Structured Multi-class Feature Selection for Effective Face Recognition

Giovanni Fusco, Luca Zini, Nicoletta Noceti, and Francesca Odone

DIBRIS - Università di Genova
via Dodecaneso, 35
16146-IT, Italy

{Giovanni.Fusco, Luca.Zini, Nicoletta.Noceti, Francesca.Odone}@unige.it

Abstract. This paper addresses the problem of real time face recognition in unconstrained environments from the analysis of low quality video frames. It focuses in particular on finding an effective and fast to compute (that is, sparse) representation of faces, starting from classical Local Binary Patterns (LBPs). The two contributions of the paper are a new formulation of Group LASSO for structured feature selection (MCGroup LASSO) to cope directly with multi-class settings, and a face recognition pipeline based on a representation derived from MC-GrpLASSO. We present an extensive experimental analysis on two benchmark datasets, MOBO and Choke Point, and on a more complex dataset acquired in-house over a large temporal span. We compare our results with state-of-the-art approaches and show the superiority of our method in terms of both performances and sparseness of the obtained solution.

Keywords: Regularized feature selection, Group LASSO, multi-class feature selection, face recognition.

1 Introduction

With the diffusion of video-surveillance systems on one side and mobile devices equipped with video cameras on the other side, it is more and more common to address the problem of recognizing faces in video frames acquired in cluttered and unconstrained environments. These settings pose many challenges: low-quality images, relatively small faces, motion blur, variable lighting conditions, subjects in non standard poses. Moreover, these domains usually call for real-time performances, thus the computational efficiency of the methods plays a key role.

In this paper we present an efficient face recognition method, which learns a very sparse face representation from data by means of structured regularized feature selection. We are interested in unconstrained settings where people are free to move as they like and are captured by conventional video-surveillance cameras or by mobile devices. We start from an overcomplete description of face images based on Local Binary Patterns (LBP) [18,1], which have been shown effective for analyzing textures and patterns in low quality images. LBPs are

computed at different scales and aspect ratios, then are subject to a feature selection step. In this way, not only we obtain an effective face representation, but we also improve the computational efficiency of face recognition at run time. With this aim, we propose a new formulation of the Group LASSO functional [17] – we call *MC-GrpLASSO* – to cope directly with a multi-class setting while capturing the block-structure of LBP. The new formulation allows us to select groups of features that simultaneously discriminate among all the identities. On top of this, we build a modular and efficient face recognition pipeline, analyzing a face image and associating with it one among N previously learnt identities.

Many works in the literature cope with the problem of selecting features in the face recognition domain. Often feature selection is based on Adaboost: in [18] the authors start off computing an overcomplete set of LBP histograms computed in different regions of face images. Then, they use Adaboost to select the most significant features. The work in [6], instead, introduces the Jensen-Shannon Boosting (JSBoost) algorithm, a modification of AdaBoost based on the Jensen-Shannon (JS) divergence, that is shown to provide a more appropriate measure of distance between examples belonging to two different classes. In [14], the authors consider Gabor features to extend Adaboost by incorporating mutual information. They achieve a lower training error rate with respect to the standard implementation. The Gabor features are also adopted in [11], where Real Adaboost is introduced to perform the selection wavelets. A genetic based approach is proposed in [13], where the selection of the LBP regions is based on Genetic & Evolutionary Computing (GEC). Many works look for identity-based descriptions [4,7] and are thus more specifically tailored for the face authentication problem. Here we focus on finding a universal description, common to all identities, with an advantage in terms of scalability. To this purpose our method is specifically designed for multi-class settings. Few works in the literature are, up to now, devoted to multi-class feature selection, in particular for face recognition. We mention [3] which extends Recursive Feature Elimination (RFE) to the multi-class case.

The remainder of the paper is organized as follows. Sec. 2 describes the multi-class feature selection method we propose, while Sec. 3 focuses on the entire face recognition pipeline. The experimental analysis on the different datasets is detailed on Sec. 4, while Sec. 5 is left to a final discussion.

2 Regularized Multi-class Feature Selection

In this section we describe the multi-class feature selection method we propose. Let us first formulate the *feature selection problem*: given a training set $(\mathbf{x}_i, y_i), i = 1, \dots, n$, with $\mathbf{x}_i \in X \subseteq R^m$ $y_i \in R$, and a dictionary $\mathcal{D} = (\phi_j)_{j=1}^p$ which is a collection of atoms or features, we consider a generalized linear formulation of the input-output relationship

$$\sum_{j=1}^p \phi_j(\mathbf{x}_i) \beta_j = y_i \quad (1)$$

or, in matrix form $\Phi\beta = \mathbf{y}$, where Φ is the feature matrix defined as $\Phi_{ij} = \phi_j(\mathbf{x}_i)$ of size $n \times p$, $\mathbf{y} \in \{-1, +1\}^{n \times 1}$ is the output vector, and $\beta \in \mathbb{R}^{p \times 1}$ is the vector of weights. The goal of feature selection is to find a *sparse* β which approximates well the input-output relationship.

Group LASSO. It is an extension of LASSO [15] to select groups of features, rather than considering each feature independently [17]. In several application domains, in fact, a prior on the structure of each input datum is available, and may be profitably inserted in the feature selection step. It can be formulated as a regularized minimization problem:

$$\beta^* = \arg \min_{\beta} \left(\|\mathbf{y} - \Phi\beta\|_2^2 + \tau \sum_{g=1}^G \|\beta_{I_g}\|_2 \right) \quad (2)$$

where $g = 1, \dots, G$ is an index referring to a feature group, while $I_g \subseteq [1 \dots p]$ is the set containing the positions in the input datum of the features of group g . The first term of the functional is a *data fidelity term* while the second one is a *penalty term* which enforces the selection of features in groups. The regularization parameter τ can be tuned to obtain solutions with a different level of sparsity.

Multi-class Group LASSO. We now re-write the functional (2) to deal with the simultaneous selection of groups of meaningful features for a multi-class problem. Let us consider N binary problems $\Phi\beta_c = \mathbf{y}_c$, $c = 1 \dots N$. Similarly to above, Φ is the features matrix of size $n \times p$, $\mathbf{y}_c \in \{-1, +1\}^{n \times 1}$ is the output vector, containing 1 in correspondence of examples of class i and -1 elsewhere. Finally, $\beta_c \in \mathbb{R}^{p \times 1}$ is the weight vector for the class c . We formulate the Multi-Class Group LASSO (MC-GrpLASSO) as follows:

$$B^* = \arg \min_B \left(\|Y - \Phi B\|_2^2 + \tau \sum_{g=1}^G \|B_{I_g}\|_F \right). \quad (3)$$

Y is a $n \times N$ matrix composed as $[\mathbf{y}_1 \mathbf{y}_2 \dots \mathbf{y}_N]$, and, similarly, B is a $p \times N$ such that $B = [\beta_1 \beta_2 \dots \beta_N]$. F refers to the Frobenius norm. In this case sparsity enforces the selection of features in groups common to all the N classes. Roughly speaking, the method selects the features that best discriminate among all the N classes simultaneously.

3 The Proposed Pipeline

In this section we describe our approach to face recognition. We assume we have in input a training set $\{\mathbf{x}_i\}_{i=1}^n$ of n face images of N different identities. We first represent each image according to an appropriate dictionary of features based on Local Binary Patterns (LBP). Then, we apply MC-GrpLASSO to obtain a set of features groups meaningful for all the N classes. Finally, we train a multi-class classifier on the training set of images represented with the selected groups only.

Fig. 1. A visualization of our face descriptors, a concatenation of LBPs

This allows us to obtain, at run time, a very fast representation and classification of each face image.

Construction of the Feature Matrix. In this work we adopt an overcomplete dictionary of Local Binary Pattern (LBP) descriptors, whose effectiveness in face recognition is well known. Similarly to [18], we consider a grid of overlapping regions having different scale and aspect ratio (faces images are all rescaled to the size 40×40 pixels). We then extract uniform 8-bits LBPs [10] and quantize their values with 59-bins histograms, accordingly to the original work. Figure 1 describes how to obtain the feature vector in a simplified situation, with no overlapping regions. The final overcomplete description consists of 841 LBPs, thus each training image $\mathbf{x}_i$ is described by $841 \times 59 = 49,619$ features (which is also the number of columns in the features matrix Φ). For numerical reasons matrix Φ needs to be normalized so that each column has mean 0 and variance 1.

Multi-class Feature Selection. We specialize the functional (3) to our case by grouping the features according to the image region they are associated with (see Figure 1). To find a solution to the optimization problem we adopt an iterative scheme as in [9] specialized for the multi-class case as follows:

$$\begin{aligned} \hat{B}_{:,c}^p &= B_{:,c}^{p-1} + \frac{1}{\sigma} \Phi^T (Y_{:,c} - \Phi B_{:,c}^{p-1}) \quad \forall c \in [1, N] \\ B^p &= S_{\frac{\tau}{\sigma}} \left(\hat{B}^p \right) \end{aligned} \quad (4)$$

where σ is the iteration step size, which we set proportional to the highest eigenvalue of matrix $\Phi^T \Phi$, to achieve an optimal convergence to the solution of (3) [9]. S is the soft-thresholding operator that acts on all features of a group $(S_{\frac{\tau}{\sigma}}(\hat{B}))_{i,c} = (|\hat{B}_{I_g,:}|_F - \frac{\tau}{\sigma})_+ \hat{B}_{i,c}$. The iterative procedure can be initialized with a matrix B^0 of zeros. Once we have a solution B we select the regions

associated with the rows without zero values: in this way the selected features are meaningful to *all* the N classes.

Recognition. Once we have found a sparse and meaningful representation for our multi-class problem, we train a multi-class classifier based on SVM and a *Winner-Takes-All* strategy [8]. Notice that it would have been possible to classify data directly via the solution of the MC-GrpLASSO, but the performance obtained with SVMs are usually higher because of the well known shrinkage effect of the weights of LASSO methods [2]. Thus, we train N One-vs-All binary SVM classifiers where each image $\mathbf{x}_i$ is represented by mapping it on the sub-set of selected regions, by a projector $\pi(\mathbf{x}_i) = \{\phi_j(\mathbf{x}_i) | B_{j\cdot} \neq 0\}$. Given the binary classifier of class c , the negative examples are randomly sampled from all the classes j , $j = \dots N$, $j \neq c$.

For a new datum $\mathbf{x}$, the resulting global classifier is then formulated as a combination of the N discriminant functions $h_1 \dots, h_N$ of the binary classifiers as

$$h(\mathbf{x}) = \arg \max_{c=1 \dots N} \{h_c(\pi(\mathbf{x}))\}. \quad (5)$$

4 Experimental Analysis

In this section we present the experimental evaluations of the proposed pipeline, with particular emphasis on representation (feature selection) and recognition.

Datasets. The analysis is performed on three different datasets: two public benchmarks – *MOBO* [5] and *Choke Point* [16] – and a in-house dataset, *R309*. The *MOBO* dataset consists of low-resolution videos of 20 subjects, while they perform 4 different actions. It is characterized by small variations in the lighting condition and subject poses. The *Choke Point* dataset is characterized by a higher variability of image quality, appearance and lighting conditions. It includes videos of 29 subjects walking through two portals, acquired by an array of three cameras. The *R309* dataset (see Fig. 2) is a collection of indoor videos acquired in unconstrained conditions: the pose and the behavior of the volunteers are entirely unconstrained and very natural. The recordings have been made under various lighting conditions and in two very different environments. R309 includes 12 people with an average of 1200 frames for each identity acquired over a period of time of 7 months. It thus allows us to evaluate the robustness of the methods with respect to variations occurring over time.

Fig. 2. *R309* dataset: same subject under different poses and lighting conditions

Experimental Protocol. Each dataset has been divided in training, validation, and test, in the proportions specified later for each dataset. The validation set is used for tuning the parameters of the classifiers (that is, the regularized parameter of each SVM). The results reported in the tables have been obtained on the test set.

Assessment of the Selected Features. The solution of the functional (3) is a weight matrix B with a degree of sparsity that depends on τ . The choice of the parameter is carried out on the validation set and guided by the aim of obtaining a sparse solution producing a high recognition rate. The matrix has a characteristic structure as the one shown in Fig. 3 (left): MC-GrpLASSO selects groups meaningful for all classes, producing a stripe structure (the selected groups have non-zero weights along all the classes). It is interesting to compare the structure of this matrix with the one derived by a 1vsALL architecture. In this case we solve N binary structured feature selection problems as in Eq. (2) obtaining N solutions β_i , $i = 1, \dots, N$. Fig. 3 (right) shows a matrix obtained by combining the N solutions (one per column). Notice how the β_i do not relate to one another.

Fig. 3. (MOBO dataset) $G \times N$ binary maps: each entry is associated with an image region (row) with respect to a given class (column). Black entries are the selected groups.

With a 1vsALL approach we need to find a way to combine the N β_i s; classical choices are intersection (keep groups selected by all binary problems) and union (keep groups selected by at least one problem). In this case parameter tuning becomes more complex: we need to choose an appropriate τ for each binary problem, while controlling the cardinality and the performances of the overall problem. In particular, in the case of intersection, it is often the case that the overall solution is empty; in this case a voting rule may be adopted. Fig. 4 highlights the features selected by MC-GrpLASSO (bottom) and 1vsALL with union (top). The two representations achieve similar performances (see Table 1), but MC-GrpLASSO does it with less than 1/4 of the regions needed by the 1vsALL method.

Fig. 4. (MOBO dataset) Binary maps of the obtained representations: in black are marked the indices of the groups selected by 1vsALL union (top) and MC-GrpLASSO (bottom). On the left smaller regions, on the right larger regions.

Finally, Fig. 5 gives a qualitative impression of the kind of features produced by our method, as it compares the 10 most meaningful features selected by different approaches to multi-class regularized feature selection: MCGroup LASSO, 1vsALL with union, and LASSO. The latter have been obtained by reformulating the N class problem with a simpler binary problem of unstructured features [7,4]: $\mathbf{x}_i$ s are obtained by computing region-wise χ^2 similarities between LBP histograms of two different image pairs (positive examples are derived from image pairs of the same class). Then a region is associated with a single feature and unstructured feature selection (for instance with Adaboost or LASSO) may be applied. From the analysis of Fig. 5 it is apparent that structured feature selection methods better capture the complexity and the overall structure of the available data.

Fig. 5. MOBO dataset: the 10 most meaningful features (with respect to the corresponding value in β) selected by LASSO (I row), 1vsALL union (II row), MC-GrpLASSO (III row)

Face Recognition Assessment. On the **MOBO dataset** we consider a training set of 50 examples per identity, extracted from a pool of 5 videos (#training set = 1000). The validation set is collected following the same strategy on 5 other videos. The remaining videos are used to build the test set, made of 200 examples per subject (#test set = 4000). Table 1 reports the recognition

Table 1. Comparative analysis on the *MOBO* dataset

METHOD	# Features	TPR	FPR
fixed grid LBP [1]	49	95.6%	0.2%
LBP + LASSO	25	94.1%	0.3%
LBP + AdaBoost	25	90.1%	0,5%
LBP+ 1vsALL GrpLASSO (voting 50%.)	33	95.1 %	0.3 %
LBP + 1vsALL GrpLASSO (union)	118	97.0%	0.1 %
LBP + MC-GrpLASSO	26	96.9%	0.1%
LPB + MC-GrpLASSO	13	96.1%	0.2%

Table 2. Comparative analysis on the *Choke Point* dataset (as in [16])

METHOD	REC. RATE
MRH + MSM on image subsets [16]	86.7%
LBP + MSM on image subsets [16]	75.8%
LBP + LASSO [4]	93.1%
LBP + MC-GrpLASSO	96.9%

Table 3. Comparative analysis on the R309 dataset

METHOD	# Features	TPR	FPR
fixed grid LBP	49	71,6%	3,5%
LBP+LASSO	19	71,4%	3,6%
LBP + GrpLASSO (union)	98	72.4%	0.3%
LBP + MC-GrpLASSO	45	79,6%	2,6%

performances we obtained starting from LBP features with different feature selection approaches. The goal is to achieve comparable or better performances with a sparser representation. MC-GrpLASSO achieves very good performances with a remarkably small representation.

On the **Choke Point dataset** we adopt the data groups and the protocol as proposed in [16]. In this paper the authors propose a video-based face recognition based on selecting the most meaningful faces in a video and then using those images to perform the recognition phase. Images are represented by Multi-Region Histograms (MRH) [12] and LBP [1]. Tab. 2 summarizes our results and also includes the best performances reported in [16], although the nature of the results is different. Tab. 3 shows how the overall performances on the **R309** dataset are significantly lower than for the other datasets, due to a consistent variability on the subjects appearances since acquisitions have been made on a temporal span of months. In any case our approach shows better performances than the others. Finally, Fig 6 reports the recognition results on the three dataset as a function of the selected regions. Notice how the results are quite stable with respect to this parameter.

Fig. 6. Recognition rates function of the number of selected regions

5 Discussion

This paper presented an efficient face recognition method based on a compact and effective face representation derived by regularized structured feature selection. Starting from an overcomplete description of face images made of Local Binary Patterns (LBPs), we applied feature selection to select the discriminative information for the problem, and obtain a lower dimensional description. In particular, we proposed a new formulation of Group LASSO – the *MC-GrpLASSO* – to directly cope with a multi-class setting and exploiting the block-structure of features. Unlike previous approaches, the new formulation allows us to select groups that simultaneously discriminate among all the identities. We presented an extensive experimental analysis, both on benchmark datasets and on a complex set of data acquired in-house. The experimental analysis showed that our method outperforms other state-of-art approaches on all the three datasets. The future work will focus on two different aspects. On one hand, we will extend the pipeline from the standpoint of the description, by exploiting the temporal information we have available from the use of videos. On the other, the method we designed will be integrated on a prototype system for supporting blind users. It will be embedded on a device wore by the users, to help them in recognizing known people.

References

1. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
2. Candes, E., Tao, T.: The Dantzig selector: statistical estimation when p is much larger than n . The Annals of Statistics 35(6), 2313–2351 (2007)

3. Chapelle, O., Keerthi, S.: Multi-class feature selection with support vector machines. In: Proc. of the American Statistical Association (2008)
4. Destrero, A., Mol, C., Odone, F., Verri, A.: A regularized framework for feature selection in face detection and authentication. *IJCV* 83(2), 164–177 (2009)
5. Gross, R., Shi, J.: The cmu motion of body database. Tech. rep., Robotics Institute, Carnegie Mellon University (2001)
6. Li, H.X., Wang, S., Jensen-shannon, Y.: boosting learning for object recognition. *CVPR* 2, 144–149 (2005)
7. Hadid, A., Pietikäinen, M., Li, S.Z.: Learning personal specific facial dynamics for face recognition from videos. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) *AMFG 2007*. LNCS, vol. 4778, pp. 1–15. Springer, Heidelberg (2007)
8. Jain, B.J., Wysotzki, F.: A competitive winner-takes-all architecture for classification and pattern recognition of structures. In: Hancock, E.R., Vento, M. (eds.) *GbrPR 2003*. LNCS, vol. 2726, pp. 259–270. Springer, Heidelberg (2003)
9. Mosci, S., Rosasco, L., Santoro, M., Verri, A., Villa, S.: Solving structured sparsity regularization with proximal methods. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) *ECML PKDD 2010, Part II*. LNCS, vol. 6322, pp. 418–433. Springer, Heidelberg (2010)
10. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI* (7) (2002)
11. Ruan, C., Ruan, Q., Li, X.: Real adaboost feature selection for face recognition. In: *CSP*, pp. 1402–1405 (2010)
12. Sanderson, C., Lovell, B.C.: Multi-region probabilistic histograms for robust and scalable identity inference. In: Tistarelli, M., Nixon, M.S. (eds.) *ICB 2009*. LNCS, vol. 5558, pp. 199–208. Springer, Heidelberg (2009)
13. Shelton, J., Dozier, G., Bryant, K., Adams, J., Popplewell, K., Abegaz, T., Purring-ton, K., Woodard, D., Ricanek, K.: Genetic based lbp feature extraction and selection for facial recognition. In: *Southeast Regional Conf.*, pp. 197–200 (2011)
14. Shen, L., Bai, L., Bardsley, D., Wang, Y.: Gabor feature selection for face recognition using improved adaBoost learning. In: Li, S.Z., Sun, Z., Tan, T., Pankanti, S., Chollet, G., Zhang, D. (eds.) *IWBRS 2005*. LNCS, vol. 3781, pp. 39–49. Springer, Heidelberg (2005)
15. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1) (1996)
16. Wong, Y., Chen, S., Mau, S., Sanderson, C., Lovell, B.C.: Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In: *CVPRW*, pp. 74–81 (2011)
17. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* 68 (2006)
18. Zhang, G., Huang, X., Li, S.Z., Wang, Y., Wu, X.: Boosting local binary pattern (LBP)-based face recognition. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) *Sinobiometrics 2004*. LNCS, vol. 3338, pp. 179–186. Springer, Heidelberg (2004)