Face Recognition Using Various Machine Learning Methods

CH Sudheer [1], V. Madhuri [2]
[1]Dept of ECE
[2]Assistant Professor, Dept of ECE
[1, 2] Indira Institute of Technology and Sciences, MARKAPUR

Abstract- *Research on color face recognition in the existing literature is aimed to establish a color space that can have the most of the discriminative information from the original data. This mainly includes optimal combination of different color components from the original color space. Recently proposed discriminate color space (DCS) is theoretically optimal for classification, in which one seeks a set of optimal coefficients in terms of linear combinations of the R, G and B components (based on a discriminate criterion). This work proposes an innovative block-wise DCS (BWDCS) method, which allows each block of the image to be in a distinct DCS. This is an interesting alternative to the methods relying on converting whole image to DCS. This idea is evaluated with four appearance-based subspace state-of- the-art methods on five different publicly available databases including the well-known FERET and FRGC databases. Experimental results show that the performance of these four gray-scale based methods can be improved by 17% on average when they are used with the proposed color space.*

Keywords- unsupervised methods, supervised methods and Euclidean distance measure

I. INTRODUCTION

Machine learning is a field of software engineering those utilizations factual strategies to enable PC frameworks to "learn" (i.e., logically enhance execution on a particular assignment) with information, without being expressly programmed. The name machine learning was instituted in 1959 by Arthur Samuel. Evolved from the investigation of example acknowledgment and computational learning hypothesis in simulated intelligence, machine learning investigates the examination and development of calculations that can gain from and make expectations on data – such calculations conquer following entirely static program guidelines by making information driven forecasts or decisions through building a model from test inputs. Machine learning is utilized in a scope of processing assignments where outlining and programming unequivocal calculations with great execution is troublesome or infeasible; illustration applications incorporate email sifting, location of system gatecrashers or noxious insiders working towards an information breach, optical character acknowledgment (OCR), figuring out how to rank, and PC vision. Machine learning is firmly identified with (and regularly covers with) computational measurements, which likewise centers around forecast making using PCs. It has solid connections to scientific improvement, which conveys techniques, hypothesis and application spaces to the field. Machine learning is now and again conflated with information mining, where the last subfield concentrates more on exploratory information investigation and is known as unsupervised learning. Machine learning can likewise be unsupervised and be utilized to learn and build up gauge behavioral profiles for different entities and afterward used to discover significant oddities. Inside the field of information investigation, machine learning is a strategy used to devise complex models and calculations that loan themselves to expectation; in business utilize, this is known as prescient examination. These expository models permit specialists, information researchers, designers, and examiners to "create dependable, repeatable choices and comes about" and reveal "shrouded bits of knowledge" through gaining from recorded connections and patterns in the data. Powerful machine learning is troublesome in light of the fact that discovering designs is hard and regularly insufficient preparing information are accessible; therefore, machine-learning programs frequently neglect to convey. Machine learning undertakings are regularly ordered into two general classes, contingent upon whether there is a learning "flag" or "criticism" accessible to a learning framework. Regulated taking in: The PC is given illustration inputs and their coveted yields, given by an "instructor", and the objective is to take in a general decide that maps contributions to yields. As exceptional cases, the info flag can be just mostly accessible, or confined to unique criticism: Semi-directed taking in: the PC is given just an inadequate preparing signal: a preparation set with a few (regularly many) of the objective yields missing. Dynamic taking in: the PC can just get preparing names for a restricted arrangement of occasions (in view of a financial plan), and furthermore needs to improve its selection of articles to procure marks for. At the point when utilized intelligently, these can be exhibited to the client for marking. Fortification picking up: preparing information (in type of prizes and disciplines) is given just as criticism to the program's activities in a dynamic situation, for example, driving a vehicle or playing an amusement against an opponent. Unsupervised learning: No marks are given to the learning calculation, abandoning it all alone to discover structure in its information. Unsupervised learning can be an objective in itself (finding concealed examples in information) or methods towards an end (include learning). Another order of machine learning undertakings emerges when one considers the coveted yield of a machine-learned system. In characterization, inputs are separated into at least two classes, and the student must create a model that appoints concealed contributions to at least one (multi-name arrangement) of these classes. This is regularly handled supervised. Spam sifting is a case of arrangement, where the sources of info are email (or other) messages and the classes are "spam" and "not spam". In relapse, additionally a regulated issue, the yields are constant as opposed to discrete. In bunching, an arrangement of data sources is to be partitioned into gatherings. Dissimilar to in order, the gatherings are not known already, making this regularly an unsupervised errand. Thickness estimation finds the dispersion of contributions to some space. Dimensionality lessening rearranges contributions by mapping them into a lower-dimensional space. Theme demonstrating is a related issue, where a program is given a rundown of human dialect reports and is entrusted to discover which archives cover comparative subjects. Among different classifications of machine learning issues, figuring out how to learn takes in its own particular inductive inclination in view of past involvement. Formative learning, explained for robot learning, produces its own particular arrangements (likewise called educational programs) of learning circumstances to aggregately secure collections of novel abilities through self-sufficient self-investigation and social association with human educators and utilizing direction instruments, for example, dynamic learning, development, engine cooperative energies, and impersonation.

II. SUPERVISED LEARNING

i) Naive Bayes theorem:

Methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of independence between every pair of features. Given a class variable y and a dependent feature vector x_1 through x_n, Bayes' theorem states the following relationship:

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) P(x_1, \dots x_n \mid y)}{P(x_1, \dots, x_n)}$$

Using the naive independence assumption that

$$P(x_i | y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i | y),$$

for all i, this relationship is simplified to

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i \mid y)}{P(x_1, \dots, x_n)}$$

Since $P(x_1, \dots, x_n)$ is constant given the input, we can use the following classification rule:

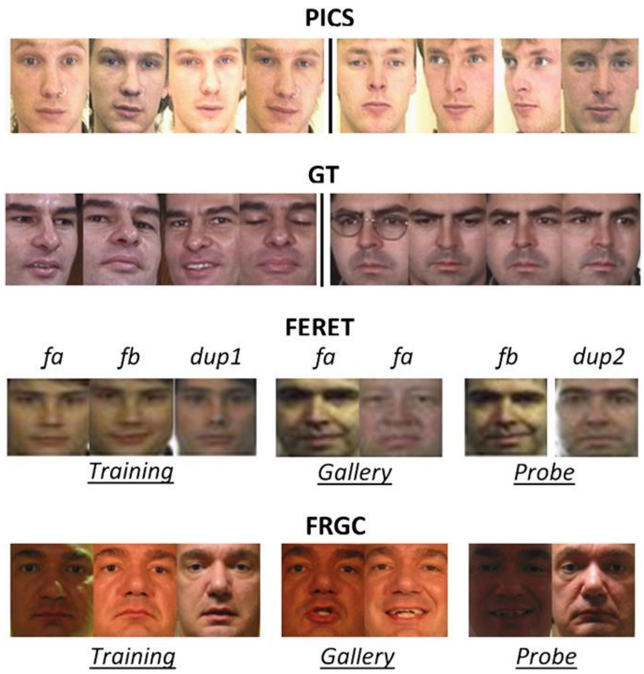
$$P(y \mid x_1, \dots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i \mid y)$$

$$\Downarrow$$

$$\hat{y} = \arg\max_y P(y) \prod_{i=1}^{n} P(x_i \mid y),$$

and we can use Maximum A Posteriori (MAP) estimation to estimate $P(y)$ and $P(x_i \mid y)$; the former is then the relative frequency of class y in the training set.

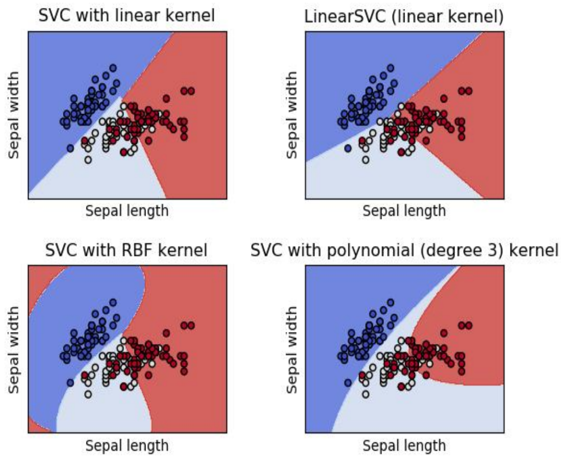
The different naive Bayes classifiers differ mainly by the assumptions they make regarding the distribution of $P(x_i \mid y)$. Regardless of their clearly finished streamlined suppositions, guileless Bayes classifiers have worked great in some certifiable circumstances, broadly report characterization and spam separating. They require a little measure of preparing information to appraise the vital parameters. (For hypothetical reasons why credulous Bayes functions admirably, and on which sorts of information it does, see the references underneath.) Naive Bayes students and classifiers can be to a great degree quick contrasted with more modern techniques. The decoupling of the class restrictive component appropriations implies that every dissemination can be freely evaluated as a one dimensional circulation. This thus lightens issues coming from the scourge of dimensionality. On the other side, albeit credulous Bayes is known as a tolerable classifier, it is known to be a terrible estimator, so the likelihood yields from predict_proba are not to be considered excessively important. Expanded example vector, which contains the reduced data of the R, G, Br picture, is additionally prepared to separate the discriminative shading highlights for confront acknowledgment. One mainstream highlight extraction strategy is discriminant investigation, which enhances the criteria of class detachability in light of disseminate grids. Discriminant investigation, in any case, regularly shows poor speculation execution due to over fitting. To break down how finished fitting happens in discriminant examination and how to conquer it to improve its speculation capacities, we now quickly survey the synchronous diagonalization of the inside class and between-class scramble lattices,



ii) Support vector machine

The benefits of help vector machines are:

Viable in high dimensional spaces. Still viable in situations where number of measurements is more prominent than the quantity of tests. Utilizations a subset of preparing focuses in the choice capacity (called bolster vectors), so it is additionally memory effective. Flexible: distinctive Kernel capacities can be determined for the choice capacity. Basic pieces are given, however it is likewise conceivable to indicate custom bits. The disservices of help vector machines include: On the off chance that the quantity of highlights is considerably more prominent than the quantity of tests, keep away from over-fitting in picking Kernel capacities and regularization term is pivotal. SVMs don't straightforwardly give likelihood gauges, these are ascertained utilizing a costly five-overlap cross-approval.



III. UNSUPERVISED LEARNING

Unsupervised learning is the place you just have input information (X) and no relating yield factors. The objective for unsupervised learning is to show the fundamental structure or dissemination in the information so as to take in more about the information. These are called unsupervised learning in light of the fact that not at all like regulated learning above there is no right answers and there is no instructor. Calculations are left to their own devises to find and present the fascinating structure in the information. Unsupervised learning issues can be additionally assembled into grouping and affiliation issues. Bunching: A grouping issue is the place you need to find the characteristic groupings in the information, for example, gathering clients by obtaining conduct. Affiliation: An affiliation manage learning issue is the place you need to find decides that portray substantial segments of your information, for example, individuals that purchase X likewise tend to purchase Y.

Proposed method has implemented below

1. Each image is decomposed as sub band.
2. Sub band is resized to the original image size.
3. Each resized image is partitioned into sub images.
4. Convert the each sub image into column data matrix. Each of them can be expressed in the order of a D-by-N. Ci = {ci1+ci2+ci3+…ciN} with i = 1, 2,…….. K. here N is the total number of images
5. Calculate mean value for each sub image.
6. Subtract the mean value from column data matrix of each sub image then obtain vertically centered column data matrix Cvi = {ĉi1+ĉi2+ĉi3+…ĉiN},i = 1, 2,……...K.
7. Rearrange the elements to get square matrix.
8. Collect Eigen values, Eigen vectors, and diagonal values of the square matrix
9. Repeat the same procedure for row data matrix.

10. Reduce the feature size as per the requirement as feature 1.
11. above steps are repeated for the whole image without generate the feature 2.
12. Feature 1 and 2 are combined to get the global feature.
13. Minkowski distance is used to retrieve the relevant images.
14. Minkowski distance is concentrated on Euclidean space, which can be considered as a generalization of both Euclidean and Manhattan distance for getting more recognition efficiency.

IV. EXPERIMENTAL RESULTS

In mathematics, the Euclidean distance or Euclidean metric is the "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula. By using this formula as distance, Euclidean space (or even any inner product space) becomes a metric space. The associated norm is called the Euclidean norm. Older literature refers to the metric as Pythagorean metric.

Definition:

The Euclidean distance between points p and q is the length of the line segment connecting them (pq).

In Cartesian coordinates, if $p = (p_1, p_2,..., p_n)$ and $q = (q_1, q_2,..., q_n)$ are two points in Euclidean n-space, then the distance from p to q, or from q to p is given by:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2} = \sqrt{\sum_{i=1}^{n}(q} \quad (1)$$

The position of a point in a Euclidean n-space is a Euclidean vector. So, p and q are Euclidean vectors, starting from the origin of the space, and their tips indicate two points. The Euclidean norm, or Euclidean length, or magnitude of a vector measures the length of the vector:

$$\|\mathbf{p}\| = \sqrt{p_1^2 + p_2^2 + \cdots + p_n^2} = \sqrt{\mathbf{p}\cdot\mathbf{p}}$$

where the last equation involves the dot product.

A vector can be described as a directed line segment from the origin of the Euclidean space (vector tail), to a point in that space (vector tip). If we consider that its length is actually the distance from its tail to its tip, it becomes clear that the Euclidean norm of a vector is just a special case of Euclidean distance: the Euclidean distance between its tail and its tip.

The distance between points p and q may have a direction (e.g. from p to q), so it may be represented by another vector, given by

$$\mathbf{q} - \mathbf{p} = (q_1 - p_1, q_2 - p_2, \cdots, q_n - p_n)$$

In a three-dimensional space (n=3), this is an arrow from p to q, which can be also regarded as the position of q relative to p. It may be also called a displacement vector if p and q represent two positions of the same point at two successive instants of time.

The Euclidean distance between p and q is just the Euclidean length of this distance (or displacement) vector:

$$\|\mathbf{q} - \mathbf{p}\| = \sqrt{(\mathbf{q}-\mathbf{p})\cdot(\mathbf{q}-\mathbf{p})}.$$

Which is equivalent to equation 1, and also to:

$$\|\mathbf{q} - \mathbf{p}\| = \sqrt{\|\mathbf{p}\|^2 + \|\mathbf{q}\|^2 - 2\mathbf{p}\cdot\mathbf{q}}.$$

One dimension:

In one dimension, the distance between two points on the real line is the absolute value of their numerical difference. Thus if x and y are two points on the real line, then the distance between them is computed as

$$\sqrt{(x-y)^2} = |x-y|.$$

In one dimension, there is a single homogeneous, translation-invariant metric (in other words, a distance that is induced by a norm), up to a scale factor of length, which is the Euclidean distance. In higher dimensions there are other possible norms.

Two dimensions:

In the Euclidean plane, if $\mathbf{p} = (p_1, p_2)$ and $\mathbf{q} = (q_1, q_2)$ then the distance is given by

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \sqrt{(p_1-q_1)^2 + (p_2-q_2)^2}.$$

Alternatively, it follows from (**2**) that if the polar coordinates of the point **p** are (r_1, θ_1) and those of **q** are (r_2, θ_2), then the distance between the points is

$$\sqrt{r_1^2 + r_2^2 - 2r_1r_2\cos(\theta_1 - \theta_2)}.$$

Three dimensions:

In three-dimensional Euclidean space, the distance is

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

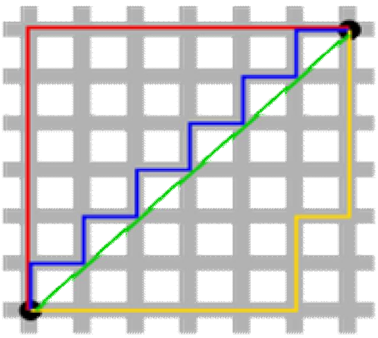
In general, for an n-dimensional space, the distance is

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2}.$$

5.3 Squared Euclidean Distance:

You may want to square the standard Euclidean distance in order to place progressively greater weight on objects that are further apart. In this case, you just square the equation.

5.4 Taxicab geometry (Manhattan distance):



Taxicab geometry versus Euclidean distance: All the lines have the same length (12) in taxicab geometry for the same route. In Euclidean geometry, the green line has length $6\times\sqrt{2} \approx 8.48$, and is the unique shortest path.

Taxicab geometry, considered by Hermann Minkowski in the 19th century, is a form of geometry. Here, in comparison to Euclidean geometry, the usual distance function or metric is replaced by a new metric in which the distance between two points is the sum of the absolute differences of their coordinates. The taxicab metric is also known as rectilinear distance, L_1 distance or ℓ_1 norm (see L^p space), city block distance, Manhattan distance, or Manhattan length, with corresponding variations in the name of the geometry. The latter names allude to the grid layout of most streets on the island of Manhattan, which causes the shortest path a car could take between two points in the borough to have length equal to the points' distance in taxicab geometry.

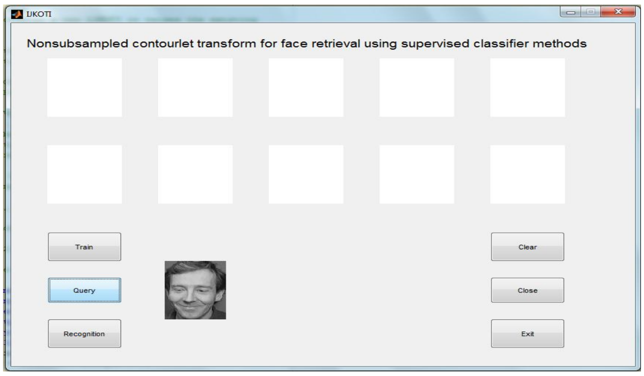


Figure 1: Test image

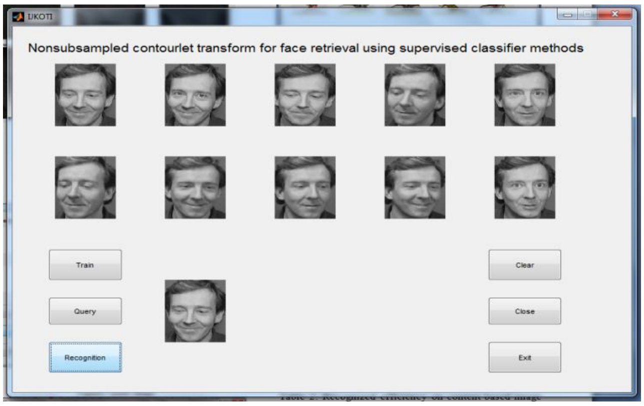


Figure 2: Retrieval images

Table1: Face recognition comparisons

Methods	Number of top matches				
	1	3	5	7	10
K-NN	100	70	63.5	60.5	55
SVM	100	75	68	62	56
PCA	100	80	75	71.5	69.5
Hybrid method (PCA+LDA) PROPOSED	**100**	**92**	**82**	**80**	**76.5**

V. CONCLUSIONS

This paper has been successfully implemented in terms as Research on color face recognition in the existing literature is aimed to establish a color space that can have the most of the discriminative information from the original data. This mainly includes optimal combination of different color components from the original color space. Recently proposed discriminate color space (DCS) is theoretically optimal for classification, in which one seeks a set of optimal coefficients in terms of linear combinations of the R, G and B components (based on a discriminate criterion). This work proposes an innovative block-wise DCS (BWDCS) method, which allows each block of the image to be in a distinct DCS. This is an interesting alternative to the methods relying on converting whole image to DCS. This idea is evaluated with four appearance-based subspace state-of- the-art methods on five different publicly available databases including the well-known FERET and FRGC databases. Experimental results show that the performance of these four gray-scale based methods can be improved by 17% on average when they are used with the proposed color space.

REFERENCES

[1] M. N. Do and M. Vetterli, "Contourlets," in Beyond Wavelets, G. V. Welland, Ed. New York: Academic Press, 2003.

[2] "The contourlet transform: an efficient directional multiresolution image representation," IEEE Trans. Image Proc., to appear, http://www.ifp.uiuc.edu/˜minhdo/publications

[3] M. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory, Special Issue on Wavelet Transforms and Multiresolution Signal Analysis*, vol. 38, no. 2, pp. 587–607, March 1992

[4] N. Do and M. Vetterli, "Pyramidal Directional Filter Banks and Curvelets", IEEE International conference on Image Processing, Vol.3, pp 158-161, Oct 2001

[5] R. H. Bamberger and M. J. T Smith, "A filter bank for the directional decomposition of images: Theory and design", IEEE Transactions on Signal Processing, Vol.40, pp. 882-893, April 1992.

[6] P. J Burt and E. H. Adelson, "The Laplacian Pyramid as a compact image code", IEEE Transactions on Communication, Vol.31, No.4, pp.532-540, April 1983

[7] See-May Phoong, Chai W.Kin, P.P. Vaidyanathan and Rahid Ansari, "A New Class of Two-channel Biortho.