Direct Energy Minimization for Super-Resolution on Nonlinear Manifolds

Tien-Lung Chang[1,2], Tyng-Luh Liu[1], and Jen-Hui Chuang[2]

[1] Institute of Information Science, Academia Sinica, Taipei 115, Taiwan
liutyng@iis.sinica.edu.tw
[2] Dept. of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan

Abstract. We address the problem of single image super-resolution by exploring the manifold properties. Given a set of low resolution image patches and their corresponding high resolution patches, we assume they respectively reside on two non-linear manifolds that have similar locally-linear structure. This manifold correlation can be realized by a three-layer Markov network that connects performing super-resolution with energy minimization. The main advantage of our approach is that by working directly with the network model, there is no need to actually construct the mappings for the underlying manifolds. To achieve such efficiency, we establish an energy minimization model for the network that directly accounts for the expected property entailed by the manifold assumption. The resulting energy function has two nice properties for super-resolution. First, the function is convex so that the optimization can be efficiently done. Second, it can be shown to be an upper bound of the reconstruction error by our algorithm. Thus, minimizing the energy function automatically guarantees a lower reconstruction error— an important characteristic for promising stable super-resolution results.

1 Introduction

In this work *super-resolution* specifically means the technique to estimate a high-resolution (HR) image from one or more low-resolution (LR) instances taken of the same scene by some imaging processes. One reason for looking into such an issue is due to the quality constraints on many existing imaging devices, especially on nowadays *digital* imaging systems. Although a large portion of them are suitable for most imaging applications, the current resolution levels by the affordable price still can not satisfy certain common needs. Take, for example, the ubiquitous security surveillance systems. To completely monitor the whole area of interest, lots of cameras are often needed. However, the quality of these cameras is generally not good enough for providing useful information. While it is always possible to physically increase the quality of sensors by investing more budgets, image processing techniques such as super-resolution provide a reasonable solution, and have been studied for years.

Super-resolution has lately become an active topic in vision research. Its application ranges from medical imaging to image compression. An extensive number of useful approaches have thus been proposed to address the problem with different aspects of consideration [1], [3], [5], [6], [7], [10], [12], [13], [15], [17], [18],

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part IV, LNCS 3954, pp. 281–294, 2006.

[19], [20], [21]. By the underlying models, these methods can be roughly divided into two categories [1], [19]: *reconstruction-based* and *recognition-based*. Typically a reconstruction-based technique tries to accomplish super-resolution with an ML or a MAP formulation [5], [10], [17], [18], [20]. To avoid causing an underdetermined system, some kind of prior information needs to be imposed for regularizing the results of super-resolution and for adding more high frequencies. On the other hand, the recognition-based methods, e.g., [1], [7], [19], often first resize an LR image into the desirable size of a target HR one, and then add the appropriate high frequencies from the training set to improve the quality of the resized image. While the reconstruction-based methods assume a more realistic model, which simulates the process that we produce an LR image and can be solved by standard optimization algorithms, the recognition-based methods indeed provide more feasible results, especially for the case that the number of given LR images is rather small.

Different from previous approaches for super-resolution, we are motivated by investigating the manifold property of LR and HR image patches, with an emphasis on the assumption that for each pair of corresponding LR and HR image patches their local neighborhoods on some proper nonlinear manifolds would be *similar*. Specifically, our method deals with the single (LR) image super-resolution problem, and uses a three-layer Markov network to realize the manifold assumption. The key contribution of the proposed approach is to explore the connection between the LR and HR manifolds without the need to explicitly construct the respective manifolds. We achieve such efficiency by establishing an energy minimization model that *directly* accounts for the expected property entailed by the implicit manifold structure. It therefore results in an optimization-based algorithm for super-resolution, and requires only a training set consisting of a small number of pairwise LR and HR image patches.

2 Previous Work

For convenience we always denote an LR image by L, and an HR image by H. The task of single image super-resolution is therefore to find a *best* H of a specified higher resolution, from which the given L can be reasonably reproduced. Indeed over the years there are many attempts to address the problem, including, e.g., direct interpolation, or frequency-domain reconstruction [12]. Our discussion here focuses only on more recent super-resolution techniques, sorted according to the following two classes.

Reconstruction-Based. Methods of this kind typically assume an observation model that describes how one can get L from H. If L and H are represented in the form of column vectors, the observation model can often be written out in a linear form:

$$L = TH + Z, \tag{1}$$

where T can be thought of as some underlying imaging system transforming H to L, and Z is the additive zero-mean Gaussian noise. As an example to illustrate,

suppose H is of size $\mathcal{X}_H$ pixels and L is of size $\mathcal{X}_L$ pixels. Then the observation model in the work of Elad et al. [5] can be stated as

$$L = DBH + Z, \tag{2}$$

where B is a blur matrix of size $\mathcal{X}_H$-by-$\mathcal{X}_H$, and D is a downsampling operator of size $\mathcal{X}_L$-by-$\mathcal{X}_H$. Notice that we have omitted the geometric motion matrix and the indices of image frames in [5] owing to that in our case there is only one LR image, namely L, as the input. With equation (1), the derivation of H can be readily casted as solving an ML (maximum likelihood) or a MAP (maximum a posteriori) problem, e.g., [5], [10], [17], [18], [20].

Still the huge dimensional characteristic of the super-resolution problem like (1) or (2) can be a challenging factor. A simple and effective technique has been proposed by Irani and Peleg [13] that approximates a *solution* of (2) based on *iterative back projection* (IBP). Their method starts with an initial guess H_0 for the HR image, projects the temporary result H_k by the same process for producing an LR image, and then updates H_k into H_{k+1} according to the projection error. These steps can be summarized by

$$H_{k+1} = H_k + B'U(L - DBH_k), \tag{3}$$

where U is now an upsampling matrix and B' is another blur matrix distributing the projection error. The IBP scheme given by (3) is intuitive and fast. However, it has no unique solution due to the ill-posed nature of equation (2), and cannot be effectively extended to include prior information on H for regularizing the solution. The concern of adding the prior information is necessary in that for the single image super-resolution problem the matrix T in (1) is inherently singular. Consequently, without using appropriate prior information on H, a reconstruction-based method for (1) could yield super-resolution results containing appreciable artifacts. Alternatively, the *projection onto convex sets* (POCS) approach [5], [20] applies set theories to the super-resolution problem, and allows constraints on additional prior information. The main disadvantages of POCS-based methods include non-uniqueness of solution, slow convergence, and high computational cost.

The ill-posedness of super-resolution noticeably hinders the performance of reconstruction-based methods, and could yield jaggy or ringing artifacts, e.g., as in the results of [19]. While adding some prior may alleviate the problem, it is generally too simple for simulating the real world texture. In fact regularizing super-resolution with prior information mostly smooths out small derivatives. When carefully done, it could produce *good* edges. However, the scheme may also suppress useful details, and is insufficient for representing complex textures.

Recognition-Based. To more naturally retain good image characteristics for super-resolution, the recognition-based techniques [1], [3], [7], [11], [19] resort to a training set of LR and HR image patches. The main idea is to use the actual HR patches to construct the results of super-resolution. Such methods usually carry out super-resolution by the following steps: divide the given LR image into

small (overlapping) patches, compare them with LR image patches in the training set, and replace them with the corresponding HR patches. In [1], Baker and Kanade discuss the limits of reconstruction-based approaches, and also establish a recognition-based super-resolution technique. Freeman et al. [7] propose a Markov model, in which overlapping patches are used to enhance the spatial continuity. However, in most of the recognition-based algorithms, the recognition of each LR patch gives a hard assignment to a specific HR counterpart in the training set. The mechanism could cause blocky effect, or oversmoothness—if image processing is performed to eliminate the blocky effect [3].

With the exception of [19], the above-mentioned recognition-based methods are restricted by the class of their collected training sets. Indeed Sun et al. [19] only replace the patches of detected primal sketches, and then apply IBP [13] to ensure the reconstruction constraint. Though the primal-sketch scheme is useful for processing a wide range of LR images, its super-resolution results may contain artifacts induced by the back projection scheme. More recently, Chang et. al [3] consider neighbor components in generating the HR image patches so that the size of the training set can be dramatically reduced.

There are some attempts to integrate the two concepts, reconstruction-based and recognition-based, for establishing a super-resolution technique that has low reconstruction error, and meanwhile enriches a resulting HR image with complex priors learned from training patches. For example, Pickup et al. [15] include the learned image prior into a MAP model for super-resolution. The way they define the image prior on an image pixel is to assume a Gaussian distribution with the mean obtained by searching the set of training patches, finding the patch most similar to the neighborhood region around this image pixel, and identifying the value from the central pixel of the resulting patch. To feasibly optimize the formulation, Pickup et al. assume that small perturbations of the neighborhood region will not affect the searching result, an assumption that is not necessary the case. In some ways the method of Sun et al. [19] also has the advantages of the two types of approaches, but it is in essence an IBP algorithm with a better initial guess (learned from the training set).

3 Manifold Ways for Super-Resolution

Given a single LR image L, which can be thought of being derived by blurring and then downsampling some HR image of a real scene, the task of super-resolution is then to approximate a high-resolution H that is similar to the original scene. In our formulation we shall split L into n overlapping patches $\{\ell_i\}_{i=1}^n$. Intuitively, for each LR patch ℓ_i, the corresponding site i on H should have an HR patch h_i that is closely related to the appearance of ℓ_i (see Figure 1a).

Note that the site correlation between ℓ_i and (a desired) h_i does not imply it needs a training set, denoted as Ω, comprising numerous pairs of LR and HR patches to construct a reasonable H. The supporting evidence could come from investigating the properties of natural image statistics. One key conclusion related to our application is that although images are typically represented as high dimensional data, the natural images are actually distributed over a relatively

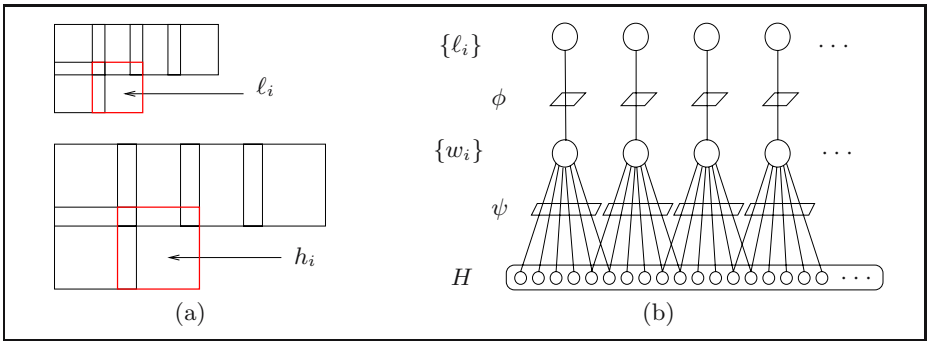

Fig. 1. (a) The corresponding low- and high-resolution patches ℓ_i and h_i (highlighted in red). (b) The 1-D illustration for the 3-layer Markov model. Note that in the output layer for describing the resulting HR image H each node represents an image pixel.

low dimensional manifold. For example, Lee et al. [14] find that the state space of natural image patches of size 3-by-3 pixels is indeed very sparse.

We assume that, among our training data, the set of HR patches and the set of corresponding LR patches respectively reside on two different nonlinear manifolds, but with similar locally-linear structure. In other words, the linear neighborhood relation of ℓ_i on the LR manifold can be used as a hint to correlate h_i and its neighbors on the HR manifold. The same assumption has been made in Chang et al. [3], and shown to produce stable super-resolution performance. Nonetheless, we emphasize a crucial difference that in [3] the locally-linear structure of the LR manifold is approximated exclusively with information from the LR patches in Ω. And each patch of the resulting HR image is independently determined via a hard assignment by imposing a similar locally-linear structure on the HR manifold. For each pixel covered by different overlapping HR patches, the average of these different values is assigned to resolve inconsistency. Such a tactic sometimes introduces oversmoothness in the results [3].

Even though our discussion so far has indicated that the proposed super-resolution algorithm involves learning two different manifolds from the training data Ω, it turns out to be more a conceptual idea. In practice there is no need to explicitly construct the LR and HR manifolds as we are only interested in exploring their underlying locally-linear property. Later in the next section we will explain how to use a Markov network to *conform* the two manifolds without knowing their structure. For now we shall give more discussions on the assumption that the two manifolds have similar locally-linear structure.

3.1 Locally Linear Assumption

Locally linear embedding (LLE) [16] is a way to map high dimensional data into a low dimensional space with the useful property of maintaining the neighborhood relationship. LLE assumes that each data point X_i and its neighbors lie on (or close to) a locally-linear patch of the manifold. This local geometry can be

characterized by linear coefficients W_{ij} that reconstruct each data point X_i from its neighbors X_j's. The coefficient matrix W is decided by minimizing

$$\sum_i \|X_i - \sum_j W_{ij} X_j\|^2, \tag{4}$$

where W_{ij} is required to be 0 if X_j is not a neighbor of X_i. The objective of LLE is therefore to construct a lower dimensional data set whose local geometry can also be characterized by W. For the LR and HR manifolds to have similar locally-linear structure, the coefficient matrix W^* minimizing the LLE formulation (4) on the HR patches should also yield a small value of (4) when the data are replaced with the LR patches, and vice versa. While it is non-trivial to verify the property analytically, the linear model (1) relating an HR image H with its corresponding LR image L suggests the assumption is indeed a reasonable one.

4 The Energy Minimization Model

We now describe how super-resolution on nonlinear manifolds can be done with the convenience of skipping constructing the manifolds. Suppose we have a pair of LR and HR image patches, respectively denoted as ℓ and h. (Each image patch will hereafter be represented as a column vector.) Let P and Q be two matrices with the same number of columns. In particular, the columns of P are ℓ's neighbors in the training set Ω, and those of Q are h's neighbors in Ω. By the similar locally-linear structure of the LR and HR manifolds, we can find a reconstruction coefficient vector w satisfying

$$\begin{bmatrix} \ell \\ h \end{bmatrix} = \begin{bmatrix} P \\ Q \end{bmatrix} w + \begin{bmatrix} \epsilon \\ \delta \end{bmatrix}, \tag{5}$$

where ϵ and δ are Gaussian noise terms. Clearly equation (5) is the mathematical interpretation for the adopted manifold assumption, and it also nicely connects pivotal *elements* in solving the single image super-resolution problem.

To realize the manifold concept embodied in (5), we consider a three-layer Markov network, shown in Figure 1b. In the input layer of the network, each node represents an LR patch, say, ℓ_i from the ith site of the given LR image L. Each node of the second (hidden) layer is a coefficient vector w_i, stating how ℓ_i and the corresponding patch h_i from the ith site of the approximated HR image H can be reconstructed from their neighbors in the training set Ω. Note that each node of the output layer consists of only one pixel of H. In the network each node ℓ_i is connected to the reconstruction coefficient vector w_i, and w_i is further connected to those nodes (image pixels) in the ith site of H. The output layer itself is a fully-connected graph, i.e., a big clique.

With (5) and the Markov network described above, we are in a position to define the energy function F for the network by

$$F(W, H; L) = \sum_i \phi(\ell_i, w_i) + \lambda_1^2 \sum_i \sum_{j \in \text{site } i} \psi(w_i, H^j) + \lambda_2^2 \zeta(H), \tag{6}$$

where there are in turn three kinds of potential functions, namely ϕ, ψ, and ζ to be specified, $W = [w_1 \cdots w_n]$ is the coefficient matrix, λ_1 and λ_2 are parameters to weigh the contributions of the three terms, and H^j denotes the jth pixel of H. We next give the definitions for each of the three potential functions. To begin with, for each LR patch ℓ_i and the connected node w_i, the network is designed to maximize the joint probability of ℓ_i and w_i. Thus, from (5), we arrive at the following definition:

$$\phi(\ell_i, w_i) = \|\ell_i - P_i w_i\|^2. \tag{7}$$

Suppose H^j is the kth pixel on site i of H (i.e. h_i). Then the potential function ψ for w_i and H^j can be defined by

$$\psi(w_i, H^j) = \|e_k^T h_i - e_k^T Q_i w_i\|^2, \tag{8}$$

where e_k is the kth coordinate vector. Notice that, from (5), minimizing the summation $\sum_{j \in \text{site } i} \psi(w_i, H^j)$ is equivalent to maximizing the joint probability of h_i and w_i. Finally, the potential function ζ in (6) is to add appropriate image prior for super-resolution, and is defined on the big clique of the whole H:

$$\zeta(H) = \|SH\|^2, \tag{9}$$

where we shall discuss the matrix S later, and here we simply treat S as the zero matrix. With (7), (8), and (9) so defined, the energy function F in (6) is convex to w_i and H. Hence the super-resolution output H^* by the Markov network can be achieved by minimizing F with respect to w_i and H, respectively and iteratively. The proposed super-resolution algorithm is summarized in Algorithm 1. (Notice that the computation of H described in line 5 of Algorithm 1 is for the convenience of presentation. Owing to the structure of the Markov network, we can indeed compute H pixelwise for a more efficient implementation.)

Algorithm 1. Direct Energy Minimization for Super-Resolution

Input : An inpute LR image L, and a training set Ω.
Output: An HR image H^*.

1 Split L into n overlapping LR image patches, $\{\ell_i\}_{i=1}^n$.
2 For each ℓ_i, find its K nearest LR neighbors in Ω, and form P_i as in (5).
3 For each P_i, take the corresponding K HR patches in Ω to form Q_i as in (5).
4 For each i, compute the initial w_i based on P_i and ℓ_i.
Repeat
5 For each i, compute h_i from (5), given w_i and Q_i.
6 For each i, compute w_i from (5), given ℓ_i, h_i, P_i, and Q_i.
Until *Convergence*

4.1 Bound the Reconstruction Error

Besides being convex for the ease of optimization, the energy function F defined in (6) can be shown to be an upper bound of the reconstruction error yielded by our algorithm. That is, since our approach is to minimize F,

a resulting super-resolution result H^* by Algorithm 1 would have a small reconstruction error (bounded by the minimal energy F^*). Thus the proposed direct energy minimization method not only possesses the convenience for not constructing the manifolds explicitly but also produces stable super-resolution results.

With (2) and a given LR image L, the reconstruction error of H^* derived by Algorithm 1 can be expressed in the following matrix form

$$\|DBH^* - L\|^2, \tag{10}$$

where B is a symmetric blur matrix and D is the downsampling matrix. We now explain why the reconstruction error in (10) will be lower than F in (6). Let B' and D' be the corresponding blur and downsampling matrices on the HR patches. It can be shown that there exists a (mask) matrix M to extract a central region within an LR patch such that $M\ell_j = MD'B'h_j$ for any pair of patches ℓ_j and h_j in Ω. We split the input L into overlapping $\{\ell_i\}_{i=1}^n$. For each ℓ_i we define M_i to select pixels from the central region defined by M such that L is a disjoint union of $\{\tilde{\ell}_i = M_i\ell_i\}_{i=1}^n$. The above procedure can be accomplished by splitting L into denser overlapping patches, or by adjusting M_i for each site i. The reconstruction error in (10) can then be rewritten as

$$\begin{aligned}\sum_i \|M_iD'B'h_i^* - M_i\ell_i\|^2 &= \sum_i \|M_iD'B'(Q_iw_i + \delta_i) - M_i(P_iw_i + \epsilon_i)\|^2 \\ &= \sum_i \|M_iD'B'\delta_i - M_i\epsilon_i\|^2 \le \sum_i (\|D'B'\delta_i\|^2 + \|\epsilon_i\|^2) \\ &\le \sum_i (\lambda_1^2\|\delta_i\|^2 + \|\epsilon_i\|^2) \le F \end{aligned} \tag{11}$$

The only restriction for (11) to be valid is that λ_1 should be larger than the downsampling ratio. One can see that F is not a tight bound for the reconstruction error. So in some cases the resulting reconstruction errors by our method are higher than those induced by the IBP. However, in our experiments the proposed algorithm often gives satisfactory results and lower reconstruction errors in fewer iteration steps than those required by IBP. (See Figure 2b.)

4.2 The Partial Gestalt Prior

We now discuss the use of prior information for super-resolution. Indeed those HR patches in the training set Ω can be considered as some kind of prior. However, due to the computation complexity, most super-resolution methods, including ours, can manage only small patches. Otherwise, the variances of image patches would cause Ω to grow into an infeasible size. Due to such limitation, there are some features of large scale as well as high frequencies cannot be recovered by techniques that work on small patches. One example can be seen in Figure 2c. To account for such artifacts, we adopt the concept of **Gestalt** [2], [4]. The Gestalt theory contains many rules to describe human visual perceptions, including symmetry, closeness, and good continuation, to name a few. In this

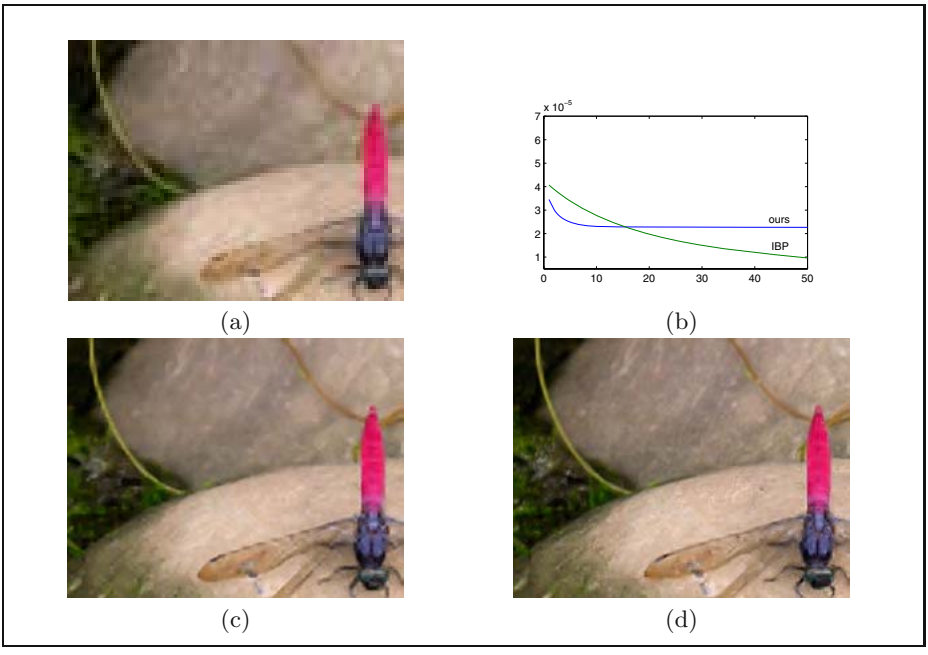

Fig. 2. Dragonfly. (a) The original low-resolution image. (b) Reconstruction errors of IBP and ours. (c) Our approach with $S = 0$. (d) Our approach with Gestalt prior.

work we adopt only the *good continuation* as the large-scale prior information for super-resolution in that properties such as symmetry are often well preserved in LR images, and can be recovered even by simple magnifying schemes.

One way to keep good continuation is to make H smoother along edges and ridges. Thus we choose to define the matrix S in (9) to be the directional derivative operator pixelwise according to the main (edge) direction multiplied by the confidence. More precisely, at pixel j, large confidence c_j means a high probability of having an edge around j in the main direction d_j. We model this claim as an attributed graph that is similar to a Gestalt field [9]. (A Gestalt field is actually a specific mixed Markov model [8] where each address node is connected to only one regular node.) In the generated attributed graph $G = (V, E)$, each node v represents a site on the desired H and has attributes $(d, c, \{a_1, ..., a_4\})$. The attributes d and c mean the main edge direction and the confidence of the patch, respectively. The attributes $\{a_1, ..., a_4\}$ are address variables whose values are neighbors of v (if an edge goes through them) or *nil*. Intuitively, d and c should be compatible with the neighbors of v in Ω as well as the neighbors indicated by $\{a_1, ..., a_4\}$ in the graph. We can therefore define two compatible functions. Like Guo et al. [9], we propose a greedy method to decide $(d, c, \{a_1, ..., a_4\})$. (See Algorithm 2.) However, we discrete the direction into 16 values, and specify the marching order to be the same as the decreasing order of confidence values. In Algorithm 2, the main directions and confidences of sampled HR patches are

Algorithm 2. Constructing Partial Gestalt Prior

Input : Q_i for each site of H.
Output: $(d, c, \{a_1, ..., a_4\})$ for each site of H.

For each site i, initialize three directions and a confidence c by referencing Q_i.
Repeat
  Find the site that has the largest confidence value c among unvisited sites.
  Decide d and $\{a_1, ..., a_4\}$ such that the local compatibility can be achieved.
  Update the main direction d and the confidence value c of this site.
  Mark this site as visited.
Until *All sites are visited*

detected by a set of the first and second Gaussian directional-derivative filters [19]. The improvement owing to adding this prior term can be seen in Figure 2d.

5 Implementation and Discussions

As described before, we optimize F in (6) with respect to w_i and to H, respectively and iteratively. Given H, the derivative of F with respect to w_i is

$$\frac{\partial F}{\partial w_i} = 2\left[P_i^T, \lambda_1 Q_i^T\right]\begin{bmatrix} P_i \\ \lambda_1 Q_i \end{bmatrix} w_i - 2\left[P_i^T, \lambda_1 Q_i^T\right]\begin{bmatrix} l_i \\ \lambda_1 h_i \end{bmatrix}. \tag{12}$$

Hence the optimization with respect to w_i can be achieved directly. Given w_i, the derivative of F with respect to H is

$$\frac{\partial F}{\partial H} = \lambda_1^2 \left[V_1, ..., V_{\mathcal{X}_H}\right]^T + 2\lambda_2^2 S^T S H, \tag{13}$$

where

$$V_j = \sum_{i \in C_j} (-2e_{g(j,i)}^T Q_i w_i + 2H^j), \text{ for } j = 1, \ldots, \mathcal{X}_{\mathcal{H}}, \tag{14}$$

C_j means the set of sites that cover jth pixel of H, and $g(j, i)$ indicates the order of the jth pixel of H in site i. If S is the zero matrix or $\lambda_2 = 0$, optimizing F according to H can also be done in one step. Otherwise, $S^T S$ is a very large matrix and may be singular. In this case we implement a conjugate gradient algorithm as suggested in [21].

5.1 Experimental Results

Due to the fact that humans are more sensitive to changes in luminance channel, we only test our method on the luminance channel, and magnify the color channel to the desired size through bicubic interpolation. Thus, after being preserved only the luminance channel, the training images are blurred with a 5-by-5 Gaussian kernel, and then downsampled into one third of the original sizes. In all our experiments the LR patches are of size 4-by-4 and the HR patches are of size 8-by-8. The parameter λ_1 is set to be $\sqrt{(4 \times 4)/(8 \times 8)}$ to balance errors induced

Fig. 3. The two training sets used for the results reported in this work

by (7) and (8). The other parameter λ_2 is set as a relatively smaller number, $0.2\lambda_1$, because we believe that the information from the training set is more important. We have run our algorithm over two classes of images. To enrich the training set we shift the training images by 0 to 2 pixels in each direction before the training set generating process, and produce nine times more patch pairs. The experimental results can be seen in Figures 2 and 4. In each case we set K (the number of neighbors) to be 20, and carry out the algorithm for 30 iterations. For comparison, we also include the results by IBP and by Chang et al. [3] in Figure 4. Overall, the super-resolution results by our method are of satisfactory quality.

5.2 Discussions

We have proposed a new model for the single image super-resolution problem. Our approach is motivated by the manifold property of LR and HR image patches, and is fortified by the use of a three-layer Markov network. Through the proposed framework, we can directly use the information from the training data, and suppress the reconstruction error in the same time. The method thus has the advantages of both recognition-based and reconstruction-based approaches. Unlike [19], our direct energy minimization formulation guarantees reasonable reconstruction errors so there is no need to worry about that the learned information may be destroyed by depressing the reconstruction error. When compared with [15], the convex energy function, defined in (6), for the Markov network ensures better convergency property. The related work by Chang et al. [3] also starts at the manifold assumption. Suppose we use the same features to measure the distances between image patches. Then the super-resolution algorithm of [3] in fact does similar effects as those produced by our algorithm at the first iteration without using additional image prior.

A direct generalization of our method could be dealing with only primal sketch patches [19]. Furthermore, since we update the high-resolution image pixelwise,

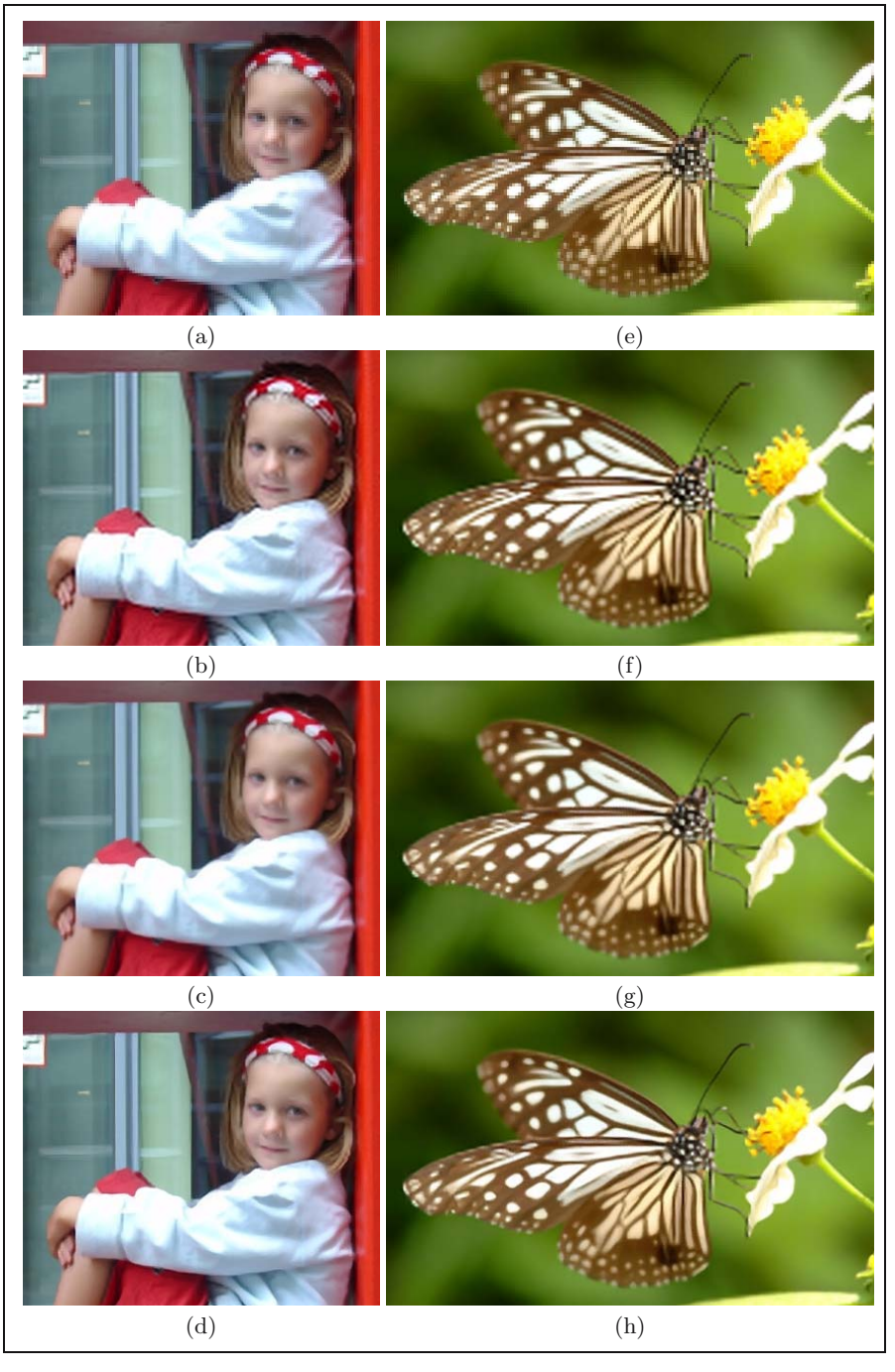

Fig. 4. From top to bottom of both columns: the low-resolution images, the results by IBP, the results by Chang et al. [3], and the results by our method

the proposed approach can be more easily extended to handle multiple image super-resolution than other recognition-based methods.

Acknowledgements

This work is supported in part by grants NSC 94-2213-E-001-020 and 94-EC-17-A-02-S1-032.

References

1. S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, September 2002.
2. F. Cao, "Good Continuations in Digital Image Level Lines," *International Conference on Computer Vision*, pp. 440–447, 2003.
3. H. Chang, D.Y. Yeung, and Y. Xiong, "Super-Resolution through Neighbor Embedding," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. I: 275–282, 2004.
4. A. Desolneux, L. Moisan, and J. Morel, "Partial Gestalts," Tech. Rep., CMLA, 2001.
5. M. Elad and A. Feuer, "Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1646–1658, December 1997.
6. S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar, "Fast and Robust Multiframe Super Resolution," *IEEE Trans. Image Processing*, vol. 13, no. 10, pp. 1327–1344, October 2004.
7. W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.
8. A. Fridman, *Mixed Markov Models*, Ph.D. thesis, Department of Mathematics, Brown University, 2000.
9. C.E. Guo, S.C. Zhu, and Y.N. Wu, "Towards a Mathematical Theory of Primal Sketch and Sketchability," *International Conference on Computer Vision*, pp. 1228–1235, 2003.
10. R.C. Hardie, K.J. Barnard, and E.E. Armstrong, "Joint MAP Registration and High Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1621–1633, December 1997.
11. A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D.H. Salesin, "Image Analogies," *Proc. of ACM SIGGRAPH'01*, pp. 327–340, 2001.
12. T. S. Huang and R. Y. Tsai, "Multi-Frame Image Restoration and Registration," *Advances in Computer Vision and Image Processing*, vol. 1, pp. 317–339, 1984.
13. M. Irani and S. Peleg, "Improving Resolution by Image Registration," *GMIP*, vol. 53, pp. 231–239, 1991.
14. A.B. Lee, K.S. Pedersen, and D. Mumford, "The Nonlinear Statistics of High-Contrast Patches in Natural Images," *Int'l J. Computer Vision*, vol. 54, no. 1-3, pp. 83–103, August 2003.
15. L. C. Pickup, S. J. Roberts, and A. Zisserman, "A Sampled Texture Prior for Image Super-Resolution," *Advances in Neural Information Processing Systems 16*, 2004.

16. S.T. Roweis and L.K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
17. R.R. Schultz and R.L. Stevenson, "Extraction of High-Resolution Frames from Video Sequences," *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 996–1011, June 1996.
18. H. Stark and P. Oskoui, "High Resolution Image Recovery form Image-Plane Arrays, Using Convex Projections," *J. Opt. Soc. Am. A*, vol. 6, pp. 1715–1726, 1989.
19. J. Sun, N.N. Zheng, H. Tao, and H.Y. Shum, "Image Hallucination with Primal Sketch Priors," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. II: 729–736, 2003.
20. D. C. Youla, "Generalized Image Restoration by the Method of Alternating Projections," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 694–702, 1978.
21. A. Zomet and S. Peleg, "Efficient Super-Resolution and Applications to Mosaics," *International Conference on Pattern Recognition*, pp. Vol I: 579–583, 2000.