FACE LOG GENERATION FOR SUPER RESOLUTION USING LOCAL MAXIMA IN THE QUALITY CURVE

Kamal Nasrollahi and Thomas B. Moeslund
Computer Vision and Media Technology (CVMT) Lab, Aalborg University, Niels Jernes Vej 14, Aalborg, Denmark

Keywords: Face Quality Measures, Super Resolution, Face Logs.

Abstract: Using faces of small sizes and low qualities in surveillance videos without utilizing some super resolution algorithms for their enhancement is almost impossible. But these algorithms themselves need some kind of assumptions like having only slight motions between low resolution observations, which is not the case in real situations. Thus a very fast and reliable method based on the face quality assessment has been proposed in this paper for choosing low resolution observations for any super resolution algorithm. The proposed method has been tested using real video sequences.

1 INTRODUCTION

Due to freedom of movement, changing in lightening and large distance between objects and cameras in surveillance situations faces are usually blurred, noisy, small and low quality. These face images aren't useful without some techniques for improving their resolution and quality. Super resolution image reconstruction is one of these techniques that attempts to estimate a higher quality and resolution image from a sequence of geometrically warped, aliased, noisy, and under-sampled low-resolution images.

Super resolution mainly consists of two important steps: registration of low resolution images and reconstruction of the high resolution image (Baker, 2000, Bannore, 2009, Chaudhuri, 2002, Chaudhuri, 2005). Registration is of critical importance. Most of the reported super resolution algorithms are highly sensitive to the errors in registration. Irani and Peleg are using an iterative method for registration under some assumptions which are only valid for small displacements between the low resolution input images (Irani, Peleg, 1991). A real one minute surveillance video capturing by a camera at 30 frp consists of 1800 frames. Due to the movement of objects usually there are large displacements between the first and last appearance of the objects in this kind of videos. It means that Irani and Peleg method as well as most of the other super resolution algorithms is not applicable to real video sequences if they are used blindly the same for still images and video sequences. Instead of applying the super resolution algorithm to all the images of a given sequence we apply this algorithm to face logs (Nasrollahi, Moeslund, 2009) of the video sequence. There are small displacements between successive faces in the face logs which are constructed based on the Face Quality Assessment.

As any other inverse problem reconstruction step in super resolution is highly sensitive to even small perturbations of the input data, i.e. it is an ill-conditioned problem. There are many published system aiming dealing with this problem and increasing the reliability of the reconstruction step. But there has not been that much attention to computational efficiency and real time applicability of super resolution. It is, thus, desirable to develop algorithms that maintain a proper balance between computational performance and the fidelity of the reconstruction (Bannore, 2009). Providing such a balance by proposing a very fast framework for super resolution image reconstruction working with video sequences is the contribution of this paper. By using a face quality assessment algorithm we reduce the number of inputs to the super resolution algorithm and by generating face logs in a specific way we force the inputs to have slight motions as is desired for the super resolution.

The block diagram of the proposed system is shown in Figure 1. In the first block, having a surveillance video faces and some of the facial

Nasrollahi K. and B. Moeslund T. (2010).
FACE LOG GENERATION FOR SUPER RESOLUTION USING LOCAL MAXIMA IN THE QUALITY CURVE.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 124-129
DOI: 10.5220/0002846601240129
Copyright © SciTePress

components are first detected. Some of the important features of face and facial components are then extracted and normalized for using them in face quality assessment. In the second block, using a fuzzy inference engine a quality curve is first plotted for the input video sequence based on the quality scores from the previous step, then for each peak in the quality curve a face log containing m-best face images is generated and finally a super resolved image is generated for each of these face logs.

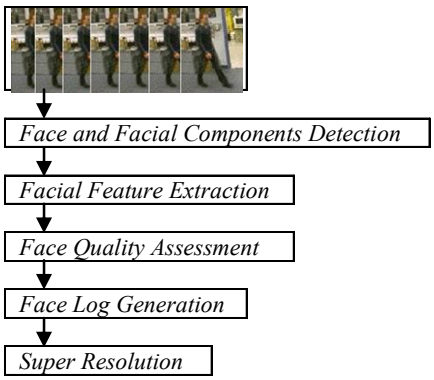

Figure 1: Block diagram of the proposed system.

The rest of the paper is organized as follows: Next section, data preparation, describes breifly face and facial components detection, features extractions, converting features values into quality scores and their normalization. Section 3 presents the fuzzy inference engine and face log generation from the quality curve. Section 4 descibes the employed super resolution algorithm. Section 5 discusses the experimental results and Section 6 concludes the paper.

2 DATA PREPARATION

Having a surveillance video sequence as the input to the system, the required processes from detecting faces in the video to representing each facial feature by a quality score are described in the following subsections.

2.1 Face and Facial Components Detection

The Vila & Jones (Viola, Jones, 2004) face detector which uses Haar-like features is employed for real time face detection. Inside each detected face, the same idea of using Haar-like features is then utilized for finding some of the facial components, i.e. eyes and mouth.

2.2 Feature Extraction

The extracted features for detected faces are: Face Yaw, Face Roll, Sharpness, Brightness and Resolution. Since faces in surveillance videos are usually of small sizes it is not easy to detect facial components and extract their features so wherever they are extractable they will be used as marginal information for improving the reliability of the system. The employed features for the facial components are the openness of the eyes and closedness of the mouth.

2.2.1 Facial Features

Extraction and normalization of the involved features of detected faces for any given sequences are as follows.

Face Yaw: The head yaw is defined (Nasrollahi, 2008) as the difference between the center of mass of the skin pixels and the center of the bounding box of the face. For calculating the center of mass, the skin pixels inside the face region should be segmented from the non-skin ones. Then using the following equation the yaw value is estimated:

$$Yaw_i = \sqrt{(x_{cm} - x_{bb})^2 + (y_{cm} - y_{bb})^2} \quad (1)$$

where, Yaw_i is the estimated value of the yaw of the ith face image in the sequence, (x_{cm}, y_{cm}) is the center of mass of the skin pixels and (x_{bb}, y_{bb}) is the center of the bounding box of the face. Since the biggest score for this feature should be assigned to the least rotated face, Equation (2) is used to normalize the scores of this feature for all the faces in the sequence:

$$S_1 = \frac{Yaw_{min}}{Yaw_i} \quad (2)$$

where, Yaw_{min} is the minimum value for the yaw in the given sequence.

Face Roll: The cosine of the angle of the line connecting the center of mass of both eyes is considered as face roll. Having extracted the values of this feature for all the faces in the sequence the following equation is used to normalize them:

Face Roll: The cosine of the angle of the line connecting the center of mass of both eyes is considered as face roll. Having extracted the values of this feature for all the faces in the sequence the following equation is used to normalize them:

$$S_2 = \frac{Roll_{min}}{Roll_i} \quad (3)$$

Sharpness: Following (Weber, 2006) the sharpness is defined as the average of the pixels of the absolute value of the difference between the image and its low passed version:

$$Sharpness_i = avg.(abs\,(image - lowPassed(image))) \quad (4)$$

Then Equation (5) is used to normalize the sharpness values for the faces in the sequence:

$$S_3 = \frac{Sharpness_i}{Sharpness_{max}} \quad (5)$$

where, $Sharpness_{max}$ is the maximum value of the sharpness in the given sequence.

Brightness: Brightness is defined as the mean of the luminance component of the face images in the $YCbCr$ color space and Equation (6) is used to normalize this feature.

$$S_4 = \frac{Brightness_i}{Brightness_{max}} \quad (6)$$

where, $Brightness_i$ is the brightness of the *i*th image in the sequence and $Brightness_{max}$ is the maximum value of brightness in that.

Resolution: Facial components are usually more visible in images with higher resolution, so these images are preferred over lower resolution ones. Equation (7) is used for normalization of this feature:

$$S_5 = \frac{Resolution_i}{Resolution_{max}} \quad (7)$$

where, $Resolution_i$ is the resolution value of the current image and $Resolution_{max}$ is the maximum value of the resolution in the given sequence.

2.2.2 Facial Components' Features

Extraction and normalization of the involved features of the eyes and mouth of the detected faces are as follows.

Eye Openness: The aspect ratio of the eye, the ratio of the eye's height to its width, is used as a straightforward method for determining its openness. The eyes are detected first by Haar features, and then converted into binary images to have a good estimation of eye boundaries. For improving this estimation and noise removal, an opening operation is applied to the obtained binary image. From this new image, the height of the eye is determined using the highest and the lowest pixels and the width using the rightmost and the leftmost pixels. Equation (8) is then used for the normalization of this feature in a given sequence:

$$S_{6a,6b} = \frac{Openness_i}{Openness_{max}} \quad (8)$$

where, S_{6a} is the openness score of the left eye, S_{6b} is the same for the right eye, $Openness_i$ is the value of this feature for the current eye for the *i*th image and $Openness_{max}$ is the maximum value of this feature for the current eye in the given sequence. The value of these two features is initially set to zero and after the above process they are averaged using the following equation:

$$S_6 = \frac{S_{6a} + S_{6b}}{Number\ of\ detected\ eyes} \quad (9)$$

Mouth Closedness: The aspect ratio of mouth which is defined as the ratio of the mouth height to its width, increases as the mouth opens. The following equation is used to normalize this feature:

$$S_7 = \frac{Aspect\ Ratio_{min}}{Aspect\ Ratio_i} \quad (10)$$

where, $Aspect\ Ratio_{min}$ is the minimum value of this feature in the sequence and $Aspect\ Ration_i$ is the value of this feature for the current mouth.

3 FACE QUALITY ASSESSMENT

In order to assign a quality score to each face in the given video sequence, all the above mentioned normalized scores are combined using a Mamdani Fuzzy Inference Engine (FIS). This FIE has seven inputs each corresponding to one of the above features and one output indicating the quality of the current face image. Each input of the FIE has two membership functions and the output has three. Figure 2 shows the rules used in this FIE. In this table F*i*, R*i*, P, G and Avg. stand for *i*th feature, *i*th rule, poor, good, and average, respectively. For more details about this FIE the reader is motivated to see (Nasrollahi and Moeslund, 2009).

Using this fuzzy interpretation similar face images get similar quality scores and they can be classified easily in classes of quality corresponding to the local maxima of the quality curve as is explained in the next section.

	F1	F2	F3	F4	F5	F6	F7	Quality
R1	P	P	P	P	P	-	-	P
R2	P	P	G	G	G	-	-	Avg.
R3	P	G	G	G	G	G	G	G
R4	P	G	P	P	P	-	-	P
R5	P	G	P	P	G	-	-	P
R6	P	G	P	G	P	-	-	P
R7	P	G	G	P	P	-	-	P
R8	P	G	P	G	G	P	P	Avg.
R9	P	G	G	P	G	P	P	Avg.
R10	P	G	G	G	P	P	P	Avg.
R11	G	P	G	G	G	G	G	G
R12	G	P	P	P	P	-	-	P
R13	G	P	P	P	G	-	-	P
R14	G	P	P	G	P	-	-	P
R15	G	P	G	P	P	-	-	P
R16	G	P	P	G	G	P	P	Avg.
R17	G	P	G	P	G	P	P	Avg.
R18	G	P	G	G	P	P	P	Avg.
R19	G	P	G	G	P	G	G	G
R20	G	P	P	G	G	G	G	G
R21	G	P	G	P	G	G	G	G
R22	G	G	G	G	G	-	-	G

Figure 2: FIE's employed rules (abbreviations in the text).

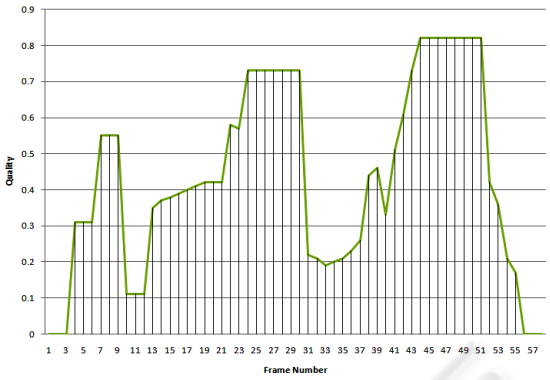

Figure 3: Quality curve for a given sequence (quality vs. frame number).

4 FACE LOG GENERATION

In order to prepare quality based classified data to the super resolution algorithm a quality curve is generated based on the output of the FIE for any input video sequence. The peaks of this quality curve are found and a face log is generated for each of them. Usually building m (m=3-5) face logs corresponding to the m highest peaks suffices. But theoretically number of the face logs can be equal to the number of local maxima in the quality curve.

Each face log contains all the images located on a specific local maximum.

The face quality curve for a video sequence containing 57 frames is shown in Figure 3. The face logs corresponding to the three highest peaks of this curve are shown in Figure 4 (a)-(c). Each local maximum contains images which are more or less similar in terms of quality. These images are temporally close to each other, i.e. rather than similarity in quality except some small variations in motion, lightening, rotation and scales they resemble each other closely and so the compensation between them can be done more easily and accurately which makes the super resolution algorithm more robust and fast.

By this way of face log constructing one can be sure that useless images which are the source for most of the errors in registration step of super resolution are ignored and faces with more spatial similarity are considered together. This spatial similarity dose not prevents images of having slight differences in rotation, scale, brightness, and motion which are necessary for super resolution.

5 SUPER RESOLUTION

The employed super resolution algorithm in this system is the one developed by Zomet and Peleg (Zomet, Peleg, 2001). They have followed (Elad, Feuer, 1999) for formulating the imaging formation as

$$\underline{x}_L^{(n)} = DB_nW_n\underline{x}_H + e_n \tag{11}$$

where $\underline{x}_L^{(n)}$ is the nth low resolution image, D is decimation matrix, B_n is the blurring matrix, W_n is the warping matrix, $\underline{x}_H$ is the high resolution image and e_n is the normally distributed additive noise in the nth image.

Stacking the above vector equations from all low resolution images:

$$\underline{x}_L = A\underline{x}_H + e \tag{12}$$

The maximum likelihood solution is then found by minimizing:

$$E(\underline{x}_H) = \frac{1}{2}||\underline{x}_L - A\underline{x}_H||^2 \tag{13}$$

The minimization is done by taking the derivative of E with respect to $\underline{x}_H$ and setting the gradient to zero:

$$\sum_{n=1}^{K} W_n^T B_n^T D^T (DB_nW_n\underline{x}_H - \underline{x}_L^{(n)}) = 0 \tag{14}$$

Figure 4: Face logs corresponding to the three highest peaks of the quality curve in Figure 3.

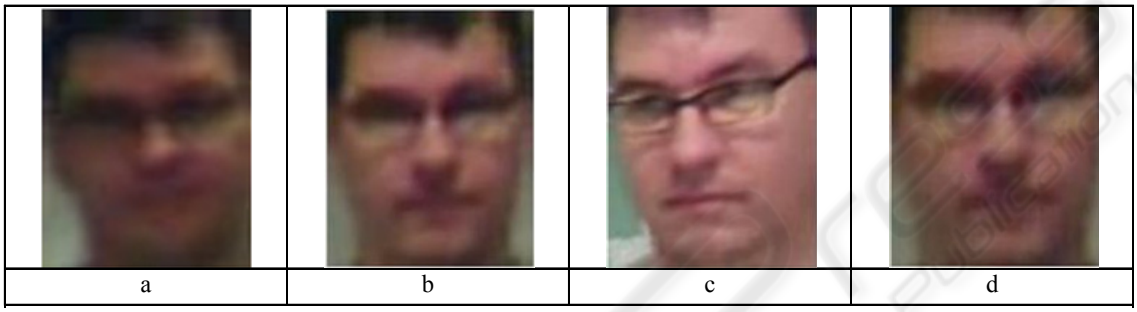

Figure 5: Results of applying the super resolution to a) first, b) second, c) third face log of the sequence of Figure 4. d) Result of the algorithm applied to all the faces in that sequence.

Since D is sub-sampling, $D^T$is up-sampling without interpolation, i.e. zero padding. If b is the convolution kernel for the blurring operator B_n, $\acute{b}$ is blurring kernel for B_n^T such that $\acute{b}(i,\ j) = b(-i,-j)$ for all i and j. W_n is implemented by backward warping so W_n^T is forward warping of the inverse motion. Following (Zomet, Peleg, 2001), the simplest implementation for this framework using Richardson (Kelley, 1995) iteration is used:

$$\underline{x}_H\{m+1\} = \underline{x}_H\{m\} + \sum_{n=1}^{K} W_n^T B_n^T D^T(\underline{x}_L^{(n)} - DB_n W_n \underline{x}_H\{m\}) \quad (15)$$

Since super resolution is an ill-posed problem a regularization term is usually considered to convert the problem into a well-posed one. The used regularization term here is the smoothness of the high resolution response. For more details regarding the super resolution algorithm, motion compensation and blurring considerations the reader is motivated to see (Chaudhuri, 2002), (Irani, Peleg, 1991) and (Rav-Acha, Peleg., 2005 and Chiang, Boult, 1997), respectively.

6 EXPERIMENTAL RESULTS

For experimental results 50 video sequences pictured by a Logitech camera are used. The average number of images in each sequence is about 300. Two tests have been performed for each of these sequences: first applying the super resolution algorithm to the face logs generated by the above mentioned method and second applying the super resolution algorithm directly to the video sequences.

Figure 5 shows the results of the both tests for the sequence given in Figure 4. As can be seen from this figure applying the super resolution to the face logs generates much better results than applying it directly to the video sequences. It means that the qualities of the images generated by the first method are much better than their counterparts generated by the second method. First images reveal more details of the face and obviously are more useful than the ones generated by the second method. The reason is that compensating for the changes inside a face log is much easier with fewer errors than doing that for the whole sequence.

In addition to the higher quality of each of the super resolved images of the first method over the results of the second method there is one more possibility for applying the super resolution

algorithm one more time to the super resolved images of the first method. For this purpose these super resolved images are first resized to the same size as the largest one using the bi-cubic interpolation and then the second round of the super resolution algorithm is applied. The result of this second application of the algorithm for the face logs of Figure 4 is shown in Figure 6. It is obvious that the super resolution algorithm for the case of the face logs is much faster. Because the number of the low resolution observations is not excessive.

Figure 6: Results of applying the second round of super resolution to the super resolved results of the face logs in Figure 4.

7 CONCLUSIONS

Super resolution algorithms have difficulties in the registration of low resolution observations. If the motion between low resolution observations be more than some specific limits these algorithms fail to compensate for the motion and blurring. Thus extending super resolution algorithms which work with still images to real video sequences without some kind of intermediate step for ignoring useless images in the sequence and classifying them based on their similarity in motion and quality is not possible. In this paper a face log generation method specifically for face super resolution has been developed and tested using real video sequences to fill the gap between the super resolution algorithms which work with still images and their application to the real video sequences. The proposed system has been tested using 50 real sequences pictured by a Logitech camera and the results are promising.

ACKNOWLEDGEMENTS

This work is funded by the BigBrother project (Danish National Research Councils-FTP).

REFERENCES

Baker, S., and Kanade, T., 2000., Limits on super resolution and how to break them. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 9.

Bannore, V., 2009. *Iterative interpolation super resolution image reconstruction*. Springer-Verlag Berlin Heidelberg.

Chaudhuri, S., 2002. *Super resolution imaging*, Kluwer Academic Publishers. New York, 2nd edition.

Chaudhuri, S, Joshi, M. V., 2005. *Motion free super resolution*, Springer Science. New York.

Chiang, M., Boult, T. E., 1997. Local blur estimation and super resolution. In *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 821-830.

Elad, M., Feuer, A., 1999. Super resolution reconstruction of image sequences, In *IEEE Transaction on Pattern Analysis and Machine Intelligence*. Vol. 21, No. 9. pp. 817-834.

Irani, M., Peleg, S., 1991. Improving resolution by image registration. In *Graphical Models and Image Processing*. Vol. 53, No. 3.

Kelley, C. T., 1995, Iterative methods for linear and nonlinear equations, *SIAM*, Philadelphia, PA.

Nasrollahi, K., Rahmati, M., Moeslund, T. B., 2008. A Neural Network Based Cascaded Classifier for Face Detection in Color Images with Complex Background. In *International Conference on Image Analysis and Recognition.*

Nasrollahi, K. Moeslund, T. B., 2009. Complete Face Logs for Video Sequences Using Quality Face Measures, In *IET International Journal of Signal Processing*, Vol. 3, No. 4, pp. 289-300.

Rav-Acha, A., Peleg., S., 2005. Two motion blurred images are better than one, In *Pattern recognition letter,* Vol. 26, pp. 311-317.

Viola, P., Jones, M. J. 2004. Robust Real Time Face Detection. In *International Journal of Computer Vision,* Vol. 57, No. 2, pp. 137-154.

Weber., F., 2006. Some Quality Measures for Face Images and Their Relationship to Recognition Performance. In *Biometric Quality Workshop*, National Institute of Standards and Technology.

Zomet, A., Peleg, S., 2001. Super resolution from multiple images having arbitrary mutual motion, In: S. Chaudhuri, Editor, *Super-resolution imaging*, Kluwer Academic, Norwell, pp. 195–209.