

# A Non-local MRF model for Heritage Architectural Image Completion

Deepan Gupta\* Vaidehi Chhajjer\* Anand Mishra† C. V. Jawahar‡

Center for Visual Information Technology, IIT Hyderabad, India

<http://cvit.iiit.ac.in/>

## ABSTRACT

MRF models have shown state-of-the-art performance for many computer vision tasks. In this work, we propose a non-local MRF model for image completion problem. The goal of image completion is to fill user specified “target” region with patches of “source” regions in a way that is visually plausible to an observer. We represent the patches in the target region of the image as random variables in an MRF, and introduce a novel energy function on these variables. Each variable takes a label from a label set which is a collection of patches of the source region. The quality of the image completion is determined by the value of the energy function. The non-locality in the MRF is achieved through long range pairwise potentials. These long range pairwise potentials are defined to capture the inherent repeating patterns present in heritage architectural images. We minimize this energy function using Belief Propagation to obtain globally optimal image completion.

We have tested our method on a wide variety of images and shown superior performance over previously published results for this task.

## Keywords

Inpainting, MRF, Belief Propagation

## 1. INTRODUCTION

Image completion is an important and challenging computer vision task. The goal of an image completion algorithm is to reconstruct the missing regions within an image in a way that is visually plausible to an observer. In most cases, the missing region (called the target region) is filled in by using the information from the rest of the image (called

\*Equal contribution

\* {deepan.gupta,vaidehi.chhajjer}@students.iiit.ac.in

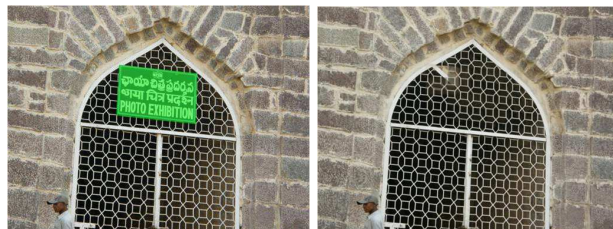
† anand.mishra@research.iiit.ac.in

‡ jawahar@iiit.ac.in

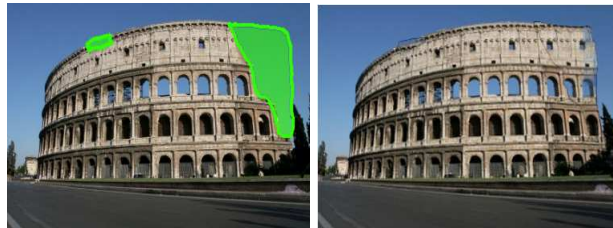
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICVGIP '12, December 16-19, 2012, Mumbai, India

Copyright 2012 ACM 978-1-4503-1660-6/12/12 ...\$15.00.



(a) Object Removal



(b) Reconstruction

**Figure 1: Many successful applications of our method. (a) Object removal: the signboard on the window is successfully removed. (b) Reconstruction: the originally broken Colosseum has been successfully reconstructed using our approach.**

the source region). Image completion is an important part of many computer vision applications such as scratch removal, object removal and reconstruction of damaged architectural parts in an image. Moreover, image completion has applications in the field of photo editing and restoration. Image completion is a highly researched area in computer vision [4, 5, 7, 8, 10, 11]. Although due to the complexity of the images, results leave a lot to be desired. In this work, we propose a non-local MRF technique to complete images. (Note that non-local MRF [15] has been successfully applied to image restoration in past). The beauty of our formulation is in its capability to use non-local repeating patterns in MRF energy minimization framework. We use Belief Propagation [12] to find the minimum of this energy *i.e.*, the optimal image completion. Few successful applications of our method are shown in Figure 1. It can be seen that our method performs well for the tasks like object removal, reconstruction etc.

India has one of the greatest architectural monuments in the world. With advances in computer vision techniques, it is now possible to capture the glory of heritage architec-

tural images for both showcasing and preservation for future generations. Our work can be considered as a small step towards this. We primarily focus on completion of images taken from Indian heritage architectures. (Although we also test our method on many other images to show the generality of the method).

**Contributions.** The contribution of this work is two fold, (1) We propose a non-local optimization framework to capture repeating patterns inherently present in the image (especially, the images of our interest *i.e.* heritage architectural images). We model these repetitions via long range potentials in MRF. (2) We prove that the proposed energy function is sub-modular as well as semi-metric. This proof guarantees that the energy function can be efficiently minimized via move making algorithm like  $\alpha$ - $\beta$  swap [2]. However, the study of energy minimization techniques is beyond the scope of this work. Hence, we restrict ourself to Belief Propagation for the implementation of our method.

**Outline of the paper.** The remainder of the paper is organized as follows. We discuss related work in Section 2. In Section 3, the image completion problem is formulated as a labeling problem. In this section, we also give the reasoning behind long range potentials and define a novel non-local MRF energy function such that its minimum corresponds to the globally optimal image completion. In Section 4, we prove that the proposed non-local MRF energy function is sub-modular and semi-metric. We then discuss experimental settings and present results of proposed method in Section 5. Finally, Section 6 concludes our work.

## 2. RELATED WORK

There has been significant research in the field of image completion. Various approaches have been put forward over the last few decades. The approaches can be grouped into four major categories: (1) Statistical Methods, (2) Partial Differentiation Equation (PDE)-Based Methods, (3) Exemplar Based Methods, (4) Global Optimization based Methods.

### *Statistical Methods.*

These methods make use of parametric statistical models for image completion. These models (wavelet coefficients [13], colour histogram [6]) are used for representation of image characteristics. The idea is to estimate the missing region and fill it using an iterative process. Initially, the output image is generated keeping the missing regions as pure noise. These regions undergo iterative noise reduction to produce the final output. Statistical methods are only useful in case of texture synthesis. Moreover, they produce blurred outputs for natural images.

### *PDE-Based Methods.*

Partial Differential Equation (PDE) based methods use diffusion process for image completion. The idea is to start the region-filling from the boundary of the missing region and then propagate towards the interior. The boundary filling uses diffusion process simulated by solving the partial differentiation equations. In [1] region-filling is done by propagating image Laplacians in the direction of the isophotes. Chan *et al.* [3] use variational model for region filling.

PDE-based approaches perform well in cases where the

missing region is smooth and non-textured. However, they fail in case of large inpainting regions.

### *Exemplar-Based Methods.*

Exemplar based techniques have been the most successful approaches in presence of large unknown regions. The visible patches of the image are used as a training set to infer the unknown parts which are then filled by simply copying the content of these known patches. Exemplar based techniques have been widely used for image completion recently. Criminisi *et al.* [4] propose a priority-based mechanism which combines texture synthesis and isophote driven inpainting for image completion. This approach, though isophote driven, is not capable of maintaining the structural consistency of the image. Hung *et al.* [8] propose Bezier curves to determine missing edge information, hence preserving structure consistency. The damaged regions are then inpainted using exemplar based methods. However, there are three major pitfalls of these methods. Firstly, the confidence map is computed based on heuristics and may not be applicable to a general case. Secondly, once a patch has been assigned to an unknown region, it cannot be changed. Finally, the greedy approach leads to a bias caused due to selection of a few incorrect patches in the priority based mechanism. These incorrect completions have a spiraling effect which destabilizes the inpainting process.

### *Global Optimization based methods.*

There has been huge interest in the discrete optimization community for image completion problem in recent years [5, 14]. However, these methods do not take advantage of repeating patterns inherently present in many architectural images. On other hand, we model the repetitions in the energy function itself and thus define a better energy function for the problem. Closest to our work is [10]. Here authors try to tackle the drawbacks of the Exemplar-based approach by posing image-filling as a discrete global optimization problem. It uses the exemplar-based framework and Markov Random Field (MRF) for image completion. The idea is to minimize the energy of the MRF using Priority-Belief Propagation (Priority-BP) optimization scheme. The approach works well for majority of the cases. However, in the images where repetitions are prominent, it produces relatively poor output. The method ignores the fact that repetition, if present, may carry critical information about the missing region. Contrary to this, we incorporate the repetition present in the image by including long-range potentials and find the global minima of the non-local MRF energy.

## 3. THE IMAGE COMPLETION PROBLEM

Given a source region  $\mathcal{S}$  and a target region  $\mathcal{T}$  the image completion problem is to fill the target region such that it agrees with its surroundings. We define the image completion problem in a labeling problem framework where overlapping spatial positions in image can be considered as a set of sites and patches of size  $w \times h$  sampled from source region can be considered as labels. In other words, site is a set  $S = \{X_0, X_1 \dots X_m\}$  where each  $X_i$  is a spatial position of size  $w \times h$  in the image. Similarly, label  $L = \{L_0, L_1, \dots, L_n\}$  is a set where each  $L_i$  is a patch of size  $w \times h$  sampled from source region  $\mathcal{S}$ . Figure 2 shows the formulation of image completion as a labeling problem.

Each site  $X_i$  can take a random value  $x_i = \{L_0, \dots, L_m\}$ . The labeling problem here is to find the optimal function  $f^* : S \rightarrow L$ . Optimality criteria is defined based on quality of the image completion. In general, this is an NP-hard problem. However it can be solved approximately by finding the minimum of the Gibbs energy (also known as MRF energy) of following form:

$$E(x) = \sum_{i=1}^m E_i(x_i) + \sum_{\mathcal{N}} E_{ij}(x_i, x_j). \quad (1)$$

$E_i(\cdot)$ ,  $E_{ij}(\cdot)$  and  $\mathcal{N}$  corresponds to data term, smoothness term and neighborhood system defined in MRF respectively. The data term measures the agreement with the available observations whereas the smoothness term is used to enforce spatial coherence. The minimum of this energy function corresponds to the optimal image completion.

### Data term.

The data term computation for image completion problem is not straightforward because only boundary sites are visible whereas interior sites are hidden. (Recall that in any image completion problem user provides a mask which needs to be filled). Without loss of generality each patch ( $x_i$ s and  $L_i$ s) can be represented as a vector of size  $w \times h$ . *i.e.* patches can be represent as a vector in a vector space as follows.

$$x_i = [x_{i_0}, \dots, x_{i_l}]^T \\ L_i = [L_{i_0}, \dots, L_{i_l}]^T,$$

where  $l = w \times h$  and each  $x_{i_l}$  is either hidden or in  $[0, 255]^3$ . Whereas  $L_{i_l} \in [0, 255]^3, \forall i$ . To distinguish visible and hidden nodes, we introduce a binary vector  $K = \{k_1, k_2, \dots, k_l\}$  such that  $k_m$  takes value 0 if  $x_{i_m}$  is hidden and 1 otherwise. We define the unary cost of random variable  $x_i$  taking label  $L_i$  as follows.

$$E_i(x_i = L_i) = \sum_{m=0}^l k_m (x_{i_m} - L_{i_m})^2. \quad (2)$$

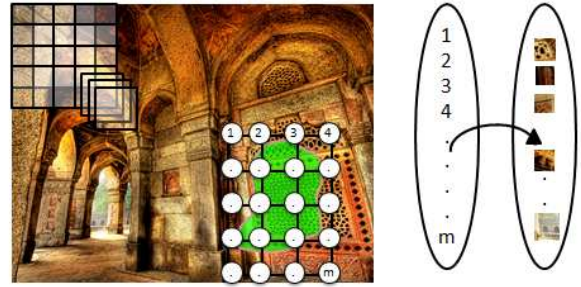
In other words, data term measures the agreement between random variable  $x_i$  and label  $L_i$  in terms of sum of squared distance (*ssd*) of known pixels. Thus, the cost of  $x_i$  taking label  $L_i$  is low if the sum of squared distance (*ssd*) between  $x_i$  and  $L_i$  is low.

### Smoothness term.

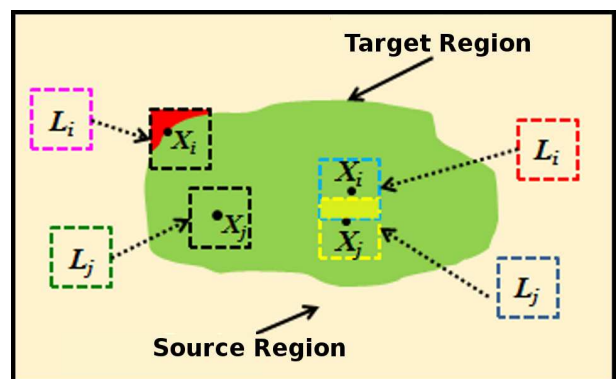
The data term alone cannot give coherent completion. To enforce coherency in the completed image, we define a smoothness term such that overlapping region of neighboring labels have least sum of squared distance. We define the smoothness term as follows.

$$E_{ij}(x_i = L_i, x_j = L_j) \\ = \sum_{m=0}^{size(\psi)} \delta(X_{i_m} \in \psi) \wedge \delta(X_{j_m} \in \psi) (L_{i_m} - L_{j_m})^2. \quad (3)$$

Here  $\psi$  is the overlapping region between sites (*i.e.* patches)  $X_i$  and  $X_j$ .  $\delta(\cdot)$  is an indicator function. The process of data and smoothness term computation is pictorially depicted in Figure 3.



**Figure 2: Image Completion as a labeling problem.** Overlapping patch positions in image and patches sampled from source region represent sites and labels respectively. The labeling problem here is to find the optimal labeling from sites to labels.



**Figure 3: Data term and smoothness term computations.** Data term is agreement from the labels to the node in terms of *ssd*. Only visible area contributes to *ssd*. Smoothness term is computed based on *ssd* in overlapping regions of labels. (note that labels here are basically a collection of patches) (Best viewed in colour).

### Long range potentials.

In addition to the data and smoothness terms, we wish to capture the inherent repetitive patterns present in heritage architectural image. To achieve this, we add an extra term in the MRF energy which we call as long range pairwise potentials. The long range pairwise potential are defined between a patch and its repeating offset at distance  $\tau$ . (We describe the repeating offset computation in the next subsection). This long range potential forces a node to take similar label to a patch at offset  $\tau$ . Mathematically, the long range potential  $E^{lr}(\cdot, \cdot)$  is defined as follows.

$$E^{lr}(x_i = L_i, x_k = L_j) = \sum_{m=1}^l k_m (x_{i_m} - L_{j_m})^2. \quad (4)$$

Note that here  $x_i$  and  $x_k$  are non-local *i.e.* they are at distance  $\tau$ . (Here  $\tau$  is a repeating offset, in other words the image has a repeating pattern at offset  $\tau$ ). This definition of long range potential ensures less penalty if  $x_i$  and  $x_k$  take similar labels.

Thus we modify Equation 1 to non-local MRF energy as

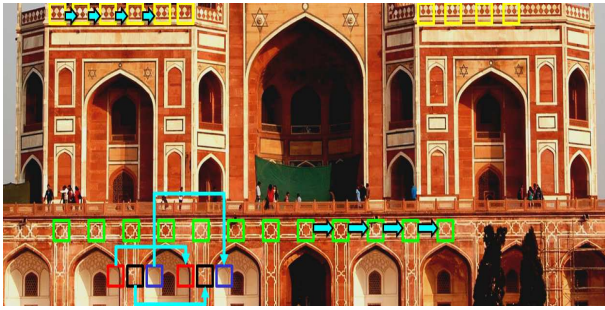


Figure 4: Many architectural images contain repeating patterns. One such example is shown here (Best viewed in colour).

follows.

$$E(x) = \sum E_i(x_i) + \sum_{\mathcal{N}} E_{ij}(x_i, x_j) + \sum_{\text{dist}(x_i, x_k) = \tau} E^{lr}(x_i, x_k). \quad (5)$$

Once the energy is formulated, the problem of image completion becomes equivalent to finding the configuration  $x^*$  corresponding to the global minima of the energy function. The graph construction corresponding to this energy function and the inference (energy minimization) are discussed in Section 3.2.

### 3.1 Repeating Offset Computation

Many archaeological monument images contain repeating patterns. One such example is shown in Figure 4. These repeating patterns vary in complexity which makes image completion a challenging task. The repetitive pattern may carry significant information about the region which is to be completed (inpainted). Hence, the idea is to make use of various repetitions present in the input image and boost the region-filling.

The repetitions can be of any size and along any direction. In order to capture both we make use of  $(p, q)$  offset  $(\tau)$ , repeating offset) pairs which correspond to the  $x$ -direction and  $y$ -direction repetition offsets respectively.

In this step, we compute offsets that can effectively represent the inherent repetition in the image. We use the fact that patches which are part of the repetition will repeat with some common offset. The distance between a patch and its nearest similar patch will account for the repetition offset. Offset Generation consists of following steps.

#### 1. Finding Nearest Similar Patches.

For every patch in the source region of the image we find the nearest most similar patch. For every patch  $P$  belonging to the source region,

$$\tau(x) = \arg \min_{\tau} \|P(x) - P(x + \tau)\|^2; |\tau| > \theta.$$

Here  $\tau$  is a  $2D$ -coordinate offset  $(p, q)$ ,  $P(x)$  is the patch centered at  $x$ .  $P(x + \tau(x))$  is the nearest similar patch.  $\tau$  represents the offset obtained for the pixel  $x$ . The similarity is defined using the sum of squared differences ( $ssd$ ) between the patches. Lesser  $ssd$  corresponds to higher similarity. The parameter  $\theta$  represents the threshold radius for the nearest patch. The patch must lie outside this range.

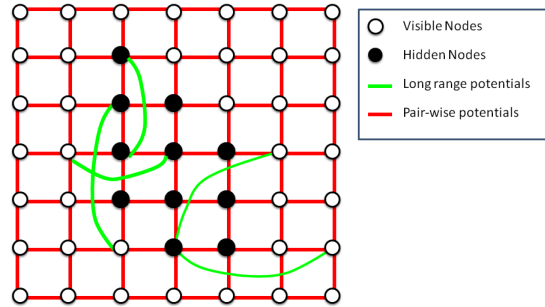


Figure 5: The proposed graphical model. There are two types of nodes in the graph: visible (in boundary and source region) and hidden (in interior region of target). Hidden nodes are shown by filled circles. Local pairwise potentials are shown via red edges and non-local long range potentials are shown via green edges. We use loopy belief propagation for inference in this graphical model (Best viewed in colour).

This is just to ignore the nearby patches which are likely to be similar but do not contribute towards the repetition offset. This threshold varies with the images based on their sizes. In our testing,  $\theta$  was set to be  $1/15^{th}$  of the maximum of image height and image width.

The brute force search for the nearest patch can be computationally expensive as for each patch we need to traverse entire source region. Therefore, to overcome the high computation cost, we use Approximate Nearest Neighbour(ANN). Using approximate nearest neighbour,

$$\tau(x) = ANN(P(x)); |\tau| > \theta.$$

#### 2. Histogram and Offset Generation.

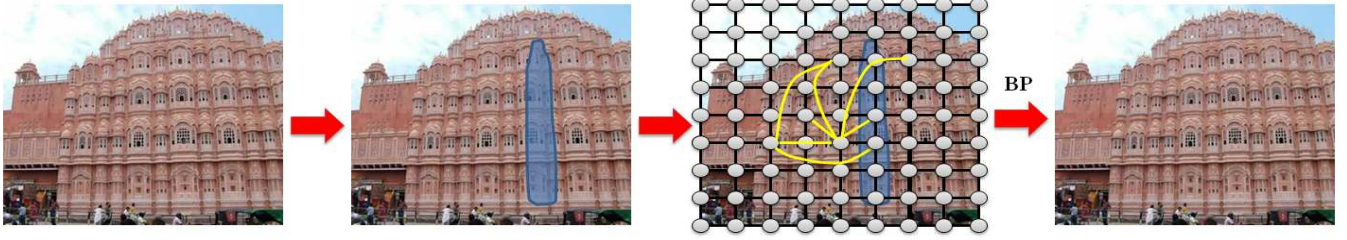
Once we obtain the offsets corresponding to each pixel in the source region, we need to combine the results in order to obtain the correct repetition offsets. To achieve this we represent  $\tau$  as a  $2D$ -plane and generate histogram count of the all  $\tau$  offsets *i.e.*

$$H(\tau) = \sum_{\tau(x)} \delta(\tau(x) = \tau).$$

$H(\tau)$  gives the count of the number of patches having their individual offset as  $\tau$ . Now, to get the prominent repetition offsets, we analyze the histogram counts of the offsets and select offsets with highest count. Since the image may contain many repetitive patterns which are prominent, thus we generate the top  $C$  offsets to capture varied repetitions ( $C = 10$  was used in our experiments).

### 3.2 Graph Construction and inference

We solve the energy minimization problem on a corresponding graph, where each random variable is represented as a node in the graph. Nodes in 4-neighborhood system are connected via edges. To capture repeating patterns in the image, we also join non-local nodes at offset  $\tau$ . (This offset is computed based on the procedure described Section 3.1). We further group nodes in this graph into two categories: visible and hidden. The nodes belonging to source region



**Figure 6: Overview of our method.** First user selects a mask. Based on user selected mask a graph is constructed where each node represent an overlapping spatial position in the image. These nodes are connected via a 4-neighborhood system  $\mathcal{N}$ . Moreover, to capture inherent repeatability, two nodes at distance of repeating offset are also connected (these edges are shown in yellow colour). To find repeating pattern in the image we use approximate nearest search (ANN). After graph construction, graph is labeled using popular inferencing technique: BP. The final output of our method is shown in the right most image (Best viewed in colour).

and boundary of target region are visible, however interior nodes of the target regions are hidden. Each node takes a label from label set  $L = \{L_0, L_1, \dots, L_n\}$  where each  $L_i$  is a patch of size  $w \times h$ . The cost of a node taking some label  $L_i$  is determined by the unary cost defined in Equation 2. Further, the joint cost of two neighboring and non-local nodes taking label  $L_i$  and  $L_j$  defined in Equation 3 and 4 respectively, give the weights to edges. Similar to [10], if a node is highly likely to take some label, we declare that node “committed” and give higher priority to it for sub-sequent inference procedure. The proposed graphical model is shown and explained in Figure 5.

### Inference.

For Inference of the proposed graphical model, we use popular message passing based inferencing algorithm known as loopy Belief propagation (BP). Belief propagation was first proposed in [12]. It iteratively tries to find the Maximum-a-Posteriori (MAP) estimate by propagating messages (beliefs) from nodes to its neighbors. (Recall that in MAP-MRF framework MAP is equivalent to the global minima of the MRF energy). Although theoretically loopy BP does not guarantee convergence for grids, but experimentally it has been shown that it yields a strong local minima for a wide range of computer vision problems [17].

The proposed method is summarized in Figure 6.

## 4. SUB-MODULARITY AND METRICITY

In this section, we prove that the non-local MRF energy function defined in Equation 5 is sub-modular and semi-metric.

**Statement 1.** The energy function defined in Equation 5 is a sub-modular function for every pair of labels.

**PROOF.** A function of single variable is trivially a sub-modular function [9]. Thus, it would suffice if we prove that the pairwise terms  $E_{ij}(\cdot, \cdot)$  and  $E_{ij}^{lr}(\cdot, \cdot)$  are sub-modular for every pair of labels. To prove the sub-modularity, we need to prove the following:

$$\sum_i E(L_i, L_i) \leq \sum_{i \neq j} E(L_i, L_j), \quad \forall i, j.$$

Since  $E(\cdot, \cdot)$  is a sum of squared distance between two vectors, thus  $E(L_i, L_i) = 0, \forall i$ . Moreover, sum of squared distance between any two not-equal vectors is always positive,

which implies,  $E(\cdot, \cdot)$  is a sub-modular function. This proof of sub-modularity can be easily extended to long range potentials without loss of generality. Further, since the sum of sub-modular functions is a sub-modular function [16], the energy function defined in Equation 5 is a sub-modular energy function for every pair of labels.  $\square$

**Statement 2.** The energy function defined in Equation 5 is a semi-metric.

**PROOF.** The energy function defined in Equation 5 is basically composed of sum of squared distance ( $ssd$ ) between two vectors, thus it would be sufficient to prove  $ssd$  as a semi-metric. Here we show that the sum of squared distance ( $ssd$ ) has all the three necessary and sufficient properties to be a semi-metric. We also show that  $ssd$  does not hold triangular inequality always, thus is not a metric.

*1. Non-negativity:*  $ssd$  between two vector is always greater than zero *i.e.*

$$ssd(L_i, L_j) \geq 0, \quad \forall i \neq j.$$

*2. Identity of indiscernibles:*  $ssd$  between two vector is equal to zero iff both the vectors are equal, *i.e.*,

$$ssd(L_i, L_j) = 0 \iff i = j, \quad \forall i, j.$$

*3. Symmetricity:*  $ssd$  between two vector is a symmetric function, *i.e.*

$$ssd(L_i, L_j) = ssd(L_j, L_i), \quad \forall i, j.$$

*4. Triangular in-equality:* In order to attempt to prove this let us first start with the Euclidean distance between two vectors. Let  $dist(L_i, L_j)$  be Euclidean distance between vectors  $L_i$  and  $L_j$ . Then, since Euclidean distance holds triangular in-equality, we can write.

$$dist(L_i, L_j) \leq dist(L_i, L_k) + dist(L_k, L_j), \quad \forall k.$$

Squaring above equation yields,

$$ssd(L_i, L_j) \leq ssd(L_i, L_k) + ssd(L_k, L_j) +$$

$$2 dist(L_i, L_k)dist(L_k, L_j).$$

Recall that  $ssd$  is a square of Euclidean distance. Now since Euclidean distances are non-negative, *i.e.*  $2 dist(L_i, L_k)dist(L_k, L_j) \geq 0$ , in other words we can always find  $L_i, L_j$  and  $L_k$  such that,

$$ssd(L_i, L_j) \geq ssd(L_i, L_k) + ssd(L_k, L_j).$$

Thus,  $ssd$  between two vector does not hold the triangular in-equality. In other words,  $ssd$  is not a metric.

From above arguments, we prove that the energy function defined in equation 5 is a semi-metric.  $\square$

The proof of sub-modularity and semi-metricity of the energy functions also guarantees that popular move making algorithm  $\alpha$ - $\beta$  swap can be efficiently used to find the global minima of this energy with a constant approximation [2]. However, the study of various energy minimization techniques is beyond the scope of this work. Reader is encouraged to see [2] for details of the move making algorithms.

## 5. EXPERIMENTS AND RESULTS

In this section, we present a detailed evaluation of our method on a large collection of images captured from Indian heritage sites. To show the generality of the method, we also include few synthetic images and natural images in our test datasets. Given an image and user provided mask, our problem is to complete the masked region in a way that is visually plausible to observer. We evaluate various components of our approach to justify our choices. We compare our method with the well known exemplar based method [4].

The dataset for our experiments comprises of a large variety of images of Indian Heritage sites including Hampi, Konark, Golkonda Fort etc. We take 70 images of resolutions varying from  $350 \times 250$  to  $640 \times 480$ .

### Approximate Nearest Neighbour.

In the process of repetition offset computation, we use Approximate Nearest Neighbour<sup>1</sup> technique in order to find the most similar patches. Using ANN, we are able to bring down the processing time significantly. For a resolution of  $100 \times 100$ , a brute force method takes around 2 minutes to process the entire image and generate the offsets. With ANN, the time is reduced to 0.1 seconds. In order to make the process faster, we scale the image by a factor of 0.5 and then process for offset generation. The threshold radius ( $\theta$ ) is set to  $1/15^{th}$  of the maximum of the image width and height. We club a group of offsets if their mutual distance is less than or equal to 1.  $C = 10$  most frequent offsets are chosen for our experiments.

### Inference.

In our experiments, the patch size is set dynamically as per the image resolution and aspect ratio with the minimum dimension of  $4 \times 4$ . For all the examples, the belief thresholds for pruning and confidence is set to  $-2ssd_0$  and  $-ssd_0$  respectively, where  $-ssd_0$  represents a predefined mediocre  $ssd$  between the patches. We use independent implementation of Priority-BP [10] for inference.

Figure 7 shows the results of object removal using our method. In the first row, hand and head of the person is successfully removed. In second row, we show an image from Golkonda where the signboard is removed with a minute defect. On the other hand, greedy algorithm like [4] clearly fails in removing these objects. We demonstrate the performance of proposed method in ruined wall reconstruction in Figure 5. The purposed method which models repeating

patterns via long range potentials clearly shows superiority over exemplar based inpainting for this task. Apart from object removal and ruined wall reconstruction we also use our method for an interesting application known as background replacement. The results of background replacement is shown in Figure 5. In this example the background of the Taj Mahal is replaced by a natural scene from a different image. These promising results can be applicable in showcasing heritage images in many visually attractive ways.

We also study the importance of long range potentials. Introduction of long range potentials not only yields us a better energy function to minimize, but one can also see noticeable difference in many visual results. We show a couple of them in Figure 5.

The current algorithm is not designed to capture repeating patterns having a rotation relationship and hence the proposed method fails in these cases, as shown in Figure 5.

Reader is encouraged to refer our supplementary material for additional results and comparisons.

## 6. CONCLUSIONS

In this work we address the problem of image completion. The image completion problem is formulated in a principled framework. We model the repeating patterns inherently present in images using long range potentials and solve the problem in non-local MRF framework. We prove that the proposed MRF energy is sub-modular and semi-metric. Experimental results on a wide collection of images show that we clearly outperform popular technique like exemplar based inpainting [4].

### Acknowledgements.

This work was partly supported by Indian Digital Heritage project of DST. Anand Mishra is supported by Microsoft Corporation and Microsoft Research India under the MSR India PhD fellowship award.

## 7. REFERENCES

- [1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *SIGGRAPH*, 2000.
- [2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *TPAMI*, 2001.
- [3] T. F. Chan, J. Shen, and H.-M. Zhou. Total variation wavelet inpainting. *Journal of Mathematical Imaging and Vision*, 2006.
- [4] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *TIP*, 2004.
- [5] K. He and J. Sun. Statistics of patch offsets for image completion. In *ECCV*, 2012.
- [6] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH*, 1995.
- [7] H.-F. Hsin, J.-J. Leou, C.-S. Lin, and H.-Y. Chen. Image inpainting using structure-guided priority belief propagation and label transformations. In *ICPR*, 2010.
- [8] J. C. Hung, C.-H. Huang, Y.-C. Liao, N. C. Tang, and T.-J. Chen. Exemplar-based image inpainting base on structure construction. *JSW*, 2008.
- [9] P. Kohli, M. P. Kumar, and P. H. S. Torr. P3 & beyond: Solving energies with higher order cliques. In *CVPR*, 2007.

<sup>1</sup><http://www.cs.umd.edu/~mount/ANN/>



Figure 7: Successful object removal. (a) The original image with selected mask. (b) Output of exemplar based inpainting mask. (c) Output of proposed approach. We see that the exemplar based approach [4] being local is not able to synthesize the missing region effectively whereas ours generates results with better coherency by performing a global optimization.



Figure 8: Successful reconstruction of ruined walls of Golkonda fort. (a) The original image with selected mask. (b) Output of exemplar based [4] inpainting mask. (c) Output of proposed approach.

- [10] N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *TIP*, 2007.
- [11] X. Li. Image recovery via hybrid sparse representations: A deterministic annealing approach. *J. Sel. Topics Signal Processing*, 2011.
- [12] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [13] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV*, 2000.
- [14] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift-map image editing. In *ICCV*, 2009.

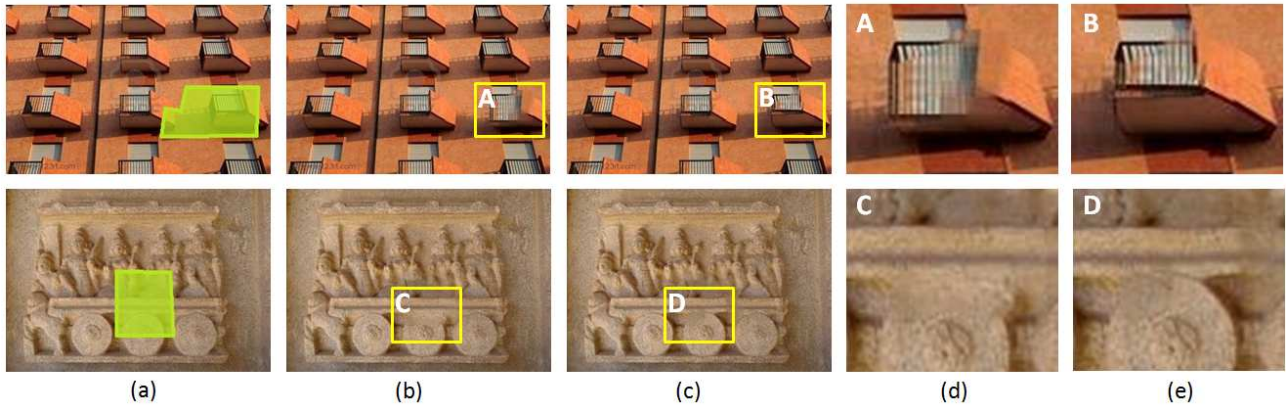


Figure 9: Importance of long range potentials. (a) The original image with selected mask. (b) Output image without long range potentials. (c) Output of proposed approach(with long range potentials). (d-e) The regions of interest (shown in yellow rectangles) are zoomed (Best viewed in colour).



Figure 10: Background replacement. (a) An Indian lady in front of Taj Mahal. The background is marked by the user. (b) A natural scene image taken as source region  $S$  for image completion. (c) The background of the Taj Mahal is replaced by natural scene by the proposed method.



Figure 11: Failure case. (a) The original image with selected mask. (b) Output of exemplar based approach (c) Output of proposed approach. The current algorithm does not take care of patterns having a rotation relationship and hence fails for the above case.

[15] J. Sun and M. F. Tappen. Learning non-local range markov random field for image restoration. In *CVPR*, 2011.  
 [16] Z. Svitkina and L. Fleischer. Submodular approximation: Sampling-based algorithms and lower bounds. In *FOCS*, 2008.  
 [17] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler,

V. Kolmogorov, A. Agarwala, M. F. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *TPAMI*, 2008.