[Graduate Research in Engineering and Technology (GRET)](https://www.interscience.in/gret)

Volume 1 | Issue 2 Article 14

October 2013

AN EFFICIENT METHOD TO FEED HIGH RESOLUTION IMAGES TO FACIAL ANALYSIS SYSTEMS

ROOPA R
CSE, Acharya Institute of Technology, Banglore., roopar@gmail.com

MRS. VANI.K .S
CSE, Acharya Institute of Technology, Banglore., vaniks@gmail.com

Recommended Citation
R, ROOPA and .S, MRS. VANI.K (2013) "AN EFFICIENT METHOD TO FEED HIGH RESOLUTION IMAGES TO FACIAL ANALYSIS SYSTEMS," *Graduate Research in Engineering and Technology (GRET)*: Vol. 1 : Iss. 2 , Article 14.
DOI: 10.47893/GRET.2013.1032
Available at: https://www.interscience.in/gret/vol1/iss2/14

AN EFFICIENT METHOD TO FEED HIGH RESOLUTION IMAGES TO FACIAL ANALYSIS SYSTEMS

ROOPA R[1], MRS. VANI.K.S[2], MRS. NAGAVENI.V[3]

[1]Student, M-tech, [2,3]Asst.Professor, CSE, Acharya Institute of Technology, Banglore.

Abstract- Image Processing is any form of signal processing for which the image is an input such as a photograph or video frame. The output of image processing may be either an image or a set of characteristics or parameters related to the image. In many facial analysis systems like Face Recognition face is used as an important biometric. Facial analysis systems need High Resolution images for their processing. The video obtained from inexpensive surveillance cameras are of poor quality. Processing of poor quality images leads to unexpected results. To detect face images from a video captured by inexpensive surveillance cameras, we will use AdaBoost algorithm. If we feed those detected face images having low resolution and low quality to face recognition systems they will produce some unstable and erroneous results. Because these systems have problem working with low resolution images. Hence we need a method to bridge the gap between on one hand low-resolution and low-quality images and on the other hand facial analysis systems. Our approach is to use a Reconstruction Based Super Resolution method. In Reconstruction Based Super Resolution method we will generate a face-log containing images of similar frontal faces of the highest possible quality using head pose estimation technique. Then, we use a Learning Based Super-Resolution algorithm applied to the result of the reconstruction-based part to improve the quality by another factor of two. Hence the total system quality factor will be improved by four.

Index Terms- *face-log generation, super-resolution, surveillance video.*

I. INTRODUCTION

In recent technologies, the Face is used as one of the most important remote biometrics and in many facial analysis systems, like face recognition, human–computer interaction, and so on, it is widely employed. One of the real-world challenges of these systems is that they have problem working with low-resolution (LR) images. This is the reason that for example surveillance applications in public places like airports need human operators to identify suspected people. Therefore, having an automated system working with LR and low-quality face images is desirable. However, low-quality images do not have enough high-resolution (HR) details for facial analysis systems and using them directly in these systems is not reliable. Thus, there is a need for a mechanism for bridging this gap between LR images and facial analysis systems. Super-resolution (SR) is one of such mechanisms for obtaining an HR image from one or more LR input images. SR algorithms are broadly classified into two classes: reconstruction-based SR (RBSR, known as classical multiframe SR) and learning-based SR (LBSR, known as hallucination, or example based SR).

RBSR algorithms usually work with more than one LR input image. These LR inputs must have intra-image subpixel misalignments. The algorithm uses these misalignments to reconstruct the missing HR details of the inputs. The improvement factor of the RBSR algorithms is practically limited to factors close to two.

LBSR algorithms mostly learn the relationship between some HR training images and their corresponding LR versions. This system works with the pixel values of the LR inputs. Unlike RBSR algorithms there is no guarantee that the hallucinated HR images produced by LBSR algorithms provide true HR details .

In this paper, we deal with the real world problems of a SR system working with faces coming from a surveillance video sequence. Such a system has several problems: the slight-motions restriction of the objects (which becomes harder to control for longer video sequences), the ill-posed and ill-conditioned nature of the HR response, slow convergence of the system, and small magnification factors (usually two).

The block diagram of our system is shown in Fig. 1. In the first block of the system, face-log generation, the input video sequence is summarized to one [up to three face-log(s)].

The images inside the face-logs are very similar to each other and are of better quality compared to the other images of the sequence. It means they are good inputs to the next block of the system, which is a cascaded SR. In this block, an RBSR is applied to the generated face-log(s) and produces an HR image.

The quality of this HR image is improved by a factor of two compared to the LR images in the video sequence. This HR image is then fed to an LBSR to improve it even more. The rest of this paper is organized as follows. Face-log generation is described in the next section. Section III explains the details of the employed SR algorithms and finally the conclusions future trends are given in Section IV.

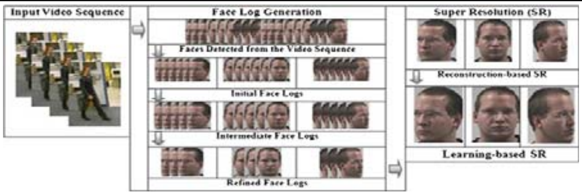

Fig. 1

II. FACE LOG GENERATION

In this section, we first explain the idea of face-log generation. Then, all the required processes for producing a facelog(s) from a video sequence are described. These processes are: face detection, facial feature extraction, and face quality assessment (FQA).

A. Face-Log(s)
Face-logs are considered as a concise and/or complete representation of a video sequence. The content of face-logs depends on the application. The proposed system in this paper uses one to three face logs. The most important one is a face-log containing frontal and semi-frontal face images. The other two face-logs are associated with two side-view face images of the subject. For constructing these three face-logs, we first use a head-pose estimation method that is developed for LR images. Based on the value of this feature, we classify face images of the input video sequence into three face-logs containing frontal, left side-view, and right side-view face images. These face logs are denoted as initial face-logs. Each of these initial face logs goes into the same computations, separately. Thus, we describe the rest of this paper only for the frontal face-log.

B. Face Detection
We use the AdaBoost idea of employing Haar-like features of the integral images of the input video sequence to detect face images.

C. Facial Feature Extraction
Extracting facial features in LR images is difficult. However, we use some simple methods that produce rough estimates of the facial features. These rough estimates fulfill our needs because what we need is to compare different LR face images of a person against each other rather than having exact numerical values for their facial features. We have used four facial features: head-pose, sharpness, brightness, and resolution. We mainly concentrate on the head-pose estimation. Head-Pose Estimation: People in a typical surveillance video sequence move freely around the camera and look in different directions. It allows the head-pose of a specific person to change in a wide range during his/her appearance in front of the camera[7]. Therefore, it is important to find the least rotated face image of an object of interest in a video sequence and give the best quality score of head-pose to this face. The head-pose is defined as the difference between the center of mass of the skin pixels (inside the face region) and the center of the face's bounding box. This method works fine with LR images but tends to be sensitive to changes in the environmental conditions.

Head-pose is determined by pan and tilt angle. Each angle varies between −90° and +90°, with a step of 15° for pan, and 30 and 15 for tilt. In this paper, we only use pan information. The results of the head-pose estimation are used to classify the images of the input video sequence to three initial face logs. If the head-pose of a face is between −15° and 15° it will be added to the initial frontal face-logs, if it is more than 15° it will be added to the right side-view initial face-log and if it is less than −15° it will be added to the left side-view initial face-log.

D. Face Quality Assessment
In order to compare a face image of a specific person with the other face images of the same person from a video sequence, we need to assign a quality score to each face. To do so, we have combined the normalized value of the above-explained features into a quality score for each face. Equation (1) is used to normalize the head-pose value

$$Q1Xi = Pmin / PXi \qquad (1)$$

where Xi is the ith face image in the given video sequence (i is changing from one to the size of the initial face-log, m1), PXi is its head-pose value, and Q1Xi is its first quality score. Pmin is the minimum value of the head-pose feature in the face-log.

E. Producing the Face-Log
Suppose that the initial frontal face-log contains m1 face images. The FQA technique reduces its size to m2 images, such that m2< m1. This reduced face-log contains the m2 best images of the initial face-log and is denoted intermediate face log. The image with the highest quality score in this intermediate face-log is found and considered as the reference image (best image). Then, two similarity measures are calculated between the reference image and the other face images in the intermediate facelog. These two similarity measures are correlation coefficient and structural similarity measure. For calculating these similarity measures, we need to resize the images to the same size as the reference image and convert them to grayscale.

III. SUPER-RESOLUTION

The employed SR algorithm in this paper cascades both types of reconstruction-based and learning-based SR algorithms. The size of the smallest face that our face detector can detect is 24 × 24. Since the subjects are moving in the video sequence, we have faces of different sizes in the obtained face log. However, in order to be able to apply the RBSR to the images in the refined face-log we resize all of them to 46×56 pixels after face quality assessment. The RBSR produces a HR image of size 92×112 from the LR images of the refined face-log. Then, the proposed system feeds this image (of size 92×112) to the learning-based part to improve its quality even more. Before applying the RBSR, the images in the refined face log need to be first registered to compensate for their misalignments. The employed registration algorithm[6] is described in the following subsection.

A. Face Image Registration
This approach takes into account horizontal shift a, vertical shift b and rotation angle θ between the LR images in the refined face-log [3]. Suppose Y is the reference image in the refined face-log and Xi is the ith face image in the log and it is going to be registered with Y.

B. Reconstruction-Based Super-Resolution
In order to reconstruct the HR image from the LR images of the refined face-log we assume that these images have been produced from the HR image by following an imaging model. Based on the imaging model each LR image has been created by warping, blurring and down-sampling the HR image[1] . It means that each Xi, i = {1, 2, . . . , m3} LR images in the refined face-log have been obtained by

$$Xi = DBiWiH + ni \qquad (2)$$

where D, Bi, and Wi are the down-sampling, blurring, and warping matrix, respectively, H is the HR image and ni is the introduced noise to the imaging process for producing the ith LR image from the HR image H.

C. Learning-Based Super-Resolution
The employed LBSR algorithm in this paper is a MLP. The benefit of using neural networks is their auto-ability in learning the complicated space of the human faces. To prepare the training data for the network, first, we have extracted the face areas of the HR images and converted them to grayscale. Then, an LR image is created for each of these HR images by down-sampling the images by a factor of two. Then, all of these LR/HR pairs are fed to the network as he training samples and the network learns the relationship between them. The original HR face images in this database are all resized to 184×224 and their corresponding LR images are of 92 × 112 pixels. This three-layered MLP has 92 × 112 neurons in its input layer, each corresponding to one of the pixels of the input LR image, ten neurons in the hidden layer, and 184×224 neurons in the output layer, each corresponding to one of the pixels of the output HR image. It shows that the result of cascading these SR algorithms is better than using them separately.

IV. EXPERIMENTAL RESULTS

To show the efficiency of the proposed system in real-world situations, it has been tested for real video sequence.

In our Super resolution application, the multiple source images can be given as input to produce a high resolution image or else we can first create low resolution image from high resolution image, then producing high resolution image.

Fig 2. shows the obtained LR image from the HR image.

Fig 2

Fig 3. shows obtained HR image from the LR image.

Fig 3

V. CONCLUSION AND FUTURE TRENDS

Typically the videos from a surveillance camera is of low and poor quality, where as Facial analysis applications like Face Recognition requires high quality frontal face images. Hence, we need a mechanism to bridge gap between Facial analysis system and low quality & low resolution face extracts from video sequence. Super resolution is one of the techniques used for this purpose. The current project deals with real-world problems of super-resolution systems working in surveillance video sequences. The supers - resolution systems is prone to error because of registration error. The free movement of face in video increases registration errors.

Face log-generation technique uses a Face Detection and Head pose Estimation technique to classify the face images of similar motions to same class by discarding the useless images in provided input video sequence. This technique reduces the registration errors and improves the quality of the resulting system. When typical reconstruction-based super-resolution is used with Learning-based super-resolution it improves the quality of images further to 2 and were stable. Finally it increases the improvement factor of the system to almost four. Whereas when improvement factor is close to two, we had improved quality and compromised stability.

In our project we have taken only the frontal face images, for further applications like identifying a person in a video, who has shown only his left or right side of face image, we can generate left or right side face images database using our project. The proposed technique efficiently and effectively identifies a single person in a video and converted to a high resolution image, if a video contains more than single person, still it's a challenging task.

REFERENCES

[1] V. Bannore, Iterative Interpolation Super Resolution Image Reconstruction. Berlin/Heidelberg, Germany: Springer-Verlag, 2009.

[2] S. Chaudhuri, Super Resolution Imaging, 2nd ed. New York: Kluwer Academic Publishers, 2002.

[3] T. S. Huang and R. Tsai, “Multi-frame image restoration and registration,” Adv. Comput. Vis. Image Process., vol. 1, no. 2, pp. 317–339, 1984.

[4] R. R. Schultz and R. L. Stevenson, “Extraction of high-resolution frames from video sequences,” IEEE Trans. Image Process., vol. 5, no. 6, pp. 996–1010, Jun. 1996.

[5] S. Lertrattanapanich and N. K. Bose, “High resolution image formation from low resolution frames using Delaunay triangulation,” IEEE Trans. Image Process., vol. 11, no. 12, pp. 1427–1441, Dec. 2002.

[6] M. E. Tipping and C. M. Bishop, “Bayesian image super-resolution,” Adv. Neural Inform. Process. Syst., vol. 15, pp. 1303–1310, 2002.

[7] L. Zhang, H. Zhang, H. Shen, and P. Li, “A super-resolution reconstruction algorithm for surveillance images,” Signal Process., vol. 90, no. 3, pp. 848–859, 2010.

[8] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in Proc. Int. Conf. CVPR, 2008.

◈◈◈