Scene coherence analysis method by means of a Trace Transform based global descriptor for face verification support

Gabriela N. Domide[1], Igor G. Olaizola[1], Naiara Aginako[2] and Basilio Sierra[2]

***Abstract*— Face verification tries to determine whether a person is or not who she/he honestly claims to be. This is known to be a very challenging task in many computer vision applications due to different parameters variations like illumination, pose or expression. Therefore, scene coherence analysis for face verification aims to offer another perspective for the problem and, associated with the global image contextual information, from semantic domain inference based on a trace transform (DITEC) method, gives promising results. The images are processed in an efficient manner by means of global descriptors, for they provide inexpensive yet valuable information compared to local descriptors. In this paper, we focus on learning a distance metric for facial verification from the robust descriptors. The proposed approach is tested working on real-world face verification data.**

I. INTRODUCTION

Recent researches in the face recognition area try to develop image analysis mechanisms based on feature extraction and matching processes so that it can perform face detection and verification in an intelligible way as a human being achieves it. The image representation methods found a wide field of efficient applications from supervised user authenticity, image search and retrieval, augmented reality until show boundary ones. The core idea of many algorithms programmed to certify whether or not a person is truthful about its identity relies on the detection of facial features followed by one to one validation of faces images. Researchers in the computer vision field find it challenging working with changes in illumination direction, facial expression and pose. A lot of work has been done to provide essential information from images by extracting features either globally or locally. These features need to provide robustness and invariance to certain spatial transformations. The images are represented by vectors filled with descriptors and the most used methods rely on a face detection technique and on evaluating the descriptors based on the pixels in the place where the face was detected. Depending on the application, the length of the feature might be a parameter to take into account when it is wanted to achieve real-time performance. Viola et al. [35] proposed a face detection method based on boosting, known as the first algorithm that could be used in a real-time application due to the classifier learning with AdaBoost and the cascade structure. They thought of training a classifier with AdaBoost containing positive and negative examples, images of faces and respectively images that do not contain faces and running a classifier on a sliding window over the image. Modern technology came with robust feature extraction methodologies, such as Scale Invariant Feature Transform (SIFT) [16], Histogram of oriented Gradients (HoGs) [10], Speeded Up Robust Features (SURF) [5] and DAISY [31]. Another feature extractor is the Local Binary Patterns (LBP) [21]. The majority of these algorithms present the following problem: they can fail to detect the face, which is a critical step in all these methods. Therefore they need a backup artificial intelligence algorithm in order to minimize the errors.

We have not found in the literature works which deal with this backup problem. Therefore, in this paper we present a method that learns a Mahalanobis, Cosine and Correlation metric in the original space, where the images are represented by vectors containing global information based on Trace Transform. Being able to apply global features to different kinds of image retrieval applications and classification algorithms makes it an advantage for our proposed method.

II. RELATED WORK

Nearly all face recognition methods make use of relevant pixels or regions that show particular characteristics. Viola-Jones gave another perspective on how to detect a face in an image, introducing a set of efficient features that can be further analyzed, a feature selection algorithm called AdaBoost and an effective learning method based on a cascade architecture. [36] uses Gabor features, recalls the Viola-Jones face detection technique and creates discriminative descriptors for every keypoint independently. When it comes to context categorization, local features are widely used, and the most influential ones are SIFT, Gabor [14] and LE [8]. Another well-known feature set robust to illumination variations is the local binary patterns (LBP) [1], [21], which proved to be very effective for face recognition tasks [2]. Later, in [10] a similar method called histogram of oriented gradients (HoG) was proposed, which has become a very popular feature for human/pedestrian detection [38]. Sengupta et al. [26] makes use of different types of descriptors such as HoG, LBP and Fisher Vector [29]. Extracting features at specific location through key-points are presented in [3], [37], [39]. While there is a growing and flourishing trend for the face images characterization, the state of the art methods accuracy decreases as the angle of the face inclination increases. Based on the analysis of low-level features, Olaizola et al. [22] propose a framework for hypothesis reinforcement for context categorization and further hypothesis formation.

[1]Vicomtech–IK4, Mikeletegi Pasealekua 57, Donostia–San Sebastián, 20009, Gipuzkoa (Spain) nadia.domide@gmail.com, iolaizola@vicomtech.org

[2]University of the Basque Country, Manuel Lardizabal, 1-20018 Donostia–San Sebastián, 20008, Gipuzkoa (Spain) naiara.aginako@ehu.eus, b.sierra@ehu.eus

Global features applied to image retrieval systems describe the image as a whole, regardless of the content of isolated pixels and have the advantage of being efficient and simple. Huang et al. [12] proposed rather than engineer new image descriptors by hand an automatically representation through unsupervised feature learning with deep belief networks.

Torralba et al. [32] proposed using the scene context to extract global information in order to solve local uncertainty, thus combining global and local features in order to detect if a car is present or absent in an image and precisely locate where it appears in the image.

Zuo et al. [40] combines local and global features trying to improve the classification performance making use of the region detector, histograms of gradient, and scale invariant feature transform.

In the literature one can find different global descriptors like histograms of several local features [6], texture features, self similarity [28], GIST [33]. Cerra et al. [9] reflects on how the complexity of an image is associated with the context in which it may belong. A successfully applied method used as a global feature for image categorization and handwritten character recognition is the Ridgelet transform [19].

Some computer vision applications used the trace transform and one of them is included in the MPEG-7 [17] standard specification for image fingerprinting. But, specifically in the face recognition area are the applications: [11], [15]. When dealing with global descriptors, one has to take into account the final number of attributes to work with for they can increase the complexity at the analysis step. For example, the number of attributes can be reduced by applying Principal Component Analysis (PCA) in order to lower the dimensionality of the feature space through selecting the most relevant components. This approach should take into account that the complexity of the algorithm will grow because the covariance matrix information is formed with all the previous examples.

Nguyen and Bai [20] apply cosine similarity metric learning (CSML) to face verification, after extracting features by means of different methods: pixel intensity, LBP, and Gabor representation. In the following section we implement our approach for face verification based on the global features characteristics extraction DITEC [23], we analyze it and validate it experimentally through a real dataset. DITEC transforms the image into a parameter space. It has great preformance regarding higher discriminative global descriptors at low dimensionality.

III. Proposed approach

Computer vision mission is to investigate a 3D scene by understanding its 2D images through their properties. We propose an application that can overcome or intervene when the problem of a local image descriptor failing to detect a face arise. In Fig. 1 it can be seen that in a face verification application based on local descriptors it is important to accomplish the step in which the face is detected, thereafter being able to accept or not the classifying condition yielding a verified or not verified identity. Looking at Fig. 2, we introduce an automatic assistant that is able to detect if a change appeared in the new image.

We propose a state-of-art image analysis contribution relying on global descriptors matching based on similarity measures. The hierarchical framework consists of image database processing layer through DITEC resulting in vectors filled with global descriptors and the action towards these vectors layer through implementing distinct similarity measures between them.

For a better extraction of information from images, a previous action is usually made and that is transforming an image from the spatial domain into an alternative one, in which one can easily extract the needed information. Fourier transform, Laplace transform or Hough transform represent the spatial domain image transformation mainly used in the image processing area.

We apply the method proposed by Olaizola et al. [24] called DITEC to extract efficient global image descriptors. In order to transform from image space into a parameter space, the mentioned method uses the Trace transform, which is a generalization of the Radon transform, with the change that it sweeps the integral function in (2) with one of the functional Ξ defined in [25].

$$R(\phi,\rho) = \int\int f(x,y)\delta(x\cos\phi + y\sin\phi - \rho)\,dx\,dy \quad (1)$$

The chosen functionals are computed along lines moved tangentially to a circle radius that intersect the image space (represented in polar coordinates). ρ and ϕ are the parameters portrayed by each line with ρ being the distance from the origin of the image space to the line (the circle radius) and ϕ can take values up to 2π. While intersecting lines l along the ranges of ρ and ϕ , with the plane S, a functional T can be applied. After applying the transformation on an image, the result is a 2D signal made of sinusoidal functions that encode the original image depending on the chosen functional, thus different trace transforms can be obtained from the same image using different functionals. Radon transform begun time ago to be used in image fingerprinting [27]. Trace transform was use to create hash codes for image fingerprinting used in the MPEG-7 standard specification [17]. Also, it was successfully applied in face recognition task, [11], [15], character recognition ones [18] and sign recognition [34].

After applying DITEC the obtained system workflow is represented in Fig. 3. In order to obtain the image descriptors and knowing that the implementation of the trace transform is allowing to choose which functional to apply, we choose the functional number three (3), as explained in Srisuk et al. [30].

$$T(f(t)) = \int_c^\infty (t-c)^2 f(t)\,dt \quad (2)$$

$$c = \frac{1}{S}\int_0^\infty t|f(t)|\,dt \quad (3)$$

$$S = int_0^\infty |f(t)|\,dt \quad (4)$$

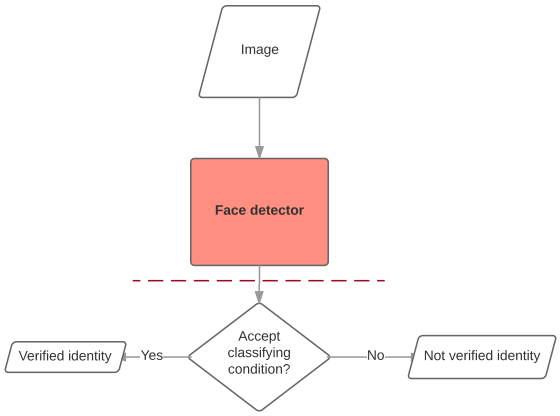

Fig. 1. Face detector based on local descriptors

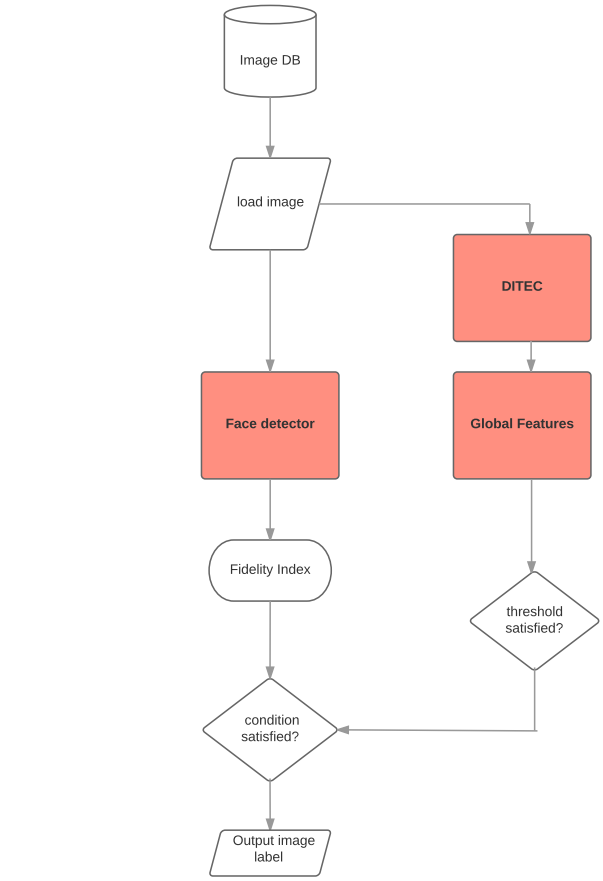

Fig. 2. Framework

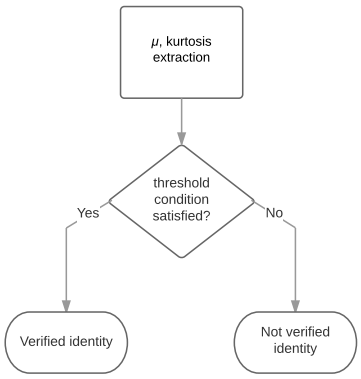

Fig. 3. System workflow

After extracting the descriptors, we want to reduce the dimensionality in order to improve the efficiency of the similarity measure process by applying principal component analysis. PCA has shown to be a very useful statistical technique with applications in different areas such as image compression and face detection and recognition.

We will discuss further each step of the application.

Step 1 Dataset preparation. The images are passed through DITEC. Here, the global descriptors are extracted, the mean and the kurtosis of the elements located orthogonally to the principal diagonal of every transformed image.

Step 2 These vectors are processed through PCA, which gives us the option to work in a lower dimensionality space, through removing the eigenvectors with the lowest eigenvalues. We create a cluster, Fig. 4, with a group of vectors and then we go to the next step.

Step 3 We learn multiple distance metrics between the rest of the vectors and the cluster. The computed distances are the Mahalanobis, Cosine and Correlation ones.

Step 4 The values obtained, in forms of distances, are thresholded to determine if the new vectors should be contained in that cluster or not. The optimal threshold depends on the similarity measure applied and on the dimensionality of the vectors.

Let us denote each image as v = $\{v_{m1}, v_{k1}, \ldots, v_{n1}\}$, with $v_{m1}, v_{k1}, \ldots, v_{n1}$ being the mean and kurtosis of every line of the principal diagonal of the images. Given the initial dataset, we form a cluster with a pair of samples. With the descriptors extracted above we have to decide if a new vector, that is not included in the cluster, comes from the same distribution, by calculating the distance between the vector itself and the cluster. We propose a multiple distance learning method, combining the Mahalanobis distance and the Cosine distance. If the distance d is smaller than a certain threshold, then, the compared vectors are from the same subject.

The Cosine Similarity c_S is defined as:

$$c_S(x, y) = \frac{x^T y}{||x||||y||}, \tag{5}$$

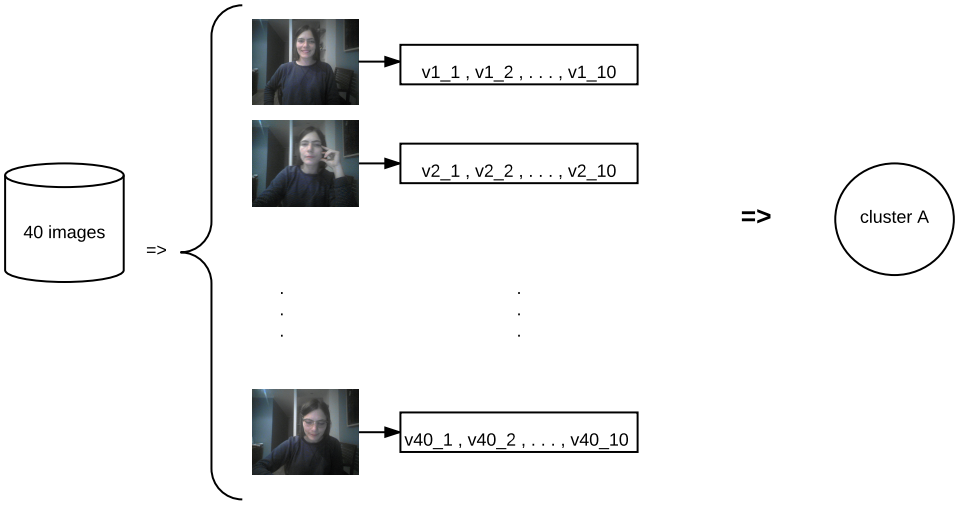

Fig. 4. Cluster build

where x and y are two row vectors. The usefulness of applying this similarity measure comes from its special property that its values are always between -1 and +1.

Being v a multidimensional vector, the Mahalanobis distance d_M is used to measure the divagation of the vector from the mean of a distribution represented by its covariance structure. It is defined as:

$$d_M = \sqrt{(X-\mu)' \cdot \sum^{-1} \cdot (X-\mu)} \tag{6}$$

where $\sum$ is the covariance matrix among the sample vectors and $\sum^{-1}$ is the inverse matrix of it. The Mahalanobis distance takes into account the correlation between variables. The shape of the distances can be an ellipse in a 2-dimensional space or, if more than two variables are used, an ellipsoid or hyperellipsoid. Firstly, we compute the covariance matrix of the sample vectors that forms the cluster, and use this to define a vector space in which the Mahalanobis distance can be employed.

In order to analyze how far a test vector is from the target group of data it is relevant to look at the cosine similarity to see the underlying structure of the samples and taking into account the dimension of the data, thus the associated covariance structure. There is high probability we can re-observe feature in the elliptical zone formed when employing the covariance matrix.

IV. Experimental results

Our goal is not to solve a classification problem, instead we verify the hypotheses of element 'a' belonging to the set (cluster) "A". In [13], Irigoien et al., considered implementing one-class classification (OCC) based on a typicality test to decide if a new observation belongs to a target class.

We analyze and validate our theoretical method by going through several experiments on a real dataset, for which we have ground truth models. The dataset contains 1216 images of faces collected in the same environment. It is composed of 15 different people with approximately 80 images each, with variability in face pose and expression. The attributes after applying DITEC have the length 274. For a better understanding of the feature space, we apply a Gaussian Naive Bayes classifier (in a cross validation test) and plot the accuracy regarding the number of principal components in Fig. 6.

Here we choose to work with 10 dimensions. Modifying the number of dimensions will influence the further threshold value.

In Fig. 5 we plot the information provided by the Bhattacharyya distance in order to see the qualitative behaviour of the method. We use the OpenOrd layout from Gephi [4], an interactive visualization platform, to envision in a 2D plane the different people. This layout is based on a force directed layout where the nodes repulse each other while the edges are ruled by an attraction force. Also, the width of each edge represents a classification of the distances from a node to all the other nodes for which the distance value is direct proportional with the width. Looking closely, it can be observed that node M for example has small distances towards other nodes, while node N is further. From here, we can obtain an approximation about the distance among different classes and we can make an idea on where we might encounter false positive values when we apply the threshold.

The main property of DITEC is being able to describe an image extracting the most robust features for the interpretation of the content. Our face verification algorithm has access to a pair of input images and needs to determine whether or not the image test pair come from the same person. We evaluate the performance of the algorithm in a one versus all

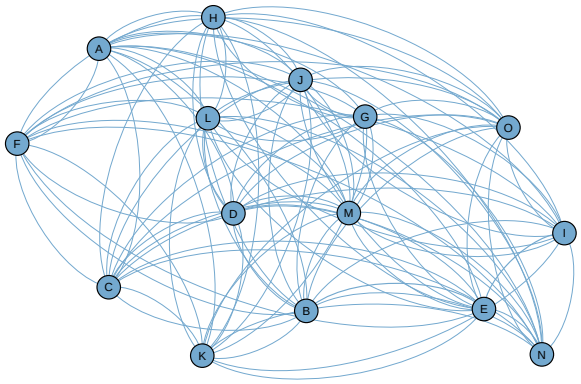

Fig. 5. Bhattacharyya distance

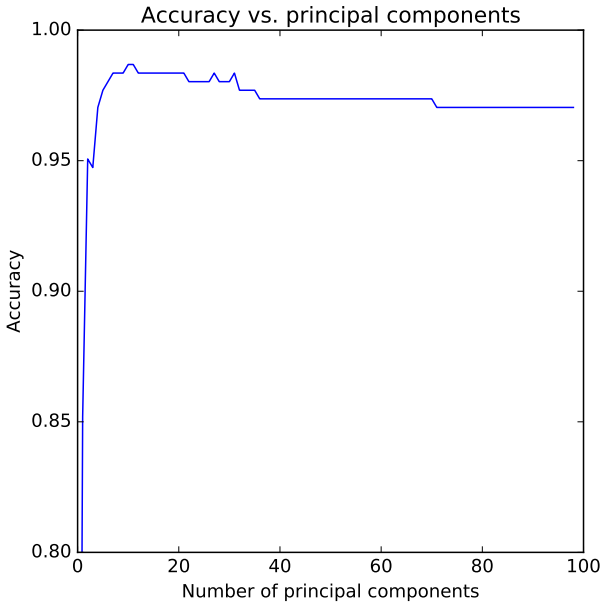

Fig. 6. Accuracy

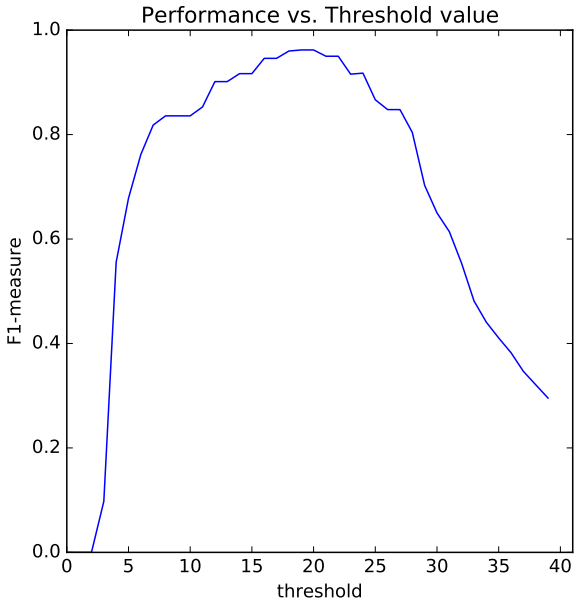

Fig. 7. F1 measure

cross validation procedure. We choose 40 faces of the same person to create the covariance matrix for the Mahalanobis distance and for the centroid.

The F1 measure, the harmonic mean of precision and recall, is represented in Fig. 7. We run through different threshold values and in the end, we confirm that the threshold value we use is in the area where the F1 measure is above 0.95%. Fig. 8 groups the asserting rate of the images evaluated for every person in the dataset. The results are improved when new true positive (TP) images of one person are added, one by one, to the cluster. Also, the algorithm implemented is a parameterized one in such a way that one can choose the dimension of the global descriptors to work with and where to set the threshold regarding of how important are the precision and recall values.

V. Conclusions and Future Work

We have presented an approach for analyzing global image descriptors in order to verify the hypotheses that an element "a" belongs to a cluster "A", relying on different similarity measures. We propose a method that combines state of the art face detection and recognition methods (based on local features) with global features that analyze the scene coherence to reinforce the local feature analysis results. We have evaluated the behavior with a real dataset for which we have the ground truth model. The dimensionality reduction results has shown that the set of features is very robust and have discriminative power, being able to verify the belonging of a person to an historic group of samples. Using the Mahalanobis distance and the Cosine similarity among the

	Same person (%)	Different people (%)
P1	0.70	1.0
P2	1.0	1.0
P3	0.97	1.0
P4	0.85	1.0
P5	1.0	1.0
P6	1.0	1.0
P7	1.0	1.0
P8	0.97	1.0
P9	1.0	1.0
P10	0.55	1.0
P11	0.82	1.0
P12	0.97	1.0
P13	0.90	1.0
P14	0.37	1.0
P15	0.92	1.0
Mean	**0.87**	**1.0**

Fig. 8. Asserting rate with regard to the same person or different ones

dataset gives good results taking into account the proposed solution. With regard to future work, we plan to setup a framework with local descriptors based on Trace Transform, through binding state-of-the-art interest point detectors.

VI. Acknowledgments

This research work has been done in a strong collaboration with Smowltech, a company specialized in online authentication. We want to show our gratitude to Mikel Labayen (CTO of Smowltech) for his support in providing the experimental dataset and real business contraints to define proper scientific metrics.

References

[1] Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." Pattern Analysis and Machine Intelligence, IEEE Transactions on 28.12 (2006): 2037-2041.

[2] Ahonen, Timo, Abdenour Hadid, and Matti Pietikäinen. "Face recognition with local binary patterns." Computer vision-eccv 2004. Springer Berlin Heidelberg, 2004. 469-481.

[3] Asthana, Akshay, et al. "Robust discriminative response map fitting with constrained local models." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.

[4] Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

[5] Bay, Herbert, et al. "Speeded-up robust features (SURF)." Computer vision and image understanding 110.3 (2008): 346-359.

[6] Bouker, Mohamed Ali, and Eric Hervet. "Retrieval of Images Using Mean-Shift and Gaussian Mixtures Based on Weighted Color Histograms." Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on. IEEE, 2011.

[7] Brasnett, Paul, and Miroslaw Bober. "Fast and robust image identification." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.

[8] Cao, Zhimin, et al. "Face recognition with learning-based descriptor." Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010.

[9] Cerra, Daniele, et al. "Algorithmic information theory based analysis of earth observation images: An assessment." IEEE Geoscience and Remote Sensing Letters 7.1 (2010): 8-12.

[10] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.

[11] Fahmy, Suhaib A. "Investigating Trace Transform Architectures for Face Authentication." FPL. 2006.

[12] Huang, Gary B., Honglak Lee, and Erik Learned-Miller. "Learning hierarchical representations for face verification with convolutional deep belief networks." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.

[13] Irigoien, I., Sierra, B. and Arenas, C., 2014. Towards Application of One-Class Classification Methods to Medical Data. The Scientific World Journal, 2014.

[14] Liu, Chengjun, and Harry Wechsler. "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition." Image processing, IEEE Transactions on 11.4 (2002): 467-476.

[15] Liu, Nan, and Han Wang. "Modeling images with multiple trace transforms for pattern analysis." Signal Processing Letters, IEEE 16.5 (2009): 394-397.

[16] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91-110.

[17] MPEG-7 Overview, ISO/IEC JTC1/SC29/WG11 Std., Rev. 10, October 2004. [Online]. Available: http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

[18] Nasrudin, Mohammad F., Maria Petrou, and Leonidas Kotoulas. "Jawi character recognition using the trace transform." Computer Graphics, Imaging and Visualization (CGIV), 2010 Seventh International Conference on. IEEE, 2010.

[19] Nemmour, Hassiba, and Youcef Chibani. "Handwritten Arabic word recognition based on Ridgelet transform and support vector machines." High Performance Computing and Simulation (HPCS), 2011 International Conference on. IEEE, 2011.

[20] Nguyen, Hieu V., and Li Bai. "Cosine similarity metric learning for face verification." Computer Vision–ACCV 2010. Springer Berlin Heidelberg, 2010. 709-720.

[21] Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.7 (2002): 971-987.

[22] Olaizola, Igor G., et al. "Architecture for semi-automatic multimedia analysis by hypothesis reinforcement." Broadband Multimedia Systems and Broadcasting, 2009. BMSB'09. IEEE International Symposium on. IEEE, 2009.

[23] Olaizola, Igor G., et al. "DITEC-Experimental Analysis of an Image Characterization Method based on the Trace Transform." VISAPP (1). 2013.

[24] Olaizola, Igor G., et al. "Trace transform based method for color image domain identification." Multimedia, IEEE Transactions on 16.3 (2014): 679-685.

[25] Petrou, Maria, and Alexander Kadyrov. "Affine invariant features from the trace transform." Pattern Analysis and Machine Intelligence, IEEE Transactions on 26.1 (2004): 30-44.

[26] Sengupta, Soumyadip, et al. "Frontal to profile face verification in the wild." 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016.

[27] Seo, Jin S., et al. "A robust image fingerprinting system using the Radon transform." Signal Processing: Image Communication 19.4 (2004): 325-339.

[28] Shechtman, Eli, and Michal Irani. "Matching local self-similarities across images and videos." Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007.
[29] Simonyan, Karen, et al. "Fisher Vector Faces in the Wild." BMVC. Vol. 5. No. 6. 2013.
[30] Srisuk, Sanun, et al. "Face authentication using the trace transform." Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 1. IEEE, 2003.
[31] Tola, Engin, Vincent Lepetit, and Pascal Fua. "Daisy: An efficient dense descriptor applied to wide-baseline stereo." Pattern Analysis and Machine Intelligence, IEEE Transactions on 32.5 (2010): 815-830.
[32] Torralba, Antonio, Kevin P. Murphy, and William T. Freeman. "Using the forest to see the trees: exploiting context for visual object detection and localization." Communications of the ACM 53.3 (2010): 107-114.
[33] Torralba, Antonio, Rob Fergus, and Yair Weiss. "Small codes and large image databases for recognition." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008.
[34] Turán, Ján, et al. "Invariant image recognition experiment with trace transform." Telecommunications in Modern Satellite, Cable and Broadcasting Services, 2005. 7th International Conference on. Vol. 1. IEEE, 2005.
[35] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.
[36] Vukadinovic, Danijela, and Maja Pantic. "Fully automatic facial feature point detection using Gabor feature based boosted classifiers." Systems, Man and Cybernetics, 2005 IEEE International Conference on. Vol. 2. IEEE, 2005.
[37] Xiong, Xuehan, and Fernando Torre. "Supervised descent method and its applications to face alignment." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.
[38] Zhu, Qiang, et al. "Fast human detection using a cascade of histograms of oriented gradients." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006.
[39] Zhu, Xiangxin, and Deva Ramanan. "Face detection, pose estimation, and landmark localization in the wild." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.
[40] Zuo, Yuanyuan, and Bo Zhang. "Robust hierarchical framework for image classification via sparse representation." Tsinghua Science & Technology 16.1 (2011): 13-21.