

INTRA-CLASS MULTI-OUTPUT REGRESSION BASED SUBSPACE ANALYSIS

S. Karthikeyan^{*} Swapna Joshi^{*} B.S. Manjunath^{*} Scott Grafton[†]

^{*} Department of ECE, University of California Santa Barbara

[†] Department of Psychology, University of California Santa Barbara

^{*}{karthikeyan,swapna,manj}@ece.ucsb.edu

[†]{grafton}@psych.ucsb.edu

ABSTRACT

A common challenge when dealing with heterogeneous tasks such as face expression analysis, face and object recognition is high dimensionality and extreme appearance variations within each class. To handle such scenarios, we formulate a supervised Non-negative Matrix Factorization (NMF) based subspace learning technique that simultaneously preserves the intra-class regression information (local) and enhances inter-class discrimination (global) in the low dimensional embedding. Our method leverages the multi-dimensional image labels that quantify the within class regression to learn the subspaces for recognition. In addition, our formulation includes a novel multi-output regression based NMF algorithm.

Index Terms— Intra class regression, Subspace Analysis, Face recognition across pose

1. INTRODUCTION

In many applications of computer vision and pattern recognition, images can be considered as points in a high-dimensional space. However, a specific pattern of interest, within this unnecessary high dimensional data, can reside in a low-dimensional subspace. Subspace analysis aims in discovering this representation by expressing the data as a (linear or nonlinear) combination of a small number of underlying signals.

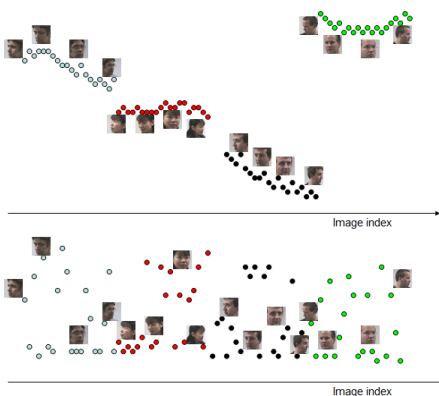


Fig. 1. Encoding coefficient subspace corresponding to a single basis obtained from our framework (top) and NMF (bottom) depicting the pan angle (-90 to +90) regression within each class (different color cluster) and the separation between each of the classes in the Pointing'04 dataset. (Image best viewed in color.)

In this paper we work with the basic idea of subspace analysis, mainly matrix factorization. However, in our method we incorporate the label information to learn a lower dimensional embedding

that represents both intra-class regression and inter-class discrimination. Especially when dealing with problems such as face recognition across pose, not only do we need to understand the differences between different people but also study the variation of a particular face across multiple pose labels. Figure 1 shows a low dimensional embedding corresponding to a particular basis, of images from the Pointing'04 dataset. As can be noticed, unlike original NMF (bottom), our embedding (top) clearly reflects the internal trend within a class, while maintaining good separation between classes. Furthermore, we handle multi-dimensional pose (regression) labels. Moreover, learning low dimensional embedding while preserving multi-dimensional regression information (here on referred as multi-label regression) is rarely studied in the literature.

Most of the prevalent subspace reduction techniques target regression or classification individually. Primarily these techniques are statistically, geometrically or intuitively motivated. Principal component analysis (PCA), Linear discriminant analysis (LDA) [1] aim in finding projections that are statistically significant for preserving maximum information content. Manifold learning algorithms [2] are geometrically motivated non-linear reduction methods; recently, [3] proposed a supervised manifold embedding extension. The other class of algorithms are motivated to obtain features which capture intuitive parts of objects. In this regard, non-negative matrix factorization (NMF) [4], and localized non-negative matrix factorization (LNMF) [5] are unsupervised parts based subspace reduction techniques.

A fundamental limitation of the above mentioned approaches is that they do not learn the variations within each data class which can create a better low-dimensional representation. Recently, regression based non-negative matrix factorization (RNMF) [6] aims to obtain subspaces which highlight the regression information in the low dimensional embedding. Our approach, which understands multi-dimensional intra-class regression patterns, is motivated by RNMF. First, we extend RNMF to handle multiple outputs called Multi-output RNMF, followed by the proposed classification technique called Local multi-label Regression Global Discrimination (LRGD) based subspace analysis.

Therefore, our contributions are two fold:

- A new supervised NMF for classification (LRGD), that learns a salient representation of the internal regression structure within the different classes of the data while simultaneously performing inter-class discrimination. This modification directs the encoding coefficients to reflect the local trend of the regression as well as classification labels.
- A novel multi-output regression framework using NMF (MRNMF), by distributing the different regression trends

among different set of bases and simultaneously reconstructing the input data.

2. MULTI-OUTPUT REGRESSION BASED NON-NEGATIVE MATRIX FACTORIZATION (MRNMF)

Most regression approaches, such as support vector regression (SVR) [7], can only deal with single dimensional outputs. Extending them to handle multidimensional outputs is nontrivial and is rarely studied. In this section we propose an extension to RNMF that can handle multi-dimensional outputs. Our technique simultaneously aims to reconstruct the data and learn different sets of bases corresponding to different regression labels. A similar idea was proposed in [8] where the distinguishing information relevant for every class was captured in separate set of bases.

We work with the basic non-negative matrix factorization framework where the input data matrix (\mathbf{V}) is represented as a product of the basis matrix (\mathbf{W}) and a coefficient matrix (\mathbf{H}). Here, the aim is to learn \mathbf{W} and \mathbf{H} such that $\mathbf{V} \approx \mathbf{WH}$, all the matrices are non-negative. The individual high dimensional data points (\mathbf{v}_i), basis images (\mathbf{w}_i) and coefficients or low dimensional representations (\mathbf{h}_i) are represented in the columns of \mathbf{V} , \mathbf{W} and \mathbf{H} matrices respectively. We use this framework to model multi-output regressions due to the efficacy of RNMF in single output parts based regression analysis [6].

In our approach, we divide the basis \mathbf{W} for capturing separate regressions as

$$\mathbf{W} = [\mathbf{W}^1 | \dots | \mathbf{W}^k | \dots | \mathbf{W}^n] \quad (1)$$

Since every column \mathbf{w}_i has a one-to-one correspondence with i^{th} row of \mathbf{H} , we divide \mathbf{H} as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}^1 \\ \vdots \\ \mathbf{H}^k \\ \vdots \\ \mathbf{H}^n \end{bmatrix}, \mathbf{h}_i = \begin{bmatrix} \mathbf{h}_i^1 \\ \vdots \\ \mathbf{h}_i^k \\ \vdots \\ \mathbf{h}_i^n \end{bmatrix} \quad (2)$$

where \mathbf{h}_i is a column of \mathbf{H} , while \mathbf{h}_i^k is a block vector corresponding to the k^{th} regression pattern.

Let every input \mathbf{v}_i have a corresponding output vector denoted by $\mathbf{y}_i = [y_i^1 \dots y_i^k \dots y_i^n]$, where n is the number of regression labels. Now the regression information for the k^{th} output variable of \mathbf{v}_i , y_i^k is reflected in the bases set \mathbf{W}^k as in (1). Hence, the regression cost for output y_i^k is imposed on the corresponding coefficients \mathbf{H}^k . The lower dimensional embedding \mathbf{h}_i of \mathbf{v}_i is divided into subsections as shown in (2). The multi-output regression cost S_{Rmo} can therefore be written as

$$S_{Rmo} = \min \sum_k \sum_{i,j} \frac{f(y_i^k, y_j^k) (\mathbf{h}_i^k - \mathbf{h}_j^k)^T (\mathbf{h}_i^k - \mathbf{h}_j^k)}{\sum_{i,j} f(y_i^k, y_j^k)} \quad (3)$$

where $f(y_i^k, y_j^k) = \exp\left(\frac{-|y_i^k - y_j^k|}{t^k}\right)$, t^k depends on the range of the labels, so as the distance between data increases in the dependent variable space, the weight decreases accordingly. This constraint aims to keep \mathbf{h}_i^k and \mathbf{h}_j^k close to each other if the corresponding y_i^k and y_j^k are close and vice versa. Similar to [6], the constrained divergence cost function now becomes

$$D(\mathbf{V}||\mathbf{WH}) = \min \left[\sum_{i,j} \left(v_{ij} \log \frac{v_{ij}}{r_{ij}} - v_{ij} + r_{ij} \right) + \alpha S_{Rmo} + \beta S_O \right] \quad (4)$$

where the orthogonality constraint $S_O = \sum_{i,j} u_{ij}$, $\mathbf{U} = \mathbf{W}^T \mathbf{W}$. The updates for \mathbf{W} and \mathbf{H} accounting for multi-dimensional outputs is obtained as

$$w_{kl} \leftarrow \frac{w_{kl} \sum_j v_{kj} \frac{h_{lj}}{\sum_{k'} w_{kk'} h_{k'l}}}{\sum_j h_{lj} + \beta \sum_j w_{kj}}, \quad (5)$$

If element h_{kl} belongs to the p^{th} output and denoting $f(y_i^p, y_j^p)$ as f_{ij}^p , then the update rule for h_{kl} is obtained as

$$h_{kl} \leftarrow \frac{-b + \sqrt{b^2 + \left(\sum_i v_{il} \frac{w_{ik} h'_{kl}}{\sum_{k'} w_{ik'} h_{k'l}} \right) \left(\frac{16\alpha \sum_k f_{kl}^p}{\sum_{i,j} f_{ij}^p} \right)}}{\left(\frac{8\alpha \sum_k f_{kl}^p}{\sum_{i,j} f_{ij}^p} \right)} \quad (6)$$

where

$$b = 1 - \frac{4\alpha}{\sum_{i,j} f_{ij}^p} \sum_{k'} (h_{kk'} f_{k'l}^p). \quad (7)$$

These iterative update rules are obtained by building an auxiliary function from the cost function and optimizing it similar to [6].

3. EXPERIMENTS AND RESULTS

We test the proposed MRNMF algorithm with a synthetic dataset and a face expression dataset.

3.1. Synthetic Data

Dataset: We begin our analysis on a synthetic data-set as illustrated in Figure 2, as a proof of concept. The dataset consists of 200 images, where each image is of size 200×200 pixels. Every image I consists of four distinct square parts p_i , $i \in \{1, 2, 3, 4\}$, locally varying in intensity with a specific regression pattern. Each image has a two-dimensional label $\mathbf{y} = (y_1, y_2)$, where $y_1 \in \{1 \dots 200\}$, $y_2 \in \{1 \dots 200\}$. The following are the variations formulated for each square block in $I(y_1, y_2)$.

- 1) $I_{p_1}(y_1, y_2) = y_1$
- 2) $I_{p_2}(y_1, y_2) = 100 - |y_1 - 100|$
- 3) $I_{p_3}(y_1, y_2) = |y_2 - 100|$
- 4) $I_{p_4}(y_1, y_2) = 200 - y_2$

where noise is added with uniform distribution as $\mathcal{U}(0, 2\mu)$ in all images, and a constant intensity pattern surrounds the four parts (teal color in Figure 2). For example, p_1 refers to the red block which increases with label y_1 . We randomly select 200 combinations of (y_1, y_2) , the training images are the corresponding $I(y_1, y_2)$.

Performance: The main focus of this synthetic experiment is to understand and isolate the different regression patterns in the data. Here we have compared MRNMF with RNMF. We use two basis images per regression pattern for MRNMF, thus using four bases for both MRNMF and RNMF. As RNMF cannot handle multidimensional output labels, we performed RNMF on each output dimension independently. As seen in Figure 2, MRNMF is able to segregate each regression pattern into individual basis images, while RNMF fails to do so, as it is handling each dimension independent of the other.

3.2. Expression Analysis

Dataset: In order to test the above described system we use the publicly available JAFFE (Japanese Female Facial Expressions) dataset [9]. This particular dataset includes 213 grayscale images, each 256×256 pixels in size. The JAFFE data set includes facial images of 10 different female models, each assuming seven distinct poses. The images containing neutral poses and fear were not used in this study. We partition the data into 10 random sets, where 120 are selected for training and 60 for testing. The aim of this experiment is

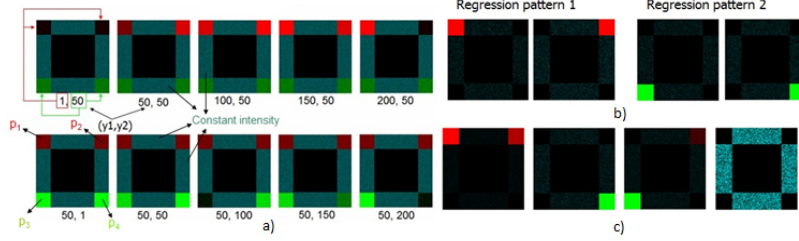


Fig. 2. a) Example of synthetic data for $I(y_1, y_2)$ where $y_1 \in \{1 \dots 200\}$, $y_2 \in \{1 \dots 200\}$. b) Four bases obtained by MRNMF. c) Four bases obtained by RNMF. Images best viewed in color

to predict the different coefficients associated with each test image using MRNMF. The expression coefficients range from 1 to 5.

Results: Our focus is to estimate the 5 dimensional expression output (happy, sad, surprise, anger, disgust) given a particular face image. In order to compute the expression output for a given test image, we first project the training and the test data onto the basis images in \mathbf{W} and obtain $\tilde{\mathbf{H}}_{train}$, $\tilde{\mathbf{H}}_{test}$, respectively. These encoding coefficients are sorted according to the regression outputs k , as shown in (2). The k^{th} output dimension is then determined by performing kernel regression on the corresponding $\tilde{\mathbf{H}}_{train}^k$. Since we are isolating the encoding coefficients according to the regression pattern, we are able to analyze the variation within every expression separately and calculate the expression label more accurately. Other methods including RNMF do not give us this flexibility.

We compare our method with PCA, RNMF, SVR, LNMF and NMF. Since RNMF and SVR can only handle single-output regression problem, we decoupled the multi-output regression problem to isolated single output runs. As we can see in the Figure 3 MRNMF outperforms all the other methods. We intuitively expect MRNMF to perform better than NMF, LNMF and PCA as these are unsupervised learners. The performance of RNMF, as expected, is lower as it considers each output dimension separately as opposed to MRNMF. In addition MRNMF outperforms SVR which achieves an average error of 0.76 (Figure 3 does not contain SVR since we have no control on the number of bases). We also note that while all methods performed well for simple expressions like ‘surprise’, our method achieved better prediction on complex expressions like ‘anger’ and ‘disgust’.

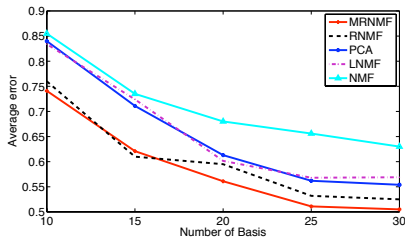


Fig. 3. Average recognition error for the JAFFE dataset when number of bases are varied. (Image best viewed in color.)

4. LOCAL MULTI-LABEL REGRESSION GLOBAL DISCRIMINATION (LRGD) BASED SUBSPACE ANALYSIS

This section details on how regression information can be leveraged for classification problems. Traditional supervised subspace analysis techniques for classification such as [1, 10] are based on maximizing between class separation and within class compactness. However, in our algorithm we learn the intrinsic variation within each class while

maximizing the inter-class separation. For example, in face recognition across different poses, we take advantage of the face pose label (pan,tilt) information to obtain a meaningful subspace. Furthermore, using a discriminative constraint, the class labels can be used to ensure strong separation between classes.

In order to capture the intra-class regression with respect to the multidimensional variable, we use a formulation similar to MRNMF, where the regression labels within each class drive encoding coefficients. We define the indices of $[\mathbf{v}_j]$ to be j ($j \in \{1, 2 \dots n_t\}$). Let l be the number of classes and the indices belonging to the class i be denoted by the set C_i , $i \in \{1, 2 \dots l\}$. The within-class regression preserving cost S_{WRmo} to be minimized is denoted by

$$S_{WRmo} = \min \frac{\sum_d \sum_k \sum_{i,j \in C_d} f(y_i^k, y_j^k) (\mathbf{h}_i^k - \mathbf{h}_j^k)^T (\mathbf{h}_i^k - \mathbf{h}_j^k)}{\sum_d \sum_k \sum_{i,j \in C_d} f(y_i^k, y_j^k)} \quad (8)$$

where $f(y_i^k, y_j^k)$ is defined in Section 2 with parameter t^k .

Since \mathbf{h}_j has a one-to-one correspondence with \mathbf{v}_j the between class separation is as follows:

$$S_B = \max \frac{1}{n_{dif}} \sum_d \sum_{i,j \notin C_d} (\mathbf{h}_i - \mathbf{h}_j)^T (\mathbf{h}_i - \mathbf{h}_j) \quad (9)$$

where n_{dif} is the number of data point pairs $(\mathbf{v}_i, \mathbf{v}_j)$ not belonging to the same class.

Therefore, the final cost function is a constrained divergence formulation which is expressed as

$$D(\mathbf{V} || \mathbf{WH}) = \min \left[\sum_{i,j} \left(v_{ij} \log \frac{v_{ij}}{x_{ij}} - v_{ij} + x_{ij} \right) + \alpha S_{WRmo} - \beta S_B + \gamma S_O \right] \quad (10)$$

where $\mathbf{WH} = \mathbf{X} = [x_{ij}]$, S_O is similar to (4) and constants $\alpha, \beta, \gamma > 0$. The update rule for \mathbf{W} is same as in (5). As every h_{kl} is associated with a particular output label (p) and class (r), the update for \mathbf{H} can be written as

$$h_{kl} \leftarrow \frac{-b + \sqrt{b^2 + 4a \left(\sum_{i,l} v_{il} \frac{w_{ik} h'_{kl}}{\sum_{k'} w_{ik'} h'_{k'l}} \right)}}{2a} \quad (11)$$

where

$$b = 1 - 4\alpha \sum_{i \in C_r} \lambda_{il}^p h_{ki} + 4 \frac{\beta}{n_{dif}} \sum_{i \notin C_r} h_{ki} \quad (12)$$

and

$$a = \left(4 \left(\alpha \sum_{i \in C_r} \lambda_{il}^p - \beta N_r \right) \right) \quad (13)$$

$$\lambda_{kl}^p = \frac{f_{kl}^p}{\sum_d \sum_k \sum_{i,j \in C_d} f_{ij}^k} N_r = \frac{n_t - |C_r|}{n_{dif}} \quad (14)$$

where $f(y_i^p, y_j^p)$ is denoted as f_{ij}^p , n_t is the number of training samples, $|C_r|$ is the cardinality of class r , n_{dif} as defined above, and $\alpha > 0, \beta > 0$. For $a > 0$ we are guaranteed non-negative update rules. The convergence proof is similar to that in [6].

5. EXPERIMENTS AND RESULTS

We tested the efficacy of LRGD on popular face and object recognition datasets. A brief description of the datasets follows.

Pointing'04 Dataset: The Pointing'04 dataset [11] consists of 15 sets of images. Each set contains two series of 93 images of the same person at different poses. The pose, or head orientation is determined by 2 angles (pan,tilt), which varies from -90° to $+90^\circ$. We preprocess the dataset by using Viola-Jones [12] frontal and profile face detectors to isolate the faces. Our evaluation procedure uses 675 training images (45 from every class) and 724 testing images. The pan and tilt angles are used as the two dimensional regression labels for this scenario.

CMU-PIE Dataset: The CMU-PIE dataset [13] consists of face images of 68 subjects with 43 different illumination conditions across 13 poses. All extracted face images are aligned and normalized to size of 32×24 pixels. We select faces of 59 subjects, for every subject we randomly select 6 illuminations for each of the 9 pan angles (excluding c31 c25 c07 c09). Out of these 54 images per subject we randomly select 36 faces for training and the rest for testing. The pan-angles are used as regression labels.

ETH 80 Dataset: This dataset [14] contains 80 objects from 8 categories. There are 10 objects per category; each object is represented by 41 images from viewpoints spaced equally over the upper viewing hemisphere. Every image is quantified by a pan and tilt angle depending on its viewpoint. We use all views from the first 7 objects for training and all views from the remaining 3 objects for testing, in each class.

Performance: Our focus is to perform face and object recognition invariant to pose changes. We compare our method (LRGD) with PCA, NMF, LNMF, FNMF [10], MFA [15] and LSDA [16]. PCA, NMF and LNMF are unsupervised approaches. The rest are supervised approaches which utilize class label information. In particular MFA and LSDA understand the intra-class structure in a data-driven manner. In all these approaches, after projecting the test and train data to the lower dimension by learning a basis we use the k -nearest neighbor classifier. Table 1 shows the classification accuracy for the different methods across various number of basis components.

We observe that LRGD outperforms all the other methods by a significant amount, especially with unaligned images with extreme pan and tilt variations (Pointing'04). We can see that understanding the pose regression of each individual face not only helps in projecting the new data in the right class but also place it well on the latent regression curve learnt in that particular class. In addition, we analyzed the encoding coefficients, and notice that the original NMF (Figure 1 (bottom)) hardly conveys information regarding the local regression patterns within classes and the separation between classes is inadequate. However, Figure 1 (top), the LRGD encoding coefficient for the Pointing'04, reflects the regression pattern of the pan angle within each class, along with sufficient inter-class distance.

Acknowledgements : Our research was supported by NSF award III-0808772, MacArthur Foundation and public health service grant NIMH 1 R01 MH070539-01. We also thank Michael Stephens for his contribution in the implementation of the algorithm.

Methods	Dataset		
	Pointing '04	CMU-PIE	ETH80
PCA	75.2	82.4	77
NMF	76.3	83.1	78.8
LNMF	72.5	83.3	79.3
FNMF	72.9	83.5	79.7
MFA	81.4	86.3	81.5
LSDA	81.8	84.5	81.7
LRGD	84.8	88.5	82.3

Table 1. Classification accuracy (%) of different algorithms on three datasets, Pointing'04, CMU-PIE and ETH-80.

6. CONCLUSIONS AND FUTURE WORK

We have presented a novel matrix factorization method (LRGD) for classification using intra-class regression information. We have also extended the previous RNMF algorithm to handle multi-dimensional output labels (MRNMF). Experimental results on synthetic and JAFFE dataset show superior performance of MRNMF in terms of prediction accuracy and visual perception. The quantitative and visual results of LRGD presented on several databases show that understanding the local regression pattern along with discrimination can prove extremely effective for classification problems.

7. REFERENCES

- [1] PN Belhumeur et al., "Eigenfaces vs. Fisherfaces: recognition using class specific linearprojection," *IEEE TPAMI*, 1997.
- [2] Tenenbaum et al., "A global geometric framework for nonlinear dimensionality reduction," *Science*, 2000.
- [3] V. Balasubramanian et al., "Biased manifold embedding: A framework for person-independent head pose estimation," in *IEEE CVPR'07*.
- [4] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, 1999.
- [5] S.Z. Li et al., "Learning spatially localized, parts-based representation," in *CVPR*. IEEE, 2001.
- [6] S. Joshi, S. Karthikeyan et al., "Anatomical parts-based regression using non-negative matrix factorization," in *CVPR*. IEEE, 2010.
- [7] V.N. Vapnik, *The nature of statistical learning theory*, 2000.
- [8] Q. Chen et al., "Factorization towards a classifier," in *CVPR*, 2010.
- [9] M.J. Lyons et al., "Automatic classification of single facial images," *IEEE TPAMI* 1999.
- [10] Y. Wang et al., "Fisher non-negative matrix factorization for learning local features," in *ACCV*, 2004.
- [11] N. Gourier et al., "Estimating face orientation from robust detection of salient facial structures," in *FG Net Workshop on Visual Observation of Deictic Gestures (POINTING)*, 2004.
- [12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, 2001.
- [13] T. Simon et al., "The CMU pose, illumination, and expression (PIE) database," in *FG, 2002. IEEE*.
- [14] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *CVPR*. IEEE, 2003.
- [15] S. Yan et al., "Graph embedding and extensions: A general framework for dimensionality reduction," *TPAMI* 2007.
- [16] D. Cai et al., "Locality sensitive discriminant analysis," in *IJCAI* 2007.