# Periocular Region Classifiers

D.R. Ambika, K.R. Radhika, and D. Seshachalam

BMS College of Engineering, Bangalore-19, India
amusatish@gmail.com

**Abstract.** Biometrics is the science of establishing human identity based on the physical or behavioral traits of an individual such as face, iris, ear, hand geometry, finger print, gait, knuckle joints and conjunctival vasculature among others. The enormous attention drawn towards the ocular biometrics during the recent years has led to the exploration of newer traits such as the periocular region. With the preliminary exploration of the feasibility of periocular region to be used as an independent biometric trait or in combination of face/iris, research towards periocular region is currently gaining lot of prominence. Over the last few years many researchers have investigated various techniques of feature extraction and classification in the periocular region. This paper attempts to review a few of these classifier techniques useful for developing robust classification algorithms.

**Keywords:** classifiers, periocular region.

## 1 Introduction

With the exploration of periocular region as a useful biometric trait, periocular region is drawing lot of attention in research studies [1, 2, 14]. It is experimented that periocular region is one of the most discriminative feature in the human face. Periocular biometrics requires the analysis of periocular images for compliance to the security related applications. To enhance the research studies in this area, periocular databases such as FRGC (Facial Recognition Grand Challenge), FERET (Facial Recognition Technology), MBGC (Multiple Biometrics Grand Challenge) and UBIRIS.V2 collected at different spectral range, lighting conditions, pose variations and different distances are available. From these periocular images, the region of interest is procured using segmentation process and fed to the feature extractor algorithm. Feature extraction is a robust process involved to seek distinguishing features of texture, color or size that are invariant to irrelevant transformations of the image. A feature extractor yields a representation to characterize the image. Various feature extraction techniques such as Gradient Orientation Histogram (GOH), Local Binary Patterns (LBP) [2, 16], Gabor Filters, Color Histograms [17], Walsh and Laws' mask, DCT, DWT, Force Field Transform and SURF are explored in periocular biometric studies. The feature vectors provided by these feature extractors are used by the classifiers to assign the object to a category. The abstraction provided by the feature-vector representation enables the development of a largely domain independent theory of classification. The degree of difficulty of the classification

V.V. Das and J. Stephen (Eds.): CNC 2012, LNICST 108, pp. 297–300, 2012.

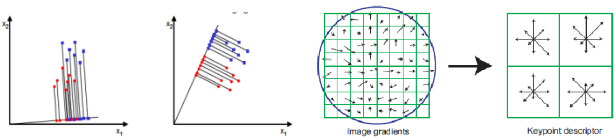
problem depends on the variability in the feature values for periocular images in the same category relative to the difference between feature values in different categories. The next section focuses on the different classification techniques.

# 2 Classification Techniques

Classifier analyzes the numerical properties of the image features and organizes it into categories. Classification algorithms typically employ two phases of processing: *training* and *testing*. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification category, *i.e. training class*, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features.

## 2.1 Different Classification Techniques

**Support Vector Machine (SVM).** SVM is a powerful learning machine extensively useful for binary classification. It intends to map the input vectors into a high dimensional feature space $Z$ through a non-linear mapping chosen a priori. A linear decision surface, known as the hyperplane (or a set of hyperplanes) is constructed in this space with special properties that ensure generalization ability of the network. Intuitively, a good separation is achieved by the optimal hyperplane that has the largest distance to the nearest training data points of any class, since in general the larger the margin the lower the generalization error of the classifier. An optimal hyperplane is defined as the linear decision function with maximal margin between the vectors of the two classes [13].




LDA hyperplane SIFT Descriptor

**Fig. 1.** (Left) A scalar y is obtained by projecting all the samples of x onto a line $y=w^T x$ and select that line which maximizes the separability of the scalars as the hyper plane for LDA. SIFT keypoint descriptor (Right) is created by computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location, as shown on the left. These are weighted by a Guassian window which is indicated by the overlaid circle. The samples are then accumulated to form orientation histograms summarizing as shown on the right. The length of each arrow corresponds to the sum of the gradient magnitudes near that direction within the region.

**Scale Invariant Feature Transform (SIFT). SIFT** transforms an image into a large collection of feature vectors as shown in figure 1 (right), each of which is invariant to image translation, scaling, and rotation, partially invariant to illumination changes and robust to local geometric distortion. These SIFT features are extracted using Difference of Guassian functions from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to the database and determining candidate matching features. The best candidate match for each keypoint is obtained by identifying its nearest neighbor in the database of keypoints from training images. The nearest neighbor is that keypoint with minimum Euclidean distance for the invariant descriptor vector [4, 5].

**Linear Discriminant Analysis (LDA).** LDA searches for the vectors in the underlying space of independent data feature that best discriminate among classes (rather than those that best describe the data itself). It projects data on a hyperplane that minimizes the *within-class* scatter and maximizes the *between-class* scatter as shown in the figure 1. Mathematically, these two measures are defined as *within-class* scatter matrix, and *between-class* scatter matrix given by the equations 2 and 3 [8]

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \tag{2}$$

where $x_i^j$ is the *i*th sample of the class *j*, $\mu_j$ is the mean of class *j*, *c* is the number of classes and $N_j$ is the number of samples in class *j*.

$$S_b = \sum_{j=1}^{c} (\mu_j - \mu)(\mu_j - \mu)^T \tag{3}$$

where $\mu$ represents the mean of all classes.

**Principle Component Analysis (PCA).** PCA is a standard technique used to approximate the original data with lower dimensional feature vectors. It is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possible correlated variables into a set of uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has as high a variance as possible and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to the preceding components [6, 7]. In principle, common properties implicitly existing in the training set; like gender, race, age and usage of glasses can be observed from these components.

**Multilayer Perceptron (MLP).** MLP is a feedforward (no recurrent nodes) network, which maps a set of input data onto a set of output through multiple layers of nodes in a directed graph . Each node (except for the input node) is a neuron with a nonlinear activation function. MLP is trained through the backpropogation algorithm (BP). The input vector $\boldsymbol{x^0}$ is transformed to the output vector $\boldsymbol{x^L}$. The difference between the desired output *d* and actual output $\boldsymbol{x^L}$ is computed as the error signal and is propagated backwards through the entire network by updating the synaptic weights W and biases b. This updating yields the actual output $\boldsymbol{x^L}$ closer to the desired output *d* [9].

**JointBoost Algorithm.** The idea of this algorithm is that at each boosting round of a classification algorithm (C) such as AdaBoost, various subsets of classes, S ⊆ C are examined and considered to fit a weak classifier such that this subset is distinguished from the background. The subset is picked up such that it maximally reduces the error on the weighted training set for *all* the classes. The best weak learner h(v, c) is then added to the strong learners H(v, c) for all the classes c ∈ S, and their weight distributions are updated so as to optimize the multiclass cost function $J = \sum_{c=1}^{C} E[e^{-z^c H(v,c)}]$ , $z^c$ is the membership label (±1) for class c. [11].

**Probabilistic Boosting Tree (PBT).** PBT is a learning framework proposed for learning two-class and multi-class discriminative models. It constructs a tree in which each node combines a number of weak classifiers into a strong classifier. The conditional probability is computed at each tree node based on the learned classifier, which guides the probability propagation in its sub-tree. The top node of the tree therefore outputs the overall posterior probability by integrating the probabilities gathered from its sub-trees [12].

# 3 Conclusion

This work presents a review of various classifier schemes suitable for categorizing the identity of the claimed, using the periocular region. It investigates different classifier schemes such as independent learning machines and fusion of classifiers which form boosting algorithms to aid in boosting the performance of weak classifiers.

# References

1. Park, U., Jillela, Ross, Jain, A.K.: Periocular Biometrics in the Visible Spectrum. IEEE Transactions on Information Forensics and Security 6(1) (March 2011)
2. Park, U., Ross, A., Jain, A.K.: Periocular biometrics in the visibility spectrum: A feasibility study. In: Proc. Biometrics: Theory, Applications and Systems (BTAS), pp. 153–158 (2009)
3. Lyle, J.R., Miller, P., Pundlik, S., Woodard, D.: Soft Biometric Classification using Periocular Region Features. IEEE Transactions (2010)
4. Distinctive Image Features from Scale- Invariant Keypoints
5. Hollingsworth, K., Bowyer, K.W., Flynn, P.J.: Identifying useful features for recognition in Near Infrared Periocular Images. IEEE Transactions (2010)
6. Zhao, W., Arvindh, Rama, Daniel, John: Discriminant Analysis of Principal components for face recognition
7. Koray, Volkan: PCA for gender estimation: Which Eigenvectors contribute? In: ICPR 2002 (2002)
8. Aleix, M.: Avinash: PCA versus LDA. IEEE Transactions on PAMI (2001)
9. Seung, S.: Multilayer perceptrons and backpropogation learning (2002)
10. Jerome, Trevor, Robert: Additive Logistic Regression: A statistical view of boosting. The Annals of Statistics (2000)
11. Antonio, Kevin, William: Sharing features: efficient boosting procedures for multiclass object detection
12. Zhuowen: Probabilistic Boosting Tree: Learning Discriminative Models for Classification. Recognition and Clustering
13. Corinna, Vladimir: Support Vector Networks. Machine Learning (1995)
14. Woodard, D., Pundlik, S., Miller, P., Lyle, J.R.: Appearance-based periocular features in the context of face and non-ideal iris recognition. Springer
15. Merkow, J., Jou, B., Savvides, M.: An Exploration of Gender Identification using only the periocular region. IEEE Transactions (2010)
16. Merkow, J., Jou, B., Savvides, M.: An Exploration of Gender Identification using only the periocular region. IEEE Transactions (2010)
17. Woodard, D., Pundlik, S., Lyle, J.R., Miller, P.: Periocular Region AppearanceCues for Biometric Identification. IEEE Transactions (2010)