



Face Clustering Utilizing Scalable Sparse Subspace Clustering And The Image Gradient Feature Descriptor

Mingkang Liu, Qi Li, Zhenan Sun, Hongwen Zhang and
Qiyao Deng

EasyChair preprints are intended for rapid
dissemination of research results and are
integrated with the rest of EasyChair.

May 7, 2018

Face Clustering Utilizing Scalable Sparse Subspace Clustering and the Image Gradient Feature Descriptor

Mingkang Liu, Qi Li, Zhenan Sun, Hongwen Zhang, Qiyao Deng

Center for Research on Intelligent Perception and Computing (CRIPAC)
National Laboratory of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences (CASIA)
Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), CAS
{znsun,qli}@nlpr.ia.ac.cn
{mingkang.liu,qiyao.deng,hongwen.zhang}@cripac.ia.ac.cn

Abstract. Face clustering is an important topic in computer vision. It aims to put together facial images that belong to the same person. Spectral clustering-based algorithms are often used for accurate face clustering. However, a big occlusion matrix is usually needed to deal with the noise and sparse outlier terms, which makes the sparse coding process computationally expensive. Thus spectral clustering-based algorithms are difficult to extend to large scale datasets. In this paper, we use the image gradient feature descriptor and scalable Sparse Subspace Clustering algorithm for large scale and high accuracy face clustering. Within the image gradient feature descriptor, the scalable Sparse Subspace Clustering algorithm can be used in large scale face datasets without sacrificing clustering performance. Experimental results show that our algorithm is robust to illumination, occlusion, and achieves a relatively high clustering accuracy on the Extended Yale B and AR datasets.

Keywords: face clustering, scalable Sparse Subspace Clustering, image gradient feature descriptor

1 Introduction

With the development of Internet, the problem of image organization and management has become an important issue. Naturally, most Internet images contain human faces. To better understand and manage face images, face clustering becomes an essential task. Face clustering aims to group faces which refer to the same people together. It has many applications, such as a preprocessing step for face retrieval and face tagging. Many face clustering algorithms have been reported in recent years [1–3]. Although great progress has been achieved, face clustering algorithms still suffer from large variations in illumination, expression and occlusion, *etc.* Spectral clustering-based algorithms can handle noise and outliers in data samples and thus have drawn much attention in recent years for face clustering.

Enhanced Sparse Subspace Clustering (ESSC) is used to process complicated face images under variant expressions, illumination or disguise [4]; Conditional Pairwise Clustering (ConPaC) method learns the adjacency matrix directly from a given similarity matrix. And ConPaC's k-NN variant can cluster millions of face images [5]. The subspace clustering method based on orthogonal matching pursuit not only has computational efficiency, but also give a subspace-preserving affinity under broad conditions. The application of face clustering shows that this method achieves the best trade off between accuracy and efficiency [6]. The regularized RPCA algorithm based on spatiotemporal sparse spectral clustering is used for efficient background modeling. The main advantage of this algorithm is that it can generate an accurate background model even if there are occlusion, confusion, jitter, and abrupt intensity variations [7].

The first step of spectral clustering-based algorithms is to construct an affinity matrix, and then Normalized Cuts [8] is employed to segment the data samples into different clusters. Various spectral clustering-based algorithms have been proposed based on sparse and low-rank representation. The main difference of these algorithms is construction of the affinity matrix. Sparse Subspace Clustering algorithm [9, 10] uses the l_1 norm regularization on the coefficient matrix to find the sparsest representation of each data sample. Low-Rank Representation algorithm [11–13] employs nuclear norm to seek the lowest rank representation of all data samples. Least Squares Regression algorithm [14] encourages l_2 norm regularization on the coefficient matrix to obtain a block diagonal solution. In particular, Sparse Subspace Clustering algorithm is well supported by theoretical analysis and achieves state-of-the-art results on many publicly available datasets [15].

There are still some issues to be further addressed for clustering various face images when using spectral clustering-based algorithms. Most of the spectral clustering-based algorithms are difficult to extend to large scale datasets because of the noise and sparse outlying terms in the objective function. For example, if the dimensionality of features used in Sparse Subspace Clustering algorithm is 4096, then an extra 4096×4096 noise matrix is needed. Such a big noise term will make the sparse coding process computationally expensive and sometimes even prohibitive when the dimensionality of features is high. While the clustering accuracy will dramatically down without the noise and sparse outlying terms. What's more, the ability to cope with various illumination and occlusion for spectral clustering-based algorithms still needs to be improved.

Feature descriptors provide a possible solution to deal with the above problems. A low dimensional yet powerful feature descriptor is very helpful for face clustering. Recent studies have shown that the image gradient feature descriptor is widely used in many applications: face recognition, face alignment, visual tracking [16–18]. Compared with pixel based methods, the image gradient feature descriptor is less sensitive to variations of illumination and occlusion. The dimensionality of the image gradient feature descriptor is lower compared with other frequently used feature descriptors. In this paper, we propose a scalable Sparse Subspace Clustering algorithm for face clustering utilizing the image gra-

dient feature descriptor. Our algorithm achieves a satisfied clustering accuracy and can be applied to large scale face datasets.

The main contributions of our work are summarized as follows.

1) We proposed a scalable Sparse Subspace Clustering algorithm for face clustering utilizing the image gradient feature descriptor. Our algorithm achieves a better clustering performance than other spectral clustering-based algorithms. Besides, the computational cost of our algorithm is low, which provides a promising solution for large scale and high accuracy face clustering problem.

2) The image gradient feature descriptor is first introduced to cluster large scale face datasets. A name of sorted local gradient pattern is grayscale inversion and rotation invariant descriptors for texture classification. Image rotation and linear or non-linear grayscale-inversion changes are highly discriminative and robust [19]. Different feature descriptors are compared for face clustering. Experimental results show that image gradient feature descriptor is very simple but very competitive compared with other feature descriptors, *e.g.*, HOG, LBP and Gabor. Another novel color image inpainting algorithm. The novelty lies in the use of gradient features combined with color features. When the image coexists with focus and blurry regions, the use of a gradient feature can improve the results obtained only by the color features [20].

2 Image Gradient Based Subspace Clustering Algorithm

2.1 Image Gradient Feature Descriptor

Given a face image $I_i \in \mathbb{R}^{w \times h}$, we compute the gradient and the corresponding gradient orientation:

$$\Phi_i = \arctan \frac{H_y * I_i}{H_x * I_i}, \quad (1)$$

where H_x and H_y are the differential filters along the horizontal and vertical face image axis respectively, $H_x * I_i \in \mathbb{R}^{w \times h}$ and $H_y * I_i \in \mathbb{R}^{w \times h}$ denote the horizontal and vertical convolution of the face image. We write Φ_i in lexicographic order and stack it as a m dimensional vector ϕ_i , then we have the image gradient feature descriptor as follows:

$$f(\phi_i) = \frac{1}{\sqrt{m}} \left[\cos(\phi_i)^T + j \sin(\phi_i)^T \right]^T, \quad (2)$$

where $\cos(\phi_i) = [\cos(\phi_i(1)), \dots, \cos(\phi_i(m))]^T$, $\sin(\phi_i) = [\sin(\phi_i(1)), \dots, \sin(\phi_i(m))]^T$.

Next, we will illustrate that the image gradient feature descriptor is a simple yet powerful feature descriptors to measure image similarity and thus it is useful for face clustering. With the image gradient feature descriptor, the correlation of two face images can be expressed as:

$$c(f_i, f_j) = f_i^T f_j = \frac{1}{m} \sum_{k=1}^m \cos(\Delta\phi_{ij}(k)), \quad (3)$$

where $\Delta\phi_{ij} = \phi_i - \phi_j$ is the difference of the feature descriptor between face image I_i and I_j . The difference of two face images can be written as:

$$d^2(f_i, f_j) = \frac{1}{2} \|f_i - f_j\|_2^2 = 1 - \frac{1}{m} \sum_{k=1}^m \cos(\Delta\phi_{ij}(k)). \quad (4)$$

Definition 1. *If two images I_i and I_j are dissimilar, then $\Delta\phi_{ij}(k) \sim U[0, 2\pi]$.*

Definition 1 has already been verified by [18]. From Eqn.(3) and Eqn.(4), we can see that if two face images are similar to each other, then $c \rightarrow 1$, $d \rightarrow 0$. Note that similar feature descriptor has also been used in [18, 21]. This feature descriptor is robust to outliers and there is no need to add the noise term in the sparse coding process.

2.2 scalable Sparse Subspace Clustering

We denote the i -th face image as vector $y_i \in \mathbb{R}^m$ ($m = w \times h$) by stacking its columns. Then the dataset of n face images can be represented as a matrix $Y = [y_1, \dots, y_n] \in \mathbb{R}^{m \times n}$. According to [22], y_i can be represented by a linear combination of other images:

$$y_i = Yx_i, \quad x_{ii} = 0, \quad (5)$$

where $x_i = [x_{i1}, \dots, x_{in}]^T \in \mathbb{R}^{n \times 1}$, $x_{ii} = 0$ avoids writing a face image as a linear combination of itself. Because the solution is not unique, l_0 norm is used as a constraint to seek the sparse solution of Eqn.(5). Considering the non-convexity of l_0 norm, l_1 norm is replaced with the l_0 norm and Eqn.(5) becomes:

$$\min \|x_i\|_1 \quad s.t. \quad y_i = Yx_i, \quad x_{ii} = 0. \quad (6)$$

We rewrite Eqn.(6) in matrix form:

$$\min \|X\|_1 \quad s.t. \quad Y = YX, \quad \text{diag}(X) = 0. \quad (7)$$

In order to deal with the corrupted or occluded face images, the noise term is usually added in Eqn.(7):

$$\min \|X\|_1 + \|E\|_1 \quad s.t. \quad Y = YX + E, \quad \text{diag}(X) = 0. \quad (8)$$

Eqn.(8) is first proposed in [9, 10]. However, the noise term prohibits this algorithm using in large scale face datasets.

If we remove the noise term from Eqn.(8) directly, the clustering performance will decrease dramatically. Fortunately, the powerful image gradient feature descriptor provides an opportunity to remedy this drawback. With the image gradient feature descriptor, there is no need to use the noise term. Based on the above analysis, the proposed scalable Sparse Subspace Clustering (sSSC) is formulated as:

$$\min \|X\|_1 \quad s.t. \quad Y = YA, \quad A = X - \text{diag}(X). \quad (9)$$

Augmented Lagrange Multiplier method is used to solve this problem. The augmented Lagrangian function of Eqn.(9) is defined by:

$$\begin{aligned} \mathcal{L}(A, X, \Delta_1, \Delta_2) = & \min \|X\|_1 + tr(\Delta_1^T (Y - YA)) \\ & + tr(\Delta_2^T (A - (X - \text{diag}(X)))) \\ & + \frac{\mu}{2} \left(\|Y - YA\|_F^2 + \|A - (X - \text{diag}(X))\|_F^2 \right), \end{aligned} \quad (10)$$

where Δ_1^T, Δ_2^T are the Lagrangian multipliers, tr denotes the trace operator of a matrix, μ is the penalty term. Eqn.(10) can be minimized through an alternative strategy with respect to A and X by fixing the other variables and then we update Δ_1^T, Δ_2^T and μ as the following forms:

$$A^{k+1} = \arg \min_A \mathcal{L}_{\mu^k}(A, X^k, \Delta_1^k, \Delta_2^k), \quad (11)$$

$$X^{k+1} = \arg \min_X \mathcal{L}_{\mu^k}(A^{k+1}, X, \Delta_1^k, \Delta_2^k), \quad (12)$$

$$\Delta_1^{k+1} = \Delta_1^k + \mu^k (Y - YA^{k+1}), \quad (13)$$

$$\Delta_2^{k+1} = \Delta_2^k + \mu^k (A^{k+1} - C^{k+1}), \quad (14)$$

$$\mu^{k+1} = \rho \mu^k. \quad (15)$$

Eqn.(11) can be solved by computing the derivative of \mathcal{L} with respect to A and setting it to zero:

$$\begin{aligned} A^{k+1} = & (\mu (Y^T Y + I)) \\ & (\mu Y^T Y + \mu (C - \text{diag}(C)) + \Delta_1 - \Delta_2). \end{aligned} \quad (16)$$

Eqn.(12) can be solved by the soft threshold method:

$$\begin{aligned} X^{k+1} = & J - \text{diag}(J) \\ J = & \mathcal{S}_{\frac{\mu}{2}} \left(A^{k+1} + \frac{\Delta_2^k}{\mu} \right), \end{aligned} \quad (17)$$

where \mathcal{S} is the soft-thresholding operator. After obtaining the representation parameter, we define the affinity matrix as $(|X| + |X^T|) / 2$. Then Normalized Cuts [8] is used to segment the image datasets. In summary, given face images, we extracted the image gradient feature descriptor (IG). Then we put the feature vectors to the scalable Sparse Subspace Clustering algorithm (sSSC) to cluster the face images.

3 EXPERIMENTS

Experiments are conducted on two datasets: the Extended Yale B and the AR datasets. The Extended Yale B [23] dataset consists of 2,414 frontal face images from 38 subjects under various lighting conditions. The cropped and normalized 192×168 face images are captured under various controlled lighting conditions [24]. The AR dataset consists of over 4,000 face images from 126 subjects.

For each subject, 26 face images are taken in two separate sessions. These images suffer from different facial expressions (neutral, smile, anger, and scream), illumination variations (left light on, right light on, and all side lights on), and occlusion by sunglasses or scarf.

The clustering result is evaluated by the **error rate**:

$$error = 1 - \frac{\sum_{i=1}^n \delta(y_i, map(s_i))}{n}$$

where y_i and s_i is the obtained cluster label and the ground truth label, δ is the delta function, $map(s_i)$ is the permutation function that maps each cluster label s_i to the equivalent label in y . All of the experiments are implemented using MATLAB on a Intel Core i7-2600 3.40GHZ machine with 16GB memory.

3.1 Experimental Results on the Extended Yale B Dataset

The first 10 subjects of Extended Yale B dataset are used to validate the clustering accuracy of our algorithm. Different feature descriptors are extracted from the original cropped images (192×168) and then sent to the sSSC algorithm for face clustering. Similar to other sparse representation algorithms for data processing, we use l_2 norm to normalize all of the feature descriptors. Because of the high dimensional feature descriptors, PCA is used to project all of the feature descriptors to $k \times 6$, where k is the number of subspaces (here is the number of person). The parameter μ is tuned empirically to have the best clustering results across all of the 10 subjects. The compared feature descriptors are listed as follows.

Local Binary Patterns (LBP) [25]: The standard uniform LBP operator $LBP_{8,2}^{U_2}$ is used with the MATLAB source code from [25, 26]. Cell size of 8×8 is used to form a local histogram of 59 uniform patterns. Histograms of all the cells are combined to represent the whole image, resulting a 29736 ($192/8 \times 168/8 \times 59$) dimensional feature descriptor. Gabor Energy Filters (Gabor) [27]: For the Gabor feature descriptor, similar to [28], a filter bank of 5 scales and 8 orientations are used. We also down-sample the obtained feature descriptor by a factor of 16. Then the combined 40 filters result in a 80640 ($192 \times 168 \times 40/16$) dimensional feature descriptor. Histogram of the Oriented Gradient (HOG) [29]: For the HOG feature descriptor, we use the toolbox from [30]. The Spatial bin size is 8×8 . Four different normalizations are computed using adjacent histograms, resulting in a 9×4 length vector for each region. Thus a total number of 18144 ($192/8 \times 168/8 \times 36$) dimensional feature descriptor is extracted from the cropped image.

The clustering accuracy of different feature descriptors is shown in Fig. (1). From Fig. (1), we can see that image gradient feature descriptor achieves the best clustering performance. Besides, LBP feature descriptor performs better than HOG and Gabor feature descriptors. This may be because LBP feature descriptor is more robust to illumination changes than HOG and Gabor feature

descriptors. Although HOG feature descriptor also utilizes image gradient information, the quantization process may lose some useful information and lead to a poor performance on face clustering task.

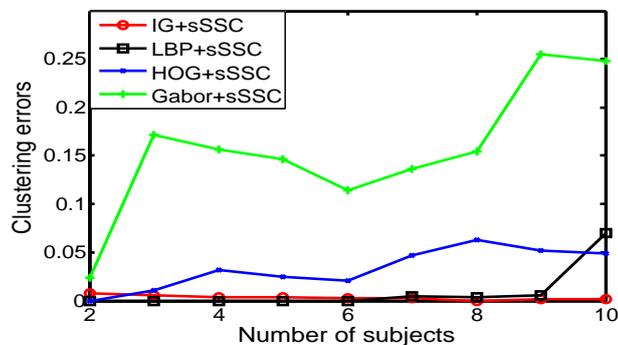


Fig. 1. Clustering errors of different feature descriptors on the Extended Yale B dataset (the less, the better). IG: image gradient feature descriptor, HOG: HOG feature descriptor, LBP: LBP feature descriptor, Gabor: Gabor feature descriptor.

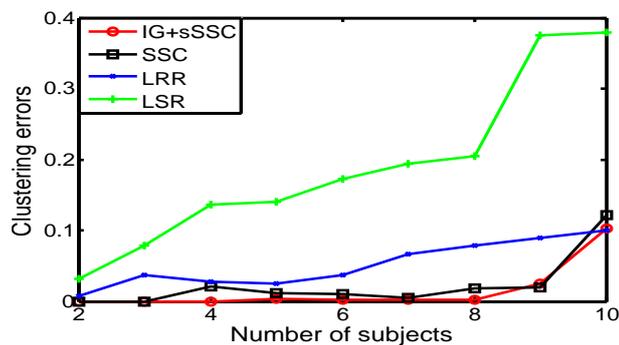


Fig. 2. Clustering errors of different algorithms on the Extended Yale B dataset (the less, the better). LBP, Gabor and HOG feature descriptors have poor clustering performances on the down-sampled images. So we haven't reported their clustering accuracy.

Besides, we also compare our algorithm with other subspace clustering algorithms. The compared algorithms are: Sparse Subspace Clustering algorithm (SSC) [9], Low-rank Recovery algorithm (LRR) [12] and Least Squares Regression algorithm (LSR) [14]. In order to have a fair comparison, we test all of the algorithms (including our algorithm) on the down-sampled images of reso-

lution 48×42 without performing PCA projection. All of the parameters are tuned carefully across the 10 subjects to have the best clustering accuracy. The clustering accuracy of different algorithms is shown in Fig. (2). From Fig. (2), we can see that our algorithm has the lowest clustering error among all of the compared algorithms. This result further validates that our algorithm is robust to illumination changes.

The computational time of different algorithms are listed in Table 1. As shown in Table 1, the computational time of our algorithm is lower compared with SSC, LRR. This advantage is exaggerated when the number of subjects becomes large. Although the computational time of LSR is much lower, its clustering accuracy is also much lower than other algorithms. Compared with other algorithms, our algorithm achieves a satisfied clustering results while still maintaining the low computational cost.

Algorithm	SSC	LRR	LSR	IG+sSSC
2 subjects	22	6	0.02	4
5 subjects	46	26	0.08	14
8 subjects	71	68	0.24	29
10 subjects	117	135	0.27	43

Table 1. The computational time (sec.) of different algorithms on the Extended Yale B dataset.

3.2 Experimental Results on the AR Dataset

In this experiment, we evaluate the robustness of our algorithm on the AR dataset with different occlusions. A subset of 50 male and 50 female subjects are selected for this experiment. It contains two separate sessions. In each session, each subject has 7 face images with different facial variations, 3 face images with sunglasses occlusion and 3 face images with scarf occlusion. All of the images are resized to a resolution of 32×32 .

Because of the limited speed of LRR and SSC algorithms, we can't cluster all of the images at a time. So two sessions are used separately which is the same experimental setting as [31]. The first 2 normal face images and 3 face images with sunglasses of each subject are used for sunglasses occlusion. The first 2 normal face images and 3 face images with scarf of each subject are used for scarf occlusion. The clustering accuracy of different algorithms is shown in Table 2. From Table 2 we can see that the clustering accuracy of our algorithm is better than SSC, LSR and LRR algorithms. Our algorithm performs better than [31] for scarf occlusion. While for sunglasses occlusion, [31] performs better than our algorithm. The possible reason is that there are limited number of face images for each subject. If the number of images per subject is increased, both SSC and our algorithm can achieve a better clustering performance.

Algorithm	Session 1		Session 2	
	Sunglasses	Scarf	Sunglasses	Scarf
IG+sSSC	0.200	0.182	0.184	0.178
LSR	0.218	0.284	0.200	0.274
LRR	0.228	0.278	0.204	0.254
SSC	0.562	0.594	0.722	0.598
rCIL2	0.148	0.216	0.136	0.188
CIL2	0.188	0.246	0.146	0.210

Table 2. Clustering errors of different algorithms on the AR dataset (the less, the better). The clustering performance of LSR, LRR and SSC, CIL2 and rCIL2 is reported in [31].

4 Conclusions

In this paper, we have proposed an efficient scalable face clustering algorithm utilizing the image gradient feature descriptor. Our algorithm has the advantage of relatively low computational cost and high clustering accuracy, which provides a promising solution for large scale face clustering problem.

References

1. N. Vretos, V. Solachidis, and I. Pitas, "A mutual information based face clustering algorithm for movie content analysis," *Image and Vision Computing*, vol. 29, no. 10, pp. 693–705, 2011.
2. B. Wu, Y. Zhang, B.-G. Hu, and Q. Ji, "Constrained clustering and its application to face clustering in videos," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3507–3514, 2013.
3. S. Xiao, M. Tan, and D. Xu, "Weighted block-sparse low rank representation for face clustering in videos," in *European Conference on Computer Vision*, pp. 123–138, 2014.
4. J. Ren, S. Zhao, K. Yang, and B. Zhao, Nlong, "A novel and robust face clustering method via adaptive difference dictionary," pp. 627–632, 2017.
5. Y. Shi, C. Otto, and A. K. Jain, "Face clustering: Representation and pairwise constraints," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1626–1640, 2018.
6. C. You, D. P. Robinson, and R. Vidal, "Scalable sparse subspace clustering by orthogonal matching pursuit," pp. 3918–3927, 2017.
7. S. Javed, A. Mahmood, T. Bouwmans, and S. Jung, Ki, "Backgroundforeground modeling based on spatiotemporal sparse subspace clustering," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 5840–5854, 2017.
8. J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
9. E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

10. E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2790–2797, 2009.
11. G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *IEEE International Conference on Computer Vision*, pp. 1615–1622, 2011.
12. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.
13. P. Favaro, R. Vidal, and A. Ravichandran, "A closed form solution to robust subspace estimation and clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1801–1807, 2011.
14. C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *European Conference on Computer Vision*, pp. 347–360, 2012.
15. V. M. Patel, H. Van Nguyen, and R. Vidal, "Latent space sparse subspace clustering," in *IEEE International Conference on Computer Vision*, pp. 225–232, 2013.
16. G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Sparse representations of image gradient orientations for visual recognition and tracking," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 26–33, 2011.
17. G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Robust and efficient parametric face alignment," in *IEEE International Conference on Computer Vision*, pp. 1847–1854, 2011.
18. G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Subspace learning from image gradient orientations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2454–2466, 2012.
19. T. Song, L. Xin, C. Gao, G. Zhang, and T. Zhang, "Grayscale-inversion and rotation invariant texture description using sorted local gradient pattern," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 625–629, 2018.
20. A. Jurio, D. Paternain, J. Fernandez, L. Miguel, De, and H. Bustince, "Image inpainting using colour and gradient features," *Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems*, pp. 1–6, 2017.
21. T. Cootes and C. Taylor, "On representing edge structure for model matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1114–1119, 2001.
22. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
23. A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.
24. K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.
25. T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
26. M. Heikkil and T. Ahonen, "A general local binary pattern (lbp) implementation for matlab." <http://www.cse.oulu.fi/CMV/Downloads/LBPmatlab/>.

27. C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.
28. M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *European Conference on Computer Vision*, pp. 448–461, 2010.
29. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–893, 2005.
30. P. Dollár, "Piotr's Image and Video Matlab Toolbox (PMT)." <http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html>.
31. C. Lu, J. Tang, M. Lin, L. Lin, S. Yan, and Z. Lin, "Correntropy induced l2 graph for robust subspace clustering," in *IEEE International Conference on Computer Vision*, pp. 1801–1808.