Fast Face Hallucination with Sparse Representation for Video Surveillance

Zhen Jia
United Technologies Research Center (China) Ltd.
Room 3502, 35/F, Kerry Parkside Office, No. 1155 Fang Dian Road
Shanghai, 201204, P. R. China

Hongcheng Wang, Ziyou Xiong and Alan Finn
United Technologies Research Center
411 Silver Lane, East Hartford
CT, 06108, USA

***Abstract*—In this paper we propose a novel face hallucination algorithm to synthesize a high-resolution face image from several low-resolution input face images. As described in Liu et al. [8]'s work, face hallucination uses two models: a global parametric model which synthesizes global face shapes from eigenfaces, and a local parametric model which enhances the local high frequency details. We follow a similar process to develop a robust face hallucination algorithm. Firstly, we obtain eigenfaces from a number of low resolution face images extracted from a video sequence using a face tracking algorithm. Then we compute the difference between the interpolated low-resolution face and the mean face, and use this difference face as the query to retrieve approximate sparse eigenfaces representation. The eigenfaces are combined using the coefficients obtained from sparse representation and added into the interpolated low-resolution face. In this way, the global shape of the interpolated low resolution face can be successfully enhanced. Secondly, we improve the example-based super-resolution method [7] for local high frequency information enhancement. Our proposed algorithm uses the Approximate Nearest Neighbors (ANN) search method to find a number of nearest neighbors for a stack of queries, instead of finding the exact match for each low frequency patch as presented in [7]. Median filtering is used to remove the noise from the nearest neighbors in order to enhance the signal. Our proposed algorithm uses sparse representation and the ANN method to enhance both global face shape and local high frequency information while greatly improving the processing speed, as confirmed empirically.**

I. Introduction

Super-resolution is desirable in many scenarios where objects of interests are not clear to the observers due to far distance or blurriness. Among different super-resolution techniques, of particular interest is to infer high-resolution (abbr. high-res) images from low-resolution (abbr. low-res) ones. This problem was introduced by Baker and Kanade [4] as hallucination. This technique has many applications in image enhancement, image compression and object recognition. The focus in our paper is for face super-resolution from surveillance videos. Face super-resolution is especially useful in surveillance applications where the resolution of a face image is limited by the resolution of the surveillance video, and the details of facial features may be crucial for identification and forensic analysis.

Recently, Liu et al. [8] proposed a two-step approach to hallucinate low-res face images by decomposing face appearance into a global PCA eigenface model [9] and a local patch Markov network model. One critical factor in their face hallucination algorithm is that faces need to be accurately registered. To solve the face registration problem, they designed a low-res face registration tool to follow face detection so high-res faces can be automatically hallucinated from low-res images with no manual intervention. Although this algorithm yields good results, the holistic Principle Component Analysis (PCA) model tends to yield results similar to the mean face, and the probabilistic local patch model is complicated and computationally demanding. Also the face registration step is computationally expensive and the quality of the face hallucination results is highly sensitive to the accuracy of face registration process. If faces are not well aligned, there will be serious "ghost effects" in the resulting images. When many faces are used for PCA training, there will be numerous eigenfaces, resulting in lengthy computation times to find coefficients for global shape reconstruction.

In this paper, we followed a process similar to Liu's, but their approach can't be directly applied here because their face hallucination method creates a "new" face using the mean face and eigenfaces from a prior training dataset, which is not appropriate for forensic applications and also very time consuming. Our algorithm hallucinates faces using information from previous face images extracted from the same video employing a face detection and tracking algorithm. The speed of our method is also much fast. Therefore, these advantages make our method more suitable for security applications.

Recently sparse representation has been successfully applied to many other related inverse problems in image processing, such as denoising and restoration [5], often improving on the state-of-the-art. In Yang et al. [11], they present a new approach to conduct single-image super-resolution based on sparse signal representation. Research on image statistics suggests that image patches can be well-represented as a sparse linear combination of elements from an appropriately chosen over-complete dictionary. Inspired by this observation, they seek a sparse representation for each patch of the low-resolution input, and then use the coefficients of this representation to generate the high-resolution output.

In our case, for global face enhancement, we can also consider eigenfaces as the global image patches and they are over-complete because one query face can only match a very small number of eigenfaces with similar shape while from

PCA training a large number of eigenfaces will be generated if the training face datasets have certain variations. If not considering this sparsity of eigenfaces representation, then global enhancement will generally introduce "ghost effects" because less similar eigenfaces will be also employed for enhancement. However, for local face enhancement, since the number of local patches is very large, estimating the sparse representation like Yang et al. [11] for local patches is very time consuming. In our research, we develop one efficient database retrieval method for local enhancement instead of using sparse representation.

Inspired by Liu and Yang's approaches, we propose a face hallucination algorithm that has the following advantages:

1) The algorithm leverages the power of sparse representations for better global face reconstruction to reduce over-enhancement "ghost effects".
2) The algorithm improves the retrieval speed of local patches for example-based super-resolution.
3) In order to avoid errors and reduce the computational burden introduced by image registration, our proposed algorithm uses face tracking to track a group of similar faces and then uses them for face hallucination. Since only similar faces will be tracked together, these faces will have very similar poses, shapes and illumination conditions, enabling the elimination of the image registration step.
4) Different from the state-of-the-art approaches like Liu et al. [8], we do not use dictionaries with high resolution faces for hallucination. What we do as one major novelty is to combine similar face features from several low resolutions tracked faces to enhance the target face. In this way, the algorithm does not require a prior training database, since tracked faces are used to generate eigenfaces. The number of eigenfaces is thereby reduced to reduce computational cost and also avoide the introduction of additional information from other faces.

II. Proposed Algorithm

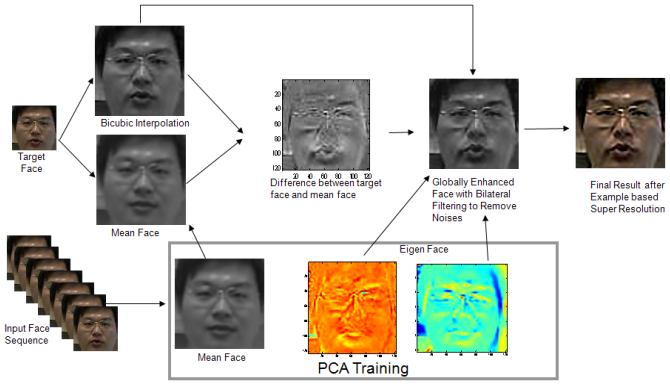


Fig. 1. Face Hallucination Algorithm Flow Chart

The proposed algorithm is composed of the following steps (Fig. 1):

1) Faces are detected from a video sequence. Then these faces are tracked and similar faces are grouped together for face hallucination.
2) For a sequence of similar faces, PCA training [9] is applied to generate the mean face μ and eigenface database B.
3) The low resolution face is interpolated (e.g. bicubic interpolation) and deblurred using Eq.1, where the observed low resolution image I_L is a blurred and downsampled version of the high resolution image I_H. Based on [6], I_L is the input low resolution face, C is the inverse blurring matrix and H is the inverse decimation matrix. Here, we use the interpolated low resolution image to approximately represent the high resolution face image. Later the interpolated mean face $\bar{\mu}$ (same size as $\bar{I_H}$) is subtracted from the approximated high resolution face $\bar{I_H}$ to generate the difference face $Diff_{Face}$. This difference face shows the lost global face shape for the target face.

$$\begin{aligned} I_L &= C \cdot H \cdot I_H \\ \bar{I_H} &\approx C^T \cdot H^T \cdot I_L \end{aligned} \quad (1)$$

4) With the difference face, we estimate the sparse coefficients (α) [11] to combine eigenfaces in order to approximate the difference face (Eq. 2). Eq. 2 shows this process. Since during PCA training eigenfaces are generated based on differences between every training face and the mean faces, we estimate the sparse representation for the difference face instead of the original face.

$$\begin{aligned} \bar{I_H} - \bar{\mu} &= Diff_{Face} \approx \alpha \cdot B \\ I_{H,GlobalEnhanced} &= \bar{I_H} + \alpha \cdot B \end{aligned} \quad (2)$$

5) Since the combined eigenfaces magnify the lost details of the target face with additional information from the training face database, we add the sparsely combined eigenfaces $\alpha \cdot B$ into the interpolated face ($\bar{I_H}$) to get the globally enhanced face ($I_{H,GlobalEnhanced}$).
6) The globally enhanced face is then filtered by bilateral filtering to remove noise and artifacts. The bilateral filtering is used before example-based super-resolution to avoid enhancing noise and artifacts.
7) After noise reduction, the globally enhanced face ($I_{H,GlobalEnhanced}$) is further enhanced by our improved example-based super-resolution method, and returned as the final hallucinated face image. In this step we don't further increase the resolution but simply add more high frequency information into $I_{H,GlobalEnhanced}$.

For sparse representation, we use the method presented in [11]. The following equation is optimized to get the sparse representation coefficients α for global face shape enhancement.

$$min \quad \lambda \|\alpha\| + \frac{1}{2} \left\| B \cdot \alpha - \bar{I_H} \right\| \quad (3)$$

where the parameter λ is a constant. More details about sparse representation for super resolution can be found in [11].

Generally our approach works better when hallucinating very low resolution face images such as those extracted from surveillance video. This is one major advantage of our method compared with other face super resolution approaches because we can enhance both global face shape information together with local face high frequency information. In Fig. 2, it is

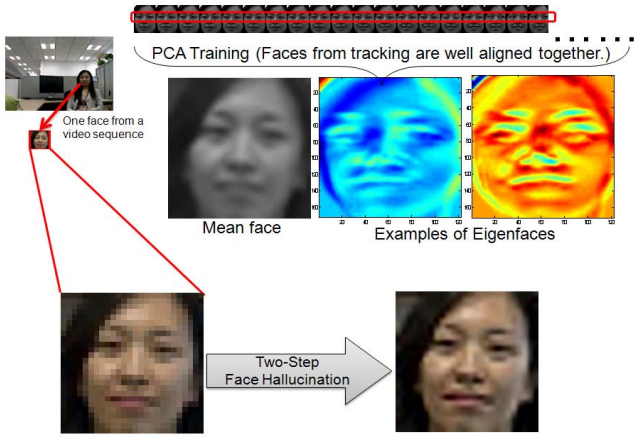


Fig. 2. Results of the proposed algorithm applied to a low resolution face input. The input faces have very low resolution. However using our proposed method, we can see that the face image quality is greatly improved.

difficult to identify the person in the raw video, however, using our proposed approach the face is both globally and locally enhanced and suitable for identification and recognition purposes.

A. Fast Example-based Super-Resolution with Approximate Nearest Neighbors Search

In our proposed algorithm, Freeman's example-based super-resolution algorithm [7] is another major step. However there is one major limitation in Freeman's approach. The patch searching step is very time-consuming. Each patch from the interpolated image is searched against the whole training database and the dimension of patch data is also high. Therefore if the training data is large and the size of the interpolated image is large, this iterative searching process will take a very long time. In Freeman's algorithm, they use a tree-based, Approximate Nearest Neighbors search. They also use a best-first tree search to find a good match. However, their method still needs many iterations and the speed is not fast enough for real-time processing. Therefore, in order to conduct example-based super-resolution for visual surveillance applications, we still need to improve the searching speed while maintaining good super resolution results.

From the literature, Approximate Nearest Neighbors searching [3] [1] is a well-known database indexing and searching method, which quickly and accurately retrieves nearest neighbors from a database. The basic idea of Approximate Nearest Neighbors searching is that the following: in some applications it may be acceptable to retrieve a "good guess" of the nearest neighbor. In those cases, we can use an algorithm which does not guarantee to return the actual nearest neighbors in every case, in return for improved speed or memory savings. Often such an algorithm will find the nearest neighbors in a majority of cases, but this depends strongly on the dataset being queried. Algorithms which support the Approximate Nearest Neighbors search include the Best Bin First and the Kd-Tree. An ϵ-approximate nearest neighbors search is becoming an increasingly popular tool for dealing with the curse of dimensionality.

In this paper, we develop a new algorithm to employ Approximate Nearest Neighbors search to efficiently retrieve high frequency information from training data, which can dramatically increase the speed of Freeman's example-based super-resolution method.

In order to improve the searching speed, one obvious idea is to reduce the iteration times. Therefore, different from patch by patch searching, we store all the patches into one stack (Fig.3) and use this stack as a query to search against the training database. Then for each individual patch the corresponding patch high frequency information is extracted from Approximate Nearest Neighbors. In this way we only need to do one-time searching and get the Approximate Nearest Neighbors for all the query patches, which can greatly reduce the time required to perform a search.

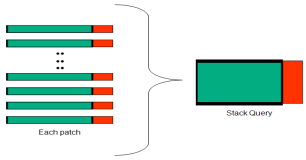


Fig. 3. Store all the patches from the interpolated image into one stack as the query. Here the green part is the low frequency information and the red part is the high frequency information. We search the low frequency information in the database and then get the corresponding high frequency information for local enhancement.

Although the speed is greatly increased based on the above idea, there is one trade-off that is important to keep in mind. Because we are using ANN searching on the full image patches, the returned nearest neighbors for the whole image are not necessarily accurate for each individual patch. In order to improve the accuracy, we propose another improvement. According to the theory of ANNs [3] [1], the time to find one exact nearest neighbors of a query is similar to the time for finding a number of nearest neighbors of a query. Therefore, we find a number of nearest neighbors and then use a median filter to smooth all the high frequency information from the retrieved nearest neighbors, and in this way we can filter out noise and artifacts and at the same time enhance the high frequency signal. This new process is depicted in Fig. 4.

In next section, we will present more results to demonstrate the performance of our algorithm.

III. Experimental Results

In order to evaluate the performance of the proposed face hallucination algorithm, we test the algorithm's performance using surveillance videos that show people walking toward the camera. Due to the length limitation of this paper, here

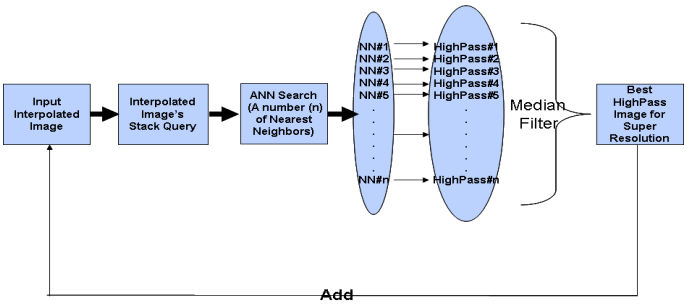


Fig. 4. Full Image ANN Search Process based on Median Filtering to Reduce Noise and Artifacts for Example-based Super-Resolution.

we listed some key results to demonstrate the effective of our proposed algorithm. Before conducting face hallucination, the system follows the steps illustrated in Fig. 5. The first

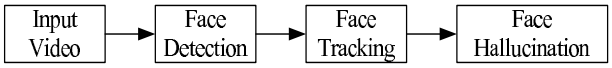
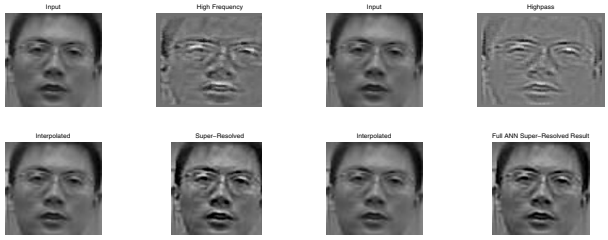


Fig. 5. Face Hallucination Experimental Flow Chart

Fig. 6. One example video for our testing case. A person is walking towards the camera. His face is detected and tracked for face hallucination.

step in the process is to conduct face detection. There are many face detection algorithms [12]. Among them, Viola and Jones' approach [10] is proven to be one of the most robust and efficient approaches, and consequently we use this method in the empirical studies presented in this paper. After faces are detected, we track these faces to distinguish different people's faces and temporarily associate the same person's faces into one group. We can choose from many different tracking methods (more details can be found from [13]) to track faces. In our research, we use detected faces as measurements, then use the Multiple Hypotheses Tracking (MHT) method for data association and finally we use Kalman Filtering for faces status filtering. In order to achieve a better data association, we use a face color histogram and color layout information, such as the Color Layout Descriptor from MPEG-7 [2] for a similarity measure. From our testing, our face tracking method works very well and we can successfully track different faces through a video sequence. After face tracking, we have grouped one person's low resolution input faces together for face hallucination.

Next, we evaluate the performance of our proposed algorithm to use ANN to improve the speed of the example-based super-resolution algorithm. As a baseline, we show the original example-based super-resolution performance of Freeman's algorithm. As a note, we mention that during our testing, we will only process the Y channel of the YUV color face images to enhance the contrast only, but not the color information. This can help us to reduce the computational cost and also avoid color over-enhancement artifacts. The time

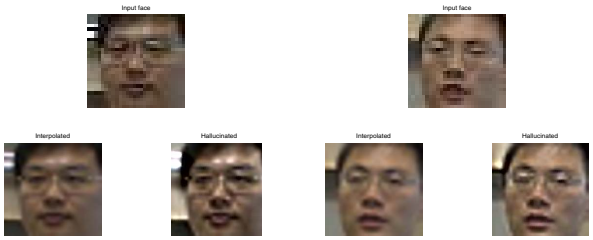


(a) (b)

Fig. 7. Fig.7(a) shows the example-based single image super resolution result based on Freeman's algorithm. The patch by patch searching is conducted by the Kd-Tree method. Fig.7(b) shows the super resolution result using our proposed method. Full ANN here means that we use the full image patches as one query for the ANN-based search. Here the ϵ value is 0, which indicates the maximum accuracy with slowest speed

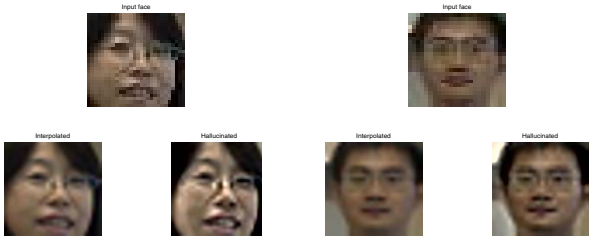
of Freeman's original example-based method to enhance one single face is 32.47s. Here our testing is done with a Matlab implementation, the input image size is 45 x 45 and the super resolution factor is 4. The computer for our testing DELL M4400 engineering laptop, with Intel Core Duo CPU T9600 (2.80GHz) and 3.5GB of RAM. Fig.7(b) is the result of our proposed ANN-based algorithm. We can see that by using the new ANN search process the high frequency information is enhanced with stronger details and the noise and artifacts are smoothed out as well. The final super resolution result is almost visually identical to the Kd-Tree-based patch by patch searching result. At the same time, the computational speed is much faster (6.55s, almost 5x faster than the patch by patch searching method). There is one parameter ϵ in ANN method, which controls the upper bound on the searching error. Typically, ϵ controls the trade-off between efficiency and accuracy. When ϵ is set larger, the approximation is less accurate, and the search completes faster. This ϵ is the major parameter that affects the performance and speed of ANN search.

Additional testing of our algorithm produces similar results for face hallucination. In Fig. 8 and 9, faces are detected and tracked from the input video. Consolidating all the tracked, low resolution faces together, we use our proposed approach for face hallucination. Here our processing is done on the Y channel of the YUV color space. Comparing experimental results, we can see that our face hallucination results are much better than those generated using the bi-cubic interpolation method. Here we note that both the global face shape information and the local face details are much improved. Consequently, this hallucinated face can provide better information to a security guard. Additionally, unlike other face hallucination approaches, our output faces can be used in a face recognition



(a) (b)

Fig. 8. Face Hallucination Results. 10 low resolution tracked faces (size: 25x25) are used for face hallucination tests. The super-resolution increase factor is 4.



(a) (b)

Fig. 9. Face Hallucination Results. 10 low resolution tracked faces (size: 25x25) are used for face hallucination testing. The super-resolution increase factor is 4.

system for forensic applications because we only use tracked faces' information.

We also compare our method with some state-of-the-art face super-resolution/hallucination methods like [8] and [7]. Due to the paper length limitation, we don't present all the results here. In our future journal version paper, we will present more experimental comparison results together with quantitative performance analysis.

IV. Conclusion

In this paper, we propose a novel face hallucination algorithm for enhancement of face images extracted from security video surveillance. Building upon the work on face hallucination and super-resolution published by Liu et al.[8] and Yang et al.[11] we present one novel algorithm that improves the state of the art in the following two aspects: First, the use of sparse representation to estimate coefficients to fuse eigen faces from face training database for face global shape enhancement. Second, the use of Approximate Nearest Neighbors (ANN) searching together with stack query idea and median filtering method to conduct local example based super solution for face local high frequency information enhancement.

In detail, our proposed approach has the following advantages:

- Detected and tracked faces from a video sequence are used for face hallucination. The faces are all from the same person. No additional information such as prior training database is used from other sources. This is appropriate for forensic applications.
- Tracked faces are very similar to each other and we don't need to conduct face alignment. The computational cost is reduced.
- Global face features are approximately determined by the difference between the interpolated face and a mean face and enhanced by the combination of eigen-faces. In this way, face global shape information can be enhanced.
- The database size for PCA training is rather small since we use tracked faces. This can reduce the overall computational cost.
- A sparse representation method is used to estimate the coefficients for eigen-face fusion. In our approach, we leverage the power of sparse representation to avoid globally over-enhancement to reduce "ghost effects".
- Using our new example based super resolution method with ANN, the algorithm speed is greatly improved (almost $20x$ speed improvement over the traditional Freeman's algorithm). At the same time, the result from our proposed approach is almost visual identical to the original result by Freeman's algorithm.
- Bilateral filtering is used before the final example-based super-resolution. In this way, we can avoid noises or artifacts over-enhancement.

From experimental testing, the quality of the low resolution faces is greatly improved. This technology is suitable for face image super resolution in a non-cooperative environment for security applications.

References

[1] *http*://en.wikipedia.org/wiki/nearest_neighbor_search.
[2] *http*://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm.
[3] *http*://www.cs.umd.edu/∼mount/ann/.
[4] S. Baker and T. Kanade. Hallucinating faces. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, March 2000.
[5] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15:3736C374, 2006.
[6] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar. Robust real-time face detection. *IEEE Transactions on Image Processing*, 13(10):1327–1344, 2004.
[7] W. T. Freeman, T. R. Jones, , and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Application*, 22(2):56–65, 2002.
[8] C. Liu, H. Y. Shum, and W. T. Freeman. Face hallucination: theory and practice. *International Journal of Computer Vision*, 75(1):115–134, 2007.
[9] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, June 1991.
[10] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, May 2004.
[11] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 2010.
[12] M.-H. Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
[13] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys*, 38(4), 2006.