International Journal of Recent Engineering Research and Development (IJRERD)
Volume No. 02 – Issue No. 05, ISSN: 2455-8761
www.ijrerd.com, PP. 35-40

Pose-Invariant Face Recognition on Video with Image Super-Resolution

Krishnasree P.A
M Tech student,Dept.of CSE
Thejus Engineering College, Vellarakkad, Kerala, India

Abstract: TheVideo-based face recognition has received major consideration in the past few years. However, the facial images in a video sequence attained from a distance are normally small in size and have low visual quality. The small size images make the recognition task difficult in real world applications and affect the accuracy of face tracking. Also it usually contains significant pose variation, which significantly reduces the performance of frontal face recognition. Here proposes a face identification framework capable of handling the full range of pose variations and uses learning-based single image super resolution(SISR) method to obtain a high resolution (HR) face image from a single given low resolution (LR) face image. This system first extracts the face image from the video sequence and perform super resolution. Then converts the problem of pose-invariant face recognition into a partial frontal face recognition problem. The feature transformation degrades recognition ability by transforming the features from the gallery and probe images to a general discriminative subspace. Finally, face matching is performed.

Keywords: Pose-Invariant Face Recognition, Super-Resolution, Single Image Super-Resolution

1. Introduction

Human face gives a completely unique feature of human being, and recognizing a person by this feature appears to be an easy task for the human brain. Automatically identifyinghuman faces, however, has come to be an important issue in the last two decades for numerous real life applications. Face recognition (FR) is the process of automatically identifying or verifying a person from his facial image. The most common approach to do that is by comparing a given image of an unknown individual called 'probe image' with a large set of images 'gallery 'of suspected individuals, and match it to the most similar image from the gallery. Human face recognition constitutes a very active area of research recently because of two main reasons. The primary reason is the extensivevariety of commercial and law enforcement applications, and the second reasons is that the trouble of machinerecognition of human faces has been attracting researches from different disciplines such as, image processing, neural network, computer vision, pattern recognition, and psychology. Although there are many reliable methods of biometric human identification such as, retinal or iris scans, and fingerprint analysis, these systems need cooperation between the person and the identification system for reliable decisions. Also, the database of this type of recognition systems is difficult to effort. Therefore, face recognition is thought to have great potential to make the automatic recognition of human more convenient.

In real life application where the images might be taken in uncontrolled environment, machine face recognition encounters great difficulties. Many FR challenges have been addressed, such as, poor illumination, pose variation, and sever expression variation. However, in many real-world applications, such as, video surveillance cameras, where it is often difficult to obtain good quality recordings of observed human faces, face recognition is still challenging task. One important factor evaluating the quality of the recordings is the resolution of the images which is usually much lower than the resolution that is needed for the typical face recognition. Therefore, addressing the performance of FR systems when image resolution decreases, as well as attempting to increase theresolution of low-resolution (LR) images is becoming a very crucial demand.

2. Literature survey

In this section, firstly the main problems of face recognition across pose variations are described and several recent studies on pose-invariant face recognition are introduced. Then super-resolution methods for both visual enhancement and face recognition are discussed.

2.1 Pose-Invariant Face Recognition

Existing methods for face recognition across pose can be roughly divided into two broad categories:(1) techniques that rely on 3D models and (2) 2D techniques.

Tae-Kyun Kim and Josef Kittler [3] have advised that the problem of pose-invariant facerecognition centred on a single model image. Face images at a arbitrary pose can be mapped toa orientation pose by the model yielding view-invariant representation. Such a model typicallybe subject to thick correspondences of

different view face images, which stayed hard to establish in training. Errors in the correspondences completely harm the accuracy of any recognizer.

Hui-Fuang Ng [4] proposed that the Face recognition security frameworks had turn outto be necessary for many submissions such as automatic access control and video surveillance.Most face recognition security frameworks required right frontal views of a person, and thoseframeworks was failed if the individual to be expected does not face the camera properly. Intheir paper, they suggest a novel approach for robust pose-invariant face recognition. Theirrecognition system was indifferent to viewing information and it requires just a single sampleview per individual. Their technique make use of the evaluations of a face image against a setof faces from a training set at the same view to introduce pose-invariant representations of anindividual in various poses.

Hyung-Soo Lee and Daijin Kim [5] have offered to recognize human faces were one ofthe most important areas of research in biometrics. Their papers propose generating frontalview face image using linear transformation in feature space for face recognition. They extractfeatures from a posed face image using the kernel PCA. After that, they make over the posedface image into its equivalent frontal face image using the transformation matrix determined bylearning. Then, the generated frontal face image was recognized by three different intolerancemethods such as LDA, NDA, or GDA.

Eidenberger [6] have described that novel algorithm for the recognition of faces from image samples. Their algorithm used the Kalman filter to identify considerable facial character.Their papers discuss Kalman faces extraction, application; tunable parameters, experimentalresults and related work on Kalman filter application in face recognition.

Ting Shan et al. [7] have described that majority face recognition systems only work wellunder quite constrain environment. In particular, the illumination conditions, facial expressionsand head pose must be strongly embarrassed for good recognition performance. They developeda face model and a rotation model which can be used to take facial features and create realisticfrontal face images when given a single novel face image. They use a Viola-Jones based facedetector to detect the face in real-time and thus solve the initialization problem for our ActiveAppearance Model search.

Hongzhou Zhang et al. [8] have projected that face recognition has varied applicationsparticularly as an identification solution which can meet the lament needs in security areas. Appearancebased approach was proposed. Face recognition was implemented by reconstructingfrontal view features using linear transformation.

2.2 Image Super-Resolution

Although wide variety of super-resolution literature is available, it is still an open topic toinvestigate. Following subsections describe some of the existing basic image super-resolutionschemes.

Jianchao Yang et al.[9] considered the sparse signal representation of an image. Based onprevious research on image statistics the image patches can be well-represented as a sparse linearcombination of elements from an appropriately chosen over-complete dictionary. Motivatedby this, they proposed a sparse representation for each patch of the low-resolution input. Thecoefficients of this representation are used to generate the high resolution output. Theoreticalresults from compressed sensing suggest that the sparse representation can be correctly recoveredfrom the down-sampled signals under mild conditions. Similarity of sparse representationsbetween the low-resolution and high-resolution image patch pair with respect to their own dictionariesis enforced, by jointly training two dictionaries for the low and high-resolution imagepatches. So, the sparse representation of a low-resolution image patch is being applied with thehigh-resolution image patch dictionary to generate a high-resolution image patch. They showedthe effectiveness of such a sparsity prior for both general image super-resolution (SR) and thespecial case of face hallucination. This algorithm can handle SR with noisy inputs in a moreunified framework because the local sparse modeling is naturally robust to noise.

Xiao Zeng and Hua Huang[10] presented a regression basedmethod that can successfully recognize the identity given all these difficulties. They built a radialbasis function in subspace by canonical correlation analysis to nonlinear regression modelsfrom the specific nonfrontal low resolution image to frontal high resolution features.

A. Maalouf and M. C. Larabi [11] proposed the idea of generating a super-resolution (SR)image from a single multi-valued low-resolution (LR) input image. This problem approachesfrom the perspective of image geometry-oriented interpolation. They computed the grouplettransform to obtain geometry of the LR image. Geometric grouplets is constructed by orthogonalmultiscale grouping with weighted Haar lifting to points grouped by association fields. Topreserve the sharpness of edges and textures SR image is synthesised by an adaptive directionalinterpolation using the extracted geometric information. This method showed improvementsover existing geometrically driven interpolation techniques on a subjective scale, and in manycases with an improvement in psychovisualcolor difference.

J Tatem Andrew et al. [12]used their idea of super-resolution for target identification in remotely sensed images. Fuzzy classification improves the accuracy of land cover target identificationmake robust and better for spatial representation of land cover. The Hopfield neuralnetwork converges to a minimum of an energy function, defined as a goal and several constraints.The energy minimum represents a best guess map of the spatial distribution of classcomponents in each pixel. They used two goal functions to make the output of a neuron similarto that of its neighboring neurons. The first goal function aims to increase the output of centerneuron to 1. The second goal function aims to decrease the output of the center neuron to 0.They showed that, by using a Hopfield neural network, more accurate measures of land covertargets can be obtained compared with those determined using the proportion images alone.By the results, the Hopfield neural network used in this way represents a simple, robust, andefficient technique, and suggest that it is a useful tool for identifying land cover targets fromremotely sensed imagery at the subpixel scale.

Yizhen Huang and Yangjing Long [13] proposed a optimal recovery based neural-network Super-Resolution algorithm. This method evaluated on classical SR test images with both genericand specialized training sets, and compared with other state-of-the-art methods. Motivated bythe idea that back propagation neural network are capable of learning complex nonlinear functionthey proposed a neural network approach that produces better results in high-frequencyregions. They integrated an optimal recovery based approach with in a neural network frameworkand, if so, two different branches of algorithms complement each other to offer a betteralgorithm. Using this algorithm in a two-pass way generates visual results that are very similarregardless of the initial interpolation step, and more times of iteration only waste the computingresource but yield negligible performance gain.

3. Proposed System

3.1 System Architecture

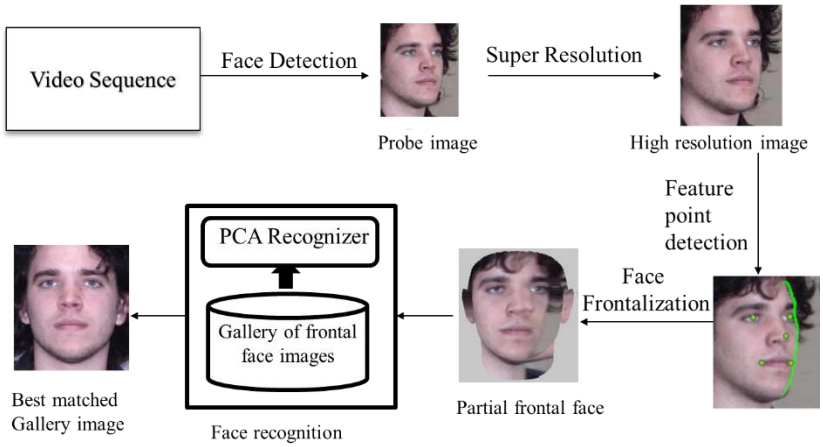

Figure 1: Proposed System Architecture

The Figure 1 shows the proposed architecture. It consist of several steps .The first step is image acquisition. Image acquisition can be accomplished by digitally scanning an existing video or by using an electro-optical camera to acquire a live video of a subject. Next, face detection is used to find the candidate face region. The system can operate on static images, where this procedure is called face localization and dealing with videos,
procedure is called face tracking. Face tracking techniques include head tracking, where the head is viewed as a rigid object performing translations and rotations. While the tracking module finds the exact position of facial features in the current frame based on an estimate of face or feature locations in the previous frame(s).Then we get a set of face images. However, the facial images in a video sequence acquired from a distance are usually small in size and their visual quality is low. The small size images make the recognition task difficult. So super resolution technique can be used to estimate high resolution image from a low resolution image. For this learning based single image super resolution technique can be used.

The final super-resolved probe face image is used as the identity of the gallery. The next step is pose estimation. The pose estimator classifies a face pattern into one of the view (pose) groups. The pose estimate is used to verify the detection of multi-view faces. If the detected face image is a profile view, the next step is

frontal face reconstruction. For this, first detect the facial feature points and using this feature a partial frontal face is reconstructed. This is used for recognition. The face recognizer matches the faces in the input and those known faces in the database, and outputs the identity information of the seen faces. Here principal component analysis is used as a face recognizer.

3.2 Methodology

This system involves 4 main phases:

- Face Detection
- Super-Resolution
- Face Frontalization
- Face Recognition

3.2.1 Face Detection

Face detection is used to find the candidate face region. Face detection based on videos consists of generally three main processes.

- Detect a Face to Track: Before begin tracking a face, first need to detect it. vision. Cascade Object Detector detect the location of a face in a video frame. The cascade object detector makes use of the Viola-Jones detection algorithm and a trained classification model for detection. By default, the detector is configured to detect faces, but it can be configured for other object types. You can use the cascade object detector to track a face across successive video frames. However, when the face tilts or the person turns their head, you may lose tracking. This limitation is due to the type of trained classification model used for detection. To avoid this issue, and because performing face detection for every video frame is computationally intensive.
- Identify Facial Features to Track:Once the face is located in the video, the next step is to identify a feature that will help to track the face. Choose a feature that is unique to the object and remains invariant even when the object moves.
- Track the Face with the skin tone selected as the feature to track, you can now use the vision. Histogram Based Tracker for tracking the face image.

3.2.2 Super-Resolution

The system uses learning based single image super resolution method to obtain a high resolution (HR) image from a single given low resolution (LR) image. SISR is a highly local problem led to the following ideas[2]:

- Use a larger input LR patch than the one that maps to the HR patch being estimated: To ensure accurate estimation of the HR patch that gives an LR patch upon downsampling, simply using this parent LR patch is not enough of an input because many combinations of pixel values of the HR patch can be downsampled to give the same parent LR patch.

- Use the smallest input size where SR performance saturates: We expected the SR performance to saturate with increasing n for a given m due to the highly local nature of SISR problem. Limiting LR input to this point of saturation will lead to more efficient learning.
- Use the smallest possible output size that gives good SR performance. Since the dimension of the input required for accurate estimation of the output depends on the output dimension, the learning efficiency (time and sample requirement) can be increased by keeping the output dimension to a minimum.

(a)ZCA Whitening

ZCA whitening is a linear transform that amplifies visually salient high-frequency components from images by normalizing the variance of the data in the direction of each eigenvectorof its covariance matrix. It returns vectors of the same size as the input vectors. Before ZCA whitening, the HR images can be downsampled to get

the same LR image.It is performed on both vectorized LR and HR images. For recreating an image, ZCA transform can be inverted by using an inverse of the saved ZCA rotation matrix, and adding back the saved patch means. ZCA can be applied to K vectorized LR patches as follows:

- Compute and save the mean of a vectorized LR patch
- Compute mean centered vector by subtracting mean from each element of vector
- Arrange all K mean centred vectors in columns and compute the covariance
- Compute eigen values and eigen vectors
- Compute and save the ZCA rotation matrix that whitens the data by redistributing the energy of eigenvalues
- Return transformed data matrix, rotation matrix, mean vector

(b) Training

PNNs are usually trained using group method of data handling (GMDH) algorithm, which attempts to obtain a hierarchically more complex and accurate estimation of the desired mapping g(y). The training step is as follows:

- For each pixels,construct the first layer
- Randomly split the data set into training and validation sets
- Add another layer where each neuron is connected to two outputs of previous layer
- For each neuron in layer

  – assume transfer function as a polynomial of order two

$$P = w_0 + w_1 a_i + w_2 a_j + w_3 a_i a_j + w_4 a_i^2 + w_5 a_j^2$$

  – Train: Estimate its weights using least squares

  – Validate: Compute the mean squared error

  – If MSE > e then prune neuron else update the error

  – Use the output of layer to create a new local copy of input matrix

- Choose the best unit having minimum MSE

(c) Testing

The trained PNNs, each having n x n LR pixel input and one HR pixel output, are applied to test images as follows:

- Extract vectorised patches from given test LR image with one pixel overlap, apply ZCA whitening using the saved rotation matrices, and use ZCA whitened vectorized patches to form the input matrix for the PNNs.
- Compute the desired child HR pixels using vectorized patches as input to the trained PNN
- Invert the ZCA whitening, re-arrange vectorized HR pixels at their respective locations, and return the reconstructed HR image.

3.2.3 Face Frontalization

Face frontalization composed of three steps: face pose normalization, unoccluded facial texture detection, and patch-wise feature extraction.

A standard 3D method is adopted for face pose normalization[1]. The five most stable facial feature points, i.e., the centers of both eyes, the tip of the nose, and the two mouth corners, are first detected automatically or manually. For profile faces, the coordinates of the occluded facial feature points are estimated. Using the orthographic projection model and the detected five facial feature points, a 3D generic shape model is aligned to the 2D face image.The 2D face image is then back-projected to the 3D model, and a frontal face image is rendered with the textured 3D model. The normalized face image is first divided into M x N overlapped patches.

Next, each of the unoccluded patches is split into J x J cells. Principal Component Analysis (PCA) is applied to each patch to project its feature into a subspace with dimension D, by which the noise is blocked.The set of patch-level DCP features following PCA processing from all unoccluded patches forms the representation of the face image.

3.2.4 Face Recognition

In this step face matching can be performed based on the Euclidean distance. It is calculated at patch level. The image with minimum Euclidean distance is returned as identified image.

4. Conclusion

Face recognition across pose is a challenging task because of the significant appearancechange caused by pose variations.In this project, the performance of face recognition using images with different poses and different resolutions is evaluated. For experiments on the video images in which the LR images are generated by down- sampling original images, the recognition rate is steady for images with resolution higher than a certain number while the accuracy declines suddenly when image resolution is lower. Obviously, the important information for classification is missing when images are so small. For surveillance image database, low resolution is not the only problem of the images captured at a distance. Other problems, take pose variation for example, also influence the performance. The recognition rates are all very low. The learning-based single image super-resolution method is used for improving the low resolution face recognition performance. This method require a training set containing HR and LR image pairs to learn a relationship or mapping from LR feature space to HR feature space, and the SR images or features are reconstructed by this relationship or mapping. In our experiments, the images were aligned using the eye coordinates, but the scale of the faces were different and nothing was done to deal with the pose problem. Better methods to align the images is expected to improve the recognition performance.

References

[1] Changxing Ding, Chang Xu, and DachengTao.”Multi-task pose-invariant face recognition”. *IEEE Trans. Image Process.* 24, 3 (2015), 980--993.

[2] Neeraj Kumar and Amit Sethi,“Fast Learning-Based Single Image Super-Resolution”,IEEE Trans Pattern Anal. Mach. Intell., vol. 25, no. 9, pp. 1063-1074,2015

[3] T.K Kim and J.Kittler,“Design and fusion of pose-invariant faceidentificationexperts, ”*IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 16, No. 9, pp. 1096-1106, September 2006.

[4] H. F.Ng, “Pose-invariant face recognition security system, ”Asian Journal of Health and Information Sciences, Vol. 1, No. 1, pp. 101-111, 2006.

[5] H. S. Lee and D. Kim, “Generating frontal view face image for pose invariant face recognition, ”Pattern Recognition Letters, Vol.27, pp.747-754, 2006.

[6] H.Eidenberger,“Kalman filtering for pose-invariant face recognition, ”In Proc. of the IEEE International Conference on Image Processing, .Atlanda,pp.2037-2040, 2006.

[7] T. Shan, B. C. Lovell, and S. Chen, “Facerecognitionrobust to head pose fromonesampleimage, ”In Proc. of the 18th International ConferenceonPatternRecognition, (ICPR 2006), pp.515-518, 2006.

[8] H. Zhang, Y. Li, L. Wang, and C. Wang, “Pose insensitivefacerecognitionusingfeaturetransformation, ”IJCSNS International Journal of ComputerScience and Network Security, Vol.7 No.2, February 2007.

[9] JianchaoYang,John Wright and Thomas S. Huang,Yi Ma. Image Super-Resolution Via Sparse Representation. In IEEE Transaction on Image Processing, 19(11):2861-2873, November 2010.

[10] Xiao Zeng and Hua Huang. Super-Resolution Method for Multiview Face Recognition From a Single Image Per Person Using Nonlinear Mappings on Coherent Features. In IEEE Signal Processing Letters, 19(4):195-198, April 2012.

[11] A. Maalouf and M.C. Larabi. Colour image super-resolution using geometric grouplets. In IET Image Processing, 6(2):168-180, 2012.

[12] Andrew J. Tatem, Hugh G. Lewis, Peter M. Atkinson and Mark S. Nixon. Super-Resolution Target Identification from Remotely Sensed Images Using a Hopfield Neural Network. In IEEE Transaction on Geoscience and Remote Sensing, 39(4):781-796, April 2001.

[13] Yizhen Huang and Yangjing Long. Super-resolution using neural networks based on the optimal recovery theory. In Springer Journal Computational Electronic, 5:275-281,2006.