Image Inpainting as an Evolving Topic in Image Engineering

Yu-Jin Zhang
Department of Electronic Engineering, Tsinghua University, China

INTRODUCTION

Image inpainting (also known as image interpolation or video interpolation) is an evolving topic in image engineering (IE), especially in image processing (IP) layer (Zhang, 2009). It aims to automatically improving the quality of image, by filling in the holes of image that are to be removed for artistically or confidential considerations. Research in this field is on the rise and many applications are created (Abraham, 2012; Janarthanan, 2012; Modi, 2012).

How can we make computer modify an image in a visually undetectable way analogous to sophisticated artists? This is quite an amazing topic in both art and image processing areas, with applications from the restoration of damaged paintings and photographs to the removal of selected objects. Image inpainting exactly aims to reconstitute the missing region (called the target region) using information from the remaining image areas (called the source region).

Until now, there are two main categories of image inpainting approaches in the literature: diffusion-based (or PDE-based) approach (Chan, 2001; Tibshirani, 1997) and exemplar-based approach (Criminisi, 2003; Shen, 2009; Wong, 2008; Xu, 2010). In the following, two recent works, one from each category, are introduced as example (some other related methods can be found in references). One uses a patch-wise inpainting algorithm with sparse representation over a redundant dictionary. It is suitable both to deal with large holes and to preserve image details while taking less risk. To ensure the visually plausibility and consistency constraints between the filled hole and the surroundings, a redundant dictionary is constructed by directly sampling from the intact source region of current image. Then, the sparse representation for each incomplete patch at the boundary of the hole is sequentially computed and recovered until the whole hole is filled. Another uses a patch propagation-inpainting algorithm based on Weighted Sparse Non-negative Matrix Factorization (WSNMF). In this approach, the inpainting task is casted as a sequential low-rank matrix recovery and completion problem, where the incomplete data matrix consists of the image patch to be inpainted and several similar intact candidate patches under the assumption that they can be described using a low-dimensional linear model.

BACKGROUND

They differ from each other on the focused image levels, and thus have disparate performance. The former tackles the filling-in problem by diffusing the image from the known surrounding regions into the missing region at the pixel level by using the variational principles and partial differential equations (PDE). Hence, it is superior for structure propagation or relatively smaller missing region, yet poor in handling textured or large region due to the introduction of smoothing effect. The latter propagates the image information at the patch level based on the texture synthesis technique. By incorporating the patch priorities to determine the filling order, the exemplar-based method can deal with structure propagation as well as texture propagation, and hence outperforms the diffusion-based one with respect to large missing region.

One of the core stages in exemplar-based inpainting algorithms is how to synthesize the needed texture by exploiting the known candidate patches from the source region. One typical approach (Criminisi, 2003) considers the structure propagation as well as texture synthesis by computing patch priorities for determining the filling order. It selects the most suitable patch from the source region at each step. Some improvements have also been proposed (Chen 2007).

The challenge of image inpainting roots deeply in the nature of real-world scene photographs, which often

DOI: 10.4018/978-1-4666-5888-2.ch122

consists of 1-D or 0-D linear structures, such as edges and corners, and 2-D pure or composite textures. The boundaries (or called the fill front) between the target and source regions are a complex product of the mutual influences of these factors. In this sense, propagating the spatially interacted multiple textures while preserving the structures becomes the core concern.

SPARSE REPRESENTATION

Between the two types of inpainting approaches, it has found that the exemplar based approaches work well to synthesize the texture in target areas, thus are more suitable to deal with large regions, such as in the cases of removing background persons from a picture, keeping details in the filled regions of image, etc. However, as it always selects the most suitable patch for the current place, it is a greedy method, which results in a risk of introducing unwanted object or artifact to the area to inpaint. Some examples will be given in the section of experiment results.

To solve the above-mentioned problem, sparse representation technique has been incorporated in the inpainting process to bridge the gap between sparse representation and texture synthesis. Signal sparse representation means that the signal admits a sparse representation over a redundant dictionary; see below for discussion that is more detailed. Thus, the inpainting problem is viewed as the recovery of incomplete image signals, with each signal corresponding to a patch. The hole is filled patch-wisely based on the sparse representation of each patch.

Lasso

Lasso is a regression method (Tibshirani, 1997), in which a L_1 norm penalty to the loss function of ordinary least square regression, which results in the sparsity of the coefficient.

Given the dictionary $\boldsymbol{x} = [x^1, x^2, \ldots, x^N]$ and input signal $\boldsymbol{y} = [y_1, y_2, \ldots, y_P]^T$. For convenience, assume the data $\boldsymbol{x}$ and $\boldsymbol{y}$ are normalized. The Lasso algorithm is to estimate the coefficient $\boldsymbol{\beta}$ of a signal over the given dictionary by

$$\hat{\boldsymbol{\beta}} = \arg\min\left\{\left\|\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right\|_2^2 + \lambda\left\|\boldsymbol{\beta}\right\|_1\right\} \tag{1}$$

The term $\|\boldsymbol{\beta}\|_1$ encourages the sparsity of the fitted coefficient vector, and the parameter λ controls the tradeoff between the reconstruction error and the sparsity. This formulation is based on the model that $\boldsymbol{y} = \boldsymbol{x\beta}$ and $\boldsymbol{\beta}$ is sparse, i.e. only a few components of $\boldsymbol{\beta}$ are nonzero.

It is more interesting, when some components of signal are corrupted, which means the model is modified into

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \mathbf{e} \tag{2}$$

where $\boldsymbol{e}$ is the error and the e_i is nonzero if and only if the y_i is corrupted. The corrupted signal component index set is denoted by I, $I = \{i | e_i \neq 0\}$. Let $\boldsymbol{y}_{\backslash I}$ denotes the vector obtained by removing the components whose indexes are in I from $\boldsymbol{y}$, i.e. $\boldsymbol{y}_{\backslash I}$ is made up by all the non-corrupted components of $\boldsymbol{y}$. $\boldsymbol{x}_{\backslash I}$ is the corresponding dictionary matrix, which is obtained by removing all the columns whose indexes are in I from $\boldsymbol{x}$. Now the sparse coefficient $\boldsymbol{\beta}$ can be estimated as follows.

$$\hat{\boldsymbol{\beta}} = \arg\min\left\{\left\|\mathbf{y}_{\backslash I} - \mathbf{x}_{\backslash I}\boldsymbol{\beta}\right\|_2^2 + \lambda\left\|\boldsymbol{\beta}\right\|_1\right\} \tag{3}$$

Then, the corrupted signal can be we recovered via (3):

$$\hat{y}_i = \begin{cases} y_i & i \notin I \\ (\mathbf{x}^{\hat{\beta}})_i & i \in I \end{cases} \tag{4}$$

As a result, by combining (3) and (4), the corrupted signal can be recovered.

Algorithm

The main tasks in the algorithm for image inpainting via sparse representation are as follows.

Filling Order

A crucial technique in exemplar-based inpainting algorithm is how to determine the filling order so as to balance the recovery of both texture and structure. Given an input image, the user selects the target region, which is to be removed and filled. Then usually the