A LANDMARK-BASED APPROACH TO FACE HALLUCINATION

Atharva Gokhale, Somesh Daga, Somesh Gupta, and Xiaomin Yan

University of Waterloo

ABSTRACT

Face Hallucination is an area of significant academic and industry interest due to its anticipated benefits in the analysis of low-resolution (LR) face images, particularly for Low-Resolution Face Recognition (LRFR). In this paper, we present a novel learning-based face hallucination method utilizing face landmarks and demonstrate its performance on face recognition for low-resolution face images obtained from two public datasets. Additionally, we provide an analysis of face recognition confidence against traditional face hallucination quality measures of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

1. INTRODUCTION

Face Hallucination, also known as Face Super-Resolution (FSR), refers to the creation of high-resolution (HR) face images from their low-resolution counterparts. The consensus for the size of low-resolution images in this field, are those less than or equal to 32x32 pixels. LRFR serves as the most prominent application for face hallucination, whereby hallucinated face images are compared to their high-resolution face images through face recognition measures.

State-of-the-art in the field of face hallucination is largely dominated by deep-learning methods, such as [1],[2]. However, classical learning-based methods typically offer advantages of reduced training time and data requirements, as well as lower computational complexity. In turn, they are more susceptible to variations in face alignment and extrinsic variables such as lighting conditions.

In this paper, we focus on a novel, classical learning-based face hallucination method, and demonstrate its effectiveness on LRFR. In this section, we provide an overview of related works, the motivations surrounding the development of our novel method and the contributions to the field of face hallucination resulting from our work. Section 2 details the formulation of our method, and Section 3 demonstrates its performance in LRFR and studies the factors contributing to the same. Lastly, we conclude this paper by summarizing our contributions and proposing directions for future research.

1.1. Related Works

Through their landmark paper, Wang et al. [3] were the first to apply Principal Component Analysis (PCA) to the hallucination of low-resolution face images, and demonstrated their results on face recognition tasks. This inspired a number of future PCA-based methods operating on local structures of images using patched-based schemes such as [4], [5], [6].

Shi et al. [7] investigated the use of similarity measures based on face geometry and landmarks for face recognition, and formulated the "Refined Procrustes Distance" as the leading similarity measure for face recognition.

More recently, the use of face landmark priors have been adopted in deep-learning methods ([1],[2]) yielding state-of-the-art results in the face hallucination space.

1.2. Motivations and Contributions

The inspiration for a landmark-based approach to face hallucination, in the realm of classical learning-based methods, comes from the increasing importance attributed to the role of face landmarks in face recognition ([8]) and face super-resolution ([1],[2]). Having chosen to focus on improving LRFR performance, it seems only natural to utilize features for the face hallucination process that lend well to face recognition tasks.

Given that face landmarks operate on a global context of a face image, we have chosen to the work of Wang et al. [3] as the baseline for comparison, which similarly super-resolves face images using a global scheme. Succeeding works of [3] such as [4] and [6] have yielded state-of-the-art results for classical learning-based methods, however, utilizing an exclusively local, patched-based approach to face hallucination.

Recent super-resolution methods ([1],[2]) have called into question the use of classical image quality metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as a measure of face super-resolution quality. While we acknowledge the inconclusive analysis of [9] which investigated the effect of PSNR against face recognition, we provide an alternative analysis with both SSIM and PSNR measures.

In summary, our contributions are as follows:

1. We present a novel, landmark-based face hallucination method that is, to the best of our knowledge, the first to utilize face landmarks for face hallucination in a traditional learning-based scheme
2. We compare the face recognition performance of our method against [3] on two public datasets, and conduct an analysis of the impact of PSNR and SSIM metrics on face recognition confidence

2. METHOD

2.1. Landmark-Based Face Hallucination

We build on the face hallucination approach of Wang et. al. [3], by incorporating face landmarks into the hallucination process, with the aim of improving face recognition performance for LR images.

First, we construct a low-resolution face image that is to be hallucinated, by down-sampling the corresponding high-resolution image. We compile a set of HR images to serve as a training dataset. Next, we proceed to locate face landmarks for the test image. In order to do so, the test image is bicubically interpolated to the dimensions of the HR image, as face landmark detection is typically not feasible on images with dimensions less than 100x100 pixels. We use an online platform, Face++ [10], to obtain face landmarks like mouth, nose, eyes and face contour in terms of x-y co-ordinates. We have used 18 landmarks for mouth, 10 for nose, 10 per eye, 8 per eyebrow, and 19 for the face contour. Similarly, we obtain the same face landmarks for all HR images in the training set.

Next, we find those HR images in the training set which closely resemble the input LR image in terms of face landmarks. The origin of the landmarks for the training images is translated to the nose tip to construct the "pre-shaped space", in order to use the "Refined Procrustes Distance" [7] (refer to section 2.2). Employing this measure, for each of the training images, we obtain a N by M matrix, $[\vec{g_1}, \vec{g_2}, ..., \vec{g_M}]$ that contains the distance of all the landmarks from the nose tip, where N is number of landmarks per image and M is number of training images.The mean landmark distance vector across the training images is then computed and given by

$$\vec{m} = \frac{1}{M} \sum_{i=1}^{M} \vec{g_i}$$

We then carry out a PCA analysis of the face landmarks as defined by the "Refined Procrustes Distance" [7] measure. A covariance matrix is computed and given as follows

$$C = \sum_{i=1}^{M} (\vec{g_i} - \vec{m})(\vec{g_i} - \vec{m})^T = GG^T$$

The covariance matrix is then decomposed into its eigen-values and eigen-vectors which represent the principal components of the landmark feature space. It is sufficient to retain only a small subset of these eigen-vectors with the highest eigen-values i.e. the k dominant eigen-vectors, as noted by [11].

The eigen-vectors are then normalized by dividing them by the square-root of their corresponding eigen-values, yielding the eigen-value weighted eigen-vector matrix defined as

$$C' = \left[\frac{\vec{c_1}}{\lambda_1}, \frac{\vec{c_2}}{\lambda_2}, ..., \frac{\vec{c_k}}{\lambda_k} \right]$$

where each $\vec{c_i}$ is an eigen-vector and λ_i its associated eigen-value.

The mean-adjusted landmark distance vectors for the test image, computed in the same way, and the training images are then projected onto the feature space represented by C' yielding a vector of k weights for each image. Finally, we obtain a reduced set of K training images by choosing the ones whose weights are closest to that of the training image, in a mean-squared sense.

We restrict the size of our reduced training set to 50 images ($K = 50$) and only retain the 5 ($k = 5$) most dominant eigen-vectors in the landmark feature space represented by C'. With this reduced training set, we employ the method of Wang et al. [3] to super-resolve the test image. Notably, we deviate from the method of [3] by utilizing only the bicubic interpolated test image and the high-resolution training images in the PCA analysis.

2.2. Refined Procrustes Distance

The "Refined Procrustes Distance" is a similarity measure for landmark-based face models developed by [7]. It is a modified form of the classical Procrustes distance that operates in a reduced feature space obtained through principal component analysis. We utilize this measure in our work as mentioned above, to constrain the training set of images to those most similar to the test image with regards to face landmarks. We introduce a minor variation to this measure, where we utilize the absolute distance in the "pre-shaped space" [7] as opposed to using complex relative positions in the computation, as it was observed to yield better face matches.

3. EXPERIMENTS

This section provides an overview of the datasets used for our analysis and details our experiments and results.

We engage the use of a commercial face recognition software, Face++ [10], for generating face recognition confidence (FRC) percentages between hallucinated and HR images.

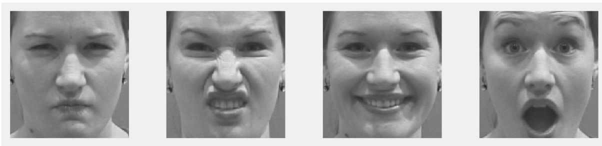

Fig. 1. Sample images from the CAFE [12] dataset

3.1. Datasets

We make use of two public datasets for our analysis and evaluations, the *California Facial Expressions database (CAFE)* [12] and the *FEI Face database* [13].

The CAFE [12] dataset contains a total of 139 images of 24 individuals with a diverse variety of facial expressions. The rationale for choosing this dataset is that it presents a more challenging set of training images due to the variations in face landmark positions, for the same individual, as a result of the different expressions. In line with the procedure adopted by [3], we normalize the images to a center eye distance of 50 pixels, and crop these images to a size of 125x117 to form the high-resolution training images, and further downsample them to 25x23 pixels for creating the low-resolution test images. An example of some of the high-resolution images can be seen in figure 1.

We have chosen a smaller segment of the spatially normalized faces from the FEI Face [13] database of 200 images, out of the total of 400 images. These 200 test images are comprised of two faces for 100 individuals, where one face contains a neutral expression while the other exhibits a happy expression. This dataset has been used by numerous traditional learning-based face hallucination methods such as [4],[6] and hence we have included it in our analysis. High-resolution training images are represented by images of size 120x100 pixels while their low-resolution test counterparts are downsampled to a size of 30x25 pixels to adhere to the accepted definition of a low-resolution image in the face super-resolution field.

3.2. Face Recognition using Landmark-Based Face Hallucination

In this experiment, we resolve test images to their super-resolved versions using the approach of Wang et al. [3] and our landmark-based method. We utilize a "Leave-One-Out" scheme, whereby training images are comprised of all high-resolution images in a dataset other than the one being tested upon. This is consistent with the approach taken by many face super-resolution methods including [3].

As can be inferred from figure 2, finer details are reconstructed in our method as compared to the method of Wang et al. [3]. We alert the reader to the fact that we

Method	Eigenfaces [14]	PSNR (dB)	SSIM	FRC (%)
Bicubic	-	26.38	0.746	-
Wang et al. [3]	50	25.12	0.742	75.8
Wang et al. [3]	100	25.34	0.747	79.1
Ours	50	26.38	0.721	83.7

Table 1. CAFE [12] Dataset Face Hallucination Results: Averaged PSNR, SSIM and Face Recognition Confidence of hallucinated images with respect to their high-resolution images

Method	Eigenfaces [14]	PSNR (dB)	SSIM	FRC (%)
Bicubic	-	25.82	0.782	-
Wang et al. [3]	50	23.05	0.736	68.5
Wang et al. [3]	100	23.55	0.757	75.9
Ours	50	25.19	0.743	80.5

Table 2. FEI Face [13] Dataset Face Hallucination Results: Averaged PSNR, SSIM and Face Recognition Confidence of hallucinated images with respect to their high-resolution images

utilized 100 "eigenfaces" [14] in evaluating results on the method for [3] while only utilizing 50 "eigenfaces" for our method, an upper bound imposed by the reduced number of training images derived from the "Refined Procrustes Distance" [7] similarity measure. Utilization of a higher number of "eigenfaces" [14] typically corresponds to the restoration of higher-frequency details, as noted by [3]. However, we are able to obtain finer details despite the lower number of "eigenfaces" [14] by constraining the learning process to faces that are most similar, and thereby encapsulating low and high frequency details in a reduced feature space.

Averages of PSNR, SSIM and face recognition confidence obtained for the method of Wang et al. [3] and our method, on the CAFE [12] and FEI Face [13] datasets are presented in tables 1 and 2 respectively. For both the CAFE [12] and FEI Face [13] datasets, it is seen that our method registers higher averages for face recognition confidence compared to [3], in cases where the number of "eigenfaces" [14] utilized are the same or twice of what our method uses. Moreover, we observe a marked improvement in the average PSNR values for our method across both datasets. While we acknowledge the marginally lower structural similarity indexes of our method in some cases, an analysis of the relation between PSNR, SSIM and FRC measures presented in section 3.3 suggests that the marked improvement in PSNR of our method offsets the small differences in SSIM.

Fig. 2. Face Hallucination Results: The first two rows depict the results on the CAFE [12] and FEI Face [13] datasets respectively, with improved recognition confidences for our landmark-based method, while the bottom two rows exhibit cases of reduced recognition confidence. Images: (a) HR (b) Bicubic (c) Wang's [3] (d) Our's

3.3. Analysis of Face Hallucination Metrics

We reserve this section to provide an analysis of the relations between the measures of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) and Face Recognition Confidence (FRC).

To analyze these measures, we propose computing differences in PSNR, SSIM and FRC quantities that are based on the same image, to compute the $\Delta PSNR$, $\Delta SSIM$ and ΔFRC values for individual images, which we refer to as the delta quantities. In the context of our work, we obtain two sets of PSNR, SSIM and FRC values for each image, from the super-resolved images obtained via the method of [3] and our landmark-based method. Hence, we define the delta quantities as follows:

$$\Delta PSNR = PSNR_{Ours} - PSNR_{[3]}$$
$$\Delta SSIM = SSIM_{Ours} - SSIM_{[3]}$$
$$\Delta FRC = FRC_{Ours} - FRC_{[3]}$$

The rationale for computing delta quantities and finding the correlation between them, rather than using the PSNR, SSIM and FRC quantities outright is due to the fact that the individualistic characteristics of face images e.g. a very distinct beard, may make a face easier to recognize despite having inferior PSNR and SSIM measures. Computing the delta quantities based on the same face images removes the bias that may arise from such individualistic characteristics.

Hence, we compute the Pearson Correlation Coefficients (PCC) between the delta quantities, to determine the linear correlation between them. The PCC of any two quantities is a scalar value between -1 to +1, where -1 denotes a total negative correlation, 0 denotes no correlation and +1 denotes a total positive correlation.

From figure 3 we can see a positive correlation between changes in Face Recognition Confidence, SSIM and PSNR. As might be expected, a stronger correlation of 0.6 is seen between ΔFRC and $\Delta SSIM$ as opposed to a correlation coefficient of 0.39 between ΔFRC and $\Delta PSNR$, given that the SSIM measure is based on a perception-based model and was invented to overcome the shortcomings of PSNR. Applying this understanding to our results from tables 1 and 2, we see insignificant changes in average SSIM values between our method and that that of [3], however the marked increase in average PSNR values leads to higher average FRC values across the datasets.

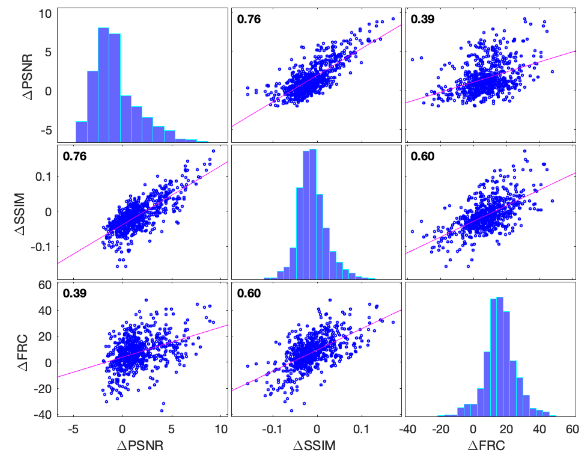

Fig. 3. Linear Correlation using Pearson Correlation Coefficients between PSNR, SSIM and FRC measures

4. CONCLUSION

In summary, we have developed a novel approach to classical learning-based face hallucination using face landmarks and demonstrated a marked improvement in the task of Low Resolution Face Recognition, compared to the state-of-the-art for a similar global hallucination scheme. Moreover, we have provided a conclusive analysis of the effect of PSNR and SSIM on face recognition confidence. For future research, we propose exploiting the use of face landmarks in a patch-based scheme.

5. REFERENCES

[1] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "Fsrnet: End-to-end learning face super-resolution with facial priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2492–2501, 2018.

[2] A. Bulat and G. Tzimiropoulos, "Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 109–117, 2018.

[3] Xiaogang Wang and Xiaoou Tang, "Hallucinating face by eigentransformation," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, pp. 425–434, Aug 2005.

[4] J. Jiang, C. Chen, K. Huang, Z. Cai, and R. Hu, "Noise robust position-patch based face super-resolution via tikhonov regularized neighbor representation," *Information Sciences*, vol. 367-368, pp. 354 – 372, 2016.

[5] H. Chen and S. Chien, "Eigen-patch: Position-patch based face hallucination using eigen transformation," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, July 2014.

[6] X. Pei, Y. Guan, P. Cai, and T. Dong, "Face hallucination via gradient constrained sparse representation," *IEEE Access*, vol. 6, pp. 4577–4586, 2018.

[7] J. Shi, A. Samal, and D. Marx, "How effective are landmarks and their geometry for face recognition?," *Computer Vision and Image Understanding*, vol. 102, no. 2, pp. 117 – 133, 2006.

[8] A. Juhong and C. Pintavirooj, "Face recognition based on facial landmark detection," in *2017 10th Biomedical Engineering International Conference (BMEiCON)*, pp. 1–4, Aug 2017.

[9] X. Xu, W. Liu, and L. Li, "Face hallucination: How much it can improve face recognition," in *2013 Australian Control Conference*, pp. 93–98, Nov 2013.

[10] M. T. Limited, "Face++ ai." `https://www.faceplusplus.com/`, 2012.

[11] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognition—a review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103 – 135, 2005.

[12] M. N. Dailey, G. W. Cottrell, and J. Reilly, "CAlifornia Facial Expressions (CAFE)," 2001.

[13] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902–913, 2010.

[14] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America. A, Optics and image science*, vol. 4, pp. 519–24, Apr 1987.