Image Inpainting with a Learned Guidance Vector Field

Yuan Ren Loke and Surendra Ranganath
Department of Electrical and Computer Engineering
National University of Singapore
Singapore
{g0500069, elesr}@nus.edu.sg

Abstract—Image inpainting is one of the challenging problems in image restoration. To recover the missing region, we can only rely on the information in the uncorrupted region of the input image and some prior knowledge. The latter can be learned from suitable training data or implemented through some smoothness constraints. In this paper, a new approach for image inpainting is proposed. Here, we iteratively learn a guidance vector field from training data and recover the missing region by solving the Poisson equation using the learned guidance vector field with Dirichlet boundary conditions. In addition, we also propose a method to select the best training set by using the correlation between neighboring patches of the damaged input image and training images. The experimental results on face images show that the new approach yields smooth and visually pleasing results.

I. INTRODUCTION

Image inpainting was initially used to restore deteriorated artwork. In image processing, this technique can be adopted to not only recover the damaged region, but can also be extended to texture synthesis [1], disocclusion [2], and super-resolution [3]. In [4], Bertalmio et al. also applied their inpainting approach for super-resolution. Liu et. al. [5] used an inpainting technique on an image to remove a fence which was occluding a target object.

Digital image inpainting techniques can be generally categorized into three approaches, PDE-based, exemplar-based and learning-based. PDE-based image inpainting propagates the image Laplacian in isophote directions from the exterior to fill in the missing region, and was introduced by Bertalmio et al. [6]. Later, Bertalmio et al. [4] extended this by applying the Navier-Stokes equation from fluid dynamics. Chan et al. [7] also applied a Euler-Lagrange method for image inpainting. Pérez et al. [8] solved the Poisson partial differential equation with Dirichlet boundary conditions and a given guidance vector field (GVF). All these smoothness constraint based image inpainting approaches work well on small regions for removing subtitles, watermarking, etc. because these approaches only rely on the smoothness constraint functions and boundary information. No other information about the missing region is used. Thus, undesirable blur usually occurs when inpainting larger missing regions, as for example in Fig. 1 (c).

Exemplar-based image inpainting [9], [10], [1], [11], techniques are another group of approaches which synthesize the missing region from patches obtained from the uncorrupted regions of the image or a database. Basically, they search for optimum patches from the available ones to fill in the missing region such that the boundaries between neighboring patches in the inpainted image are smooth. These techniques are usually used for removing a bigger foreground object appearing over a background such as grass, sky, water etc. and inpainting the object-removed region to merge smoothly with existing background; this did not consider any structural information. The technique was extended by Criminisi et al. [12] and Sun et al. [13] to consider the structural information around the missing region by imposing constraints. This information helps to inpaint the missing region on more general background scenes. However, a unique structure such as a human face is difficult to learn directly from the uncorrupted region and manually imposing hard constraints is not trivial and perhaps impractical. An example of applying exemplar-based inpainting of [12], is shown in Fig. 1 (d), from which it can be seen that it is not always easy to seek similar patches from the available region to fill in the missing portion. Therefore, some kind of training process is required for inpainting specific object classes.

![Fig. 1.](image)

Learning-based image inpainting can be considered as a more advanced type of exemplar-based image inpainting. It does not directly find the best patch from the dataset to fill
in the missing region. Rather, it learns the structure from the uncorrupted region and a training data set to synthesize a new patch for the missing region. The new patch may not be exactly the same as in the training data set. In [14], Levin et. al. estimated the missing region by using the image statistics of the training images. Roth and Black [15] used Markov random field to learn the missing region from small image patches. Turaga and Chen [16] used a mixture of eigenspaces to recover corrupted video frames.

Since information is lost, it is not possible in general to recover it exactly. Thus, image inpainting is generally more concerned with recovering a realistic, visually pleasing reconstruction instead of exactly filling in the missing region. This is the main difference from other image restoration problems such as noise reduction which seek to recover the original information as faithfully as possible from the noisy data.

II. MOTIVATION AND OVERVIEW

In this paper, we seek to inpaint face images containing missing regions using training images and the undamaged region of the given image. Our approach is based on the iterative projection onto convex sets (POCS) algorithm. POCS can iteratively reconstruct a signal by incorporating multiple convex constraints. It was initially proposed by Papoulis [17] and Gerchberg [18] for signal extrapolation and is also known as the Gerchberg-Papoulis algorithm. It has since been applied in various image restoration and image enhancement problems. Stark and Oskoui [19] applied POCS with data constraints and prior knowledge for super-resolution of images. Tekalp et al. [20] also applied POCS to restore out-of-focus blurry images. Bandwidth constraints, spatial support contraints, consistency, positivity, etc. are typical examples of the convex constraints used in POCS approaches.

For generating a realistic and visually pleasing face from a corrupted face image, we need to satisfy two criteria. The first is to retain the structure of the face faithfully. Faces have unique structure with regions such as eyes, nose and mouth, and if these unique structures are missing, there is no way to find a similar patch directly from the undamaged region of the image. Thus, we need to rely on the help of training images to learn the structure, and use the learned model to provide the structure of the missing region.

The second criterion is the smoothness between the recovered region and its surroundings. To have a realistic and visually pleasing effect, smoothness or continuity of the textural details is as important as the structure of the image. Thus, we seek the solution to satisfy a set of Poisson equations with Dirichlet boundary conditions. These two criteria can jointly retain the structure of the image as well as maintain continuity between the inpainted region and its surroundings.

The details of our approach are discussed in Section III. In short, high frequency or structural details of the face are learned from the training data set by Principal Component Analysis. This is used in conjunction with the solution of the Poisson equation with Dirichlet boundary conditions, in an iterative POCS algorithm. In section IV, we also propose a method for selecting the best patches for learning the model to improve the effectiveness and robustness of our approach. Section V gives the experimental results, and the conclusions are in Section VI.

III. THE APPROACH

A. PCA Based Image Inpainting

We can use Principal Component Analysis (PCA) based image inpainting which learns the PCA model of the missing region from training images and then recovers the missing region from the model. In this technique, a region of interest (ROI), \( \mathcal{R} \), is defined as shown in Fig. 2 to enclose the missing region, \( \mathcal{M} \), within a bounding band of known pixels, \( \mathcal{B} \). The corresponding ROI vectors, \( \mathbf{F} \) from the training images are used to extract a PCA model, \( \mathcal{L} \), represented by a set of orthonormal bases, \( \mathbf{I}_i \) and mean, \( \mathbf{F} \). An ROI, \( \mathcal{F} \), can be reconstructed from \( \mathcal{L} \) as

\[
\hat{\mathbf{F}} = \sum_i \alpha_i \mathbf{I}_i + \mathbf{F}
\]  

where \( \alpha_i \) are the projection weights of \( \mathbf{F} \) onto the basis vectors. \( \hat{\mathbf{F}} \) will differ from the original \( \mathbf{F} \) in two ways: firstly, \( \mathcal{M} \) will be filled in by the model, which is desired, and secondly the original uncorrupted pixels in \( \mathcal{B} \) will be altered. The latter effect is undesirable, and hence it is necessary to impose a consistency constraint and replace all the \( \mathcal{B} \) pixels in \( \hat{\mathbf{F}} \) by the original pixels. This results in an updated \( \hat{\mathbf{F}} \), but there is a noticeable discontinuity between the filled region, \( \mathcal{M} \), and the bounding pixels, \( \mathcal{B} \). In order to obtain a better estimate for \( \mathcal{M} \), reconstruction of the ROI, \( \hat{\mathbf{F}} \), from \( \mathcal{L} \) and imposition of the consistency constraint can be iterated in a POCS algorithm until convergence. This will improve the estimate of pixels in \( \mathcal{M} \); however, the discontinuity between \( \mathcal{M} \) and \( \mathcal{B} \) is ignored in this process, and hence may be present even though the inpainted region looks acceptable. The results are shown in Section V. Hence, we propose to learn a GVF from the training data and use it in a Poisson equation based solution, and obtain obtain a smooth transition.

![Illustration of missing region, \( \mathcal{M} \), and bounding band, \( \mathcal{B} \), in ROI.](Image 390x202 to 441x263)

B. Guidance Vector Field Image Inpainting

Guidance Vector Field (GVF) image inpainting was originally used for seamlessly editing image regions by Pérez et. al. [8]. The solution is obtained by solving a set of Poisson equations with Dirichlet boundary conditions with a prescribed GVF. Here, the problem can be formulated as a minimization problem:
\[
\min_{f|_M} \sum_{i,j} \mathbb{1}_{M \neq 0} (f_i - f_j - v_{ij})^2 \\
\text{s.t.} \quad f_i = f^*_i, \quad \forall i \in \partial M
\]  

(2)

where \(f_i\) denotes the value of the pixel at location \(i\), \(M\) (as before) denotes a closed missing region, \(f|_M\) denotes the values of the pixels in \(M\), \(v_{ij}\) denotes the gradient at the mid point of location \(i\) and \(j\), \(\partial M\) denotes the boundary subset and \(f^*_i\) denotes the given input value of pixel \(i\). The solution can be obtained through the Gauss-Seidel solver. This inpainted solution will have a gradient field that is close to the given GVF, with a seamless transition between \(M\) and the boundary, \(\partial M\).

However, the structure of the missing region depends highly on the GVF chosen. In the case of faces which have a unique structure in different regions, it is not possible to obtain the GVF from an undamaged region. The GVF can only be learned from other sources, and here, we propose a new approach to iteratively learn the GVF from training images.

C. Iterative Learning of Guidance Vector Field for Image Inpainting

Our proposed method for learning the GVF for image inpainting is a POCS based iterative image inpainting technique which seeks the solution which satisfies two different constraints. The first is a structure constraint which extends the concept of the PCA based image inpainting in Section III A. For this, we model the high frequency details of the ROI, \(F\), by a set of orthonormal bases obtained by PCA. It is defined as:

\[
\nabla \hat{F} = \sum_i \beta_i h_i + \bar{H}
\]

(3)

where \(\bar{H}\) denotes the mean of the gradient of the training ROIs, \(\nabla \hat{F}\) denotes the reconstructed gradient of \(F\), \(h_i\) denote the orthonormal basis vectors of the high frequency detail space and \(\beta_i\) are the projection weights.

The second constraint imposes smoothness through GVF based inpainting as discussed above. The \(\nabla \hat{F}\) obtained from the first (PCA model) constraint is used as the GVF for inpainting. Then, the solution obtained from this second constraint is input to the PCA model and reconstructed. This process is iteratively repeated until convergence. The solution will satisfy the facial structure constraints, and also have a seamless boundary between \(M\) and \(B\). The algorithm is summarized in Algorithm 1. The algorithm complexity is \(O(m^2)\) where \(m = |F|\). Two example results are shown in Fig. 3.

IV. PATCH SELECTION FOR LEARNING PCA MODEL

The face is a non-rigid structure which changes significantly with facial expression. For example, there is a large difference between the happy face (open mouth with teeth visible) and the neutral face (mouth closed). To account for this variation, we manually mark points on all the training images (this can be easily automated, for example, by using active shape model), and form a mesh by Delaunay triangulation. Following this, all training images are warped onto the triangulated mean face mesh. In this paper, we assume that only one triangular patch is corrupted and needs to be inpainted (though it is easy to generalize this). Hence, a corrupted patch will have three neighboring triangular. Here, instead of using the neighboring patches from all the training images, we propose a patch selection algorithm to select the appropriate training images from which the patches will be used for learning.

Let \(M\) and \(\{B_i|i = 1, \ldots, m\}\) (here \(m = 3\)) denote the missing region of the input image and its neighboring patches, respectively, and let \(M^j\) and \(B_i^j\) denote the corresponding patches in the training image \(j\). The subscript \(i\) of \(B_i\) denotes the index of the patch. The distance between each \(B_i^j\) and \(B_i\) is measured by the L2-norm, \(d(B_i^j, B_i)\). The \(k\) (we use \(k = 5\)) training images with the smallest \(d\) are selected and the corresponding indexes of these \(k\) training images form an index set, \(\mathcal{L}_i\). The indexes of the images used to form the training set is given by \(\bigcap \mathcal{L}_i\). The algorithm is summarized in Algorithm 2.

V. EXPERIMENTS

In our experiments, 15 aligned images obtained from the Yale Face Database B [21] and 31 images from an image sequence with different expressions were used. Four experiments
Input: ROI from $n$ training images, $L_j$, $j \in \{1, \ldots, n\}$, ROI of a corrupted image, $F$
Output: Selected training images, 
\[\{L_s|\{s\} \subset \{j\}, |\{s\}| \leq k\}\].

forall $i$ do
forall $j$ do
    Compute $d(B_i^j, B_i)$.
end
1) Sort the training data based on $d$.
2) $I_i \leftarrow \{\text{corresponding indexes of the } k \text{ nearest training images}\}$
end

Images with index in $\cap I$, $\{L_s\}$, are selected as training images for Algorithm 1.

Algorithm 2: Algorithm for patch selection.

Fig. 4. Image inpainting when a subject is in the training database or not: (a) the corrupted images, (b) inpainted images when the subject is not in the training database, (c) inpainted images when the subject is in the training database and (d) original images.

were performed to evaluate our proposed image inpainting approach.

In the first experiment, the results when a test subject was included or excluded from the training set were compared. In the former case, all 15 aligned faces were used for training and a test subject was selected from one of them. For the latter case, 14 of the 15 aligned faces were used for training and the remaining image was used for test. The experimental results are shown in Fig. 4. It can be seen that regardless of whether the subject is in the training database or not, the inpainting is realistic, though when the subject is in the training set, the inpainted region is closer to the original.

In the second experiment, our approach in Algorithm 1 is compared with image inpainting obtained based on PCA only and only on Poisson image inpainting. Here, 14 of the 15 aligned images are used for training and the remaining image is used for test. Since the GVF for Poisson image inpainting is unknown, we simply assumed the GVF as zeros. The results are shown in Fig. 5. For PCA-based image inpainting, the boundary of the filled missing region can be clearly seen although the structure of the missing region is well synthesized from the face eigenspace. With the Poisson image inpainting, the inpainted boundary is seamless compared to the PCA-based image inpainting but it fails to recover the structure of the missing region because the GVF is not known. The results produced by our approach not only retain the structure of the face, but also have a smoother and visually pleasing seamless boundary.

In the third experiment, 31 images from a video sequence with different expressions were used to evaluate our inpainting method with patches selected for the PCA model as discussed in Section IV. 55 markers were labeled on the 31 images, and the mean of the 55 markers over the training images was used to form a face mesh by Delaunay triangulation. All 31 face images were warped onto the triangulated mean face mesh with 55 markers, and this was used to form the training set. Similar to the second experiment, comparison between a test image included or excluded from the training set was done. When the test image was included in the training database, all 31 warped image were available for training, and in the other case, the test was excluded from the training set. One of the face images and its corresponding warped face image are shown in Fig. 6.

One of the triangular patches was removed from a test image to simulate damage. The adjacent patches of the damaged patch, $B_i$ and the corresponding patches in the training images, $B_i^j$ were extracted and the patch selection method of Section IV was applied to choose the best patches for learning. The selected training data obtained from Algorithm 2 was used in our inpainting approach to recover the damaged region. The results are shown in Fig. 7. It can be seen that with our approach, the quality of inpainted images is realistic even if the subject is not in the training database. The experiment also
We first generated resolution simultaneously by a POCS-based technique [22]. The results are shown in Fig. 8. This experiment demonstrated by experiments. A patch selection scheme for the learning model significantly reduced the complexity of modeling the missing region space and improved the effectiveness of the new approach.

In this paper, a new POCS based image inpainting technique was proposed to iteratively learn the guidance vector field for solving the Poisson equation with Dirichlet boundary conditions. Experiments showed that the performance of the new approach is much better than solely applying PCA-based image inpainting or GVF image inpainting. The new approach not only retains the structure of the missing region, but also smoothens the boundary between the missing region and its neighbours. The robustness of the new approach has been demonstrated by experiments. A patch selection scheme for the learning model significantly reduced the complexity of modeling the missing region space and improved the effectiveness of the new approach.

VI. CONCLUSION

In this paper, a new POCS based image inpainting technique was proposed to iteratively learn the guidance vector field for solving the Poisson equation with Dirichlet boundary conditions. Experiments showed that the performance of the new approach is much better than solely applying PCA-based image inpainting or GVF image inpainting. The new approach

Fig. 7. Image inpainting by patch-based learning model: (a) the original images, (b) corrupted images, (c) inpainted images when the test image is not in the training set and (d) inpainted images when the test image is in the training set.

Fig. 8. Image inpainting and super-resolution: (a) the original high resolution image, (b) the damaged low resolution image, (c) the interpolated inpainted image and (d) the super-resolved inpainted image.

showed that the computational time is significantly reduced because fewer images were used for training the PCA model. Besides that, better representation of missing region improves the inpainting result, even though facial expression differ.

In the last experiment, we considered inpainting and super resolution simultaneously by a POCS-based technique [22]. We first generated 60 × 50 low resolution images from the 15 120 × 100 high resolution Yale faces by 2 × 2 averaging, and replacing those pixels by the average value. A portion of the low resolution test image was removed to simulate damage. In this experiment, we seek not only to recover the damaged region, but also to super-resolve the image back to high resolution with sharp texture details. For this, we first inpainted the damaged region in low resolution by our proposed approach and then the inpainted image was super-solved [22]. The results are shown in Fig. 8. This experiment shows that our inpainting and super-resolution approach can be applied to recover and enhance low quality images captured by CCTV cameras as commonly used in surveillance systems.

REFERENCES