A Generalised Formulation for Collaborative Representation of Image Patches (GP-CRC)

Tapabrata Chakraborti[1]
tapabrata@cs.otago.ac.nz
Brendan McCane[1]
mccane@cs.otago.ac.nz
Steven Mills[1]
steven@cs.otago.ac.nz
Umapada Pal[2]
umapada@isical.ac.in

[1] Computer Science Department
University of Otago
Dunedin, NZ

[2] CVPR Unit
Indian Statistical Institute
Kolkata, India

Abstract

Collaborative Representation based Categorization (CRC) represents query samples collaboratively as an optimal weighted average of training samples across all classes. It has been shown to be effective for recognition problems, but its performance degrades in presence of high variation in image background. We present a generalization mathematical reformulation of a patch based CRC approach. The proposed method (GP-CRC) analytically overcomes the problem in the cost function itself and provides a closed form solution. Experiments are carried out on two face recognition (AR and LFW) and two species recognition (Oxford-102 Flowers and Oxford-IIIT Pets) benchmarks. The proposed method outperforms the original CRC as well as basic patch based CRC consistently across all the datasets (with statistical significance in majority of the cases) and comparable or marginally higher accuracy than the state of the art probabilistic CRC. It is also demonstrated experimentally that our method is more robust to background variations than its competitors.

1 Introduction

Collaborative Representation based Categorization (CRC) [17] was introduced and applied originally to the face recognition problem. CRC represents the test image as an optimal weighted average of all training images across classes. The weights are optimised analytically, so as to generate least residual for the predicted class. Thus in the CRC paradigm, the samples of the entire space co-operate irrespective of class to generate a suitable weight distribution for optimal classification. In this regard, collaborative representation is quite novel compared to the traditional discriminative recognition approach. Collaborative representation has the additional advantage of being both a feature representation and classification framework. Thus any descriptor(s) or ensemble thereof may be fed to the CRC pipeline for representation and classification. Also, CRC has a closed solution rather than taking an iterative, heuristic or empirical approach.

Since its first introduction, CRC based methods have performed consistently well in face recognition problems [18]. This is because human faces have subtle inter-class differences and significant similarities between classes. The fact that CRC collaboratively uses cross class samples to reconstruct the query, takes advantage of these similarities and improves robustness of representation and hence classification.

However, most of the experiments have been on traditional face recognition benchmarks having well aligned face focussed images with minimal background inclusion or variation. Even the few works which have used face datasets having humans in natural scene backgrounds, like the Labeled Faces in the Wild (LFW) dataset [7], have mostly employed pre-processing steps to align the face and then crop it out, thus removing the effect of the natural setting.

A possible way to overcome this challenge, without resorting to segmentation, is to use image patches of suitable size to predict the output class through majority voting. Instead of using the whole image, patches are used for training and testing, to reduce the effect of background variation. This however, makes the underlying assumption that the major portions of the images are foreground information, that is, the face to be recognized. If the majority of the image is background, this patch-based approach would still not be completely effective, as it will affect the majority decision. Similar problems will occur if the test images have new or rare backgrounds.

Zhu *et al*. [18] proposed a patch based CRC method (PCRC) which achieved good results on the early face recognition datasets (like AR [11] benchmark) as well as a cropped and aligned version of LFW dataset, LFW-a [16]. But our experiments with PCRC on the original LFW [7] show considerable degradation in performance.

We present a new and generalised formulation for patch based collaborative representation (GP-CRC) of images for robust classification against background outliers. This is incorporated analytically in the cost function itself and experiments have yielded encouraging results on two face recognition datasets (AR [11] and LFW [7]) and two species recognition datasets (Oxford 102 Flowers [13] and Oxford-IIIT Pets [15]). The reason for selecting species recognition as a candidate problem is because it particularly poses the problem of high background variation, and collaborative representation may be effective there as recently shown in [2]. The proposed GP-CRC outperforms the original patch based CRC (PCRC) and has comparable results with the state of the art Probabilistic CRC (ProCRC) [1] by Cai *et al*.

2 Methodology

This Section presents the CRC framework briefly as well as the original patch based method (PCRC). Then it goes into the details of the proposed method (GP-CRC) along with analytical proof and logical justification of its effectiveness.

2.1 Collaborative Representation based Categorization (CRC)

The mathematical framework for Collaborative Representation based Classification (CRC) [17] is described in brief here. Consider a training dataset with images in the feature space as $X = [X_1, \ldots, X_c] \in \mathbb{R}^{d \times N}$ where N is the total number of samples over c classes and d is the feature dimension per sample. Thus $X_i \in \mathbb{R}^{d \times n_i}$ is the feature space representation of class i with n_i samples such that $\sum_{i=1}^{c} n_i = N$.

The CRC model reconstructs a test image in the feature space $y \in \mathbb{R}^d$ as an optimal collaboration of all training samples, while at the same time limiting the size of the reconstruction parameters, using the Lagrangian multiplier λ.

The CRC cost function is given as

$$\hat{\alpha} = \arg\min_{\alpha} (\|y - X\alpha\|_2^2 + \lambda\|\alpha\|_2^2) \tag{1}$$

where $\hat{\alpha} = [\hat{\alpha}_1, \ldots, \hat{\alpha}_c] \in \mathbb{R}^N \mid \hat{\alpha}_i \in \mathbb{R}^{n_i}$ is the reconstruction matrix corresponding to class i. A least-squares derivation yields the optimal solution as

$$\hat{\alpha} = (X^T X + \lambda I)^{-1} X^T y \tag{2}$$

The representation residual of class i for test sample y can be calculated as:

$$r_i(y) = \frac{\|y - X_i\hat{\alpha}_i\|_2^2}{\|\hat{\alpha}_i\|_2^2} \; \forall i \in 1, \ldots, c \tag{3}$$

The final class of test sample y is thus given by

$$C(y) = \arg\min_{i} r_i(y) \tag{4}$$

The optimal value of λ, for which $G(\lambda)$ is minimum, is evaluated using gradient descent.

2.2 Patch based CRC (PCRC)

Zhu *et al*. [18] introduced a patch-based framework for collaborative representation (PCRC).

Let the query image y be divided into q overlapping patches $y = \{y_1, \ldots, y_q\}$. From the feature matrix X, a local feature matrix M_j is extracted corresponding to location of patch y_j. Thus the modified cost function becomes:

$$\hat{p}_j = \arg\min_{p_j} (\|y_j - M_j p_j\|_2^2 + \lambda\|p_j\|_2^2) \tag{5}$$

where $M_j = [M_{j1}, \ldots, M_{jc}]$ are the local dictionaries for the c classes and $\hat{p}_j = [\hat{p}_{j1}, \ldots, \hat{p}_{jc}]$ is the optimal reconstruction matrix for the j^{th} patch. The class of a patch in the test image is predicted as:

$$C(y_j) = \arg\min_{k} r_{jk}(y) \tag{6}$$

where

$$r_{jk} = \frac{\|y_j - M_{jk}\hat{p}_{jk}\|_2^2}{\|\hat{p}_{jk}\|_2^2} \; \forall i \in 1, \ldots, c \tag{7}$$

The classification of the entire test sample y is determined by majority voting of the classification labels of the patches y_j.

2.3 Generalised Patch based CRC (GP-CRC)

The proposed method addresses some of the inherent issues with the original patch based CRC formulation (PCRC). These have been touched upon in brief in the introduction, but here we deal with them in detail and provide concrete mathematical solutions. In essence, we develop a more generalised and robust patch based CRC, we call it GP-CRC.

Case 1: According to Equation 7, in PCRC, the dictionary is extracted at the same corresponding position of the test patch, which means that the test patch is only represented as a weighted collaboration of the training patches at the exact same location in the images. This is quite a restricted formulation and in fact, unless the foreground object in the image is well centred, aligned and covers most of the image, this location matching of patches will rarely occur. This disadvantage can affect the recognition rate to quite an extent.

Solution: We generate an augmented dictionary M considering all patches across all classes. p_j is the weight matrix for the image patch y_j and hence Mp_j is the representation of y_j as a weighted collaboration of all training image patches.

Thus the modified cost function becomes:

$$\hat{p}_j = \arg\min_{p_j} (||y_j - Mp_j||_2^2 + \lambda||p_j||_2^2) \tag{8}$$

Other symbols have same meaning as in PCRC, with dimensions of corresponding matrices being adjusted appropriately.

Case 2: If the test sample(s) happens to contain background patches that are significantly different (new or rare) from the majority of patches in the training set, then there is a high probability of wrong classification.

Solution: To address this we add an additional term to the cost function. M_j is a sub-matrix of M consisting of all the training image patches (can be in a transformed domain) at the same location j of the incoming test patch y_j in the test image y. p_{jj} is a sub-matrix of p_j constructed similarly and hence $M_j p_{jj}$ is the representation of y_j as a weighted collaboration of all training image patches which are at the same position j with respect to their corresponding full images.

We compare the representation of the current patch, $M_j p_{jj}$ to the overall representation of the dataset Mp_j and the distance $||Mp_j - M_j p_{jj}||_2^2$ gives a cost metric of how different the patch is from the overall nature of the image dataset. This reduces the effect of outlier background patches on the system.

Thus the modified cost function becomes:

$$\hat{p}_j = \arg\min_{p_j} (||y_j - Mp_j||_2^2 + \lambda||p_j||_2^2 + \gamma||Mp_j - M_j p_{jj}||_2^2) \tag{9}$$

Case 3: If the images of the dataset are such that significant portions of the images are background, then many of the generated patches will only contain background, which may confound the system.

Solution: In general, the patch based collaborative representation framework is quite robust against this challenge, due to the very fact that since the test image is represented as a co-operation across all samples, it strengthens the larger number of recurring objects, that is the foreground, rather than the changing background which occurs in fewer patches. So it performs better than purely discriminative approaches. In our formulation, the term introduced for Case 2 also strengthens this aspect further.

Derivation of closed form solution: Let M'_j be a matrix of same dimension as that of M with only the values in M_j being the non-zero elements in M'_j. Then $M_j p_{jj}$ may be replaced by $M'_j p_j$, the two expressions being equivalent. Let $\bar{M}_j = M - M'_j$. So we have,

$||Mp_j - M_j p_{jj}||_2^2 = ||Mp_j - M'_j p_j||_2^2 = ||(M - M'_j)p_j||_2^2 = ||\bar{M}_j p_j||_2^2 = \bar{M}_j^T \bar{M}_j ||p_j||_2^2$

Thus Equation 9 can be rewritten as

$$\hat{p}_j = \arg\min_{p_j} (||y_j - Mp_j||_2^2 + \lambda||p_j||_2^2 + \gamma \bar{M}_j^T \bar{M}_j ||p_j||_2^2) \tag{10}$$

Differentiating Equation 10 with respect to p_j and equating to zero to find $\hat{p}_j$, we have

$$-M^T(y_j - Mp_j) + \lambda p_j + \gamma \bar{M}_j^T \bar{M}_j p_j = 0 \tag{11}$$

Hence,

$$\hat{p}_j = (M^T M + \gamma \bar{M}_j^T \bar{M}_j + \lambda I)^{-1} M^T y_j \tag{12}$$

The projection matrix $T_j = (M^T M + \gamma \bar{M}_j^T \bar{M}_j + \lambda I)^{-1} M^T$ can be computed off-line and the solution $\hat{p}_j = T_j y_j$ can thus be computed efficiently in a closed form. The class of test sample is predicted similar to PCRC, but M_{jk} being replaced by M_k in Equation 7. Optimal values of γ and λ are found by gradient descent.

Summary: In earlier work only patches at same locations were compared, which only works well if the images are well aligned without much background variation. So, in the present work, instead of constructing M with patch features at same locations, we construct an augmented M with features of all patches over all training images (in Case 1). This solution handles the case of misaligned foreground objects, but raises the chances of the representation learning the background. To compensate, we have the second term of the cost function (in Case 2) which compares it to other location matched patches (Mj) in order to have a penalty if the query patch is too dissimilar to other patches at same location. This is like regularization, trading off misaligned foreground objects with the risk of learning the background- all integrated in the cost function.

3 Experimental Setup

3.1 Datasets

Experiments have been carried out on two face recognition datasets (AR and LFW) and two species recognition datasets (Oxford 102 Flowers and Oxford -IIIT Pets).

The face recognition benchmarks are chosen to compare directly with the CRC methods under inspection, which have reported results on these benchmarks in published articles. However, the AR dataset [11] has well aligned front facing face images with minimal background. The LFW dataset [7] does have human images from natural settings, but the reported experiments with CRC [18] were performed on aligned and cropped face images from the original datasets (LFW-a) [16]. Experiments are performed on both versions. To further investigate the effect of background variation on performance, experiments are also carried out on two popular species recognition benchmarks (Oxford 102 Flowers [13] and Oxford-IIIT Pets) [15].

AR Face Dataset: The AR dataset [11], developed at the Ohio State University, contains more than 400 color face images of 126 people with changes in illumination, emotion and occlusion. For comparison to patch based CRC (PCRC), as in [18], a subset of 50 male and 50 female subjects were chosen and the images are resized to 32×32 for our experiments.

LFW Face Dataset: The Labeled Faces in the Wild (LFW) dataset [7], by the University of Massachusetts Amherst, contains unconstrained images of 5749 individuals. LFW-a [16] is an front aligned variation of it, which is used in this work. The setup is constructed as in PCRC for direct comparison. 158 subjects having at least 10 sample images each are chosen. The images are resized to 121×121 pixels, keeping background information intact.

Oxford 102 Flowers Dataset: It has 8,189 images of 102 flowers, with at least 40 images per class. It was developed by the Robotics Group at Oxford University [13]. It

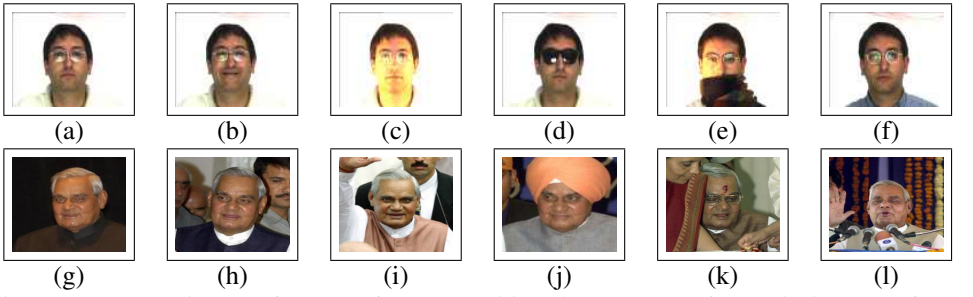

Figure 1: Sample images from the face recognition datasets: (a)-(f) are six images of one individual from the AR dataset, showing variations in color of clothing, facial expression, illumination, and partial covering of face; (g) to (l) are images of one individual from the LFW dataset, showing variation in clothing, headgear, other faces in image, varying background and partial covering of face.

is an expansion of the earlier dataset [17] by the same group with 17 flower types with 80 images per class.

Oxford IIIT Pets Dataset: This dataset [15] compiled by the Oxford Robotics Group and IIIT Hyderabad, consists of 37 categories of pet cats and dogs with around 200 images belonging to each class.

3.2 Features

For the face recognition experiments, to maintain similar experimental conditions with reported results for fair comparison, we have used the raw images as input. For species recognition, on the other hand a feature space transformation is more the norm, mainly to properly encode and exploit repeated local patterns, like in flowers.

So for the species recognition experiments, an ensemble of two popular global descriptors, GIST and HOG, are used for evaluation. Pixel-wise global descriptors are chosen so as to generate two dimensional feature matrices of same dimension as the input image. This is because the proposed method works on image patches, rather than feature vectors. The resulting GIST and HOG features together generate a Bag of Words and multi-scale Spatial Pyramid [9] per patch, for ensemble classification prediction.

GIST: Global Invariant Scale Transform [14] (here referred as GIST) is a global feature that describes the spatial envelope of the image using directional properties. It extracts dense multi-scale overlapping patches.

HOG: Histogram of Oriented Gradients [4] (HOG) are global descriptors and the features are extracted in 2×2 and 3×3 cells which are concatenated at each grid location to generate the descriptor.

3.3 Classifiers

Besides the proposed method (GP-CRC), a group of classifiers have been carefully chosen for meaningful evaluation of GP-CRC.

CRC based classifiers: A few of the available CRC based classifiers are employed based on the following selection criteria.

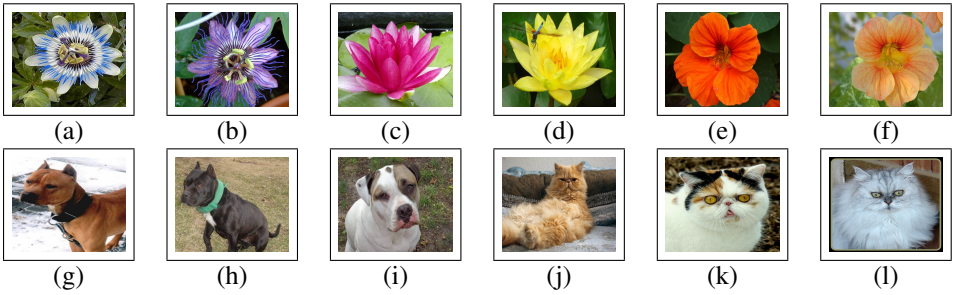

Figure 2: Sample images from the species recognition datasets: (a)-(f) are six images of different flowers from the Oxford 102 dataset showcasing the challenge of low inter-class variation; (g) to (l) are three images each of one breed/class of dogs and that of cats, showcasing the challenge of high intra-class variation.

Basic CRC: First, we consider the basic CRC method [17] for baseline performance reporting.

Original PCRC: The second is the original patch based method PCRC [18] to showcase the improvement over CRC and also to have the reference to compare our generalised patch based method (GP-CRC) with.

ProCRC: Third, we report results on the state of the art Probabilistic CRC (ProCRC) method [1] to compare the performance of the proposed scheme.

Non-CRC based Classifiers: It is also important to evaluate the accuracy of the CRC based strategies against a few standard non-CRC based classifiers. For this purpose, three popular non-linear classifiers have been chosen, namely SVM, AdaBoost and RDF.

Support Vector Machines: SVM [3] with χ^2 kernel have been employed. Multiclass categorization is performed with the binary SVM classifier in a one-versus-all fashion.

AdaBoost: AdaBoost [5] has two major stages: first a weak classifier and then a boosting part. The weak classifier generates a threshold optimised to one data dimensions in order to classify the data into two classes. The boosting part calls the classifier iteratively, and updates the weights of misclassified examples after each iteration. This creates a cascade of "weak classifiers" which act as a "strong classifier" in a repeated one-versus-all strategy.

Random Decision Forest: RDF [6] is an ensemble of Decision Trees. Since individual decision trees are prone to over-fitting, bootstrap aggregated (bagged) Classification and Regression Trees (CART) are used to achieve better generalisation.

4 Results and Discussions

For the face recognition datasets, raw images have been used as input to the classifiers to compare directly with results in the existing literature on CRC. For the AR benchmark, images have been resized to 32×32, while for the LFW-a benchmark the aligned face images are cropped to 121×121 to eliminate background and then resized to 32×32. Again these conditions are maintained to have parity with reported results in earlier articles on CRC. Experiments are also carried out on the original LFW images with just resizing to 121×121. 10×10 is used as patch size, which is same as that used in earlier work.

Sample images from the face recognition datasets are presented in Fig1. Fig1 (a)-(f) are six images of one individual from the AR dataset. If (a) is considered as the reference

image, then (b) shows change in expression, (c) has change in illumination, (d) and (e) has partial covering of face, that is eyes and mouth respectively, (f) has change of clothing. Fig1 (g)-(l) are six images of one individual taken from the LFW dataset. If (a) is taken as the reference image, then each of the rest present multiple variations including one or more among background change, headgear, change in clothing, other faces in the image, and partial covering of face.

For the species recognition datasets, an ensemble of GIST and HOG features are used (to exploit repeated local patterns) and input images are resized to 121×121, preserving the original background. A patch size of 10×10 is used here as well. GIST and HOG are concatenated and fed through a Bag of Words (BoW) and Spatial Pyramid [9] pipeline before the classifier stage.

Sample images from the species recognition datasets are presented in Fig2. Fig2 (a)-(f) are images from the Oxford 102 Flowers dataset. It is to be noted that all six are different flowers, but there is only fine-grained differences between (a) and (b), between (c) and (d) and between (e) and (f). In fact, there is only subtle geometrical difference between these pairs, the only major variation being that of colour. Fig 2 (g)-(l) are images of dogs and cats (3 each) from the Oxford-IIIT Pets dataset. The first three images are those of the same breed of dog- American Bulldog and the last three are of the same breed of cat- Persian. Here it can be observed that although all three dogs are of same breed, there is remarkable difference in appearance. The same may be noted for the three cats. Thus the two Species benchmarks present the exact opposite challenge: the Flowers dataset poses the challenge of low inter-class differences and the Pets dataset poses the challenge of high intra-class differences.

For each dataset, ten-fold cross validation is performed and the average recognition accuracy is tabulated in percentage form. Results of experiments on the AR Faces, Oxford 102 Flowers, Oxford-IIIT Pets, LFW Faces, and LFW-a Faces datasets are presented in Table 1. Similar trends are noticeable across all the datasets, which are as follows:

•Among the CRC-based methods, the basic CRC [17] has the least accuracy compared to the modified versions, which is expected. The proposed improvement of the patch based approach (GP-CRC) consistently outperforms the original patch based CRC (PCRC) [18] by a significant margin (the consistent and statistically better performance of the proposed method is discussed later). In fact, GP-CRC has comparable and even slightly better results than the state of the art probabilistic version of CRC (ProCRC) [1].

•Comparing with the non-CRC methods, we observe that CRC has comparable results with Support Vector Machines (SVM) [3], whereas basic patch based CRC (PCRC) performs comparably with AdaBoost [5] and Random Decision Forests (RDF) [6]. The proposed method GP-CRC and the state of the art ProCRC outperform all other methods consistently and significantly.

To test the **statistical significance** of these improved results, unpaired t-tests were carries out between GP-CRC and each of the other methods for each dataset using the mean accuracy and standard deviations reported in Table 1. The two-tailed p values thus obtained are also mentioned in the same table and highlighted in bold if significant at a threshold of $p = 0.05$ (5% level of significance).

The results on the LFW and LFW-a benchmarks show the strong effect of background variation on classifier performance. The LFW-a results when the images are cropped as to exclude background information and the faces are aligned to front using commercial software. The results on the original LFW images preserve background, with only resizing of entire image.

Classifier	AR Faces	Oxford-102 Flowers	Oxford-IIIT Pets	LFW Faces	LFW-a Faces
SVM	88.5 ± 6.4 p = **0.0437**	61.2 ± 5.5 p = **0.0006**	56.4 ± 4.3 p = **0.0012**	25.0 ± 3.6 p = **0.0001**	41.7 ± 2.9 p = **0.0153**
AdaBoost	91.1 ± 5.9 p = 0.2323	63.1 ± 5.2 p = **0.0026**	58.5 ± 4.5 p = **0.0135**	26.6 ± 3.1 p = **0.0001**	42.2 ± 2.7 p = **0.0296**
RDF	90.6 ± 6.0 p = 0.1721	63.6 ± 5.6 p = **0.0052**	58.1 ± 4.6 p = **0.0095**	26.2 ± 3.2 p = **0.0001**	42.5 ± 2.8 p = 0.0532
CRC	90.0 ± 6.1 p = 0.1176	60.5 ± 5.7 p = **0.0004**	56.3 ± 4.4 p = **0.0012**	25.6 ± 3.8 p = **0.0001**	42.1 ± 3.0 p = **0.0330**
PCRC	92.3 ± 6.3 p = 0.4749	65.8 ± 5.1 p = **0.0266**	60.5 ± 4.6 p = 0.0954	27.5 ± 3.4 p = **0.0001**	42.96 ± 2.5 p = **0.0497**
ProCRC	93.9 ± 5.8 p = 0.9052	69.4 ± 5.9 p = 0.3909	63.6 ± 4.3 p = 0.7718	35.0 ± 3.5 p = **0.0336**	44.4 ± 3.0 p = 0.6384
GP-CRC	**94.2 ± 5.3**	**71.7 ± 5.8**	**64.2 ± 4.8**	**38.5 ± 3.3**	**45.0 ± 2.6**

Table 1: Percentage accuracies with standard deviation and statistical significance. The p values are computed against GP-CRC with a threshold value of $p = 0.05$.

Results indicate substantial degradation of performance of all classifiers when significant amount of background is present, except for the proposed GP-CRC and the state of the art ProCRC, for which the reduction in accuracy is comparatively low. ProCRC works well because its probabilistic framework takes into account the likelihood of a test image to belong to a given set. This compensates any new or rare background in the query sample(s), though not in a direct analytical fashion as in the proposed method. Thus better performance of both our method GP-CRC, besides ProCRC, in presence of significant background data, demonstrates the validity of the claim of this work.

It can be observed from Table 1 that the state-of-the-art ProCRC and the proposed GP-CRC have similar performance and the slight improvement in average accuracy is not statistically significant in most cases. So, to further look into the comparative performance of these two methods, we perform more statistical analysis. For every class of each dataset (AR, LFW-a, Flowers and Pets), the total number of available images are randomly divided into 10 equal parts and then 10-fold cross validation is carried out with same partitions for both ProCRC and GP-CRC. In the cross-validation, GP-CRC performs better than ProCRC 8 out of 10 times for AR dataset and 7 out of 10 times for each of LFW-a Faces, Oxford-102 Flowers and Oxford-IIIT Pets datasets. Thus GP-CRC beats ProCRC a total number of 29 out of 40 times, that is 72.5% of the times.

Next we perform a sign binomial test on this result with the null hypothesis that the two methods are almost equally good and hence any one may be assumed to give better result than the other equal number of times. It is calculated that given the assumption that both methods are equally good, then the probability of GP-CRC having better performance than ProCRC 72.5% or more times is 0.32% (one-tail p-value of 0.0032). The two-tail p-value is 0.0064. Considering a level of significance of $\alpha = 0.05$, we have to apply the Bonferroni adjustment, since we are applying several methods (4 nos.) simultaneously. So

we divide the 5% level of significance by 4 to get adjusted $\alpha = 0.0125$. Since the one-tail and two-tail p-values obtained are both less than 0.0125 (i.e., 1.25%), it may be concluded that though improvement in average accuracy of GP-CRC over ProCRC is marginal, it is still be significant considering the frequency of out-performance.

5 Conclusion

The premise of the present work is that although collaborative representation can effectively encode fine-grained classes, it can be susceptible to high within class background variations compared to foreground objects having subtle differences between classes. One way to solve this is instead of taking the classification prediction from whole images, the images are partitioned into patches of suitable size. Then these patches individually cast classification predictions and the output class is chosen from the majority of votes across all patches. This strategy only works if the foreground object covers most of the images and the background is limited. If not, the background patches may confound the majority decision. This is especially the case, if the test image(s) contain new or rare backgrounds.

We demonstrate the above effect experimentally on the Labeled Faces in the Wild (LFW) dataset, where we observe major change in performance between cropped face images and face images in natural settings. We also introduce a generalised formulation of collaboratively representing image patches which incorporates a regulation term to suppress outlier backgrounds. We test our methods on several benchmark datasets (2 face recognition and 2 species recognition) with encouraging results. We achieve major improvement over the original patch based method and comparable results with the state of the art probabilistic CRC approach.

In future we will incorporate the proposed generalised patch based CRC method (GP-CRC) into the probabilistic CRC method (ProCRC) and determine whether there is any further improvement in performance. Furthermore, a Probabilistic-Kernel CRC method (PKCRC) has recently been proposed by Liu *et al*. [11] which is the latest addition to the range of CRC based methods. It incorporates the Kernel trick into the Probabilistic CRC (ProCRC) formulation and it would be interesting to investigate the effect of the patch based approach in PKCRC as well.

6 Acknowledgement

We list here the publicly available source codes used for this work. The codes for the GIST and HOG descriptors are from Aditya Khosla's GitHub repository [8]; those for the CRC based classifiers (CRC, PCRC and ProCRC) can be found at Lei Zhang's academic page; those for AdaBoost and Random Decision Forests are available in MATLAB File Exchange. Statistical tests have been used using the tools available online at GraphPad. The code for the proposed method (GP-CRC) is expected to be made available in due course at the lead author's webpage: https://tapabrata-chakraborti.github.io/

References

[1] S. Cai, L. Zhang, W. Zuo, and X. Feng. A probabilistic collaborative representation based approach for pattern classification. In *Proc. CVPR*, 2016.

[2] T. Chakraborti, B. McCane, S. Mills, and U. Pal. Collaborative representation based fine-grained species recognition. In *Proc. IVCNZ*, 2016.

[3] C. Cortes and V. Vapnik. A support vector networks. *Machine Learning*, 20(3):273–297, 1995.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. CVPR*, 2005.

[5] Y. Freund and R. E. Schapire. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5):771–780, 1999.

[6] T. K. Ho. Random decision forests. In *Proc. ICDAR*, 1995.

[7] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments, 2007. University of Massachusetts, Amherst, Technical Report.

[8] A. Khosla, W. A. Bainbridge, A. Torralba, and A. Oliva. Modifying the memorability of face photographs. In *Proc. ICCV*, 2013.

[9] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. CVPR*, 2006.

[10] J. Liu, Z. Wu, J. Li, A. Plaza, and Y. Yuan. Probabilistic-kernel collaborative representation for spatial spectral hyperspectral image classification. *IEEE Trans. Geoscience and Remote Sensing*, 54(4):2371–2384, 2016.

[11] A.M. Martinez and R. Benavente. The ar face database, 1998. CVC Technical Report No.24.

[12] M-E. Nilsback and A. Zisserman. Delving into the whorl of flower segmentation. In *Proc. BMVC*, 2007.

[13] M-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In *Proc. ICVGIP*, 2008.

[14] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001.

[15] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. V. Jawahar. Cats and dogs. In *Proc. CVPR*, 2012.

[16] Y. Taigman, L. Wolf, and T. Hassne. Multiple one-shots for utilizing class label information. In *Proc. BMVC*, 2009.

[17] L. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In *Proc. ICCV*, 2011.

[18] P. Zhu, L. Zhang, Q. Hu, and Simon C.K. Shiu. Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. In *Proc. ECCV*, 2012.