Identifying Vandalized Regions in Facial Images of Statues for Inpainting*

Milind G. Padalkar[1], Manali V. Vora[1], Manjunath V. Joshi[1], Mukesh A. Zaveri[2], and Mehul S. Raval[1]

[1] Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India-382007
{milind_padalkar,manali_vora,mv_joshi,mehul_raval}@daiict.ac.in
[2] Sardar Vallabhbhai National Institute of Technology, Surat, India–395007
mazaveri@coed.svnit.ac.in

Abstract. Historical monuments are considered as one of the key aspects for modern communities. Unfortunately, due to a variety of factors the monuments get damaged. One may think of digitally undoing the damage to the monuments by inpainting, a process to fill-in missing regions in an image. A majority of inpainting techniques reported in the literature require manual selection of the regions to be inpainted. In this paper, we propose a novel method that automates the process of identifying the damage to visually dominant regions viz. eyes, nose and lips in face image of statues, for the purpose of inpainting. First, a bilateral symmetry based method is used to identify the eyes, nose and lips. Textons features are then extracted from each of these regions in a multi-resolution framework to characterize both the regular and irregular textures. These textons are matched with those extracted from a training set of true vandalized and non-vandalized regions, in order to classify the region under consideration. If the region is found to be vandalized, the best matching non-vandalized region from the training set is used to inpaint the identified region using the Poisson image editing method. Experiments conducted on face images of statues downloaded from the Internet, give promising results.

Keywords: texton, inpainting, vandalism, damage-detection, heritage.

1 Introduction

Heritage sites are essential sources of precious information. Cultural heritage conservation helps a community to preserve its history and gives a sense of continuity and identity. However, because of a variety of factors such as weather, antisocial elements, etc., monuments are sometimes ruined or vandalized. Renovating such sites is a very sensitive activity and requires great expertise. Renovation not only poses danger to the undamaged monuments but may also introduce changes in

* This work is a part of project sponsored by Department of Science and Technology (DST), Govt. of India (Grant No: NRDMS/11/1586/2009/Phase-II).

A. Petrosino, L. Maddalena, P. Pala (Eds.): ICIAP 2013 Workshops, LNCS 8158, pp. 208–217, 2013.

the damaged ones that notably deviate from the historical existence. An obvious solution that avoids physical contact is to digitally renovate these monuments, which may be accomplished using image restoration and inpainting techniques [2,3,10]. Image restoration refers to the recovery of an original image from its degraded version, while image inpainting refers to the process of modifying image contents imperceptibly, such as adding or removing object.

Image inpainting has been an active area of research for over a decade. A number of techniques have been proposed during this period. Notable among these works are those based on level lines [2,10], exemplar [3], solving Poisson's equation [12], global image statistics [9], multiple views of a scene [18], multiple images that are semantically similar [6], texture and structure information [1] and many more. All the above listed inpainting techniques are, however, not fully automatic i.e. they require the target regions for inpainting to be either known or selected manually. Work proposed in this paper addresses this issue by automating the region detection process for inpainting face images of damaged statues. Globally, it is observed that vandalism by humans is a major reason for statues being damaged. The facial regions like eyes, nose and lips are the visually the most dominant ones when one looks at a statue and are therefore more likely to get damaged when vandalized. A damage to such regions diminishes the attractiveness of the statues. Although other facial regions may be vandalized, the damage may not be as noticeable as that to the eyes, nose or lips, when it comes to visual attractiveness. We therefore intend to detect and inpaint any damaged eye, nose or lips in a given facial image of a statue.

Recently, Parmar et al. [11] proposed a method on similar lines. Their method relies on the use of edge based features and template matching to detect the damaged regions in frontal-face images of statues. However, the method is highly dependent on the scaling, rotation and pose in the facial images. Moreover, images of different statues may have different sizes and shapes of eyes and lip regions, while the templates used are of fixed size and shape, leading to unreliable results. Further, their method does not address the classification of nose regions. Our proposed approach also takes care of these limitations.

The motivation behind our proposed work is twofold. First, it can be a useful tool to continuously monitor the heritage site using a surveillance system and alert the authorities if any damage to monuments takes place. This damage may be unintentional due to the curiosity of visitors to know details or it may be intentional. After the alert is sounded, a corrective action can be initiated and monuments can be protected from further damage. Secondly, since our approach automatically detects the target regions for inpainting, the subjectiveness associated with the manual selection of target regions for inpainting will be eliminated. Results of an inpainting algorithm are highly influenced by selection of target regions. If the target regions are manually selected, a fair comparison of various inpainting algorithms might not be possible. Thus, a common automatically detected target region for inpainting may prove helpful here. Our work may be used to add up a new feature in digital cameras to create immersive navigation effect. The work can also be useful to replace the manual selection of regions by automatic detection in a photo editing software, for correcting / replacing a damaged part of a given photograph.

Texture analysis of several facial images of statues downloaded from the Internet [5] revealed that the texture of a damaged region is different from that of an undamaged region. This study motivated us to use texture as a feature for region classification. Textons, which are cluster centres in a filter response space, have been extensively used for texture classification [14,15]. In the proposed method, we try to identify a region being vandalized or non-vandalized by comparing its textons with those of a training set consisting of true vandalized and non-vandalized regions. Moreover, our method extracts these textons in a multi-resolution framework to characterize both the regular and irregular textures. The organization of this paper is as follows. Section 2 presents the proposed approach. In section 3, experimental results are shown and finally, we conclude in section 4.

2 Proposed Approach

Our approach for automatic inpainting the face images of monument is divided into following major steps viz. (a) extraction of potential regions of interest, (b) identification of vandalized regions and (c) inpainting vandalized regions.

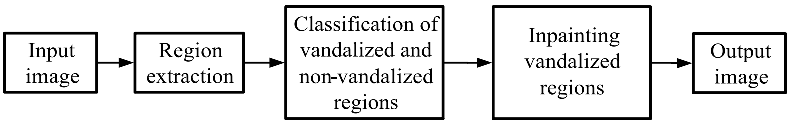


Fig. 1. Block diagram of the proposed approach

Figure 1 illustrates our proposed approach. Initially, major regions of interest from given face image are extracted using bilateral symmetry of gradient directions. Vandalized and non-vandalized regions are identified using the texture based method, out of which the vandalized regions are inpainted using Poisson image editing technique [12]. Here, we have considered eyes, nose and lips as the potential region of interest because of the tendency of human beings to target these regions while vandalizing. The inputs are assumed to be frontal face images. It may be noted that our algorithm works well even when there is slight deviation from the frontal pose. For large deviation due to complex distortions, one may think of using image registration as a preprocessing step. However, given a single image with complex distortions, registration itself is a difficult problem and involves pixel interpolation, affecting texture classification.

2.1 Extraction of Potential Regions of Interest

Given a face image of a statue, we intend to identify visually attractive regions like eyes, nose and lips. Such regions have a common property viz. symmetry.

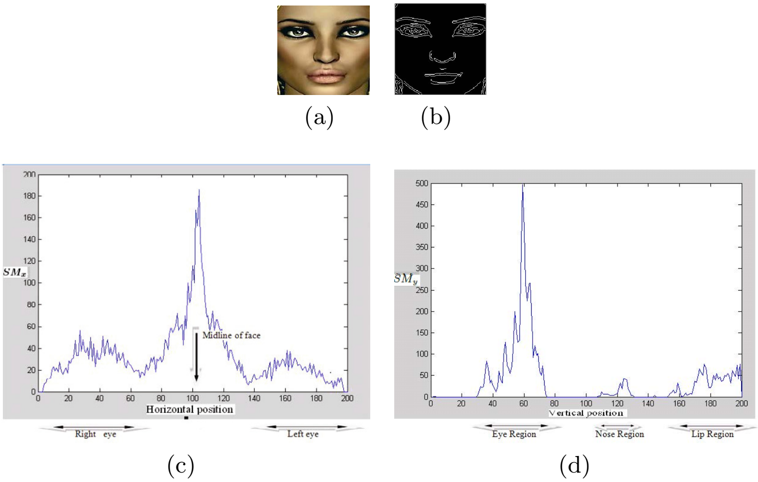


Fig. 2. To extract potential regions of interest; (a) input image, (b) image containing edges, (c) plot of symmetry measure SM_x, (d) plot of symmetry measure SM_y

A method to identify eyes, nose and lips that uses symmetry as a cue has been proposed by Katahara and Aoki [7]. In our approach, we use symmetry based method with noise removal as a preprocessing. This leads to better edge preservation and hence the potential regions are extracted with better accuracy. The use of single scale retinex (SSR) algorithm [16] makes our method invariant to illumination changes. Following are main steps of face region extraction.

(a) Make the input image illumination invariant using SSR algorithm [16].
(b) Apply Perona-Malik edge preserving diffusion [17] to smooth the image.
(c) Extract various outlines present in the smoothed image.
(d) Compute the symmetry measure of each pixel $b_m(x, y)$ [7].
(e) Project the symmetry measure onto X-axis, i.e. SM_x

$$SM_x = \sum_{j=1}^{M} b_m(j, x), \tag{1}$$

where, M is number of columns in an image and b_m represents symmetry measure around horizontal axis. Once SM_x is computed, its peak value is used to find mid-line about which the face is nearly symmetric.

(f) Project the symmetry measure onto Y-axis, i.e. SM_y

$$SM_y = \sum_{i=1}^{N} b_m(y, i), \tag{2}$$

where, N is number of rows in an image. The various peaks in SM_y are used to find eyes, nose and lips.

Figure 2. shows an input edge image and the plots of symmetry measure on the X and Y axis, respectively.

2.2 Identification of Vandalized Regions

Our approach for classifying a region as vandalized or non-vandalized is based on the work of Varma and Zisserman [15], where the materials to be classified were of uniform texture and thus enabled effective modeling of different texture classes. The proposed method aims to detect damage having natural texture. We therefore avoid building the model of different texture classes, which in turn reduces the complexity of the proposed method. Moreover, our method automatically calculates the number of clusters representing the texture classes as opposed to the approach in [15], which uses fixed number of clusters. In addition, our method explores multi-resolution framework to address the issue of irregularities in natural texture at different resolutions; a property characterized by stone-work and monument surfaces.

Textons are basic entities which represents a particular class of texture. They are cluster centres in the filter response space. The textons are extracted by first convolving each extracted potential region interest with the maximum-response-8 (MR8) filter bank [14]. The MR8 filter bank consists of 38 filters which include Gaussian, Laplacian of Gaussian, edge and bar filters, out of which 8 responses are considered [14]. Each pixel of the input region is now transformed into a vector of size 8 as there are 8 filter responses. K-means algorithm is then applied on these vectors where the K cluster centres are known as textons. In the proposed method, selection of K for K-means is done automatically as described later in section 2.3. The texton are extracted individually for each region of interest viz. eyes, nose and lips. Figure 3 illustrates extraction of textons from a nose region.

Since the face images of monuments have natural texture which is complex, it is difficult to extract any repetitive pattern at a single scale. However, irregular patterns and structures in nature have been successfully represented using fractals in computer modeling [4,8]. A fractal is a geometric pattern that is repeated at smaller scales to produce irregular shapes and surfaces that cannot be represented by classical geometry. This property of fractals motivated us to apply a multi-resolution framework for obtaining textons that would better represent the natural texture of the photographed monuments. In order to incorporate the multi-resolution framework, the procedure of texton extraction is repeated for two coarser scales as well. Thus, the actual region and the corresponding two coarser resolution versions are the three different scales at which the MR8 filter response is computed. To obtain coarser versions of an actual regions (i.e. extracted nose), the image is low-pass filtered before down sampling using Gaussian filter.

The training set that consists of true vandalized and non-vandalized regions is now used in a learning stage. This stage is used to extract textons that represent the true vandalized and non-vandalized regions. Thus, all the training images containing true vandalized eye are considered together to extract the textons representing vandalized eye. Similarly, textons for non-vandalized eye, vandalized and non-vandalized nose and lips are extracted from the training set. Figure 4 illustrates the learning of textons for vandalized nose region.

In the classification stage, the Euclidean distance between the textons of extracted regions of the test image and those from the corresponding vandalized

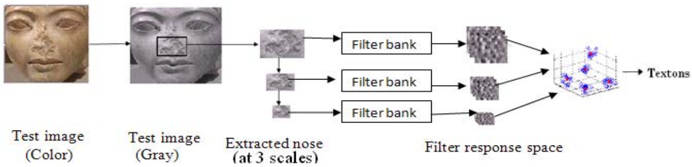


Fig. 3. Extraction of textons using the MR8 filter bank

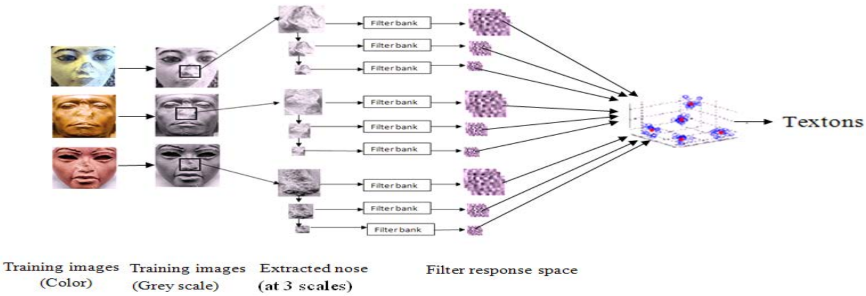


Fig. 4. Learning of texton for vandalized nose region using the MR8 filter bank

and non-vandalized region of the training images is computed. The minimum distance criteria is used to classify the region as either vandalized or non-vandalized. It may be noted that for each extracted region, viz. eyes, nose and lips, classification is performed independently. This enables detection and inpainting of more than one damaged regions in an image simultaneously. Since the textons being robust features are used to represent the training images containing a variety of textures, the vandalized and non-vandalized regions in different statues can be successfully identified.

2.3 Automatic Selection of Number of Clusters

Many of the existing clustering algorithms require the number of clusters (i.e. textons represented as vectors) to be known in advance. However, it may not be possible to pre determine the number of clusters. In our work, we use an uncomplicated approach to estimate the number of clusters. Here, we obtain clusters by varying the value of K in the K-means algorithm. We then plot a two dimensional evaluation graph, where X-axis shows number of clusters (K) and Y-axis shows a function of the pooled within cluster sum of squares around the cluster means (W_k) calculated as follows [13],

$$W_K = \sum_{r=1}^{K} \left(\sum_{\forall i,i' \in C_r} d_{i,i'} \right) \tag{3}$$

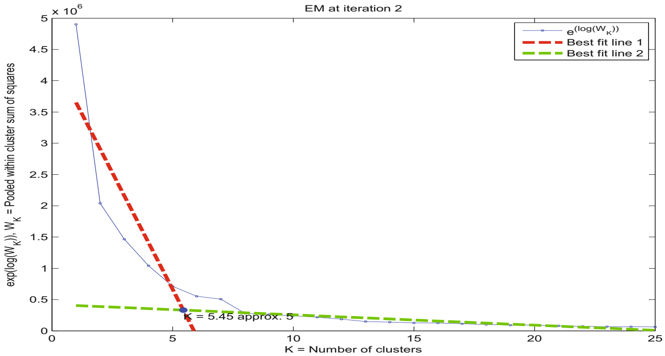


Fig. 5. Auto-selection of number of clusters K by fitting two straight lines to the data

where $d_{i,i'}$ is the squared Euclidean distance between members (i, i') of cluster C_r.

For such curves Tibshirani et al. [13] have shown that the point at which the monotonic decrease flattens markedly is taken as the best K. However, if the curve is smooth (like the one shown in figure 5) where it is difficult to determine where exactly the decrease flatters, then we have a challenging task to obtain the best value of K. To overcome this difficulty, we try to best fit two straight lines to the curve using expectation-maximization (EM) algorithm. The point of intersection of the two best fit lines then gives the approximate point at which the curve starts to flatten. We take this point to be the best K. This is illustrated in figure 5.

2.4 Inpainting

After the identification of vandalized and non-vandalized regions, final step is to inpaint vandalized regions using the corresponding non-vandalized regions from the same image or from different images available in the database. The source region that replaces the vandalized region is selected from the training set of true non-vandalized images. Here the selection criteria is the extent of similarity of the undamaged regions in the vandalized image with the true non-vandalized images. The extent of similarity is measured as the distance in the Euclidean space. Once the vandalized region and the non-vandalized source image are identified, we use Poisson image editing technique for seamless blending as described in [12].

Let f denote the test image, Ω denote the vandalized region in the test image and $\delta\Omega$ denotes the boundary of vandalized region. The minimization problem to be solved for replacing vandalized region with corresponding non-vandalized region can be written as

$$\min_f \iint_{\Omega} |\nabla f - \nabla f^*|^2 \; with \;\; f|_{\delta\Omega} = f^*|_{\delta\Omega}, \tag{4}$$

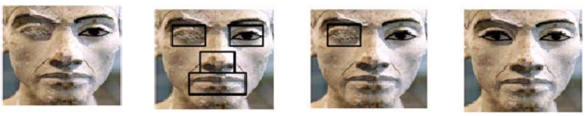
where, f^* represents the non-vandalized source, ∇ is the gradient operator.

3 Experimental Setup and Results

Our database consists of 40 facial images of statues having vandalized and non-vandalized regions. The inputs used are frontal face images downloaded from the Internet [5]. However, the proposed method can be extended for non-frontal faces by using homography with keypoints invariant to changes in scale, illumination and view point, which we intend to address in future. The spatial resolution of the images is adjusted such that all images are of the same size. A mean correction is applied to the images so that, they have the same average brightness. To select the source image from which the non-vandalized regions are chosen for inpainting, we use minimum of the squared distance between the non-vandalized regions of the test image and the database images. A manual selection of the non-vandalized source image would be required in case all the regions in the test image classified as vandalized. Figure 6-9 show results of our proposed approach.

(a) (b) (c) (d) (e)

Fig. 6. Auto-inpainting on vandalized nose; (a) input image, (b) extracted potential regions of interest, (c) detected vandalized nose, (d) inpainted image, (e) source image.



(a) (b) (c) (d)

Fig. 7. Auto-inpainting on single vandalized eye; (a) input image, (b) extracted potential regions of interest, (c) detected vandalized eye, (d) inpainted image.

Training for the eyes, nose and lips regions was done independently. For training, we have used 10 images each for vandalized and non-vandalized regions. Testing was carried out on all the images from the database including those used for training. Results are shown in figures 6-9. In figure 7, the reflected version of non-vandalized left eye has been used to inpaint vandalized right eye. However, in figure 8, since both eyes are vandalized, the face image from the training set which gives minimum squared Euclidean distance for the non-vandalized regions when compared to the test image, is used as the source image. Figure 9 shows a result where our proposed work fails to detect vandalized region. As the input image contains nose having small, the corresponding textons match those of a non-vandalized nose. This happens due to the similarity in the extracted statistics of the vandalized and non-vandalized regions, and our method fails to differentiate the two regions.

Fig. 8. Auto-inpainting on vandalized eyes; (a) input image, (b) extracted potential regions of interest, (c) detected vandalized eyes (d) inpainted image, (e) source image

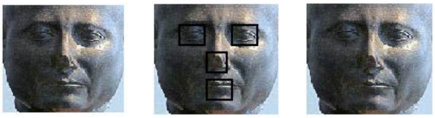


Fig. 9. Failed vandalism detection; (a) input image, (b) extracted potential regions of interest, (c) unable to detected the vandalized nose

The performance of our method is discussed in terms of precision and recall metrics defined as follows

$$Recall = \frac{|Ref \bigcap Dect|}{|Ref|}, \qquad Precision = \frac{|Ref \bigcap Dect|}{|Dect|}, \tag{5}$$

where *Ref* are the regions declared to be damaged or undamaged by volunteers and *Dect* are the regions detected as damaged or undamaged by the proposed technique. From a set of 40 images, 50 regions were found to be damaged, while 50 were undamaged. Out of 50 damaged regions 47 were correctly classified, while all 50 undamaged regions were correctly classified. For source region selection 49 out of 50 regions were correctly selected in accordance to the volunteers. The performance in terms of these metrics is summarised in table 1.

Table 1. Performance evaluation

Region type	# regions	Recall	Precision
Damaged	50	0.9400	1.0000
Undamaged	50	1.0000	1.0000

The source selection method used in our approach in not comparable with content based image retrieval (CBIR) techniques. This is because for a large vandalized region, a CBIR system may not find adequate amount of non-vandalized content to retrieve the best match relevant for inpainting. The proposed method is developed for images of statues. However, since statues and natural face images, both have same facial characteristics, the method can certainly be effective for facial regions in natural images.

4 Conclusion

In this paper, we have presented a texture based approach to automatically detect vandalized regions in face images of statues for inpainting. The results show that facial regions viz. eyes, nose and lips in images can be effectively repaired. In future we aim to incorporate invariance to rotation, illumination and view point so that vandalized regions in non-frontal face images can also be inpainted. Also we would extend our work such that damage in non-facial region can be auto-detected for inpainting.

References

1. Bertalmio, M., Vese, L., Sapiro, G., Osher, S.: Simultaneous structure and texture image inpainting. Trans. Img. Proc. 12(8), 882–889 (2003)
2. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Computer graphics (SIGGRAPH 2000), pp. 417–424 (2000)
3. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. Trans. Img. Proc. 13(9), 1200–1212 (2004)
4. Emerson, C., Lam, N., Quattrochi, D.: Multi-scale fractal analysis of image texture and patterns. Photogrammetric Engg. and Remote Sensing 65(1), 51–62 (1999)
5. Google Images (March 2012), `http://www.images.google.com`
6. Hays, J., Efros, A.A.: Scene completion using millions of photographs. ACM Transactions on Graphics (SIGGRAPH 2007) 26(3) (2007)
7. Katahara, S., Aoki, M.: Face parts extraction windows based on bilateral symmetry of gradient direction. In: Solina, F., Leonardis, A. (eds.) CAIP 1999. LNCS, vol. 1689, pp. 489–497. Springer, Heidelberg (1999)
8. Legrand, P.: Local regularity and multifractal methods for image and signal analysis. In: Scaling, Fractals and Wavelets. Wiley (January 2009)
9. Levin, A., Zomet, A., Weiss, Y.: Learning how to inpaint from global image statistics. In: Int. Conf. on Computer Vision, vol. 1, pp. 305–312 (October 2003)
10. Masnou, S., Morel, J.M.: Level lines based disocclusion. In: Int. Conf. on Image Processing, vol. 3, pp. 259–263 (October 1998)
11. Parmar, C.M., Joshi, M.V., Raval, M.S., Zaveri, M.A.: Automatic image inpainting for the facial images of monuments. In: Proceedings of Electrical Engineering Centenary Conference 2011, December 14-17, pp. 415–420 (2011)
12. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. In: ACM SIGGRAPH 2003 Papers, SIGGRAPH 2003, pp. 313–318 (2003)
13. Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistic. Journal of Royal Stat. Soc., B 63(2), 411–423 (2001)
14. Varma, M., Zisserman, A.: A statistical approach to material classification using image patch exemplars. IEEE Trans. PAMI 31(11), 2032–2047 (2009)
15. Varma, M., Zisserman, A.: Classifying images of materials: Achieving viewpoint and illumination independence. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 255–271. Springer, Heidelberg (2002)
16. Štruc, V., Pavešić, N.: Illumination invariant face recognition by non-local smoothing. In: Proceedings of the BioID MultiComm., pp. 1–8 (2009)
17. Weickert, J.: Theoretical foundations of anisotropic diffusion in image processing. Computing, Suppl. 11, 221–236 (1996)
18. Whyte, O., Sivic, J., Zisserman, A.: Get out of my picture! internet-based inpainting. In: Proceedings of the 20th British Machine Vision Conference, London (2009)