Universitat Autònoma de Barcelona

escola d'enginyeria

6153

Face Recognition for Long-Term Interaction

Memòria del Projecte Fi de Carrera
d'Enginyeria en Informàtica
realitzat per

Ms. Narayanaraju Sravanajyothi

i dirigit per

Dr. Bogdan Raducanu

Bellaterra, 5 de Junio de 2015

El sotasignat, Dr. Bogdan Raducanu

professor/a de l'Escola d'Enginyeria de la UAB,

CERTIFICA:

Que el treball a què correspon aquesta memòria ha estat realitzat sota la seva direcció per en/na Narayanaraju Sravanajyothi

I per a que consti firma la present.

Raducanu B.

Signat: Dr. Bogdan Raducanu

Bellaterra, 5 de junio de 2015

Abstract

Face recognition is one of the few biometric methods that possess both accuracy and intrusiveness. For this reason it has drawn attention of many researchers and numerous algorithms have been proposed. Various fields such as network security, surveillance benefits from the face recognition because it provides more efficient coding scheme. Since the face recognition is a real world problem and there are cases when not all the input data is not known beforehand. In this project the focus is on the online learning strategy.We implemented online nonparametric discriminant analysis methodology for long-term face recognition problem. The advantage of using NDA over LDA is explained briefly. Besides reviewing the online version of NDA, we propose an optimized version based on 'affective forgetting'. In order to guarantee real-time response, the online learning strategy has been extended with a pruning mechanism which gets rid of the oldest samples. Experimental results on the FRIENDS dataset demonstrated that the performance of classification is not affected by replacing the former samples with new ones.

Acknowledgements

I would like to express my deepest appreciation to all those who provided me the possibility to complete this report. A special gratitude to my supervisor, Dr.Bogdan Raducanu, for his timely help, valuable suggestions, guidance and encouragement throughout the completion of the project especially in writing this report.

I would forever remain grateful to SASTRA UNIVERSITY, for giving me an opportunity to do this project.

I express my earnest and humble thanks to my colleague Suman Ghosh who helped me throughout my project.

Last but not least, I also thank all my colleagues who have directly or indirectly helped me by extending their moral support and encouragement for completion of this project.

Contents

1 Introduction

1.1 Scope of Project

Face recognition systems used in present days work very well under constrained conditions, although all systems work much better with frontal mug-shot images and constant lighting. All current face recognition algorithms tends to fail under few varying conditions under which humans need to and are able to identify other people. Next generation person recognition systems are expected to recognize people in real-time and in much less constrained situations.Considering all the requirements and the constraints, identification systems that use face recognition and speaker identification seem to us to have the most potential for wide -spread application.

Several approaches such as [7] have been proposed to solve face and recognition problem. Here in this project we have followed the 'Incremental Non-parametric Discriminant Analysis' (NDA) strategy in order to address the problem long-term face recognition. The reason behind choosing this method is that it is more realistic in case of real time applications. NDA is known to cope well with nonlinear, non-gaussian distributions, This can be used robustly in applications which deal with significant variations in data input. Our contribution consists in an optimisation step for the Incremental NDA based on an 'active forgetting' mechanism which periodically replaces the oldest samples by new ones, thus guaranteeing a real-time response.

1.2 Motivation

Face recognition has been grabbing attention of the scientific community for several decades. The motivation resides in the fact that faces represent our main gateway to express our emotions and at the same time serve as the main cue to identify people. Faces represent a particular type of objects, with a very well defined structure but which are subject to elastic deformations. From this perspective, they share some similarities with hand-written characters, for instance. Although studied for a very long time, face recognition is only a partially solved problem, achieving a high recognition rate in strictly controlled conditions (i.e. passport like images). Other than these, there are a couple of factors that makes face recognition a difficult problem: illumination conditions as discussed in c , presence of artefacts (beard, glasses, facial make-up, etc.), changes of facial appearance over time and pose. In [3] , they have proposed a face recognition method which is pose-independent and similar case is briefly discussed in [10]. Face recognition is desirable in the applications where the system has to identify the person without having the need of people interacting with the system like giving fingerprints or talking in the case of voice recognition.Face recognition is used in many areas (like human-computer interaction, biometrics, social robotics, etc.)

Due to the difficulty of the face recognition task, the number of techniques is large and diverse. There are many popular algorithms used for face recognition like Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), also known as Fisher discriminant analysis, neural networks. Each method has it own perks. Like PCA, it primarily uses Eigenfaces, but in a supervised way. It is used to reduce the dimension of the data which eventually drops the irrelevant information and decomposes data into orthogonal components called eigenvectors. Knowing the eigenspace, when a new test face image is presented, the image is projected on the eigenvectors and using a classifier (e.g. nearest neighbour) we could determine the identity of the face. Coming to the LDA method, it is a supervised method, which takes into account class label information to build the eigenspace.. In this case we try to maximise the between class variance and the minimise the within class variance. To be more formal if given a number of independent features relative to which the data is described, LDA generates a linear combination of these features which yields the largest mean difference between desire classes of the data.

Many face recognition systems nowadays work on standard datasets, consisting of a limited number of classes and few samples per class (tens). In this project we will focus on the problem of face recognition on larger datasets, where the accent is not on the number of persons, but rather on the number of samples for each person (thousands), in order to see how our face recognition system could cope and adapt to changes in facial appearance over a long period of time (weeks).

1.3 Outline

The rest of the paper is organised as follows: In the next section, we will briefly describe the different types of face recognition i.e., image based and video based recognitions and also the challenges faced in this field. In section 3 we discuss both versions of NDA i.e., batch NDA and incremental NDA and its optimisation via ‘active forgetting’. Section 4 is dedicated to present the experimental results describing the algorithm performance . Application of this methodology to social robots will be contained in section 5 . And finally section 6 contains the conclusion of the report and the guidelines for future work.

2 State of Art

Despite tremendous effort to solve the face recognition problem and to design a system close to human performance, it still remains a challenge. Design of robust face recognition systems require new perspectives of pattern recognition approaches.There are many factors like poor illumination that affect the recognition rate of the system and here is the place where we shall be requiring some non trivial engineering efforts to increase its robustness. However , adapting the model to specific task seems to be a realistic and important problem. Data acquisition of the training images is of main concern in any system. The argument here is that the manual encoding of person's details along with the facial images is not only complex but also highly impractical because it requires a tedious process. To avoid this issue in [14], they proposed a method called online learning where the learning process starts with empty database and it keeps on learning incrementally in online fashion. This automatic online learning process is a bit complex since the number of clusters i.e., number of individuals is unknown in the beginning.

The multi-view face recognition as stated in the reference source [4] , deals with the image aquisition from the multiple cameras simultaneously.In the case of video data, multi-camera system guarantees the acquisition of multi-view data at any moment,the chance of obtaining the equivalent data by using a single camera is highly unpredictable. Such things makes vital difference in non-cooperative applications such as surveillance.When compared to still image based recognition and video based recognition, smaller number of approaches are in daily use.In [15], they talked about multi-view face recogntion system face recognition, in a strict sense, only refers to situations where multiple cameras acquire the subject(or scene) simultaneously. They also discussed about video based recognition where multiple cameras are used to obtain multi view data. However the image acquisition from this multi camera might not be similar to that of image from single camera. This might pose a problem in applications like surveillance.

In the next two sections we will briefly review the state of the art in still and video-based face recognition

2.1 Still-Image based recognition

Lot of research can be seen in image based multi view face recognition area. Existing algorithms are based on view synthesis like in [10], 3D model construction as in [2], regularized regression, stereo matching and local feature matching as stated in the reference [5] . Due to their effectiveness in handling pose variations, local patch/feature based approaches have become popular in present day scenario such as [3]. Given a pair of face images to verify, they look up in the collection to "align" the face part's appearance in one image to the same pose and illumination of the other image. This method will also require the poses and illumination conditions to be estimated for both face images. This "generic reference set" idea has also been used to develop the holistic matching algorithm , where the ranking of look-up results forms the basis of matching measure.There are also works which handles pose variations implicitly , without an explicit representation. For example in [1], they showed the modeling of location-augmented local descriptors using a Gaussian Mixture Model, they performed probabilistic elastic matching on some face images even when large pose variations existed.

2.2 Video based recognition

The complexity increases when it comes to video since video contains more information than still images. A smart and straightforward way to handle single-view videos is to take advantage of the data redundancy and perform view selection. Some of the algorithms employed a combination of skin-color detector and edge feature-based SVM regression to localize face candidates so that estimation of their poses is no longer a difficult task. Then, for each of the candidates, a face detector specific to that pose is applied to determine if it is a face. Only the frontal faces are retained for recognition. Few algorithms made use of SVM to select frontal faces from video for recognition as described in [16]. The idea of modeling face pose manifolds has inspired the continuity of pose variation in video. The classical method is to cluster the frames of similar pose and train a linear subspace to represent each cluster with different pose. Modelling of face videos are being done with Gaussian Mixture models by assuming that all the face images of a single person sits on a manifold. Single view videos have been modelled by Hidden Markov models as per [11], and the output 3D model has been used as model based algorithm.

In [1], they stated that the continuity of pose variation in video has inspired the idea of modeling face pose manifolds. To represent each pose cluster they have formed different clusters for each pose and trained the linear subspace. Some of them like in [17], made the linear subspace grow gradually from a seed sample to include more and more nearest neighbors, until the linearity condition is violated.The linearity according to them is measured as the ratio of geodesic distance to Euclidean distance, and the distances are calculated between a candidate neighbor and each existing sample in the cluster. In [13] ,they developed a fully automatic end-to-end system for video face recognition, which includes face tracking and recognition leveraging information from both still images for the known dictionary and video for recognition. They have also proposed Mean Sequence Sparse Representation-based Classification (MSSRC), which performs a joint optimization over all faces in the track at once.

2.3 Challenges

Although lot of research has been done in face recognition field, more effort is required in order to develop a robust system with a potential close to human performance. Nevertheless to say , achievement of 100 percent recognition is a desirable, but hard to reach objective. Unstructured environments characterised by variations in illumination , clutter are still a challenge for researchers.

Considering the illumination issue, the variations between the images of the same face due to the change in the direction are greater than the variations due to the change in face identity as stated in [17]. The variations can also be caused by the factors like facial expression, hair styles and also ageing. Since the problem of face recognition falls under standard pattern recognition or machine learning method, many geometric based methods have used several properties and relations between facial features like eyes, mouth to perform recognition. Many of the existing methods for pattern recognition have been proved ineffective because of their insensitivity to large variations in pose , lighting and viewpoint.

Other appearance based methods which uses low dimensionality representation of face images as stated in [18], have demonstrated their power in terms of efficiency and also accuracy. However these appearance based methods fail miserably under poor illumination conditions and this in turn have huge effect on the overall performance of the system. In order to overcome these obstacles, 3D based face recognition method is proposed in [8]. Here we build a 3D face model of the person is formed with the help of shading in the face image using parametric model of face shape. Furthermore, there are few complications arising due to voluntary changes in the facial expressions and the less adverse effects due to aging.

Coming to the feature based methods , they use the distance between landmarks such as nose,eyes for the feature extraction and these are not reliable and not robust due to the inconsistency mainly due to pose variations and are highly inaccurate. However there are some parameters to determine the efficiency of any approach like FAR (False Acceptance Rate) , FRR (False Recognition rate) and equal error rates. By taking into account these parameters we can choose the method which is most desirable for the application being carried out.

In this section some of most related works of recent past in the above said areas are reviewed.

3 Nonparametric Discriminant Analysis

Several approaches have been proposed to render the feature selection technique in high dimensional classification problems. If we take the most popular method LDA(Linear Discriminant Analysis) [6], we are interested in features which reduce the dimensionality while preserving class separability. The objectives of LDA comes in two folds, i) predict the lowest possible dimension of the subspace which is also called as intrinsic discriminant dimension in such a manner that classification accuracy is maximum and ii) to extract the features with maximal discriminative power. This method is most suitable for linearly separable data and is less effective for non linearly separable data.

LDA, is based on the assumption that it is dealing with Gaussian distribution. The main disadvantages of this approach are i) It is based on the assumption that all the classes share Gaussian distribution. and ii) it is ineffective in capturing boundary class information because the between class scatter matrix takes only the centres of the classes into account while computing the matrix.

In [12], they have proposed a nonparametric technique to solve the aforementioned problems. The Nonparametric Discriminant Analysis (NDA) overcomes the main limitations of LDA, being suitable to deal with nonlinear, non-gaussian distributions (see Figure 1 left). This is due to the fact that NDA is able to capture the differences between class boundaries, in contrast with LDA which takes into account only the class centers. In the next section we review both the classical NDA (called here Batch NDA) and its online version.

3.1 Batch NDA

In classical or batch NDA is passive learning approach which assumes that entire data is presented in advance and is truly informative. To be more formal learning happens in single step. Here the within class scatter matrix(Sw) and between class scatter matrix(Sb) are used to calculate class separability. In [14],they have proposed that one of the criterias is to maximize the following expression :

$$\zeta = tr(S_w^{-1} S_b) \tag{1}$$

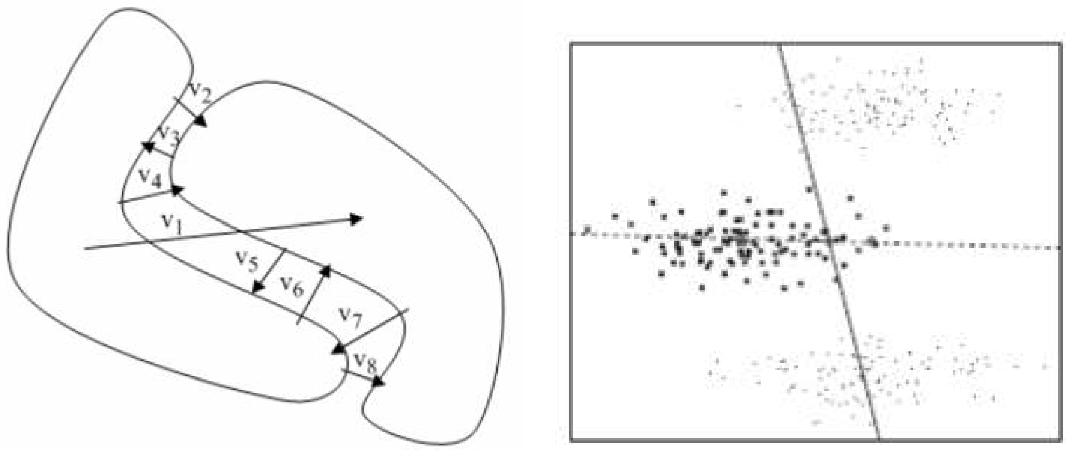

Figure 1: Showing NDA vs LDA where NDA is effective in obtaining boundary value information and LDA proving to be ineffective with non Gaussian distribution

The scatter matrices are defines as follows

$$S_w = \sum_{i=1}^{C_N} \sum_{j \in C_i} (x_j - x^{-C_i})(x_j - x^{-C_i})^T \tag{2}$$

$$S_b = \sum_{i=1}^{C_N} \sum_{j \neq i}^{C_N} \sum_{t=1}^{n_{C_i}} W(C_i, C_j, t)(x_t^i - \mu_{Cj}(x_t^i))(x_t^i - \mu_{Cj}(x_t^i))^T \tag{3}$$

where

$$\mu_{cj}(x_j^i) = \frac{1}{k} \sum_{p=1}^{k} NN_p(x_t^i, C_j) \tag{4}$$

3.2 Online NDA

In [14] ,they proposed online version of NDA which is more concrete and they sequentially updated the eigenspace representation. It is highly impractical to calculate the entire scatter matrices with the incoming samples each time. So here as per proposed method, we updated the already calculated space and scatter matrices. The proposed method starts with minimum two classes by calculating scatter matrices sb and sw and updating them with the later on sequentially added samples . In this way we overcome the drawbacks of classical NDA and also the overhead of computational time can be minimised. Following equations are used to update the scatter matrices :

$$S_b^{'} = S_b - S_b^{in}(C_L) + S_b^{in}(C_L^{'}) + S_b^{out}(y^{C^L}) \tag{5}$$

$$S_b^{in}(C_L) = \sum_{j=1 j \neq L}^{C_N} \sum_{i=1}^{n_{Cj}} W(C_j, C_L, i)(x_i^j - \mu_{CL}(x_i^j))(x_i^j - \mu_{CL}(x_i^j))^T \tag{6}$$

$$S_b^{out}(y^{C_L}) = \sum_{j=1 j \neq L}^{C_N} (y^{C_L} - \mu_{C_j}(y^{C_L}))(y^{C_L} - \mu_{C_j}(y^{C_L}))^T \tag{7}$$

$$S_w^{'} = \sum_{j=1 j \neq L}^{C_N} S_w(C_j) + S_w(C_L^{'}) \tag{8}$$

and

$$S_w(C_L^{'}) = S_w(C_L) + \frac{n_{CL}}{n_{CL}+1}(y - x^{-C_L})(y - x^{-C_L})^T \tag{9}$$

3.3 Optimisation via active forgetting

Learning is an innate cognitive ability which has empowered humans with intelligence and power.As dicussed in [9]. It is achieved by acquiring new knowledge and use it to adapt to the situations for survival. Inspired by the learning paradigm which takes place in the biological systems, present day technology opens the door for the analysis of huge amount of data in a similar, incremental way. Online learning can be considered crucial for classification since it is one of those learning algorithms which iteratively selects the distinctive information essentially important for the classifier. Here we performed the active learning under supervised settings by discriminative selective sampling criteria which reduces the computational cost. As told in [9], learning is a process in which organised representation of experience is constructed whereas forgetting is the process in which these organised parts are either discarded or rearranged. During the learning phase, the amount of knowledge that could be acquired is very large and this might increase the execution time, thus affecting the system's performance.

In consequence, additional optimisation process might be required in order to maintain system's response time, but without affecting significantly its performance. The solution is again provided by the learning mechanisms that take place in the biological systems. In other words, this optimisation process could be considered analogous to forgetful property of human brain. This way we can overcome the memory limitations and meet the optimization requirements. Here the memory is updated by active learning and optimization is achieved through 'forgetfulness' or 'degradation' of the memory.We sequentially update the eigenspace by the adding face images and when the data size reach some threshold, we discard the samples which became obsolete and lost significance in data representation. We do it in such a way that face recognition accuracy is not affected.

To implement this method ,here we used a timestamp variable associated with every sample and when the total number of samples exceed some limit i.e., threshold we discarded the samples with old timestamp values. Usually human memory is organised in such a manner that we tend to forget old memories or instances. This is the motivation behind the current experiments where old samples are removed from the knowledge base in order to make room for the new instances. The main reason behind using this active forgetting technique is to find a trade off between the execution time and overall performance and also to reduce the memory cost.

4 Experimental Results

4.1 Dataset description and experimental setup

To examine the efficiency of the algorithm, we selected 'F.R.I.E.N.D.S' dataset which consists of 6543 samples with the face images taken from the sitcom 'FRIENDS'. The face images of 6 main characters of the sitcom forms this database of 6543 face images with an average of 1000 images for each character. Images in this database comes with different pose, illumination and directions and the size of each face image is 4848. Our choice for this dataset has been motivated by the objective of our experiment. We needed a dataset with a relative small number of classes but a significant number of samples per class in order to simulate the 'long-term' effect. From several standard datasets we have considered, this is the one which fulfills our needs.

To implement the method, the whole dataset is partitioned into training and test datasets. 80% of the database is dedicated for training and the remaining 20% is used for testing the algorithm.From the training dataset ,we used 9% (473 face images with six classes) to train the initial batch NDA and also to build the initial eigenspace .Singularity problem has been avoided by performing the PCA step before hand. This helped to reduce the dimension of the data from 2304 to 200. The remaining part (5432 face images approx.) of the dataset is used for incremental learning stage by adding each image sequentially thus updating eigenspace. After addition of every instance , we evaluated performance at every learning stage. The classification accuracy is examined based on nearest neighbor rule (k-NN one). Each time eigenspace is updated and the eigenvectors of the test samples are projected and the error rate is calculated. This error rate is further used to determine the recognition rate. There is gradual increase in the recognition rate with each instance being added to update the eigenspace construction. Few samples from the database are shown in figure 2.

4.2 Algorithm Performance

To test the performance , we carried out the experiments on the database with different ratios of training and test datasets and the results are examined. Recognition rates for each instance are calculated and plotted to compare . Timestamp

Figure 2: Samples from 'FRIENDS' database with 6 different classes of face images having size of 48×48 each.

variable is used here as a criterion to model 'active forgetting': the oldest samples are discarded from the eigenspace representation. The main reason behind implementing active forgetting is to find the trade off between the execution time and recognition rate. Since the algorithm has practical uses, the execution time of the method is crucial and should be kept optimal to make best use of it.

We have done the experiment with different ratios of training and test datasets and performance has been evaluated. Different permutations of the samples are used to implement the algorithm and the average recognition rates are plotted using both the methods of classifiers i.e., Nearest Neighbor rule and multisvm. We have used a 5-fold cross-validation strategy. Ratios including 40% training 60% test , 60%training and 40% test and finally 80% training ,20%test are used to observe the behaviour of algorithm. In all cases, we have used an initial percentage of 10% of the training data to build NDA eigenspace. The 'refresh' of the NDA eigenspace by removing of oldest samples occurred after every 500 additions.

Figures 3 and 4 depicts the Incremental NDA algorithm with ratio 40-60 using kNN and multi svm classifiers respectively. It can be concluded that multi svm classifier is better than kNN since the gap between the curves of Batch NDA and Incremental NDA is minimum with multi svm and this could be explained for need to keep a near real-time performance for our system.

As previously stated above same kind of experiments are done with ratio 60-40 i.e., 60% training and 40% test dataset. Average recognition is calculated and Figure 6 represents such curve when the classifier is kNN. The curve is showing maximum recognition rate as 96% which is good enough in terms of real time

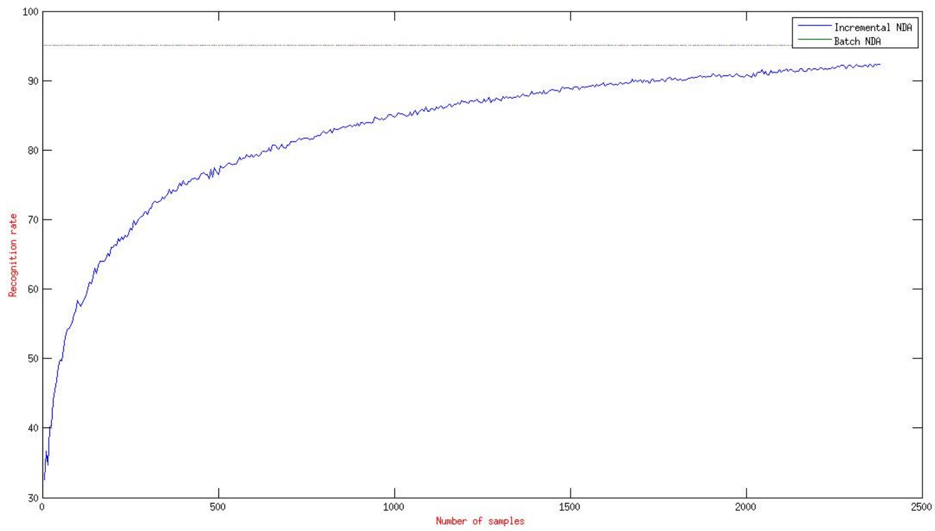

Figure 3: Average recognition rate curve with 40-60 ratio using k-NN classifier with active forgetting employing timestamp method. Here X-axis shows the number of samples and Y-axis depicts the percentage of recognition rate.

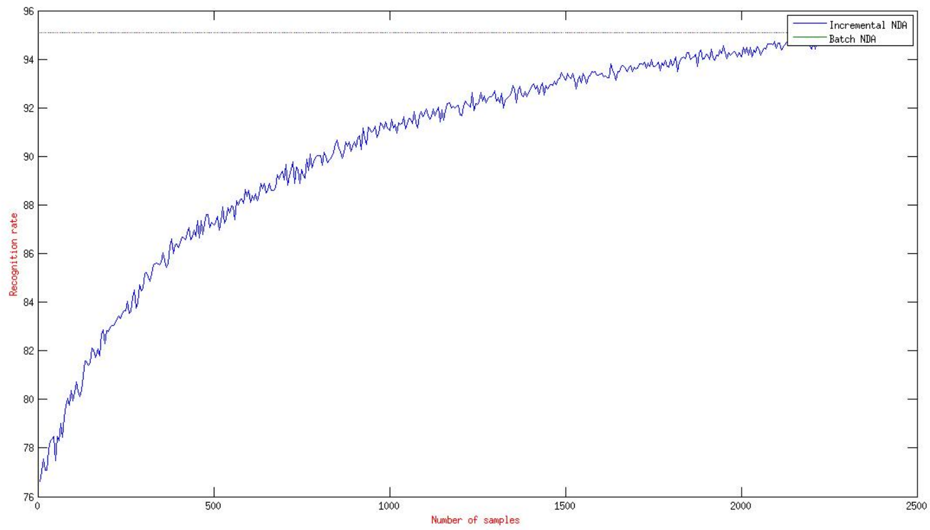

Figure 4: Average recognition rate curve with 40-60 ratio of dataset using multisvm as classifier by using timestamp variable for active forgetting .Here X-axis shows the number of samples and Y-axis depicts the percentage of recognition rate.

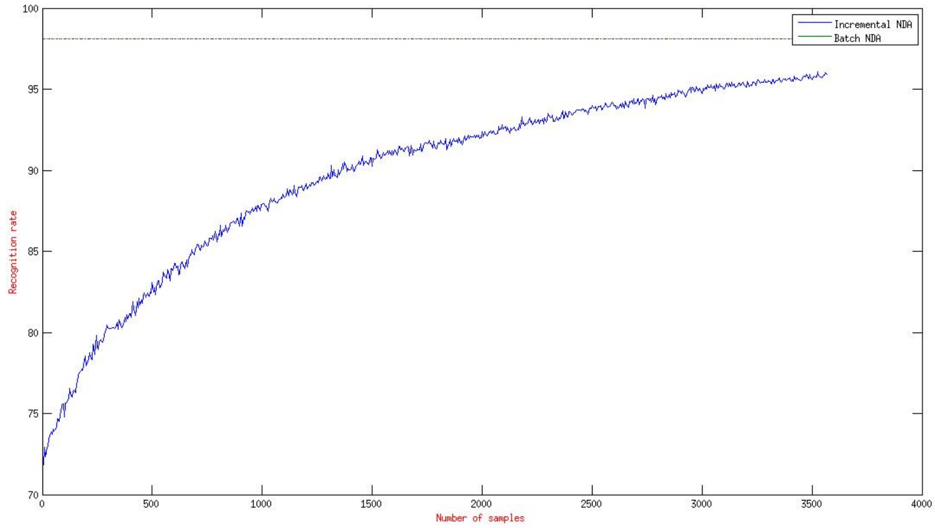

Figure 5: Curve showing average recognition rate with 60-40 ratio using k-NN classifier with active forgetting by using timestamp variable.

applications.

In Figure 5 , we have depicted the algorithm with the curve representing average recognition rate using the Nearest Neighbor as classifier.It could be appreciated that the maximum recognition rate achieved is around 96% .For comparison purposes, we plotted the recognition rate obtained by executing Batch NDA algorithm using classifier multi svm. This is done for showing the performance of the incremental algorithm. In figure 6 we have depicted the performance of the algorithm using multisvm classifier.

Finally, figures 7 and 8 depict the use of 80-20 ratio for computing the recognition rate, using Nearest Neighbor and multisvm classifier, respectively. From the graph it is evident that convergence of the curve with multisvm classifier is greater than that of curve when k-NN is used. The maximum recognition rate of the Incremental NDA algorithm that can be deducted from the graph is 96 approximately.

From all the above curves, we could appreciate that even with removing samples, the recognition rate still converges towards the batch more, thus confirming our hypothesis that discarding irrelevant information, without statistical relevance, does not affect the system performance significantly. In figure 9, we summarize all

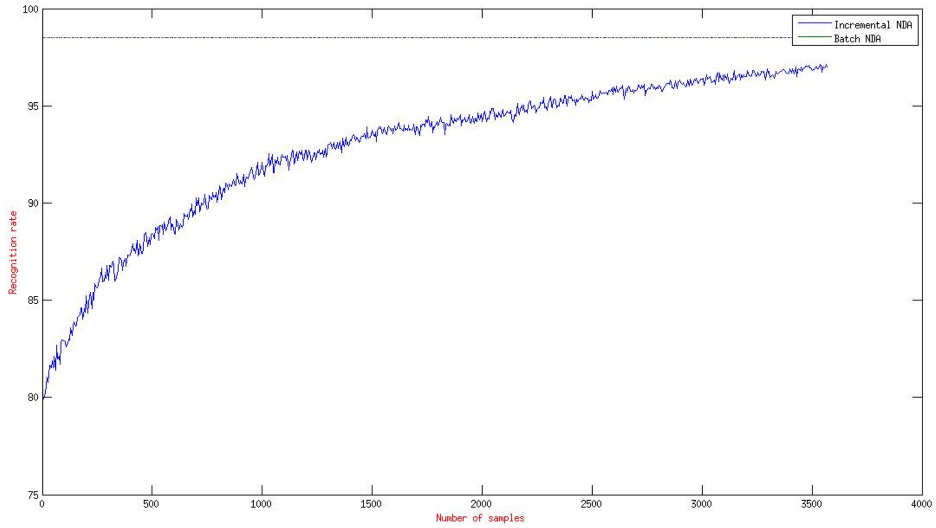

Figure 6: Average recognition rate curve with 60-40 ratio using multisvm classifier with discarding samples using timestamp method.

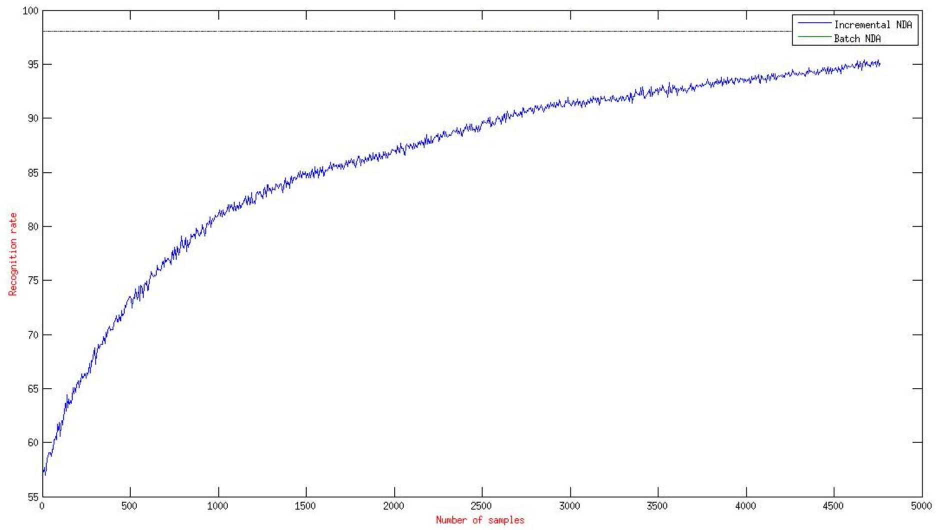

Figure 7: Average recognition rate with ratio of 80 (training) - 20 (test) using k-NN classifier with active forgetting using timestamp.

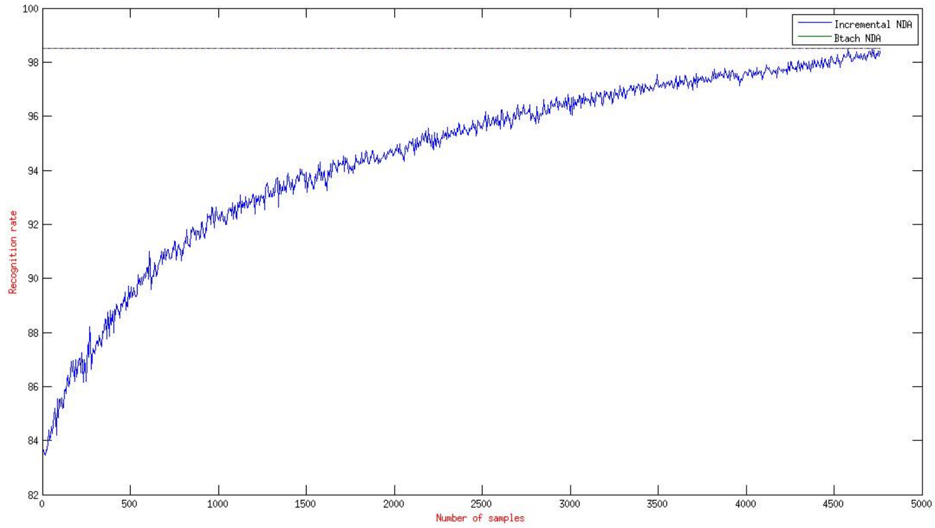

Figure 8: Smooth curve representing the average recognition rate obtained with 80%training and 20% test dataset using multisvm classifier with active forgetting using timestamp.

the results presented so far, by mentioning the maximum recognition rate achieved in each case. This gives us clear picture of the efficiency of the proposed optimisation.

In Figure 10, we have represented the time needed to update the scatter matrices after each added sample. It could be appreciated that keeping all the samples is not feasible if we want to keep the real-time performance of our system. For this reason we came up with a simple policy to remove the 'oldest' samples, from each class, after every 500 iterations. For this purpose, each sample has been 'labelled' with a 'timestamp' (i.e. its index). An almost equal number (80%) of

Modality	Nearest Neighbor		SVM	
	Batch	Online	Batch	Online
40-60	95.0	91.21	95.15	94.95
60-40	97.91	95.21	98.10	97.12
80-20	98.10	96.00	98.50	98.10

Figure 9: Maximum recognition rate for all 3 modalities using the Nearest Neighbor and multisvm classifiers (in percentage)

Number of samples	500	1000	1500	2000	2500	3000	3500	4000	4500
Incremental NDA	2.9	4.3	8.1	12.2	17	25	32	40	50
Incremental NDA with removal	2.9	3.1	3.2	3.4	3.3	3	3.4	3.4	3.4

Figure 10: showing execution time in seconds for Incremental NDA with and without 'active forgetting'.

samples are removed from each class in order to avoid the unbalanced effect. It could be appreciated that this way the time needed to update is kept relatively low, which makes the online learning algorithm with 'active forgetting' suitable for real-time performance. This is another way to demonstrate the efficiency of our optimisation process.

5 Conclusion

In this project we have implemented Incremental NDA algorithm using two different classifiers which are kNN and multi svm along. An optimisation process has been proposed based on 'active forgetting'. This mechanism is inspired from the biological learning systems, and refers to discarding old information which is no longer relevant. Timestamp has been used as a criterion to model 'active forgetting'.We demonstrated the performance of the algorithm on the FRIENDS dataset, which consists of a small number of class, but with a large variation inside each class. Experimental results demonstrated that our solutions maintain the system's performance near real-time without a significant loss in recognition rate. Future work will be devoted to deploy the current algorithm to a robotic platform and thus test it in a real-world scenario.

References

[1] Shakhnarovich G. Fisher J. Cipolla R. Darrell T. Arandjelovic, O. Face recognition with image sets using manifold density divergence,pp. 581–588. 2005.

[2] Kim K. Kienzle W. Scholkopf B. Breuer, P. and V. Blanz. Automatic 3d face reconstruction from single images or video, pp. 1–8. 2008.

[3] Yin Q. Tang X. Sun J. Cao, Z. Face recognition with learning-based descriptor , pages 2707-2714. In *CVPR2010*, 2010.

[4] Li W. Xu D. Shan S. Chen X Cui, Z. Fusing robust face region descriptors via multiple metric learning for face recognition , pp. 3554-3561. In *CVPR2013*, 2013.

[5] Sankaranarayanan A.C. Chellappa Du, M. Robust face recognition from multi-view videos, 23(3):1105-1117. 2014.

[6] O. Hart PE. Stork DG. Duda, R. Pattern classification. 2001.

[7] Ekenel H.K Stiefelhagen R. Fisher, M. Person re-identification in tv series using robust face recognition and user feedback, 55(1):83-104. 2011.

[8] Hu Y. Yan S. Zhang L. Zhang H. Gao W. Jiang, D. Efficient 3d reconstruction for face recognition, pp. 787–798. 2005.

[9] Paul D.S. KaufmannShaul, M. The role of forgetting in learning. 1998.

[10] D. Kim. Generating frontal view face image for pose invariant face recognition, pp. 747–754. 2006.

[11] Chen T. Liu, X. Video-based face recognition using adaptive hidden markov models, pp. 340–345. 2003.

[12] Tang X. Liu W, Lin D. Nonparametric subspace analysis for face recognition, pp 961–966. 2005.

[13] Wright A. Shah M. Ortiz, E.G. Face recognition in movie trailers via mean sequence sparse representation-based classification,pp.3531-3538. In *CVPR2013*, 2013.

[14] Vitrià J. Raducanu, B. Online nonparametric discriminant analysis for incremental subspace learning and recognition, 11(3-4):259-268. 2008.

[15] Koterba S. Xiao J. Hu X. Matthews I. Baker S. et al. Ramanathan, K. Multiview aam fitting and construction,vol. 76, pp. 183–204. 2008.

[16] C. Shan. Face recognition and retrieval in video,pp.235-260. 2010.

[17] Shan S. Chen X. Wen G. Wang, R. Manifold-manifold distance with application to face recognition based on image set ,pp. 1–8. 2008.

[18] Yang A.Y. Ganesh A. Sastry S.S. Ma Y. Wright, J. Robust face recognition via sparse representation, 31(2):210-227. 2009.