Surveillance Face Super-Resolution via Shape Clustering and Subspace Learning

Lu Tao[1,2], Hu Ruimin[1,3], Han Zhen[1,3] , Xia Yang[1] and Gao Shang[1]

[1]*National Engineering Research Center on Multimedia Software, Wuhan University, Wuhan 430072, China*

[2]*Hubei Province Key Laboratory of Intelligent Robot, College of Computer Science and Engineering Wuhan Institute of Technology, Wuhan,430073,China*

[3]*School of Computer, Wuhan University, Wuhan 430072, China*

{lut,rmhu,zhenh,xiay,gaos}@whu.edu.cn

Abstract

In a learning-based super-resolution algorithm, suitable prior from the training database is a key issue. A novel face hallucination algorithm based on shape clustering and subspace learning for adaptive prior is proposed in this paper. We define face shape metrics with point distribution model by Hausdorff Distance, then a framework of adaptive prior and subspace learning is proposed to enhance the performance of surveillance face super-resolution. Linear regression is used to learn the relationship between low and high image systhesis coefficients. Experiments show that the face super-resolution algorithm based on shape classification can improve the subjective and objective quality of the input low-resolution face images and outperform many state-of-the-art global-based face super-resolution methods.

Keywords: *Surveillance, Super-resolution, Face shape, K-means clustering,, Subspace Learning*

1. Introduction

Super-resolution is widely used to enhance the details in low resolution images, and it is an effective tool in areas of criminal investigation, surveillance, remote sensing [1, 13, 14, 15] and so on. It can be especially useful in a surveillance system where a potential high-resolution image may be crucial for identification and further analysis. In recent years, many learning based super-resolution methods which have pairs of high and low resolution training samples are proposed, these methods have better performance than many interpolation methods due to the detailed face features that are learned from the samples.

The key issue of learning based super-resolution method is to use accurate prior for high-resolution image reconstruction. In order to use image prior from the samples, there are two ways: global approach and local patch based approach. For example, In [2], Wang and Tang propose a hallucination method using eigentransformation, Principle Component Analysis is used to fit the input face image as a linear combination of the low-resolution face images in the training set. Their hallucination algorithm is based on global sample faces. This method is robust to noise due to the global face prior, but it depends on face alignment and has ghosting effect on the face edges. To overcome this problem, patch based prior methods have been proposed, for example, locally linear embedding (LLE) based hallucination method is proposed in [3].

In this method, images are divided into overlapped patches, for each patch, the nearest neighbors are used as the prior to generate the target high resolution patches. Compared to global method, this approach has better subjective quality due to more smooth face image, but this method is not robust to noise. In [4], Liu proposes a global parameter and local non-parameter model for super-resolution, a patch-based non-parametric Markov network is used to model the residue between original low-resolution image and the reconstructed high-resolution image. This method combines the global prior and local prior together to get more satisfictory results. Motivated by the two step framework, Huang [5] proposes a new hallucination with canonical correlation analysis which maximizes the correlation between the local neighbor relationship of high and low resolution images. The samples in the training database are all used.

In order to use more accurate prior for super-resolution reconstruction, adaptive prior framework has been proposed. For example, Ogawa, et al., [6] propose an adaptive example-based super-resolution using kernel PCA with classification approach. In order to exploit intrinsic characteristics of the samples, training database are classified into different clusters, and for an input patch, use kernel PCA adaptively to find the most suitable clusters as the prior to reconstruct the output patch. While another position-based local hallucination method [7] uses the same position patch to infer the high-resolution patch. Dong [8] uses a framework of adaptive sparse domain selection with adaptive regularization. Different from previous sparse prior based method, in this method he divides the training database into many sub-training databases and learn the sparse representation dictionary of them, and for each input, most accurate prior can be adaptive selected as adaptive regularization. Most of global approaches impose the constraint that all of the training images should be globally similar, especially in real surveillance situation. The input images often suffer from strong noise and blur, and most of algorithms hardly give good results. On the other hand, local face methods are not robust to the real noise. In order to improve the robustness of surveillance face super-resolution methods, we propose a novel hallucination method that uses k-means cluster algorithm to classified the samples into different clusters, and adaptive prior by subspace learning can be used in high-resolution image reconstruction. As we know, in computer vision, shape is a very important attribute for object recognition. Compared to face texture, shape is more robust to noise. So in this paper, we use face shape as a robust intrinsic feature, then classify the samples into different clusters as adaptive prior.

The remainder of the paper is organized as follows: in Section 2, we describe face shape metrics based on Hausdoff distance; in Section 3, face super-resolution based on shape classification is introduced, and the last section presents the experiment.

2. Face Shape Metrics based on Hausdoff Distance

2.1 Points Distribution Model

In order to describe the shape, many models have been proposed, for example, curvature model, points model and math representation model. In1994, T. F. Cootes [9] proposed the active shape model (ASM), which have been widely used in feature point's location. So far, the speed and accuracy of ASM have made progress. As the same to the ASM, we describe the shape of the object by a set of feature points. For a calibrated image, the shape vector can be calibrated by each feature point coordinate, the coordinates of all feature points are arranged to form shape vector, and a sample database get shape vectors corresponding of faces. These shape vectors are through the translation and rotation and normalized to the same axis, the shape of the sample database has been set as:

$$L = \{s_i \mid i = 1, \cdots, n; s = (x_1, y_1, x_2, y_2, \cdots, x_m, y_m)^T\} \quad (1)$$

Here, n is the number of samples, m indicates the number of shape characteristic points, (x, y) is the coordinate of a feature point. As shown in Figure 1, we calibrate the location values of the face shape by hand. There are 62 feature points with the edge of the face contour to represent the face shape.

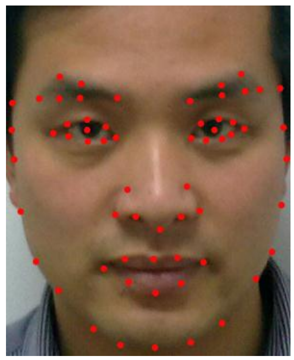

Figure 1. Face Shape Feature Points

2.2 Shape Metrics based on Hausdoff Distance

After calibrating shape feature points, we define face shape metrics on Hausdoff distance between shapes. Hausdorff distance is a measurement of two-point sets, and it is an effective similarity measurement for object matching, to more effectively characterize the similarity between the edge contours [10]. Without creating a relationship between two sets, it can be calculated in real-time. When the feature point extraction method used to extract the contour feature to calculate the Hausdorff distance, we can greatly reduce the computational complexity.

Given two point sets $A = \{a_1, a_2, ..., a_{Na}\}$ and $B = \{b_1, b_2, ..., b_{Nb}\}$, where N_a and N_b are the numbers of point sets respectively,the Hausdorff Distance (HD) between these two point sets is defined as:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (2)$$

In this equation,

$$h(A, B) = \max_{a \in A} \min_{b \in B} \| a - b \| = \max_{a \in A} D(a, B) \quad (3)$$

$$h(B, A) = \max_{b \in B} \min_{a \in A} \| b - a \| = \max_{b \in B} D(a, B) \quad (4)$$

3. Adaptive Prior Face Super-resolution based on Subspace Learning

In this paper, we propose a new hallucination method based on shape clustering and subspace learning. Firstly, we define point distribution model to describe the face shape, then k-means clustering method is used to cluster the training data into sub-training datas due to the face shape, secondly we use linear representation with global faces for a new input, and select most proper sub-training data with adaptive prior. At last, different from previous global methods, we adapt linear regression to build the high and low resolution representation coefficients. The framework is described in Figure 2.

3.1 Clustering of Training Database

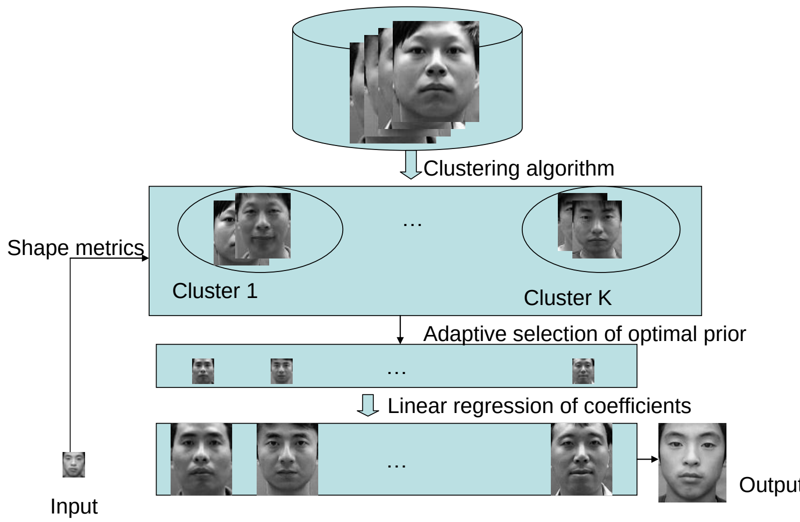

Figure 2. Outline of the Subspace Learning based Adaptive SR Algorithm

In this subsection, clustering of training database into K clusters is described. As the face shape is a robust intrinsic feature, we cluster the database in shape feature space. K-Means, a dynamic clustering algorithm, is a rather simple but well known algorithm for grouping objects and clustering. This method can display the most essential attribute in the initial recognition or classification of objects in the initial classification. As an unsupervised learning algorithm, the main idea is to define K centroids. These centroids are gat by iteration, the purpose of clustering is to minimize the difference in-cluster and maximize the difference between classes. We use the algorithm in [11], details are as follows:

Step 1: Define K points as initial group centroids.

n is the number of faces in training database, K is the number of clusters , c_j is the centroid of each cluster.

Step 2: Assign all faces to the cluster that has the closest centroid.

Step 3: Recalculate the positions of K centroids

Step 4: Repeat step 2 and step 3 until the centroids no longer move.

In order to obtain the optimal classification, the membership needs to be optimized, the objective function is constructed as follows:

$$J = \sum_{j=1}^{K} \sum_{i=1}^{n} \| l_i^{(j)} - c_j \|^2 \qquad (5)$$

Where $\| l_i^{(j)} - c_j \|$ is the Hausdoff Distance between input shape and cluster centers.

In this paper, we use Chinese face database CAS-PEAL-R1 [12] from Institute of Computing Technology, Chinese Academy of Sciences. It contains 30 863 images of 1040 subjects. These images belong to two main subsets: the frontal and non-frontal subsets. Here we choose 1040 frontal and no-expression faces to compose the database. After getting the feature points of the face shape by hand, the size of the images is 96×112 pixels with alignment. When $K = 5$, centroids are listed below in Table 1.

Table 1. Centroid of Each Cluster

Category	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Quantity	315	204	115	167	199
Centroid					

3.2 Subspace Learning based Super-resolution

The face shape classification is applied to the face super-resolution algorithm in this paper. We first use the input low-resolution images of the contour shape to get the face feature points, and then classify the sample database. These classifications of the sample database are applied as a priori information on face image reconstruction.

For each input low-resolution face image, we first use face alignment algorithm by the distance between human two eyes. Then we use ASM algorithm to get feature point's location which is shape vector coordinate I_s. By affine transformation, we obtain a new shape vector $I_s^{'}$ with the same resolution of the samples. In the previous section, the face database is divided into five categories, as the center of each cluster is clear now, we select the face shape database $t = \{\mathrm{i}, | \min\{H(I_s^{'}, c_i), i = 1 \cdots K\}\}$, so that in face shape sample database t, we select K minimum Hausdorff Distances with the input shape, these faces make up the sample face database, where K is less than or equal to the number of face images.

Let $N \times K$ matrix $[f_1, f_2 \cdots f_K]$ represents the certain sample database, where N is the image pixel, K is the number of samples. Calculating average face of the samples:

$$\bar{m} = \frac{1}{K}\sum_{i=1}^{K} f_i \qquad (6)$$

Removing the mean face from each image, we have

$$L = [f_1 - \bar{m}, \cdots f_K - \bar{m}] \qquad (7)$$

The ensemble covariance matrix $C = LL^T$. Directly computing the eigenvectors of C is not practical because of the large size of the matrix. Alternatively, the eigenvectors of a smaller matrix $L^T L$ can be computed easily.

$$(L^T L)V = V\Lambda \qquad (8)$$

Here V is the eigenvector matrix and Λ is the eigenvalue matrix. The orthonormal eigenvector matrix of C is as follows:

$$E = LV\Lambda^{-\frac{1}{2}} \qquad (9)$$

For any face I , we have:

$$I = LV\Lambda^{-\frac{1}{2}}w + \bar{m} \qquad (10)$$

Where a weight vector w can be computed by projecting the face onto the eigenfaces.

We assume that the weight synthetic coefficient in low resolution samples is the same as the high ones, so from the low-resolution face synthetic coefficients, high-resolution face images can be inferred. Let $b = V\Lambda^{-\frac{1}{2}}w$.

$$I = Lb + \bar{m} = \sum_{i=1}^{K} b_i f_i + \bar{m} \qquad (14)$$

3.3 Coefficients Transformation by Linear Regression

Many methods [2, 4] assume that the Principal Component Analysis (PCA) coefficients of low-resolution and high-resolution are consistent, so the weights of low-resolution are transformed into high-resolution space directly. We use linear regression to learn the mapping between the low and high PCA coefficients. In Eq(14), for the each training face can express as weights b , we can get all training LR weights B_L and HR weights B_H . If matrix A represents the transformation form LR coefficients to HR ones, then we have:

$$B_H = AB_L \quad (15)$$

This equation can easily get a least square solution as $A = B_H B_L^T (B_L B_L^T)^{-1}$. For an input low resolution image, the corresponding high resolution is

$$X = \sum_{i=1}^{K} Ab_i F_i + \bar{M} \qquad (16)$$

Where X is the high resolution image, $F_i, \bar{M}$ are high resolution training database and the mean face respectively. Optimizing X by smoothing and denoising, the output quality can be improved.

4. Experiments

The hallucination experiment is conducted on different databases. From the above section, all faces in the database are classified into 4 categories. For each face we choose different database for super-resolution. In Chinese face database CAS-PEAL-R1, there are 1040 individuals with one front face image for each individual. We use 1000 images for clustering, the rest 40 images as test images. The high resolution image size is fixed at 96×112 pixels. All images are blurred and downsampled to low resolution images of 24×28 pixels. For each input face image, we use shape similarity to classify the input images to certain category, and then adaptively select corresponding database for super-resolution reconstruction. Compared with the input image and nearest interpolation, our results provide more details, as show in Figure 3.

In this paper, for fair competition, we set all sub-training databases to be the same number, 98% principal components are preserved in each method. In the coefficients regression training, we use other cluster images as training database. We compare our method with Liu' method, Wang' method and interpolation-based method, our results have higher subjective image quality. As shown in Figure 3.

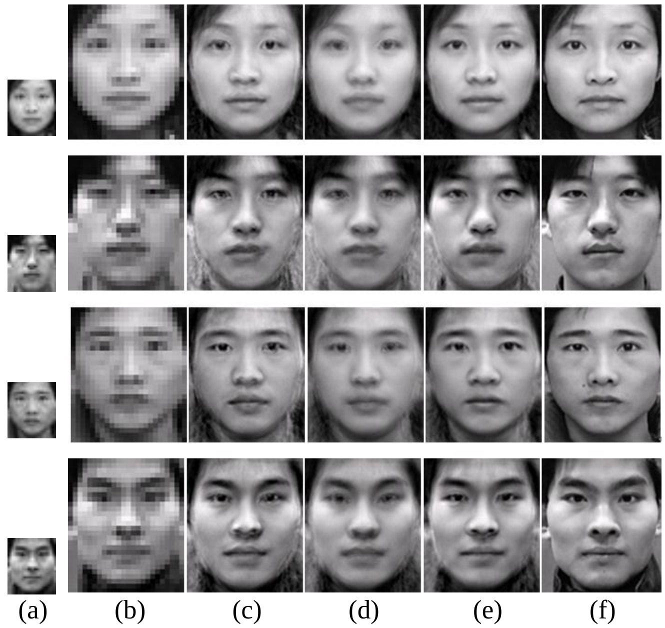

Figure 3. Face Hallucination with Different Methods. (a) The input low resolution face (b) Nearest interpolation (c) Hallucination based on eigentransformation [2] (d) Hallucination by two steps [4] (e) Hallucination by adaptive prior (e) Original high-resolution images

From the above, the algorithm in this paper can use the shape information as clustering feature, and the shape features are robust to noise. On the other hand, our adaptive prior selection framework can reduce the edge blur. Different from the compared methods, the adaptive prior selection and coefficients regression can improve the performance of our algorithm in real situation.

On the other hand, to quantify the deviations of the results from the ground truth, we also compared the PSNR of each method when testing with 4 face images in Figure 4.

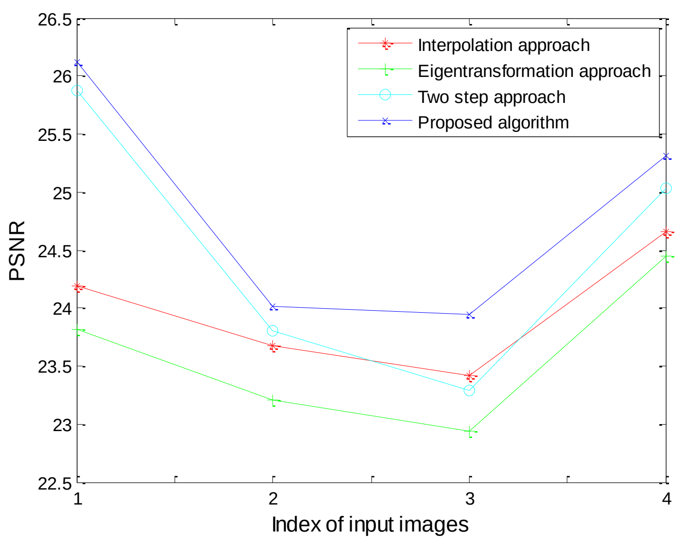

Figure 4. PSNR of Different Face Super-resolution Methods

(a) (b) (c) (d) (e)

Figure 5. Super-resolution for Real Surveillance Face Images

As shown in Figure 4, for the horizontal axis, we choose 4 different types of faces as input images, and the vertical axis is value of PSNR. From the Figure 2, we can know that our method get better objective quality than other methods, such as Wang's method and Interpolation method, comparing to Wang's method, the proposed method can improve 0.45db.

To further verify the efficiency of our method, we perform the approach on real images. As show in Figure 5, we get image (a) from the real video surveillance system, and (b) is cubic interpolation, (c) is Wang's method (d) is Liu' method (e) is our result. We can notice that our method have better details than other methods.

5. Conclusion

In this paper, we use the shape intrinsic feature as clustering metric, then framework of adaptive prior selection and PCA coefficient regression has been proposed. Through experimental results, we can know that face super-resolution algorithm based on the classification of face shape reconstruction can improve the quality of the reconstructed images; in particular, some of the face edges have been significantly improved than some other methods. On the other hand, face shapes as an important attribute play an important role for the super-resolution reconstruction. So how to maintain the shape in the reconstruction process is a problem worth to explore in future.

Acknowledgements

This paper was supported by the major national science and technology special projects (2010ZX03004-003-03), the National Natural Science Foundation of China (61172173, 60832002, 61070080, 60970160, 61003184), the Natural Science Foundation of Hubei Province of China (2009CDA134, 2010CDB05103), the Fundamental Research Funds for the Central Universities (201121102020009, 3101011), the Hubei Province Key Laboratory of Intelligent Robot(HBRI 200907), the youths science foundation of Wuhan institute of technology, the excellent youth science and technology innovation team project of the Educational Department of Hubei Province of China (No. T201206).

References

[1] T. Jing and K. K. Ma, "A survey on super-resolution imaging", Signal, Image and Video Processing, **(2011)**, pp. 1-14.
[2] X. G. Wang and X. O. Tang, "Hallucinating Face by Eigentransformation", IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, vol. 35, no. 3, **(2005)**, pp. 425-434.
[3] H. Chang, D. Yeung and Y. Xiong, "Super-resolution through neighbor embedding", Proceedings of IEEE Conference of Computer Vision and Pattern Recognition, **(2004)**, Washington, DC, USA, pp. 275-282.
[4] C. Liu, Shum and H. Y. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **(2001)**, Kauai, HI, USA.
[5] H. Huang, H. T. He, X. Fan and J. P. Zhang, "Super-resolution of human face image using canonical correlation analysis", Pattern Recognition, vol. 43, **(2010)**, pp. 2532–2543.
[6] T. Ogawa and M. Haseyama, "Adaptive example-based super-resolution using kernel PCA with a novel classification approach", EURASIP Journal on Advances in Signal Processing, vol. 1, **(2011)**.
[7] X. Ma, J. Zhang and C. Qi, "Hallucinating face by position-patch", Pattern Recognition, vol. 43, no. 6, **(2010)**, pp. pp. 3178–3194.
[8] W. S. Dong, L. Zhang, G. M. Shi and X. L. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization", IEEE Ttransactions on Image Processing, vol. 20, no. 6, **(2011)**, pp. 1838-1857.
[9] Cootes, T. F. Hill, A. Taylor and C. J. Haslam, "Use of active shape models for locating structures in medical images", Image and vision computing, vol. 12, (1994), pp. 355-356.
[10] D. P. Huttenlocher, G. A. Klanderman and W. A. Rucklidge, "Comparing images using the Hausdorff distance", IEEE Transactions on pattern analysis and machine intelligence, vol. 15, no. 9, **(1993)**, pp. 850-860.
[11] S. N. Sulaiman, "Adaptive fuzzy-k-means clustering algorithm for image segmentation", IEEE Ttransactions on Consumer Electronics, vol. 56, no. 4, **(2010)**, pp. 2661-2668.
[12] W. Gao, B. Cao and S. G. Shan, "The CASPEAL Large-scale Chinese face database and baseline evaluations", IEEE Trans on Systems, Man, and Cybernetics, Part A: Systems and Humans, vol. 38, no. 1, **(2008)**, pp. 149-161.
[13] A. P. Anil and S. H.Jyoti, "Discrete Curvelet Transform Based Super-resolution using Sub-pixel Image Registration", International Journal of Signal and Image Processing, vol. 4, **(2011)**, pp. 41-50.
[14] K. Rama and L. Reddy, "Face Recognition Based on Eigen Features of Multi Scaled Face Components and Artificial Neural Network", International Journal of Security and Its Applications, vol. 5, **(2011)**, pp. 23-44.
[15] D. R. Kisku, P. Gupta and J. K. Sing, "Face Recognition using SIFT Descriptor under Multiple Paradigms of Graph Similarity Constraints", International Journal of Multimedia and Ubiquitous Engineering, vol. 5, no. 4, **(2010)**, pp. 1-8.