Face Recognition and Human Action Recognition using L_1-Principal Component Analysis

Gaurav Karki
Department of Computer Engineering
Santa Clara University
California, USA
gkarki@scu.edu

***Abstract*— This paper develops upon two main tasks: face recognition using the L_1-principal and then using similar approach for human action recognition. To calculate the L_1-Principal Component for rank k, we used the individual method [1] and matched the testing images to training images using Nearest Subspace Algorithm. L_1-subspace works better than the L2-subspace under the effect of outliers [1]. The experiments in this paper supports this and verifies as results in human action recognition where improved when we used individual method [1].**

Keywords—Action Recognition, Face Recognition, Principal Components.

I. Introduction

The actions performed by the humans are complex and sophisticated, hence there is lack of techniques available to identify them correctly .Human Action Recognition is used at various sciences like Robotics, Human Surgery, Biology and many more. Since a long time, L_2-PCA has been preferably used over L_1-PCA [4], it was calculated either using SVD of data or EVD of auto covariance of data [4]. The computation of L_1-Principal Component for N signal samples of dimension D such that $X \in R_{N \times D}$ was considered as NP hard until fast algorithm[4] was introduced which has time complexity of $O(N^3)$.

It is known that L_1-PCA outperforms L_2-PCA in presence of outliers[1], but L_1-PCA didn't utilized the information from the class labels of training data. The Individual method[1] describes how to include the class labels in L_1-PCA by computing the "individual" subspaces for each class using the training images of that class. Our paper extended the use of Individual method[1] to perform human action recognition on the Weizmann Space-Time Classification Database[6].

We implemented the L_1-PCA algorithm as described in the Individual method[1] and tested upon the Extended Yale face database and ORL face database to observe the error percentage over the six ranks of L_1-subspace. The test images from different classes where used to find the nearest subspace in the trained L_1-subspace. Further, we used similar approach to recognize the human action in the Weizmann Database by converting the videos into a motion history image, which represents a silhouette of the human action over black background. We trained these motion history images using the same algorithm used above and compared the test sample of videos from classes [walk, jack, bend] with nearest subspace of trained motion history images based on training classes labels.

In further sections, we will show related work done in the field, we will also describe our algorithm for human action recognition, show results of the experiments and make conclusions.

II. Related Work

A. *Face Recognition with L_1-norm Subspaces*

The L_1-norm subspaces could perform better than the L_2-norm with outside outliers, but they didn't used information from class labels. Individual L1-subspace solves that problem.

Algorithm for Individual L1-subspace Face Representation:

- Make the training data zero centered.
- Find the rank-K L_1-subspace of each class j using the fast algorithm[4].
- Remove the contribution of this class q1(j) from the entire dataset.
- The updated training ensemble X(j) is used to calculate the next principal component and the representation
- continues until the desired number of components K is reached.
- Classification using N.S and Subtracting the mean of j^{th} class from x(t).
- Finding the minimum distance from all the classes for test face image.

B. *Fast Computation of the L_1-Principal Component*

For $X \in R_{D \times N}$ of N signal samples of dimension D, the computation of the L1 principal component of the data by maximum L1-norm projection is **NP-hard.** But if the signal D is fixed such that D<=N, L1 PCA can be computed in Polynomial time $O(N^{rank(X)})$, rank(X) $\leq$ D. It had the drawback that for moderate to larger values of rank(X)<=D, it is still unsuitable for practical implementation.

Optimal Algorithm by Markopoulos et al. [4], provided the L1 PCA in polynomial time using greedy single-bit-flipping (SBF).

C. *Weizman Human Actions as Space-Time Shape*

The human action can be seen as a 2D matrix over Space-Time plane[6]. The algorithm builds up a silhouette from movement of the body parts like legs, arms and torso. They calculated the volumetric space-time action shapes. The Poisson equation's solution was used to extract the space-time features such as local space-time saliency, action dynamics, shape structure and orientation. They used simple human actions like walking, jumping, galloping and bending to perform the analysis.

III. METHODOLOGY

We have used Individual L1-subspace method to model both face recognition and human action recognition's training data. After that, we classified some test samples by calculating the minimum distance to training classes' labels by using the Nearest Subspace Algorithm.

A. *Individual L1-subspace Face Representation*[1]

They propose a novel method to include the class labels information along with L_1-subspace, by computing the subspace for each class individually from the training samples of face images. Consider that there are C classes of training samples and each class has N images of size D pixels, then our training matrix $X^{(j)} = [x_1, x_2, x_3, \dots, x_N] \in R_{N \times D}$, where j = 1,2,…,C is the class index and each column is a vectorized training image of class j that has D pixels. First step is to remove the sample mean from all columns of $X^{(j)}$ so that training data is zero centered. Then find the rank-K L_1-subspace of each class j using the fast algorithm, [4] which is as follows:

$$\mathbf{q}_1^{(j)} = \arg \max_{\mathbf{q}\in\mathbb{R}^D, \|\mathbf{q}\|_2=1} \left\|\mathbf{q}^T\mathbf{X}^{(j)}\right\|_1 .$$

We can also use bit-flipping algorithm instead of fast algorithm to find principal component, which is as follows:

1.) It focus on the maximization problem for single component D.

$$\mathbf{r}_{L_1} = \arg \max_{\mathbf{r}\in\mathbb{R}^D, \|\mathbf{r}\|_2=1} \|\mathbf{X}^{\mathrm{T}}\mathbf{r}\|_1$$

2.) Find the b_{opt} using alternative bit-flipping algorithm.

$$\mathbf{b}_{\mathrm{opt}} = \arg \max_{\mathbf{b}\in\{\pm 1\}^N} \mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}.$$

3.) Calculate the principal component as shown below:

$$\mathbf{r}_{L_1} = \frac{\mathbf{X}\mathbf{b}_{\mathrm{opt}}}{\|\mathbf{X}\mathbf{b}_{\mathrm{opt}}\|_2}$$

Next step is to remove the contribution of each class j from the entire training dataset, update the $X^{(j)}$ to calculate next principal component, using shown equation:

$$\mathbf{X}^{(j)} = \mathbf{X}^{(j)} - \mathbf{q}_1^{(j)}\mathbf{q}_1^{\mathrm{T}(j)}\mathbf{X}^{(j)}.$$

Repeat the previous step until required number of rank or k principal components have been calculated. Compile all the $q^{(j)}$ into in single $Q_{L1}^{(j)} = [q_1^{(j)}, q_2^{(j)},\dots,q_k^{(j)}]$ for the class j = 1,2…C.

Next step to find out the minimum distance from trained subspace to the testing image using the nearest subspace algorithm.

$$\widehat{j} = \arg \min_{1\le j\le C} \left\|(\mathbf{x_t} - \mu^j) - \mathbf{Q}_{L_1}^{(j)}\mathbf{Q}_{L_1}^{\mathrm{T}(j)}(\mathbf{x_t} - \mu^j)\right\|_2$$

The j^{th} class which provides the minimum value for $\hat{j}$ would be matched to the testing image.

B. *Proposed Human Action Recognition Algorithm*

The program was coded in Python3 using the OpenCV library. Following is the algorithm:

1. Read the video for actions j = [walk, jack, bend] using the OpenCV library in python.
2. Capture frames from the video and resize them to 60x48.
3. Consider the frame for human action j = [walk, jack, bend], then 2D frame $X = [\ \ ]_{60x48}$
4. Calculate the absolute difference between each 4^{th} frame and convert it into grayscale.
5. Pass the grayscale image through a threshold into order to remove the background surroundings from moving object.
6. The resulting object is a motion history image which is similar to a silhouette[5].
7. Flatten the resulting Motion History Image for class j = [walk, jack, bend] and calculate the L_1-Principal Components using Individual method[1].
8. Using greedy approach, find $Q_{L1}^{(j)} = [q_1^{(j)}, q_2^{(j)},\dots,q_k^{(j)}]$ for all j = [walk, jack, bend].
9. Test some random sample videos using the Nearest Subspace Algorithm.
10. Test the Robust Database = [walking with dog, walking with bag, ….]$_{1x11}$ using Nearest Subspace Algorithm.

IV. EXPERIMENTAL STUDIES

We experimented on two databases Extended Yale and ORL database, to observe the performance of individual L_1-subspace algorithm:

A. *Extended Yale Database*

The Extended Yale Database has 38 classes representing different individuals and each class has 25 images of size 50x50 pixels.

The experiment was performed 10 times individually on 8 random classes from the set. The training sample included 8 images and the testing sample included 17 images both from each class.

Figure A.1: Sample classes/person in different expression and angles from the Extended Yale database

Rank of L_1- subspace	Total Images Tested	Images Identified Incorrectly	Error %
1	1224	86	7.03
2	1224	26	2.12
3	1224	8	0.65
4	1224	5	0.41
5	1224	2	0.16
6	1224	3	0.24

Figure A.2: Results after running 10 individual test experiments for Extended Yale

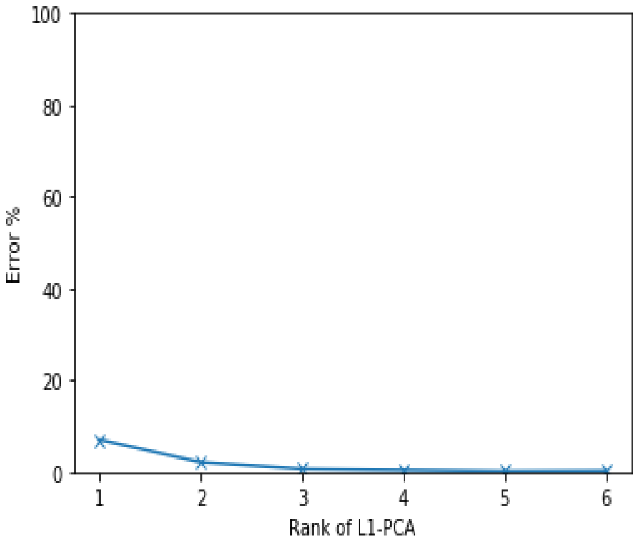


Figure A.3: Plot showing how error % vs Rank of L_1-PCA

B. ORL Database

The ORL Database has 40 classes representing different individuals and each class has 10 images of size 50x50 pixels.

The experiment was performed 10 times individually on 8 random classes from the set. The training sample included 7 images and the testing sample included 3 images from each class.

Figure 2: Sample classes/person in different expression and angles from the ORL database

Rank of L_1- subspace	Total Images Tested	Images Identified Incorrectly	Error %
1	240	32	13.33
2	240	2	0.83
3	240	2	0.83
4	240	2	0.83
5	240	2	0.83
6	240	2	0.83

Figure B.2: Results after running 10 individual test experiments for Extended ORL

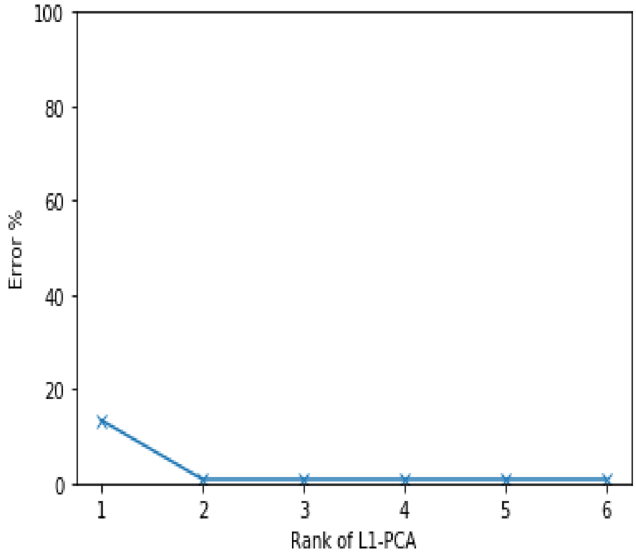


Figure B.3: Plot showing how error % vs Rank of L_1-PCA

C. Weizmann Database

The Weizmann Database consist of 11 classes from which we are have used "walk", "bend" and "jack". Each class represents a human performing certain motion and contains 9 videos each of different persons.

The experiment was performed on above mentioned classes by using 7 videos for training data and 2 videos for testing data. We converted the training images into Motion History Images as shown in fig C.1 .

1.)Walk 2.) Jack 3.) Bend

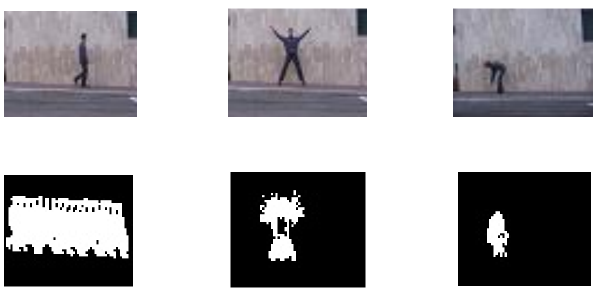
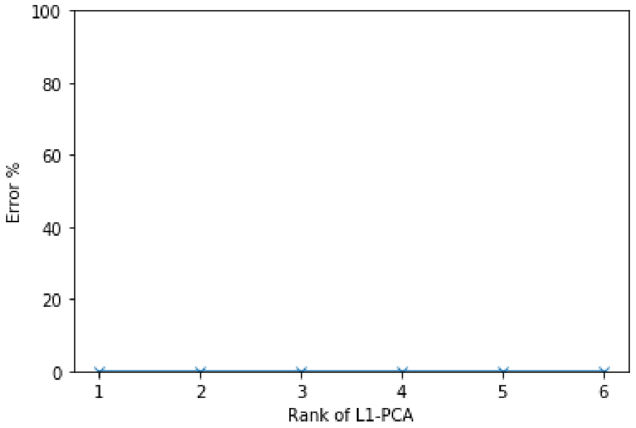


Figure 2: Motion History Images for walk, jack and bend

Rank of L_1-subspace	Total Images Tested	Images Identified Incorrectly	Error %
1	7	0	0
2	7	0	0
3	7	0	0
4	7	0	0
5	7	0	0
6	7	0	0



We also experimented for the robustness using the 11 classes like "walking with dog", "walking with bag", and others. The model was trained using previous classes i.e. walk, bend and jack. Upon testing with robustness database, we got 100% accuracy in results, as it was able to identify all walking even in the presence of outliers like dog, bag and pole.

a.)Walk with dog b.)Walk with skirt c.)Walk with pole

V. Conclusions

In this paper, we proposed that Human Action Recognition can be done by training motion images using the individual L_1-PCA method[1]. We observed that Individual L_1-PCA has excellent performance for face recognition. We also observed that its error percentage reduces with increase in the rank K. The human action recognition gave 0 % error in the robustness testing of walk action using L_1-PCA.

References

[1] F. Maritato, Y. Liu, D. A. Pados, and S. Colonnese, "Face recognition with L1-norm subspaces," in *Proc. SPIE Commercial + Scientific Sensing and Imaging*, Baltimore, MD, Apr. 2016.

[2] P. P. Markopoulos, G. Karystinos, and D. A. Pados, "Optimal algorithms for L-subspace signal processing,"

[3] N. Kwak, "Principal component analysis based on L1-norm maximization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, pp. 1672-1680, Sep. 2008.

[4] S. Kundu, P. P. Markopoulos, and D. A. Pados, "Fast computation of the L1-principal component of realvalued data," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP), Florence, Italy, May 2014, pp. 8028-8032.

[5] Y. Kong and Y. Fu, "Human action recognition and prediction: a survey," arXiv preprint arXiv:1806.11230 (2018).

[6] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in Proc. ICCV, 2005. http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html - Classification Database.