Multimed Tools Appl (2015) 74:1845–1862
DOI 10.1007/s11042-013-1721-4

A two-step face hallucination approach for video surveillance applications

Zhen Jia · Jianwei Zhao · Hongcheng Wang · Ziyou Xiong · Alan Finn

Published online: 12 November 2013

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. Face hallucination normally uses two models: a global parametric model which synthesizes the 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. First, we obtain eigenfaces from a number of low resolution face images segmented from a video sequence using a face tracking algorithm. Then we compute the difference between an interpolated low-resolution face and a mean face, and use this difference as the query to retrieve an approximate sparse eigenface representation. The eigenfaces are combined using the coefficients obtained from the 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. Second, we improve the example-based super-resolution method 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. Median filtering is used to remove the noise from the nearest neighbors in order to enhance the signal. Our proposed algorithm uses a 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.

Z. Jia (✉) · J. Zhao
United Technologies Research Center (China) Ltd.,
Room 3502, 35/F, Kerry Parkside Office, No. 1155 Fang Dian Road,
Shanghai, 201204, People's Republic of China
e-mail: jiaz@utrc.utc.com

H. Wang · Z. Xiong · A. Finn
United Technologies Research Center,
411 Silver Lane, East Hartford, CT 06108, USA

Keywords Face detection · Face tracking · Face recognition · Face hallucination · Super-resolution · Sparse representation · Visual surveillance

1 Introduction

Super-resolution is desirable in many scenarios where objects of interests are not clear to the observers due to 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 [1] 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.

1.1 Related work

Liu et al. [12] proposed a two-step approach to hallucinate low-res face images by decomposing face appearance into a global PCA eigenface model and a local patch Markov network model [15]. 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 Principal 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 time to find coefficients for global shape reconstruction.

Freeman's example-based super-resolution algorithm [5] uses local information enhancement for face hallucination. However there is one major limitation – the patch search 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 search process takes a very long time. Freeman uses a best-first tree search to find a good match. However, their method still needs many iterations for each patch 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 search speed while maintaining good super resolution results.

Recently sparse representations have been successfully applied to many problems in image processing, such as de-noising and restoration [3] and face recognition [17], often improving on the state-of-the-art. Yang et al. [18] presents a new approach to conduct single-image super-resolution based on a 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. Yang also employs a two-step approach similar to Liu's [12], i.e., first find a suitable subspace for human faces and apply the reconstruction constraints to recover a medium resolution image, then they recover the local details using the sparsity prior for image patches.

Another recent approach is by Hu et al. [9], which developed a new face hallucination framework called 'local pixel structure to global image super-resolution' (LPS-GIS). Based on the assumption that two similar face images should have similar local pixel structures, their new framework first uses the input low-resolution face image to search a face database for similar high-resolution faces in order to learn the local pixel structures for the target high-res face. It then uses the input low-res face and the learned pixel structures as priors to estimate the target high-res face.

Besides the above mentioned state-of-the-art face hallucination approaches, there are other approaches such as [10, 11, 14]. We will not go to the details of all these approaches since they are in fact not for visual surveillance applications, which is our goal.

1.2 Contributions

In this paper, we follow a process similar to Liu's [12], Hu's [9] and Yang's [18] approaches to develop a two-step procedure that considers both local detail enhancement and global face shape enhancement, since in this way the face hallucination can achieve the maximum enhancement. However, these state-of-the-art approaches can't be directly applied here because their face hallucination methods created a "new" face shape 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 global face shapes using information from previous face images from the same person extracted from the same video employing a face detection and tracking algorithm. The speed of our method is also fast. Therefore, these advantages make our method more suitable for potential security applications.

In our case, for global face shape enhancement, we can also consider eigenfaces as global image patches and they are over-complete because one query face can only match a very small number of eigenfaces with similar shapes 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 eigenface representation, then global enhancement will generally introduce "ghost effects" because less similar eigenfaces will also be 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. [18] for local patches is very time consuming. In our research, we have developed an efficient database retrieval method for local enhancement instead of using sparse representation.

Inspired by Liu and Yang's approaches [12] and [18], we propose a face hallucination algorithm that has the following advantages:

1. The algorithm leverages the power of sparse representations for better global face shape 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 (which naturally includes segmentation) to track a group of similar faces and then uses them for face hallucination. Since only similar faces will be tracked, these faces will have similar poses, shapes and illumination conditions, enabling the elimination of a separate image segmentation and registration step.
4. The global face shape enhancement algorithm does not require a prior training database, since tracked faces are used to generate eigenfaces. In this way, we reduce the number of eigenfaces thereby reducing computational cost and also avoiding the introduction of additional shape information from other faces.

2 Proposed algorithm

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 [15] is applied to generate the mean face μ and eigenfaces B. The eigenfaces B contains all the eigenvectors of the tracked low-resolution faces. For example, the tracked low resolution face has a size of 900 (30×30) pixels (transformed to a 1-D vector for processing) and if we have 10 eigenvectors from the PCA training, then the eigenfaces B's size is 900×10.

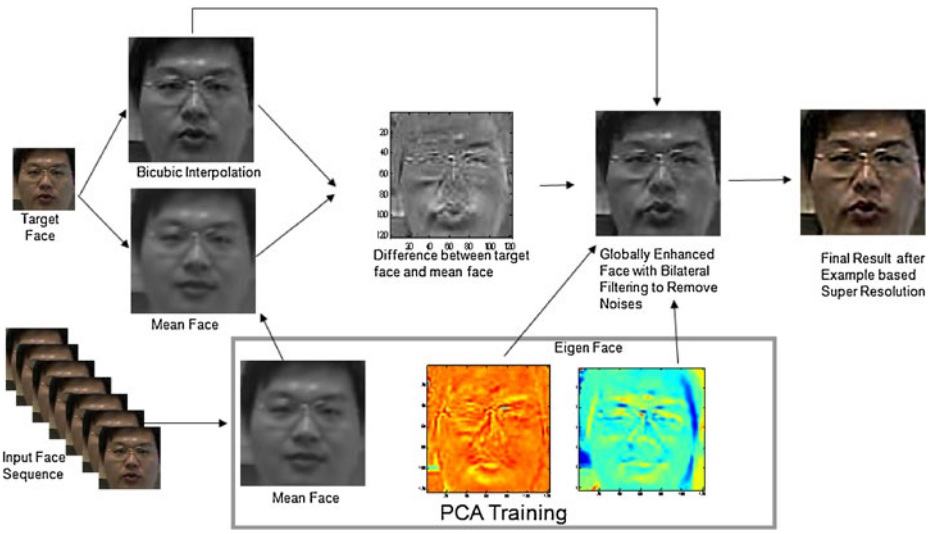

Fig. 1 Our proposed face hallucination algorithm flow chart

3. The low resolution face is interpolated (e.g. bicubic interpolation) and deblurred using equation (1), where the observed low resolution image I_L is considered to be a blurred and down-sampled version of the high resolution image I_H. Based on [4], I_L is the input low resolution face, C is the blurring matrix and H is the decimation matrix which transforms the high resolution image I_H to I_L. Here, we use the interpolated low resolution image to approximately represent the high resolution face image with the inverse of C and H. 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 $Diff_{\text{Face}}$. This difference face shows the lost global face shape for the target face.

$$\begin{aligned} I_L &= C \cdot H \cdot I_H \\ \bar{I_H} &\approx C^{-1} \cdot H^{-1} \cdot I_L \end{aligned} \tag{1}$$

4. With the difference face, we estimate the sparse coefficients (α) [18] to combine eigenfaces in order to approximate the difference face. Equation (2) shows this process. Since PCA training eigenfaces are generated based on the differences between every training face and the mean face, we estimate the sparse representation for the difference face instead of the original face.

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

   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 ($I_{H,\text{GlobalEnhanced}}$).
5. The globally enhanced face is then filtered by bilateral filtering to remove noise and artifacts. The bilateral filtering is used before local example-based super-resolution to avoid overenhancing noise and artifacts.
6. After noise reduction, the globally enhanced face ($I_{H,\text{GlobalEnhanced}}$) 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 $I_{H,\text{GlobalEnhanced}}$.

For the sparse representation (Step 4 above), we follow Yang's method presented in [18]. Yang used the following equation to estimate the sparse coefficients to find high resolution patches from the training database.

$$\min_{\alpha} \quad \frac{1}{2}||FD\alpha - Fy||_2^2 + \lambda||\alpha||_1 \tag{3}$$

where F is a feature extraction operator, D is the high-resolution patch training database, y is the observed low-resolution image and λ balances sparsity of the solution and fidelity of the approximation to y.

Following the same procedure as above, the following equation is optimized to get the sparse representation coefficients α for the global face shape enhancement.

$$\min_{\alpha} \quad \lambda \|\alpha\|_1 + \frac{1}{2} \left\| B \cdot \alpha - \bar{I_H} \right\|_2^2 \tag{4}$$

where the parameter λ is a similar constant. As in the previous example, if B's size is 900×10 and $\bar{I_H}$'s size is 900, then α will be a 10×1 vector of the eigenface weights.

Here the computed sparse representation adaptively selects the most relevant eigenfaces from the training database to best represent the given interpolated low-resolution face $\bar{I_H}$. This leads to superior performance, both qualitatively and quantitatively, compared with the traditional local example based face hallucination/super resolution approaches such as [5, 12], which normally use a fixed number of nearest neighbors or eigenfaces. Our method can generate sharper edges and clearer textures. In additional, the sparse representation is robust to noises and thus our algorithm is more robust to noise in the input low-resolution image, while most other algorithms don't perform denoising and super-resolution at the same time. More details about sparse representation for computer vision applications can be found in references such as [17, 18].

Here, similar as Yang's approach [18], we also first use sparse representation to reconstruct constraints to recover a medium high-resolution face image $I_{H,\text{GlobalEnhanced}}$, but this solution is searched only in the face subspace (eigenfaces B) and the next step (explained in the next section) is to use the local sparse model to recover the image details. Different from the local patch-by-patch based image resolution approach mentioned in [18], this step in our method uses the eigenfaces to enhance the global face shapes, which helps to maintain and enhance the face's global shape structure.

Generally our approach works better when hallucinating very low resolution face images such as those directly 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 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.

2.1 Fast example-based super-resolution with approximate nearest neighbors search

Approximate Nearest Neighbors (ANN) search [6, 8] is a well-known database indexing and search method, which quickly and accurately retrieves nearest neighbors from a database.

In the nearest neighbor problem a set P of data points in d-dimensional space is given. These points are preprocessed into a data structure, so that given any query point q, the nearest (or generally k nearest) points of P can be retrieved efficiently. One difficulty with exact nearest neighbor search is that for virtually all methods other than brute-force search, the running time or space grows exponentially as a function of dimension. Consequently these methods are often not significantly better than brute-force search, except in fairly small dimensions. However, it has been shown [8] that if the user is willing to tolerate a small amount of error in the search (returning a point that may not be the nearest neighbor, but is not significantly

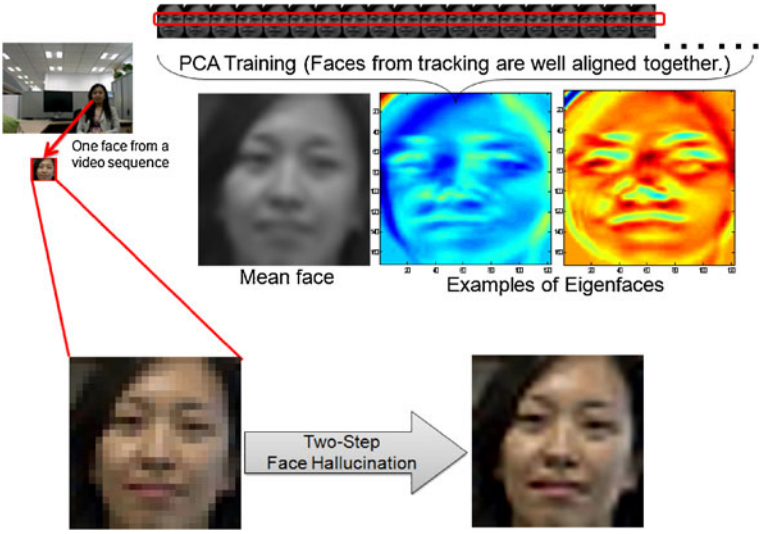

Fig. 2 Results of the proposed algorithm applied to a low resolution face input. The input faces have very low resolution. However using our proposed method, we can see that the face image quality is greatly improved with much more global and local face details

further away from the query point than the true nearest neighbor) then it is possible to achieve significant improvements in running time. ANN is such a system designed for obtaining nearest neighbor queries both exactly and approximately.

From literature comparisons [13], ANN performs efficiently for point sets ranging in size from thousands to hundreds of thousands, and in dimensions as high as 20. The basic idea of Approximate Nearest Neighbors search is that in some applications it may be acceptable to retrieve a "good guess" of the nearest neighbors. 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 example-based super-resolution methods. Here we will not further describe the ANN details since it is out of the scope of our paper. For more details of the approach, please refer to [6, 8, 13].

In order to improve the search speed, one obvious idea is to reduce the iteration times. Therefore, different from patch by patch search, we can 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 high frequency information is extracted from the training dataset with the ANN search method. In this way we only need to do a one-time search and get the Approximate Nearest Neighbors for all the query patches. This can greatly reduce the time required to perform a search.

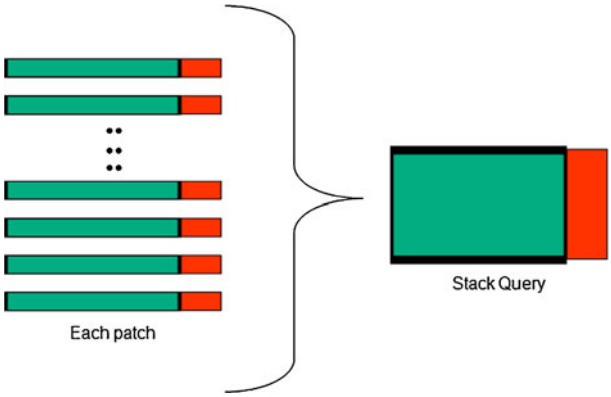

Fig. 3 Store all the patches from the interpolated image into one stack as the query. Here the *green part* is the low frequency information and the *red part* is the high frequency information. We search the low frequency information in the database and then get the corresponding high frequency information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

Although the speed is greatly increased based on the above idea, there is one important trade-off. Because we are using ANN search 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 [6, 8], the time to find one exact nearest neighbor 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.

Having introduced the details of the two-step face hallucination approach, we will next present results to demonstrate the performance of our algorithm.

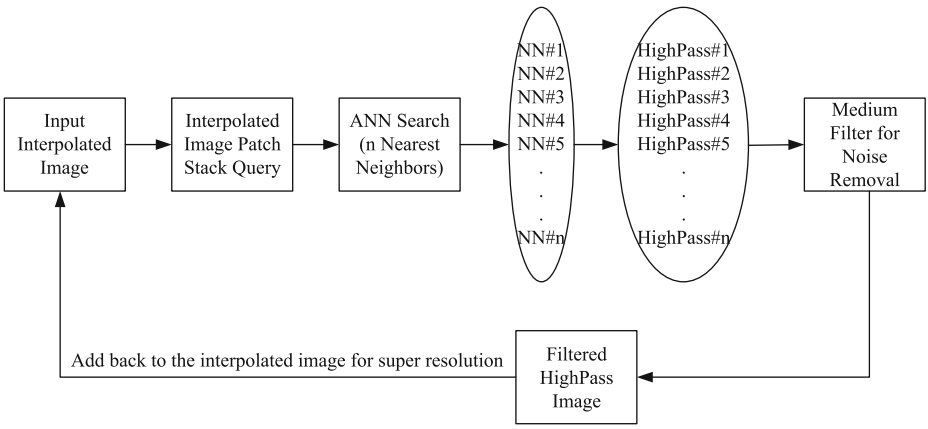

Fig. 4 Full image ANN search process based on median filtering to reduce noise and artifacts for the local example-based super-resolution

3 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 towards the camera. Before conducting face hallucination, the system follows the steps illustrated in Fig. 5. The first step in the process is to conduct face detection (Fig. 6). There are many face detection algorithms [19]. Among them, the Viola and Jones' approach [16] is proven to be one of the most robust and efficient methods, 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 in [20]) to track faces. In our research, we use detected faces as measurements, then use the Multiple Hypotheses Tracking (MHT) [2] for data association and finally we use Kalman Filtering for face status filtering. In order to achieve a better data association, we use face color histogram and color layout information, such as the Color Layout Descriptor from MPEG-7 [7] for similarity measures. 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 faces together for face hallucination. In this paper, we mainly focus on introducing our new face hallucination algorithm, as a result we do not present much on the face detection and tracking. For more details of these approaches refer to the above references.

Next, we evaluate the performance of our proposed algorithm. As a baseline, we show the original example-based super-resolution performance of Freeman's algorithm [5]. 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 of Freeman's original example-based method to enhance one single face is 32.47 s (Fig. 7). Here our testing is done with a Matlab implementation, the input image size is 45×45 and the super resolution factor is 4. The computer for our testing is DELL M4400 engineering laptop, with Intel Core Duo CPU T9600 (2.80 GHz) and 3.5 GB of RAM.

Next, we conduct a parametric study for the ANN method. There is one parameter ϵ, which controls the upper bound on the search 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. There are some other parameters in the ANN search algorithm including:

- *use_b dtree*: whether to use box-decomposition tree (default = false).
- *b ucket_size*: the size of each bucket in the tree (default = 1).
- *split*: the name of the splitting rule in kd-tree construction (default = 'suggest').
- *shrink*: the name of the shrinking rule in bd-tree construction (default = 'suggest').

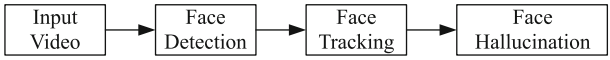

Fig. 5 Face hallucination experimental testing flow chart

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

- *search_sch*: the search scheme to use (default = 'std').
- *radius*: the maximum distance between the neighbors and the query point.

We have also studied all the above parameters. Except ϵ, other parameters do not greatly change the performance.

Figure 8 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 search result. At the same time, the computational speed is much faster (6.55 s, almost 5× faster than the Kd-Tree based patch by patch search method).

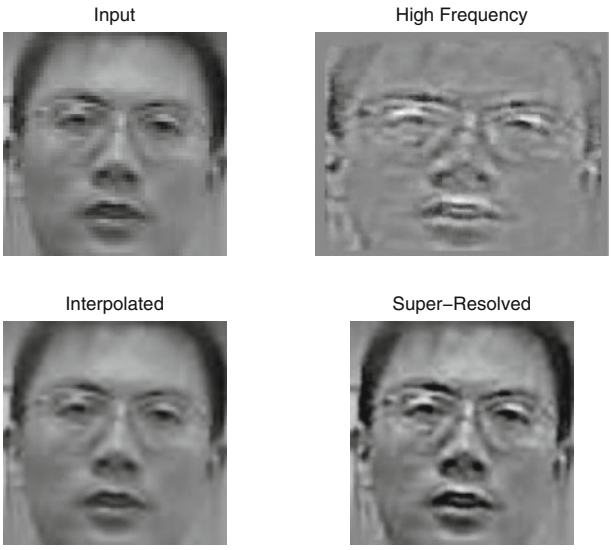

Fig. 7 Example-based single image super resolution result based on Freeman's algorithm. The patch by patch search is conducted by the Kd-Tree method

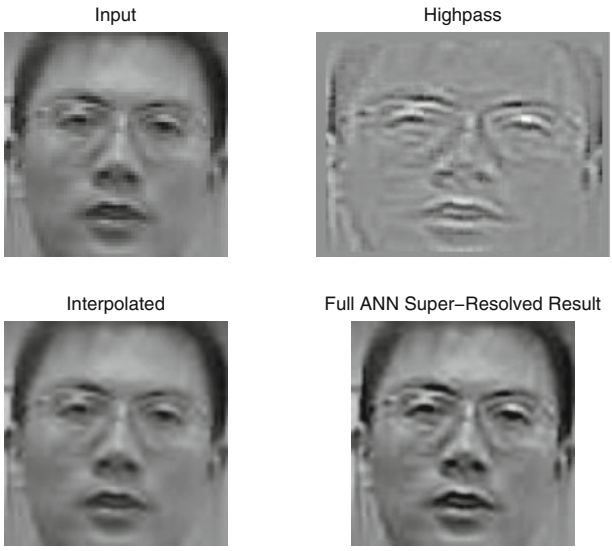

Fig. 8 The super resolution result using our proposed method. Full ANN here means that we use the full image patches as one query for the ANN-based search. Here the ϵ value is 0, which indicates the maximum accuracy, but with slowest speed

In Fig. 9 we show the results when we tested the performance and speed of the ANN algorithm with different values for ϵ. When the ϵ value is very large (larger than 5), the computational speed is almost constant. From Fig. 10, we can see that when ϵ is very large the super resolution performance degrades, while for small ϵ the super resolution results are almost visually identical. The reason for such small performance degradation is that we estimate multiple nearest neighbors and then the median filter can help to filter out noise while also enhancing the signal. Based on experimental testing, we chose $\epsilon = 4$. As the result shown in Fig. 9 attests, we can achieve a very low search times (here we get 1.58 s, almost 20× faster than the original Kd-Tree patch by patch search result) while maintaining a very good super resolution result (almost visually identical to the original Kd-Tree patch by patch search result).

Additional testing of our algorithm produces similar good results for our two-step face hallucination approach. In Figs. 11 and 12, faces are detected and tracked from the input video. Consolidating all the tracked, low resolution faces together, we use

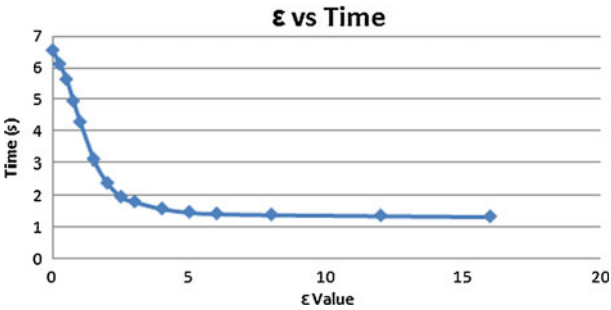

Fig. 9 Super resolution computational time using different ϵ values for ANN search

Fig. 10 Super resolution results using different ϵ values

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 traditional 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 system for forensic applications because we only use tracked faces' information for the global face shape enhancement.

One critical step in face hallucination is to align faces. In our proposed approach, we use face tracking to omit the face alignment step. The tracked faces are cropped out, and normalized for the face hallucination processing. However, from our testing, even with tracked faces, these faces may often have slight variations due to facial expression or pose changes. As a result, we may still have some "ghost effect" in the hallucinated results as shown in Fig. 13. The elimination of these "ghost effects" is planned for future work.

In our experimental testing, our two-step face hallucination approach has quite a few parameters such as the ones for face detection/tracking, ANN-based nearest neighbors search, bi-cubic interpolation, Kd-trees, sparse representation and so on. All these parameters have to be properly tuned to achieve good performance. Currently, these parameters are tuned based on experiments. In the future, how

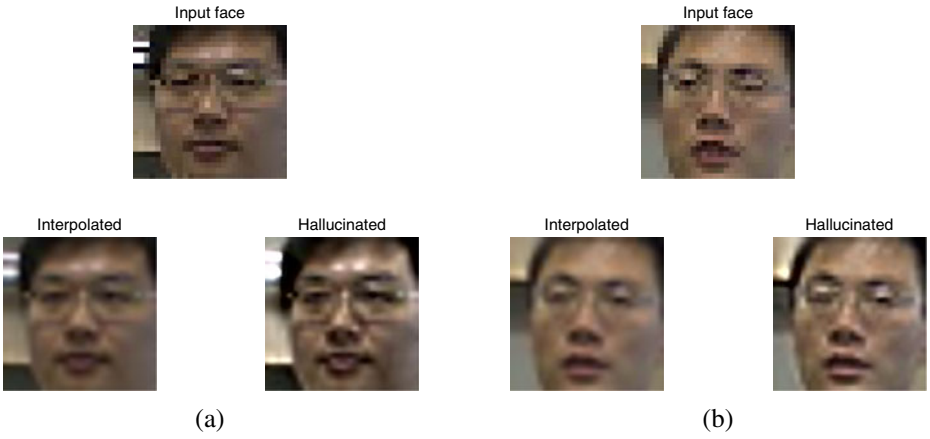

Fig. 11 Face hallucination results. 10 low resolution tracked faces (size: 25 × 25) are used for face hallucination tests. The super-resolution increase factor is 4

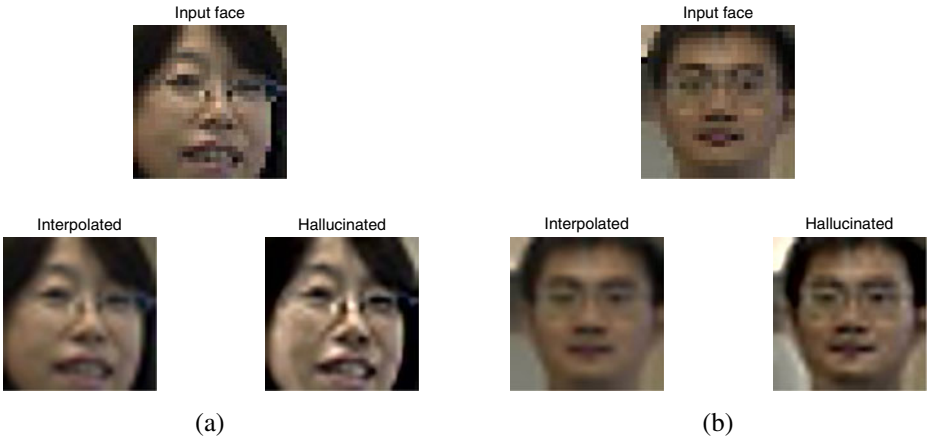

Fig. 12 Face hallucination results. 10 low resolution tracked faces (size: 25 × 25) are used for face hallucination testing. The super-resolution increase factor is 4

to automatically set up these parameters or make these parameters be adaptively adjusted based on the scene will be considered.

Computational cost is also one important factor for experimental results verification. In our testing, we mainly compare the time reduction results using the ANN approach for the speed improvement of the local example-based super-resolution algorithm. With our approach and Matlab implementation, we can achieve around 6.55 s to process one image with the size of 45 × 45 and super resolution factor of 4. On the other hand, for the global face shape hallucination, it will take longer time depending on the number of tracked faces for the eigenfaces creation and also the sparse representation (L_1-minimizer)'s optimization time. Here based on our testing, the computational time for the global face shape hallucination is on the order of 10 seconds with our Matlab implementation. Here we want to emphasize that our algorithm and Matlab code are developed for research testing purposes without any further algorithm and code. In the future, if we want to put the algorithm into practical applications, more work related to algorithm and code optimization needs to be conducted to further reduce the computational cost.

In this paper, we also try to compare the algorithm performance between our proposed approaches and also the approaches published in [12, 18] and [9], since we have conducted improvements over these state-of-the-art face hallucination approaches. However, for these approaches one key step is to align faces. Face alignment requires

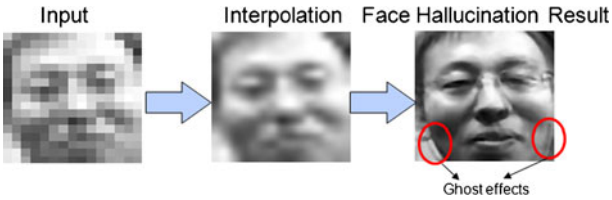

Fig. 13 "Ghost effects" in the hallucinated face image

face features detection such as eye detection for face position correction. In our approach, one of the advantages is that we don't need to align faces but use face tracking (which includes segmentation) to group similar faces together for face hallucination. Also, prior approaches require a pre-defined face database for training both for eigenfaces generation and dictionary training for sparse representation. In our approach, we avoid using an existing face database for training, because first the trained database may be large and this will introduce additional computational cost and second additional face information may make the hallucinated face (with the global face shape enhanced) not suitable for forensic applications due to additional information introduced from other persons' faces. Therefore, we did not compare our proposal approaches with these state-of-the-art approaches and mainly present the results from our research to show the effectiveness and efficiency for face hallucination.

In our future work, we will provide more quantitative analyses for different face hallucination approaches. In this paper, we mainly show some qualitative comparison results to demonstrate the efficiency and effectiveness of our proposed approach.

4 Conclusion

In this paper, we propose a novel face hallucination algorithm for enhancement of face images extracted from surveillance video. Building upon the work on face hallucination and super-resolution by Liu et al. [12], Yang et al. [18] and Hu et al. [9] we present a 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 eigenfaces from a captured face dataset for face global shape enhancement and second, the use of Approximate Nearest Neighbors (ANN) search together with the stack query idea and median filtering to conduct local example based super solution for face local high frequency information enhancement.

In detail, our proposed approach has the following advantages:

- We use the detected and tracked faces from a video sequence for face hallucination. The faces are all from the same person. No additional information is used from other sources. No prior training database is needed for the face global shape enhancement. This is appropriate for forensic applications if a number of low resolutions faces are provided for our proposed face hallucination algorithm.
- Since we are using tracked faces for face hallucination, 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 an interpolated mean face and then enhanced by the combination of eigenfaces. In this way, face global shape information can be enhanced.
- The faces used for PCA training are faces tracked for a short time. The database size is rather small, which can reduce the overall computational cost.

- A sparse representation method is used to estimate the coefficients for eigenface fusion. In our approach, we leverage the power of sparse representation to avoid globally over-enhancement to reduce "ghost effects" and noise.
- Using our new example based super resolution method with ANN, the algorithm speed is greatly improved (for a test image of 45×45 and super resolution factor is 4, the speed increase is almost 20 times over the traditional Freeman's algorithm's implementation with Kd-Tree). At the same time, the result from our proposed approach is almost visually 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 for potential security applications.

References

1. Baker S, Kanade T (2000) Hallucinating faces. In: Proceedings of IEEE international conference on automatic face and gesture recognition
2. Blackman S (2004) Multiple hypothesis tracking for multiple target tracking. IEEE Aerosp Electron Syst Mag 19(1):5–18
3. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15:3736–374
4. Farsiu S, Robinson MD, Elad M, Milanfar P (2004) Fast and robust multiframe super resolution. IEEE Trans Image Process 13(10):1327–1344
5. Freeman WT, Jones TR, Pasztor EC (2002) Example-based super-resolution. IEEE Comput Graph Appl 22(2):56–65
6. http://en.wikipedia.org/wiki/nearest_neighbor_search
7. http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm
8. http://www.cs.umd.edu/~mount/ann/
9. Hu Y, Lam, K-M, Qiu G, Shen T (2011) From local pixel structure to global image super-resolution: a new face hallucination framework. IEEE Trans Image Process 20:433–445
10. Jian M, Lam K-M, Dong J (2013) A novel face-hallucination scheme based on singular value decomposition. Patt Recogn 46(11):3091–3102
11. Jung C, Jiao L, Liu B, Gong M (2011) Position-patch based face hallucination using convex optimization. IEEE Signal Process Lett 18:367–370
12. Liu C, Shum HY, Freeman WT (2007) Face hallucination: theory and practice. Int J Comput Vis 75(1):115–134
13. Liu T, Moore AW, Yang K, Gray AG (2004) An investigation of practical approximate nearest neighbor algorithms. In: Advances in neural information processing systems, pp 825–832
14. Ma X, Huang H, Wang S, Qi C (2010) A simple approach to multiview face hallucination. IEEE Signal Process Lett 17:579–582
15. Turk M, Pentland A (1991) Face recognition using eigenfaces. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 586–591
16. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154
17. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
18. Yang J, Wright J, Huang T, Ma Y (2010) Image super-resolution via sparse representation. IEEE Trans Image Process 19:2861–2873
19. Yang M-H, Kriegman D, Ahuja N (2002) Detecting faces in images: a survey. IEEE Trans Pattern Anal Mach Intell 24(1):34–58
20. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv. doi:10.1145/1177352.1177355

Zhen Jia is a Group Leader for the Computational Intensive Signal Processing group at United Technologies Research Center (China) Ltd. Dr. Jia joined UTRC in 2006. Since then, he has been working on advanced technology development and innovation in the area of video surveillance and security. His research interests are computer vision, image processing, pattern recognition, machine learning and robotics. Prior joining UTRC, Dr. Jia had research experiences with French National Research Institute of Computer Science and Control (INRIA), University of Technologies, Sydney (UTS), and Singapore Technologies Kinetics (STK) in 2004 and 2005. Dr. Jia has published ten journal papers and over 25 conference papers, and he also has seven patent applications. Dr. Jia obtained the B.E. degree from Shanghai Jiao Tong University (China) in 2002 and Ph.D. degree from Nanyang Technological University (Singapore) in 2006, both in electrical engineering. In 2012, he also completed his MBA degree from the joint program between Tongji University (China) and ENPC (France). He is a member of IEEE.

Jianwei Zhao is the engineering manager for video technology at United Technologies Climate, Controls and Security Systems Shanghai Design Center. She joined UTRC in 2006, and has been working on advanced technology development and innovation in the area of video surveillance and security. In 2012 she joined UTC CCS for her current position. Her research interests are computer vision, video processing, pattern recognition and machine learning. She obtained her PhD degree in pattern recognition and intelligent systems from Shanghai Jiao Tong University in 2003, and she received her MBA from the joint program between Tongji University (China) and ENPC (France) in 2009.

Hongcheng Wang is a Staff Scientist and a Principle Investigator at United Technologies Research Center (UTRC), East Hartford, Connecticut, USA. He has more than 10 years of experience in computer vision, machine learning and computer graphics. Since he joined UTRC in 2006, he has been working and leading innovation or government projects on video surveillance, UAV perception, and visual inspection. Dr. Wang had short-term research experience with Mitsubishi Electric Research Laboratories (MERL), Siemens Corporate Research (SCR) and IBM Almaden Research Center on various computer vision projects in 2003, 2004 and 2005, respectively. Dr. Wang has published 30+ papers in premium computer vision/graphics and machine learning journals and conferences, such as ICCV, CVPR, ICRA, IJCV, SIGGRAPH, etc. He holds 7 U.S. patents and 15 U.S. patent applications. He is serving as an Associate Editor (AE) for the Visual Computer Journal and an Area Chair (AC) for IEEE WACV 2014. He has served regularly as a Program Committee (PC) member or a reviewer for premium conferences and journals, such as ICCV, CVPR, ECCV, ACCV, PAMI, etc. He received his M.S. and Ph.D. in Electrical and Computer Engineering from University of Illinois at Urbana-Champaign (UIUC) in 2005 and 2006 respectively. He also received M.S. degree in Industrial Engineering from UIUC and B.S. in Manufacturing Automation from Northeastern University (NEU), P. R. China. He is a senior member of IEEE.

Ziyou Xiong received his B.S. degree from Wuhan University, Hubei Province, China, in July 1997. He received his M.S. degree in electrical and computer engineering from University of Wisconsin, Madison in December 1999 and Ph.D. degree in electrical and computer engineering from University of Illinois at Urbana Champaign (UIUC) in October 2004. He has also been a research assistant with the Image Formation and Processing Group of the Beckman Institute for Advanced Science and Technology at UIUC from January 2000 to August 2004. In the summers of 2003 and 2004, he worked on sports audiovisual analysis at Mitsubishi Electric Research Labs, Cambridge, Massachusetts. Since September 2004, he has been with the United Technologies Research Center as

a researcher/scientist in East Hartford, Connecticut. His current research interests include image and video analysis, video surveillance, computational audio-visual scene analysis, pattern recognition, machine learning and related applications.

Alan Finn is a Research Fellow at United Technologies Research Center (UTRC) in East Hartford, Connecticut, USA. He is responsible for technical leadership, innovation, and research quality in digital signal processing and video analytics. He has additional expertise in diagnostics and prognostics, embedded controls, embedded computer architecture, performance optimization, communications, fault-tolerance, and machine learning. Previously at UTRC he led the Estimation and Decision Group and has been a project leader and individual contributor. He has published 26 papers, holds 47 US patents, and has received 8 Outstanding Achievement Awards from UTRC. He received a Ph.D. in Electrical Engineering from Cornell University in 1983, an M. Eng. In Electrical Engineering from Cornell University in 1980 and two undergraduate degrees: a B.S. in Electrical Engineering and a B.S. in Mathematics from Rensselaer Polytechnic Institute, 1977.