Dynamic Batch Mode Active Learning

Shayok Chakraborty, Vineeth Balasubramanian and Sethuraman Panchanathan
Center for Cognitive Ubiquitous Computing (CUbiC), Arizona State University
(schakr10, vineeth.nb, panch)@asu.edu

Abstract

Active learning techniques have gained popularity to reduce human effort in labeling data instances for inducing a classifier. When faced with large amounts of unlabeled data, such algorithms automatically identify the exemplar and representative instances to be selected for manual annotation. More recently, there have been attempts towards a batch mode form of active learning, where a batch of data points is simultaneously selected from an unlabeled set. Real-world applications require adaptive approaches for batch selection in active learning. However, existing work in this field has primarily been heuristic and static. In this work, we propose a novel optimization-based framework for dynamic batch mode active learning, where the batch size as well as the selection criteria are combined in a single formulation. The solution procedure has the same computational complexity as existing state-of-the-art static batch mode active learning techniques. Our results on four challenging biometric datasets portray the efficacy of the proposed framework and also certify the potential of this approach in being used for real world biometric recognition applications.

1. Introduction

The application of learning frameworks in real-world contexts necessarily requires a large amount of labeled data in the training phase. The rapid escalation of technology and the widespread emergence of modern technological equipments have resulted in the generation of humungous amounts of digital data. However, while gathering vast quantities of digital data is cheap and easy, annotating them with class labels entails significant human labor. This has set the stage for research in the field of active learning.

In addition to the large quantities of data that are generated each day (for example, YouTube videos), the presence of multiple labeling agents (for example, the vast consumer population) necessitates a scheme to simultaneously select and learn from multiple data points. To address this need, active learning techniques, which attempt to select a batch of data points at one shot from an unlabeled set, have been proposed in recent years. Sample applications of such batch mode active learning (BMAL) techniques include content based image retrieval (CBIR) [10], medical image classification [12] and text classification [11].

BMAL algorithms are of paramount importance in applications involving video data. Modern video cameras have a high frame rate and consequently, the captured data has high redundancy. Selecting batches of relevant frames from a superfluous frame sequence in captured videos is a significant and valuable challenge. Due to its wide usage, we focus on face based biometric recognition systems as the exemplar application in this paper to explain our framework. Although validated on biometric data, the proposed framework is generic and can be used in any application where it is required to select a number of representative entities simultaneously from repetitious samples.

An ideal BMAL system can be conceptualized as consisting of two main steps: (i) deciding the batch size (number of data points to be queried from a given unlabeled set of points) and (ii) selecting the most appropriate data points from the unlabeled pool once the batch size has been determined. Both these steps are critical in ensuring maximum generalization capability of the learner with minimum human labeling effort, which is the primary objective in any active learning application. However, the existing few efforts on batch mode active learning (see Section 2) focus only on the second step of identifying a criteria for selecting informative batches of data samples and require the batch size to be specified in advance by the user. In an application like face based biometric recognition, deciding on the batch size (number of relevant frames in a video) in advance and without any knowledge of the data stream being analyzed, is impractical. The batch size should depend on the quality and variability of the images in the unlabeled stream and also on the level of confidence of the current classifier on the unlabeled images.

In this paper, we propose a novel strategy for batch mode active learning, which adaptively selects samples based on the particular data stream being analyzed. We exploit numerical optimization based techniques to simultaneously

decide the batch size as well as identify the informative data points for manual annotation, through a single framework. Our method has the same computational complexity as state-of-the-art static BMAL technique, where the batch size is pre-specified by the user.

The rest of the paper is organized as follows: in Section 2, we present existing work on active learning; Section 3 details the mathematical formulation of our approach, along with an intuitive unsupervised learning technique for dynamic batch size selection for comparison of performance; the results of our experiments are presented in Section 4; and we conclude with discussions in Section 5.

2. Related Work

Active learning methods can be broadly categorized as *online* and *pool-based*. In online active learning, the learner encounters the data points sequentially over time and at each instant it needs to decide whether the current point has to be queried for its class label [1],[9],[20]. In contrast, in pool-based active learning, the learner is exposed to a pool of unlabeled data points and it iteratively selects instances for manual annotation.

Pool-based methods can be sub-categorized as *serial query based*, where a single point is queried at a time and *batch mode*, where a batch of points is queried simultaneously before updating the classifier. Majority of the active learning techniques have been applied in the serial query-based setting and can be divided into 4 categories - (i) SVM based approaches, which decide the next point to be queried based on its distance from the hyperplane in the feature space [25], (ii) Statistical approaches, which query points such that some statistical property of the future learner (eg the learner variance) is optimized [4], (iii) Query by Committee, which chooses points to be queried based on the level of disagreement among an ensemble of classifiers [2],[6], [14] and (iv) Other miscellaneous approaches [19].

Amidst batch mode (BMAL) techniques, Brinker [3] proposed a strategy which queried a diverse batch of points where diversity was measured as the angle between the hyperplane of the selected point to all the other hyperplanes of the already selected points. Hoi *et al.*[10], [11],[12] used the Fischer information matrix as a measure of model uncertainty and proposed to select a batch of points which reduced the Fischer information. Guo and Schuurmans [8] formalized the problem by defining an objective function and selecting a set of unlabeled points which optimized the value of that function. This approach has a well-defined mathematical basis as compared to the other heuristic techniques and was found to be the best performing BMAL scheme so far.

All the aforementioned techniques of batch mode active learning, including [8], concentrate only on the development of a selection criteria assuming the batch size is chosen by the user in advance. In an application like face-based biometric recognition, this is not a practical assumption. We would expect the number of relevant frames to be large when the active learner is exposed to an unlabeled video containing many new identities unknown to the learner, and the number to be low when the unlabeled video contains images similar to the training data. Thus, there is a strong need for the active learner to adapt to different contexts and dynamically decide the batch size as well as the specific instances to be queried. In this paper, we propose an optimization technique to address this issue. The strategy is similar to the work of Guo and Schuurmans [8], which however, has a different objective and is restricted to static scenarios where the batch size is user specified. With the same computational complexity as [8], we simultaneously solve for both the batch size and the specific points to be selected for a given unlabeled pool. We now describe the mathematical formulation of our approach.

3. Dynamic Batch Mode Active Learning: Mathematical Formulation

3.1. Optimization based Dynamic Batch Mode Active Learning

Consider a BMAL problem which has a current labeled set L_t and a current classifier w^t trained on L_t. The classifier is exposed to an unlabeled video U_t at time t. The objective is to select a batch B from the unlabeled stream in such a way that the classifier w^{t+1}, at time $t+1$, trained on $L_t \cup B$ has maximum generalization capability. An efficient method to judge the generalization capability of the updated learner is to compute its entropy on the remaining set of $U_t - B$ images after batch selection (given that future data is unknown). To ensure high generalization power of the future learner, we need to minimze the entropy of the updated learner on the remaining $|U_t - B|$ images.

From a data geometry point of view, it is possible that an objective function with only the entropy criterion will select images from high-density regions in the space of the unlabeled data points. This is because, the set of $U_t - B$ images may be dominated by samples from such high-density regions constituting a large portion of the data. However, considering the specific challenges of face based biometrics, we would like to ensure that our learner, in addition to learning from frames in high-density regions, also learns from informative visages made briefly by the subjects (eg a sudden smile or a sudden eyebrow raise). These images lie away from the main body of points, possibly in low density regions. To address this issue, we impose a condition in the objective function which selects images from low-density regions in the data space, i.e. images that have a high distance from the remaining set.

Let C denote the total number of classes and ρ_j denote

the average Euclidean distance of an unlabeled image x_j from other images in the video U_t. Greater values of ρ_j denote that the point is located in a low-density region. The two conditions mentioned previously can be satisfied by defining a score function as follows:

$$f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) \quad (1)$$

The first term denotes the sum of the average distances of each selected point from other points in the unlabeled video, while the second term quantifies the sum of the entropies of the learner on each remaining point in the unlabeled stream. λ_1 is a tradeoff parameter.

The problem therefore reduces to selecting a batch B of unlabeled images which produces the maximum score f(B). Let the batch size (number of images to be selected for annotation) be denoted by m, which is an unknown. Since there is no restriction on the batch size m, the obvious solution to this problem is to select *all* the images in the unlabeled video, leaving no image behind. Then, the entropy term becomes 0, and the density term attains its maximum value. Consequently, f(B) will also attain its maximum score. However, querying all the images for their class labels is not an elegant solution and defeats the basic purpose of active learning. To prevent this, we modify the score function by enforcing a penalty on the batch size as follows:

$$\tilde{f}(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) - \lambda_2 m \quad (2)$$

The third term essentially reflects the cost associated with labeling the images, as the value of the objective function decreases with every single image that needs to be labeled. The extent of labeling penalty can be controlled through the weighting parameter λ_2. Defining the score function in this way ensures that any and every image is not queried for its class label. Only images for which the density and entropy terms outweigh the labeling cost term, get selected.

We therefore need to select a batch B of unlabeled images so as to maximize $\tilde{f}(B)$. Since brute force search methods are prohibitive, we employ numerical optimization techniques to solve this problem. We define a binary vector M of size $|U_t|$ where each entry denotes whether the corresponding point is to be queried for its class label. We rewrite the objective function in Equation 2 into an equivalent function in terms of the defined vector M:

$$\max_{M,m} \sum_{j \in U_t} \rho_j M_j - \lambda_1 \sum_{j \in U_t} (1 - M_j) S(y|x_j, w^{t+1}) - \lambda_2 m \quad (3)$$

subject to the constraint:

$$M_j \in [0, 1] \quad (4)$$

In this formulation, note that if an entry of M is 1, the corresponding image will be selected for annotation and if it is 0, the image will not be selected. The number of images to be selected, is therefore equal to the number of non-zero entries in the vector M, or the zero-norm of M. Hence,

$$m = ||M||_0 \approx ||M||_1 = \sum_j M_j \quad (5)$$

Here, we have replaced the zero norm of M by its tightest convex approximation, which is the one-norm of M (similar to [26]). Also, from constraint 4, the one-norm is simply the sum of the elements of the vector M. Substituting m in terms of M, the formulation becomes:

$$\max_M \sum_{j \in U_t} \rho_j M_j - \lambda_1 \sum_{j \in U_t} (1 - M_j) S(y|x_j, w^{t+1}) - \lambda_2 \sum_j M_j$$

subject to the constraint: $M_j \in [0, 1]$. The above optimization is an integer programming problem and is NP hard. We therefore relax the constraint to make it a continuous optimization problem:

$$\max_M \sum_{j \in U_t} \rho_j M_j - \lambda_1 \sum_{j \in U_t} (1 - M_j) S(y|x_j, w^{t+1}) - \lambda_2 \sum_j M_j \quad (6)$$

subject to the constraint: $0 \leq M_j \leq 1$.

This problem is solved using the Quasi Newton method [21] [1]. The final value of M is used to govern the number of points and the specific points to be selected for the given data stream (by greedily setting the top m entries in M as 1 to recover the integer solution, where $m = \sum_j M_j$). Hence, solving a single optimization problem helps in dynamically deciding the batch size and selecting the specific points for manual labeling.

While the proposed framework combines batch size and data sample selection in a single formulation, it is also possible to think of an intuitive approach to solve this problem using a clustering-based batch size selection step, followed by application of a traditional static BMAL algorithm (such as [8]). For purposes of comparison of performance, we present below an alternative clustering-based approach for selecting the batch size in the latter case.

3.2. Clustering-based Batch Size Selection: An Alternative Approach

An obvious strategy to decide the batch size is to use a clustering algorithm to segment the images in the unlabeled video stream into relatively pure clusters (in terms of class labels) followed by a method to compute the batch size. Since the number of subjects (and hence, the number of clusters) in the data stream is an unknown, we need to exploit the spatial distribution of the unlabeled points for clustering (and cannot use algorithms like k-means which require the number of clusters as an input). This motivates the application of the DBSCAN algorithm (which can automatically determine the number of clusters for a given set of points) to isolate high density regions as separate clusters. For details about this method, please refer [23]. Our initial

[1] For details about the solution process and the approximation of the future unknown classifier w^{t+1}, please refer to the Supplemental File

experiments (not presented here for brevity) confirmed the efficacy of DBSCAN in isolating images of different subjects into separate clusters.

The Silhouette Coefficient (based on the cohesion and separation measures of a cluster) is a natural choice to decide the number of points to be queried from each cluster. It can attain a maximum value of 1, where a high value denotes a compact and well separated cluster. Intuitively, we would like to select few points for a compact and well-separated cluster and more points otherwise. Thus, the number of points to be selected from a cluster should be proportional to (1 - the Silhouette coefficient). Also, we would like to select more points from larger clusters. If m is the total number of points, m_i is the number of points in cluster i, SC_i is the Silhouette coefficient of cluster i and C is a constant, the number of points to be selected from cluster i can thus be defined as:

$$N_i = C * \frac{m_i}{m} * (1 - SC_i) \quad (7)$$

This operation is performed for each of the identified clusters to compute the number of points to be selected (the sum of the values obtained across all clusters provides the overall batch size). The dynamically computed batch size for each cluster can now be passed as an input to any standard static BMAL procedure for selecting the required number of points from the corresponding cluster. In our experiments, we used the state-of-the-art technique proposed in [8] to decide the specific points to be queried from a cluster, where we replaced the terms in the objective function with the density and entropy terms, similar to our formulation, for fair comparison.

4. Experiments and Results

Our study consisted of three experiments to validate the efficiency of our framework. Using preliminary experiments, the parameters λ_1 and λ_2 were empirically set to 1 and C to 50 in the study. The entropy term in the objective function necessitates a classifier which can provide probability estimates of an unlabeled point with respect to all classes. So, we used the Gaussian Mixture Models (GMMs) as the classifier in our experiments. GMMs have been successfully used in face recognition [13] in earlier work.

4.1. Datasets

We used four challenging biometric datasets in our different experiments:

(i) The VidTIMIT dataset [22], which contains video recordings of subjects reciting short sentences under unconstrained natural conditions.

(ii) The MOBIO dataset [18], which was recently created for the MOBIO (Mobile Biometry) challenge to test state-of-the-art face and speech recognition algorithms. It contains recordings of subjects under challenging real world conditions, captured using a hand-held device.

(iii) The MBGC (Multiple Biometric Grand Challenge) dataset [24], collected by the National Institute of Standards and Technology (NIST), which is the leading dataset to test commercial biometric recognition algorithms and contains video recordings of subjects under uncontrolled indoor and outdoor lighting.

(iv) The FacePix dataset (*www.facepix.org*), which contains 181 (-90 degree to 90 degree) pose images of each of 30 subjects in one degree increments. It also contains frontal images of each subject under varying illumination, where a spotlight was moved in one degree increments. The dataset has been used to study the effects of varying poses and illumination angles in face recognition [15].

The VidTIMIT, MOBIO and MBGC datasets represent videos captured under different real world settings (stationary, using handheld device and under uncontrolled lighting respectively). FacePix contains calibrated measurements of pose and illumination variations, which were useful to study the efficacy of our framework. [2]

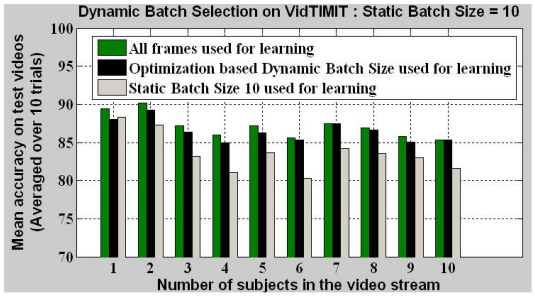
4.2. Experiment 1: Dynamic vs Static BMAL

The purpose of this experiment was to demonstrate the efficacy of dynamic batch selection over static selection in applications like face recognition. The VidTIMIT and the MOBIO biometric datasets were used in this experiment. 25 subjects were randomly selected from each dataset. Our preliminary experiments (not presented here due to lack of space) confirmed that the Discrete Cosine Transform (DCT) feature could effectively differentiate the subjects and hence was used in this experiment (for details about the feature extraction process, please refer [5]). The feature extraction step was followed by PCA to reduce the dimension.

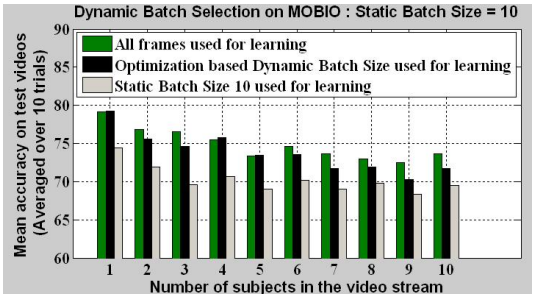
A classifier was induced with 1 training video of each of the 25 subjects, used in this experiment. Unlabeled video streams were then presented to the learner. To demonstrate the generalizability with different subject combinations, the number of subjects in each unlabeled stream was varied between 1 and 10. For each stream, the batch size and the specific points were selected simultaneously using the proposed optimization strategy (Equation 6). The classifier was updated with the selected points and tested on test videos containing the same subject(s) as in the corresponding unlabeled videos.

To illustrate the usefulness of dynamic batch size selection, the accuracy was compared against the case when *all the frames* in the unlabeled video were used for learning and also when the batch size was static and predetermined. The static batch size was selected as 10 (the effect of this parameter is studied later) and the optimization scheme (as

[2] Our purpose was to test the performance of active learning and so, for the MBGC and MOBIO datasets, we did not follow the protocols specified in the actual challenge which were intended for face recognition.



(a) VidTIMIT dataset



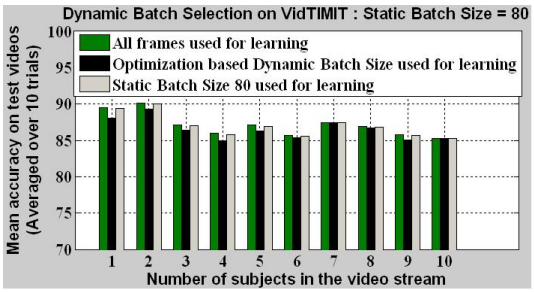
(b) MOBIO dataset

Figure 1. Dynamic vs Static BMAL on 100 unlabeled video streams from the VidTIMIT and MOBIO datasets (static batch size = 10).

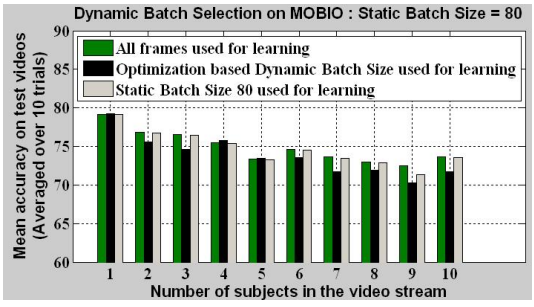
Number of subjects in the video stream	1	2	3	4	5	6	7	8	9	10
VidTIMIT dataset	58%	68.1%	55.58%	45.46%	51.11%	35.71%	35.88%	41.38%	51.93%	47.43%
MOBIO dataset	54.4%	48.4%	45.4%	45.4%	45.1%	46.8%	46.5%	46.8%	46.4%	44.6%

Table 1. Mean percent increment in labeling cost using static selection with batch size 80 against optimization based dynamic selection.

outlined in Section 3.2) was used to select the 10 points, for fair comparison. The results are shown in Figure 1 and are averaged over 10 runs to rule out effects of randomness. We see that, in both datasets, the accuracy obtained with dynamic batch selection very closely matches that obtained when trained on all the frames. This emphasizes the efficiency of the framework to properly identify a batch size and the specific points so that the resulting classifier is comparable to the one trained using all the images. We also note that the classifier obtained when the batch size is static and pre-determined does not attain good generalization capability compared to dynamic selection.



(a) VidTIMIT dataset



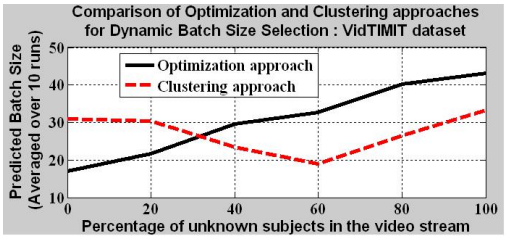
(b) MOBIO dataset

Figure 2. Dynamic vs Static BMAL on 100 unlabeled video streams from VidTIMIT and MOBIO (static batch size = 80).

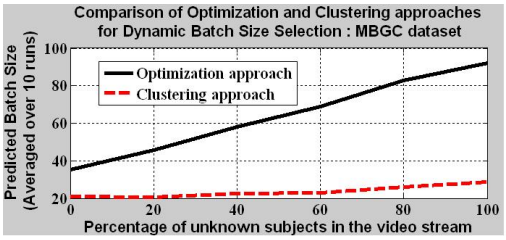
In general, we can expect that if we select a greater number of images from an unlabeled set, the updated learner will perform better on a test set containing the same subjects. Thus, if we select a higher value of the batch size in a static BMAL learner, then the selection is expected to perform better than in Figure 1. This is depicted in Figure 2 where the static batch size was taken as 80 instead of 10. We see that the static selection performs almost as well as the learner obtained when trained on all frames. However, to achieve this performance, the static selection required a significantly greater number of images to be labeled than dynamic selection. Table 1 shows the mean percentage increment in the number of images that had to be labeled using the static selection with batch size 80 against optimization based dynamic selection. It is evident that for both the datasets, the static framework required a much greater number of images to be labeled to marginally outweigh dynamic selection. Hence, by selecting a number at random, the static batch selection strategy can sometimes query too few points leading to poor generalization power of the updated learner, while in some cases it can entail considerable labeling cost to attain an insignificant increment in accuracy. The dynamic selection strategy, on the other hand, computes the batch size by exploiting the level of confidence of the future learner on the images in the current unlabeled video and thus provides a concrete basis to decide the batch size.

4.3. Experiment 2: Proposed Dynamic BMAL vs Clustering-based Dynamic BMAL

Having demonstrated the superiority of dynamic batch size selection over static selection, we performed a comparative study of the proposed optimization framework with the two step process of clustering followed by static BMAL



(a) Experiment with unknown subjects from the VidTIMIT dataset



(b) Experiment with unknown subjects from the MBGC dataset

Figure 3. Comparison of Proposed and Clustering-based batch size selection on the VidTIMIT and MBGC datasets.

Proportion of new identities	0%	20%	40%	60%	80%	100%
Accuracy using proposed approach	87.1%	79.9%	82.8%	84.6%	86.6%	81.8%
Accuracy using clustering approach	84.1%	68.4%	61%	53.4%	63.8%	63.9%

Table 2. Test set accuracies using Proposed and Clustering based BMAL on the MBGC dataset with increasing proportions of new identities.

(Section 3.2), for dynamic batch selection. We used the VidTIMIT and MBGC datasets for this experiment. Contrary to the previous experiment, where all the 25 subjects were present in the training set, the subjects in this experiment were divided into two groups - a "known" group containing 20 subjects and an "unknown" group containing the remaining 5 subjects. A classifier was induced with 1 video of each of the known subjects. Unlabeled video streams were then presented to the learner and the batch size decided by the two schemes were noted. The proportion of unknown subjects in the unlabeled video was gradually increased from 0% (where all the subjects in the unlabeled video were from the training set) to 100% (where none of the subjects in the unlabeled video were present in the training set) in steps of 20%. The learner was not given any information about the composition of the video stream. Also, the size of each video stream was kept the same to facilitate fair comparison. The DCT feature, followed by PCA, was used again.

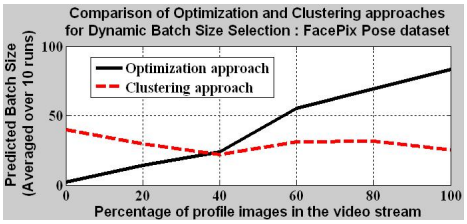
The results of the aforementioned experiment (averaged over 10 trials) are shown in Figure 3. The x-axis denotes the percentage of atypical images in the unlabeled pool and the y-axis denotes the batch size predicted using both the proposed and clustering-based strategies. We note that in both the experiments, as the proportion of salient images in the unlabeled stream increases, the uncertainty term outweighs the cost term in Equation 6 and the proposed algorithm decides on a larger batch size. This matches our intuition because, with growing percentages of atypical images in the video stream, the confidence of the learner on those images decreases and so it needs to query more images to attain good generalization capability. The clustering based scheme, on the other hand, does not consider the training set in deciding the batch size and so, it fails to reflect the uncertainty of the classifier. The batch size, therefore, does not bear any specific trend to the percentage of atypical images in the unlabeled set. Thus, while the clustering scheme decides the number of points to be queried based on a score computed from the spatial distribution of the unlabeled points, the optimization based technique provides a more logical ground to decide the batch size by considering the performance of the updated learner.

Besides the predicted batch size, it is equally important to analyze the accuracy obtained on test sets with similar compositions as the unlabeled videos. Since the proposed scheme appropriately reflects the uncertainty of the learner and queries points accordingly, it is expected to have a better accuracy on test videos as compared to the clustering technique. This is confirmed in Table 2 which shows the accuracy obtained on test videos from the MBGC dataset using the two strategies. It is evident that the proposed scheme achieved significantly better generalization as compared to the clustering based approach with varying proportions of new identities in the unlabeled stream. The result on the VidTIMIT dataset was similar and is not presented due to lack of space.

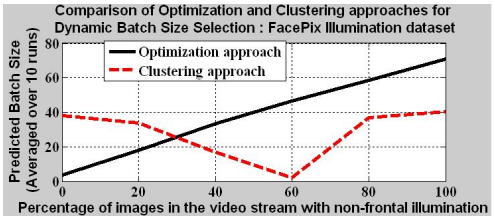
To further demonstrate the usefulness of the proposed approach in batch size selection under changing conditions, we conducted experiments with unlabeled videos containing images with different poses and illumination conditions compared to the training images. These are detailed below:

Presence of images with unknown pose angles: The FacePix dataset was used in this experiment. The training set contained frontal images (-10 degree to 10 degree) of 25 randomly chosen subjects. Unlabeled sets of images of the same 25 subjects were presented to the learner and the percentage of profile images (-45 degree to -90 degree and 45 degree to 90 degree) was gradually increased from 0% (where all the unlabeled images were frontal) to 100% (where the unlabeled video contained only profile images) in steps of 20%. The Gabor feature was used here (as in [7]) and PCA was used to reduce the dimension.

Presence of images under unknown illumination: The FacePix dataset was used in this experiment also. As before, the training set contained images of 25 subjects where



(a) Experiment with varying poses from the FacePix dataset



(b) Experiment with varying illumination from the FacePix dataset

Figure 4. Comparison of Proposed and Clustering-based batch size selection on the FacePix dataset.

the illumination angle was -10 degree to 10 degree. Unlabeled images of the same subjects were presented to the learner and the percentage of images where the angle of illumination varied between -45 degree to -90 degree and 45 degree to 90 degree was gradually increased from 0% to 100% in steps of 20%. The Gabor feature, followed by PCA, was used in this experiment (as used in [16]). The results are shown in Figure 4 and further corroborate the conclusions drawn in the previous experiment.

4.4. Experiment 3: Active Learning Performance

Here, we study the practical applicability of the end-to-end system by analyzing its performance under real world settings. The VidTIMIT and the MOBIO datasets, representing challenging real-world conditions, were used in this experiment. A classifier was induced with 1 training video of each of 25 randomly chosen subjects. Unlabeled video streams (each containing about 250 frames) were then presented to the classifier sequentially. The images in the video streams were randomly chosen from all 25 subjects and did not have any particular proportion of subjects in them, to mimic general real-world conditions. For each video, optimization based dynamic BMAL was used to query images. The selected images were appended to the training set, the classifier updated and then tested on a test video containing about 5000 images spanning all the 25 subjects.

The proposed approach was compared with three heuristic BMAL schemes - (i) Random Sampling, (ii) Diversity based BMAL, as proposed by Brinker [3] and (iii) Uncertainty Based Ranked Selection, where the top k uncertain points were queried from the unlabeled video, k being the batch size. For each video stream, the dynamically computed batch size was noted and used for the corresponding unlabeled video in each of the heuristic techniques, for fair comparison. The performance was also compared against the two step process of clustering followed by static BMAL (Section 3.2).

The label complexity (number of batches of labeled examples needed to achieve a certain level of accuracy) was used as the metric for quantifying performance in this experiment. The average time taken by each approach, to query images from an unlabeled stream, was also noted. The results are shown in Table 3. As evident from the running time figures, the proposed approach is computationally intensive compared to the heuristic BMAL techniques. However, the label complexity values, to attain a test accuracy of 85%, is markedly lower for the proposed approach. This asserts the fact that the proposed scheme succeeds in selecting the salient and prototypical data points as compared to the heuristic approaches and attains a given level of accuracy with significantly reduced human labeling effort. The clustering scheme followed by the optimization framework achieves comparable label complexity as our approach. However, it is a two step process and therefore involves more computation than our approach which is depicted in the running time values. [3]

5. Conclusions

In this work, we proposed a novel optimization scheme for dynamic batch mode active learning. Our framework solves for the batch size as well as the specific points to be queried through a single formulation and has the same computational complexity as the state of the art static BMAL algorithm [8]. The results showed immense promise in using the proposed approach in real-world batch mode active learning applications. The proposed algorithm is flexible and it is straightforward to extend it for dynamic batch selection in situations where multiple sources of information (eg. audio and video data) are available. For instance, Equation 6 can be modified by appending relevant terms from the respective sources, together with a penalty on the batch size:

$$\max_{M} \sum_{j \in U_{t1}} \rho_j M_j - \sum_{j \in U_{t1}} (1 - M_j) S(y|x_j, w^{t+1}) + \sum_{j \in U_{t2}} \rho_j M_j - \sum_{j \in U_{t2}} (1 - M_j) S(y|x_j, w^{t+1}) - \sum_{j} M_j$$

Moreover, if contextual information is available (eg location of a subject, whether at home or in office), it can be used to

[3] A visual illustration of the images selected from a given video, using each of the techniques, is shown in Appendix B of the Supplemental File. The FacePix dataset was chosen to demonstrate the effectiveness of the framework under pose and illumination variations.

	VidTIMIT Label Complexity	VidTIMIT Time(seconds)	MOBIO Label Complexity	MOBIO Time(seconds)
Proposed Approach	8.67	105.66	20.67	157.67
Diversity based BMAL	27.67	1.3	63.33	1.38
Uncertainty based BMAL	23.67	13.98	44.67	21.46
Random Sampling	31.33	0.01	61.67	0.01
Clustering based BMAL	11.33	122.11	22.67	174.28

Table 3. Number of batches of labeled images required to achieve 85% accuracy and the time taken (in seconds) to query a batch of images from an unlabeled pool with 250 images. The results have been averaged over 3 runs with different orders of the unlabeled video streams

construct a prior probability vector depicting the chances of seeing particular acquaintances in a given context. The entropy term can then be computed on the posterior probabilities obtained by multiplying the likelihood values returned by the classifier with the context aware prior. Thus, subjects not expected in a given context (eg. a home acquaintance in an office setting) will have low priors and consequently, the corresponding posteriors will not contribute much in the entropy calculation. The framework can therefore be extended to perform context-aware adaptive batch selection. Our preliminary experiments in these directions have shown promising results.

Our future work will mainly include handling scaling issues of the proposed algorithm; the quadratic programming problem which needs to be solved as a part of the optimization procedure can be a bottleneck in dealing with large scale data. However, there have been recent efforts [17] to efficiently solve QP problems by using a pivoting algorithm and the KKT conditions to significantly reduce computations. This can be judiciously used in our approach, making it feasible and meritorious even for large-scale data.

References

[1] V. Balasubramanian, S. Chakraborty, and S. Panchanathan. Generalized query by transduction for online active learning. In *OLCV Workshop at ICCV*, 2009. 2650

[2] Y. Baram, R. El-Yaniv, and K. Luz. Online choice of active learning algorithms. *JMLR*, 5, 2004. 2650

[3] K. Brinker. Incorporating diversity in active learning with support vector machines. *ICML*, 2003. 2650, 2655

[4] D. Cohn, Z. Ghahramani, and M. Jordan. Active learning with statistical models. *JAIR*, 1996. 2650

[5] H. Ekenel, M. Fischer, Q. Jin, and R. Stiefelhagen. Multi-modal person identification in a smart environment. In *IEEE CVPR*, 2007. 2652

[6] Y. Freund, S. Seung, E. Shamir, and N. Tishby. Selective sampling using the query by committee algorithm. *Machine Learning.*, 1997. 2650

[7] B. Gokberk, L. Akarun, and E. Alpaydin. Feature selection for pose invariant face recognition. In *IEEE ICPR*, 2002. 2654

[8] Y. Guo and D. Schuurmans. Discriminative batch mode active learning. In *NIPS*, 2008. 2650, 2651, 2652, 2655

[9] S. Ho and H. Wechsler. Query by transduction. *IEEE TPAMI*, 2008. 2650

[10] S. Hoi, R. Jin, J. Zhu, and M. Lyu. Semi-supervised SVM batch mode active learning for image retrieval. In *IEEE CVPR*, 2008. 2649, 2650

[11] S. C. H. Hoi, R. Jin, and M. R. Lyu. Large-scale text categorization by batch mode active learning. In *International Conference on World Wide Web*. ACM, 2006. 2649, 2650

[12] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu. Batch mode active learning and its application to medical image classification. In *ICML*, 2006. 2649, 2650

[13] J. Y. Kim, D. Y. Ko, and S. Y. Na. Implementation and enhancement of GMM face recognition systems using flatness measure. In *Robot and Human Interactive Communication*, 2004. 2652

[14] R. Liere and P. Tadepalli. Active learning with committees for text categorization. *ICAI*, 1997. 2650

[15] G. Little, S. Krishna, J. Black, and S. Panchanathan. A methodology for evaluating robustness of face recognition algorithms with respect to changes in pose and illumination angle. In *ICASSP*, 2005. 2652

[16] D. H. Liu, K. M. Lam, and L. S. Shen. Illumination invariant face recognition. In *Pattern Recognition*, 2005. 2655

[17] Y. Liu and Z. Zhang. A fast algorithm for linearly constrained quadratic programming problems with lower and upper bounds. In *International Conference on Multimedia and Information Technology*, 2008. 2656

[18] S. Marcel, C. McCool, and P. Matejka. Mobile biometry (mobio) face and speaker verification evaluation. *Idiap Research Institute, Technical Report*, 2010. 2652

[19] A. McCallum and K. Nigam. Employing EM and Pool-Based active learning for text classification. *ICML*, 1998. 2650

[20] C. Monteleoni and M. Kaariainen. Practical online active learning for classification. In *IEEE CVPR*, 2007. 2650

[21] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, 1999. 2651, 2657

[22] C. Sanderson. *Biometric Person Recognition: Face, Speech and Fusion*. VDM Verlag, June 2008. 2652

[23] P. Tan, M. Steinbach, and V. Kumar. Introduction to data mining. 2006. 2651

[24] M. Tistarelli and M. Nixon. Advances in biometrics: Icb. *SpringerLink*, 2009. 2652

[25] S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *JMLR*, 2000. 2650

[26] J. Weston, A. Elisseeff, B. Scholkopf, and M. Tipping. Use of the zero norm with linear models and kernel methods. In *JMLR*, 2003. 2651