Face Behind Makeup

Shuyang Wang[1] **Yun Fu**[1,2]
[1]Department of Electrical & Computer Engineering,
[2]College of Computer & Information Science,
Northeastern University, Boston, MA, USA
{shuyangwang, yunfu}@ece.neu.edu

Abstract

In this work, we propose a novel automatic makeup detector and remover framework. For makeup detector, a locality-constrained low-rank dictionary learning algorithm is used to determine and locate the usage of cosmetics. For the challenging task of makeup removal, a locality-constrained coupled dictionary learning (LC-CDL) framework is proposed to synthesize non-makeup face, so that the makeup could be erased according to the style. Moreover, we build a stepwise makeup dataset (SMU) which to the best of our knowledge is the first dataset with procedures of makeup. This novel technology itself carries many practical applications, e.g. products recommendation for consumers; user-specified makeup tutorial; security applications on makeup face verification. Finally, our system is evaluated on three existing (VMU, MIW, YMU) and one own-collected makeup datasets. Experimental results have demonstrated the effectiveness of DL-based method on makeup detection. The proposed LC-CDL shows very promising performance on makeup removal regarding on the structure similarity. In addition, the comparison of face verification accuracy with presence or absence of makeup is presented, which illustrates an application of our automatic makeup remover system in the context of face verification with facial makeup.

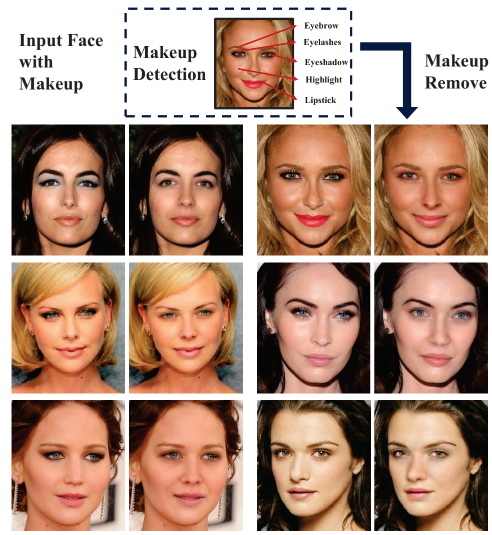

Figure 1: Framework and examples of our system. It first tells what and where the cosmetics are. Furthermore, the system can remove the detected cosmetics to virtually recover the original face (right image of each pair).

Introduction

Facial makeup has been ubiquitous in our daily life and social networks. It is quite common for women to wear cosmetics to hide facial flaws and appear to be more attractive. What information can we learn from face makeup and what if we can virtually remove the makeup? Take the security application for example, the use of makeup poses a significant challenge to biometric systems. Facial makeup is capable to alter and hide one's original appearance, leading to some more difficult verification tasks. Recently, Dantcheva et al. discussed the negative impact introduced by facial cosmetics to the face recognition problem (Dantcheva, Chen, and Ross 2012). Therefore, a system for detecting, analysing and even removing the makeup is eagerly needed.

In this work, we design a system to first detect and locate each makeup region, then further decompose each region of cosmetics to recover the original face. This function leads to lots of applications, for example, user can then virtually try on other cosmetics and the performance of face verification with makeup could also be improved. Regarding above system requirements, the most basic and also challenging tasks are makeup detection and makeup removal. Corresponding approaches to above two sub-problems will be introduced.

The first step of our system is to detect and locate the makeup region in a face. Different from previous two makeup detection works (Chen, Dantcheva, and Ross 2013; Guo, Wen, and Yan 2014), which use SVM and Adaboost as classifiers, our makeup detection method employs locality-constrained low-rank dictionary learning. Since makeup basically has several specific styles, learning a locality-constrained dictionary could give different styles of makeup a better representation.

Some digital makeup clone researches (Guo and Sim 2009; Du and Shu 2013) have been conducted to add virtual makeup to bare face images. The result seems encouraging by using image processing techniques like gradient editing and alpha blending on different color channels. Compared to adding makeup to naked face problem, however, we focus on a totally different one, makeup removal. Unlike covering human skin with virtual cosmetics, we design a method to recover the occluded original face, which is actually an ill-posed problem. For makeup removal, we consider it as a multi-step cross-modal problem, that is each makeup status is a modal and non-makeup modal can be synthesized from makeup modal stepwise. To this end, we propose a locality-constrained coupled dictionary learning (LC-CDL) algorithm. Besides, since this is the first work trying to do the makeup reversion, we collect a new dataset called Stepwise Makeup Dataset (SMU). Compared to other three existing makeup datasets, SMU dataset is the first one contains multiple steps of makeup images and each image was labeled with makeup status information in sub-region of face.

The main contributions of our work are: 1) we develop a system to detect facial makeup and reverse it automatically. To our best knowledge, we are the first to recover original face from makeup face; 2) we collect a stepwise makeup (SMU) dataset for the sake of makeup reversion and conduct several experiments on this dataset; 3) a new LC-CDL algorithm based on SCDL (Wang et al. 2012) is deployed to remove makeup, and a postprocess with Poisson editing and ratio-image merging is applied to produce a more realistic synthesis result; and 4) dictionary learning algorithm is introduced to detect facial makeup.

Related Works

In this section, we briefly discuss related works from both application and techniques. Facial beauty has recently begun to draw attentions in pattern recognition and computer vision areas. However the research topics are barely focus on face beauty estimation (Wang, Shao, and Fu 2014) while the research related with facial makeup is still quite limited. The research on makeup recommendation system arises these two years. In (Liu et al. 2013), Liu, et al. developed a system for hairstyle and facial makeup recommendation and synthesis. They applied candidate makeup into an original face and recommend to users the makeup which result a highest beauty score. This system produces appealing results but still has a lot of limitation, such as they can only deal with a non-makeup face.

Compared with works on makeup recommendation, the research dealing with makeup face is even rare. Dantcheva, et al. was the first work to explicitly establish the impact of facial makeup on a face recognition system (Dantcheva, Chen, and Ross 2012). They assembled two datasets, YouTube MakeUp database and Virtual MakeUp database, then tested the recognition performance before and after makeup with three face recognition features: Gabor wavlets, Local Binary Pattern and the commercial Verilook Face Tookit. Based on this work, there are two papers focusing on makeup face. In (Chen, Dantcheva, and Ross 2013), they detected the presence of makeup in face images based on a feature vector which contains shape, texture and color information. The other one (Guo, Wen, and Yan 2014) dealt with the verification problem. They extracted features from both makeup and non-makeup face then do the face matching based on correlation mapping. Above makeup face related works focus on the makeup's affection with face recognition and involve in a pre-processing of makeup detection, whereas our system further performing the makeup remove. Since makeup remove is essentially a novel cross-modal synthesis problem, which shares some basic assumptions with image super-resolution (Walha et al. 2014) or sketch-photo synthesis (Zhang et al. 2015) problem and dictionary learning based methods have been proved to be successfully in dealing with this kind of tasks. Thus, we briefly review some dictionary learning (DL) works.

Recent researches have contributed to the rapid growth in the theory and application of DL (Yang, Zhang, and Feng 2011) and low-rank regularization (Liu, Lin, and Yu 2010; Zhao and Fu 2015). The performance of problems like image classification has been improved dramatically with a well-adapted discriminative low-rank dictionary (Ma et al. 2012; Li, Li, and Fu 2014). In the cross-modal DL literature, Wang et al. (Wang et al. 2012) proposed semi-coupled dictionary learning to conduct sketch-photo synthesis. DL methods for cross-domain problems (Huang and Wang 2013; Ding, Ming, and Fu 2014) have been proposed in recent literature, aiming to capture more intrinsic structure and achieve a better representation.

In sparse based methods, the local structure was ignored with the assumption that each point has independent linear representation. To specifically encourage the coding to rely on the spatial consistency of neighbor sample points, Local Coordinate Coding (LCC) was presented recently (Wang et al. 2010). Motivated by above algorithms, we explore the makeup detection problem by adding locality regularization on low-rank dictionary learning, meanwhile, introduce locality constraint coupled dictionary learning (LC-CDL) to perform makeup remove. Thus, the makeup style information can be exploited in more thorough manner.

Makeup Detection

The few existing works related to automatic makeup detection are all based on SVM and Adaboost classifier (Chen, Dantcheva, and Ross 2013; Guo, Wen, and Yan 2014), which ignores the intrinsic local information of makeup. To find a better representation of makeup, we use locality-constrained low-rank dictionary learning algorithm given following three observations:

1. Most of the data related to makeup are unconstrained wild images which vary on pose, expression and illumination, sometimes even have occlusion on makeup region. To this end, error term is necessary to exorcise the sparse noises.
2. Cosmetics hide facial flaws, meanwhile inevitably cover some individual facial characteristics, which reduces the within-class identity variation. Therefore low-rank is introduced to discover a pure and compact dictionary.
3. Inspired by (Wang et al. 2010; Yu, Zhang, and Gong 2009) , under certain assumptions, locality plays more essential roles than sparsity. That means similar samples should be represented with similar features.

Locality-constrained low-rank dictionary learning

Define matrix $X = [X_1, X_2, \ldots, X_c]$ as the training data from c different classes (c equals to 2 as makeup and non-makeup number in this paper), $X \in \mathbb{R}^{d\times N}$, with d feature dimension and N total training samples. For i-th class samples X_i, a sparse representation matrix A_i is yield from a group of learned sub-dictionaries $D = [D_1, D_2, \ldots, D_c]$ and used to facilitate future classification task. The training data then can be represented by $X_i = DA_i + E_i$ with E_i as the sparse noises.

In this task dealing with makeup faces, the within-class samples usually have same style (e.g. red lips, black eye shadow) and lie in a low dimensional manifold. Therefore, the dictionary is expected to be learned with the most concise atoms by minimizing the matrix rank norm. As (Candès et al. 2011) suggested, the nuclear norm $\|D_i\|_*$ can be used as a surrogate for the rank of D_i, where $\|.\|_*$ denotes the sum of singular values of the matrix. In addition, as suggested by LCC (Yu, Zhang, and Gong 2009), locality plays more essential roles than sparsity given the fact that locality must guarantee sparsity but not necessarily vice versa. Specifically, the locality regularization uses the following criteria:

$$\min_a \|l_k \odot a_k\|_2^2, \text{ s.t. } \mathbf{1}^{\mathrm{T}} a_k = 1, \forall k, \tag{1}$$

where l_k is the locality adaptor for k-th coefficient a_k, and $\odot$ denotes the element-wise multiplication. Specifically,

$$l_k = \exp\left(\frac{[\mathrm{dist}(x_k, d_1), \ldots, \mathrm{dist}(x_k, d_n)]^{\mathrm{T}}}{\sigma}\right). \tag{2}$$

where $\mathrm{dist}(x_k, d_j)$ represents the Euclidian distance between input sample x_k and each basis vector d_j. Accordingly, locality constraint assigns different freedom to dictionary atoms proportional to their similarity with input data x_k.

Considering the low-rank constraint term on the sub-dictionaries as well as the locality regularization on the coding coefficients, each sub-dictionary have the following objective function:

$$\begin{aligned} &\min_{D_i, A_i, E_i} R(D_i, A_i) + \alpha\|D_i\|_* + \beta\|E_i\|_1 \\ &+\lambda \textstyle\sum_{k=1}^{n_i} \|l_{i,k} \odot a_{i,k}\|_2^2 \quad \text{s.t. } X_i = DA_i + E_i, \end{aligned} \tag{3}$$

where α and β are two balanced parameters. Note that $R(D_i, A_i)$ endows sub-dictionary D_i with the discriminability to well represent samples from i-th class by minimizing following:

$$R(D_i, A_i) = \|X_i - D_i A_i^i - E_i\|_{\mathrm{F}}^2 + \sum_{j=1,\neq i}^{c} \|D_i A_j^i\|_{\mathrm{F}}^2, \tag{4}$$

where A_i^j is the coefficient matrix of X_i over D_j.

Classification method

A linear classifier $\hat{W}$ is obtained using the multivariate ridge regression model (Zhang, Jiang, and Davis 2013):

$$\hat{W} = \arg\min_W \|Y - WA\|_2^2 + \lambda\|W\|_2^2, \tag{5}$$

where Y is the class label matrix, and A is the locality-constrained coefficients of X. This yields $\hat{W} = YA^{\mathrm{T}}(AA^{\mathrm{T}} + \lambda I)^{-1}$. For each incoming testing sample x_{test}, the predicted label vector is computed as $\hat{W} a_{\text{test}}$, where a_{test} is the coefficient vector learned from proposed model. Specifically, the position of the largest value in the label vector is used as the final class label assigned to the testing data.

Makeup Decomposition

To the best of our knowledge, this is the first work to recover a non-makeup face by automatically removing the cosmetics from a makeup face . The makeup decomposition problem can be formulated as follows: given an makeup image X_m, how to recover the associated nuke face X_n? This is totally different from previous published virtual makeup papers, which add makeup to a nuke face using makeup example. In this problem, the original face is almost covered up by cosmetics which makes it an ill-conditioned problem. A well-known fact is that makeup has some standard styles, which could benefit our makeup decomposition by learning these styles from training data with local structure. To solve this challenging problem, we propose a locality-constrained coupled dictionary learning method called LC-CDL.

Preprocessing

Accurate pixel-wise alignment is necessary for realistic face synthesis in our task, since we learn pair-wise dictionaries which require corresponding face region in before and after makeup. To establish a unified training set, all face images with the size of 150×130 are automatically aligned with 83 facial landmarks extracted through Face++[1] (Inc. 2013). These fiducial points define an affine warp, which is then used in Thin Plate Spline (Donato and Belongie 2002) method to warp the images into a canonical form.

As we can see, the makeup styles are usually complicated or varied in the dataset and practical application. Some face may only have lipsticks while some could apply eye shadow and face foundation. That makes it impossible to recover all kinds of makeup by training only one pair of dictionaries. Therefore, different pairs of dictionaries should be assigned to different face regions. For this reason, we separate whole face into three region (facial skin, upper/lower mouth lips, left/right eye) in the preprocessing step.

Locality-constrained coupled dictionary learning

Consider the situation in the same style makeup procedure, all subjects should have similar changes from previous status to the next, which suggests that the styles can be converted between each other in a hidden space. We first illustrate the procedure in a simple two-steps situation. Denote X_m and X_n as the sample sets formed by the image patch pairs of makeup and non-makeup. Since each pair of images indicates the similar makeup effect, it is reasonable to

[1]The Face++ Research Toolkit can be downloaded from http://www.faceplusplus.com/.

Algorithm 1 Updating coefficients A_m

Input: Training data X_m, Dictionary D_m, Projection P coupled coefficient A_n, Parameters δ, γ

1. **for** i = 1 to N **do**
2. **for** i = 1 to number of atoms in (D_m) **do**
   $l_{m,k} \longleftarrow \exp(\|x_{m,k} - d_{m,i}\|^2/\sigma)$
3. **end for**
4. **end for**
5. $l_m \longleftarrow$ **normalize**$_{(0,1]}(l_m)$
6. $A_m \longleftarrow \arg\min\limits_{A_m} \|X_m - D_m A_m\|_F^2 + \gamma \|l_m \odot a_m\|^2$
7. **remove** $D_m(id)$, $\{id \mid a_{m,k}(id) \le 0.01\}$
8. $A_m \longleftarrow \arg\min\limits_{A_m} \|[X_m; \delta P A_n] - [D_m; \delta I] A_m\|_F^2$

Output: A_m

assume that there exists some kind of transformation from makeup to non-makeup for each sample. To avoid addressing the flexibility of face structures in each status with a too strong assumption, we assume there exists a dictionary pair D_m and D_n over which the two representations A_m and A_n have a stable mapping P. In LC-CDL, we further employ locality constraint to seek for the projection between each adjacent status meanwhile exploit the makeup style information in the dataset's manifold. Once the projections between each pair of coefficients are learned, we can perform the makeup decomposition rely on the relationship in the learned sparse coefficients. The desired coupled dictionaries are obtained by minimizing the following objective function:

$$\begin{aligned} &\min_{D_m, D_n, P} \|X_m - D_m A_m\|_F^2 + \|X_n - D_n A_n\|_F^2 + \\ &\gamma(\sum_{k=1}^N \|l_{m,k} \odot a_{m,k}\|_2^2 + \sum_{k=1}^N \|l_{n,k} \odot a_{n,k}\|_2^2) \\ &+ \delta \|A_m - P A_n\|_F^2 + \lambda_P \|P\|_F^2 \\ &\text{s.t. } \|d_{m,i}\|_{l_2} \le 1, \|d_{n,i}\|_{l_2} \le 1, \forall i, \end{aligned} \tag{6}$$

where γ, δ and λ_P are regularization parameters to balance the terms in the energy function and $d_{m,i}, d_{n,i}$ are the atoms of D_m and D_n. N is the number of samples.

The above Eq.(6) can be alternatively optimized using iterative algorithm. Specifically, fix other variables, update 1) A_m, A_n, 2) D_m, D_n, 3) P in an iterative way. In these processes, we mainly modified the update of coefficients A_m, A_n as below:

$$\begin{aligned} &\min_{A_m} \|\bar{X}_m - \bar{D}_m A_m\|_F^2 + \gamma \sum_{k=1}^N \|l_{m,k} \odot a_{m,k}\|_2^2 \\ &\min_{A_n} \|\bar{X}_n - \bar{D}_n A_n\|_F^2 + \gamma \sum_{k=1}^N \|l_{n,k} \odot a_{n,k}\|_2^2, \end{aligned} \tag{7}$$

where $\bar{X}_m = [X_m; \delta P A_n]$, $\bar{X}_n = [X_n; \delta A_m]$, $\bar{D}_m = [D_m; \delta I]$, $\bar{D}_n = [D_n; \delta P]$ and I is an identity matrix. Take the update of A_m for example. Since there is a combination matrix of X_m and $\delta P A_n$, while the locality should be found between X_m and D_m, we modified the LLC coding procedure as Algorithm 1. To elaborate, we first prune the dictionary by only keeping the set of basis in D_m whose corresponding weights are larger than a predefined constant. The corresponding weights are calculated with locality constraint between X_m and D_m. Then both X_m, D_m and $\delta P A_n, \delta I$ are used to update A_m.

Synthesis with Poisson Editing and Ratio-image Merging

Since preprocessing step will warp the face into a canonical form and only retain the central part of the face, to make the result more realistic, it requires a warping back and a seamless blending procedure to replace the reconstructed part in original image. In this work, a Poisson Image Editing (Pérez, Gangnet, and Blake 2003) method is used to blend the makeup removal face into original image. In addition, this process leads to an advantage that the extent of makeup removal could be adjusted freely through a blending parameter.

One phenomenon has been observed in the experiments is that some individual facial textures like wrinkles will be smooth out in the makeup removal face due to the dictionary reconstruction. Therefore, after the above makeup removal image has obtained, one more technique called ratio-image merging (Liu, Shan, and Zhang 2001) is introduced to solve the problem and make the final results more likely to the original subject. This method was proposed by Liu et.al. to tackle with facial expression mapping problem.

Given one subject's with and without makeup face surfaces **A** and **B**, for any point p on surface **A**, there exists a corresponding point on **B** which has the same meaning. Assume there are m point light sources and each has the light direction from p denoted as $d_i, 1 \le i \le m$, and its intensity denoted as I_i. Suppose the surface is diffuse, according to the Lambertian model, the intensity at point p is:

$$I = \rho \sum_{i=1}^m I_i n \cdot d_i, \tag{8}$$

where, n denotes its normal vector and ρ is the reflectance coefficient at p.

After the surface is deformed, which could be considered as a face with wrinkles,the intensity at p becomes

$$I' = \rho \sum_{i=1}^m I_i n' \cdot d_i', \tag{9}$$

From Eq. (8) and Eq. (9), we have

$$\frac{I_a'}{I_a} = \frac{\sum_{i=1}^m I_i n_a' \cdot d_{ia}'}{\sum_{i=1}^m I_i n_a \cdot d_{ia}}, \quad \frac{I_b'}{I_b} = \frac{\sum_{i=1}^m I_i n_b' \cdot d_{ib}'}{\sum_{i=1}^m I_i n_b \cdot d_{ib}}, \tag{10}$$

for surface **A** and **B** at each point.

In our case, we transfer wrinkles between same subject but with or without makeup. The surface normals at the corresponding positions are about the same, that is, $n_a \approx n_b$ and $n_a' \approx n_b'$. And since the two images are in the same pose, the lighting direction vectors are also the same, that is, $d_{ia} = d_{ib}$ and $d_{ia}' = d_{ib}'$. Under this assumption, we have

$$B'(x, y) = B(x, y) \frac{A'(x, y)}{A(x, y)}, \tag{11}$$

Table 1: Average makeup detection accuracy(%) of different algorithms on YMU, MIW, VMU and SMU datasets, the input is raw pixel. Dictionary initialized with **PCA**.

Conf	YMU	MIW	VMU	SMU		
				Facial	Eye	Mouth
LRC	79.29	87.73	92.73	77.37	71.71	74.12
LDA	89.61	52.74	91.67	50.52	62.54	72.34
SVM	87.08	89.14	91.33	77.34	80.90	79.57
Ours	**91.59**	**91.41**	**93.75**	**82.40**	**84.36**	**86.51**

where (x, y) denote the coordinates of a pixel in the images.

In summary, given a person's makeup image $\mathbf{A}$, we first do smooth filter on some regions without makeup but usually has wrinkles e.g. eyebags, corner of the mouth, forehead to get $\mathbf{A}'$. Once the makeup removal image $\mathbf{B}$ is obtained, the final image with more detailed texture could be set pixel by pixel through Eq. (11).

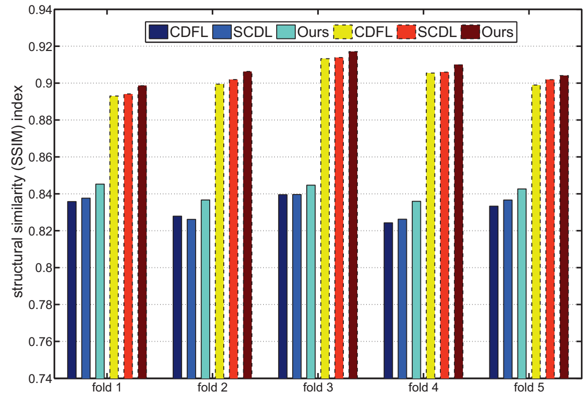

Figure 2: Average SSIM of different methods on VMU (solid line) and SMU (dash line) with 5-fold training test split.

Experiments

In order to evaluate our system's performance on makeup detection and makeup decomposition, we employ three kinds of experiments. First, we conducted makeup recognition experiments to ascertain the effectiveness of our DL method compared with some other classification methods. Next, the performance of LC-CDL on makeup decomposition is demonstrated. This also provides insights into our method through visual examples. Finally, we further illustrate the impact of our makeup decomposition by performing face verification on with or without makeup removal procedure. We utilize the databases[2] introduced by Dantcheva et al. and Chen et al. (Dantcheva, Chen, and Ross 2012; Chen, Dantcheva, and Ross 2013) which are YouTube MakeUp (YMU) database, Virtual MakeUp (VMU) database and Makeup in the wild database (MIW). However, these databases only have before and after makeup images for each subjects. In order to facilitate this study of makeup decomposition, we assembled a face dataset with stepwise makeup labeled for every sub-regions makeup status (SMU)[3].

Makeup detection

A 5-fold cross-validation scheme is employed to evaluate the performance of the proposed makeup detection algorithm. There is no overlap between the training and testing set in terms of subjects. For three existing datasets, we do the makeup detection on the entire face since there are no label information on the region makeup. For our collected dataset, we do makeup detection on sub-regions.

YMU, MIW and VMU Databases. In this section, three existing databases are deployed to evaluate our makeup detector's performance. Here, the YMU dataset is separated into 5 folds, each has 120 images; similarly, the MIW with approximately 30 images in each fold, while the VMU dataset is approximately 40 images in each fold. In Table (1), we show the comparison results of our LC-LRD, LRC (Naseem, Togneri, and Bennamoun 2010), LDA, and SVM on raw pixel data. As presented, our dictionary learning method performs better for all the datasets. Besides, previous work in Chen et.al. extracted lots of features to classifying, that means our method has a lot of room for improvement.

Collected SMU Databases. For our collected SMU dataset, we do the makeup detection on three sub-regions where are most commonly applied cosmetics, 108 images in each fold of eye region, 47 images in each fold of mouth region and 100 images in each fold of skin region. Table (1) shows the detection results of these regions, our method performs best in all cases. After above experiments, we show our system's acceptable ability on detecting the makeup and locating the cosmetic regions.

Makeup decomposition

Learning only one pair of dictionaries and an associated projection matrix is often not enough to cover all variations of makeup decomposition. For example, some faces may apply strong eye makeup while others may only use lipsticks, which varies the mapping function significantly. Therefore multi-modal should be learned to enhance the robustness, that is different pairs of dictionaries should be assigned to different face regions. What't more, due to the various of makeup styles, we also need to introduce locality constraint to explore the dataset's manifold structure. Intuitively, LC-CDL is conducted to train a coupled dictionary with local information in order to learn the linear mapping more stably. Altogether, we first separate whole face into three regions, for each region we run LC-CDL separately.

We consider the most representative methods CDFL (Huang and Wang 2013) and SCDL (Wang et al. 2012) for comparisons. For the above methods, we apply the codes provided in their websites and modify them to our makeup remove scheme. For fair comparisons, the same preprocessing is applied to all of the above methods. 51 pairs of faces in VMU and 97 pairs in SMU were used in this experiment and split into 5 folds. Fig.2 compares the results of different algorithms in terms of structural

[2,3] The detailed information refers to http://antitza.com/makeup-datasets.html, and http://www.northeastern.edu/smilelab/.

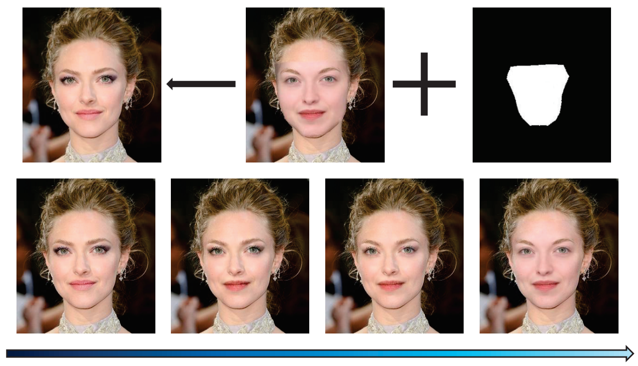

Figure 3: Example of poisson editing. Upper row shows the target, insert and mask images for blending. The second row gives an example of adjusting makeup extent through poisson editing.

similarity index measure (SSIM) on VMU(solid line) and SMU(dash line) datasets. The SSIM (Wang et al. 2004) is a well-known measurement of the similarity between two images which considered to be associated with the human visual perception. It can be observe that our approach achieves the highest SSIM values for both two datasets in all folds.

For the poisson editing part, we need a mask image to assign the blending part of insert image. Since we already have fiducial landmarks for each image, this mask image could be automatically generated using the landmarks constrain. As we mentioned before, using poisson editing could even let us modify the makeup extent easily by choosing different value of blending parameter. One example of changing makeup extent is presented in Fig.3.

At the last step of makeup removal, we introduce the ratio-image merging method to produce more realistic results by adding wrinkle to non-makeup face. The top-left sample in Fig.4 illustrates the wrinkles copied by ratio-image merging. We can see that the wrinkles around the eyes and mouth have been exactly copied to the output makeup removal image, which makes the results more likely to the original person. And the rest of the Fig.4 shows more results for makeup remove both on VMU and SMU datasets. Besides, we also use SMU as training set and test our system on web makeup image of celebrities (Fig.1).

Face Verification with Makeup Removal

In this section, we discuss the use of the proposed makeup detection and removing system in the context of face verification. In (Eckert, Kose, and Dugelay 2013), it is shown that the recognition rate of face verification drops when the makeup faces are matched with the corresponding non-makeup images. To address this problem, we exploit the makeup decomposition as a pre-processing routine to suppress the effect of makeup by using a synthesised non-makeup image to help.

Figure 4: Ratio-image mergion (top-left) and more examples of makeup remove results on VMU (left) and SMU (right). For these seven examples, left column is input makeup image, right column is output makeup removal image, middle column is ground truth non-makeup image.

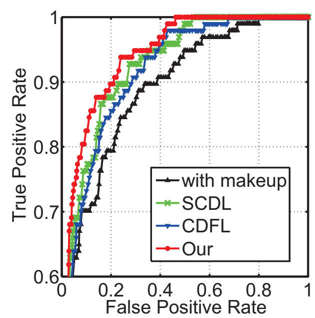

Figure 5: ROC curve.

M vs **N**	SVM(%)
Before	79.18
CDFL	83.81
SCDL	84.85
Our	**87.73**

Table 2: Verification rates of with or without our makeup remove pre-processing on SMU dataset.

The performance of the makeup detector is reported using Receiver Operating Characteristic (ROC) curve (Fig.5), here, the true positive rate is plotted as a function of the false positive rate. Face verification performances before and after applying the proposed face pre-processing scheme. We do experiments on SMU datasets, where we have ten-fold subsets each has 97 positive pairs and 97 negative pairs. The mean verification rates with SVM are reported in Table (2). We can see all three makeup removed scheme reach at least 5% higher than baseline, and our method produces the highest results. Note that this is a simple test of our method's application on face verification. We neither use any low-level features nor intend to outperform other state-of-art methods. Actually our method could work as a preprocessing of any advanced algorithm to enhance the performance.

Conclusion

In this work, we proposed a novel automatic makeup detector and remover system. For detector, we used a locality-constrained dictionary learning to determine the presence of makeup. For the makeup remover, we proposed a novel

LC-CDL framework for non-makeup image synthesis. The learned dictionary pair ensures the status-specific data fidelity, meanwhile uncover the latent spaces which generate stable mapping between status. We also collected a stepwise makeup dataset which, to the best of our knowledge is the first dataset with steps of sub-region makeup. The proposed LC-CDL showed very competitive performance on makeup removal. The output of the makeup detector and remover was then used to perform adaptive pre-processing in the context of face verification between makeup images and their non-makeup counterparts. The recognition results indicated an improvement by applying the proposed pre-processing routine. In the future study, we will collect more diverse data and extend this system to different ethnic groups.

References

Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM (JACM)* 58(3):11.

Chen, C.; Dantcheva, A.; and Ross, A. 2013. Automatic facial makeup detection with application in face recognition. In *Proc. of International Conference on Biometrics (ICB)*, 1–8.

Dantcheva, A.; Chen, C.; and Ross, A. 2012. Can facial cosmetics affect the matching accuracy of face recognition systems? In *Proc. of International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 391–398.

Ding, Z.; Ming, S.; and Fu, Y. 2014. Latent low-rank transfer subspace learning for missing modality recognition. In *AAAI*.

Donato, G., and Belongie, S. 2002. Approximation methods for thin plate spline mappings and principal warps. In *Transformation of Datasets in Linear-based Map Conflation Framework. Surveying and Land Information Systems*. Citeseer.

Du, H., and Shu, L. 2013. Makeup transfer using multi-example. In *Proc. of International Conference on Information Technology and Software Engineering*, 467–474. Springer.

Eckert, M.-L.; Kose, N.; and Dugelay, J.-L. 2013. Facial cosmetics database and impact analysis on automatic face recognition. In *Proc. of International Workshop on Multimedia Signal Processing (MMSP)*, 434–439.

Guo, D., and Sim, T. 2009. Digital face makeup by example. In *CVPR*, 73–79.

Guo, G.; Wen, L.; and Yan, S. 2014. Face authentication with makeup changes. *Circuits and Systems for Video Technology, IEEE Transactions on* 24(5).

Huang, D.-A., and Wang, Y.-C. F. 2013. Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. In *ICCV*, 2496–2503.

Inc., M. 2013. Face++ research toolkit. www.faceplusplus.com.

Li, L.; Li, S.; and Fu, Y. 2014. Learning low-rank and discriminative dictionary for image classification. *Image and Vision Computing*.

Liu, L.; Xu, H.; Xing, J.; Liu, S.; Zhou, X.; and Yan, S. 2013. Wow! you are so beautiful today! In *Proc. of ACM MM*, 3–12.

Liu, G.; Lin, Z.; and Yu, Y. 2010. Robust subspace segmentation by low-rank representation. In *Proc. of ICML*, 663–670.

Liu, Z.; Shan, Y.; and Zhang, Z. 2001. Expressive expression mapping with ratio images. In *Proc. of the 28th annual conference on Computer graphics and interactive techniques*, 271–276.

Ma, L.; Wang, C.; Xiao, B.; and Zhou, W. 2012. Sparse representation for face recognition based on discriminative low-rank dictionary learning. In *CVPR*, 2586–2593.

Naseem, I.; Togneri, R.; and Bennamoun, M. 2010. Linear regression for face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32(11):2106–2112.

Pérez, P.; Gangnet, M.; and Blake, A. 2003. Poisson image editing. In *ACM Transactions on Graphics (TOG)*, volume 22, 313–318.

Walha, R.; Drira, F.; Lebourgeois, F.; Garcia, C.; and Alimi, A. M. 2014. Sparse coding with a coupled dictionary learning approach for textual image super-resolution. In *Proc. of ICPR*, 4459–4464.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on* 13(4):600–612.

Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; and Gong, Y. 2010. Locality-constrained linear coding for image classification. In *CVPR*, 3360–3367.

Wang, S.; Zhang, D.; Liang, Y.; and Pan, Q. 2012. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *CVPR*, 2216–2223.

Wang, S.; Shao, M.; and Fu, Y. 2014. Attractive or not?: Beauty prediction with attractiveness-aware encoders and robust late fusion. In *Proc. of ACM MM*, 805–808.

Yang, M.; Zhang, D.; and Feng, X. 2011. Fisher discrimination dictionary learning for sparse representation. In *ICCV*, 543–550.

Yu, K.; Zhang, T.; and Gong, Y. 2009. Nonlinear learning using local coordinate coding. In *Advances in neural information processing systems*, 2223–2231.

Zhang, S.; Gao, X.; Wang, N.; Li, J.; and Zhang, M. 2015. Face sketch synthesis via sparse representation-based greedy search. *Image Processing, IEEE Transactions on*.

Zhang, G.; Jiang, Z.; and Davis, L. S. 2013. Online semi-supervised discriminative dictionary learning for sparse representation. In *Computer Vision–ACCV 2012*. Springer. 259–273.

Zhao, H., and Fu, Y. 2015. Semantic single video segmentation with robust graph representation. In *Proc. of IJCAI*, 2219–2225.