15-29

15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Skin Beautification detection using Sparse Coding

Tianyang Sun
Harbin Institute of Technology
2 Wenhuaxi Road, Weihai, 264209, China
tianyang_sun@foxmail.com

Xinyu Hui
Harbin Institute of Technology
2 Wenhuaxi Road, Weihai, 264209, China
xiinyu_hui@foxmail.com

Zihao Wang
Harbin Institute of Technology
2 Wenhuaxi Road, Weihai, 264209, China
wangatestech@gmail.com

Shengping Zhang
Harbin Institute of Technology
2 Wenhuaxi Road, Weihai, 264209, China
s.zhang@hit.edu.cn

Abstract

In the past years, skin beautifying softwares have been widely used in portable devices for social activities, which have the functionalities of turning one's skin into flawless complexion. With a huge number of photos uploaded to social media, it is useful for users to distinguish whether a photo is beautified or not. To address this problem, in this paper, we propose a skin beautification detection method by mining and distinguishing the intrinsic features of original photos and the corresponding beautified photos. To this aim, we propose to use sparse coding to learn two sets of basis functions using densely sampled patches from the original photos and the beautified photos, respectively. To detect whether a test photo is beautified, we represent the sampled patches from the photo using the learned basis functions and then see which set of basis functions produces more sparse coefficients. To our knowledge, our effort is the first one to detect skin beautification. To validate the effectiveness of the proposed method, we collected about 1000 photos including both the original photos and the photos beautified by a software. Our experimental results indicate the proposed method achieved a desired detection accuracy of over 80%.

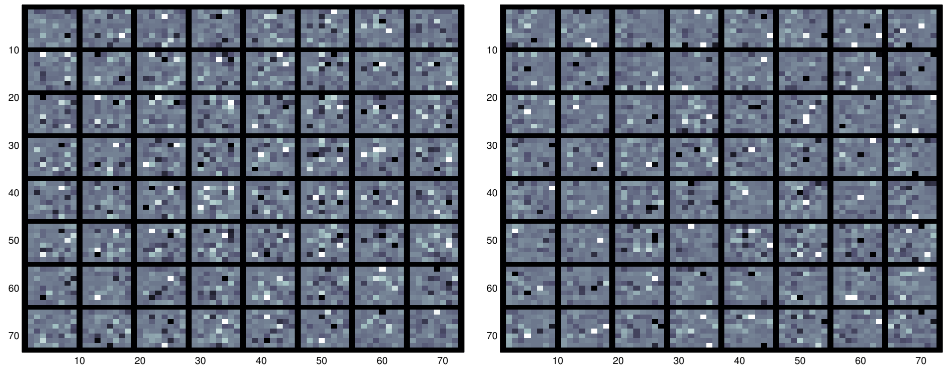
(a) Original photo (b) Beautified photo

Figure 1. Examples of an original photo and the corresponding beautified photo.

1 Introduction

In recent years, skin beautifying softwares have been widely used in portable devices for social activities [3, 4, 5], which have the functionalities of turning one's skin into flawless complexion. For example, users can choose to whiten their skin, widen their eyes, and remove the flaw on their faces simply by touching the screen on their portable devices. There is an application called "Primo" developed by a Japanese undergraduate, which is one of the most well-known photo beautifying restoration softwares, while it simply doctors the prints to uglify the portraits without the detection of skin-beautifying. When a facial image is uploaded to this software, "Primo" will simply greaten your face and smallen your eyes, no matter the facial image was beautified or not. The effectiveness of skin-beautifying detection is far from satisfaction.

A large number of uploaded photos to social software makes it critical for users to distinguish whether a photo is beautified or not. For example, a beautified image posted on a website have attracted millions of page views and the person in the image became a network star overnight. However, the posted image was so distinct from the original image that the crowd believed that they were cheated after the original image was concealed. Therefore, it is very necessary to design a method to detect whether a photo is an original photos without any beautification or a beautified photo. However, achieving this purpose is not very easy because it is very difficult to extract useful features to distinguish the original photos and the beautified photos given the fact that the photos are recorded in very challenging environments including illumination changes, pose variations and noises. In particular, when the photos are recorded by mobile phones, the quality of the recorded photos may be worse than the photos recorded by professional cameras. In this paper, inspired by the successful applications of sparse coding in different vision tasks [6, 7], we proposed a novel photo beautification detection method by mining the intrinsic features of original photos and beautified photos using sparse coding. The proposed method includes two stages. In the training stage, training photos are divided into overlapped patches. The resulting patches from the original photos and the patches from the beautified photos are used as input to learn two dictionaries. The first dictionary is capable of representing the intrinsic features of the original photos. The second dictionary is capable of representing the intrinsic features of the beautified photos. In the testing phase, a test photo is also divided into overlapping patches using the same procedure as like in the training stage. The resulting patches are encoded using the two dictionaries, respectively. There is a reasonable assumption that the representation coefficients should be more sparse if the used dictionary is learned from the same class as the test photo. Therefore, we



(a) $\boldsymbol{D}_1$ (b) $\boldsymbol{D}_2$

Figure 2. Visualization of the learned dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$.

can compare how sparse the coefficients are when two dictionaries are used to represent the test photo and then make a decision based on this comparison. To compare the sparseness of representation coefficients, we adopt the kurtosis measurement to calculate the sparseness of each basis function and then calculate the average kurtosis over the whole dictionary. The final decision is made based on the comparison between the average kurtosis of two dictionaries. To validate the effectiveness of the proposed method, we collect about 1000 photos including both the original photos and the beautified photos. The beautified photos were processed from the original photos using some professional skin beautification softwares. The number of beautified images is larger than the one of the original images because more than one kind of beauty effects can be added to each original image. Experimental results have shown that our skin beautification detection algorithm achieves a desired accuracy of over 80%.

2 Background of Sparse coding

The essence of sparse coding is using a few basis vectors to effectively and neatly represent a signal. Let $\boldsymbol{x} \in \mathbb{R}^m$ be the signal to be encoded, the sparse vector is $\boldsymbol{a} \in \mathbb{R}^n$ $(m << n)$, the relation of the input signal and the sparse vector is

$$\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{a} \tag{1}$$

where $\boldsymbol{D} \in \mathbb{R}^{m \times n}$ is the dictionary matrix, whose column vectors $\boldsymbol{d}_i, i = 1, 2, \ldots, n$ are basis functions. The sparse vector $\boldsymbol{a}$ is also called sparse coefficients. The reason why we use sparse coding to mine the intrinsic features of face photos is that most sensory-collected data, for example, the face images, can be expressed as the combination of few basis functions.

There are two issues in sparse coding. The first one is to compute the sparse coefficients given a signal $\boldsymbol{x}$ and a dictionary $\boldsymbol{D}$, which can be obtained by optimizing the following objective

$$\boldsymbol{a} = \arg\min \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{a}\|_2 + \lambda\|\boldsymbol{a}\|_1 \tag{2}$$

which can be easily solved by some existing methods [1, 2]. The second issue is how to learn the dictionary $\boldsymbol{D}$ using training samples $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L$, which can also be expressed by an optimisation problem

$$\{\boldsymbol{D}, \{\boldsymbol{a}_l\}_{l=1}^L\} = \arg\min_{\boldsymbol{D}, \{\boldsymbol{a}_l\}_{l=1}^L} \sum_{i=1}^{L} \|\boldsymbol{x}_l - \boldsymbol{D}\boldsymbol{a}_l\|_2 + \lambda\|\boldsymbol{a}_l\|_1 \tag{3}$$

which can be solved by iteratively optimizing the coefficient vectors $\{\boldsymbol{a}_l\}_{l=1}^L$ given the fixed dictionary $\boldsymbol{D}$ and optimizing the dictionary $\boldsymbol{D}$ given the fixed coefficient vectors $\{\boldsymbol{a}_l\}_{l=1}^L$. In our experiments, we use the publicly available SPAMS toolbox [1] to solve these two problems.

3 Proposed Method

To distinguish the beautified skin photos and non-beautified photos, the proposed method learns two dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ from the beautified skin photos and non-beautified photos, respectively, at the training stage, which are used to mine the intrinsic features of the beautified skin photos and non-beautified photos. At the testing stage, a test skin photo is encoded using the learned dictionaries, respectively and a decision is made based on which dictionary is capable of producing more sparse representation coefficients. In this section, we present the technique details of both the training and testing stages.

3.1 Learning Sparse Coding Dictionaries

In the training stage, a number of non-beautified original photos $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$ and the beautified photos $\mathcal{I}' = \{I'_1, I'_2, \ldots, I'_M\}$ processed by some professional softwares are used as training samples. It should be noted that the number of beautified photos M may be larger than the number of original photos N because an original photo may produce several beautified photos by using different softwares or using the same software but with different settings. From each training image, a set of 2049 8×8 patches are densely sampled, which are allowed to have partial overlapping. Let $\mathcal{P} = \{P_1, P_2, \ldots, P_T\}$ and $\mathcal{P}' = \{P'_1, P'_2, \ldots, P'_S\}$ are the densely sampled patches from the original photos and the beautified photos, respectively. Two dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ can be learned from $\mathcal{P}$ and $\mathcal{P}'$, respectively, by solving the optimisation problem Eq. 3. The learned dictionaries are shown in Fig. 2.

[1] http://spams-devel.gforge.inria.fr

Figure 3. Examples of the collected original photos (the first row) and the corresponding beautified photos (the second row).

3.2 Beautification Detection

In the testing stage, a set of patches are also densely sampled from a test photo using the same procedure as like in the training stage. Let $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_Q\}$ be the sampled patches. Each patch is sparsely encoded by the dictionaries learned at the training stage. Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_Q$ and $\boldsymbol{a}'_1, \boldsymbol{a}'_2, \dots, \boldsymbol{a}'_Q$ be the sparse coefficients of representing the patches $\mathcal{Y}$ using the dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, respectively. Let $\boldsymbol{A}$ and $\boldsymbol{A}'$ be a matrix with $\boldsymbol{a}'_1, \boldsymbol{a}'_2, \dots, \boldsymbol{a}'_Q$ and $\boldsymbol{a}'_1, \boldsymbol{a}'_2, \dots, \boldsymbol{a}'_Q$ as their columns, respectively. The coefficients in the k-th row are associated with k-th basis function of the dictionary $\boldsymbol{D}_1$. If we consider the k-th basis function as a neuron, all these coefficients can be considered as the neuron's responses. As inspired by the theory of sparse coding, the neuron's responses should be sparse. Therefore, the k-th row of $\boldsymbol{A}$ should have a large kurtosis. The similar observation can be obtained from the coefficients in the k-th row of $\boldsymbol{A}$. On the other hand, if the test photo is an original photo without any beautification operators, the sampled patches from this test photo should be represented using the dictionary $\boldsymbol{D}_1$ with more sparse coefficients than using the dictionary $\boldsymbol{D}_2$. Based on these observations, we can design an approach to beautification detection by analysing the sparseness of the coefficients using the dictionary $\boldsymbol{D}_1$ or $\boldsymbol{D}_2$.

In particular, we first calculate the kurtosis f_k and f'_k of the k-th row of $\boldsymbol{A}$ and $\boldsymbol{A}'$, respectively, using the following equations

$$f_k = \frac{\frac{1}{Q}\sum_{i=1}^{Q}(A_{k,i} - \mu_k)^4}{\left(\frac{1}{Q}\sum_{i=1}^{Q}(A_{k,i} - \mu_k)^2\right)^2} \tag{4}$$

$$f'_k = \frac{\frac{1}{Q}\sum_{i=1}^{Q}(A_{k,i} - \mu'_k)^4}{\left(\frac{1}{Q}\sum_{i=1}^{Q}(A_{k,i} - \mu'_k)^2\right)^2} \tag{5}$$

where $\mu_k = \frac{1}{Q}\sum_{i=1}^{Q} A_{k,i}$ and $\mu'_k = \frac{1}{Q}\sum_{i=1}^{Q} A'_{k,i}$. Then we can calculate the average kurtosis as

$$\bar{f}_k = \frac{1}{K}\sum_{k=1}^{K} f_k \tag{6}$$

$$\bar{f}'_k = \frac{1}{K}\sum_{k=1}^{K} f'_k \tag{7}$$

The test photo is classified as the original photo without beautification if $\bar{f}_k \le \bar{f}'_k$. Otherwise, it is classified as the beautified photo if $\bar{f}_k < \bar{f}'_k$. The summarised steps of the proposed method is shown in Algorithm 1.

Algorithm 1: The proposed beautification detection method

Input: The non-beautified original photos $\mathcal{I} = \{I_1, I_2, \dots, I_N\}$, the beautified photos $\mathcal{I}' = \{I'_1, I'_2, \dots, I'_M\}$ and the test photo $\mathcal{I}$
Output: Yes or No

Densely sample patches $\mathcal{P}$ and $\mathcal{P}'$ from $\mathcal{I}$ and $\mathcal{I}'$, respectively;
Learn dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ from $\mathcal{P}$ and $\mathcal{P}'$, respectively;
Densely sample patches $\mathcal{Y}$ from the test photo $\mathcal{I}$;
Calculate the coefficient matrices $\boldsymbol{A}$ and $\boldsymbol{A}'$ for patches $\mathcal{Y}$ using dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, respectively;
Initialize $\vec{a}_0 = \vec{a}_{-1} = \vec{0} \in \mathbb{R}^N$ and $q_0 = q_{-1} = 1$;
Calculate the average kurtosis $\bar{f}_k$ and $\bar{f}'_k$ from $\boldsymbol{A}$ and $\boldsymbol{A}'$, respectively;
Return No if $\bar{f}_k \le \bar{f}'_k$. Otherwise, return Yes.

4 Experiments and Analysis

4.1 Experiment Settings

To our best knowledge, there are no a public datasets for evaluating beautification detection methods. To validate the performance of the proposed method, we first collect a dataset, which consists of 1,001 photos of 19 subjects. In particular, the database contains 335 non-beautified high-resolution photos (4608 x 3456), which are captured far away 20–30 cm from the subject using a HUAWEI NXT-AL10 mobile phone with different illumination and pose variations. In order to produce the beautified images, two kinds of beauty effects including skin-smoothing and skin-whitening are manually added (separately 50% and 100%) to the collected images. By this way we have four versions of the beautified photo images (666 in amount). Several examples of the collected original photos as well as their beautified photos are shown in Fig. 3.

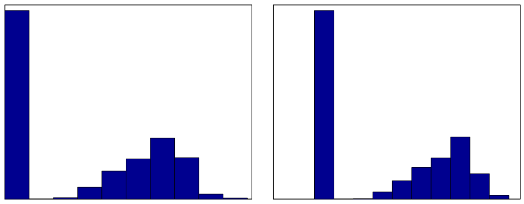
We fixed the positions of a certain area of face (mainly the forehead and cheek) of a person and named it as a namesake of the portrait image. The data of the area contains the top left and bottom left coordinates as well as the length and height of the rectangle are annotated. 167 non-beautified images and 348 beautified images were used to train the model and the rest are used for testing.

From each photo image, a total of 1024 patches with size 8×8 are randomly selected on each image with partial overlapping. The size of the dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are set to 64. The parameter λ is set to 0.01. The detection accuracy, calculated as the rate of the successfully detected photos over all test photos, is used to evaluate the performance of our method.

4.2 Results and analysis

To our best knowledge, there is no previous work on beautification detection in the literature. So in this section, we only report the results of the proposed method without comparing them to other methods. The individual accuracy of the non-beautified and beautified photos is 78.5% and 86.6%, respectively. The overall accuracy of all test photos is 81.1%. The accuracy of beautified images is a bit higher than the one of non-beautified images.

The main rational behind the proposed method is the assumption that the patches from a test photo can be represented with more sparse coefficients using the dictionary learned from training photos with the same class as the test photos than using the dictionary learned from training photos from the other class. To validate whether this assumption is reasonable, we show the histogram of the spares coefficients of representing an original photo using the dictionary learned from the original photos and the dictionary learned from the beautified photos. The resulting two histograms are shown in Fig. 4. As we can see, the coefficients shown in Fig. 4(a) is more sparse than the coefficients shown in Fig. 4(b).



(a) Original photo (b) Beautified photo

Figure 4. The histogram of the spares coefficients of representing an original photo using the dictionary learned from the original photos and the dictionary learned from the beautified photos.

5 Conclusion

In this paper, we propose a skin beautification detection method by mining and distinguishing the intrinsic features of original photos and the corresponding beautified photos. To this aim, we propose to use sparse coding to learn two sets of basis functions using densely sampled patches from the original photos and the beautified photos, respectively. To detect whether a test photo is beautified, we represent the sampled patches from the photo using the learned basis functions and then see which set of basis functions produces more sparse coefficients. To our knowledge, our effort is the first one to detect skin beautification. To validate the effectiveness of the proposed method, we collected about 1000 photos including both the original photos and the photos beautified by a software. Our experimental results indicate the proposed method achieved a desired detection accuracy of over 80%.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61672188 and the First Year Undergrad Student Annual Creative Project of HIT WH.

References

[1] J. Mairal, F. Bach, J. Ponce and G. Sapiro: "Online Dictionary Learning for Sparse Coding," *International Conference on Machine Learning*, Montreal, Canada, 2009.

[2] J. Mairal: "Optimization with First-Order Surrogate Functions," *International Conference on Machine Learning*, 2013.

[3] Lingyu Liang, Lianwen Jin, Xuelong Li: "Facial Skin Beautification Using Adaptive Region-Aware Masks',' *IEEE Trans. Cybernetics*, 44(12): 2600-2612, 2014.

[4] Lingyu Liang, Deng Liu, Lianwen Jin: "FaceMore: A Face Beautification Platform on the Cloud," *IEEE International Conference on Systems, Man, and Cybernetics*, 1798-1803, 2015.

[5] Xi Lu, Xiaobin Chang, Xiaohua Xie, Jian-Fang Hu, Wei-Shi Zheng: "Facial skin beautification via sparse representation over learned layer dictionary," *Proceedings of International Joint Conference on Neural Networks* , 2015.

[6] Peter Wilf, Shengping Zhang, Sharat Chikkerur, Stefan A. Little, Scott L. Wing, Thomas Serre: "Computer Vision Cracks the Leaf Code," *Proceedings of the National Academy of Sciences of the United States of America*, 113(12): 3305-3310, 2016.

[7] Shengping Zhang, Huiyu Zhou, Feng Jiang, Xuelong Li: "Robust visual tracking using structurally random projection and weighted least squares," *IEEE Transactions on Circuits and Systems for Video Technology*, 25(11): 1749–1760, 2015