Improving the Accuracy of Face Detection for Damaged Video and Distant Targets

Jun-Horng Chen
Department of Communication Engineering, Oriental Institute of Technology, New Taiepi City, Taiwan

Keywords: Error Concealment, Face Detection, Super-resolution.

Abstract: This work aims at improving the accuracy of face detection in two scenarios, when the video quality is deteriorated by the transmission link and when the target is far away from the camera. In block based coding, the packet loss inevitably makes the corrupted face image lacks some blocks. This work proposes the sparse modeling error concealment can coarsely recover the lost blocks, the fine texture can be obtained by diminishing the edge discontinuity, and a satisfied result for face detection can thus be recovered. Furthermore, this work utilizes the relationship learning super-resolution method to enhance the resolution in the case of face image taken from a long distance. Experimental results demonstrate that the proposed approach can effectively increase the accuracy of face detection for severely degraded and low resolution face images.

1 INTRODUCTION

As the continuous growth of ubiquitously installed cameras, the applications of computer vision techniques are rapidly developed. Over the past decades, face recognition has become one of the most popular biometric applications. The widespread surveillance systems encourage the development and establishment of face recognition in public area. Generally, face recognition systems are composed of two stages: detection stage and recognition stage, and are analyzed separately (Marciniak et al., 2013). That is, if the face can not be detected at the first stage, system with high accuracy of recognition will not function expectedly.

However, in some surveillance systems, the video signal is fed into the recognition system via transmission link. Therefore, the image quality is inevitably degraded by imperfect transmission, and the degraded face video definitely diminishes the accuracy of recognition. Generally in video communication, the error concealment technique which recovers the corrupted Macroblocks(MB) at the decoder site is proposed for maintenance of the visual quality. The sparse modeling error concealment (Lakshman et al., 2010) has been proven to be an effective way to enhance the visual quality. Accordingly, this work will utilize sparse modeling error concealment to recover the corrupted face images so that the face detection accuracy can thus be improved.

Furthermore, the impressive performance of face recognition system is usually measured in controlled conditions, such as ambient illumination, pose, resolution, *etc*. For example, in FRVT 2006 (Phillips et al., 2007) , the interpupillary distance (IPD) of some experiments can be as high as 400 pixels. It is the main reason for some deployments (Bonner, 2001)(Dempsey and Forst, 2010) did not meet the required accuracy. As for some successful deployments, the subject's cooperation and the controlled conditions are required and expected. Since the super-resolution (SR) process is proposed to enhance resolution image from one or multiple low resolution images, this work will utilize an effective SR approach to estimate a high resolution image from a very low resolution image which is taken by a camera located at a long distance away from target.

2 SPARSE MODELING ERROR CONCEALMENT

The sparse modeling error concealment technique which recovers the corrupted or lost blocks at the decoder site is proposed for maintenance of the image visual quality in imperfect transmission link. In contrast to the traditional error resilience techniques e.g. FEC and ARQ, the error concealment is expected to diminish the channel effect without the over-

Chen J..
Improving the Accuracy of Face Detection for Damaged Video and Distant Targets.
DOI: 10.5220/0005161603510355
In *Proceedings of the International Conference on Neural Computation Theory and Applications* (NCTA-2014), pages 351-355
ISBN: 978-989-758-054-3

head bandwidth. The authors of (Kaup et al., 2005) (Lakshman et al., 2010) used sparse modeling to extrapolated corrupted image data. By referring to the available neighbor data, the recovered data is a linear combination of a set of basis functions. The criterion of MMSE (minimizing mean-squared-error) of the available image data is used to determine the coefficient of each basis iteratively. Let $\mathbf{x} \in \mathcal{R}^N$ be an interested region, which contains a known part $\mathbf{x}_a$ and an unknown part $\mathbf{x}_b$, $\mathbf{x}$ can be represented as a linear combination of a linearly independent set $\Phi = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N\}$. That is, in the n-th iteration, the approximated $\tilde{\mathbf{x}}^{(n)}$ will be given by

$$\tilde{\mathbf{x}}^{(n)} = \sum_{\mathbf{u}_k \in \Phi^{(n)}} c_k \mathbf{u}_k \,. \tag{1}$$

Accordingly, in the $(n+1)$-th iteration, $\Phi^{(n+1)} = \Phi^{(n)} \bigcup \left\{\mathbf{u}^{(n+1)}\right\}$, where $\mathbf{u}^{(n+1)}$ is the new chosen basis, and

$$\tilde{\mathbf{x}}^{(n+1)} = \tilde{\mathbf{x}}^{(n)} + c_{n+1}\mathbf{u}^{(n+1)} \,, \tag{2}$$

where c_{n+1} is determined by minimizing the error,

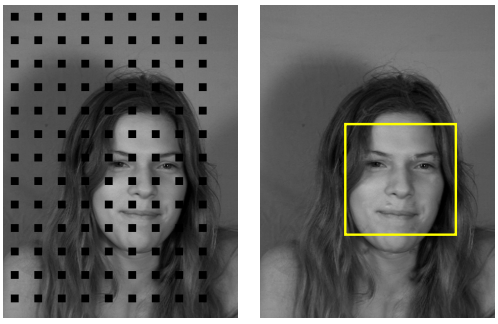
$$E^{(n+1)} = \|\tilde{\mathbf{x}}_a^{(n+1)} - \mathbf{x}_a\|^2 \,. \tag{3}$$

In (Chen, 2011), c_{n+1} can be determined by

$$c_{n+1} = \frac{\left[\tilde{\mathbf{x}}_a^{(n+1)} - \mathbf{x}_a\right]^t \cdot \mathbf{u}_a^{(n+1)}}{\left[\mathbf{u}_a^{(n+1)}\right]^t \cdot \mathbf{u}_a^{(n+1)}} \,, \tag{4}$$

and $\mathbf{u}^{(n+1)}$ can be determined by

$$\mathbf{u}^{(n+1)} = \arg\max_{\mathbf{u}_k \in \Phi} \left[E^{(n+1)} - E^{(n)}\right] \,. \tag{5}$$



(a) Corrupted Image. (b) Recovered Image.

Figure 1: (a) The face in corrupted image can not be detected. (b) The face in sparse modeling recovered image can be detected.

Figure 1 shows the sparse modeling error concealment can improve the detection accuracy when images are corrupted by transmission link. The example image is drawn from the MUCT face database (Milborrow et al., 2010). The corrupted image as shown in Fig. 1(a) is the simulation result of the image coded in H.264/AVC with *Flexible Macroblock Ordering*(FMO) error resilience technique suffers packet loss during transmission. The face detection is conducted by the oft-used *Haar Cascade Classifier*.

Although the pixels inside the corrupted image block can be estimated by sparse modeling, it is noticed that the edge across the boundary may not be continued in Fig. 1(b). This is because the MMSE approach solves Eq. (4) and (5) without consideration of edge continuity. It is proven (Chen, 2011) that the edge can be well extended across the boundary by minimizing a pre-defined cost function of discontinuity.

In (Chen, 2011), the parametric cost function was defined by the absolute magnitude difference of the gradient across the block boundary in four directions, and represented by the product of a sparse matrix $\mathbf{D}$ and a target region which is represented by a column vector. That is,

$$J(\mathbf{c}) = \|\mathbf{D} \cdot \hat{\mathbf{x}}\|^2 \,, \tag{6}$$

where $\hat{\mathbf{x}}$ is the column vector form of estimated image vector by sparse modeling error concealment. The coefficient vector $\mathbf{c}$ is the projection vector of $\hat{\mathbf{x}}$ on a set of linearly independent basis vectors Φ, that is,

$$\hat{\mathbf{x}} = \sum_{i=1}^{N} c_i \mathbf{u}_i = \Phi \cdot \mathbf{c} \,. \tag{7}$$

In this work, the 2D-DCT kernel functions are used as the basis functions. Then, by the steepest descent approach iteratively, the coefficient vector $\mathbf{c}$ is moved towards $\hat{\mathbf{c}} = \mathbf{c} + \delta\Delta\mathbf{c}$ such that the cost function J in Eq. (6) has a maximum reduction. That is, the moving vector $\Delta\mathbf{c}$ is

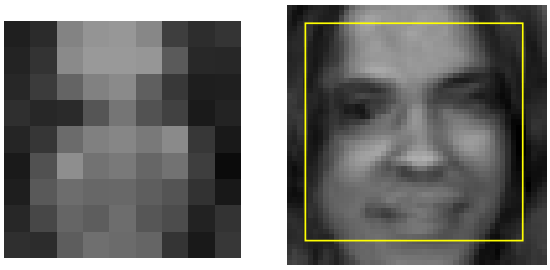
$$\Delta\mathbf{c} = -\nabla J(\mathbf{c}) \,, \tag{8}$$

and can be analytically determined by (Chen, 2011)

$$\Delta\mathbf{c} = 2(\mathbf{D} \cdot \sum_{i=1}^{N} c_i \mathbf{u}_i)^t \cdot \mathbf{D} \cdot \Phi \,. \tag{9}$$

3 RELATIONSHIP LEARNING BASED SUPER-RESOLUTION

In surveillance systems, the face image might be impoverished because the object is located at a large distance away from the camera. Generally, the oft-used face detection approach can not detect such low resolution face image whose resolution is lower than 20×20. The super-resolution based image inpainting (Meur and Guillemot, 2012) has been proven to



(a) LR Image. (b) HR Image.

Figure 2: (a) The LR face image with size of 9×9. (b) The estimated HR image with size of 50×50 .

be an effective way to estimate the missing region in an image. This work propose that the face super-resolution methods can be utilized to enhance the resolution of the images, and hopefully the detection accuracy can be accordingly improved for long distant targets. In (Wilman and Yuen, 2010), the authors improve the existing learning-based super-resolution approaches by modeling the super-resolution problem as a regression problem. By optimizing the constraint on high resolution image space, the proposed relationship learning based super-resolution provides more detailed and discriminative information, which makes the resulting face image can be more accurately detected by Haar Cascade Classifier.

In relationship learning super-resolution, a set of training HR and LR images pairs are used to determine the relationship between HR and LR pairs at the training stage. At the query stage, the relationship is accordingly used to estimate an HR face image from a given LR image. Let $\mathcal{S} = \{(\mathbf{x}_1^L, \mathbf{x}_1^H), (\mathbf{x}_2^L, \mathbf{x}_2^H), \cdots, (\mathbf{x}_N^L, \mathbf{x}_N^H)\}$ be a training set, and $\mathbf{R} \in \mathbb{R}^{m \times n}$ be the relationship matrix, the regression model can be represented as:

$$\mathbf{x}_i^H = \mathbf{R} \cdot \mathbf{x}_i^L + \mathbf{e}\,, \quad \text{for } i = 1, 2, \cdots, N\,, \tag{10}$$

where m and n are the dimensionalities of HR and LR images respectively, and $\mathbf{e}$ is the regression noise. Therefore, the relationship matrix $\mathbf{R}$ can be determined by minimizing the regression error,

$$\mathbf{R} = \arg\min_{\mathbf{R}} \sum_{i=1}^{N} \|\mathbf{x}_i^H - \mathbf{R} \cdot \mathbf{x}_i^L\|^2\,. \tag{11}$$

Equation (11) can be iteratively solved by *gradient descent* approach. In each iteration, the the relationship matrix $\mathbf{R}$ can be determined by

$$\mathbf{R}^{(n+1)} = \mathbf{R}^{(n)} - \delta \sum_{i=1}^{N} \nabla_{\mathbf{R}} \|\mathbf{x}_i^H - \mathbf{R}^{(n)} \cdot \mathbf{x}_i^L\|^2\,, \tag{12}$$

where δ is the adjustment step size and the gradient of regression error can be given by

$$\sum_{i=1}^{N} \nabla_{\mathbf{R}} \|\mathbf{x}_i^H - R \cdot \mathbf{x}_i^L\|^2 = \sum_{i=1}^{N} -2(\mathbf{x}_i^H - \mathbf{R} \cdot \mathbf{x}_i^L)\left(\mathbf{x}_i^L\right)^t\,. \tag{13}$$

As shown in Fig. 2, the estimated HR face image can be detected by Haar Cascade Classifier. In (Wilman and Yuen, 2012), the authors demonstrate the relationship based super-resolution outperforms the existing super-resolution algorithms in terms of visual quality and recognition performance. However, the region of interest should be located before resolution enhancement. Since the relationship matrix is trained from face images pairs, even non-face images may be mapped to face-like images. This work proposes an inverse verification process to filter out non-face images,

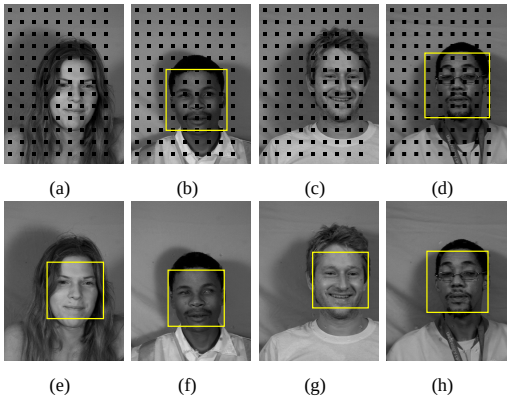
$$\|\mathfrak{T}\left(R \cdot \mathbf{x}^L\right) - \mathbf{x}^L\|^2 < \varepsilon\,, \tag{14}$$

where $\mathfrak{T}$ is the down-sampling process which reduces the resolution of the estimated HR image, and ε is a preset threshold which controls the tolerable error.

4 EXPERIMENT RESULTS

4.1 Face Detection For Corrupted Face Image

In order to verify the performance of sparse modeling error concealment, this work assumes the images in the MUCT face database (Milborrow et al., 2010) are compressed in H.264/AVC with *Flexible Macroblock Ordering*(FMO) error resilience technique. When some packets are lost, the image will lack some blocks with size of 16×16, as shown in Fig. 3(a)-3(d). It can be seen that some of corrupted face images can not be detected, the detection accuracy is 49.9% in



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

(e) (f) (g) (h)

Figure 3: (a)-(d) The corrupted images. (b) The recovered images.

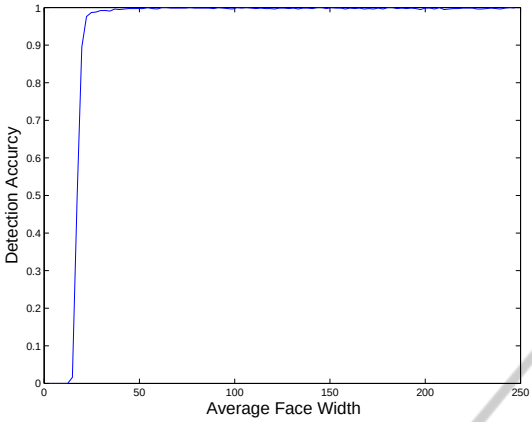


Figure 4: The accuracy of face detection with various sizes of face width.

this experiments. However, with the help of sparse modeling error concealment, the detection accuracy can be up to 99.8%.

4.2 Face Detection For Low Resolution Face Image

As shown in Fig. 4, the face images with resolution lower than 20×20 are difficult to be detected by Haar Cascade Classifier. Therefore, this work uses the images in the MUCT face database (Milborrow et al., 2010) to build the LH and HR training images pairs. Each face image detected in the original image is resized to 50×50 as the HR images, and is resized to 5×5 and 9×9 as the LR images, as shown in Fig. 5 and 6 respectively. The experimental results show that there is no face image with resolution 5×5 or 9×9 can be detected by Haar Cascade Classifier. However, by using the proposed approach, the detection accuracy can be improved to 57.26% and 97.87% respectively.

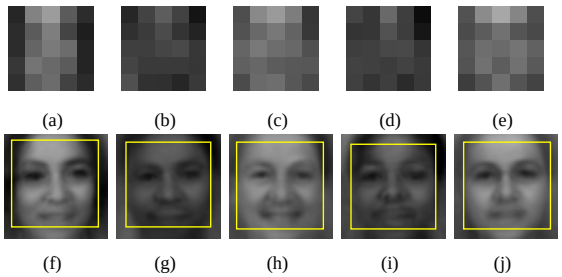


Figure 5: (a)-(e) The LR face images with size of 5×5. (b) The estimated HR images with size of 50×50.

5 CONCLUSIONS

This work proposes the approaches to improve the accuracy of face detection in two scenarios, corrupted images and low resolution images. By sparse modeling error concealment, the face images of which some blocks are lost during transmission can be also detected. The experimental results demonstrate the accuracy of detection can be significantly improved to 99.8%. Furthermore, this work proposed the relation learning super-resolution with inverse verification can effectively improve the face detection for face images with very low resolution when the objects are located in a long distance away from the camera. The experimental results demonstrate the proposed approach makes the low resolution face images be detectable with 57.26% and 97.87% of detection accuracy for face images with sizes of 5×5 and 9×9 respectively.

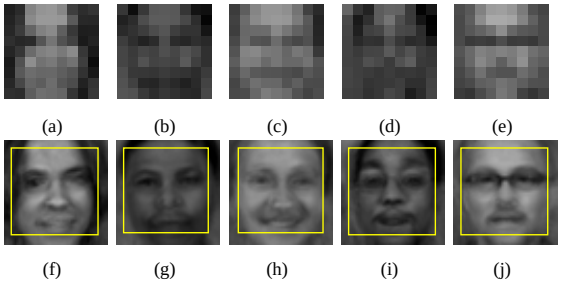


Figure 6: (a)-(e) The LR face images with size of 9×9. (b) The estimated HR images with size of 50×50.

REFERENCES

Bonner, J. (2001). Looking for faces in the super bowl crowd. *Access Control & Security System*.

Chen, J.-H. (2011). An improved error concealment by diminishing the edge discontinuity. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2213–2216.

Dempsey, J. S. and Forst, L. S. (2010). *An Introduction to Policing*. DELMAR CENGAGE Learning, 5 edition.

Kaup, A., Meisinger, K., and Aach, T. (2005). Frequency selective signal extrapolation with applications to error concealment in image communication. *AEUE - International Journal of Electronics and Communications*, 59:147–156.

Lakshman, H., Köppel, M., Ndjiki-Nya, P., and Wiegand, T. (2010). Image recovery using sparse reconstruction based texture refinement. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 786–789.

Marciniak, T., Chmielewska, A., Weychan, R., Parzych, M., and Dabrowski, A. (2013). Influence of low resolution of images on reliability of face detection and recognition. *Multimedia Tools and Applications*.

Meur, O. L. and Guillemot, C. (2012). Super-resolution-based inpainting. *Lecture Notes in Computer Science*, 7577:554–567.

Milborrow, S., Morkel, J., and Nicolls, F. (2010). The MUCT Landmarked Face Database. *Pattern Recognition Association of South Africa*. http://www.milbo.org/muct.

Phillips, J., Scruggs, W. T., OToole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., and Sharpe, M. (2007). Frvt 2006 and ice 2006 large-scale results. *NISTIR 7408*.

Wilman, W. Z. and Yuen, P. C. (2010). Very low resolution face recognition problem. In *Proceedings of Fourth IEEE International Conference on Biometrics: Theory Applications and Systems*, pages 1–4.

Wilman, W. Z. and Yuen, P. C. (2012). Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, 21.