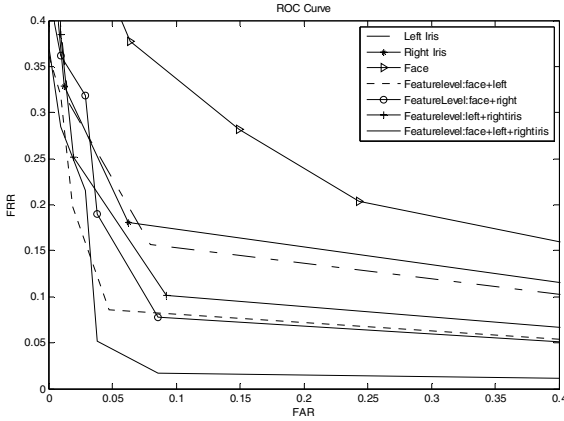
For testing the application of SIFT representation to the iris biometric, 125 clients from the CASIA version 3 database [18] with ten left and ten right iris images are used as mentioned in 4.1. The iris images are normalized and segmented as discussed in [17]. Left and Right iris are confirmed to contain discriminatory information for each individual so their combination is a source of complementary information [20]. Thus feature level fusion of left and right iris using SIFT is performed as described in section 3 and it was evaluated against left and right uni-modal traits.

The performance is evaluated as proposed in section 4.1. In total,  $125x9=1125$  client scores and  $125x124x9=139500$  impostor scores are generated for each iris. The ROC curves obtained from the matching scores of the left and right iris images using SIFT features and feature level fusion are shown in Fig. 3. As it can be noted from the ROC curves, the feature level fusion of these complementary information sources considerably enhances the identification capabilities.

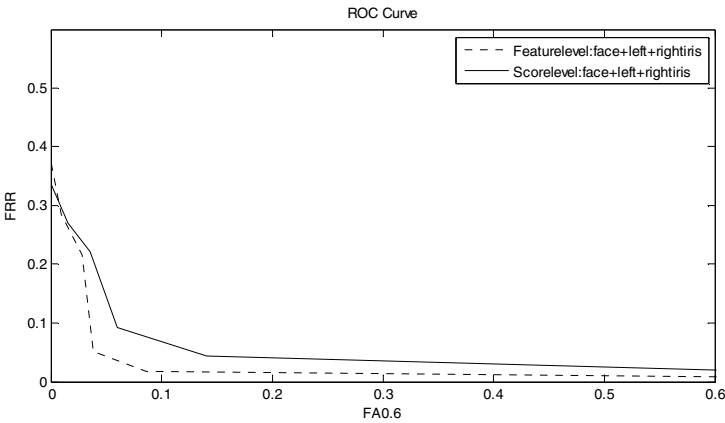
## 4.3 Multi-modal with Multi-unit Fusion: Face, Left and Right Irises

To test the performance of matching the fused SIFT representation based on face and iris, a chimerical database is obtained by coupling 57 clients images from CASIA [18] and Equinox face database [19]. The SIFT features are extracted from the normalized face and iris images. The SIFT representation of the face and iris biometric traits are fused at feature level as detailed in section 3. The performance is evaluated as

reported in section 4.1. In total  $57 \times 9 = 513$  client scores and  $57 \times 56 \times 9 = 28728$  impostor scores are generated for the uni-modal face matching and the face and iris fused vector matching. The ROC curves representing the error distributions of the feature level fused matching are shown in Fig 4. The curves are generated from the scores obtained by matching the mono-modal face representations and the multi-modal feature level fusion with left and right iris images.



**Fig. 4.** ROC curves representing the matching performance of individuals based on the SIFT representation extracted from the face, the left and the right iris and the feature level fusion of the three traits



**Fig. 5.** Comparison of the feature level fusion, of the SIFT representations of the face, the left and the right irises, with the score level fusion of the matching scores of the individual modalities

It is worth noting that the error distributions are lowered by combining more sources of information at the feature level. The performance of the feature level fusion of face and the two irises is compared with the score level fusion in Fig 5. As it can be noted, the data fusion performed at the earlier stage, i.e. feature level, produces better performances in terms of error distributions than the fusion at the score level.



## 5 Conclusion

The fusion of face and iris biometrics is a natural process in a multi-biometric system which, in principle, can be implemented without introducing extra additional cost for acquisition. At the same time, the two biometrics are strongly complementary. In fact, while face biometric samples are easier to acquire with a conventional camera, iris has demonstrated very superior performances. This suggests an image acquisition system with a camera acquiring face images at a high resolution to provide shape and textural information for both the two irises and the face. Current camera devices allow to sample up to 15 M pixels images at a rate of 30 frames per second and deliver the data stream over a fast Ethernet channel. By properly controlling the image acquisition with a fast face detection algorithm to keep the head within one third of the total image area, the resulting images can provide iris images up to 64 K pixels. Several systems to acquire iris from a distance have been proposed and remarkably demonstrated the possibility to acquire both the face and the iris at the same time [21].

In this paper a novel approach to feature level fusion of face and iris has been proposed. The main advantages of the proposed fusion method are the ease of implementation and the robustness of the resulting representation. Due to the scale-space analysis, SIFT features proved to be very stable and almost insensitive to illumination variations while providing a scale and translation invariant representations. At the same time, the adoption of a common feature representation greatly simplifies the normalization, concatenation and matching processes in the feature level fusion, which is generally an issue related to fusion at this level. From the experiments performed on a mixed database obtained by combining face images from the Equinox database and iris images from the CASIA v.3 database the representation based on feature level fusion demonstrate superior matching performance with respect to uni-modal systems and score level fusion. Some issues are still under investigation, such as the optimal feature selection scheme based on quality driven analysis and the evaluation on large multi-modal database.

## References

1. Hong, L., Jain, A., Pankanti, S.: Can Multi-biometrics Improve performance. In: Proc. of AutoID 1999, pp. 59–64 (1999)
2. Jain, A.K., Ross, A.: Multi-biometric systems. *Communications of the ACM* 47(1), 34–40 (2004)
3. Ross, A., Jain, A.K.: Information Fusion in Biometrics. *Pattern Recognition Letters* 24, 2115–2125 (2003)
4. Chibelushi, C.C., Mason, J.S., Deravi, F.: Integration of acoustic and visual speech for speaker recognition. In: *EUROSPPEECH 1993*, pp. 157–160 (1993)
5. Duc, B., Maître, G., Fischer, S., Bigün, J.: Person authentication by fusing face and speech information. In: Bigün, J., Borgfors, G., Chollet, G. (eds.) *AVBPA 1997*. LNCS, vol. 1206. Springer, Heidelberg (1997)
6. Hong, L., Jain, A.: Integrating Faces and Fingerprints for Personal Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(12), 1295–1307 (1998)

7. Ross, A., Govindarajan, R.: Feature Level Fusion Using Hand and Face Biometrics. In: Proc. of SPIE Conference on Biometric Technology for Human Identification II, Orlando, USA, pp. 196–204 (2005)
8. Zhou, X., Bhanu, B.: Feature fusion of face and Gait for Human Recognition at a distance in video. In: International Conference on Pattern Recognition, Hong kong, (2006)
9. Singh, S., Gyaourova, G., Pavlidis, I.: Infrared and visible image fusion for face recognition. In: SPIE Defense and Security Symposium, pp. 585–596 (2004)
10. Wang, Y., Tan, T., Jain, A.K.: Combining Face and Iris Biometrics for Identity Verification. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 805–813. Springer, Heidelberg (2003)
11. Zhang, Z., Wang, R., Pan, K., Li, S.Z., Zhang, P.: Fusion of near infrared face and iris biometrics. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 172–180. Springer, Heidelberg (2007)
12. Son, B., Lee, Y.: Biometric Authentication System Using Reduced Joint Feature Vector of Iris and Face. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 513–522. Springer, Heidelberg (2005)
13. Gan, J., Liang, Y.: A Method for Face and Iris Feature Fusion in Identity Authentication. IJCSNS, 6 (2B) (2006)
14. Lowe, David, G.: Object recognition from local scale invariant features. In: International Conference on Computer Vision, Corfu, Greece, pp. 1150–1157 (September 1999)
15. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the use of SIFT features for face authentication. In: Proc. of Int. Workshop on Biometrics, in association with CVPR (2006)
16. Park, U., Pankanti, S., Jain, A.K.: Fingerprint Verification using SIFT Features. In: Proc. of SPIE Defense and Security Symposium, Orlando, Florida (2008)
17. Ross, A., Shah, S.: Segmenting Non-ideal Irises Using Geodesic Active Contours. In: Proc. of Biometrics Symposium (BSYM), Baltimore, USA (2006)
18. <http://www.cbsr.ia.ac.cn/english/IrisDatabase.asp>
19. <http://www.equinoxsensors.com/products/HID.html>
20. Wu, X., Wang, K., Zhang, D., Qi, N.: Combining left and right irises for personal authentication. In: Yuille, A.L., Zhu, S.-C., Cremers, D., Wang, Y. (eds.) EMMCVPR 2007. LNCS, vol. 4679, pp. 145–152. Springer, Heidelberg (2007)
21. Matey, J.R., Naroditsky, O., Hanna, K., Kolczynski, R., LoIacono, D.J., Mangru, S., Tinker, M., Zappia, T.M., Zhao, W.Y.: Iris on the Move: Acquisition of Images for Iris Recognition in Less Constrained Environments. Proc. of the IEEE 94(11), 1936–1947 (2006)