Exemplar based Inpainting using Autoregressive Parameter Estimation

Milind G. Padalkar* and Manjunath V. Joshi†
DA-IICT, Gandhinagar,
Gujarat, India
Email: *milind_padalkar@daiict.ac.in
†nv_joshi@daiict.ac.in

Mukesh A. Zaveri
Computer Engineering Department,
SVNIT, Surat,
Gujarat, India
Email: mazaveri@coed.svnit.ac.in

Chintan M. Parmar
Information Technology Department,
V.V.P. Engineering College,
Rajkot, Gujarat, India
Email: chintan.4287@gmail.com

Abstract—In this paper, we propose an iterative exemplar based inpainting technique. A set of exemplars is automatically searched considering a window around every pixel in the region to be inpainted. The novelty of our technique lies in the use of parameters of an autoregressive (AR) model that are estimated using the non-negatively constrained least squares (NNLS) method. Unlike simply copying into the target or damaged pixels (i.e. pixels to be inpainted), values from the best matching exemplar as is done by many of the exemplar based methods, we use the AR parameters along with the best matching exemplar to modify the values of the damaged pixels. The AR parameters for a set of candidate exemplars suggest the contribution of values of neighboring pixels towards the respective center pixel of every $3 \times 3$ region in that set. Determining a pixel value with the knowledge of spatial relationship of pixels and information from an exemplar avoids discontinuity that arises due to direct copying of pixels from the exemplar. Results are illustrated with the help of true images captured from the world heritage site Hampi, Karnataka.

Index Terms—Exemplar; Inpainting; Autoregressive model; NNLS;

I. INTRODUCTION

Image inpainting is the process of restoring or modifying the image contents imperceptibly. Given an image and a region of interest in it, the task of an inpainting process is to fill up the pixels in this region in such a way that either the original content is restored or the region is visually plausible in the context of the image. Image inpainting can be used for a number of applications that require automatic restoration or retouching some region of a photograph. In fact, the term inpainting is derived from the art of restoring damaged images in museums by professional restorers [1]. Although restoration and inpainting are used interchangeably, inpainting can be considered as a superset of restoration. In general, restoration refers to undoing of degradation, while inpainting also allows creating special effects such as removing / adding objects.

There are various heritage sites across the world which are of immense interest to human beings to know their history. However at many sites monuments or statues are ruined due to various reasons. Hence it would be interesting if one looks at these sites by digitally restoring the monuments as they existed in the past and use it as means for attracting the tourists. This would also serve as a tool to educate the general public on the culture of ancient civilization. Towards this end, we propose a method based on exemplar search with estimation of autoregressive (AR) parameters. A set of candidate exemplars or matching patches is searched automatically based on the sum of squared differences (SSD) criteria. The AR parameters are estimated from the set of candidate exemplars using the NNLS approach. The exemplar with least SSD, along with the estimated parameters are used for modifying the pixel values in the damaged regions.

Image inpainting has been an active area of research for the past decade. Although during the 1990s many researchers addressed the problem of interpolating missing pixel values, the first attempt towards inpainting was done by Masnou and Morel [2]. They connected the level lines (i.e. curves of constant intensity) arriving at the boundary of an occluded region. With this technique the occluded region was filled up, but the desired curving of the level lines was not obtained. A major contribution to inpainting was later provided by Bertalmmio et al. [1]. Their algorithm not only connected the level lines arriving at the boundary of the occluded regions, but also enabled their curving inside the occluded region in a visually plausible manner. The periodical curving of the level lines that also avoids crossings, is achieved using anisotropic diffusion [3, 4]. This algorithm is successful in propagating structure, but fails in the presence of large textured regions. Other works based on level lines are presented in [5–8].

Pérez et al. [9] introduced the Poisson image editing technique for adding objects / texture from other images by using them as guidance vector field. Thus, if a user supplies a region of interest to be edited and a region from another (or same) image from which information is to be transferred (guidance vector field), a seamless blending is achieved with Poisson image editing. However, if this technique is to be used for inpainting, the guidance vector field needs to be manually selected, leading to highly subjective results. Moreover, the objective function to be minimized determines the inpainted pixel values as the average of neighbouring pixel values, which is not true for all images. This motivates us to use an AR model based approach for inpainting.

For propagating texture, block replicating methods have been suggested in [10, 11]. These techniques exploit the fact
that for a small patch at the boundary of the occluded region, a similar patch (considering only the non-occluded part of the patches) can be found from the surrounding. The occluded part of the patch is then filled up by copying corresponding pixels from the similar patch or the exemplar. They emphasize on order of selecting the patch to be filled up, which allows the propagation of structure as well as texture. However, due to simply copying of the pixels, discontinuities do appear where even the most similar exemplar is considerably dissimilar from the patch to be filled up.

Our work addresses limitation of the widely used inpainting algorithms. We propose a method that is based on [10], capable of automatically searching a set of exemplars similar to a patch under consideration. For an image, because of the spatial correlation of a pixel value on its neighbours, an AR model can be used to express this dependence where a pixel value is a linear combination of values of its neighboring pixels [12]. We assume the neighbourhood to be the first order neighbourhood. Therefore, once the set of candidate exemplars is available we extract AR parameters considering the dependence of a pixel on its first order neighbours. Using the estimated AR parameters and the best matching exemplar as a guidance vector field, the pixel values in the target region of the patch under consideration are determined in such a way that they obey the derived relationship with the neighbours. Since the AR parameters suggest the contribution of neighbours, a NNLS method [13, 14] is used to calculate the values of these parameters. A simple least squares (LS) method is unsuitable for this purpose as this may give negative values for the AR parameters which does not correspond to correct fractions.

The rest of the paper is organized as follows. In section II we describe the details of our proposed algorithm. We present the experimental results in section III followed by conclusion in section IV.

II. PROPOSED APPROACH

Our proposed technique is shown in figure 1. We start with an input image $I$ and a user defined mask that gives the region to be inpainted or the target region $\Omega$. The boundary of the target region $\Omega$ is denoted by $\delta \Omega$. Similar to [10], for every pixel $p$ in $I$, a confidence term $C(p)$ and data term $D(p)$ are calculated. Initially, the confidence term (initial confidence $c(p)$) is calculated as

$$c(p) = \begin{cases} 1, & \forall p \in I - \Omega \\ 0, & \forall p \in \Omega \end{cases}$$

Now considering a fixed size neighbourhood $\Psi_p$ around every pixel $p$, the confidence term $C(p)$ is calculated as

$$C(p) = \frac{\sum_{q \in \Psi_p} c(q)}{|\Psi_p|}$$

where $|\Psi_p|$ is the area of the neighbourhood $\Psi_p$ of pixel $p$. The data term $D(p)$ is calculated as

$$D(p) = \frac{|\nabla I_p^\perp \cdot n_p|}{\alpha}$$

where $\nabla I_p^\perp$ is orthogonal to the gradient $\nabla I_p$ at a pixel $p$, $n_p$ is unit normal to the target boundary $\delta \Omega$ and $\alpha$ is the normalization factor taken to be 255 (for grey-level images). The priority $P(p)$ associated with every pixel $p$ in $\delta \Omega$ is given by

$$P(p) = C(p)D(p)$$

Once these priorities are calculated, a patch $\Psi_{\hat{p}}$ around the pixel $\hat{p}$ of maximum priority is considered for filling up. Since the pixels on the boundary of the target region get more priority, the selected patch $\Psi_{\hat{p}}$ will always consist of both the target and non-target pixels. Information from a patch (exemplar) similar to $\Psi_{\hat{p}}$ can be used to fill the target pixels in it. An exemplar can be found by comparing a patch $\Psi_q$ around every pixel $q$ in the entire image with $\Psi_{\hat{p}}$. Therefore, as the image size increases, the time taken to find an exemplar also increases. However, we observe that a patch is similar to patches that are in its surrounding region. Therefore, a considerably large sized window $W_{\hat{p}}$ around the patch to be filled up can be used to search the best matching patch or exemplar. By doing so, the number of computations required to search the exemplar are considerably reduced. We measure the similarity of patches $\Psi_{\hat{p}}$ and $\Psi_q$ by comparing the pixels $\hat{p}_i$ in patch $\Psi_{\hat{p}}$ that are outside the target region (i.e. $\hat{p}_i \notin \Omega$) with corresponding pixels in every patch $\Psi_q$ (where $\Psi_q \in W_{\hat{p}}$ and $\Psi_q \cap \Omega = \phi$). The patch $\Psi_{\hat{q}}$ which gives the minimum sum of squared differences (SSD) is considered as the exemplar $E_{\hat{p}}$.

Once the exemplar is available, we intend to transfer the information from the exemplar into the corresponding target pixels in patch $\Psi_{\hat{p}}$. One way of transferring information from the exemplar into the target pixels of the patch $\Psi_{\hat{p}}$ is by simply copying the corresponding pixels from the exemplar $E_{\hat{p}}$ as done in [10]. The method works well if the pixel intensities in the window $W_{\hat{p}}$ do not vary much. In such cases, the SSD error obtained for the best matching exemplar will be small and therefore the patch $\Psi_{\hat{p}}$ and $E_{\hat{p}}$ will be very similar. Thus,
copying the pixel values from $E_p$ to $\Psi_p$ is justified. However, sometimes due to variation in illumination or contrast within $W_p$, the SSD obtained for $E_p$ will be very high, i.e. even the most similar patch will also be considerably different from $\Psi_p$. In such cases, if pixel values are copied from the $E_p$ into $\Psi_p$, the modifications in $\Psi_p$ do not appear to be uniform, making the seam clearly visible in the inpainted patch. Therefore, one can think of smoothly blending the pixel values from $E_p$ to $\Psi_p$ for the target pixels as another way to transfer information.

Pérez et al. [9] have demonstrated that blending of images can be achieved seamlessly by solving for unknown pixel values using discrete Poisson equations i.e. the values of every pixel $p$ in the target region $\Omega$ satisfies the following equation

$$|N_p|f_p - \sum_{q\in N_p\cap \Omega} f_q = \sum_{q\in N_p\cap \delta \Omega} f_q^* + \sum_{q\in N_p} v_{pq}$$  \hspace{1cm} (5)

where $f_p$ is a value of pixel $p \in \Omega$, $f_q^*$ is a value of pixel $q \in I\setminus\Omega$, $f_q$ is a value of pixel $q \in \Omega$, $q \in N_p$ where $N_p$ is the set of neighbours of pixel $p$ and $|N_p|$ is the number of pixels in the set $N_p$. $v_{pq}$ is difference between values of corresponding pixels $p$ and $q$ in the source region (used as guidance vector field). One may note that both the source and target regions are selected manually as the technique is primarily for image editing. Also, the blending is done considering all the target pixels at once and not using patches. In our technique, the source region can be considered as the automatically obtained exemplar for a corresponding target patch. If we re-arrange equation (5), we get

$$f_p = \frac{\sum_{q\in N_p\cap \Omega} f_q + \sum_{q\in N_p\cap \delta \Omega} f_q^* + \sum_{q\in N_p} v_{pq}}{|N_p|}$$  \hspace{1cm} (6)

From figure 2 and taking $|N_p| = 4$, the equation (6) can be be written as

$$f_p = \frac{(f_{q_1} + f_{q_2}) + (f_{q_3} + f_{q_4}) + 2(f_{q_5} - (f_{q_3} + f_{q_4} + f_{q_5} + f_{q_6}))}{4}$$  \hspace{1cm} (7)

where $\sum f_q = f_{q_1} + f_{q_2}$, $\sum f_q^* = f_{q_3} + f_{q_4}$ and $v_{pq} = g_p - g_q$. The above equations (6) and (7) show that the value $f_p$ is determined by taking the average of its neighbours and by computing the average in the corresponding source patch selected manually. However, for many images, it is not true that a pixel value is an average of its neighbours. One may say that the neighbouring pixel values contribute in some proportions towards the value of the center pixel. In other words, a pixel value can be considered to be a linear combination of the first order neighbouring pixel values, and for this reason we modify the equation (7) to be useful for plausibly blending patches from corresponding exemplars.

For a small region, the coefficients associated with each neighbouring pixel in the linear combination remain constant. Also, these coefficients remain constant for similar regions. Since $E_p$ is similar to patch $\Psi_p$ we can write,

$$f_p = k_t f_{q_t} + k_r f_{q_r} + k_b f_{q_b} + k_l f_{q_l},$$

$$g_p = k_t g_{q_t} + k_r g_{q_r} + k_b g_{q_b} + k_l g_{q_l},$$  \hspace{1cm} (8)

where $k_t$, $k_r$, $k_b$ and $k_l$ are the coefficients corresponding to top, right, bottom and left neighbouring pixels respectively. If in some way, we are able to obtain these coefficients, then they can be used in the following equation to calculate the values of all the target pixels in the considered patch.

$$f_p = \frac{(k_t f_{q_t} + k_r f_{q_r} + k_b f_{q_b} + k_l f_{q_l}) + g_p - (k_t g_{q_t} + k_r g_{q_r} + k_b g_{q_b} + k_l g_{q_l})}{k_t + k_r + k_b + k_l}$$  \hspace{1cm} (9)

It is worth to note that when $k_t = k_r = k_b = k_l = \frac{1}{4}$, equation (9) reduces to equation (7). Simultaneously solving equations for every $f_p$ will determine the value of each pixel $p \in \Omega$.

In order to determine $k_t$, $k_r$, $k_b$ and $k_l$ we arrange all the patches $\Psi_q \in W_p$ in ascending order of the SSD error and consider only the first $L$ patches to form a set $S_p$. As we know that the coefficients associated with the neighbouring pixels remain constant for a small region or similar regions, for all the $3 \times 3$ regions in $S_p$ we can write

$$g_p = k_t g_{q_t} + k_r g_{q_r} + k_b g_{q_b} + k_l g_{q_l} + n_p$$  \hspace{1cm} (10)

Here $g_p$ is the center pixel value of every $3 \times 3$ region in $S_p$. The equation (10) represents the equation for AR model [15] for the pixel $g_p$. Here $n_p$ represents the independent identically distributed (i.i.d.) Gaussian noise with mean 0 and variance $\sigma_p^2$. $k_t$, $k_r$, $k_b$ and $k_l$ are the parameters or coefficients of the AR model. As the values of $g_p$ and corresponding first order neighbours are known, we find the best fitting values of $k_t$, $k_r$, $k_b$ and $k_l$.

Using least squares (LS) method to determine the AR parameters, it is possible that some of the values obtained are negative. But since we say that the coefficients associated with neighbouring pixels denote the proportion of the respective neighbour’s contribution, a negative value is unacceptable. We therefore use non-negatively constrained least squares (NNLS) method [13, 14] to obtain the values of $k_t$, $k_r$, $k_b$ and $k_l$, which assures that the obtained values are non-negative. The NNLS method iteratively categorizes the constraints into Active and Passive sets. The constraints corresponding to a negative or zero regression coefficients are included in the active set and the remaining constraints constitute the passive set. The
solution is then the unconstrained least squares solution using only the variables corresponding to the passive set, setting the regression coefficients corresponding to the active set to zero. With these calculated values of the AR parameters and using $E_p$, equation (9) can be solved to obtain the values of target pixels in the patch $Ψ_p$. Once a patch is processed, its pixels are excluded from the target $Ω$. The modified target is then used for the next iteration of filling up the most prioritized patch. Thus, after each iteration, the target region shrinks. The algorithm is terminated when the target is empty.

The number of best matching exemplars $L$ to be considered for AR should be at least equal to the number of AR parameters to be determined. As $L$ increases, we obtain better values of AR parameters. However, if $L$ is very large, then many exemplar candidates with larger SSD values get involved in calculating the AR parameters. As a result, the estimated values do not represent the true spatial relationship of the pixels in the patch $Ψ_p$. Therefore, one has to heuristically choose the number $L$.

While considering the size of patch $Ψ_p$, one has to take care that the size is not very large. If a patch with large size is considered, the size of corresponding candidate exemplars (that are used to calculate the AR parameters) will also be of large. Over a large region, the spatial relationship of the pixels may change. Thus, if a large region is considered then we may not be accurately representing the spatial relationship by the AR parameters, an in turn reducing the effectiveness of the algorithm.

III. EXPERIMENTAL RESULTS

We present the results of our technique on data collected from the world heritage site, Hampi, Karnataka. The images were captured using a Samsung ES55 digital camera. The data consists of a number images of monuments, having both vandalized and non-vandalized regions. Three such images shown in figures 3(a), 4(a) and 5(a), respectively, were taken up for experiment. In all the images, fairly large cracks are visible and we intend to restore the images as if they had no cracks at all.

We compare the results of our proposed algorithm with that of the algorithm presented in [10]. The algorithms are implemented in Matlab. Once an image is given as an input, the user selects the target or the cracked region that has to be filled up. These regions for the experimental images are shown in black color in figures 3(b), 4(b) and 5(b), respectively. The algorithms then iteratively process distinct patches in the target region to obtain the final results. The results using algorithm in [10] are shown in figures 3(c), 4(c) and 5(c), respectively, and those of the proposed method are displayed in figures 3(d), 5(d) and 5(d), respectively. The patch $Ψ_p$ is taken to be of size $9 \times 9$, the window $W_p$ is taken to be of size $37 \times 57$. In our experiment we consider the number of candidate exemplars (used for estimating the AR parameters) $L$ as 30.

From the figures 3, 4 and 5 one can clearly see seams at the inpainted regions of the images in figures 3(c), 4(c), 5(c), respectively, obtained using the technique proposed in [10]. At the same time, the results obtained using our technique shown in figures 3(d), 4(d) and 5(d), respectively, appear seamless and plausible.

Since no reference images were available, we present the quantitative performance of the two algorithms using the popularly used metrics in the area of image fusion and image restoration [16] [17] [18], when no reference images are available. The first metric, entropy, quantifies the information contained in an image. More the entropy, better is the resultant image. Entropy $H$ is calculated as

$$H = - \sum_{i=0}^{N-1} p(i) \log_2(p(i))$$  \hspace{1cm} (11)

where $p(i)$ is the probability or the normalized histogram of $i^{th}$ gray level with the image having $N$ gray levels. The metric standard deviation $SD$ [17] tells about the contrast in an image. An image with high contrast will have higher value of $SD$. With the above conventions, $SD$ is calculated as

$$SD = \sqrt{\sum_{i=0}^{N-1} (i - i')^2p(i), \quad i' = \sum_{i=0}^{N-1} ip(i)}$$  \hspace{1cm} (12)

The local orientation dominance $R_i$ value indicates the presence of an edge in a local region $i$ [18]. We have considered the local region $i$ to be of size $5 \times 5$. $R_i$ is calculated using singular values $s_1$ and $s_2$ of the local gradient vectors as,

$$R_i = \frac{s_1 - s_2}{s_1 + s_2}, \quad s_1 \geq s_2, \quad 0 \leq R_i \leq 1$$  \hspace{1cm} (13)

It may be noted that for a noisy image, the value $R = \text{avg}(R_i)$ is low.

The results of the two algorithms based on these measures are summarized in table I. From the observations once can conclude that our method performs better in terms of perceptual quality as well as quantitative measures (except for image3 in terms of $R_i$) in comparison with the algorithm stated in [10].

### TABLE I

<table>
<thead>
<tr>
<th>Images</th>
<th>Entropy</th>
<th>SD</th>
<th>$R_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_{new}$</td>
<td>$T_{old}$</td>
<td>$T_{new}$</td>
</tr>
<tr>
<td>image1</td>
<td>7.5935</td>
<td>7.5455</td>
<td>47.9363</td>
</tr>
<tr>
<td>image2</td>
<td>6.8473</td>
<td>6.8435</td>
<td>30.4150</td>
</tr>
<tr>
<td>image3</td>
<td>7.0426</td>
<td>7.0424</td>
<td>32.1703</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

In this paper we have presented an automatic exemplar search based inpainting technique that uses spatial dependence of a pixel with its neighbours as a cue to blend information from the exemplar into the target pixels of the patch under consideration. Assuming the neighbourhood to be of the first
order, the spatial dependence is represented using an AR model, the parameters of which are estimated from a set of candidate exemplars using the NNLS method. The proposed technique enables seamless blending that was otherwise not possible by direct copying of pixels from the exemplar to the target pixels of the corresponding patch under consideration. The reported results are promising. We conclude that by estimating the spatial dependence of a pixel with its neigh-
bourhood using the AR model, vandalized regions in images of monuments can be inpainted plausibly.

ACKNOWLEDGMENT

This work is a part of the Indian Digital Heritage (IDH) project. The authors would like to thank Department of Science and Technology (DST), Govt. of India for funding this project.

REFERENCES


