Evaluating the Quality of Super-resolved Images for Face Recognition

Xiaoli Zhou and Bir Bhanu
Center for Research in Intelligent Systems
University of California at Riverside, Riverside, CA 92521, USA

Abstract

The widespread use of super-resolution methods, in a variety of applications such as surveillance has led to an increasing need for or quality assessment measures. The current quality measures aim to compare different fusion methods by assessing the quality of the fused images. They consider the information transferred between the super-resolved image and input images only. In this paper, we propose an objective quality evaluation algorithm for super-resolved images, which focuses on evaluating the quality of super-resolved images that are constructed from *different conditions* of input images. The proposed quality evaluation method combines both the relationship between the super-resolved image and the input images, and the relationship between the input images. Using the proposed measure, the quality of the super-resolved face images constructed from videos are evaluated under different conditions, including the variation of pose, lighting, facial expressions and the number of input images.

1 Introduction

Super-resolution for image and video is emerging as a vital technology in recent years. The aim of image and video super-resolution is to create new images that are more suitable for human/machine perception. In many application scenarios, a super-resolution algorithm is only an introductory preprocessing stage to other tasks. Therefore, the quality of super-resolved images need to be measured in terms of the performance improvement for the subsequent tasks.

In the literature, several mathematically defined objective image quality measures have been suggested for their ease of computation and independence of viewing conditions and individual interpreters. Among the quality indices, the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), signal to noise ratio (SNR), and peak signal to noise ratio (PSNR) are widely employed for comparing a distorted image with an ideal image in full-reference quality assessment approaches. However, for super-resolution applications, the ideal reference image is normally unknown. Designing objective image fusion metrics for cases without an ideal or a reference image is a very difficult task, but such metrics are highly desired.

Table 1 presents a summary of the recent work and compares it with the work presented in this paper for the image quality evaluation. The first three quality measures (SSIM, UQI and MCQI) need a reference. The quality of the fused image is evaluated based on the similarity between the fused image and the reference image. The non-reference image quality measures [5] [4] [10] [6] [8] [7] [9] mainly focus on comparing the fusion results obtained with different algorithms. The quality of the fused image is evaluated based on the similarity between the fused image and the input images.

The proposed quality measure aims to evaluate the quality of super-resolved images constructed from different conditions by analyzing the factors which may influence the quality of super-resolved face images. It can be computed independent of the subsequent task. In comparison to the previous work, the contributions of this paper are as follows:

- An objective quality evaluation algorithm is proposed for the super-resolved images, which does not require a ground-truth or reference image. It focuses on comparing the quality of super-resolved images that are constructed under different conditions.

- The quality of the super-resolved face images constructed from real video data are evaluated using the proposed quality measures under different conditions, including the variation of pose, lighting, facial expression and the number of input images. The influence of different conditions on the quality of the super-resolved face images is analyzed based on the experimental results.

- The relationship between the quality of the face image and the performance of face recognition is addressed. Face images of 45 people are con-

978-1-4244-2340-8/08/$25.00 ©2008 IEEE

Table 1: Sample recent work for the image quality evaluation vs this paper.

Authors	Methodology	Boundedness	Need for Reference Image	Parameterization
Wang and Bovik (2002) [1] (2004) [2]	Combination of the luminance, contrast, and structure. Universal Image Quality Index (UQI) [1] is a special case of the Structural Similarity Index (SSIM) [2]	UQI: $[-1, 1]$ SSIM: $[-\infty, 1]$	Yes	UQI: No SSIM: Yes
Cornet and Binard (2004) [3]	Multi-Criteria Quality Index (MCQI): Linear Combination of three standardized criteria	$[-3, 3]$	Yes	No
Xydeas and Petrovic (2000) [4]	Edge information	$[0, 1]$	No	Yes
Qu et al. (2002) [5]	Mutual information (MI) measure	$[0, \infty]$	No	No
Piella and Heijmans (2003) [6]	Combine local salient information and edge information with UQI index	$[-1, 1]$	No	Yes
Tsagaris and Anastassopoulos (2004) [7]	Mutual information and conditional information	$[0, 1]$	No	No
Cvejic et al. (2005) [8]	Combine local similarity with UQI	$[0, 1]$	No	No
Cvejic et al. (2006) [9]	Mutual information and Tsallis entropy	$[0, \infty]$	No	Yes
This paper	Combine the relationship between the super-resolved image and the input image, and the relationship between the input images	$[-1, 1]$	No	Yes

structed using the different numbers of input images from a video database. The quality evaluation and face recognition experiments are conducted on these super-resolved images. Experimental results show the effectiveness of the proposed quality measure for the super-resolved image and the necessity of super-resolution for face recognition under low-resolution conditions.

2 Proposed Image Quality Index

The proposed quality evaluation method is based on the intensity value of a pixel in a image. For the color images, they are first transformed into YIQ representation and the luminance (Y) component [11] is used. The integrated quality measure Q_{int} takes the form as

$$Q_{int} = f(Q_g, Q_e, Q_i)$$

where Q_g is the gray scale based quality, Q_e is the structure based quality, and Q_i is the similarity between input images.

2.1 Gray Scale Based Quality (Q_g)

The gray scale based quality takes into account the integration of information transferred from all the input images to the super-resolved image. The form of this integration is similar to the form in [6]. We define the quality based on the gray value as Q_g. It is given by

$$\begin{aligned} Q_g(\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n, \mathbf{F}) &= \sum_{\mathbf{w} \in \mathbf{W}} \kappa(\mathbf{w})[\alpha_1(\mathbf{f}_1|\mathbf{w})Q(\mathbf{f}_1, \mathbf{F}|\mathbf{w}) \\ &+ \alpha_2(\mathbf{f}_2|\mathbf{w})Q(\mathbf{f}_2, \mathbf{F}|\mathbf{w}) + ... \\ &+ \alpha_n(\mathbf{f}_n|\mathbf{w})Q(\mathbf{f}_n, \mathbf{F}|\mathbf{w})] \qquad (1) \end{aligned}$$

where $\mathbf{f}_i, i = 1, 2, ..., n$ is the input image, $\mathbf{F}$ is the super-resolved image, $\mathbf{w}$ is the analysis window and $\mathbf{W}$ is the family of all windows. The parameter α_i and $\kappa(\mathbf{w})$ are defined as

$$\begin{aligned} \alpha_i(\mathbf{f}_i|\mathbf{w}) &= \frac{\sigma^2(\mathbf{f}_i|\mathbf{w})}{\sigma^2(\mathbf{f}_1|\mathbf{w}) + \sigma^2(\mathbf{f}_2|\mathbf{w}) + ... + \sigma^2(\mathbf{f}_n|\mathbf{w})} \\ i &= 1, 2, ..., n. \end{aligned}$$

$$\kappa(\mathbf{w}) = \frac{\max_{i=1}^{n} \sigma^2(\mathbf{f}_i|\mathbf{w})}{\sum_{\mathbf{w}' \in \mathbf{W}} \max_{i=1}^{n} \sigma^2(\mathbf{f}_i|\mathbf{w}')} \qquad (2)$$

where $\sigma^2(\mathbf{f}_i|\mathbf{w})$ denotes the variance of image $\mathbf{f}_i$ in window $\mathbf{w}$.

$Q(\mathbf{f}_i, \mathbf{F}|\mathbf{w}), i = 1, 2, ..., n$ evaluates the similarity between the images, $\mathbf{f}_i$ and $\mathbf{F}$, within the sliding window $\mathbf{w}$. It takes the same form as proposed by Wang and Bovik [1]. Let $\mathbf{f} = \{x_i|i = 1, 2, ..., N\}$ and $\mathbf{F} = \{y_i|i = 1, 2, ..., N\}$ be the input and the super-resolved image signals, respectively. Q is defined as

$$Q = \frac{4\sigma_{xy}\mu_x \cdot \mu_y}{(\sigma_x^2 + \sigma_y^2)[(\mu_x)^2 + (\mu_y)^2]} \qquad (3)$$

where

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2,$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu_y)^2$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$$

The dynamic range of Q is $[-1, 1]$. The best value 1 is achieved if and only if $y_i = x_i$ for all $i = 1, 2, ..., N$. The lowest value of -1 occurs when $y_i = 2\mu_x - x_i$ for all $i = 1, 2, ..., N$. The definition of Q can be decomposed as a product of three components

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\mu_x \mu_y}{(\mu_x)^2 + (\mu_y)^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^y} \quad (4)$$

This quality index models any distortion as a combination of three different distortions: loss of correlation, luminance distortion, and contrast distortion. Q is used to quantify the structural distortion between two images. In fact, the value $Q = Q(\mathbf{f}, \mathbf{F})$ is a measure for the similarity of images $\mathbf{f}$ and $\mathbf{F}$. The first component is the correlation coefficient between $\mathbf{f}$ and $\mathbf{F}$ and its dynamic range is [-1, 1]. The second item measures how close the mean luminance is between $\mathbf{f}$ and $\mathbf{F}$ and it has a dynamic range of [0, 1]. The third item measures how similar the contrast distortion is and its dynamic range is also [0, 1].

Since images are generally non-stationary signals, it is appropriate to measure Q over local regions and then combine the different results into a single measure Q. Wang and Bovik propose to use a sliding window [1]: starting from the top left corner of the two images, $\mathbf{f}$ and $\mathbf{F}$, a sliding window of fixed size moves pixel by pixel over the entire image until the bottom-right corner is reached. For each window $\mathbf{w}$, the local quality index $Q(\mathbf{f}, \mathbf{F}|\mathbf{w})$ is computed for the pixels within the sliding window $\mathbf{w}$.

2.2 Structure Based Quality (Q_e)

Generally, an image with stronger edges is regarded to have a better quality. Therefore, we take into account the edge strength, which is associated to some important visual information of the human visual system. The Sobel operator performs a 2D spatial gradient measurement on an image. For the input image $\mathbf{f}_i$ and the super-resolved image $\mathbf{F}$, we could get the corresponding edge strength image $\hat{\mathbf{f}}_i$ and $\hat{\mathbf{F}}$. The structure based quality of the super-resolved image Q_e is evaluated as

$$\begin{aligned} Q_e(\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, ..., \hat{\mathbf{f}}_n, \hat{\mathbf{F}}) &= \sum_{\mathbf{w} \in \mathbf{W}} \lambda(\mathbf{w})[\beta_1(\hat{\mathbf{f}}_1|\mathbf{w}) Q(\hat{\mathbf{f}}_1, \hat{\mathbf{F}}|\mathbf{w}) \\ &+ \beta_2(\hat{\mathbf{f}}_2|\mathbf{w}) Q(\hat{\mathbf{f}}_2, \hat{\mathbf{F}}|\mathbf{w}) + ... \\ &+ \beta_n(\hat{\mathbf{f}}_n|\mathbf{w}) Q(\hat{\mathbf{f}}_n, \hat{\mathbf{F}}|\mathbf{w})] \quad (5) \end{aligned}$$

The parameters β_i and $\lambda(\mathbf{w})$ in Equation (5) are obtained using the same method as the parameters α_i and $\kappa(\mathbf{w})$ in Equation (1), where the σ^2 corresponds to the variance of edge image $\hat{\mathbf{f}}_i$ in window $\mathbf{w}$. They are defined as

$$\begin{aligned} \beta_i(\hat{\mathbf{f}}_i|\mathbf{w}) &= \frac{\sigma^2(\hat{\mathbf{f}}_i|\mathbf{w})}{\sigma^2(\hat{\mathbf{f}}_1|\mathbf{w}) + \sigma^2(\hat{\mathbf{f}}_2|\mathbf{w}) + ... + \sigma^2(\hat{\mathbf{f}}_n|\mathbf{w})} \\ i &= 1, 2, ..., n. \end{aligned}$$

$$\lambda(\mathbf{w}) = \frac{\max_{i=1}^n \sigma^2(\hat{\mathbf{f}}_i|\mathbf{w})}{\sum_{\mathbf{w}' \in \mathbf{W}} \max_{i=1}^n \sigma^2(\hat{\mathbf{f}}_i|\mathbf{w}')} \quad (6)$$

$Q(\hat{\mathbf{f}}_i, \hat{\mathbf{F}}|\mathbf{w}), i = 1, 2, ..., n$ is computed using Equation (3) to evaluate the similarity between the edge images, $\hat{\mathbf{f}}_i$ and $\hat{\mathbf{F}}$, within the sliding window $\mathbf{w}$.

2.3 Similarity between Input Images (Q_i)

The current quality methods focus on evaluating the fused image quality by directly assessing the similarity between the fused image and the input images [4] [6] [7] [8] [9]. They never directly consider the relationship between the input images. Moreover, they assume that all the input images are perfectly registered, even though it is not the case in most of the time in real-world application, especially for the construction of super-resolved images from video where the input images from video frames have to be registered before any resolution enhancement. Considering the important role of input images in super-resolved image construction, we want to measure the relationship between input images explicitly. Without loss of generality, we assume the input images $\mathbf{f}_i, i = 2, ..., n$ are all aligned to the input image $\mathbf{f}_1$. The relationship between input images is defined as:

$$\begin{aligned} Q_i(\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n) &= \frac{1}{|\mathbf{W}|} \sum_{\mathbf{w} \in \mathbf{W}} [\gamma_1(\mathbf{f}_1, \mathbf{f}_2) Q(\mathbf{f}_1, \mathbf{f}_2|\mathbf{w}) \\ &+ \gamma_2(\mathbf{f}_1, \mathbf{f}_3) Q(\mathbf{f}_1, \mathbf{f}_3|\mathbf{w}) + ... \\ &+ \gamma_{n-1}(\mathbf{f}_1, \mathbf{f}_n) Q(\mathbf{f}_1, \mathbf{f}_n|\mathbf{w})] \quad (7) \end{aligned}$$

$\gamma_i, i = 1, 2, ..., n-1$ is the weight, which represent the registration quality between the two input images. $\mathbf{w}$ is the analysis window and $\mathbf{W}$ is the family of all windows. $Q(\mathbf{f}_1, \mathbf{f}_i|\mathbf{w}), i = 2, ..., n$ is computed using Equation (3) to evaluate the similarity between the input images, $\mathbf{f}_1$ and $\mathbf{f}_i$, within the sliding window $\mathbf{w}$.

Cvejic et al. [8] use cross-correlation to indicate the similarity between images in the spatial domain. However, cross-correlation is not appropriate as the weight because of its potentially negative values. It is also difficult to interpret and inconvenient to use it for more than two input images. The methods based on the maximization of the mutual information (MI) are the leading techniques for multi-modal image registration. MI, originating from the information theory, is a well studied measure of statistical dependency between two data sets. Based on MI, the weight γ_i in Equation (7) is defined as:

$$\gamma_i(\mathbf{f}_1, \mathbf{f}_{i+1}) = \frac{I(\mathbf{f}_1, \mathbf{f}_{i+1})}{\sum_{i=1}^{n-1} I(\mathbf{f}_1, \mathbf{f}_{i+1})} \qquad i = 1, 2, ..., n-1 \quad (8)$$

where I is the mutual information (MI), which describes the common information between two input images and indicates the registration quality. I between two images $\mathbf{f}_1$ and $\mathbf{f}_{i+1}$ is defined as:

$$I(\mathbf{f}_1, \mathbf{f}_{i+1}) = H(\mathbf{f}_1) + H(\mathbf{f}_{i+1}) - H(\mathbf{f}_1, \mathbf{f}_{i+1}) \qquad (9)$$

where

$$\begin{aligned} H(\mathbf{f}_1) &= \sum_{a \in \mathbf{f}_1} p(a) \log p(a) \\ H(\mathbf{f}_{i+1}) &= \sum_{b \in \mathbf{f}_{i+1}} p(b) \log p(b) \\ H(\mathbf{f}_1, \mathbf{f}_{i+1}) &= \sum_{a \in \mathbf{f}_1 b \in \mathbf{f}_{i+1}} p(a, b) \log p(a, b) \end{aligned}$$

$H(\mathbf{f}_1)$ and $H(\mathbf{f}_{i+1})$ are the Shannon entropies of gray level value distributions of images $\mathbf{f}_1$ and $\mathbf{f}_{i+1}$, and $H(\mathbf{f}_1, \mathbf{f}_{i+1})$ is the Shannon entropy of the joint distribution of gray level values of the images $\mathbf{f}_1$ and $\mathbf{f}_{i+1}$. a and b denote gray level values of the images $\mathbf{f}_1$ and $\mathbf{f}_{i+1}$. I is dependent on the similarity in spatial domain between the two input image. In this sense, we are able to measure the relationship between different input images accurately and conveniently using Equation (7) even when the number of the input images is large.

2.4 Integrated Quality Measure (Q_{int})

The current quality measures shown in Table 1 are based on evaluating how much of the information contained in each of the input images has been transferred into the fused image. The fused images for comparison are all constructed from the same input images. They do not consider the quality difference of the fused image due to the changing conditions, such as the changing of the quality of the input image and the number of input images. However, the relationship between the quality of the super-resolved images and the condition of input images are very important. To compare the quality of super-resolved images that are constructed from images acquired under different conditions, the proposed quality evaluation method for the super-resolved images combines not only the relationship between the super-resolved image and input images, but also the relationship between different input images. It takes the final form as:

$$Q_{int} = (1-\theta) \times \frac{(Q_g + Q_e)}{2} + \theta \times Q_i \qquad (10)$$

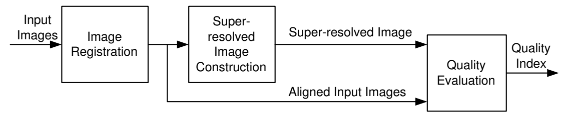


Figure 1: The quality evaluation of the super-resolved image.

where θ is a parameter, $0 < \theta < 1$. Usually, the value of θ should be no more than 0.5. Q_g, Q_e and Q_i are defined in Equation (1) (5) and (7), respectively. The proposed measure has a dynamic range of $[-1, 1]$. The first item is the similarity between the input images and the super-resolved image. The second item is the similarity between the input images.

3 Experimental Results

Using the proposed quality measure defined by Equation (10), the quality of the super-resolved face images constructed from video is evaluated under different conditions. These conditions include the variation of pose, lighting, facial expression and the number of input images used for constructing the super-resolved image. The process of the quality evaluation of the super-resolved images is shown is Figure 1.

3.1 Experiments

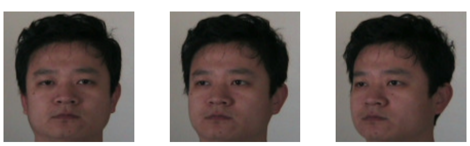
We use an iterative method [12] to construct the super-resolved face image from the aligned input images from the video. An elastic registration algorithm [13] is used for motion estimation of the input face images. The reference (ideal) image is the original face image directly obtained from the video frame, which is used for comparison with the super-resolved image. The input images, which are used for constructing the super-resolved face image, are obtained by downsampling by a factor of 2 the corresponding original face images. The proposed quality measure is computed using the super-resolved face image and the input face images in each case. We choose the size of the sliding window as 8×8 and θ as the inverse of the number of the input images in Equation (10). Moveover, we compute the UQI and PSNR between the super-resolved image and the reference image for comparison. In real-world applications, we do not have access to the reference image in super-resolution scenarios, therefore, the UQI and PSNR values are provided just as references. Finally, different numbers of input images are used to construct the super-resolved images. The relationship between the quality of super-resolved image quality and the recognition performance is tested on a real-video database of 45 people acquired under outdoor conditions.

3.1.1 Experiment 1: Influence of Pose Variation on the Super-resolved Face Image

This experiment tests the influence of pose variation on the quality of the super-resolved face image. Fig. 2 shows three input face images with the same pose. Fig. 3 shows three input face images with perceptibly different poses. Fig. 4 shows the two super-resolved face images constructed from the images in Fig. 2 and 3, respectively, and the ideal reference image (for comparison only). The size of the input images is 128×128 and the size of the super-resolved image is 256×256. The images in Fig. 2(a) and 3(a) are the same and all the other input images are aligned to this image before the construction of the super-resolved image. Therefore, the reference image is the same for the two super-resolved images and it is the original face image corresponding to the image in Fig. 2(a) and 3(a). Table 2 shows the quality of the super-resolved image evaluated using the proposed quality measure, UQI and PSNR.

(a) (b) (c)

Figure 2: Three input face images with the same pose.



(a) (b) (c)

Figure 3: Three input face images with different poses.

(a) (b) (c)

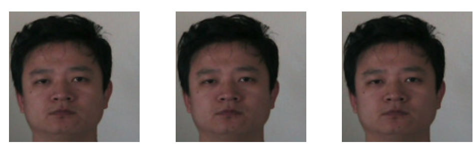
Figure 4: (a) The super-resolved face image from the input images in Fig. 2. (b) The super-resolved face image from the input images in Fig. 3. (c) The ideal reference image shown for comparison. It is directly obtained from the original video.

Table 2: The Effect of Pose Variation on Quality.

Measure	Q (proposed) (No reference is needed)	UQI [1] (reference is needed)	PSNR(dB) (reference is needed)
Fig. 4(a)	0.8067	0.5595	28.1384
Fig. 4(b)	0.6791	0.4918	27.1954

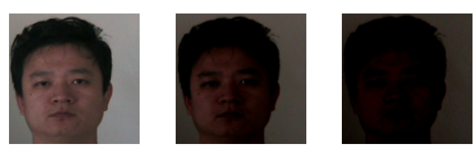
3.1.2 Experiment 2: Influence of Lighting Variation on the Super-resolved Face Image

This experiment tests the influence of lighting variation on the quality of the super-resolved face image. Fig. 5 shows three input face images with the same lighting condition. Fig. 6 shows three input face images with perceptibly different lighting conditions. Fig. 7 shows the two super-resolved face images constructed from the images in Fig. 5 and 6, respectively, and the ideal reference image (for comparison only). The images in Fig. 5(a) and 6(a) are the same and all the other input images are aligned to this image before the construction of the super-resolved image. Therefore, the reference image is the same for the two super-resolved images and it is the original face image corresponding to the image in Fig. 5(a) and 6(a). Table 3 shows the quality of the super-resolved image evaluated using the proposed quality measure, UQI and PSNR.



(a) (b) (c)

Figure 5: Three input face images with the same lighting condition.



(a) (b) (c)

Figure 6: Three input face images with different lighting conditions.

3.1.3 Experiment 3: Influence of Facial Expression Variation on the Super-resolved Face Image

This experiment tests the influence of expression variation on the quality of the super-resolved face image.

(a) (b) (c)

Figure 7: (a) The super-resolved face image from the input images in Fig. 5. (b) The super-resolved face image from the input images in Fig. 6. (c) The ideal reference image shown for comparison. It is directly obtained from the original video.

Table 3: The Effect of Light Variation on Quality.

Measure	Q (proposed) (No reference is needed)	UQI [1] (reference is needed)	PSNR(dB) (reference is needed)
Fig. 7(a)	0.8075	0.5408	28.2127
Fig. 7(b)	0.5094	0.4720	21.9090

Fig. 8 shows three input face images with the same expression. Fig. 9 shows three input face images with perceptibly different facial expressions. Fig. 10 shows the two super-resolved face images constructed from the images in Fig. 8 and 9, respectively, and the ideal reference image (for comparison only). The images in Fig. 8(a) and 9(a) are the same and all the other input images are aligned to this image before the construction of the super-resolved image. Therefore, the reference image is the same for the two super-resolved images and it is the original face image corresponding to the image in Fig. 8(a) and 9(a). Table 4 shows the quality of the super-resolved image evaluated using the proposed quality measure, UQI and PSNR.

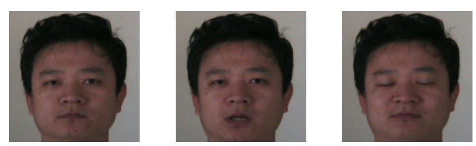
3.1.4 Experiment 4: Influence of the Number of Images Used for Constructing the Super-resolved Face Image for Face Recognition

In this experiment, the quality of the super-resolved images, which are constructed from the different number of input images, are evaluated and compared. Also the relationship between the quality and the recognition rates is examined. Ninety video sequences of 45 people are used. Each subject walks in an outdoor condition and exposes a side view to the camera. Therefore, the super-resolved side-face images are constructed. Each person has two video sequences and each video sequence includes one person. For each video sequence, we construct 10 super-resolved side-face images by using different numbers of input images. The number are 3, 5, 7, 9, 11, 13, 15, 17, 19 and 21. Fig. 11 shows sample low-resolution input side-face images of 45 people. We name 45 people from 1 to 45. Fig. 12 shows a sample of super-resolved side-face, which is constructed from nine low-resolution input images. The size of the low-resolution input images is 68×68 and the size of the super-resolved images is 136×136.

Since each of 45 people has two video sequences, we use one for training and the other one for testing. Face recognition is based on Principal Component Analysis (PCA). It compares the super-resolved face images from the training video sequences with those from the testing video sequences. Table 5 shows the average quality index of 45 people using the proposed measure and the recognition rates for these super-resolved face images. We also plot the recognition rates versus the different number of images to obtain the super-resolved images in Fig. 13(a) and the quality versus the different number of images to obtain the super-resolved images in Fig. 13(b).

(a) (b) (c)

Figure 8: Three input face images with the same expression.



(a) (b) (c)

Figure 9: Three input face images with different expressions.

(a) (b) (c)

Figure 10: (a) The super-resolved face image from the input images in Fig. 8. (b) The super-resolved face image from the input images in Fig. 9. (c) The ideal reference image shown for comparison. It is directly obtained from the original video.

3.2 Discussion

Through the four experiments, it is clear that the conditions of the input image have much effect on the quality of the super-resolved images. To achieve a super-resolved image with the improved quality, it should be constructed from multiple input images, which must be of the same object, taken from slightly different angles, but not so much as to change the overall appearance of the object in the image. The experimental results demonstrate that if the difference between the input images are too large, the quality of the super-resolved images will degrade. Among them, the variation of pose and lighting will cause much effect on the quality of the super-resolved image since they are more likely to bring the overall appearance changes of the face in the image. These global changes make the registration between different input images hard and information in the input images work as noise instead of complementary cues to each other. Comparatively, the facial expression variation is a local change of the face in the image. The overall appearance of the face will not be degraded too much since most information still works complementarily except the regions where the changes are large. Moveover, for the super-resolved image construction, it is clear that the quality will be improved with the increase of the number of input images on the condition that the information of input images are complementary but not too different to each other. Using the proposed measure, we can quantify the quality difference and, therefore, choose the appropriate number of the input images. The experimental results show the effectiveness of the proposed measure in the quality assessment of image/video super-resolution.

Table 4: The Effect of Facial Expression Variation on Quality.

Measure	Q (proposed) (No reference is needed)	UQI [1] (reference is needed)	PSNR(dB) (reference is needed)
Fig. 10(a)	0.8087	0.5485	28.0497
Fig. 10(b)	0.7925	0.5431	27.9933

4 Conclusions

In this paper, a non-reference objective measure is proposed, which aims to evaluate the quality of the super-resolved images that are constructed under different conditions. Different from the current non-reference quality measure that only uses the the relationship between the super-resolved image and the input images, the proposed quality evaluation method combines both the relationship between the super-resolved image and the input images, and the relationship between the input images.

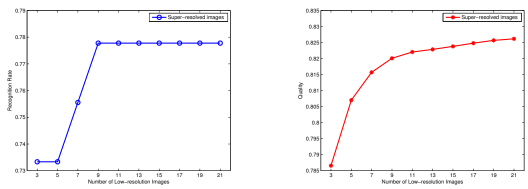
Figure 11: Sample low-resolution input face images taken from videos of 45 people. They are numbered 1 to 45 (from left to right and top to bottom).

Face recognition experiments are conducted based on the super-resolved images. Experimental results demonstrate that the variation of pose, lighting and expression have different effects on the quality of the super-resolved face images and that the recognition performance is in agreement with the quality of the tested images.

Figure 12: An example of super-resolved face images of people from number 13 to 24. They are constructed from nine low-resolution input images.

Table 5: The Effect of the Number of Input Images on the Quality of Super-resolved image.

Number of Input Images	3	5	7	9	11
Q (proposed)	0.7866	0.8070	0.8157	0.8201	0.8220
Recognition Rate	73.3%	73.3%	75.6%	77.8%	77.8%
Number of Input Images	13	15	17	19	21
Q (proposed)	0.8229	0.8238	0.8248	0.8257	0.8262
Recognition Rate	77.8%	77.8%	77.8%	77.8%	77.8%



(a) (b)

Figure 13: Results from 90 video sequences of 45 people: (a) Recognition rate vs. number of input images. (b) Quality vs. number of input images.

References

[1] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[3] Y. Cornet and M. Binard, "Which metrices to assess and compare the quality of image fusion products?" in *Remote Sensing of Land Use and Land Cover, Proc. of the EARSeL Workshop*, Dubrovnik, Croatia, 2004.

[4] C. Xydeas and V. Petrovic, "Objective image fusion performance measure," in *Electronics Letters*, vol. 36, 2000, pp. 308–309.

[5] G. H. Qu, D. L. Zhang, and P. F. Yan, "Information measure for performance of image fusion," *Electronics Letters*, vol. 38, no. 7, pp. 313–315, 2002.

[6] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. Int'l Conf. Image Processing*, 2003, pp. 173–176.

[7] V. Tsagaris and V. Anastassopoulos, "Information measure for assessing pixel-level fusion methods," in *Image and Signal Processing for Remote Sensing X. Proc. of the SPIE*, vol. 5573, 2004, pp. 64–71.

[8] N. Cvejic, A. Loza, D. Bull, and N. Canagarajah, "A novel metric for performance evaluation of image fusion algorithms," *Trans. Engineering, Computing and Technology*, vol. 7, pp. 80–85, 2005.

[9] N. Cvejic, C. Canagarajah, and D. Bull, "Image fusion metric based on mutual information and tsallis entropy," in *Electronics Letters*, vol. 42, 2006, pp. 626–627.

[10] V. Petrovic and C. Xydeas, "Sensor noise effects on signal-level image fusion performance," *Information Fusion*, vol. 4, no. 3, pp. 167–183, 2003.

[11] H. Gharavi and A. Tabatabai, "Application of quadrature mirror filtering to the coding of monochrome and color images," in *Proc. IEEE Int'l Conf. Acoustic, Speech, and Signal Processing*, 1987, pp. 2384–2487.

[12] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *J. Visual Communication and Image Representation*, vol. 4, pp. 324–335, 1993.

[13] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Trans. Medical Imaging*, vol. 22, no. 7, pp. 865–874, 2003.