Face Super Resolution in Reduced Spaces by Using Shape and Texture

Aydın Akyol
Istanbul Technical University, Turkey
akyolayd@itu.edu.tr

Muhittin Gökmen
Istanbul Technical University, Turkey
gokmen@itu.edu.tr

Abstract

The problem of inferring a missing face image which is at much higher resolution from lower observations is called as Face Super Resolution or Hallucination problem. Mostly the problem is approached in spatial domain by using the aligned textural information of the observation. However the ignorance of the shape information limits the performance of these approaches. In Resolution Aware Fitting (RAF) algorithm it was successfully shown that superior results could be obtained by utilizing both shape and texture components together. Though the RAF algorithm provides more satisfactory results, warping and deformation operations on high resolution image during the optimization could undermine its effectiveness in real world applications. As a remedy in this work we propose a faster alternative by effectively transforming the problem into reduced dimensions and making image warping only at low resolution. Experimentally it was shown that better reconstructions could be obtained faster than the RAF algorithm.

1. Introduction

High Resolution (HR) images are critical for image analysis and the posterior applications using this analysis. However it is known that the optics of an imaging system limits the amount of information that is received by the imager device and the imaging system yields blurred and under-sampled images. At that point Super Resolution (SR) techniques are used to overcome the limitations of imaging systems.

A forward model is assumed to represent the image formation and the most common form used for this purpose is

$$I_L = HI_H + n \tag{1}$$

where I_H denotes the HR image and I_L is the deformed Low Resolution (LR) version of I_H under the deformation operator H, which presumably consists of blurring B and decimation D operators; $H = DB$. Also n represents the observation gap in formation. In SR problem it is intended to approximate the inverse of this forward. Reaching to the exact backward model would not be possible due to the ill-posed nature of the deformation H. SR techniques approximate to the exact solution by regularizing the Least Squares (LS) solution, $|I_L\text{-}HI_H|_2$. This generic approach can be expressed in Bayesian formulation as the MAP estimation of I_H

$$\hat{I}_H = \arg\max_{I_H} \{ p(I_L \mid I_H) p(I_H) \}. \tag{2}$$

The first term $p(I_L|I_H)$ refers to the LS solution (called also as Maximum Likelihood solution – ML solution) and $p(I_H)$ defines the a priori information. Depending on the needs, this generic problem definition can be restricted by making assumptions on these components of the problem. In order to align with the right literature, it is important to state the exact problem setup under consideration. In this work we assume that the image domain is restricted to the face images and the deformation operator H is known. This problem setup is also known as Face Hallucination problem in literature. Hallucination is first declared by Baker & Kanade [6] to describe the problem of inferring a missing face image which is at much higher resolution from LR observations. Within the scope of this work Hallucination definition is restricted to the case where single observation exists.

Image analysis in constraint domains, such as face, high frequency components (or called as facial details) are critically important. Minor errors on these details might be significant both for human and machine perception. It is expected that an effective face hallucination technique can bring enough high frequency content to maximize the identity of the subject under processing. Basing on the fact that the LS solution could mostly provide the low frequencies, the high frequency content could only be gained via regularization.

In literature general tendency is to benefit from the textural priors in order to regularize the solution. Though there are plenty of regularizers proposed [9], here we are contended with mentioning only on a few good representatives which are not only successful but also close to our proposal.

In [2] Gunturk et.al. define one of these successful regularizers. The subspace projection statistics of the texture data is used as the prior information. Though it is possible to use other projection techniques as in [5], due to its computational simplicity PCA is preferred. In [1] Liu et.al. state that subspace statistics would bring only the mid frequencies, and in order to add higher frequencies more customized constraints are required. In addition to the subspace projection statistics they use also a Markov Random Field (MRF) in order to define a joint locality model. Though wealthier content can be obtained with this non-parametric step, the results suffer from unrealistic texture caused by global discontinuity. This experiment shows that even restricted image domains could have excessive variety which could not be represented by even complex locality models.

As an alternative to texture models, a relatively new trend in literature is to utilize the shape information in

addition to the texture. The main information that could be extracted from the image data are shape and texture.

The texture-centric approaches assume that the shape information is known beforehand and they obtain the texture information in both resolutions by aligning the image on to a common ground via this shape information. Another assumption in texture-centric approaches is the equivalence of the shape information in both resolutions. However in most scenarios also the LR shape information may require improvement during the estimation of HR texture. The excessive deformation of the texture would create deviations on the shape information. So as in [4] the shape and texture information of the LR observation is used simultaneously in order to estimate their HR counterparts. The Resolution Aware Fitting (RAF) algorithm, described in [4], can be considered as the state-of-the-art among these new approaches. The RAF algorithm is a generative approach and iteratively estimates the synthesis parameters. In order to avoid the effects of asymmetry [8] the cost function is defined on the LR image space. In other words the synthesis is first warped back in HR and then deformed as

$$(\hat{s},\hat{t}) = \arg\min_{s,t} \left\{ \left| I_L - HW(M_H t_H, T_H^{-1}(N_H s_H)) \right|_2^2 \right\} \qquad (3)$$

where s_H and t_H are subspace representations of shape and texture components, N_H and M_H are subspace transform operators, W refers to image warping, and T_H is the spatial mapping operator which defines the correspondence between the image shape information and the mean shape. The satisfactory results of RAF algorithm show that the use of shape and texture information in a generative way may bring significant high frequency content into the solution.

In this work we propose a faster alternative for the RAF algorithm following the same principles: "utilize both shape and texture information", "use a generative structure in order to obtain realistic high frequency content". Since real world scenarios always quest for fast techniques, the use of RAF algorithm could be problematic in real-world cases due to iterative deformation and warping operations on spatial domain HR image. In order to relieve this computational load we suggest working fully in subspace by using quadratic structures. Especially when quadratic structures are chosen, the solution can be reached analytically.

2. Approach

First we provide the decomposition of the images and then the generic SR approach (2) is re-defined individually for each component of the decomposition. Later we transform these relations into subspaces by using an effective linear transformation and the relations are re-organized in order to represent the backward model in terms of subspace variables. Note also that these steps are followed individually for both low and high resolutions.

If we use I_H and I_L to represent spatial domain images in LR and HR, then their decompositions in terms of shape X and texture G components can be given as

$$I_L = X_L + G_L \quad , \quad I_H = X_H + G_H \qquad (4)$$

The reflection of this decomposition on forward and backward relations can be given for the texture component as

$$G_L = H_G G_H + n_G \qquad (5)$$

$$\hat{G}_H = \arg\max_{G_H} \{ p(G_L \mid G_H) p(G_H) \} \qquad (6)$$

and similarly for the shape component they will be

$$X_L = H_X X_H + n_X \qquad (7)$$

$$\hat{X}_H = \arg\max_{X_H} \{ p(X_L \mid X_H) p(X_H) \} . \qquad (8)$$

After the decomposition, now these components are transformed onto reduced spaces and these relations are re-given in terms of subspace representations of the components. In order to transfer the components into subspaces we prefer PCA. Though it is also possible to use other types of transformations [5], PCA is not only advantages in terms of theoretical and computational simplicity but also there are effective techniques, such as Active Appearance Model (AAM) [3], which automatically utilize these models on the input. We use individual AAMs for each resolution and obtain the 4 PCA transformations, for each component in each resolution. When M_L and M_H are used to denote the textural transformations in different resolutions and similarly N_L and N_H are used for shape components, then the component projections can be given as

$$\begin{aligned} G_H &= M_H t_H + \bar{G}_H + e_H \quad G_L = M_L t_L + \bar{G}_L + e_L \\ X_H &= N_H s_H + \bar{X}_H + \varepsilon_H \quad X_L = N_L s_L + \bar{X}_L + \varepsilon_L \end{aligned} \qquad (9)$$

where s's and t's are subspace representations, $\bar{G}$'s and $\bar{X}$'s are means, and (e,ε)'s are representational gaps. Transformation of the forward models of the components; given in (6) and (8); into subspace can be obtained by putting the projections of (9) on (5) and (7). The resulting forward models in terms of subspace projections are

$$\begin{aligned} t_L &= M_L^T H_G M_H t_H + M_L^T v_T \\ s_L &= N_L^T H_X N_H s_H + N_L^T v_S \end{aligned} \qquad (10)$$

where $v_T = H_G e_H + n_G$ denotes the total error in texture formation and similarly $v_S = H_X \varepsilon_H + n_X$ is the total gap in shape formation. Note that in (10) it is assumed that the errors (v_T and v_S) are orthogonal to the transform domains. After the transformation of the forward models, the backward models of components, (6) and (8), are re-defined in terms of subspace representations as follows;

$$\hat{t}_H = \arg\max_{t_H} \{ p(t_L \mid t_H) p(t_H) \} \qquad (11)$$

$$\hat{s}_H = \arg\max_{s_H} \{ p(s_L \mid s_H) p(s_H) \} .$$

The ML solutions, $p(t_L|t_H)$ and $p(s_L|s_H)$, are approximated by the probability distribution of the projected total errors, v_T and v_S,

$$M_L^T v_T = t_L - M_L^T H_G M_H t_H \qquad (12)$$
$$N_L^T v_S = s_L - N_L^T H_X N_H s_H .$$

It is assumed that both noise terms have Gaussian forms,

$$p(t_L \mid t_H) \sim N(M_L^T \mu_{v_T}, M_L^T \Sigma_{v_T}^{-1} M_L) \qquad (13)$$
$$p(s_L \mid s_H) \sim N(N_L^T \mu_{v_S}, N_L^T \Sigma_{v_S}^{-1} N_L)$$

and the parameters of these models are estimated from the sample statistics of the training data. The remaining terms of (11) are $p(t_H)$ and $p(s_H)$. We use projection statistics of the samples to model these regularization terms. For simplicity again it is assumed that these projected samples constitute a Gaussian form as follows.

$$p(t_H) \sim N(\mu_T, \Sigma_T) \qquad (14)$$
$$p(s_H) \sim N(\mu_S, \Sigma_S)$$

Since we model all the terms in (11) quadratic ally, the solution is obvious and can be expressed analytically

$$\hat{t}_H = \frac{\left[\left(M_L^T H_G M_H\right)^T M_L^T \Sigma_{v_T}^{-1} M_L \left(M_L^T H_G M_H\right) + \left(\Sigma_T^{-1}\right)\right]^{-1}}{\left[\left(t_L{}^T - M_L^T \mu_{v_T}\right)^T M_L^T \Sigma_{v_T}^{-1} M_L \left(M_L^T H_G M_H\right) + \left(\mu_T^T \Sigma_T^{-1}\right)\right]} \qquad (15)$$

$$\hat{s}_H = \frac{\left[\left(N_L^T H_X N_H\right)^T N_L^T \Sigma_{v_S}^{-1} N_L \left(N_L^T H_X N_H\right) + \left(\Sigma_S^{-1}\right)\right]^{-1}}{\left[\left(s_L{}^T - N_L^T \mu_{v_S}\right)^T N_L^T \Sigma_{v_S}^{-1} N_L \left(N_L^T H_X N_H\right) + \left(\mu_S^T \Sigma_S^{-1}\right)\right]}.$$

2.1 Shape and Texture Deformation Operators

During our derivations we assumed that the deformation model of the image components, H_G and H_X, are given. As recall from Section 1, actually only the image deformation operator H is given, but nothing told about its decomposition: $H = H_G + H_X$. The decomposition of the deformation operator is made as parallel to the image decomposition.

H is assumed same for all images and not dependent to the input, I_H. Also, the shape component H_X shares this characteristic of H. On the other hand H_G is different and has dependency on the input. Because the texture information requires warping and warping is determined by the shape information. So H_G is structured with the HR shape information of X_H. Considering this dependency it is possible to approximate to the corrected form of texture deformation operator as

$$H_G \cong W(W(H, T_H), T_L) \qquad (16)$$

where first the rows H are warped and then the columns of the intermediate operator are warped. Since H is same for the whole solution space and X_H varies in a fixed interval, alternative H_G's can be calculated offline and stored beforehand easily. The space complexity of this operation would be quite low since H is highly sparse.

3. Experiments

Results of the experiments are demonstrated both quantitatively and qualitatively. Quantitative results are critical in order to evaluate the method for machine perception. For quantitative comparisons the distances to the actual subspace representations of the shape and texture components are used. On the other hand qualitative results are good to make evaluation in terms of human perception. We provide the synthesis results of the models with the estimated projection coefficients.

A selective set of images from the FERET database [7] were used for the experiment. The data set consists of total 100 different subject faces in the resolution of [360x360]. The shape information was built by manually annotating the images with 103 landmarks. In order to create the lower counterpart [45x45] of the dataset, 8 factor decimation was applied by adding random noise with 0.01 variance for texture and 0.0001 variance for shape (noises were applied on 0-1 normalized values). The data set was divided into two; 75 for training, and 25 for testing. We train two AAMs for low resolution and high resolution images individually. Models represent the %95 percentage of dataset domain and were adjusted to search around +/-3σ. The results were compared with the results of RAF algorithm [4].

The quantitative results, Fig.1 and Fig.2, show that the proposed method approximates the performance of the RAF algorithm by spending much less computational resources. Most of the computational load of the RAF algorithm is caused by the inverse-warping and deforming the HR image during the optimization iterations. On the other hand the main computational load of the proposed method is caused by warping the LR input during LR AAM fitting, and the cost of the optimization stage is nothing, because it is solved analytically with coefficient size operator multiplications. In other words the increase in the speed between RAF and the proposed approach is equal to the difference between model fitting in HR and model fitting in LR. Since model fitting has $O(n^2)$ complexity [3] (where n refers to number of pixels), especially when excessive amount of decimation exists; such as 8 factor as in the experiment, this difference is quite significant.

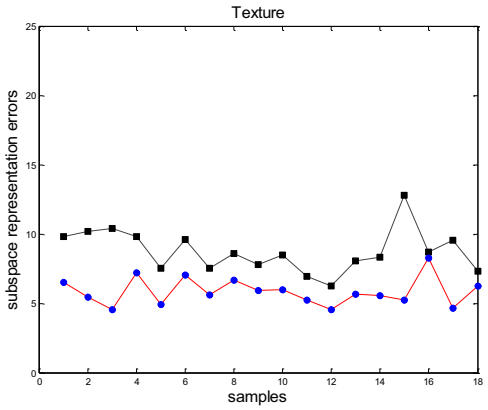

Figure 1. Error in the estimation of HR texture subspace representation. Black squares refer to RAF reconstructions, and blue circles refer to the reconstructions by the proposed method.

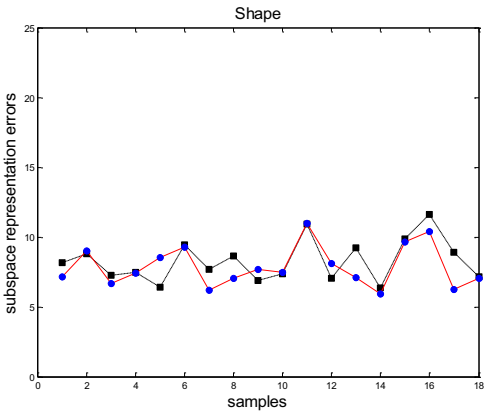

Figure 2. Error in the estimation of HR shape subspace representation. Black squares refer to RAF reconstructions, and blue circles refer to the reconstructions by the proposed method.

Also as parallel with the quantitative results (qualitative results are obtained by synthesizing from the same transform models, and quantitative results are the input for the synthesis) the qualitative results in Fig.3 are close.

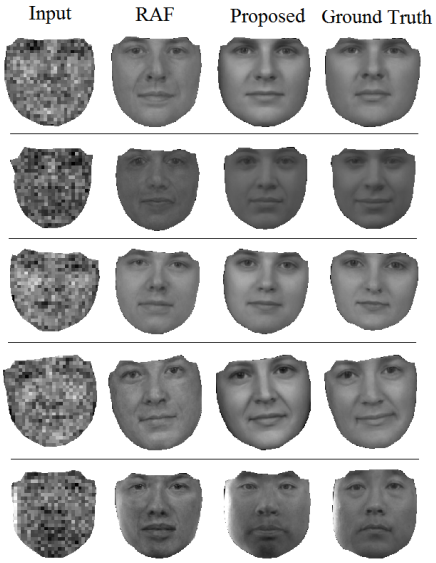

Figure 3. Selected examples from the test set. The reconstruction syntheses are shown for RAF and Proposed algorithms in second and third columns respectively.

4. Conclusion

In this work we proposed a fast solution for the face super resolution problem by using generative models and utilizing shape and texture components together. Experimentally we showed that the performance of other approaches utilizing both shape and texture information, such as RAF, could be reached faster. We stated that the saving in time complexity is exactly the same with the computational saving between model fitting in HR and model fitting in LR. Especially when excessive amount of decimation exists the proposed method would be much more effective and applicable compared to the others.

References

[1] C. Liu, H. Y. Shum, W. T. Freeman. Face hallucination: theory and practice. International Journal of Computer Vision (IJCV), Vol. 75, No. 1, pp. 115-134, October, 2007.

[2] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes III, R. M. Mersereau. Eigenface-domain super-resolution for face recognition. IEEE Trans. Image Processing, vol. 12, no. 5, pp. 597-606, May 2003.

[3] T. F. Cootes, G.J. Edwards, C.J. Taylor. Active Appearance Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 6, June, 2001, pp. 681-685.

[4] G. Dedeoglu, S. Baker, T. Kanade. Resolution-Aware Fitting of Active Appearance Models to Low Resolution Images. ECCV, 2006.

[5] O. G. Sezer, Y. Altunbasak. Face recognition with independent component based super-resolution. SPIE Visual Comm. and Image Proc. Conf. 2006.

[6] S. Baker, T. Kanade. Hallucinating faces. In IEEE International Conference on Automatic Face and Gesture Recognition, March 2000.

[7] P. Philips, H. Moon, P. Pauss, S. Rivzvi. The feret evaluation methodology for face recognition algorithms. In Proc. IEEE Conf. Computer Vision and Pattern Recognition, pages 137–143, 1997.

[8] G. Dedeoglu, T. Kanade, S. Baker. The Asymmetry of Image Registration and its Application to Face Tracking. Robotics Institute CMURI- TR-06-06, Carnegie Mellon University, February, 2006

[9] S. C. Park, M. K. Park, M. G. Kang. Super-resolution image reconstruction: a technical overview. IEEE Signal Processing Magazine, 20(3):21–36, May 2003.