To appear Int. Conference on Pattern Recognition, Cambridge, UK, August 2004.

Resolution Enhancement by AdaBoost

Abstract

This paper proposes a learning scheme based still image super-resolution reconstruction algorithm. Super-resolution reconstruction is proposed as a binary classification problem and can be solved by conditional class probability estimation. Assuming the probability takes the form of additive logistic regression function, AdaBoost algorithm is used to predict the probability. Experiments on face images validate the algorithm.

1 Introduction

Advances in technology have increased the need for high-resolution images. However, due to physical constraints such as limited storage space, bandwidth, and ubiquitous low resolution image sensors on handheld devices, these needs cannot always be satisfied. In super-resolution image reconstruction, high resolution images are generated from lower-resolution samples, thereby allowing higher resolution imagery to be used in an increasing number of applications.

There are two main approaches to still image super-resolution reconstruction: interpolation and statistical learning methods. Traditional interpolation methods such as bilinear interpolation suffer from over-smoothing and hence are not visually satisfying. Statistical learning methods such as those proposed by [1, 2, 3, 4] use prior statistical information to enhance the resolution of images in a perceptually favorable manner. These methods assume that images from the same category bear certain similar statistics. By predicting the lost information and fusing it back into the original low resolution image, the reconstructed image acquires more detail. For example, the lost high-frequency component can be predicted from the prior knowledge learned from ground-truth high resolution images. Because the learned components are correct only in the statistical sense, the results are susceptible to aliasing.

In this paper, the super-resolution problem is fit into the context of classical binary classification and modeled as a conditional class probability estimation problem. Assuming the insufficient resolution is a result of the intrinsic low-pass characteristics of the data collection device, the conditional class probability is used to determine the consistency between the given low-resolution image and the estimated high-frequency component. Greater probability implies more confidence in the estimated high-frequency parts. AdaBoost is used to estimate the probability with the assumption that this probability lies in the logistic regression category. Experimental results validate our algorithm.

The remain part of the paper is organized as follows: in section 2, the algorithm and its theoretical basis are presented. In section 3, experimental evaluation is presented to validate of the proposed algorithm. Section 4 concludes the paper.

2 Algorithm

2.1 Problem description

We assume that insufficient resolution are results from lacking of high-frequency component. Therefore, compensating the high-frequency component back will enhance the resolution. Different from the Markov Random Field (MRF) assumption in [1], in this paper, still image super-resolution reconstruction is modeled as a binary classification problem. Class conditional probability is evaluated to predict the corresponding high-frequency component.

Local regions from the same category of images exhibit certain consistent statistics. For example, conditional probability $\Pr(DHF|LF)$ is highly predictable from examples, where LF is the low frequency part of an image patch and DHF is the directional high-frequency counterpart. We introduce a concept *consistency*, which means the probability that a low frequency feature LF and high frequency features DHF are from the same image region. *Consistency* relates the super-resolution reconstruction problem with the binary classification language.

Suppose we have an image database with sufficient high resolution. Randomly select image patches

$\{\mathbf{X}_i, i = 1, \cdots, N\}$ from the database. All $\mathbf{X}_i$ have two parts: the low frequency part $\mathbf{Xl}_i$ and the directional high frequency counterpart:

$$[\mathbf{Xh}_{i,\theta_1}, \mathbf{Xh}_{i,\theta_2}, \cdots, \mathbf{Xh}_{i,\theta_K}].$$

θ_k is the orientation index. We will talk more about it later in this section. Then *consistency* represents the following probability:

$$p(c_{t,k} = 1|\mathbf{Xl}_t, \mathbf{Xh}_{j,\theta_k}),$$

where $c_{t,k}$ is an assigned class label. $c_{t,k} = 1$ if $\mathbf{Xl}_t$ and $\mathbf{Xh}_{j,\theta_k}$ are from the same image patch and $c_{t,k} = -1$ otherwise. The given low resolution image are treated as a set of incomplete testing examples $\mathbf{Y}_g$, with the directional high frequency information missing. If we can recover the most consistent directional high-frequency components for every element in $\{\mathbf{Y}_g\}$, the super-resolved image can be reconstructed.

The training sample set $\{(\mathbf{Y}_{t,k}, c_{t,k})\}$ is constructed as follows (t is the sample index, k is the orientation index):

$$\mathbf{Y}_{t,k} = [\mathbf{Xl}_t{}^{\mathrm{T}}, \mathbf{Xh}_{j,\theta_k}{}^{\mathrm{T}}]^{\mathrm{T}}, k = 1, \cdots, K; t = 1, \cdots, T. \quad (1)$$

$c_{t,k}$ is the class label. Positive sample ($c_{t,k} = 1$) means low-frequency part $\mathbf{Xl}_t$ and corresponding directional high-frequency part $\mathbf{Xh}_{j,\theta_k}$ are from the same image patch ($t = j$). For negative samples, we require that $\mathbf{Xh}_{j,\theta_k}$ has significant difference with the true corresponding directional high-frequency component of $\mathbf{Xl}_t$. The training examples have two functions. First, prior knowledge about the *consistency* is learned from $\{(\mathbf{Y}_{t,k}, c_{t,k})\}$. Second, it is used as the candidacy dictionary $\mathbb{D}$ for high frequency features.

The *consistency* between the given low frequency feature $\mathbf{Y}_g$ and a high frequency feature $\mathbf{Xh}_{j,\theta_k}$ from the dictionary $\mathbb{D}$ is evaluated. The one that matches best is the candidate of the lost high-frequency component. Hence, the high-resolution image reconstruction problem actually becomes a conditional class probability estimation problem:

$$\widehat{\mathbf{Xh}}_{g,\theta_k} = \arg \max_{\mathbf{Xh}_{m,\theta_k} \in \mathbb{D}} p(c_{t,k} = 1|(\mathbf{Y_g}, \mathbf{Xh}_{m,\theta_k})), \quad (2)$$

If assuming that the class conditional probability takes logistic regression function form, AdaBoost actually provide a way to evaluate the probability quantitatively.

The steerable pyramid method [5] is used to decompose the image into low frequency component and directional high frequency components. Steerable pyramids provide orientational localized basis filters, which decompose high frequency components into different orientations. In every orientation, the directional high frequency component is estimated separately and therefore each component receives equal weight. As a result, the high frequency component in the dominant orientation cannot outweigh the weaker ones. For every set of orientational high frequency components with orientation θ_k, we construct a training sample set and the most matching directional high frequency component is learned.

2.2 AdaBoost for conditional class probability estimation

It is reasonable to assume the class conditional probability is modeled by logistic regression function:

$$p(c = 1|\mathbf{Y}) = \frac{e^{F(\mathbf{Y})}}{1 + e^{F(\mathbf{Y})}}. \quad (3)$$

It can also be written as:

$$\log \frac{p(c = 1|\mathbf{Y})}{p(c = -1|\mathbf{Y})} = F(\mathbf{Y}). \quad (4)$$

For additive logistic regression model, we have $F(\mathbf{Y}) = \sum_{m=1}^{M} f_m(\mathbf{Y})$. In [6], the author shows that AdaBoost algorithm is actually a stagewise estimation procedure for fitting an additive logistic regression model. This statement can be proved from the following two aspects. AdaBoost procedure finds an $F(\mathbf{Y})$ that minimize $\mathtt{E}\{e^{-cF(\mathbf{Y})}\}$; while $\mathtt{E}\{e^{-cF(\mathbf{Y})}\}$ is minimized at:

$$F(\mathbf{Y}) = \frac{1}{2} \log \frac{p(c = 1|\mathbf{Y})}{p(c = -1|\mathbf{Y})}, \quad (5)$$

which is unique with the logistic regression model up to a factor of 2.

AdaBoost has become one of the most popular learning algorithms in the last decade or so. The most attractive advantage is its simplicity. Good performance can be achieved even if the underlying weak learner is very simple. Also, for separable data, the convergence can be guaranteed. The main idea of AdaBoost is that training is focused on *hard examples*. The following steps summarizes AdaBoost:

1. Assign a set of weights $D_1(t), t = 1, \cdots, T$ over the training set, where T is the size of the training set. Initially they are set equal for all examples.

2. For $s = 1, \cdots, S$:

- Train weak learners by distribution $D_s(t)$:

$$h_s : \{\mathbf{Y_t}\} \mapsto \{1, -1\};$$

- Get weak hypothesis $h_s(\mathbf{Y_t})$;
- Choose the weight α_s of the current weaker learner h_s according to its error rate ε_s

$$\alpha_s = \frac{1}{2}\ln(\frac{1-\varepsilon_s}{\varepsilon_s}),$$

with $\varepsilon_s = \Pr_{i\sim D_s}(h_s(\mathbf{Y}_t) \neq c_t)$.

- Update $D_s(t)$ to $D_{s+1}(t)$. Normalize it by Z_s so that it is still a distribution.

$$D_{s+1}(t) = \frac{D_s(t)e^{-\alpha_s c_t h_s(t)}}{Z_s}.$$

3. Final output is:

$$H(\mathbf{Y}) = \mathrm{sign}(F(\mathbf{Y})),$$

where:

$$F(\mathbf{Y}) = \sum_{s=1}^{S} \alpha_s h_s.$$

Therefore, the class conditional probability is:

$$p(c=1|\mathbf{Y}) = \frac{e^{2(\sum_{s=1}^{S}\alpha_s h_s)}}{1+e^{2(\sum_{s=1}^{S}\alpha_s h_s)}}. \quad (6)$$

Since different regression models are trained individually for different orientation θ_k, there are K training sample sets and K regression models in the total. The training samples have high dimension. It will be computational expensive to exploit these samples directly. SVD is exploited on the low and orientational high frequency components separately to decrease the dimension. The final super-resolution image is then become:

$$\widehat{\mathbf{Ih}} = [\mathbf{Xl}_g + \beta\sum_{k=1}^{K}\widehat{\mathbf{Xh}_{g,\theta_k}}]_{g=1,\cdots,G}, \quad (7)$$

where β is the blending factor.

2.3 Blocky effect elimination

The above procedure will give a blocky high frequency component estimation. Two additional steps are taken to solve the problem. The first is the normalization of the samples. All the training samples are normalized to have norm of 1. Also the estimated high-frequency component is rescaled according to its low frequency counterpart. Second, overlapped patches are used. For the overlapped pixels, there will be multiple estimates. The final estimate is obtained by the following fusion procedure. Let the multiple estimates be $\tilde{\mathbf{xh}} = \{xh_{t,k}(1),\cdots,xh_{t,k}(P)\}$. The final estimation is:

$$\widehat{xh}_{t,k} = \begin{cases} \mu_{t,k} + \mathrm{b}\sigma_{t,k}, & \mathrm{if MA}_{t,k} \geq \mu_{t,k} + \mathrm{b}\sigma_{t,k} \\ \mu_{t,k} - \mathrm{b}\sigma_{t,k}, & \mathrm{if MI}_{t,k} \leq \mu_{t,k} - \mathrm{b}\sigma_{t,k} \\ \mu_{t,k}, & \text{otherwise} \end{cases} \quad (8)$$

where:

$$\mu_{t,k} = \frac{1}{P}\sum_{p=1}^{P} xh_{t,k}(p);$$

$$\sigma_{t,k} = \frac{1}{P}\sqrt{\sum_{p=1}^{P}(xh_{t,k}(p)-\mu_{t,k})^2};$$

$$MA_{t,k} = \max\{xh_{t,k}(1),\cdots,xh_{t,k}(P)\};$$

$$MI_{t,k} = \min\{xh_{t,k}(1),\cdots,xh_{t,k}(P)\};$$

and b is a parameter which shows the tolerance to the variance. Normally we set b > 1.

3 Experimental evaluation

In this section, the proposed algorithm is evaluated experimentally on face images. Face is an important subject in many research fields and real applications. Also, humans are perceptually sensitive to the details on faces. The face database we used is from [7]. The database is split into two parts. Part one is used as the groundtruth high-resolution images. Images from part two are down-sampled to half by half, blurred by a 3×3, $\sigma = 0.6$ Gaussian and then used as the testing images. We use 15×15 image patch. Smaller patch size will make the training procedure harder since less structural information is preserved. We use 8 orientations for the high frequency components. For every training set (at orientation θ_k), there are 20000 examples. 5000 positive samples and 5000 negative samples are used for training and the remaining 10000 samples are used for testing. *Stump* [8] is used as the basic weak learners for AdaBoost. Fig.1 gives the the testing error rate with the iteration times ($\theta_k = 0°$). 20000 iterations are used and hence 20000 weak hypotheses altogether. It shows that the procedure converges fast and consistently, which means the samples constructed in this manner have good separation.

Images from both parts of the database are tested for performance. Same down-sample and blurring procedure is applied on all test images. Bicubic spline interpolated images are used as the low-frequency input for our algorithm. Experimental example from first part of the database is shown in Fig.2. Example from the second part of the database is shown in Fig.3. Fig.2(a) and Fig.3(a) are the original images. Fig.2(b) and fig.3(b) are results from bicubic spline interpolation. Fig.2(c) and fig.3(c) are results from the proposed algorithm. Fig.2(d) and fig.3(d) are the estimated high frequency components. Results validate that the proposed algorithm is perceptually favorable since more details are present. However,

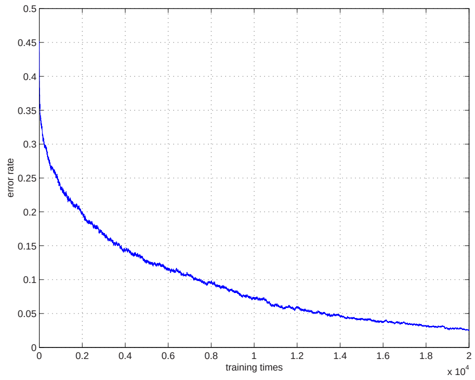
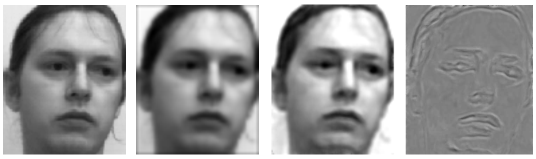


Figure 1: Testing error rate from AdaBoost for $\theta_k = 0°$. The horizontal axis are the iteration times. The vertical axis are the error rates.



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

Figure 2: Example from the first part of the database. (a): Original high-resolution image; (b): High resolution image from bicubic spline interpolation; (c): High-resolution image from the proposed algorithm; (d): Learnt high frequency component.

there is still more work to be done to preserve the brightness of the images.

4 Conclusion

In this paper, a super-resolution reconstruction algorithm for still images is proposed. By fitting the super-resolution reconstruction problem into the binary classification context, this problem is solved by class conditional probability estimation. Using the assumption that the probability taking a logistic regression function form, AdaBoost algorithm can be used to predict the probability. Experiments on face images validate the algorithm.

References

[1] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *International Journal of Computer Vision.*, 40(1):24–57, October. 2000.

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

Figure 3: Example from the second part of the database. (a): Original high-resolution image; (b): High resolution image from bicubic spline interpolation; (c): High-resolution image from the proposed algorithm; (d): Learnt high frequency component.

[2] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 24(9):1167–1183, September. 2002.

[3] F. M. Candocia and J. C. Principe. Superresolution of images based on local correlations. *IEEE Trans. Neural Networks.*, 10(2):372–380, March. 1999.

[4] A. J. Storkey. Dynamic structure super-resolution. In *Advances in Neural Information Processing Systems 15 (NIPS2002).*, pages 1295–1302, 2003.

[5] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multi-scale transforms. *IEEE Trans. Information Theory, Special Issue on Wavelets.*, 38(2):587–607, March. 1992.

[6] J. H. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. Technical report, Stanford University., 1998.

[7] D. B. Graham and N. M. Allinson. Characterizing virtual eigensignatures for general purpose face recognition. In *Face Recognition: From Theory to Applications , NATO ASI Series F, Computer and Systems Sciences.*, volume 163, pages 446–456. H. Wechsler and P. J. Phillips and V. Bruce and F. Fogelman-Soulie and T. S. Huang, 1998.

[8] G. Lugosi and N. Vayatis. A consistent strategy for boosting algorithms. In *Computational Learning Theory : 15th Annual Conference on Computational Learning Theory, COLT 2002.*, pages 303–318, July 2002.