Low Resolution Face Recognition with Pose Variations Using Deep Belief Networks

Miaozhen Lin
School of Software
Dalian University of Technology
Dalian, China

Xin Fan
School of Software
Dalian University of Technology
Dalian, China

***Abstract*—In practice face recognition sometimes encountered by low resolution (LR) face images with varying poses, which degrade the performance significantly. To address this problem, we propose an approach that applies deep belief network (DBN) to handle the non-linearity caused by pose variations. The manifold assumption states that point-pairs from high resolution (HR) manifold share the topology with the corresponding LR manifold. Inspired by this assumption, we learn the relationship between HR manifold and LR manifold by sending both HR images and LR images to a deep architecture. High performance is achieved in the experiment on ORL and UMIST, in which great facial pose variations present.**

Keywords-face recognition; pose variation; low resolution; deep belief network

I. INTRODUCTION

Face recognition has been attracting more and more attentions recently. The recognition rate of most face recognition systems declines sharply on non-frontal faces despite of their acceptable rates under strictly controlled conditions (e.g. frontal faces under designated lighting). Furthermore, the low quality of face images dramatically degrades the performance of face recognition [1].

With the increasing importance attached to public safety, wide range of surveillance systems have been deployed in public places. Taking the cost of facility and storage into account, people tend to use cheap devices. On the other hand, cameras are often far from subjects, resulting in quite small faces in images, with the purpose of enlarging the surveillance scope and the people's self-preservation consciousness. Besides the configuration and quality of capturing devices, the subjects are also conducting uncontrolled movements in practice that yield non-frontal faces in the field of views. Pose variations along with LR place an obstacle to the wide application of automatic face recognition to real-world surveillance systems. In this paper, we present a novel method that is able to simultaneously handle both pose and LR problems. The performance of face recognition based on classical features (e.g. Eigen-face [10] and Fisher-face [11]) for LR images can be improved by using the constraint that LR images and their original counterparts of HR share the common intrinsic geometries [2,5,6,9]. Unfortunately, pose variations bring nonlinearity to the underlying geometry, which disables classical linear feature extractors. Approximation by multiple locally linear mappings [5] or enforcement of non-linear mappings on linear features [9] might be helpful for dealing with limited pose changes. However, a wide range of pose changes inevitably lower down the recognition rates of these approaches. We employ the advanced deep belief networks (DBNs) in order to capture the nonlinear connections between LR/HR face images with pose variations. Attributing to the non-linearity in the networks, artificial neural networks (ANN) are attractive for non-linear classification problems. Theoretical and neurophysiological studies show that it requires deep architectures, i.e. those consisting of several layers of nonlinear processors. Multilayer neural networks are perhaps the best example of such models with deep architectures. Back-propagation (BP) [16] was one of the first learning algorithms for these deep networks that could learn multiple layers representation. But BP does not work well in particle when training models that contain more than a few layers [17]. Recently, Hinton et al. [14] introduce a fast learning algorithm for deep, hierarchical probabilistic model called deep belief networks [14, 15]. This type of networks does not only act as a classifier but also a non-linear feature. This characteristic makes it possible to discover the common non-linear structures of LR face images and their corresponding HR ones. Significantly high recognition rate is achieved by our approach based on DBNs as shown in the experiments on ORL and UMIST data sets in which great pose variations present.

II. RELATED WORK

Researchers exert themselves in the struggle for improving the performance of face recognition. Compared with HR images, LR images lose a great deal of information, which is important to distinguish face identities. Reconstructing the lost information of LR face images came very naturally to solve the problem of LR face recognition, which actually contains two steps, i.e., super-resolution (SR) as the first step, followed by a generic face recognition process for the super-resolved face images. Baker and Kanade estimate the HR face image from an input LR counterparts based on face priors in [2]. Freeman et al. [3] use a Markov network to model the relationship between the LR images and the HR counterparts. Like[4], Liu et al. [5] propose to integrate a holistic model represented by principal component analysis (PCA) and a local model using Markov Network between LR and HR for SR reconstruction. In recent years, a series of SR methods based on a manifold assumption (called locally linear embedding (LLE)-like methods [4]) have

978-1-4244-9306-7/11/$26.00 ©2011 IEEE

been presented. The assumption claims that point-pairs from the LR manifold (LRM) and the corresponding HR manifold (HRM) possess similar local geometry. Chang et al. [6] propose an algorithm based on neighborhood embeddings.

Unfortunately, most SR algorithms are not designed for recognition but for visual enhancement of images. Explicit SR reconstruction of facial textural details in pixel domain may not be able to significantly improve the performance of recognition. Therefore, some algorithms attempt to avoid explicit SR in the image or pixel domain. Gunturk et al. [8] first proposed an approach to perform SR reconstruction in the feature domain by accumulating the information from a series of LR observations. P. Hennings et al. [12] propose a joint objective function, which explores the information given training LR and HR images under constraints between HR and LR images. Li's method shows [5] fairly good performance by coupled locality preserving mapping (CLPM). Huang et al. [9] enhance the correlation between HR and LR features by using Canonical correlation analysis (CCA) increasing the rate of neighborhood preserving and achieves pretty good performance. The success of these methods attributes to the rationality of manifold assumption once more. But when it comes to face images with pose variant, the performance of method [12] decreases drastically, which may ascribe to the lack of handling the differences caused by pose; also, it is quite time-consuming since optimization executed for each test image with regard to each enrollment.

We use DBNs to obtain the nonlinear features shared by LR/HR images and the recognition of a probe image runs fast once the networks are well trained.

III. THE DEEP BELIEF NETWORK

In this section, we will first review the basic block of the model called restricted Boltzmann machine and its inference, and the holistic model DBN, and then we introduce an effective way to train a DBN greedily layer-by-layer. In addition, we explain the importance to fine-tune the parameters of all layers in the networks, and also provide strategies on making use of supervised information such as labels or inputs. Finally, we describe the details on face recognition using DBN.

A. Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is a type of Markov random field (MRF), which is a Boltzmann machine [18] under the constraint that there are no visible-to-visible and hidden-to-hidden connections. In general, we will rarely be interested in learning a fully connected Boltzmann machine; instead, we will focus on RBM.

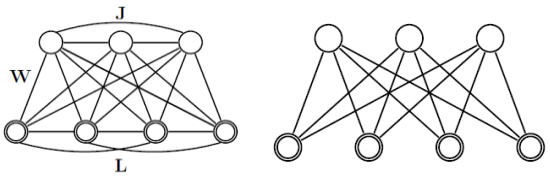

Figure 1. Left: Boltzmann machine Right: restricted Boltzmann machine

RBM is a network of symmetrically coupled stochastic binary units. It consists of a set of visible units $v \in \{0,1\}^D$ and a set of hidden units $h \in \{0,1\}^F$ (see Fig.1). The energy of the state {v, h} is defined as:

$$E(v,h;\theta) = -v^T W h - b^T v - a^T h, \tag{1}$$

where $\theta = \{W, a, b\}$ are the parameters: W represents visible-to-hidden symmetric interaction terms, a, b are bias terms. The joint distribution over the visible and hidden vector is:

$$p(v,h;\theta) = \frac{1}{z(\theta)} \exp(-E(v,h;\theta)), \tag{2}$$

$$z(\theta) = \sum_v \sum_h \exp(-E(v,h;\theta)). \tag{3}$$

The probability that model assigns to a visible vector v is:

$$p(v,h;\theta) = \frac{1}{z(\theta)} \sum_h \exp(-E(v,h;\theta)). \tag{4}$$

The conditional probability distributions over hidden and visible units can easily derived from (2):

$$p(h_j = 1 \mid v) = g(\sum_i W_{ij} v_i + a_j), \tag{6}$$

$$p(v_i = 1 \mid h) = g(\sum_j W_{ij} h_j + b_i), \tag{7}$$

where $g(x) = 1/(1+\exp(-x))$ is the logistic function. Easily to derive the log-likelihood with regard to the parameters:

$$\frac{\partial \log P(v;\theta)}{\partial W} = \alpha(E_{data}[v h^T] - E_{model}[v h^T]), \tag{8}$$

$$\frac{\partial \log P(v;\theta)}{\partial a} = \alpha(E_{data}[h] - E_{model}[h]), \tag{9}$$

$$\frac{\partial \log P(v;\theta)}{\partial b} = \alpha(E_{data}[v] - E_{model}[v]). \tag{10}$$

where α is a learning rate. $E_{data}[.]$ called data-dependent expectation, donates an expectation with regard to the completed data distribution $P_{data}(v,h;\theta) = P(h \mid v;\theta) P_{data}(v)$ with $P_{data}(v)$ representing the empirical distribution. $E_{model}[.]$ called model's expectation, donates an expectation with regard to the distribution defined by (2).

It's intractable to exact maximum likelihood learning since computation of model's expectation take time, which is exponential in D or F. For RBM, learning can be carried out efficient using Contrastive Divergence [19] (CD), which obtain exact samples from the condition $P(h|v;\theta)$, while it is intractable when learning full Boltzmann machines. Hinton et al. proposed an approach to approximate both expectations by using Gibbs sampling. In practice, learning is done by using persistent Markov chains to estimate the model's expectation and a variational approach to estimate the data-dependent expectation.

B. Deep Belief network

DBNs are probabilistic models that contain several layers of hidden layers, in which each layer capture high-order correlation between hidden units in the layer below. Top two layers of DBN consists an undirected bipartite graph and layers below top two form a directed sigmoid belief network as show in Fig. 2. The main building block of a DBN is an RBM.

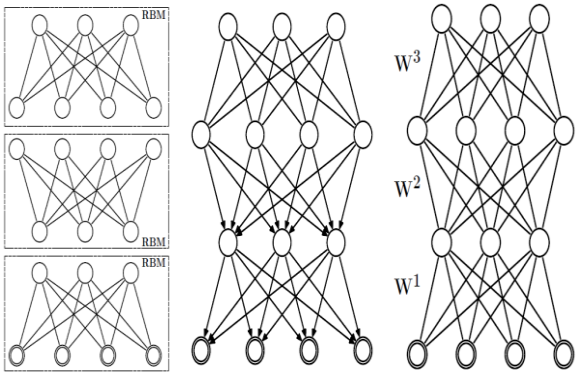

Figure 2. Left: Greedy learning a stack of RBM's in which the samples from the lower-level RBM are used as the data for training the next RBM. Middle: corresponding Deep Belief Network learned from left panel. Right: Deep Boltzmann Machine with the same hierarchy as DBN in middle.

C. Greedy learning algorithm for DBNs

The A greedy layer-wise training algorithm was proposed by Hinton et al. [20] to train a DBN one layer at a time. Bengio et al. [21] stress that greedy layer-wise learning of DBN hold well, pre-training one layer at a time in a greedy way, using unsupervised learning at each layer in order to preserve information from the input, and fine-tuning the whole network with respect to the ultimate criterion of interest. And this method shows its outstanding performance in quite a lot of tasks [22, 23]. A reasonable explanation for the apparent success of the layer-wise training strategy for DBNs is that pre-training helps to mitigate the difficult optimization problem of deep networks by better initializing the weights of all the layers.

The greedy layer-wise training algorithm for DBNs is quite simple, as illustrated as follow:

- Fit parameters W^1 by training 1st RBM taking data v as input.
- Freeze W^1 and use that trained RBM h^1 from $Q(h^1|v) = P(h^1|v,W^1)$ as the data for training the next layer of binary features with an RBM.
- Freeze W^2 that define the 2nd layer of features and use h^2 from $Q(h^2|h^1) = P(h^2|h^1,W^2)$ as the data for training the 3rd layer of binary features.
- Proceed recursively for the next layers.

D. Fine-tune

As a last training stage, it is possible to fine-tune the parameters of all the layers together. For example Hinton et al. [24] propose to use the wake-sleep algorithm to continue unsupervised training. Hinton et al. [20] also propose to optionally use a mean-field approximation of the posteriors. It is also possible to use this as initialization of all except the last layer of a traditional multi-layer neural network, using gradient descent to fine-tune the whole network with respect to a supervised training criterion. In the paper, we focus on the last one. We train each new hidden layer as the hidden layer of a one-hidden layer supervised neural network (taking as input the output of the previously trained layers), and then throw away the output layer and use the trained hidden layer as pre-training initialization, to map the output of the previous layers to a hopefully better representation. The final fine-tuning is done by adding a logistic regression layer on top of the network and training the whole network by stochastic gradient descent on the cross-entropy with respect to the target classification.

E. Face recognition using DBNs

In general, we mix up HR face images with LR face images, which interpolated by simple method according with size of HR images. Interpolation is employed for adjusting the size of HR and LR images, but not for recovering any high frequency information of LR images. And then we send face images with different resolution with pose variants into the DBNs as visible inputs. Then we learn millions of parameters of the model layer-by-layer greedily, which supervised by the reconstruction error between the layer and the layer below. At last, we fine-tune parameters in all layers by standard BP to minimize the cross-entropy error with respect to the target classification. Fig. 3 shows the model for recognition.

Compared with standard BP algorithm, the fine-tuning converges in rather short time. In our experiments, it only needs less than 20 iterations to tune-up all the parameters, while standard BP who has the same hierarchy as DBNs without pre-training, may not converge after thousands of iterations, which takes intolerant time in practice. What is worse, it usually trapped in local minimal.

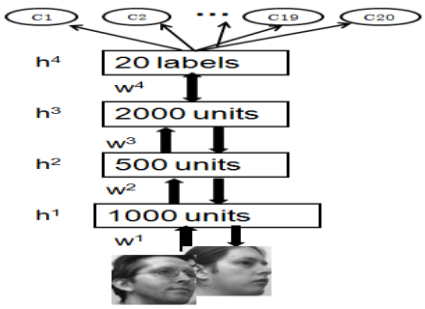

Figure 3. Face recogniton using DBN

IV. EXPERIMENTS

We carry out our experiments on UMIST database and ORL database. The face images in UMIST and ORL present significant pose variations. Given LR input faces (14×11 for UMIST, 7×6 for ORL), the recognition rate is as high as 100% and 100% respectively. The performance outperforms SR algorithms based on manifold assumption including Huang's method [9] and CLPM [5], Wang's method [25] and classical Gunturk's method [8], and works even better than the eigen-face recognition on original HR face images [10]. We will explain the experiment for each database in details as follows.

A. UMIST Database for Recognition

The UMIST database consists of 575 images over 20 individuals, each covering a wide range of multi-viewed face images, from profile to frontal views. The influence of pose variant will be studied in these experiments. We evaluate our method based on the training set including 10 images for each individual and the testing set choosing other 5 images for each individual. The train and test LR images with size 14× 11 pixels are generated by the operation of smoothing and down-sample, while HR face images with size 56×46 pixels. We resize the train LR images by nearest neighborhood interpolation to adjust the number of the visible units. The comparative methods are carried out with similar configuration. Our model consists of a DBN with structure like (56×46)-500-500-20. During the pre-training, we iterate 200 times for each RBM in DBN, then iterate 20 times in fine-tuning stage. Fig. 4 is images of an individual.

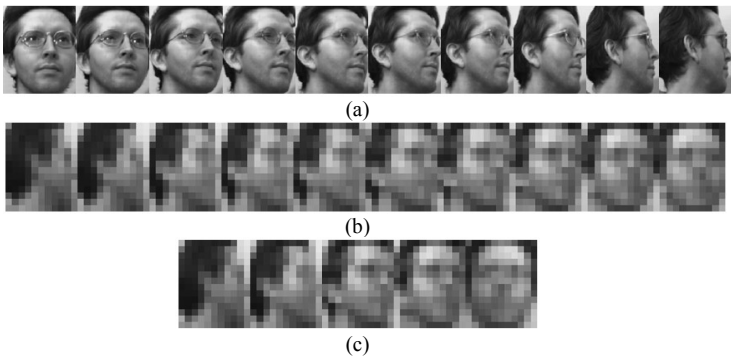

Figure 4. images of an individual in UMIST: (a) the HR training images with size 56×46, (b) the LR training images with size 14×11, (c) the LR testing images with size 14×11.

Face recognition rate among different methods have been shown in Fig. 5. Our method show its superiority compared with other methods. We select the best performance and performance with 100D features for each method among 70D, 90D, 100D, 120D and 154D to compare with our method. The best recognition rate of Huang's method, CLPM, Wang's method, Gunturk's method, LR-PCA, HR-PCA are 93%, 88%, 91%, 49%, 90%, 92% respectively, while with 100D features, CLPM, HR-PCA, LR-PCA all decline in performance. The performance of CLPM method and Gunturk's method is even worse LR-PCA methods. The reason lies in that CLPM method applies linear mapping to obtain the recognition features which is not suitable for handling the nonlinearity caused by pose variant, and Gaussian assumption on the PCA features in Gunturk's method is not applicable to the pose variations face images.

And we try to put images with different resolution such as 56×46, 28×23, 14×11, 7×6 pixels as the input of DBN, and generate the testing images with similar operation. In this experiment, we use the model mentioned above. And we achieve high recognition rate as 100%, even with size 7×6 pixels, which is not able to be recognition by human, as shown in Fig. 6.

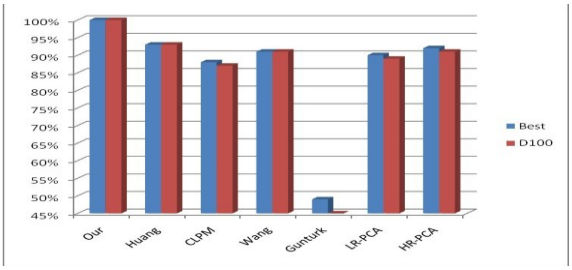

Figure 5. Recognition rate of UMIST over methods, where best denotes best perfermance for each method, and D100 denotes recognize with100D features

Figure 6. Testing images with 7× 6 pixels

B. ORL Database for Recognition

Experiment also has been conducted in ORL database. The ORL face image database consists of 40 individuals with ten images for each individual where presents lighting, expression and pose variants. In the experiment, we select five images for each individual for training, the other five for testing. LR images with size 7×6 pixels generated by the operation of smoothing and down-sample, while HR images with size 56×46 pixels. We resize the LR images by nearest neighborhood interpolate to adjust the number of the visible units. Our model consists of a DBN with structure like (56×46)-500-500-40. Because the capability of the samples is too small, we add the mirror images of all the training images to the train set. As a matter of convenience, we send images with different size 28×23, 14×11 pixels together as visible vector into the DBN. And we test our method by images with 3 different down-sampling rates with respect to HR images. To be excited that three group testing images are recognize correctly by DBN. And other method just test on images with size 8×8. Other methods are conducted with 20D, 30D, 40D, 50D and 64D features. Huang's method achieve recognition rate of 95% with 50D feature, CLPM, Wang's method, HR-PCA and Gunturk's method obtain recognition rate 91.5%, 91%, 91.5% and 91% respectively with 40D features and followed by LR-PCA with 84.5% using 64D features. The result in Fig.7 that the performance of our method is better than other comparative methods may attribute to charm of DBN, which shows attractive talent in non-linear representation.

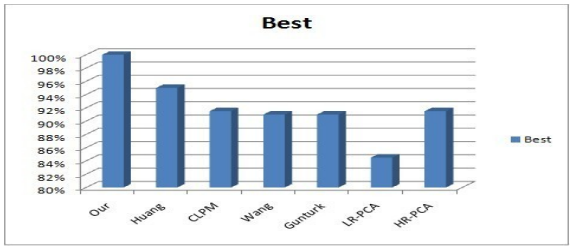

Figure 7. the best perfermance for each method in ORL

C. *Time Complexity Analysis*

And we evaluate the efficiency of our method compared with the other method in this experiment. Table I presents the mean runtime for each method on UMIST and ORL databases by 5 times. It can be seen obviously that time consuming in our method is pretty short. The time consumed by CLPM, Wang's methods are more than 1-8 times as ours, and CLPM is quite instable with the growth of the sample. Among these methods, Gunturk's takes the longest time which is not able to tolerate in real time applications. Based on the results, our method achieves best performance and best timesaving.

TABLE I. TIME COMPLEXITY FOR EACH METHOD

Method	**Databases**	
	UMIST	***ORL***
Ours	0.13	0.24
Huang	0.14	0.62
CLPM	0.33	3.21
Wang	0.64	1.56
Gunturk	9.24	183.12

CONCLUSION

In this paper, we revisit the manifold assumption which is the basis of some learning-based SR algorithms. We analyze the limitations of these methods and propose a novel approach based on Deep Belief Network to solve the problem of the LR face recognition with pose variant concurrently. Different from the traditional two-step methods, our algorithm put both HR and LR images into a same model which is under the assumption that HRM and LRM share the same topology. Owing to the non-linearity caused by pose variant, we apply the deep neural network which shows its talent in nonlinear representation. And the encouraging results verify the effectiveness of our method, as well as the rationality of our hypothesis. Take another look at it, our timesaving method, without any SR, is more suitable for real-time application.

The performance of real-world LR face recognition will be degraded by the factors of noise, geometric distortion, as well as facial expression, lighting, etc. Considering these factors with LR simultaneously, it will be effective and great potential in developing a face generative model which will be pursued in our future work.

ACKNOWLEDGMENT

This work is supported by Natural Science Foundation of China (no. 61003177) and the Fundamental Research Funds for the Central Universities (nos. DUT10ZD110 and 1600-852008).

REFERENCES

[1] O. Sezer, Y. Altunbasak, A. Ercil, "Face recognition with independent component-based super-resolution." in Proc. SPIE Visual Commun. Image Process., vol. 6077. San Francisco, CA, 2006, pp. 52-66.

[2] S. Baker and T. Kanade, "Limits on super-resolution and how to break them." PAMI, IEEE Trans. 2002, vol. 24,no. 9, pp. 1167-1183.

[3] W. Freeman, E. Pasztor, and O. Carmichael, "Learning Low-level Vision," IJCV, vol. 40, no. 1, pp. 25–47, 2000 .

[4] S. T. Roweis, L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding, " Science, vol. 290, 2000.

[5] B. Li, H. Chang, S. Shan and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings." Signal Processing Letters, IEEE, 2010, vol. 17,no. 1, pp. 20-23.

[6] H. Chang, D. Y. Yeung, et al. "Super-resolution through neighbor embedding. " CVPR, 2004.

[7] C. Liu, H. Y. Shum, et al. "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model." ,2001 .

[8] B. K. Gunturk, A. U. Batur, "Eigenface-domain super-resolution for face recognition." Image Processing, IEEE Trans. 2003, vol. 12, no. 5, pp.597-606.

[9] H. Huang, and H. He, "Super-Resolution Method for Face Recognition Using Nonlinear Mappings on Coherent Features." Neural Networks, IEEE, 2010, vol. 99, pp. 1-10.

[10] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces", IEEE 2010.

[11] P. N. Belhumeur and J. P. Hespanha, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection." PAMI, IEEE Trans.1997 vol. 19,no.7, pp. 711-720.

[12] P. Hennings-Yeomans and S. Baker, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces", IEEE 2008.

[13] X. Tan and S. Chen, "Face recognition from a single image per person: A survey." Pattern recognition, 2006, vol. 39, no. 9, pp. 1725-1745.

[14] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks." Science, 2006, vol. 313, no. 5786, pp. 504.

[15] R. Salakhutdinov and G. Hinton , "Deep Boltzmann machines", 2009.

[16] D. E. Rumelhart and G. Hinton, "Learning representations by back-propagating errors." Nature, 1986, vol. 323, no. 6088, pp. 533-536.

[17] Y. Bengio, P. Lamblin, "Greedy layer-wise training of deep networks", Technical Report, 2009, The MIT Press.

[18] D. H. Ackley, G. E. Hinton, et al., "A learning algorithm for Boltzmann machines." Cognitive science, 1985, vol.9, no. 1, pp. 147-169.

[19] G. E. Hinton, "Training products of experts by minimizing contrastive divergence." Neural computation, 2002, vol. 14, no. 8, pp. 1771-1800.

[20] G. E. Hinton, S. Osindero, et al., "A fast learning algorithm for deep belief nets." Neural computation, 2006, vol.18, no. 7, pp. 1527-1554.

[21] Y. Bengio, "Learning deep architectures for AI." Foundations and Trends(R) in Machine Learning, 2009, vol. 2, no. 1, pp. 1-127.

[22] V. Nair and G. Hinton, "3-d object recognition with deep belief nets." NIPS, 2009, vol. 22, pp. 1339-1347.

[23] R. Salakhutdinov and I. Murray, "On the quantitative analysis of deep belief networks". ICML. Helsinki, Finland, ACM, pp. 872-879, 2008.

[24] G. Hinton, P. Dayan, et al., "The wake-sleep algorithm for unsupervised neural networks." Science, 1995, vol. 268, no. 5214, pp. 1158.

[25] X. Wang and X. Tang, "Hallucinating face by eigen-transformation." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Trans. 2005, vol. 35, no. 3, pp. 425-434.