Single Image Super-Resolution using Gaussian Mixture Model

Huayong He1,2, Jianhong Li1,2,3, and Xiaonan Luo1,2,3

1National Engineering Research Center of Digital Life, State-Province Joint Laboratory of Digital Home Interactive Applications, School of Information Science & Technology, Sun Yat-sen University, Guangzhou, 510006, China
2Shenzhen Digital Home Key Technology Engineering Laboratory, Shenzhen Key Laboratory of Digital Living Network and Content Service, Research Institute of Sun Yat-sen University in Shenzhen, Shenzhen, 518057, China
3Corresponding author
ljhsuperg@126.com, lnslxn@mail.sysu.cn

Abstract

In this paper we present a novel method for single image super-resolution (SR). Given an input low-resolution image, we create a pyramid pair: the ground truth pyramid and the interpolated pyramid. This method aims to model the relationship between pixel value in ground truth pyramid and its corresponding 8-neighborhood vector in interpolated pyramid using Gaussian Mixture Model (GMM). Each pixel in final high-resolution image is predicted by its corresponding 8-neighborhood vector through the trained GMM. Unlike the previous example-based SR method, our algorithm only utilizes the information of input image rather than the external image database. Our proposed algorithm achieves much better results than the state of the art algorithms in terms of both objective measurement and visual perception.

1. Introduction

The goal of image super-resolution (SR) is to obtain the corresponding high-resolution (HR) image from the given low-resolution (LR) image. SR method can be broadly categorized into three classes: 1. Interpolation-based method; 2. Reconstruction-based method; 3. Example-based method. The Interpolation-based methods [1] are fast and easy to implement but the results are blurred. The Reconstruction-based method considers image SR as a process of solving the inverse problem, which obtains the final HR image by introducing a series of prior knowledge to constraint the inverse problem and compress the solution space [2, 3]. But this approach needs much preliminary work to obtain the prior knowledge. Example-based method generates the HR image by searching for the similar image patches from training LR images database and using their corresponding HR image patches [4, 5, 6, 7, 9]. It can produce high contrast details in HR image with powerful database searching method. But the training image database needs to be selected carefully.

In this paper, we create a single image super-resolution method that makes two significant advances: Firstly, we propose a novel SR framework and construct the image pyramid pair [5] that only uses the information of input image rather than the external image database. Secondly, we consider the relationship between pixel in ground truth pyramid and its surrounding 8-Neighborhood pixels of the corresponding pixel in interpolated pyramid. These pixels are used to train a Gaussian Mixture Model (GMM). Pixels in the final HR image are estimated via the training GMM.

In the following parts of this paper, we give a brief introduction of GMM in Section 2. In section 3, we describe the detailed process of our algorithm: Pyramids construction, GMM model training and prediction for SR. Section 4 displays the results of our algorithm and the comparison with other’s.

2. Gaussian Mixture Model (GMM)
Gaussian Mixture Model (GMM) is a parametric probability density function that is represented as a weighted sum of Gaussian densities. GMM can approximate an arbitrary probability density function accurately [8]. For a $D$-dimension random variable $Z$, we can rewrite its probability density function (PDF) by GMM as,

$$ p(Z) = \sum_{i=1}^{M} \alpha_i G(Z_i, \mu_i, \Sigma_i), $$

(1)

Where $M$ is the number of component Gaussian densities, $\alpha_i$ denotes the mixture weight and $\sum_{i=1}^{M} \alpha_i = 1$, $G(Z_i, \mu_i, \Sigma_i)$ represents the $i$th component which is a $D$-dimension Gaussian function.

We denote the parameter set of the GMM as,

$$ \Theta = \{\alpha_1, \alpha_2, ..., \alpha_M, \mu_1, \mu_2, ..., \mu_M, \Sigma_1, \Sigma_2, ..., \Sigma_M\}. $$

(2)

In this paper, we use EM algorithm [8] to estimate the parameter set $\Theta$. EM algorithm alternates the parameters by implementing the E-step and M-step. For the training set $Z = \{Z_1, Z_2, ..., Z_N\}$, the estimation of the new parameters in terms of the old parameters is as follows,

$$ p(j \mid Z_i, \Theta^{old}) = \frac{\alpha_j^{old} G(Z_i, \mu_j^{old}, \Sigma_j^{old})}{\sum_{i=1}^{M} \alpha_i^{old} G(Z_i, \mu_i^{old}, \Sigma_i^{old})} $$

(3)

$$ \alpha_j^{new} = \frac{1}{N} \sum_{i=1}^{N} p(j \mid Z_i, \Theta^{old}) $$

(4)

$$ \mu_j^{new} = \frac{\sum_{i=1}^{N} Z_i p(j \mid Z_i, \Theta^{old})}{\sum_{i=1}^{N} p(j \mid Z_i, \Theta^{old})} $$

(5)

$$ \Sigma_j^{new} = \frac{\sum_{i=1}^{N} p(j \mid Z_i, \Theta^{old})(Z_i - \mu_j^{new})(Z_i - \mu_j^{new})^T}{\sum_{i=1}^{N} p(j \mid Z_i, \Theta^{old})}. $$

(6)

Note that the equation (3) performs the E-step and the equations (4) ~ (6) perform the M-step. The algorithm uses the new derived parameters as the prediction for the next iteration.

3. GMM for Super-Resolution

Figure 1 shows the framework of our algorithm, our algorithm mainly consists of three steps: Pyramid Pair Construction, GMM Training and Prediction for SR. In Section 3.1, we utilize the information about the input low-resolution image to construct a pyramid pair. In Section 3.2, the pyramid pair is used to produce the training data set and train a GMM. In Section 3.3 it provides a way to estimate the final high-resolution image using GMM.

3.1. Pyramid Pair Construction

Motivated by [7], we construct a pyramid pair: the ground truth pyramid $P_H$ and the interpolated pyramid $P_L$. Fig 1(a) shows the structure of pyramid pair $P_H$ and $P_L$. $P_H$ consists of the image set $\{H_i\}, i = 2, ..., 1$. $H_0$ is the input low-resolution image and $H_1$ is the final high-resolution image to be estimated. The image $H_i$ is generated from $H_{i+1}$ by the following operation,

$$ H_i = (H_{i+1} \ast B) \downarrow s $$

(7)

where $\ast$ denotes the convolution operation, $B$ is a blur matrix, $\downarrow$ represents a downsampling operation and $s$ is the scale reduction factor between $H_i$ and $H_{i+1}$. The pyramid $P_L$ is made up of the image set $\{L_i\}, i = 2, ..., 1$. Image $L_i$ is generated from $H_{i+1}$ using bicubic interpolation by the scale factor $s$.
It’s obvious that the pyramid $P_n$ is corresponds to pyramid $P_L$. Since $H_1$ is the final high-resolution image, we could estimate $H_1$ by the relationship between $H_1$ and $L_i$. Compared with prior example-based methods, we do not utilize the external image database but only use the input LR image $H_0$.

3.2. GMM Training

The pixels with the similar edge tend to have similar neighborhoods whose intensities change fast in the direction perpendicular to that of the edge and pixels in a smooth region seems to have relatively invariant intensities within the neighborhood [9]. The similarity of two pixels can be reflected as the similarity of their corresponding surrounding 8-Neighborhood pixels.

In our method we consider the relationship between pixel in pyramid $P_n$ and the surrounding 8-Neighborhood pixels $X$ of the corresponding pixel in $P_L$.

Figure 1(a) shows the structure of a training data $Z$. The black point $y_{l,i}$ corresponds to $y_{l,i}$. The red points in $L_0$ represents the 8-neighborhood pixels $X$ for $y_{l,i}$. Each training data consists of $y_{l,i}$ and $X$, so $Z$ can be rewritten as,

$$Z = [Y, X], \quad (8)$$

For each image pair $H_i$ and $L_i$, $i = -2, -1, 0$. We get the training data for each pixel in $H_i$. Now we define training data as random variables $Z = \{Z_1, Z_2, ..., Z_N\}$ where $N$ is the number of training data. We can train a GMM $p(Y, X)$ for $Z$.

3.3. Prediction for SR

We consider the parameter set $\Theta$ and rewrite

$$\Theta = [\mu_m, \mu_y], \quad \Sigma_y = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix},$$

we can get

$$p(Y, X) = \sum_{j=1}^{M} \alpha \cdot G(Z, \mu_j, \Sigma_j)$$

$$= \sum_{j=1}^{M} \alpha \cdot \left[ \frac{G(Y \mid X, \mu_{j,i}, \Sigma_{j,i}) \cdot G(X, \mu_j, \Sigma_j)}{p(X)} \right]$$

Where

$$\mu_{j,i} = \mu_y - \Sigma_{xy} \Sigma_{yy}^{-1} (\mu_x - X)$$

$$\Sigma_{j,i} = \Sigma_{yy} - \Sigma_{xy} \Sigma_{xx}^{-1} \Sigma_{yx}.$$  

If $p(Y \mid X)$ is known, the optimal estimation of the pixels $Y$ in $H_i$ can be given as

$$\hat{Y} = E(Y \mid X)$$  

under the Minimum Mean Square Error (M.M.S.E)[10].

Combining the equation (9), we can easily get the $p(Y \mid X)$ by Bayesian Theorem. That is

$$p(Y \mid X) = \frac{p(Y, X)}{p(X)} = \sum_{i=1}^{M} \beta G(Y \mid X, \mu_{j,i}, \Sigma_{j,i})$$

Where

$$\beta_i = \frac{\alpha \cdot G(X, \mu_j, \Sigma_j)}{\sum_{j=1}^{M} \alpha \cdot G(X, \mu_j, \Sigma_j)}$$  

The optimal forecasting $\hat{Y}$ can be represented as the following form [10]:

$$\hat{Y} = E(Y \mid X) = \sum_{i=1}^{M} \beta_i \mu_{j,i}$$  

Where the $\mu_{j,i}$ is the same as in equation (10) and (13).

For each pixel $y$ in the final high-resolution image $H_1$, it can be predicted by equation (14) and the corresponding vector $X$.

4. Experiments

Our method has two parameters to determine. The first parameter is the number of component Gaussian densities $M$ in GMM. We set $M$ to 5 for all our experiments. The second parameter is the level $K$ of pyramid. We set $K$ to 4 for preventing the smaller size of the lowest level image.

The PSNR and SSIM values of four different test images were displayed in Table 1 (the scaling factor is 2). To compare our method with other SR methods, we consider bicubic interpolation method, Locally Linear Embedding (LLE) [7] and Yang et al. [11]. In Table 1, the first row in each method is the PSNR and the second row is the SSIM. We can see that except for flower image, our method obtained higher PSNR and SSIM value than the other three algorithms.

<table>
<thead>
<tr>
<th></th>
<th>tiger</th>
<th>pepper</th>
<th>flower</th>
<th>fruit</th>
</tr>
</thead>
<tbody>
<tr>
<td>bicubic</td>
<td>26.86</td>
<td>27.93</td>
<td>32.54</td>
<td>27.03</td>
</tr>
<tr>
<td>LLE [7]</td>
<td>0.882</td>
<td>0.922</td>
<td>0.891</td>
<td>0.880</td>
</tr>
<tr>
<td>Our method</td>
<td>0.813</td>
<td>0.874</td>
<td>0.856</td>
<td>0.830</td>
</tr>
</tbody>
</table>
Figure 2 shows the results of using different super-resolution algorithms to a tiger image for 4X magnification. We compare our method with the bicubic interpolation and Yang et al. [11]. It’s clear that our result is better especially in textual parts than the other two algorithms. Our approach produced less noise and artifacts. Figure 3 shows the comparison of Kim et al. [4] and our proposed with the scaling factor of 4. In the enlarged region, we can see that our result shows clearer carved curve than the result of Kim.

5. Conclusions

We proposed a novel algorithm for single image SR in this paper. We construct a pyramid pair: the ground truth image pyramid and the interpolation image pyramid using only the input LR image without external image database. The pixels in ground truth pyramid and the corresponding 8-neighbor pixels are used to train GMM. The final HR image is predicted through the training GMM. Compared with interpolation-based method and other learning-based method, our approach achieved better result.

6. Acknowledgement

This research is supported by NSFC-Guangdong Joint Fund (U0935004, U1135003), the National Key Technology R&D Program (2011BAH27B01, 2011BHA16B08), and the Industry-academy-research Project of Guangdong (2011A091000032).

References