ROBUST RETINA-BASED PERSON AUTHENTICATION USING THE SPARSE CLASSIFIER

Alexandru Paul Condurache, Johannes Kotzerke and Alfred Mertins

Institute for Signal Processing, University of Lübeck
Ratzeburger Allee 160, D-23538, Lübeck, Germany
phone: + 49(0)451 5005801, fax: + 49(0)451 5005802, email: condurache@isip.uni-luebeck.de

ABSTRACT

We address the problem of person authentication, including verification and identification, using the vascular network of the retina. We propose a novel feature extraction process that includes the segmentation of feature points related to anatomical characteristics of the retinal vessel-network, the description of these points with the help of the scale-invariant feature transform (SIFT) and the computation of a final feature vector related to the statistical characteristics of the SIFT-based description. After feature extraction, authentication is conducted with the help of the sparse classifier. We successfully test our methods on two databases, one publicly available and the other one (that we now make available as well) specially generated for this purpose. The results show that apart from high accuracy, the proposed algorithm enjoys a set of invariance properties that make it robust to a set of issues afflicting retina-based person authentication systems, while being fast enough to allow practical deployment.

Index Terms— sparse classifier, covariance features, retina-vessel segmentation, person authentication

1. INTRODUCTION

The pattern of vessels supplying blood to the retina is a unique feature in each eye and can be used to authenticate an individual [1, 2]. This feature is impossible to forge and the blood vessels decay too fast to allow the eye of a deceased person to be used to deceive the system. Furthermore, the vascular pattern is virtually constant over the entire life span of the target individual, the most often exceptions being pathological cases like diabetic rethinopathy. Even though it enjoys such desirable features in comparison to other biometric traits, it is by far the least used. The reason for this is the acquisition procedure that is considered intrusive and requires a relatively high degree of user cooperation (for example eyeglasses must be removed). From a historical perspective there has been a tradeoff between the quality of the acquired retinal scans and the amount of data analysis needed for the purpose of authentication. The first commercial retina-based person authentication systems acquired high-quality images in a tightly controlled environment, for which purpose they used visible light to illuminate the retina [3]. This procedure was very uncomfortable for the user and as a consequence a near-infrared light source replaced the visible-light source shortly afterwards. Later, the amount of energy radiated by this source, as well as the acquisition time, was decreased more and more with each new retinal-scan system on the market to improve acceptance. However, as a direct consequence, the quality of the obtained images deteriorated. Thus, the improved acceptance generated the need for more powerful data analysis tools to accomplish the intended authentication task. Apart from a decreasing signal-to-noise ratio (SNR), the acquired retinal scans may be afflicted by geometric transforms like rotation and translation but also a small amount of scaling due to the eye movement or head placing with respect to the sensor. Clearly, retina-based person authentication must be invariant to such disturbances. Additional difficulties are encountered with people suffering from astigmatism and under some circumstances with those wearing contact lenses.

With high-quality retinal scans conducting person identification is usually a relatively easy task. The first solution [2] simply used the Fourier transform for feature extraction, to deal with some of the imaging-related issues, followed by correlation to measure the similarity between two images. More modern approaches still use correlation [4, 5, 6, 7], but the preprocessing is different. Geometric distortions due to the image acquisition process are dealt with using polar coordinates [4], or image registration [6, 5] with vessel parts as cues. In [7], a more complicated feature extraction process is used, with a polar transform followed by a wavelet-based multiresolution analysis of segmented vessels, also modified correlation coefficients over several scales are used together to reach a decision. In other approaches [8], vessel segmentation together with a simple orientation analysis on the segmentation result using an angular partition are used for feature extraction. The feature vectors thus obtained are compared with the help of the Manhattan distance. In [9], two retinal images are matched based on a set of feature points of the vessel pattern. These feature points are anatomically interesting points of the vascular network like bifurcations and cross-overs, that can be easily detected despite low SNR. After exacting them, the optimal transform that registers the two feature-points clouds is found. Then, the number of landmark-pairs (i.e., pairs of fea-

ture points that are considered to be the representation of the same anatomical point in each image) is used to compute the similarity between two images. A similar feature-point-based strategy is used in [10] as well.

Our approach is based also on feature points, extracted from the vessel segmentation result computed with the hysteresis classifier [11], that is particularly robust to noise. We extract one feature vector per image. The features we extract exhibit all the invariance properties required by retina-based person authentication. With our feature extraction, unlike other approaches, we do not need to align the two images to find out if they come from the same eye, which clearly reduces the computational burden of our algorithm. Together with the subsequent sparse-classifier based decision, this results in a particularly robust algorithm.

2. METHODS

For each retina image a cloud of feature points is computed. The subsequent feature extraction process has as its aim to provide a vector-based description of this point cloud. Then, person authentication is conducted in this feature space with the help of the sparse classifier.

2.1. SIFT-log-covariance based point-cloud features

The feature-point cloud should offer a unique and compact description of the target retinal vessels. Highly informative points in this context are the vessel bifurcations, as they are related to anatomical characteristics of the vascular network [12] but also the vessel crossings, i.e., the points where a retinal vessel goes over or under another one. At the same time such points offer a set of invariance properties, like rotation invariance, that makes them well suited for our purposes.

Here, bifurcations and crossings are detected on binary images with vessel centerlines. The vessel centerlines are computed with the help of morphological image processing methods from segmented vessels. The vessel are segmented with the help of the hysteresis classifier. After detecting the feature points we compute the corresponding Scale Invariant Feature Transform (SIFT) descriptors [13]. Our final feature vector is related to the statistical properties of the SIFT-descriptor sample in the feature-point cloud. Working this way allows us to be robust with respect to potential outliers in the form of cloud components that are neither bifurcations nor crossings.

2.1.1. Computing the feature-point cloud

For security applications, we are interested to reliably detect crossings and bifurcations (i.e., we want to detect only points where we are highly confident that they are at a vessel crossing or bifurcation). Thus, we morphologically process [14] the hysteresis-segmentation result (see Fig. 1 (a)) such as to obtain only the large and mid-sized vessels, where we can reliably detect the feature points.

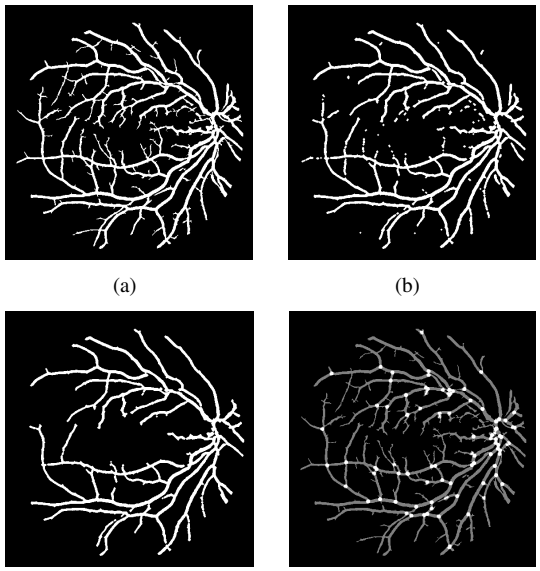

Fig. 1. Original vessel segmentation (a), result of the opening (b) and final result after elimination of the small vessels (c), and feature-point cloud (white) and the corresponding vascular network (gray) (d)

The first step to detect the feature points is to open the original segmentation with a disk-like structuring element of a radius larger than the smallest vessels (see Fig. 1 (b)). After this, we need to select only the main vessels. To this end we apply successive erosion steps to the result of the opening, until all vessel points are eliminated and then use the vessel points from the last but one erosion result as markers for the large vessels. More precisely, we use these points to select from the opening result only the large vessels. For this purpose all points linked to the markers over an eight-neighborhood (i.e., the 3×3 region of interest centered at the keypoint) are selected (see Fig. 1 (c)).

After eliminating the small vessels, we thin the large ones until they are one-pixel thick. Here, the crossings and bifurcations are easily detected by counting the object points in the eight-neighborhood of each object pixel. If the count is larger or equal to three we detect a feature point. A feature-point cloud with the corresponding vascular network is shown in Fig. 1 (d).

2.1.2. The log-covariance feature vector

After computing the locations of the points in the feature-point cloud, we extract for each component of the cloud the corresponding SIFT descriptor. The SIFT descriptor provides a unique and largely invariant representation of the local neighborhood of the corresponding point [15]. This representation is based on the local orientation and is linked

to the scale at which the point can be optimally described. As the SIFT descriptor is rather large (128 components), we first reduce its dimension with the help of PCA. This has the benefit of a smaller computational burden and improved fit to the sparse-classifier framework. The PCA is computed once, from all SIFT descriptors, from all images in the training set.

We compute the final feature vector from the diagonal of the log-covariance matrix [16] that we estimate from the available PCA-transformed feature-point cloud SIFT descriptors. The log-covariance matrix is the reconstruction of the covariance matrix from its eigen decomposition, where the eigenvalues have been replaced with their natural logarithms. This procedure is needed to ensure that our feature space is a true vector space, such that linear combinations of vectors from this feature-space leads to a vector that is himself a component of the feature space, i.e., the feature space has the property of bing closed under linear combinations. This type of feature extraction offers us an additional set of invariance properties to some simple transformation of the SIFT descriptors.

2.2. Sparse classification

Sparse-representation based classification looks for the sparsest representation of a test vector in terms of a matrix of training vectors. This representation is sparse because it should contain only vectors from the class to which the test vector belongs [17]. With the training matrix $\mathbf{T} = [\mathbf{T}_1, \ldots, \mathbf{T}_k]$, containing the class-submatrices $\mathbf{T}_i = [\boldsymbol{v}_{i,1}, \ldots, \boldsymbol{v}_{i,N_i}]$ with $i = 1, \ldots, k$, where N_i is the number of vectors in class i and k the number of classes, for each new vector $\boldsymbol{y}$ we ideally have $\boldsymbol{y} = \mathbf{T}\boldsymbol{x}$, the sparse coefficient vector $\boldsymbol{x}$ having entries different from zero only for the training-space vectors from the class i to which $\boldsymbol{y}$ belongs. For classification purposes we compute $\boldsymbol{x}$ and assign $\boldsymbol{y}$ to the class whose corresponding entries in $\boldsymbol{x}$ are different from zero. The total number of vectors in $\mathbf{T}$ is $n = \sum_{i=1}^{k} N_i$ and each vector has m entries. For computational reasons we find the sparsest $\boldsymbol{x}$ as $\hat{\boldsymbol{x}} = \arg\min \parallel \boldsymbol{x} \parallel_1 \quad \text{subject to} \quad \parallel \mathbf{T}\boldsymbol{x} - \boldsymbol{y} \parallel_2 \leq \epsilon$, with a small positive ϵ, optimizing over the l_1 instead of the l_0 norm of $\boldsymbol{x}$. The decision rule reads then: $C(\boldsymbol{y}) = \arg\min_i \parallel \boldsymbol{y} - \mathbf{T}(1_i \odot \hat{\boldsymbol{x}}) \parallel_2$, where we use the following notation: $1_i = [b_1, \ldots, b_n]^T$, $b_l \in \{0, 1\}$, $l = 1, \ldots, n$ is the selection vector for class i and $\boldsymbol{v}_1 \odot \boldsymbol{v}_2$ is the component-wise product of two vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, i.e., the Hadamard product. Thus, $1_i \odot \boldsymbol{x}$ selects the entries of $\boldsymbol{x}$ where the coefficients of class i reside. $C(\boldsymbol{y})$, with $C : \mathbb{R}^m \to \{1, \ldots, k\}$ is the function that assigns a class label to the vector $\boldsymbol{y}$.

Depending on the maximal number of examples per class, i.e., the maximal number of available retinal scans of an enrolled person, we can establish the minimal size of the final feature vector such as to ensure the appropriateness of the sparse-classifier framework for this problem. For the sparse classifier to work properly [17], we need that $m \geq c \cdot 2$, with m the dimension of the feature vector, and c the maximal number of examples per class, i.e., the largest N_i. At the same time the $m \times n$ training matrix $\mathbf{T}$ usually needs to be underdetermined, such that $n > m$.

3. EXPERIMENTS AND DISCUSSION

To test and demonstrate the qualities of our system, we have used two databases. The first database we have used is a publicly available database the VARIA [18], the second one was constructed by us from the DRIVE database [19]. The VARIA database has 233 images from 139 different individuals, out of which 59 had two or more samples. The optic-disc centered images have been acquired over a period of several years with a TopCon NW-100 model non-mydriatic retinal camera at a resolution of 768x584. These images have a high variability in contrast and illumination. Our DRIVE for Retinal Authentication (DRIVERA) database[1] contains 280 images. These were generated from the 20 images of the test set of the DRIVE database, at a resolution of 576×560 pixels. For this purpose we have created for each DRIVE image 14 more images while applying various types of distortions. These distortions are supposed to simulate different image acquisition-related problems that may appear when the same retinal vasculature is imaged at different times. We have divided the distortions into three categories: person-related, optics related and sensor related. The person-related distortions are small rotations, translations and scalings of the original image, the optics-related distortion are blurring, barrel and pincushion transforms applied to the original image and the sensor-related distortions are changes in illumination, white and "salt&pepper" noise. The new images were generated by randomly applying these distortions, such that an image may be affected by one or by several such distortions. The DRIVERA database has only gray-level images obtained from the green channel of each original image.

As previously discussed, the dimension of the feature space must be chosen in accordance with the maximal number of examples per class available in the training set. For the VARIA database this is $c = 7$ and for the DRIVERA database this is $c = 14$. Cross-validation experiments on half of the DRIVERA database led us to choosing $m = 57$ (the results were largely similar for $m > 50$). The same m was used on the VARIA database. This means that the PCA-transformed SIFT-descriptors have a dimension of 57.

During morphological processing, we would like to eliminate only the small vessels from the original segmentation. For this purpose we need the diameter of these vessels, which depends on the image acquisition hardware. In our case, this diameter was two pixels.

[1]The software to compute the DRIVERA database from the DRIVE database may be downloaded from http://www.isip.uni-luebeck.de/index.php?id=610&no_cache=1

We have conducted classifications in three scenarios: with manually selected anatomical feature points (Manual), with SIFT keypoints (SIFT) and with the vessel segmentation-based feature-point cloud, as described above (Segmentation). In each scenario we have applied three types of classification algorithms: the first one (Diff) is based on the sum of squared differences (SSD) between the scaling-aligned query image and the train-set image, the second one (Ransac) is based on the number of landmark pairs between the query image and the train-set image (similar to the algorithm from [9]) and finally the third one is the log-covariance feature vector and sparse-classification framework proposed here. The results obtained by leave-one-out cross-validation are shown in Fig. 2.

For the first and second classifiers we need landmark pairs to compute the parameters of the scaling and the maximal number of landmarks respectively. To find landmark pairs, we use a procedure similar to the RANdom Sample Consensus (RANSAC) algorithm [20]. To begin with, we search for sets of matched SIFT keypoints from the two images [21]. Assuming we find more than four matches, we use all possible sets of four matches to register the two images with a scaling transform. We then count the number of landmarks for this transform. Landmarks are feature-points pairs that after the transformation have positions that are very similar, the Euclidian distance between their position vectors being smaller than a small threshold. The number of landmarks describes the quality of the transform. We declare the optimal transform as the one with the largest number of landmarks. Each image pair where we found more than four matches is described by a certain transform and thus a number of shared landmarks. Image pairs where less than four matches were found are discarded.

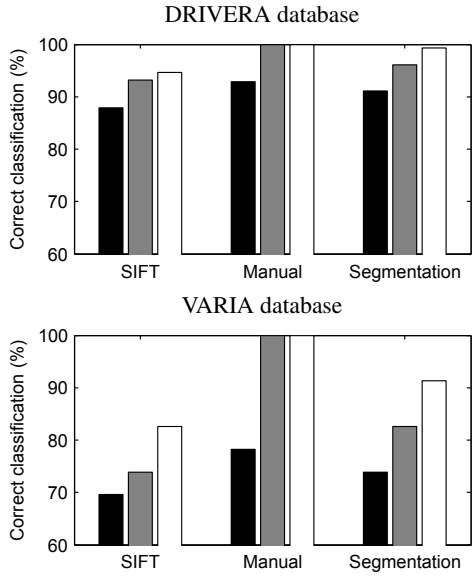

Fig. 2. Classification results obtained by the various methods: Diff (black), Ransac (grey), Sparse (white)

As the sparse classifier needs several images to work properly, we have tested only the four individuals from the VARIA database with five or more images. To ensure proper deployment of the sparse classifier, we have used all available examples when computing the sparse vector. For the DRIVERA database we have conducted experiments with the sparse classifier in two scenarios: with seven images per eye and with all 14 images. The results improved from 90.71% to 99.29% correct decisions, when using more images.

On the DRIVERA database, when the sparse classifier works on all SIFT keypoints it achieves 94.64% correct classifications, while when working only on the anatomic-relevant feature points it achieves 99.29% for the automatically segmented feature points and 100% for the manual ones. A similar behavior has been observed on the VARIA database. This shows that it is indeed advantageous to conduct our person authentication using only anatomically-relevant feature points, thus completely ignoring the background. It also shows that there is room for improvement with respect to the computation of the feature-point cloud.

Our system is fast, it returns a decision in six seconds, by comparison the RANSAC-based algorithm needs 14 seconds, while the SSD-based classification 66 seconds on an Intel Core i5 (3.1GHz) machine with 16GB RAM.

4. CONCLUSIONS AND SUMMARY

Vessel bifurcations and crossings together with many more other interesting feature points can be detected by applying SIFT directly to the original retinal image. However, in this case we obtain besides the sought points a large set of other points whose relationship to the vascular network is questionable (e.g., points in the background), even though their relationship to each analyzed image is strong. This in turn decreases the descriptive power of our covariance-based feature vector, as it will answer more to the background than to the vascular network, and thus each retinal-vessel network will appear to be similar to any other one. We are therefore interested in a feature-point selection procedure that is rather specific with respect to crossings and bifurcations. Conversely, it is possible that a few feature-point cloud components are neither crossings nor bifurcations. However, this is far from critical considering that our final feature vector is computed from the empirical covariance matrix of all SIFT descriptors. We need only a majority of cloud points to be true crossings and bifurcations.

Even though a feature vector based on anatomical feature points related to the vascular network of a retina is well suited for such tasks as person authentication, the problem until now was that a reliable vessel segmentation would take a rather long time [9, 10]. The hysteresis classifier is fast and accurate enough at to render such an approach feasible. Our feature vector is related to the second-order statistical properties of the feature vectors corresponding to the feature-point cloud.

Clearly, other statistical descriptors could be used. It remains to be investigated if relating our feature vector to higher-order statistics, or using other types of statistical descriptions than moment-based ones, represents an improvement.

The results obtained with the sparse classifier improve, the more images are available for each eye in the database. For a practically deployed retina-based authentication system, one should either acquire several images during enrollment, when setting up the system, or start with one image per eye and use, e.g., the classifier based on the number of landmark-pairs shared by two images to conduct classification, recording new images each time the respective eye is imaged, until enough images have been acquired to be able to sensefully use the sparse classifier.

Using the sparse classifier offers an easy way to deal (while the system is in use) with changes that occur in the acquired images over a long period of time, due to various non-critical problems with the image-acquisition hardware. To compensate for this one should simply update the training set of a certain eye by adding new images recorded at fixed time intervals. Furthermore, also by concentrating on the mid and large vessels and extracting the final feature vector with the help of the log-covariance descriptor (that is robust to outliers, meaning that a few additional crossings and bifurcations will not lead to significant changes), we are robust to pathological changes of the retinal vessels such as small-vessel proliferation, encountered in the case of diabetic retinopathy.

We have described a robust retina vasculature-based person authentication system. It is based on a novel feature extraction process using anatomic-relevant points on the retina vessel tree, SIFT features and a log-covariance descriptor together with a sparse classifier. Our system is fast and exhibits invariance properties that make it robust to a set of issues related to both image acquisition and pathological changes in the retina.

5. REFERENCES

[1] C. Simon and I. Goldstein, "A new scientific method of identification," *New York State Journal of Medicine*, vol. 35, pp. 901–906, 1935.

[2] R. R. Hill, *Retina identification*, Biometrics: Personal Identification in Networked Society. Kluwer Academic Press, Boston, 1999.

[3] R. B. Hill, "Apparatus and method for identifying individuals trough their retinal vascular patterns.," Tech. Rep., US patent - nr. 4109237, 1978.

[4] S. M. Raiyan Kabir, R. Rahman, M. Habib, and M. R. Khan, "Person identification by retina patern matching.," in *Proceedings of ICECE-2004*, Dhaka, Bangladesh, December 28–30 2004, pp. 522–525.

[5] C. Marino, M. G. Penedo, M. Penas, M. J. Carreira, and F. Gonzalez, "Personal authentication using digital retinal images.," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 21–33, 2006.

[6] K. Fukuta, T. Nakagawa, Y. Hayashi, Y. Hatanaka, T. Hara, and H. Fujita, "Personal identification based on blood vessels of retinal fundus images.," in *Proceedings of MI: Image Processing*, San Diego, US, February 16–21 2008, pp. 69141V–69141V–9, SPIE.

[7] H. Farzin, H. A. Moghaddam, and M.-S. Moin, "A novel retinal identification system," *EURASIP J. Adv. Sig. Proc.*, vol. 2008, pp. 1–10, 2008.

[8] W. Barkhoda, F. A. Tab, and M. D. Amiri, "Rotation invariant retina identification based on the sketch of vessels using angular partitioning.," in *Proceedings of IMCSIT-2009*, Margowo, Poland, October 12–14 2009, pp. 3–6, IEEE.

[9] M. Ortega and M. G. Penedo, "Retinal vessel tree as biometric pattern.," *Biometrics*, vol. http://www.intechopen.com/articles/show/title/retinal-vessel-tree-as-biometric-pattern, pp. 113–138, 2011.

[10] L. Latha, M. Pabitha, and S. Thangasamy, "A novel method for person authentication using retinal images.," in *Proceedings of ICICT-2010*, Tamil Nadu, India, February 12–13 2010, pp. 1–6.

[11] A. P. Condurache and A. Mertins, "Segmentation of retinal vessels with a hysteresis binary-classification paradigm," *Computerized Medical Imaging and Graphics*, vol. 36, no. 4, pp. 325–335, June 2012.

[12] A. P. Condurache, *Cardiovascular biomedical image analysis: methods and applications.*, GCA-Verlag, Waabs, 2008.

[13] D. G. Lowe, "Object recognition from local scale-invariant features.," in *Proceedings of ICCV*, Kerkyra , Greece, July 27–31 1999, vol. 2, pp. 1150–1157, IEEE.

[14] R. C. Gonzales and R. E. Woods, *Digital image processing.*, Addison-Wesley, 1993.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, pp. 91–110, November 2004.

[16] Kai Guo, Prakash Ishwar, and Janusz Konrad, "Action recognition using sparse representation on covariance manifolds of optical flow," in *Proc. AVSS*, 2010, pp. 188–195.

[17] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE TPAMI.*, vol. 31, no. 2, pp. 210–227, 2009.

[18] http://www.varpa.es/varia.html., "Varpa retinal images for authentication (varia).," .

[19] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina.," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[20] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, June 1981.

[21] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Proc. IEEE Computer Society Conf Computer Vision and Pattern Recognition*, 1997, pp. 1000–1006.