Coupling Face Registration and Super-Resolution

Kui Jia, Shaogang Gong and Alex Po Leung
Department of Computer Science
Queen Mary, University of London
London, E1 4NS, UK
{chrisjia,sgg, alex}@dcs.qmul.ac.uk

Abstract

Existing approaches to learning-based face image super-resolution require low-resolution testing inputs manually registered to pre-aligned high-resolution training models [9, 12, 13, 5]. This restricts automatic applications to live images and video. In this paper, we propose a multi-resolution patch tensor based model to automatically super-resolve and register low-resolution testing face images. Face candidates are triggered first by a face detector giving the subwindows with their coarse initial positions and scales in a large image frame. This initialises a combined registration and super-resolution process. Rather than manually aligning each coarsely detected face subwindow to some predefined template, based on its position and scale, we scan all the potential face subwindows across different positions and scales, and obtain registration and super-resolution in a simultaneous process. The super-resolution result which is optimally correlated to its original low-resolution face subwindow is also guaranteed to be the best super-resolved reconstruction. We verify our approach by experimenting on MIT+CMU face detection dataset, the promising results demonstrate the robustness of our approach on learning-based face super-resolution on real images.

1 Introduction

Super-resolution is a technique to generate high-resolution images given a single or set of low-resolution input images. In particular, existing learning-based face image super-resolution techniques use manually pre-aligned face images to construct different high-resolution training models, and super-resolution is then performed by recovering some parameters in these training model spaces [9, 10, 12, 13, 5].

More specifically, Capel and Zisserman [12] divided human face into six unrelated parts and applied PCA on them separately. Combined with MAP estimator, they can recover the result from a high-resolution eigenface space. Baker and Kanade [9] attempted to establish the model prior based on a set of training face images pixel by pixel using Gaussian, Laplacian and feature pyramids. Freeman and Pasztor [11] tried to recover the lost high-frequency information from low-level image primitives by representing images using Markov network parameters obtained from a training data set. Liu

and Shum [13] combined the PCA model-based approach and Freeman's image primitive technique to form a mixture model. Jia and Gong [5] developed a multi-modal face image super-resolution and recognition system across different views and illuminations. They constructed two training tensors in high- and low-resolution separately, and performed multiple face image super-resolutions by the inferrence of high-resolution tensor identity parameter vectors.

However, none of the existing approaches addressed the problem of automatic face image detection and registration for super-resolution. To this end, we propose in this paper a multi-resolution patch tensor based model to automatically super-resolve and register low-resolution face images. Face image subwindows in low-resolution are first detected by a face detector therefore initialising possible candidates. Rather than manually aligning each coarsely detected face candidate to some predefined template, based on its position and scale, we scan all the potential face subwindows across different positions and scales. By coupling the registration and super-resolution in a simultaneous process, an optimised face registration and super-resolution output is yielded that is best correlated to its low-resolution face input.

The rest of the paper is organized as follows. Section 2 introduces the idea of multi-resolution patch tensor for super-resolution. The coupled face image registration and super-resolution is presented in section 3. Section 4 gives experimental results before conclusions are drawn in Section 5.

2 Multi-Resolution Patch Tensor for Super-Resolution

Super-resolution requires a suitable model for generating high-resolution images given a single or set of low-resolution images. Tensor (multilinear) analysis provides an effective means to model multiple factor interactions of an image ensemble. In particular, we introduce in this section a concept of multi-resolution patch tensor and its usage for super-resolution. In essence, we decompose uniformly multi-resolution face images into small overlapped patches, and then group these patched pixels of different positions and resolutions as an emsemble. We construct a multi-resolution patch tensor for multiple resolution face image super-resolution (hallucination). In the following, we first briefly introduce some basic properties of tensor algebra.

2.1 Basics of Tensor Analysis

Tensor (multilinear) analysis [2, 4, 3] is a general extension of traditional linear subspace analysis such as PCA or matrix SVD. Instead of modelling relations within vectors or matrices, multilinear analysis provides a means to investigate the mappings between multiple factor spaces. In the following, we denote scalars by lower-case letters $(a, b, \ldots; \alpha, \beta, \ldots)$, vectors by upper-case $(A, B, \ldots)$, matrices by bold upper-case $(\mathbf{A}, \mathbf{B}, \ldots)$, and tensors by calligraphic letters $(\mathcal{A}, \mathcal{B}, \ldots)$.

Given a tensor $\mathcal{A} \in R^{I_1 \times I_2 \cdots \times I_N}$, an element of $\mathcal{A}$ is denoted as $\mathcal{A}_{i_1 \ldots i_n \ldots i_N}$ or $a_{i_1 \ldots i_n \ldots i_N}$, where $1 \leq i_n \leq I_n$. The mode-n vectors of the Nth order tensor are the I_n-dimensional vectors obtained from $\mathcal{A}$ by varying index i_n while keeping the other indices fixed. We can unfold or flatten tensor $\mathcal{A}$ by taking the mode-n vectors as the column vectors of matrix $\mathbf{A}_{(n)} \in R^{I_n \times (I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N)}$. These tensor unfoldings provide an easy manipulation in tensor algebra and if necessary, we can reconstruct the tensor by an inverse process of mode-n unfolding.

Moreover, we can generalize the product of two matrices to the product of a tensor and a matrix. The mode-n product of a tensor $\mathcal{A}$ by a matrix $\mathbf{M}$, denoted by $\mathcal{A} \times_n \mathbf{M}$, is a tensor $\mathcal{B}$. This mode-n product of tensor and matrix can be expressed in terms of unfolding matrices as $\mathbf{B}_{(n)} = \mathbf{M}\mathbf{A}_{(n)}$ for ease of usage. In singular value decompositions of matrices, a matrix $\mathbf{D}$ is decomposed as $\mathbf{U}_1 \boldsymbol{\Sigma} \mathbf{U}_2^T$, this matrix product can also be written in terms of mode-n product as $\mathbf{D} = \boldsymbol{\Sigma} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2$. We can generalize the SVD of matrices to multilinear higher-order SVD (HOSVD). An Nth-order tensor $\mathcal{A}$ can be written as the following product

$$\mathcal{A} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times \cdots \times_N \mathbf{U}_N, \tag{1}$$

where $\mathbf{U}_n$ is a unitary matrix, and $\mathcal{Z}$ is the core tensor having the property of all orthogonality, that is, two subtensors $\mathcal{Z}_{i_n=\alpha}$ and $\mathcal{Z}_{i_n=\beta}$ are orthogonal for all possible values of n, α and β subject to $\alpha \neq \beta$. The HOSVD of a given tensor $\mathcal{A}$ can be computed as follows. The mode-n singular matrix $\mathbf{U}_n$ can directly be found as the left singular matrix of the mode-n matrix unfolding of $\mathcal{A}$, afterwards, based on the product of tensor and matrix, the core tensor $\mathcal{Z}$ can be computed by $\mathcal{Z} = \mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_N \mathbf{U}_N^T$. Eq. (1) gives the basic representation of multilinear model. If we investigate the mode-n unfolding and folding, and rearrange Eq.(1), we can have

$$\mathcal{S} = \mathcal{B} \times_n V_n^T, \tag{2}$$

where $\mathcal{S}$ is a subtensor of $\mathcal{A}$ corresponding to a fixed row vector V_n^T of singular matrix $\mathbf{U}_n$, and

$$\mathcal{B} = \mathcal{Z} \times_1 \mathbf{U}_1 \cdots \times_{n-1} \mathbf{U}_{n-1} \times_{n+1} \mathbf{U}_{n+1} \cdots \times_N \mathbf{U}_N.$$

This expression is the basis for recovering original data from tensor structure. If we index into basis tensor $\mathcal{B}$ for more particular V_n^T, we can get different modal sample vector data.

2.2   Super-Resolution Using Multi-Resolution Patch Tensor

A tensor structure provides a powerful mechanism to model interaction of image ensembles with different resolutions. Benefiting from the mapping relations of multiple factor spaces inherently embedded in the tensor structure, we can recover higher resolution images given any corresponding lower resolution images. More precisely, given a training dataset of high-resolution images, of which they bear some common properties of pixel distributions, as shared by all human faces. We blur and sub-sample these high-resoluiton images with different Gaussian filters and sub-sampling factors, while keeping the image size unchanged. Furthermore, in order to perform patch-level analysis, we decompose these face images into small overlapped patches uniformly. We then obtain a hierarchical ensemble containing patch images of multiple resolutions. With these training images in place, we construct a tensor structure and use HOSVD for decomposition, expressed as

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_{idens} \times_2 \mathbf{U}_{resos} \times_3 \mathbf{U}_{patches} \times_4 \mathbf{U}_{pixels},$$

where tensor $\mathcal{D}$ groups these training patch images of multiple resolutions into a tensor structure, and the core tensor $\mathcal{Z}$ governs the interactions between the 4 mode factors. The mode matrix $\mathbf{U}_{idens}$ spans the parameter space of identities for these training images, the mode matrix $\mathbf{U}_{resos}$ spans the parameter space of different resolutions, the mode matrix

$\mathbf{U}_{patches}$ spans the parameter space of patch positions, and the mode matrix $\mathbf{U}_{pixels}$ spanning space of image pixels.

With this constructed tensor, we can perform super-resolution in a tensor parameter vector space. Based on the tensor properties given in Section 2.1, specifically as suggested in Eq. (2), image patches of different resolutions can be recovered given their tensor space single identity parameter vector on each patch. This single identity parameter vector can be computed by projecting testing resolution patches onto the multi-resolution patch tensor. More precisely, suppose we have a basis tensor

$$\mathcal{B} = \mathcal{Z} \times_2 \mathbf{U}_{resos} \times_3 \mathbf{U}_{patches} \times_4 \mathbf{U}_{pixels}, \tag{3}$$

we can index into this basis tensor at a particular resolution r and patch position p to yield a basis subtensor

$$\mathcal{B}_{r,p} = \mathcal{Z} \times_4 \mathbf{U}_{pixels} \times_2 V_r^T \times_3 V_p^T.$$

Then the subtensor containing the individual patch data can be expressed as

$$\mathcal{D}_{r,p} = \mathcal{B}_{r,p} \times_1 V^T + \mathcal{E}_{r,p}, \tag{4}$$

where V^T represents the single identity parameter row vector and $\mathcal{E}_{r,p}$ stands for the tensor modelling error for resolution r and patch position p. For ease of notation and readability, we will use the mode-1 unfolding matrix to represent tensors. Then the matrix representation of Eq.(4) becomes

$$\mathbf{D}_{r,p}^{(1)} = V^T \mathbf{B}_{r,p}^{(1)} + e_{r,p}. \tag{5}$$

Eq.(5) provides a possible solution for the single identity parameter vector V^T. Applied it on higher resolution r', the corresponding resolution patch data can be computed as

$$\mathbf{D}_{r',p}^{(1)} = V^T \mathbf{B}_{r',p}^{(1)} + e_{r',p}. \tag{6}$$

After recovering all the patch data at higher resolution r', we compose them together to form the higher resolution face image.

This multi-resolution patch tensor mechanism provides an attractive feature in super-resolution. Since the training tensor incorporates information of multiple resolution face images, given a low-resolution testing input, a hierarchical structure of multiple higher resolution face images can be simultaneously reconstructed. Even more significant from a practical usefulness perspective, the highest resolution reconstruction can be obtained given any scale in this hierarchical resolution structure as a low-resolution probe input.

3 The Coupled Registration and Super-Resolution

The multi-resolution patch tenor for super-resolution uses training face images which have been manually warped to some canonical form, this is unrealistic for automatic super-resolution of detected candidate inputs in live images or video. In reality, automatically detected potential face images are often mis-aligned and even false positives. This is especially true in low-resolution inputs due to the lack of information for feature extraction and accurate alignment, one of the primary causes for poor face recognition performance in an uncontrolled CCTV environment from a distance. To address the problem, we present here a tensor super-resolution model in a Bayesian framework to perform coupled face image registration and super-resolution given low-resolution input candidates, resulting in an optimised face alignment and super-resolution simultaneously.

3.1 Super-Resolution in a Bayesian Approach

Suppose that H is the high-resolution image to be recovered, and L is the low-resolution input of potential faces. Then the problem of super-resolution can be formulated into a Bayesian framework, and the task comes as finding the Maximum A Posterior (MAP) estimation of H given L:

$$H_{MAP} = \arg \max_{H} \log P(H|L) \tag{7}$$

By applying the Bayes rule, the probability $P(H|L)$ becomes $P(H|L) = P(L|H)P(H)$, where $P(L|H)$ is the conditional probability modelling the relations between low- and high-resolution images and $P(H)$ is the high-resolution prior. Since we decompose face images into small overlapped patches, the inference of image H will be carried out at a patch level. We factorize the likelihood $P(L|H)P(H)$ at the patch level and it becomes

$$P(L|H)P(H) = \prod_{p=1}^{N} P(L_p|H_p)P(H_p)$$

Assume that $\mathbf{A}$ is a blurring and sub-sampling operator connecting L_p and H_p in an imaging observation model, and are Gaussian, so as its prior probability. Then

$$P(L|H)P(H) =$$
$$\prod_{p=1}^{N} \left\{ \frac{1}{Z} \exp \left(-\frac{\|\mathbf{A}H_p - L_p\|^2}{\lambda} \right) \frac{1}{Z'} \exp \left(-\left(H_p - \mu_{H_p}\right)^T \Lambda^{-1} \left(H_p - \mu_{H_p}\right) \right) \right\} \tag{8}$$

where Z and Z' are normalization constants, λ scales the variance, and Λ is the covariance matrix for all the training high-resolution face images.

Based on Eq.(4) and Eq.(5), we know that the subtensor containing the pixel data at a particular resolution r and patch position p can be approximated as $\mathcal{D}_{r,p} = \mathcal{B}_{r,p} \times_1 V^T$, and the unfolded matrix representation becomes $\mathbf{D}_{r,p}^{(1)T} = \mathbf{B}_{r,p}^{(1)T} V$. Similarly we can obtain a subtensor for resolution r' of the same patch postion, which is $\mathbf{D}_{r',p}^{(1)T} = \mathbf{B}_{r',p}^{(1)T} \widetilde{V}$. Suppose $\mathbf{D}_{r,p}^{(1)T}$ and $\mathbf{D}_{r',p}^{(1)T}$ correspond to the L_p and H_p respectively, we replace them in Eq.(8) resulting in

$$P(L|H)P(H) = \prod_{p=1}^{N} \left\{ \frac{1}{Z} \exp \left(-\frac{\|\mathbf{A}\mathbf{B}_{r',p}^{(1)T} \widetilde{V} - \mathbf{B}_{r,p}^{(1)T} V\|^2}{\lambda} \right) \right.$$
$$\left. \frac{1}{Z'} \exp \left(-\left(\mathbf{B}_{r',p}^{(1)T} \widetilde{V} - \mu_{H_p}\right)^T \Lambda^{-1} \left(\mathbf{B}_{r',p}^{(1)T} \widetilde{V} - \mu_{H_p}\right) \right) \right\} \tag{9}$$

We optimize paramter $\widetilde{V}$ based on the constructing properties of our multi-resolution patch tensor, which suggest that the relation between $\mathbf{B}_{r',p}^{(1)T}$ and $\mathbf{B}_{r,p}^{(1)T}$ observes a basic imaging observation model. In reality, this is consistent with the uniqueness of the identity parameter vector on each patch. By setting $\widetilde{V} \equiv V$, we can approximate $H_p \approx \mathbf{B}_{r',p}^{(1)T} \mathbf{\Psi} L_p$ where $\mathbf{\Psi}$ is the pseudoinverse of $\mathbf{B}_{r,p}^{(1)T}$ and is equal to $(\mathbf{B}_{r,p}^{(1)} \mathbf{B}_{r,p}^{(1)T})^{-1} \mathbf{B}_{r,p}^{(1)T}$. After reconstructing all the patches at different positions, the final high-resolution reconstructions are simply composition of their corresponding overlapped patches.

The first component in Eq.(9) guarantees that the optimized parameter $\widetilde{V}$ can reconstruct a high-resolution face image which when subsampled, will be consistent with the low-resolution input governed by the basic imaging model. While the input images are not faces, e.g. some random patch images, the second component in Eq.(9), which holds a priori information about canonically warped high-resolution training face images, assures that a higher frequency layer of face images will be hallucinated and added to the original random low-resolution inputs. This phenomena is illustrated in Figure 1, and provides an effecient matching metric in next section of correlated-based registration by super-resolution. .
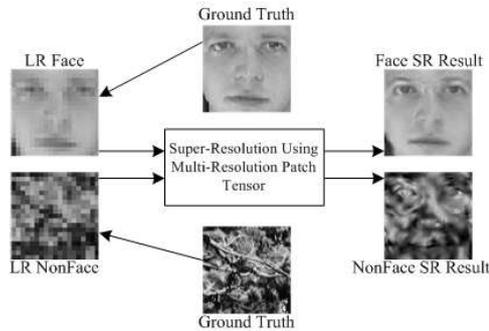

Figure 1: Comparison of face and nonface super-resolution using training face image multi-resolution patch tensor.

3.2 Correlation-based Registration by Super-Resolution

The registration process is initialized by the automatic face detector, which provides potential face subwindows to be super-resolved. As described in Section 2.2, our training multi-resolution patch tensor has a hierarchical resolution structure, and the high-resolution reconstruction can be obtained given any lower resolution scale as the testing input. Assume the potential face subwindow is W, we find the closest resolution r_W in this hierarchical structure and project on training tensor, its super-resolution result HW can be computed on patch level as

$$HW_p = \mathbf{B}_{r_{HW},p}^{(1)T} \, (\mathbf{B}_{r_W,p}^{(1)} \mathbf{B}_{r_W,p}^{(1)T})^{-1} \mathbf{B}_{r_W,p}^{(1)T} \, W_p, \qquad (10)$$

where r_{HW} is the resolution of the desired reconstruction output.

In reality this potential face subwindow may have a variety of possible sizes (resolutions), and its position and alignment are not necessarily accurate enough as required as in the canonically warped training face images, which make the super-resolution result poor in quality as compared to the high-resolution training face images. To alleviate these problems, based on the initial subwindow given by the automatic face detector, we enlarge or diminish its window size, and translate it across different positions within some predefined range. That subwindow with the best position and size for super-resolution is chosen as the optimization.

In this process, the metrics for subwindow selection is another important issue. Rather than matching the potential low-resolution subwindow to some standardly aligned face template, we integrate the matching process as a step of optimized super-resolution.

Specifically, assume $\widetilde{W}$ is the optimization in the subwindow selection process, and it has the perfect alignment to the training face images of multi-resolution patch tensor, we use it as the low-resolution face input and super-resolve it. That means the super-resolved high-resolution reconstruction has no geometric transformation with respect to this optimized subwindow input. To match the high- and low-resolution face images without geometric warping, we can safely neglect any estimation bias of image registration across different resolutions [6, 7], and choose simple correlation-based matching criteria. The correlation coefficient ($\widetilde{CC}$) between $\widetilde{W}$ and its high-resolution reconstruction $\widetilde{HW}$ is computed as:

$$\widetilde{CC} = \frac{\sum \left(\widetilde{W} - E(\widetilde{W}) \right) \left(\widetilde{HW} - E(\widetilde{HW}) \right)}{\sqrt{\sum \left(\widetilde{W} - E(\widetilde{W}) \right)^2} \sqrt{\sum \left(\widetilde{HW} - E(\widetilde{HW}) \right)^2}} \tag{11}$$

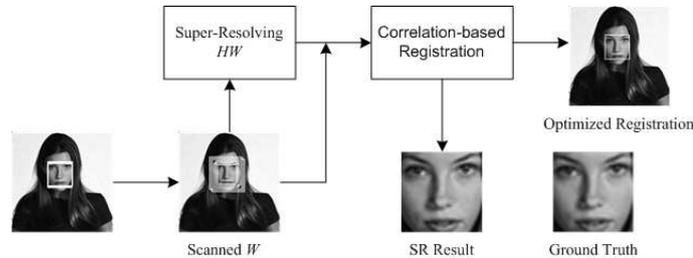where $E(\cdot)$ is the mean operator for pixel intensity values.

Figure 2: Illustration of the registration process by correlation-based matching.

This correlation-based matching criteria provides an efficient way to optimize the face subwindow registration process. Based on the initialized subwindow, we translate and resize it across different positions and scales, we project all these potential face subwindows W onto the multi-resolution patch tensor, and super-resolve their corresponding high-resolution reconstructions HW. We compute each correlation coefficient CC of W and HW, the one with the highest score is taken as the optimized registration, and its high-resolution reconstruction is natually the super-resolution result. We illustrate this process in Figure 2 with associated pseudo code shown in Algorithm 1.

Algorithm 1: Algorithm for face registration by super-resolution

input : Initialized face subwindow W by automatic face detector

output: Opitimized super-resolution result $\widetilde{HW}$

repeat

    I:

    Scan the different potential face subwindows based on W, and get the high-resolution reconstruction HW as a patch composition of
$HW_p = \mathbf{B}_{r_{HW},p}^{(1)T} \cdot (\mathbf{B}_{r_W,p}^{(1)} \mathbf{B}_{r_W,p}^{(1)T})^{-1} \mathbf{B}_{r_W,p}^{(1)T} \cdot W_p,$

    II:

    For each potential W and reconstruction HW, compute its correlation coefficient CC using Eq.(11),

until *The highest $\widetilde{CC}$ obtained* ;

---

4 Experiments

We used face images from a subset of AR, FERET and Yale databases to build our training multi-resolution patch tensor. Originally these training face images have different sizes with the centre of faces varying considerably. We aligned them manually by marking the locations of 3 points: the centers of the eyeballs and the lower tip of the nose. These 3 points define an affine warp, which was used to warp the images into a canonical form. These canonical images have a high-resolution of 76×76 pixels. We blurred and sub-sampled them with 4 different point spread functions and ratios to build a hierarchical structure of 5 layers of different image resolutions with the other four in 57×57, 38×38, 27×27, and 19×19 respectively. We then bicubically interpolated all these hierarchical face images to the resolution of 76×76 and decomposed each of them into 625 small 3×3 patches which overlapped horizontally and vertically with each other by 1 pixel (the patch size and overlapping size were experimentally decided). Given regularly decomposed patches on the manually aligned face images in this 5 layer hierarchical structure, we built the training multi-resolution patch tensor to be used later for super-resolution.
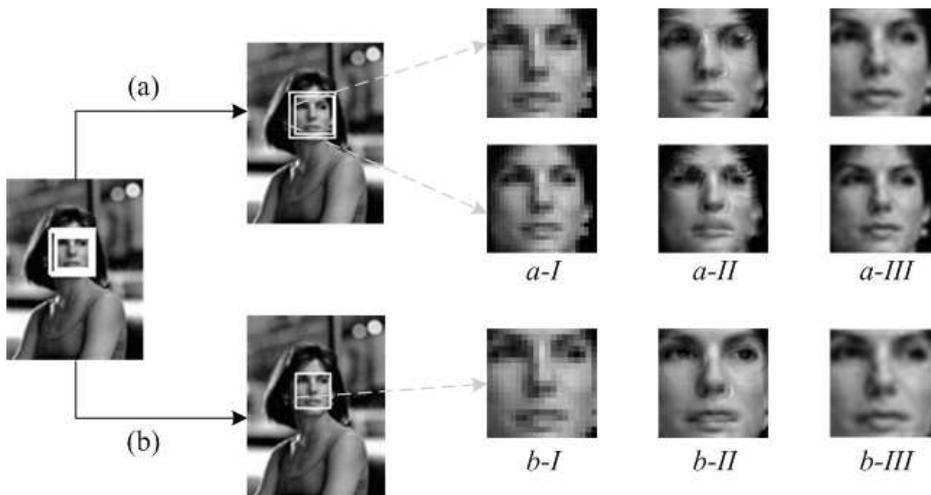

Figure 3: Experiments on face detection, registration and super-resolution: (a) Registration by averaging positions and scales of subwindows initialized by a AdaBoost face detector. Column a-I shows low-resolution face input of 20×20 and 24×24 pixels. Column a-II shows their super-resolution output of 76×76 pixels whilst column a-III gives the ground truth with resolutions in 30×30 and 36×36 respectively. (b) Coupled registration and super-resolution. b-I is low-resolution input of 19×19, b-II is 76×76 super-resolution result, and b-III is 28×28 ground truth. Note that the ground truth resolutions are smaller than those super-resolved output.

We performed our face image detection, registration and super-resolution experiments on the MIT+CMU dataset, and used AdaBoost face detector [8] to initialize this process.

Some of the original testing images in MIT+CMU dataset contain faces of higher resolutions, we subsampled them to make sure the resolutions of central faces contained in those testing images were ranging below 76×76 to even 15×15. We applied AdaBoost face detection to testing images and obtained candidate low-resolution face subwindows, we then applied our coupled face image registration and super-resolution algorithm to yield optimized subwindows (location and size). Figure 3 shows a comparison between Process-(a) where low-resolution multiple face subwindow candidates detected by AdaBoost were averaged directly to yield a final selection (in column a-I), and Process-(b) where the candidate subwindows were processed by a coupled face registration and super-resolution. It is clear that the latter corrected mis-alignment in the high-frequency occurred in the direct process obtained in column a-II.
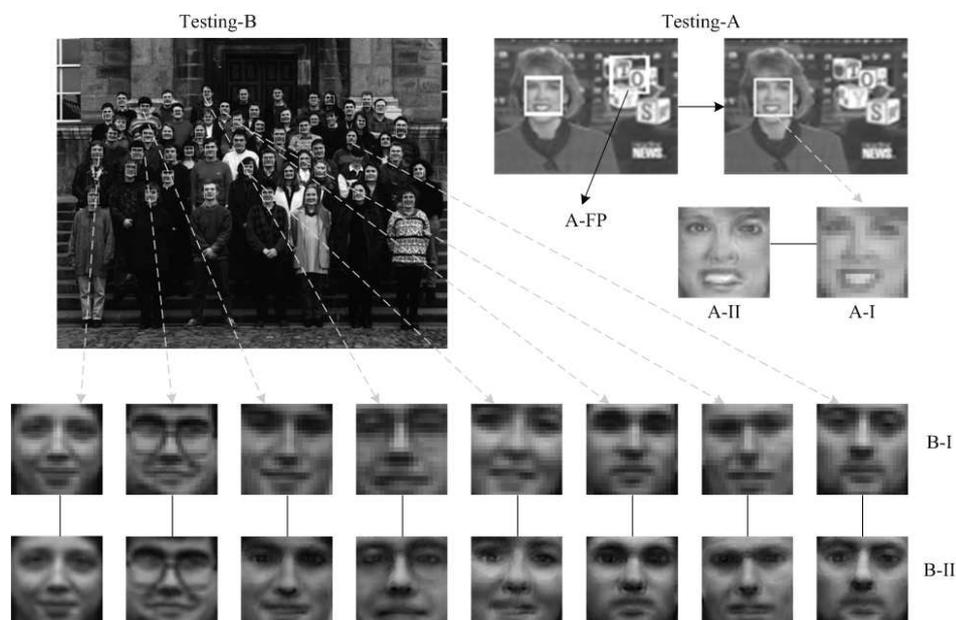

Figure 4: More experimental examples on face image registration and super-resolution. Testing-A: A-FP is false positive detections given by AdaBoost detector which were then filered out by our coupled registration and super-resolution model. A-I is the 20×20 optimally registered low-resolution input and A-II is a super-resolved result in 76×76. Testing-B: Row B-I shows optimized low-resolution example inputs resulting from our coupled registration and super-resolution process, with resolution ranging from 18×18 to 30×30. Row B-II shows the corresponding super-resolution results in 76×76.

We present more experimental results in Figure 4, in which the Testing-A image had false positive subwindows A-FP provided by Adaboost face detector. During our registration and super-resolution process, the matching correlation coefficients between A-FP and their super-resolution reconstructions were very small, by experimentally setting a correlation threshold between face and nonface candidates, we can filter out most of the false positive results passed from Adaboost face detector.

5 Conclusion and Future Work

In summary, we present in this paper a coupled automatic face image registration and super-resoltion algorithm. Given low-resolution candidate face subwindows initialised by a face detector such as AdaBoost, our model scans the subwindows across different positions and scales and perform super-resolution on them using multi-resolution patch tensor. The super-resolution result which is best correlated to a candidate face subwindow gives the optimised registration output. The registration process in our current work only concerns with zooming and translation without taking into account any geometric transformation. In future work, we will introduce a facial deformation model for automatic registration and to further improve the performace of super-resolution on real images.

References

[1] M.A.O. Vasilesescu, D. Terzopoulos, " Multilinear image analysis for facial recognition", *Proc. of International Conf. on Pattern Recognition*, 2002.

[2] M. A. O. Vasilescu, D. Terzopoulos, "Multilinear analysis of image ensembles: TensorFaces", *Proc. 7th European Conference on Computer Vision*, 2002.

[3] T.G.Kolda, "Orthogonal tensor decompositions", *SIAM Journal on Matrix Analysis and Applications*, Vol.23, pp. 243-255, 2001.

[4] L.D.Lathauwer, B.D.Moor, and J.Vandewalle, "Multilinear Singular Value Tensor Decompositions", *SIAM Journal on Matrix Analysis and Applications*, Vol.21, No.4, pp.1253-1278, 2000.

[5] K.Jia and S.Gong, "Multi-modal tensor face for simultaneous super-resolution and recognition", *Proc. IEEE International Conference on Computer Vision*, Oct, 2005.

[6] Y. Dufournaud, C. Schmid and R. Horaud, "Matching Images with Different Resolutions", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1612-1618, 2000.

[7] G Dedeoglu, S. Baker and T. Kanade, "Resolution-Aware Fitting of Active Appearance Models to Low-Resolution Images", *to appear in European Conf. on Computer Vision*, May, 2006.

[8] P. Viola and M. J. Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision*, vol. 57, No. 2, pp. 137-154, 2004.

[9] S. Baker and T. Kanade, "Limits on super-resolution and how to break them", *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*,June 2000.

[10] S. Baker and T. Kanade, "Hallucinnating Faces", *Proc. of IEEE Automatic Face and Gesture Recognition*, pp.83-90, March 2000

[11] W. Freeman and E. Pasztor, "Learning low-level vision", *7th International Conference on Computer Vision*, pp. 1182-1189,1999.

[12] D. P. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models", *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[13] C. Liu, H. Shum and C. Zhang, "A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model", *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp 192-198, 2001.