Analysis-Synthesis Dictionary Learning for Universality-Particularity Representation based Classification

Meng Yang[†], Weiyang Liu[†‡], Weixin Luo[†], and Linlin Shen[†]
[†]College of Computer Science & Software Engineering, Shenzhen University, China
[‡]School of Electronic & Computer Engineering, Peking University, China

Abstract

Dictionary learning has played an important role in the success of sparse representation. Although synthesis dictionary learning for sparse representation has been well studied for universality representation (i.e., the dictionary is universal to all classes) and particularity representation (i.e., the dictionary is class-particular), jointly learning an analysis dictionary and a synthesis dictionary is still in its infant stage. Universality-particularity representation can well match the intrinsic characteristics of data (i.e., different classes share commonality and distinctness), while analysis-synthesis dictionary can give a more complete view of data representation (i.e., analysis dictionary is a dual-viewpoint of synthesis dictionary). In this paper, we proposed a novel model of analysis-synthesis dictionary learning for universality-particularity (ASDL-UP) representation based classification. The discrimination of universality and particularity representation is jointly exploited by simultaneously learning a pair of analysis dictionary and synthesis dictionary. More specifically, we impose a label preserving term to analysis coding coefficients for universality representation. Fisher-like regularizations for analysis coding coefficients and the subsequent synthesis representation are introduced to particularity representation. Compared with other state-of-the-art dictionary learning methods, ASDL-UP has shown better or competitive performance in various classification tasks.

Introduction

With inspirations of the sparsity mechanism of human vision system (B.A. Olshausen 1996) and the success of sparse coding in image processing (M. Elad 2006)(J.C. Yang 2008), sparse representation has been widely applied to many fields, such as computer vision and pattern recognition (J. Wright 2009)(J. Wright 2010). As indicated by (R. Rubinstein 2010), the dictionary plays an important role in the success of sparse representation, which should faithfully and discriminatively represents an input signal as a sparse linear combination of dictionary atoms. Learning the desired dictionary from training data instead of using off-the-shelf bases (e.g., wavelets) has led to state-of-the-art results in many practical applications, such as image denoising (M. Elad 2006), face recognition (Q. Zhang 2010)(Z.L. Jiang 2013), and image classification (I. Ramirez 2011)(M. Yang 2011).

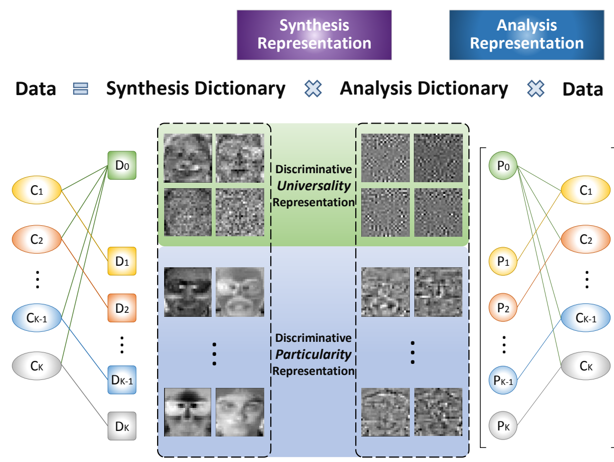


Figure 1: Framework of the proposed analysis-synthesis dictionary learning for universality-particularity representation.

One representative dictionary model for sparse representation is K-SVD (M. Aharon 2006), which learns an over-complete dictionary of atoms for image patches and has shown promising performance in image restoration. In the task of image classification, image patches of all classes could also be used to learn an dictionary, on which the coding coefficients of image patches could generate the descriptor of images (J.C. Yang 2009). Both (M. Aharon 2006) and (J.C. Yang 2009) belong to unsupervised dictionary learning approach since they ignore to use class label information by assuming that image patches are universal to all classes. With this assumption, the encoding of image patches on the learned dictionary is regarded as universality representation. Unsupervised dictionary learning for universality representation is powerful for data reconstruction, but not advantageous for classification tasks. Without class label information, the dictionary of unsupervised learning can not do particularity representation for labeled data so that the discrimination embedded in the training data can not be well exploited. For classification tasks, the class label information could guide dictionary learning to achieve a better classification ability, so most prevailing dictionary learning approaches for classification is supervised.

In supervised dictionary learning, discriminative universality representation and particularity representation have been studied for better classification. When the learned dictionary is universal to all classes, discrimination of universality representation was explored by using group sparsity (S. Bengio 2009) on coding coefficients or jointly learning a dictionary and a classifier over the coding coefficients (J. Mairal 2012; Q. Zhang 2010; Z.L. Jiang 2013; W. Liu 2015a). Due to the promising performance of class-particular dictionary representation in (J. Wright 2009), regularization for particularity representation has been introduced in the phase of dictionary learning. (e.g., low class-particular dictionary coherence (I. Ramirez 2011), good class-particular representation for some class but bad for all the other classes (J. Mairal 2008)(A. Castrodad 2012), Fisher discrimination on class-particular dictionary and coding coefficient (W. Liu 2015b), task-driven terms (M. Yang 2013), etc.) (L. Zhang 2011a) further boosts the coding efficiency using l_2 norm instead of l_1 for coding coefficients.

Most of dictionary learning approaches (e.g., the methods above) synthesize an input signal by using a linear combination of dictionary atoms, so they are called synthesis dictionary learning. Recently, Rubinstein *et al.* (R. Rubinstein 2013) proposed an analysis dictionary learning method, named analysis K-svd, for image restoration. As a dual-viewpoint of synthesis dictionary, analysis dictionary directly transforms a signal to a sparse feature space by multiplying the signal. Similar to class-particular synthesis dictionary learning, particularity discrimination has also been exploited in the analysis and synthesis dictionary learning for image classification tasks (S. Gu 2014).

Although improved performance has been reported in the existing dictionary learning approaches, there still remains several critical issues. Firstly, separately exploring discriminative universality representation and particularity representation (J. Mairal 2012)(Q. Zhang 2010)(Z.L. Jiang 2013)(I. Ramirez 2011)(J. Mairal 2008)(A. Castrodad 2012) is not optimal since the data with different class labels not only have distinctness but also share commonality (as shown in Fig.1). Dictionary learning for discriminative universality representation alone ignores the powerful class-particular representation, while there would be a big correlation of class-particular dictionaries for the single discriminative particularity representation. Secondly, conducting synthesis dictionary learning and analysis dictionary learning separately can not collaborate their advantages together. Analysis-synthesis dictionary provides a more comprehensive viewpoint for representing a signal. The analysis part projects the signal to a new feature space like feature transformation, while the synthesis part represents the signal with well reconstruction. Furthermore, how to learn a discriminative analysis-synthesis dictionary for universality-particularity representation based classification still remains to be explored.

In this paper we propose a novel model of analysis-synthesis dictionary learning for universality-particularity (ASDL-UP) representation based classification. The discrimination of universality and particularity representation is jointly exploited by simultaneously learning a pair of analysis dictionary and synthesis dictionary, as shown in Fig.1. To the best of our knowledge, it is the first time to integrate analysis-synthesis dictionary learning and universality-particularity representation into a unified model. More specifically, a label preserving term that regularizes analysis coding coefficients for universality representation is designed, while Fisher-like regularizations for analysis coding coefficient and synthesis representation are introduced to particularity representation. The proposed ASDL-UP is evaluated on action, face, and gender classification. Compared with existing state-of-the-art dictionary learning methods, ASDL-UP has better or competitive performance in various classification tasks.

Related Work

Supervised synthesis dictionary learning for universality representation (J. Mairal 2012)(Q. Zhang 2010)(Z.L. Jiang 2013)(S. Bengio 2009) and particularity representation (J. Mairal 2008)(M. Yang 2011)(A. Castrodad 2012)(I. Ramirez 2011) has been well studied recently. Moreover, universality and discriminative components have also been mined in matrix decomposition (Q. Zhang 2012). The closely related work to our proposed analysis-synthesis dictionary learning for universality-particularity (ASDL-UP) representation are synthesis dictionary learning for universality-particularity representation (N. Zhou 2012)(S. Kong 2012)(L. Shen 2013)(M. Yang 2014). Zhou *et al.* (N. Zhou 2012) proposed a discriminative regularization on the synthesis coding coefficients, while Kong *et al.* (S. Kong 2012) introduced a coherence penalty term of different synthesis sub-dictionaries. Instead of using a flat category structure, Shen *et al.* (L. Shen 2013) proposed to learn a synthesis dictionary with a hierarchical category structure; recently Yang *et al.* (M. Yang 2014) proposed a latent synthesis dictionary learning approach. Although these synthesis dictionary learning methods exploited universality and particularity representation, they ignored to fully explore the discrimination of universality-particularity representation and to learn a dual-viewpoint analysis dictionary, which could make coding more meaningful and efficient. For analysis dictionary learning, there is no supervised method but a unsupervised approach for universality representation applied to image restoration (R. Rubinstein 2013).

The study of supervised analysis-synthesis dictionary learning has started recently. Gu *et al.* (S. Gu 2014) proposed a model of analysis-synthesis dictionary learning for particularity representation. Although improved results are reported, it doesn't use intra-class discrimination and ignores universality representation, resulting in big correlation between different class-particular dictionaries. For universality representation, there is no supervised analysis-synthesis dictionary learning method proposed except a unsupervised analysis-synthesis dictionary learning approach (R.Rubinstein 2014) for image denoising.

Analysis-Synthesis dictionary learning

Compared to universality or particularity representation, the universality-particularity representation could better represent intrinsic characteristic of data, i.e., data of different classes not only have class-particular parts but also share commonality. Meanwhile, analysis-synthesis dictionary could provide a more complete view of data representation data than analysis dictionary or synthesis dictionary because analysis dictionary is a dual viewpoint of synthesis dictionary (R. Rubinstein 2013). Both universality-particularity representation and analysis-synthesis dictionary learning could benefit the representation and classification of data, therefore they should be jointly considered in a unified dictionary learning model.

Let $\boldsymbol{X} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_k, \cdots, \boldsymbol{X}_K]$, where each column of $\boldsymbol{X}_k$ is a k^{th}-class training sample. The model of analysis-synthesis dictionary learning for universality-particularity (ASDL-UP) representation is represented as

$$\begin{array}{ll} \min\limits_{\boldsymbol{D}_0,\boldsymbol{P}_0,\boldsymbol{D},\boldsymbol{P}} & \|\boldsymbol{X} - \boldsymbol{D}_0 S_\eta(\boldsymbol{P}_0\boldsymbol{X}) - \boldsymbol{D}S(\boldsymbol{P}\boldsymbol{X})\|_F^2 \\ & +\gamma R_0(\boldsymbol{D}_0, \boldsymbol{P}_0) + \lambda R(\boldsymbol{D}, \boldsymbol{P}) \\ \text{s.t.} & \left\|\boldsymbol{d}_i^0\right\|_2^2 \leq 1\ \forall i; \|\boldsymbol{d}_i^k\|_2^2 \leq 1\ \ \forall i, k \end{array} \tag{1}$$

where $\boldsymbol{P}_0$ and $\boldsymbol{D}_0$ are a pair of universal analysis and synthesis dictionaries, $\boldsymbol{D} = [\boldsymbol{D}_1, \cdots, \boldsymbol{D}_k, \cdots, \boldsymbol{D}_K]$, $\boldsymbol{P} = [\boldsymbol{P}_1, \cdots, \boldsymbol{P}_k, \cdots, \boldsymbol{P}_K]$, $\boldsymbol{P}_k$ and $\boldsymbol{D}_k$ are a pair of class-particular analysis and synthesis dictionaries for class k, and $\boldsymbol{d}_i^0$ and $\boldsymbol{d}_i^k$ are the i^{th} atoms of $\boldsymbol{D}_0$ and $\boldsymbol{D}_k$, respectively. Here $S_\eta(.)$ and $S(.)$ are two sparse operators on the coding coefficients generated by the analysis dictionary, $R_0(\boldsymbol{D}_0, \boldsymbol{P}_0)$ and $R(\boldsymbol{D}, \boldsymbol{P})$ are discriminative regularizations on ASDL for universality representation and particularity representation, respectively. These two regularizations are illustrated in detailed in the following sections.

Regularization for universality representation

The regularization of $R_0(\boldsymbol{D}_0, \boldsymbol{P}_0)$ on analysis-synthesis dictionary is designed for discriminative universality representation based classification. Although the universal dictionary atoms lose correspondence to class labels, the discrimination of coding coefficients is exploited to make the universal dictionary powerful for classification. Here we designed a label preserving term for coding coefficients to simultaneously minimize the intra-class variance and maximize the inter-class variance. To this end, we proposed a discriminative analysis coding coefficient term,

$$R_0(\boldsymbol{D}_0, \boldsymbol{P}_0) = \|\boldsymbol{Y} - \boldsymbol{W}S_\eta(\boldsymbol{P}_0\boldsymbol{X})\|_F^2 \tag{2}$$

where $\boldsymbol{Y} \in \Re^{K \times N}$ is a label indicating matrix, N is the number of training samples, $\boldsymbol{W}$ is a discrimination matrix projecting the coding coefficient to the label space, and $S_\eta(.)$ is a hard thresholding operator which sets its input entries with absolute values below η as zeros, while keep all the other entries unchanged. Here a column vector of $\boldsymbol{Y}$ is a label indicating vector for some training sample. For instance, the indicating vector for a training sample of class k is $[0, \ldots, 1, \ldots, 0]^T$ where 1 is in the k^{th} element.

Regularization for particularity representation

The regularization of $R(\boldsymbol{D}, \boldsymbol{P})$ on analysis-synthesis dictionary is designed for discriminative particularity representation. Since class-particular dictionary atoms have correspondence to class labels, both the discrimination of class-particular representation and coding coefficient is exploited by designed a suitable regularization. Here we proposed a Fisher-like regularization on the analysis coding coefficient and synthesis representation. We design $R(\boldsymbol{D}, \boldsymbol{P})$ as

$$R(\boldsymbol{D}, \boldsymbol{P}) = \sum\nolimits_{k=1}^{K} \left\|\boldsymbol{P}_k\left[\bar{\boldsymbol{X}}_k, \boldsymbol{X}_k - \boldsymbol{M}_k\right]\right\|_F^2 \tag{3}$$

where $\bar{\boldsymbol{X}}_k$ is the complementary set of $\boldsymbol{X}_k$, and $\boldsymbol{M}_k$ is a matrix with each column as the mean vector of $\boldsymbol{X}_k$. The discrimination on analysis representation is designed in the light of the Fisher criterion. The minimization of Eq.(3) not only makes $\boldsymbol{P}_k$ have less correlation with the data from other classes for a big between-class scatter, but also makes $\boldsymbol{X}_k - \boldsymbol{M}_k$ small along the subspace spanned by $\boldsymbol{P}_k$ for a small within-class scatter.

The sparse regularization, i.e., $S(.)$, is designed as $S(\boldsymbol{P}\boldsymbol{X}_k) = [\boldsymbol{0}; \cdots; \boldsymbol{P}_k\boldsymbol{X}_k; \cdots; \boldsymbol{0}]$ according to the Fisher criterion to achieve a discriminative synthesis representation. For k^{th} data $\boldsymbol{X}_k$, the sparse regularization sets zero synthesis representation for other class synthesis dictionary (i.e.,$S(\boldsymbol{P}_l\boldsymbol{X}_k) = \boldsymbol{0}$ for $l \neq k$) while keep the synthesis representation for k^{th}-class synthesis dictionary.

Model of ASDL-UP

We have proposed discriminative regularizations on AS dictionary for university representation and particularity representation. Thus the proposed model of analysis-synthesis dictionary learning for university-particularity (ASDL-UP) representation is written as

$$\begin{array}{ll} \min\limits_{\boldsymbol{D},\boldsymbol{W},\boldsymbol{P},\boldsymbol{D}_0,\boldsymbol{P}_0} \sum\limits_{k=1}^{K} & \|\boldsymbol{X}_k - \boldsymbol{D}_0 S_\eta(\boldsymbol{P}_0\boldsymbol{X}_k) - \boldsymbol{D}_k\boldsymbol{P}_k\boldsymbol{X}_k\|_F^2 \\ & +\lambda\left\|\boldsymbol{P}_k\left[\bar{\boldsymbol{X}}_k, \boldsymbol{X}_k - \boldsymbol{M}_k\right]\right\|_F^2 \\ & +\gamma\|\boldsymbol{Y}_k - \boldsymbol{W}S_\eta(\boldsymbol{P}_0\boldsymbol{X}_k)\|_F^2 \\ \text{s.t.} & \left\|\boldsymbol{d}_i^0\right\|_2^2 \leq 1\ \forall i; \|\boldsymbol{d}_i^k\|_2^2 \leq 1\ \ \forall i, k \end{array} \tag{4}$$

where the first term is the discriminative universality-particularity representation on AS dictionary based on Eq.(3), and the second and third terms are discriminative analysis coding coefficients terms of universality representation and particularity representation.

ASDL-UP representation based classification

After learning the universal and class-particular analysis-synthesis dictionaries, we conduct the classification by jointly using the discrimination of universality and particularity of analysis-synthesis dictionaries. The identity is

$$\arg\min_k \|\boldsymbol{y} - \boldsymbol{D}_0 S_\eta(\boldsymbol{P}_0\boldsymbol{y}) - \boldsymbol{D}_k\boldsymbol{P}_k\boldsymbol{y}\|_2 - \tau[\boldsymbol{W}S_\eta(\boldsymbol{P}_0\boldsymbol{y})]_k \tag{5}$$

where the first term is the particularity representation residual on analysis-synthesis dictionary for class k, the second term is class confidence based on university representation on analysis dictionary for class k, and a simple linear combination of universality-particularity discrimination is utilized in our classification model.

Optimization of ASDL-UP

The ASDL-UP objective function in Eq.(4) is divided into two sub-problems by doing universal analysis-synthesis dictionary learning (ASDL) and class-particular ASDL alternatively: updating $\boldsymbol{D}_0$, $\boldsymbol{P}_0$ by fixing $\boldsymbol{D}_k$, $\boldsymbol{P}_k$ and updating $\boldsymbol{D}_k$, $\boldsymbol{P}_k$ by fixing $\boldsymbol{D}_0$, $\boldsymbol{P}_0$.

Universal ASDL

By fixing the class-particular analysis-synthesis dictionary (i.e., $\boldsymbol{D}_k$, $\boldsymbol{P}_k$ for each class k), the ASDL-UP model of Eq.(4) becomes

$$\begin{array}{l}\min\limits_{\boldsymbol{D}_0,\boldsymbol{P}_0,\boldsymbol{W}} \sum\limits_{k=1}^{K} \left(\begin{array}{c} \|\boldsymbol{Z}_k - \boldsymbol{D}_0 S_\eta\left(\boldsymbol{P}_0\boldsymbol{X}_k\right)\|_F^2 \\ +\gamma \|\boldsymbol{Y}_k - \boldsymbol{W} S_\eta\left(\boldsymbol{P}_0\boldsymbol{X}_k\right)\|_F^2 \end{array}\right) \\ \quad \text{s.t. } \left\|\boldsymbol{d}_i^0\right\|_2^2 \le 1 \ \forall i; \end{array} \tag{6}$$

where $\boldsymbol{Z}_k = \boldsymbol{X}_k - \boldsymbol{D}_k\boldsymbol{P}_k\boldsymbol{X}_k$. Let $\boldsymbol{\Gamma}_k = \left[\boldsymbol{Z}_k; \sqrt{\gamma}\boldsymbol{Y}_k\right]$ and $\boldsymbol{O} = \left[\boldsymbol{D}_0; \sqrt{\gamma}\boldsymbol{W}\right]$, the two terms of Eq.(6) could integrate a single one,

$$\min_{\boldsymbol{O},\boldsymbol{P}_0} \|\boldsymbol{\Gamma} - \boldsymbol{O} S_\eta\left(\boldsymbol{P}_0\boldsymbol{X}\right)\|_F^2 \ \text{ s.t. } \|\boldsymbol{o}_i\|_2^2 \le 1 \ \forall i \tag{7}$$

where $\boldsymbol{o}_i = [\boldsymbol{d}_i^0; \boldsymbol{w}_i]$, $\boldsymbol{w}_i$ is the i^{th} column vector of $\boldsymbol{W}$, $\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_1, \cdots, \boldsymbol{\Gamma}_k, \cdots, \boldsymbol{\Gamma}_K]$ and $\boldsymbol{X} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_k, \cdots, \boldsymbol{X}_K]$. The constraint of Eq.(7) can not only regularize the universal dictionary but also make the solution of $\boldsymbol{W}$ more stable. For solving Eq.(7), we update the universal analysis-synthesis dictionary atoms pair by pair. At the j^{th} step, we keep all but the j^{th} pair of atoms (i.e.,$\boldsymbol{o}_j$ and $\boldsymbol{p}_j$) fixed, where $\boldsymbol{P}_0 = [\boldsymbol{p}_1^T; \cdots, \boldsymbol{p}_j^T; \cdots]$. Then we could isolate the dependence on the j^{th} atom pair and rewrite Eq.(7) as

$$\min_{\boldsymbol{o}_j,\boldsymbol{p}_j} \left\|\boldsymbol{E}_j - \boldsymbol{o}_j S_\eta\left(\boldsymbol{p}_j^T\boldsymbol{X}\right)\right\|_2^2 \quad s.t. \ \|\boldsymbol{o}_j\|_2^2 \le 1 \tag{8}$$

where $\boldsymbol{E}_j = \boldsymbol{\Gamma} - \sum_{i\neq j} \boldsymbol{o}_i S_\eta\left(\boldsymbol{p}_i^T\boldsymbol{X}\right)$.

Let $S_\eta(.)$ define a partition of $\boldsymbol{X}$ into two sets: J and $\bar{J}$, with current $\boldsymbol{p}_j$, where $S_\eta(\boldsymbol{p}_i^T\boldsymbol{X}^J) = \boldsymbol{p}_i^T\boldsymbol{X}^J$ and $S_\eta(\boldsymbol{p}_i^T\boldsymbol{X}^{\bar{J}}) = \boldsymbol{0}$. $\boldsymbol{E}_j$ is similarly spitted to the submatrices $\boldsymbol{E}_j^J$ and $\boldsymbol{E}_j^{\bar{J}}$. Then we approximate Eq.(8) to

$$\min_{\boldsymbol{o}_j,\boldsymbol{p}_j} \left\|[\boldsymbol{E}_j^J; \boldsymbol{0}] - \boldsymbol{o}_j\boldsymbol{p}_j^T\left[\boldsymbol{X}^J; \epsilon\boldsymbol{X}^{\bar{J}}\right]\right\|_2^2 \ \text{ s.t. } \|\boldsymbol{o}_j\|_2^2 \le 1 \tag{9}$$

where ϵ is simply set as a big value (e.g.,10) to minimize $\boldsymbol{p}_j^T\boldsymbol{X}^{\bar{J}}$ and make $S_\eta(\boldsymbol{p}_j^T\boldsymbol{X}^{\bar{J}}) = \boldsymbol{0}$. Similar to (R.Rubinstein 2014), the set J remains roughly when Eq.(9) is solved. The model of Eq.(9) can be efficiently solved by using the Rank-one approximation algorithm in Fig.2 of (R.Rubinstein 2014), with normalizing the energy of each synthesis dictionary atom less than 1.

By using similar procedure, all pairs of analysis-synthesis dictionary atoms for universality representation are updated. The cost of Eq.(6) can decrease in each iteration and can be approximately solved after several iterations.

Algorithm 1 Training Procedure of ASDL-UP

1: **Initialization**
All analysis-synthesis dictionary atoms are randomly initialized with unit l_2-norm energy.
2: **Class-particular ASDL**
Updating each pair of class-particular dictionaries, $\boldsymbol{D}_k$ and $\boldsymbol{P}_k$, by solving Eq.(10).
3: **Universal ASDL**
Updating each pair of universal dictionary atoms by solving Eq.(8).
4: **Output**
Return to step 2 until the values of the objective function in Eq. (4) in adjacent iterations are close enough or the maximum number of iterations is reached.
Output $\boldsymbol{D}, \boldsymbol{P}, \boldsymbol{D}_0, \boldsymbol{P}_0$, and $\boldsymbol{W}$.

Class-particular ASDL

By fixing the universal analysis-synthesis dictionary (i.e., $\boldsymbol{D}_0$, $\boldsymbol{P}_0$), the ASDL-UP model of Eq.(4) becomes

$$\begin{array}{l}\min\limits_{\boldsymbol{D},\boldsymbol{P}} \sum\limits_{k=1}^{K} \ \|\boldsymbol{Z}_{k,0} - \boldsymbol{D}_k\boldsymbol{P}_k\boldsymbol{X}_k\|_F^2 + \lambda \|\boldsymbol{P}_k\boldsymbol{V}_k\|_F^2 \\ \quad \text{s.t. } \|\boldsymbol{d}_i^k\|_2^2 \le 1 \ \forall i; \end{array} \tag{10}$$

where $\boldsymbol{V}_k = [\bar{\boldsymbol{X}}_k, \boldsymbol{X}_k - \boldsymbol{M}_k]$ and $\boldsymbol{Z}_{k,0} = \boldsymbol{X}_k - \boldsymbol{D}_0 S_\eta\left(\boldsymbol{P}_0\boldsymbol{X}_k\right)$. Then the analysis-synthesis dictionary learning for particularity representation can be conducted class by class. The formulation of Eq. (10) is same to that of Eq.(5) in (S. Gu 2014). So in this paper we solve Eq.(10) by using the algorithm of (S. Gu 2014).

ASDL-UP algorithm

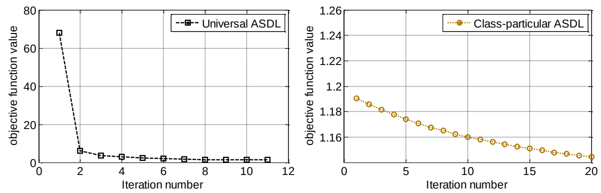
Based on the algorithms of class-particular ASDL and universal ASDL, the algorithm of ASDL for universality-particularity representation is summarized in Algorithm 1. When $\boldsymbol{O}$ is learned, the projection matrix (i.e.,$\boldsymbol{W}$) and the universal analysis dictionary $\boldsymbol{D}_0$ is solved based on $\boldsymbol{O} = \left[\boldsymbol{D}_0; \sqrt{\gamma}\boldsymbol{W}\right]$. And the final $\boldsymbol{W}$ is got by do renormalization, $\boldsymbol{w}_i = \boldsymbol{w}_i / \left\|\boldsymbol{d}_i^0\right\|_2 \ \forall i$.

Based on the proposed solving algorithm, both object function values of class-particular ASDL and universal ASDL monotonously decreases. Since the object function of ASDL-UP is lower bounded, the whole algorithm can converge. An convergence example of universal ASDL and calss-particular ASDL on UCF sport action dataset (M. Rodriguez 2008) is show in Fig. 2(a), with the convergence of the whole ASDL-UP shown in Fig. 2(b).

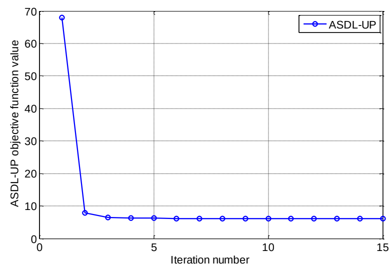
Experiments and results

In this section, we evaluate the performance of ASDL-UP on various classification task. Action classification, face recognition, gender classification are performed by using ASDL-UP and the competing methods in the following Sections. More experimental and time complexity analysis are presented in the **supplementary material**[1]. To clearly il-

[1]available at **www.yangmeng.org.cn** or **www.wyliu.com**



(a)



(b)

Figure 2: An example of ASDL-UP convergence on UCF sport action dataset (M. Rodriguez 2008).

lustrate the advantage of ASDL-UP, we compare ASDL-UP with several recent DL methods, such as Discriminative K-SVD (DKSVD) (Q. Zhang 2010), Label Consistent K-SVD (LCKSVD) (Z.L. Jiang 2013), dictionary learning with structure incoherence (DLSI) (I. Ramirez 2011), dictionary learning with commonality and particularity (COPAR) (S. Kong 2012), joint dictionary learning (JDL) (N. Zhou 2012), Fisher discrimination dictionary learning (FDDL) (M. Yang 2011), and the latest dictionary pair learning (D-PL) (S. Gu 2014). Besides, we also report sparse representation based classifier (SRC) (J. Wright 2009), collaborative representation based classification (CRC)(L. Zhang 2011b), linear support vector machine (SVM), nearest subspace classifier(NSC) and some methods for special tasks. In no specific instruction, the number of class-specific dictionary atoms in these DL methods is set as the number of training samples in the same class.

There are three parameters,λ,γ, and η, in the model of ASDL-UP (i.e.,Eq.(4)). λ is a parameter of discriminative class-particular analysis coding coefficient term, γ is a parameter of discriminative universal analysis coefficient term, and η is a parameter of hard-thresholding function. In our experiments, we fix $\lambda = 1e-3$ in all experiments and set $\gamma = 0.1$, and $\eta = 1e-5$ for face recognition, and $\gamma = 0.5$, and $\eta = 1e-4$ for all the other experiments.

UCF Action classification

The benchmark action dataset, UCF sports action dataset (M. Rodriguez 2008), is used to conduct the action classification experiment. The dataset collected video clips from various broadcast sports channels (e.g., BBC and ESP-N). The action bank features of 140 videos provided by (S. Sadanand 2012) are adopted in the experiment. These videos cover 10 sport action classes: driving, golfing, kicking, lifting, horse riding, running, skateboarding, swinging-(prommel horse and floor), swing-ing-(high bar) and walking. As the experiment setting in (Z.L. Jiang 2013), we e-valuated the ASDL-UP via five-fold cross validation. Here the dimension of the action bank feature is reduced to 100 via PCA, and the performance of some specific methods for action recognition, such as Qiu 2011 (Q. Qiu 2011), action bank feature with SVM classifier (S. Sadanand 2012) are also re-ported in Table 1. It can be observed that the proposed ASDL-UP achieves the highest rate (i.e.,95.0% accuracy), 1.4% improvement over the second best method, FDDL.

Table 1: The recognition rates (%) on UCF sports action dataset. Qiu and Sadanand denote (Q. Qiu 2011) and (S. Sadanand 2012) respectively.

Qiu	LCKSVD	DLSI	DKSVD	FDDL	Sadanand
83.6	91.2	92.1	88.1	93.6	90.7
SRC	COPAR	JDL	KSVD	DPL	**Ours**
92.9	90.7	90.0	86.8	92.9	**95.0**

Table 2: The recognition rates (%) on UCF 50 dataset. Sadanand, Wang denote the approach used in (S. Sadanand 2012) and (H.R. Wang 2012) respectively.

NSC	LCKSVD	DLSI	DKSVD	FDDL	Sadanand
51.8	53.6	60.0	38.6	61.1	57.9
SRC	COPAR	JDL	Wang	DPL	**Ours**
59.6	52.5	53.5	47.9	62.4	**62.8**

Following the experiment settings in (S. Sadanand 2012), we then evaluated ASDL-UP on the large-scale UCF50 action dataset by using fivefold group-wise cross validation, and compared it with the recent DL methods and the other state-of the-art methods, including (S. Sadanand 2012) and (H.R. Wang 2012). The results are shown in Table 2. Again, ASDL-UP achieves better performance than all the competing methods. Meanwhile, both the analysis-synthesis dictionary learning methods, including ASDL-UP and DPL, are visibly better than all the other methods. Compared with (S. Sadanand 2012), ASDL-UP has over about 5% improvement.

Face recognition

Two face databases, such as the aligned labeled face in the wild (LFWa)(L. Wolf 2009) and AR (A. Martinez 1998), are used to evaluate the performance of the proposed ASDL-UP. The AR dataset contains illumination, expression, and disguise variation. Following the experimental setting of AR in (Z.L. Jiang 2013), a set of 2,600 images of 50 female and 50 male subjects are extracted. 20 images of each subject are used for training and the remaining 6 images are used for testing. 540-d feature provided by (Z.L. Jiang 2013) is used as the facial features. The experimental results of all the competing methods are listed in Table 3. In this dataset, ASDL-UP achieves the best performance, e.g., 99.5% accuracy, and have over 1% improvements over all the other competing methods including DPL.

We then evaluate ASDL-UP on the application of face

Table 3: The recognition rates (%) on AR dataset.

NSC	SRC	DLSI	DKSVD	FDDL
92.0	97.5	97.5	89.2	97.5
LCKSVD	COPAR	JDL	DPL	**Ours**
97.8	98.3	98.5	98.3	**99.5**

Table 4: The recognition rates (%) on LFWa dataset.

SVM	SRC	DLSI	DKSVD	FDDL
63.0	72.7	73.8	65.9	74.8
LCKSVD	COPAR	JDL	DPL	**Ours**
66.0	72.6	72.8	74.6	**78.1**

recognition in the wild. LFWa(L. Wolf 2009) is a large-scale database, which contains variations of pose, illumination, expression, misalignment and occlusion, etc. 143 subjects with no less than 11 samples per subject are chosen (4174 images in total). For each person the first 10 samples are used for training data with the remaining samples for testing. Histogram of Uniform-LBP is extracted via dividing a face image into 10×8 patches. Table 4 illustrates the comparison of all methods. Similar to the results on AR, ASDL-UP achieves the best performance. Especially, the proposed ASDL-UP has over 5% improvement compared to the synthesis hybrid dictionary learning models (e.g., JDL).

Gender classification

There are total 13143 face images in the aligned labeled face in the wild (LFWa) (L. Wolf 2009). LFW is a large-scale database, which contains variations of pose, illumination, expression, misalignment and occlusion, etc. Based on gender information provided by LFW and attribute information provided by (N. Kumar 2009), we remove some extremely difficult samples (i.e., ones with absolute gender attribute values less tan 0.1) and final get a dataset of 12702 face images, of which 9900 are male samples and 2802 are female samples. Some gender examples on LFW dataset are shown in Fig. 3 We randomly select 585 male samples and 622 female samples as the training set, while the remaining 11495 face images used as the test data. Histogram of Uniform-LBP is extracted from the face images, and 320d pca+block-lda feature are finally used the descriptor of face images. Here we require that the dictionary number is 40 for all dictionary learning algorithm.

The experimental results are listed in Table 5. It can be seen that ASDL-UP is visibly better than all the other mehtods. The accuracy of ASDL-UP is not only at least 1.5% higher than COPAR and JDL, but also 1.1% higher than DPL, which is the second best method.

Testing time comparison

The classification of ASDL-UP is very efficiently conducted because there is only multiplication of vectors in its classifier (i.e., Eq.(5)). The classification of other DL methods, such as DLSI, DKSVD, FDDL, LCKSVD, COPAR and JDL, all needs to solve a sparse coding problem, which has more time complexity than ASDL-UP. The average running times on two datasets using Matlab 2011a are listed in Table 6. It can be seen that the testing time of ASDL-UP is less than or similar to that of DPL, but much faster than sparse coding in other DL methods. More time complexity and running time comparison are presented in supplementary material.

Figure 3: Some gender examples on LFW dataset.

Table 5: Gender classification accuracy (%) on LFW dataset.

CRC	SRC	DLSI	DKSVD	FDDL
83.7	90.5	87.9	87.0	90.2
LCKSVD	COPAR	JDL	DPL	**Ours**
92.4	91.4	85.2	92.0	**93.1**

Table 6: Average running time (second) on two datasets.

method	Sparse Coding	DPL	**ASDL-UP**
UCF sport action	1.3e-3	4e-6	**3e-6**
LFW gender	9.2e-2	1.4e-5	**1.4e-5**

Concluding Remarks

This paper presented a novel discriminative analysis-synthesis dictionary learning model (ASDL-UP) for universality-particularity representation based classification. In ASDL-UP, a pair of analysis dictionary and synthesis dictionary has been learned for the universality-particularity representation based classification. Because analysis-synthesis dictionary can give a more complete view of data representation, the learned dictionary pair more effectively utilizes the discrimination of universality representation and particularity representation of data. Moreover, discriminative universal analysis coding coefficient term is designed to preserve the class label information, while Fisher-like regularizations on the class-particular analysis-synthesis representation are proposed to ensure a discriminative analysis-synthesis representation. We also proposed an efficient solving algorithm for ASDL-UP. The ASDL-UP was extensively evaluated on several image classification tasks. Experimental results demonstrated that ASDL-UP outperforms some state-of-the-art methods.

Acknowledgments

This work is partially supported by the National Natural Science Foundation for Young Scientists of China (Grant no. 61402289), National Science Foundation of Guangdong Province (Grant no. 2014A030313558), National Natural Science Foundation of China (Grant no. 61272050) and Natural Science Foundation of Guangdong Province (Grant no. 2014A030313556).

References

A. Castrodad, G. S. 2012. Sparse modeling of human actions from motion imagery. *Intl Journal of Computer Vision* 100:1–15.

A. Martinez, R. B. 1998. The ar face database. Technical report, VC Technical Report.

B.A. Olshausen, D. F. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609.

H.R. Wang, C.F. Yuan, W. H. C. S. 2012. Supervised class-specific dictionary learning for sparse modeling in action recognition. *Pattern Recognition* 45(11):3902–3911.

I. Ramirez, P. Sprechmann, G. S. 2011. Classification and clustering via dictionary learning with structured incoherence and shared features. In *ICCV*.

J. Mairal, F. Bach, J. P. G. S. A. Z. 2008. Learning discriminative dictionaries for local image analysis. In *CVPR*.

J. Mairal, F. Bach, J. P. 2012. Task-driven dictionary learning. *IEEE TPAMI* 34(4):791–804.

J. Wright, A.Y. Yang, A. G. S. S. Y. M. 2009. Robust face recognition via sparse representation. *IEEE TPAMI* 31(2):210–227.

J. Wright, Y. Ma, J. M. G. S. T. H. S. Y. 2010. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE* 98(6):1031–1044.

J.C. Yang, J. Wright, Y. M. T. H. 2008. Image super-resolution as sparse representation of raw image patches. In *CVPR*.

J.C. Yang, K. Yu, Y. G. T. H. 2009. T. linear spatial pyramid matching using sparse coding for image classification. In *CVPR*.

L. Shen, S.H. Wang, G. S. S. J. Q. H. 2013. Multi-level discriminative dictionary learning towards hierarchical visual categorization. In *CVPR*.

L. Wolf, T. Hassner, Y. T. 2009. Similarity scores based on background samples. In *ACCV*.

L. Zhang, M. Yang, X. F. 2011a. Sparse representation or collaborative representation: Which helps face recognition? In *ICCV*.

L. Zhang, M. Yang, X. F. 2011b. Sparse representation or collaborative representation, which helps face recognition? In *ICCV*.

M. Aharon, M. Elad, A. B. 2006. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE TSP* 54(11):4311–4322.

M. Elad, M. A. 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE TIP* 15(12):3736–3745.

M. Rodriguez, J. Ahmed, M. S. 2008. A spatio-temporal maximum average correlation height filter for action recognition. In *CVPR*.

M. Yang, L. Zhang, X. F. D. Z. 2011. Fisher discrimination dictionary learning for sparse representation. In *ICCV*.

M. Yang, L. Van Gool, L. Z. 2013. Sparse variation dictionary learning for face recognition with a single training sample per person. In *ICCV*.

M. Yang, D.X. Dai, L. S. L. V. G. 2014. Latent dictionary learning for sparse representation based classification. In *CVPR*.

N. Kumar, A.C. Berg, P. B. S. N. 2009. Attribute and simile classifiers for face verification. In *ICCV*.

N. Zhou, J. F. 2012. Learning inter-related visual dictionary for object recognition. In *CVPR*.

Q. Qiu, Z.L. Jiang, R. C. 2011. Sparse dictionary-based representation and recognition of action attributes. In *ICCV*.

Q. Zhang, B. L. 2010. Discriminative k-svd for dictionary learning in face recognition. In *CVPR*.

Q. Zhang, B. X. L. 2012. Mining discriminative components with low-rank and sparsity constraints for face recognition. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

R. Rubinstein, A.M. Bruckstein, M. E. 2010. Dictionaries for sparse representation modeling. *Proceedings of the IEEE* 98(6):1045–1057.

R. Rubinstein, T. Peleg, M. E. 2013. Analysis k-svd: A dictionary-learning algorithm for the analysis sparse model. *IEEE TSP* 61(3):661–677.

R.Rubinstein, M. E. 2014. Dictionary learning for analysis-synthesis thresholding. *IEEE TSP* 62(22):5962–5972.

S. Bengio, F. Pereira, Y. S. D. S. 2009. Group sparse coding. In *NIPS*.

S. Gu, L. Zhang, W. Z. X. F. 2014. Projective dictionary pair learning for pattern classification. In *NIPs*.

S. Kong, D.H., W. 2012. A dictionary learning approach for classification: Separating the particularity and the commonality. In *ECCV*.

S. Sadanand, J. C. 2012. Action bank: A high-level representation of activeity in video. In *CVPR*.

W. Liu, Z. Yu, M. Y. L. L. Y. Z. 2015a. Joint kernel dictionary and classifier learning for sparse coding via locality preserving k-svd. In *ICME*.

W. Liu, Z. Yu, M. Y. 2015b. Jointly learning non-negative projection and dictionary with discriminative graph constraints for classification. *arXiv preprint arXiv:1511.04601*.

Z.L. Jiang, Z. Lin, L. D. 2013. Label consistent k-svd: Learning a discriminative dictionary for recognition. *IEEE TPAMI* 35(11):2651–2664.

Supplementary Material: Analysis-Synthesis Dictionary Learning for Universality-Particularity Representation based Classification

Meng Yang[†], Weiyang Liu[†‡], Weixin Luo[†], and Linlin Shen[†]
[†]College of Computer Science & Software Engineering, Shenzhen University, China
[‡]School of Electronic & Computer Engineering, Peking University, China

Summary

In the supplementary material, we give more analysis and experimental results for the proposed ASDL-UP. In the following sections, we will present the algorithm of dictionary atom pair updating in universal ASDL, control experiments to evaluate different components of ASDL-UP, analysis-synthesis dictionary learned from face images, recognition on the dataset of Scene-15 (S. Lazebnik 2007), and the time complexity analysis of ASDL-UP.

Dictionary atom pair updating in Universal ASDL

In universal ASDL, an analysis dictionary atom, a synthesis dictionary atom, and a column vector of the projection matrix $\boldsymbol{W}$ is solved via

$$\min_{\boldsymbol{o}_j,\boldsymbol{p}_j} \left\| [\boldsymbol{E}_j^J; \boldsymbol{0}] - \boldsymbol{o}_j \boldsymbol{p}_j^T \left[\boldsymbol{X}^J; \epsilon \boldsymbol{X}^{\bar{J}} \right] \right\|_2^2 \text{ s.t. } \|\boldsymbol{o}_j\|_2^2 \leq 1 \quad (1)$$

where $\boldsymbol{o}_i = [\boldsymbol{d}_i^0; \boldsymbol{w}_i]$. Let $\boldsymbol{F} = [\boldsymbol{E}_j^J; \boldsymbol{0}]$ and $\boldsymbol{Y} = \left[\boldsymbol{X}^J; \epsilon \boldsymbol{X}^{\bar{J}} \right]$, Eq.(1) is rewritten as

$$\min_{\boldsymbol{o}_j,\boldsymbol{p}_j} \left\| \boldsymbol{F} - \boldsymbol{o}_j \boldsymbol{p}_j^T \boldsymbol{Y} \right\|_2^2 \text{ s.t. } \|\boldsymbol{o}_j\|_2^2 \leq 1 \quad (2)$$

which is similar to Eq.(11) of (R.Rubinstein 2014) except the constraint.

Based on the solving algorithm shown in Fig.2 of (R.Rubinstein 2014), the dictionary atom pair updating algorithm in Universal ASDL is summarized in Algorithm 1. Based on the analysis of rank-one approximation in (R.Rubinstein 2014), Algorithm 1 will well solve the problem of dictionary atom pair updating in universal ASDL.

Control experiments

In this section, we present some control experiments to evaluate different components of the proposed ASDL-UP. When there is no particularity representation, ASDL-UP changes to ASDL for university representation (ASDL-U), which is a discriminative version of (R.Rubinstein 2014). When there is no universality representation, ASDL-UP changes to ASDL for particularity representation (ASDL-P), similar to DPL (S. Gu 2014). If we only require synthesis dictionary learning for university-particularity representation, ASDL-UP changes to synthesis DL-UP (SDL-UP), similar to JDL (N. Zhou 2012) and COPAR (S. Kong 2012). It is a little complicated if there is no synthesis dictionary in ASDL-UP. To avoid a trivial solution of dictionary learning, we replace the synthesis dictionary by a predefined matrix and only update the analysis dictionary. In that case ASDL-UP changes to analysis DL-UP (ADL-UP). We conduct classification experiments of ASDL-UP, ASDL-U, ASDL-P, ADL-UP and SDL-UP on AR dataset and UCF sport action dataset (some examples are shown in Fig. 1). Here the detailed experimental setting is the same to that in the main body of this paper. The experimental results are listed in Table 1. We can observe that ASDL-UP achieves the highest accuracy while ADL-UP performs worst. Without synthesis dictionary, analysis dictionary may not have a good generalization ability for testing signals. The other three dictionary learning methods can only perform well in a single test, which show that analysis-synthesis dictionary learning and universality-particularity representation are beneficial to each other.

Algorithm 1 Dictionary atom pair updating in Universal ASDL

1: **Input:** Matrices $\boldsymbol{F}$ and $\boldsymbol{Y}$
2: **Procedure**
3: compute the SVD: $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$
$\triangle = diag(s_1^{-1}, \cdots, s_n^{-1}, \cdots)$
$\hat{\boldsymbol{Y}} = \triangle \boldsymbol{U}^T \boldsymbol{Y}$
$\{\boldsymbol{o}_j, \hat{\boldsymbol{p}}_j\} = \arg\min_{\boldsymbol{o}_j,\hat{\boldsymbol{p}}_j} \left\| \boldsymbol{F}\hat{\boldsymbol{Y}}^T - \boldsymbol{o}_j \hat{\boldsymbol{p}}_j^T \right\|_2$
$\boldsymbol{p}_j^T = \hat{\boldsymbol{p}}_j^T \triangle \boldsymbol{U}^T$
if $\|\boldsymbol{o}_j\|_2 > 1$
  $\boldsymbol{p}_j^T = \boldsymbol{p}_j^T \cdot \|\boldsymbol{o}_j\|_2$
  $\boldsymbol{o}_j = \boldsymbol{o}_j \div \|\boldsymbol{o}_j\|_2$
end
4: **END**

Table 1. The recognition rates (%) of variants of ASDL-UP. 1: UCF Sport action dataset, 2: AR dataset

	ASDL-U	ASDL-P	ADL-UP	SDL-UP	**ASDL-UP**
1	93.6	92.1	89.3	90.7	**95.0**
2	96.3	98.8	75.5	98.5	**99.5**

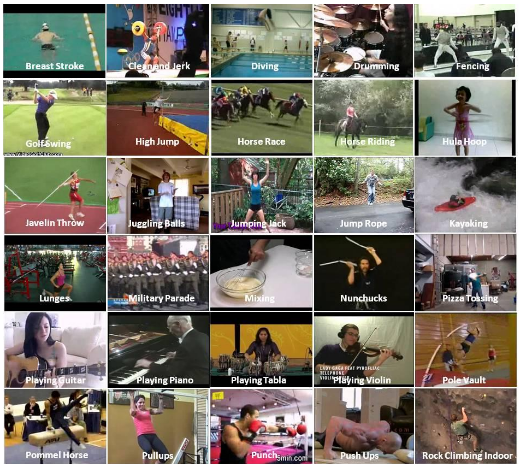


Figure 1. Some action examples on UCF 50 action dataset.

AS dictionary learned from face images

To further understand the learned dictionaries, we show the learned analysis-synthesis dictionary on AR database with 25×25 face images. Fig.2(a) shows the learned analysis-synthesis dictionary for university representation, from which we can observe some rough facial structures and some small and local distractions. The 4^{th} class analysis-synthesis dictionary and 70^{th} class analysis-synthesis dictionary for particularity representation are presented in Figs.2(b) and 2(c), respectively. We can see that class-specific synthesis dictionary consists of discriminative reconstruction bases for different subjects, while class-specific analysis dictionaries selects discriminative facial parts, in which entries of analysis dictionary have quite high/low values.

Scene categorization

The fifteen scene dataset was introduced in (S. Lazebnik 2007). Each category of natural scene has 200 to 400 images, and the average image size is about 250×300 pixels. The fifteen scenes contain bedroom, kitchen, and country scenes. Following the experimental setting in (Z.L. Jiang 2013), 100 images per category are chosen as the training data, with the rest as the testing data. The image descriptor is generated by extracting SIFT feature in local region, encoding local patch feature, and max pooling in spatial pyramid. In our pa-per, we use the 3000-d PCA feature of the spatial pyramid features provided by Compared to the traditional synthesis dictionary learning methods, such as DK-SVD, LC-KSVD, K-SVD, the proposed ASDL-UP has over 5% improvements. Compared to the synthesis dictionary learning for universality-particular representation, e.g., COPAR and JDL, the proposed ASDL-UP has about 1% improvement.

Table 2. The classification accuracy (%) on Scene-15 dataset.

CRC	SRC	KSVD	DKSVD	FDDL
97.7	91.8	86.7	89.1	97.7
LCKSVD	COPAR	JDL	DPL	**ASDL-UP**
92.9	97.5	96.7	97.7	**98.3**

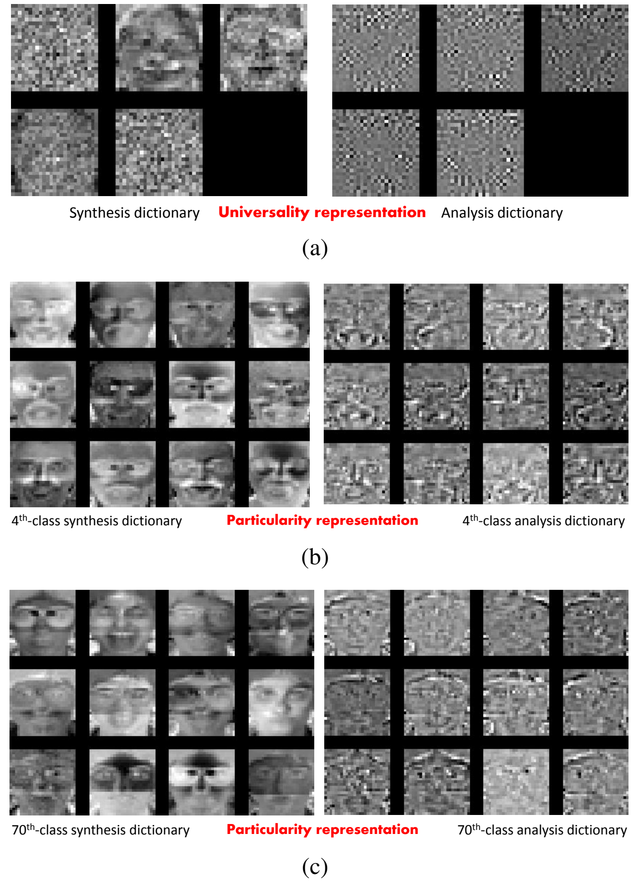


Figure 2. An example of learned dictionaries via ASDL-UP on AR database. To better visualize the atoms, we normalize the values of atoms and make them range from 0 to 255.

Time complexity analysis of ASDL-UP

The training time complexity of the proposed ASDL-UP is a little larger than that of dictionary pair learning (D-PL) (S. Gu 2014) since an additional universal analysis-synthesis dictionary learning (ASDL) needs to be solved. DPL has shown its faster training and testing speeds compared to other state-of-the-art dictionary learning approaches (S. Gu 2014). Compared to DPL, the main computation burden of universal ASDL comes from rank-one approximation to solve Eq.(9). Fortunately, we find that the rank-one approximation could be efficiently solved in most cases. We listed the training time of ASDL-UP and DPL (S. Gu 2014) using Matlab2011a with the desktop of i7 4.0GHz CPU and with 16G RAM: 0.845 second for ASDL-UP and 0.0015 second for DPL on action recognition of UCF sports action dataset, and 39.02 second for ASDL-UP and 21.04 second for DPL on gender classification of LFWa. It can be seen that the training of ASDL-UP is still fast.

When all the analysis-synthesis dictionaries are learned, the classification is very efficiently conducted. Denote by m be the dimensionality of the test image, $\boldsymbol{y}$. Let n_k and n_0 be the number of atoms in the universal dictionary and k^{th} class-particular dictionary. Therefore the time complexity of testing (i.e., Eq.(5)) is $O\left(\sum_k m\left(2n_k + c + n_0\right)\right)$, which is

Table 3. The average running time (second) on different datasets.

method	UCF sport action	LFW	Scene 15
DPL	4e-6	1.4e-5	1.16e-2
Sparse coding	1.3e-3	9.16e-2	3.0
ASDL-UP	3e-6	1.4e-5	1.18e-2

very low. The classification of other DL methods, such as DLSI, DKSVD, FDDL, LCKSVD, COPAR and JDL, all needs to solve a sparse coding problem, which has more time complexity (e.g,$O(m^2(n)^\varepsilon)$ where $\varepsilon \geq 1.2$ (S. J. Kim 2007) than the proposed ASDL-UP. The average running times on three datasets are listed in Table 3 (Table 3 is the table in supplementary martial). It can be seen that the testing time of ASDL-UP is less than or similar to that of DPL, but much faster than the sparse coding in other DL methods.

References

N. Zhou, J. F. 2012. Learning inter-related visual dictionary for object recognition. In *CVPR*.

R.Rubinstein, M. E. 2014. Dictionary learning for analysis-synthesis thresholding. *IEEE TSP* 62(22):5962–5972.

S. Gu, L. Zhang, W. Z. X. F. 2014. Projective dictionary pair learning for pattern classification. In *NIPs*.

S. J. Kim, K. Koh, M. L. S. B. D. G. 2007. A interior-point method for large-scale l_1-regularized least squares. *IEEE Journal on Selected Topics in Signal Processing* 1(4):606–617.

S. Kong, D.H., W. 2012. A dictionary learning approach for classification: Separating the particularity and the commonality. In *ECCV*.

S. Lazebnik, C. Schmid, J. P. 2007. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*.

Z.L. Jiang, Z. Lin, L. D. 2013. Label consistent k-svd: Learning a discriminative dictionary for recognition. *IEEE TPAMI* 35(11):2651–2664.