Gray level face recognition using spatial features

Md Fawwad Hussain, Haidong Wang, K.C. Santosh**, *IEEE* Senior Member

Department of Computer Science
The University of South Dakota
414 E Clark St., Vermillion, SD 57069, USA

santosh.kc@ieee.org

Abstract. Face recognition has always been an active research area with several applications, such as security access control, human-machine interface and gender classification. More often, in real world, grayscale images have been used: video surveillance, for instance. Further, difficulties in face recognition could be due to face poses, orientation, lighting, aging etc. Faces, either in color or grayscale and are having any difficulties (as mentioned earlier) can be learned through edge map and texture, where spatial properties could be learned. Inspired from the fact that face can be considered as line-rich pattern/object, we propose novel face recognition framework that helps learn/recognize via spatial arrangements of edges (and textures as complement). To exploit edge map, we use shape context (SC) and pyramid histogram of orientated gradient (PHOG), and similarly GIST as texture features. Experimental tests (on four different publicly available datasets, such as Caltech, ColorFERET, IndianFaces and ORL) conforms that spatial features are crucial in face representation and recognition.

Keywords: Gray level images, spatial features, shape context, pyramid histogram of oriented gradients, gist, biometrics, face recognition.

1 Introduction

Automated face recognition system turns out to be one of the most successful applications in image processing, pattern recognition, and computer vision. There are many reasons for the growing demand in face recognition system that include rising concerns for public security, identity verification in high-security areas, face analysis and modeling techniques in digital entertainment. Research for face recognition started in 1960, and there has been a remarkable success in this area. Many face detection and recognition systems have been developed and a number of algorithms have been designed so far by various research scientists in different domains but still, there is a number of challenges present in making an accurate and a robust face detection and recognition system in real-world environments. For many applications, the performance of face recognition systems in controlled environments has now reached a satisfactory level, but there are still

** Corresponding author

many challenges presents when dealing with uncontrolled environments. Some of the challenges are caused by variations in illumination, face pose, expression, and etc. It's a true challenge to build an automated system which equals the human ability to recognize faces. Humans can recognize faces, but too many faces sometimes being hard to get. Now, machine learning is now being improved a lot to do this crucial task. The ability to make the machine thinks like humans and recognize faces in a limited amount of time and memory is the must.

1.1 Problems

Face pose is a specifically difficult problem in this aspect simply because all faces seem similar; specifically, all faces consist of two eyes, mouth, nose, and other features that are in the same location. As human face is a dynamic object having the high degree of variability in its appearance. Most face images captured by surveillance systems are non-ideal because they are often affected by many factors such as pose, illumination, expression, occlusion, distance, weather and so on. The human face is not a unique rigid object and each of the faces have a variety of deformations. Human face varies from person to person. The gender, age, emotions, expressions and other physical characteristics have to be taken into account thereby creating a challenge for the research scientist in computer vision.

1.2 Goal

The primary goal of our research is to build an automated face recognition systems that are fast, robust, accurate, with less computational overheads, relatively simple algorithms and techniques that would be easy to understand and to implement it on a large programmable system. To achieve higher efficiency in terms of storage and computation, grayscale images are used. The primary reasons behind the use of grayscale images can be summarized as follows:

a) Many commercial applications used to process grayscale images because of its low computational complexities and overheads.
b) Many Cameras in high-alert zones used grayscale images for different identification and verification. It is relatively easier to deal with (in terms of storage) a single color channel (shades of white/black) than color channels.
c) An application like object detection would barely require information at the edges of an image, which can as well be obtained in gray scale images (also can be obtained in color images, at an expense of complexity).

1.3 Contributions

Following the importance of the spatial features, three different features are used.

a) We start with the hape context (SC) is used to compute the shape similarity based on the features points extracted from internal or external contours of face image [5]. The technique is based on the low level (edge-based) representation of an image. Edge map can give us a better understanding of the human face features and covers all aspects of human face recognition. The approach is invariant to rotation and scale and consumes very low memory requirement.
b) We also employ an extremely rich global shape descriptor which has information from all over the shape in terms of histogram of edge orientations and spatial distribution of edges [6]. The approach is insensitive to small deformations and consumes low memory requirements as well.
c) Another idea based on the spatial layout is given by Oliva and Torralba [19]. The approach is based on the low level representation of the shape called the spatial envelope. The descriptor encodes the information based on the spatial frequencies and orientations that contains enough information to identify the shape.

2 Related works

2.1 Context

Many different approaches and techniques have been developed so far in the field of face matching and recognition for the reliable and efficient system. These techniques use different methods such as the appearance based method and model-based methods. Recent work in face matching uses appearance-based approach for fast and robust face recognition system where the image is considered as a high-dimensional vector. Modern appearance-based recognition system needs a rich descriptor for shape matching. Some descriptors use all image information as a fact that all pixels in the image are equally important which leads to high computational complexities and requires a high degree of correlation between the probe and gallery images. Such descriptors contain high inconsistencies and variations in pose, scale, and illuminations.

2.2 Relevant publications

Some descriptors use non-dense image information such as edges or various key spots for feature matching. They can handle small deformations and variations in pose, scale, orientation as well. Recent work by Sema et al [8] (inspired from previous work [20]) which uses a hybrid technique by processing descriptors that includes features from both categories. They present a rotation-and-scale-invariant, line-based color-aware (RSILC) descriptor which allows matching of objects in terms of their key-lines, line-region properties, and line spatial arrangements. This general purpose descriptor is color-aware, invariant to rotation/scale and is robust to illumination changes.

Another well known local key point descriptor scale invariant feature transform (SIFT) [18] is one of the most robust key-points descriptors among local feature descriptor. It first detects a number of key points in the images and then computes a local descriptor for each point. Recognition can be performed by matching each key point descriptor in the test image against the descriptor extracted from all images in the database. SIFT key point detector and descriptor have shown remarkable success in a number of applications including object recognition, face recognition, image stitching etc. But it contains a huge computational burden for real-time systems. Later, they introduced a new variant of SIFT that is color SIFT (CSIFT) [1] which also takes color information into the account. The color invariant SIFT is more robust as compared to conventional SIFT but still, it is not rotation and scale invariant. Based on the SIFT algorithm, Liu et al [17] propose a novel method SIFT flow to align an image to its nearest neighbors in a database. The SIFT flow algorithm consists of matching densely sampled, pixel-wise SIFT features between two images, while preserving spatial discontinuities. Another well-known interest point detector and descriptor is speeded up robust features (SURF) [4] which outperforms another state of the art descriptor due to its robustness and distinctiveness. The computed descriptor is also rotation and scale invariant and provides a quicker way to detect key points. It has been successfully used in many object and shape matching application including face matching. Analyzing the importance of SURF as a local descriptor, Dreuw et al [12] proposed an approach of using SURF algorithm to face recognition problem.

Local binary pattern(LBP) [2], which is originally introduced for texture representation, has proved to be a powerful descriptor for face recognition. This new approach considered both shape and texture information of face image. The face area is first divided into small regions from which Local Binary Pattern (LBP) histograms are extracted and concatenated into a single feature histogram, spatially enhanced feature histogram efficiently representing the face image. The similarity of images can be compared by measuring the distance between their histograms. Research shows that face recognition using the LBP method provides remarkable results, both in terms of speed and discrimination performance. This technique seems to be quite robust against face images with different facial expressions, different lighting conditions, image rotation and aging of persons.

As the human recognize line drawings well, the machine can also. Gao et al [14] describe a novel descriptor for face recognition using line edge map (LEM) which extracts features line segment from a face edge map. This approach is robust and efficient in retrieval capability as compared to Eigenface [21]. Deboeverie et al. [10] proposed an approach in which faces are represented as curve edge maps. CEM describes polynomial segments with a convex region in an edge map. With these concepts in mind, in this paper, for gray level face verification we follow the works by Serge Belongie [5], Bosch et al [6] and Oliva and Torralba [19]. Serge Belongie [5] presents a novel approach for measuring the similarity between shapes for object recognition. The key-points descriptor also contains global spatial relationship of other key-point descriptors relative to it. The shape context descriptor is tolerant to all common shape deformations. This novel approach is widely used in many shape matching applications. Besides, the work by

a) b)

Fig. 1. Output images from edge detection algorithm: a) Input Image; and b) Edge based representation using Canny.

Bosch et al [6] who proposed a novel descriptor based on histogram of oriented gradients called the pyramid histogram of orientation gradients (PHOG). This descriptor represents an image by its local shape and the spatial layout of shape information. The local shape is represented by a histogram of edge orientations. The spatial layout is represented by tiling the image into regions at multiple resolutions. These are designed so that the shape correspondence between two images can be measured by the distance between their descriptor. The descriptor is rotation dependent and color blind. Oliva and Torralba [19] proposed a method that represents an image in a low dimension space that contains enough information to identify the scene, which also bypasses the segmentation and processing of individual objects or regions. That low-level representation of the scene is called spatial envelope. They define the features that separate a scene from the rest. It was mainly used for scene recognition.

3 Spatial features for face representation and recognition

3.1 Shape context (SC)

The central idea behind this approach came from the point of how to compute shape similarity and find the difference between two shapes so that category-level recognition is possible.

The SC [5] has been useful in measuring the similarity between the studied shapes by recovering from point correspondences and exploiting it for shape-based object recognition. In general, the shape context at a point captures the distribution over relative positions of other shape points and thus summarizes global shape information in a rich local descriptor. Inspired from the work [8],

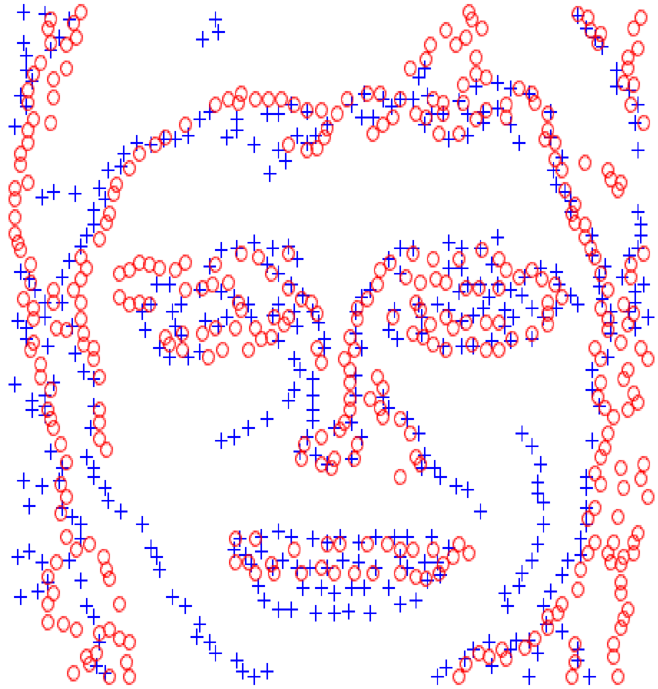


Fig. 2. Example of the original sampled points (that was set to 400) from two face shapes belonging to the same person with varying light and expression. Feature points representing in blue and red refer to two different faces.

where line-rich object, i.e face has been used, in this study, we used Canny edge detection algorithm (an industrial standard tool) to produce possible edge map (as shown in Fig. 1). With the edge map, the shape now can be sampled into a set of N feature points. To reduce the computational complexity, we have roughly uniform spacing to 400 feature points (see Fig. 2).

The shape context descriptor is now computed for each sample point (follow Fig. 2) that will describe the coarse arrangement of the rest of the shape with respect to the point. For any point p_i on the shape, the coarse histogram of the relative coordinates of the remaining $n-1$ points,

$$h_i(k) = \#\{q \neq p_i \quad : \quad (q - p_i) \in bin(k)\}.$$

The above equation is considered to be the shape context of p_i. We used a log-polar coordinate system that has 12 equally spaced angle bins and 5 equally placed log-radius bins. Fig. 3 shows an example of it. For matching two different face shapes, we follow the exact same procedure as mentioned in the original work [5].

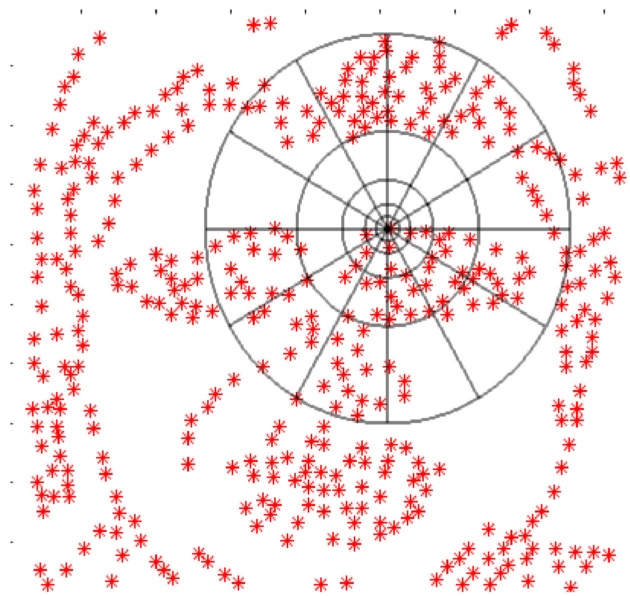


Fig. 3. Computing shape context and store them in log-polar bins.

3.2 Pyramid histogram of oriented gradients (PHOG)

Bosch et al. [6] proposed a descriptor that not only works on the local image shape but it also takes spatial relationship into account. It captures the spatial distribution of edges and stored them as 1D vector representation. With the 1D feature vector, the similarity between two images can easily be computed (just bu using distance metrics).

PHOG is an excellent image global shape descriptor inspired by the two sources: the image pyramid representation of Lazebink et al, [16] and the Histogram of Gradient Orientation (HOG) of Dalal and Triggs [9]. It consists of a histogram of orientation gradients over each image subregion at each resolution level. An image is represented by its local shape (distribution over edge orientations within a region) and its spatial layout(tiling the image into regions at multiple resolutions). It has two major aspects: a) local appearance and b) spatial layout.

Local appearance can be described by a histogram of edge orientations (quantized into K bins) within an image subregion. The edge orientations are quantized into K bins, each of which represents the number of edges which have a certain angular range orientations. The spatial layout of the shape is based on the concept of spatial pyramid matching [15]. Each image is divided into a sequence of increasingly finer spatial grids by doubling the number of grids in each axes direction. The number of points in each grid cell is then recorded. Fig. 4 illustrates an idea about PHOG computation.

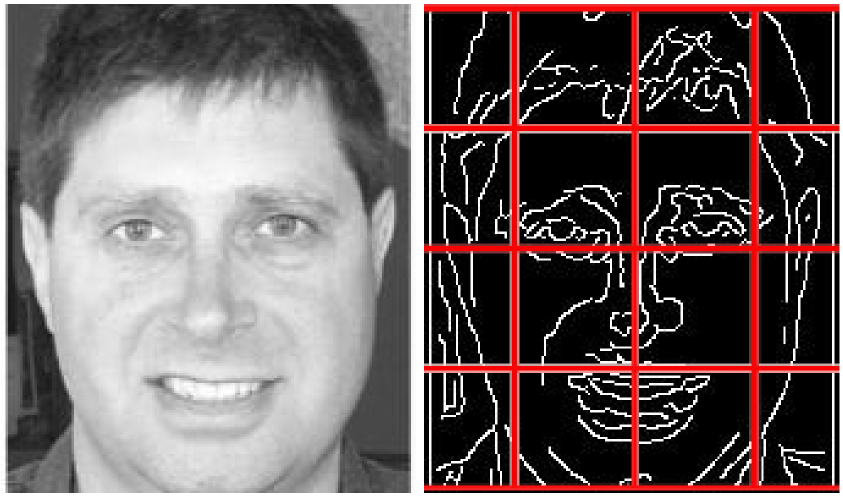


Fig. 4. Computing PHOG with level 2.

In general, the PHOG descriptor of the entire image is a vector with dimensionality $K \times \sum_{l \in L} 4^l$. For example, for level L = 1 and K = 20 bins, the size of the PHOG feature vector is of $(20 \times 4^0 + 20 \times 4^1)$ 100.

3.3 Gist: spatial envelope

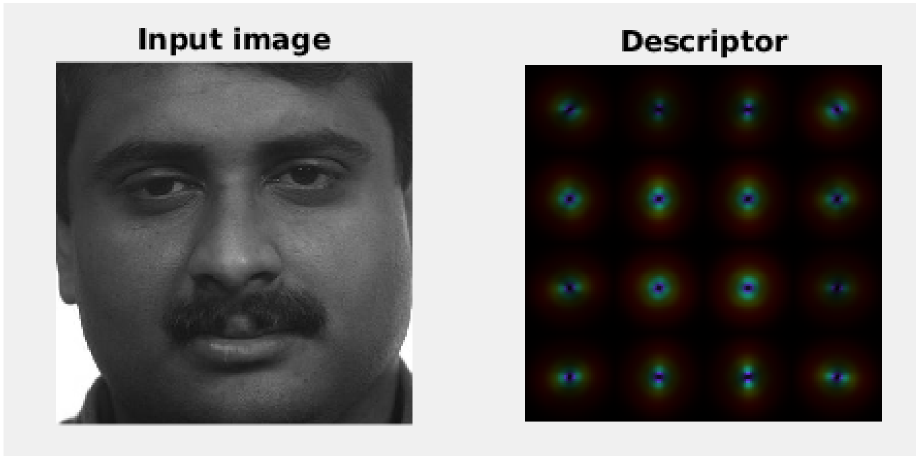
Object recognition tool that can bypass the segmentation process is always interesting. Gist descriptor, proposed by Oliva and Torralba [19] is one of such techniques. Such technique produces a very low dimensional representation of the image, which we call spatial envelope.

Spatial envelope, for image recognition, is considered as a set of global properties. At the image level, each scene property, or the whole image, can be represented by a low-dimensional vector that encodes the distribution of orientations and scales in the image along with a coarse description of the spatial layout. The spatial envelope representation that has semantic attributes about the image provides a way of computing high-level image and space similarities between 2D sequences/images. Such representation of an image is a proof that object shape or identity may not be required for image segmentation/categorization.

In our study, for an input face image, a gist descriptor is computed by:

- Convolve the image with 32 Gabor filters at 4 scales, 8 orientations, producing 32 feature maps of the same size of the input image.
- Divide each feature map into 16 regions (by a 4x4 grid), and then average the feature values within each region.
- Concatenate the 16 averaged values of all 32 feature maps, resulting in a $16 \times 32 = 512$ gist descriptor.

In Fig. 5, an illustration about how can we compute the gist feature is provided.



a) b)

Fig. 5. Computing the gist descriptor of a face image.

4 Experiments

4.1 Datasets and evaluation metric

Datasets For a through test, it is highly recommended to use standard datasets (publicly available) so that fair comparison is possible. In our study, the following datasets (see Table 1 for additional information) are used:

a) Caltech dataset;
b) AR face dataset;
c) Color FERET dataset;
d) ORL dataset; and
e) Indian face dataset.

The California Institute of Technology (Caltech) is a world-renowned science and engineering research and education institution. It has an important part in discovering new knowledge and innovations. For face recognition system, the dataset provided by Caltech contains 450 frontal face images of 27 distinct subjects with varying conditions.

The AR Face dataset was developed by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. The rich dataset contains over 3000 face images corresponding to 126 subject which includes 70 men and 56 women.

The FERET database was developed under the supervision of Face Recognition Technology (FERET) program to develop new techniques, technology, and algorithms for the automatic recognition of human faces. For FR system, the dataset consists of 500 frontal face images of 105 distinct subjects with varying conditions. In our study, we ignore color pixels.

Table 1. Datasets used in our experiments.

Datasets	Dimension	Images/subject	Total
Caltech	190 × 232	27	450
AR Face	120 × 165	11	550
ColorFERET	346 × 351	4 (min.)	500
ORL	92 × 112	10	400
Indian Faces	232 × 232	11	675

The ORL database of faces was developed under the supervision of Cambridge University Laboratory. The database consists of 400 frontal face images of 40 distinct subjects with varying lighting conditions, poses and facial expressions.

The Indian Face database was developed by IIT Kanpur. There are eleven different face images of each of 40 distinct subjects. Different orientations of the face are included for example looking front, looking left, looking right, looking up, looking up looking down along with different facial emotions.

Evaluation metric The various testing protocols have been designed so that algorithm performance can be computed. A common approach of testing the tool is to provide two sets of images to the algorithm, the probe image which is considered to be an unknown individual and training image which consist of known individuals from the gallery. The images in the testing set are the unknown facial images that have to be identified by the tool (proposed tool, let us say). The decision is made on the basis of similarity score computed by the algorithm. To test the performance of our tool, the *hit rate* is used.

The *hit rate* counts the successful top-k matches using the query set of size $|q|$. The higher the recognition rate, the greater the hit rate. We define the hit rate for top-k matches as

$$Hit\ rate(k) \ = \frac{Hit\ count(k,q)}{|q|}.$$

The system is capable of accepting or rejecting the claim.

Like most of the face recognition tools (in the literature), we follow leave one out validation protocol, where each image act as a query knows as the testing data and processed with the remaining set of images known as training data.

4.2 Results, analysis and comparison

In Table 2, our tool with SC features shows encouraging results in all the major face image databases except the Indian faces. This shows that shape context can achieve more in the field of face recognition.

Table 2. Hit rate (in %) using the SC features for different datasets.

Datasets	Top-1	Top-3	Top-5
Caltech	0.92	0.95	0.95
ColorFERET	0.79	0.89	0.91
ORL	0.94	0.96	0.98
Indian faces	0.53	0.62	0.66

Table 3. Hit rate (in %) using the PHOG features for different datasets.

Datasets	Top-1	Top-3	Top-5
Caltech	0.92	0.95	0.96
ColorFERET	0.97	0.98	0.99
ORL	0.99	0.99	1.00
Indian faces	0.61	0.66	0.70

The results using the PHOG features produce better scores (see Table 3) than the ones that uses the SC features. It is also important to note that the PHOG descriptor can outperform major state-of-the-art descriptors (follow comparison Table 5 at the end).

In Table 4, the gist features shows similar trends like previous results (Tables 2 and 3). Comparatively, all features produce better score on ORL dataset. The difficulty for all of them lies on Indian faces dataset, which is primarily due to the fact that face poses have been widely varied from the same person.

For a quick comparative study, in Table 5, we have taken baseline results reported in the literature. These are primarily based on 2D principal component analysis (2D PCA) [22], histogram of gradients (HOG) [11], PCA with linear discriminant analysis (LDA) [23], rotation- and scale-invariant, line-based color-aware descriptor (RSILC) [8], scale-invariant feature transform (SIFT) [3], local binary pattern (LBP) [2] and speeded up robust features (SURF) [13]. In this comparison table (Table 5), we observe that spatial features are worth-taking for face recognition. More importantly, the proposed method does not take color images into account to ease the process, i.e. time complexity.

Table 4. Hit rate (in %) using the gist features for different datasets.

Datasets	Top-1	Top-3	Top-5
Caltech	0.97	0.98	0.98
ColorFERET	0.93	0.96	0.97
ORL	0.98	0.99	0.99
Indian faces	0.59	0.67	0.70

Table 5. Comparison: hit rates (in %) of different approaches.[T1: Top 1, T3: Top 3, T5: Top 5]

Datasets	2D-PCA [22]	HOG [11]	PCA, LDA [23]	RSILC [8]	SIFT [3]	LBP [2]	SURF [13]	Method 1 (SC)	Method 2 (PHOG)	Method 3 (gist)
ORL	0.98	0.95	—	—	—	—	—	T1: 0.94 T3: 0.96 T5: 0.98	T1: 0.99 T3: 0.99 T5: 1.00	T1: 0.98 T3: 0.99 T5: 0.99
ColorFERET	—	—	0.95	0.98	—	0.95	0.95	T1: 0.79 T3: 0.89 T5: 0.91	T1: 0.97 T3: 0.98 T5: 0.99	T1: 0.93 T3: 0.96 T5: 0.97
Caltech	—	—	—	0.96	—	—	—	T1: 0.92 T3: 0.95 T5: 0.95	T1: 0.92 T3: 0.95 T5: 0.96	T1: 0.97 T3: 0.98 T5: 0.98
Indian faces	—	—	—	0.80	—	—	—	T1: 0.53 T3: 0.62 T5: 0.66	T1: 0.61 T3: 0.66 T5: 0.70	T1: 0.59 T3: 0.67 T5: 0.70

5 Conclusion and future works

In this paper, we have studied the usefulness of spatial features for gray level face image representation and recognition. The reason behind the use of gray level images is public video surveillances, for instance, are always in gray tone format due to limited storage capacity. The study has not been inspired from the real-world data but also has been attempted to take advantage of computational complexity (since no color information has been used), i.e. processing time without missing information about the face images. We have observed that faces can be learned/recognized by the use of spatial arrangements of edges and textures. We have used shape context (SC) and pyramid histogram of orientated gradient (PHOG), and similarly GIST as texture features. Comparative study (on four different publicly available datasets, such as Caltech, ColorFERET, IndianFaces and ORL) shows that spatial features are worth-taking for face recognition.

Since spatial features can be compared with the state-of-the-art works, our immediate plan is to integrate/combine them so that gray level face recognition can be possible with no color information it. Another idea is to work on machine learning classifier-based idea, such as active learning [7] at the time when we need real data or live data.

References

1. Abdel-Hakim, A.E., Farag, A.A.: Csift: A sift descriptor with color invariant characteristics. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. vol. 2, pp. 1978–1983. IEEE (2006)
2. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence 28(12), 2037–2041 (2006)
3. Aly, M.: Face recognition using sift features. CNS/Bi/EE report 186 (2006)
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer vision and image understanding 110(3), 346–359 (2008)
5. Belongie, S., Malik, J., Puzicha, J.: Shape context: A new descriptor for shape matching and object recognition. In: Advances in neural information processing systems. pp. 831–837 (2001)
6. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on Image and video retrieval. pp. 401–408. ACM (2007)
7. Bouguelia, M., Nowaczyk, S., Santosh, K.C., Verikas, A.: Agreeing to disagree: active learning with noisy labels without crowdsourcing. Int. J. Machine Learning & Cybernetics 9(8), 1307–1319 (2018)
8. Candemir, S., Borovikov, E., Santosh, K., Antani, S., Thoma, G.: Rsilc: Rotation- and scale-invariant, line-based color-aware descriptor. Image and Vision Computing 42, 1–12 (2015)
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 1, pp. 886–893. IEEE (2005)

10. Deboeverie, F., Veelaert, P., Philips, W.: Face analysis using curve edge maps. In: International Conference on Image Analysis and Processing. pp. 109–118. Springer (2011)
11. Do, T.T., Kijak, E.: Face recognition using co-occurrence histograms of oriented gradients. In: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. pp. 1301–1304. IEEE (2012)
12. Dreuw, P., Steingrube, P., Hanselmann, H., Ney, H., Aachen, G.: Surf-face: Face recognition under viewpoint consistency constraints. In: BMVC. pp. 1–11 (2009)
13. Du, G., Su, F., Cai, A.: Face recognition using surf features. In: Proc. of SPIE Vol. vol. 7496, pp. 749628–1 (2009)
14. Gao, Y., Leung, M.K.: Face recognition using line edge map. IEEE transactions on pattern analysis and machine intelligence 24(6), 764–779 (2002)
15. Grauman, K., Darrell, T.: The pyramid match kernel: Discriminative classification with sets of image features. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. vol. 2, pp. 1458–1465. IEEE (2005)
16. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Computer vision and pattern recognition, 2006 IEEE computer society conference on. vol. 2, pp. 2169–2178. IEEE (2006)
17. Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: Sift flow: Dense correspondence across different scenes. In: European conference on computer vision. pp. 28–42. Springer (2008)
18. Lowe, D.G.: Object recognition from local scale-invariant features. In: Computer vision, 1999. The proceedings of the seventh IEEE international conference on. vol. 2, pp. 1150–1157. Ieee (1999)
19. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision 42(3), 145–175 (2001)
20. Santosh, K.C., Lamiroy, B., Wendling, L.: Integrating vocabulary clustering with spatial relations for symbol recognition. Int. J. Document Analysis & Recognition 17(1), 61–78 (2014)
21. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on. pp. 586–591. IEEE (1991)
22. Yang, J., Zhang, D., Frangi, A.F., Yang, J.y.: Two-dimensional pca: a new approach to appearance-based face representation and recognition. IEEE transactions on pattern analysis and machine intelligence 26(1), 131–137 (2004)
23. Zhao, W., Chellappa, R., Krishnaswamy, A.: Discriminant analysis of principal components for face recognition. In: Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. pp. 336–341. IEEE (1998)