Super-Resolution of Face Images

Rui Huang RUIHUANG@PAUL.RUTGERS.EDU
Division of Computer and Information Sciences, Rutgers University, USA

Abstract

In this paper, we describe a learning-based method for the super-resolution problem, and especially the application of this method to human face images. The Markov Random Field (MRF) model is used to represent the relationship between low-resolution and high-resolution images, and then the super-resolution problem can be solved with two phases. First we learn the parameters of the MRF model from the training data, and then infer the high-resolution image corresponding to a given low-resolution image using the Belief Propagation (BP) algorithm. Experimental results show that this learning-based method outperforms the interpolation-based methods both visually and in terms of the root mean square (RMS) intensity error. We also find the result of this low-level vision problem can be used to solve a high-level vision problem: face recognition.

1. Introduction

Super-resolution is the task of generating a high-resolution image from a given low-resolution image. Various methods have been proposed to solve this problem. The interpolation-based super-resolution methods, such as nearest neighbor interpolation, bilinear interpolation, bicubic interpolation, etc., only use a single low-resolution image to generate the high-resolution image. They are called *interpolation* because the pixels in the low-resolution image are replicated according to some rules to generate the missing pixels of the high-resolution image. These methods are fast because they are based on some simple image processing techniques. However, a disadvantage of these methods is that they replicate pixels from the low-resolution image, which does not contain the high-frequency information, so they cannot essentially increase the information in the image.

The reconstruction-based super-resolution methods (Schultz & Stevenson, 1996; Hardie, Barnard, & Armstrong, 1997) estimate the high-resolution image from a sequence of slightly translated low-resolution images. This is possible if there exists sub-pixel motion between the acquired frames. The fusion of multiple low-resolution images is normally based on the constraints that the high-resolution image, when appropriately warped and down-sampled to model the image formation process, should yield the low-resolution inputs. Super-resolution is therefore posed as a *reconstruction* problem. These methods are more effective than the interpolation-based methods, but require a sequence of slightly translated images. This strong constraint limits the applications of these methods.

Freeman, Pasztor, & Carmichael (2000) recently proposed a learning-based framework for low-level vision problems, one application of which is the super-resolution problem. Unlike the reconstruction-based methods, the learning-based methods have lower requirements on input images. Although the learning-based methods need training images, they do not have strong constraints on the training data, and they only need one low-resolution input image. On the other hand, the learning-based methods are more effective than the interpolation-based methods, because they do learn the high-frequency information from the training data to add to the given low-resolution image.

Freeman et al. only applied their algorithm to generic images. In this paper, we use this method on human face images, which is more specific and more useful. Freeman et al. only visually compared the results of their algorithm to the results of other algorithms. Although this is common in many vision problems (e.g., segmentation, tracking, etc.), we are trying to give a more rigorous comparison based on the root mean square (RMS) intensity error. We have also studied the relationship between the RMS intensity error and the amount of the training data. In addition, we have preliminarily investigated the effectiveness of this learning-based super-resolution method as a face recognition method.

The rest of this paper is organized as follows: Section 2 introduces the learning-based super-resolution method; the application of this method on face recognition is described in Section 3; Section 4 shows the experimental results.

2. Learning-based super-resolution

2.1 MRF models for super-resolution

MRF models have been widely used in image analysis problems (Li, 2001), because they can capture the context of an image (i.e., dependencies among neighboring image pixels) and deal with the noise.

A typical MRF Model, as shown in Figure 1, is a graph with two kinds of nodes: hidden nodes (circles denoted by **x** in Figure 1) and observable nodes (squares denoted by **y** in Figure 1). Edges in the graph represent the relationship of nodes. In the super-resolution problem, we divide the images into small patches. The hidden nodes in MRF represent the high-resolution patches and the observable nodes represent the low-resolution patches.

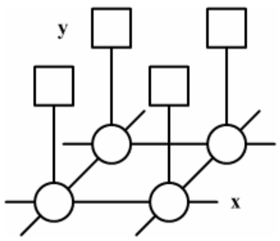


Figure 1. MRF Model

Let n be the number of the hidden/observable states (i.e., the number of patches in the image). A configuration of the hidden layer (i.e., a possible high-resolution image) is:

$$\mathbf{x} = (x_1, ..., x_n), x_i \in L, i = 1, ..., n ,$$

where L is a set of high-resolution patches. In our problem, this set is obtained from the high-resolution training images.

Similarly, a configuration of the observable layer (i.e., the input low-resolution image) is:

$$\mathbf{y} = (y_1, ..., y_n), y_i \in D, i = 1, ..., n ,$$

where D is a set of low-resolution patches. In our problem, this set is obtained from the low-resolution input image.

The relationship between the corresponding high-resolution and low-resolution patches (also known as local evidence) can be represented as the compatibility function $\phi(x_i, y_i) = P(y_i \mid x_i)$.

Similarly, the relationship between the neighboring high-resolution patches can be represented as the second compatibility function $\psi(x_i, x_j) = P(x_i, x_j)$.

Now the super-resolution problem can be viewed as a problem of estimating the MAP solution of the MRF model:

$$\mathbf{x}_{MAP} = \arg\max_{\mathbf{x}} P(\mathbf{x} \mid \mathbf{y}) ,$$

where **x** is the high-resolution image and **y** is the low-resolution image, and

$$P(\mathbf{x} \mid \mathbf{y}) \propto P(\mathbf{y} \mid \mathbf{x}) P(\mathbf{x}) \propto \prod_i \phi(x_i, y_i) \prod_{(i,j)} \psi(x_i, x_j) .$$

Solving an MRF model-based problem involves two phases: a learning phase, where the parameters of the model are learned from the training data, and an inference phase, where the MAP solution of the model is estimated. The learning of the parameters of the model is essentially the learning of the compatibility functions.

$\phi(x_i, y_i)$ is usually assumed to obey a Gaussian distribution to model Gaussian noise:

$$\phi(x_i, y_i) = \exp\left(-\frac{\left\| y_i - y_{x_i} \right\|^2}{2\sigma_l^2} \right) , \tag{1}$$

where x_i is a high-resolution patch candidate, y_{x_i} is the low-resolution version of x_i , and y_i is the low-resolution patch in the input image.

Similarly, $\psi(x_i, x_j)$ is usually defined as:

$$\psi(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2\sigma_h^2} \right) \tag{2}$$

where $d(x_i, x_j)$ is the difference of the overlapping part of the two neighboring high-resolution patch candidates. These functions can be learned from the training data.

2.2 Belief Propagation for MRF models

The exact MAP inference in MRF models is computationally infeasible, so various techniques for approximating the MAP estimation have been proposed (R.C. Dubes et al., 1990), such as Markov Chain Monte Carlo (MCMC), iterated conditional modes (ICM), maximizer of posterior marginals (MPM), etc. We solve this MRF-MAP estimation problem using the Belief Propagation (BP) algorithm.

BP is an inference method proposed by Pearl (1988) to efficiently estimate Bayesian beliefs in the network by the way of iteratively passing messages between neighbors. It is an exact inference method in the network without loops. Even in a network with loops, some successful experimental results have motivated us to use the same message-passing scheme to do an approximate inference (Weiss, 1998). There are two variants of the BP algorithm: sum-product and max-product. The sum-product message-passing rule can be written as:

$$m_{ij}(x_j) = \sum_{x_i} \psi_{ij}(x_i, x_j) \phi_i(x_i) \prod_{k \in \aleph(i) \setminus j} m_{ki}(x_i) .$$

The max-product has analogous formula, with the marginalization replaced by the maximum operator. At convergence:

$$x_{iMAP} = \arg\max_{x_i} \phi_i(x_i) \prod_{j \in \aleph(i)} m_{ji}(x_i) .$$

2.3 Implementation issues

There is a little difference in practical implementation. We assume that the input low-resolution image only contains mid-frequency and low-frequency information, and we want to infer the high-frequency information from the training data. We also assume that the low-frequency part of the input image is not useful for inferring the high-frequency part, given the mid-frequency part. So we do not really use the high-resolution and low-resolution images as **x** and **y**, but use the high-frequency part of the training images as **x**, and the mid-frequency part of the input image as **y**. See Freeman et al. (2000) for more details. The whole algorithm works as follows:

1. Pre-process the training images (including removing the low frequencies of the image, separating the high frequencies and the middle frequencies, and contrast normalizing the image), and break them (both the high-frequency and the mid-frequency images) into small patches, which form the sets of x_i 's and y_{x_i} 's.

2. Pre-process the input images (including removing the low frequencies of the image, and contrast normalizing the image), and break it into small patches with the same size of the training data, which form the set of y_i 's.

3. For each input patch y_i, find the 10 closest y_{x_i} 's, and the corresponding 10 x_i 's are the candidates for that patch. Calculate the compatibility function $\phi(x_i, y_i)$ according to Equation (1).

4. For each pair of neighboring input patches, calculate the 10×10 compatibility function $\psi(x_i, x_j)$ according to Equation (2).

5. Estimate the MRF-MAP solution using BP.

6. Read out the estimated maximum probability high-frequency patches x_{iMAP} , undo the pre-processing, and add them to the input low-resolution image to get the high-resolution image.

3. Super-resolution-based face recognition

There is a strong relationship between the super-resolution problem and the recognition problem (Baker & Kanade, 1999). An extreme case is face recognition-based super-resolution. If a low-resolution face can be firstly recognized, we can simply use the high-resolution face of the same identity in the training data as the solution of the super-resolution problem. On the other side, if a low-resolution face image can be successfully enhanced by super-resolution, then we have more useful information for recognition. We call this latter problem super-resolution-based face recognition. Given a face database and a low-resolution face image, the most direct way to do recognition could be: firstly solve the super-resolution problem using the learning-based method, and then use the estimated high-resolution face image to do recognition.

We propose to solve this problem in a more tightly coupled way. Particularly, from the result of the BP algorithm, we can easily know the identity of each high-frequency patch x_{iMAP} . Then we can use this information to determine the identity of the whole face image (e.g., the identity of the whole image could be the identity that the most patches come from). Actually, in the learning-based super-resolution method, the compatibility function $\phi(x_i, y_i)$ is a local similarity measurement of the testing sample and the training data. The process of determining the closest candidates to the input patch is essentially a process of local template matching. Then we use the other compatibility function $\psi(x_i, x_j)$ to constrain the compatibility of the neighboring local features. So, the final result could be more accurate than that of the traditional template matching-based method, which does not consider the compatibility of the local features.

In some sense, our proposed method can be viewed as an improved template-based face recognition method. The preliminary experiments gave some encouraging results.

4. Experiments

We first test our learning-based super-resolution algorithm on the Yale face database, which includes 15 subjects with 11 face images of each subject. Figure 2 depicts the 11 images of one subject.

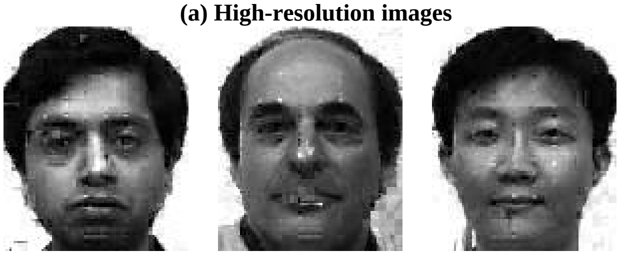
Figure 2. Yale Face Database

We randomly pick 1 image of each subject (totally 15) as the testing images, and the rest 10 of each (totally 150) as the training images. Some results are shown in Figure 3. Figure 3a shows the original high-resolution images (112×92, input images are down-sampled to 28×23). Figures 3b-3e show the results of our learning-based method and three different interpolation-based methods. The results of the learning-based method *look* obviously better than those of the interpolation-based methods.

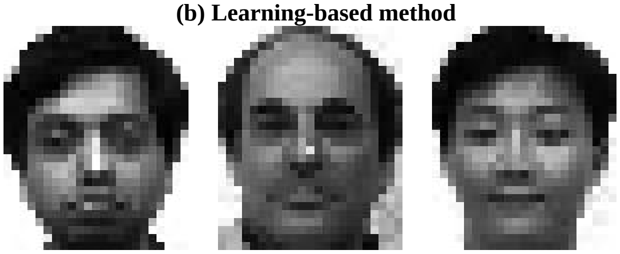
Besides the visual comparison in Figure 3, we also compute the RMS intensity error of each method. The result depicted in Figure 4a is consistent with Figure 3. Figure 4b shows that the RMS error of the learning-based method decreases when more images are used for training.

In above experiments, we have estimated one super-resolution image for each subject. For each estimated image, we count the numbers of the patches that came from different subjects, and plot the result in Figure 5. Obviously, for each estimated image, the number of the patches coming from the *correct* identity is much higher than others (*correct* means the patch came from the same subject as the input image). So, Figure 5 shows the feasibility of our proposed face recognition method based on the learning-based super-resolution method.

(a) High-resolution images

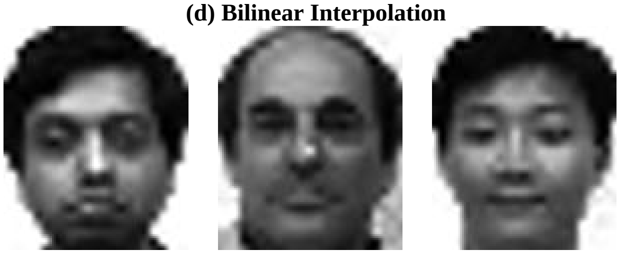


(b) Learning-based method



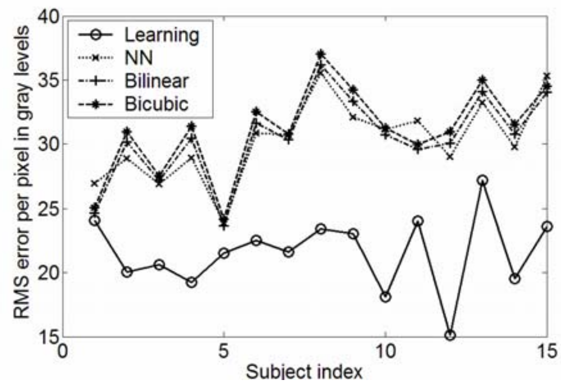
(c) Nearest Neighbor Interpolation

(d) Bilinear Interpolation

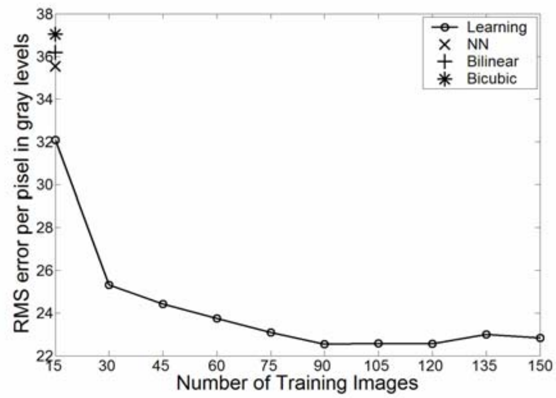


(e) Bicubic Interpolation

Figure 3. Visual Comparison of super-resolution methods



(a)



(b)

Figure 4. RMS error of super-resolution methods

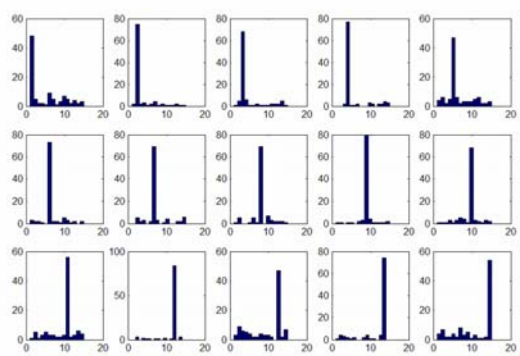


Figure 5. Super-resolution-based face recognition

References

S. Baker & T. Kanade (1999). Hallucinating Faces. *CMU-RI-TR-99-32*.

R.C. Dubes, A.K. Jain, S.G. Nadabar, & C.C. Chen (1990). MRF Model-Based Algorithm for Image Segmentation. *Proceedings of ICPR.*

W.T. Freeman, E.C. Pasztor, & O.T. Carmichael (2000). Learning Low-Level Vision. *International Journal of Computer Vision*, Vol. 40, No. 1.

R. Hardie, K. Barnard, & E. Armstrong (1997). Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images. *IEEE Transactions on Image Processing*, Vol. 6, No. 12.

S.Z. Li (2001). *Markov Random Field Modeling in Image Analysis*. Springer-Verlag.

J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, CA.

R. Schultz & R. Stevenson (1996). Extraction of High-Resolution Frames from Video Sequences. *IEEE Transactions on Image Processing*, Vol. 5, No. 6.

Y. Weiss (1998). Belief Propagation and Revision in Networks with Loops. *Technical Report MIT A.I. Memo 1616*.