



**HAL**  
open science

# Model and Dictionary guided Face Inpainting in the Wild

Reuben Farrugia, Christine Guillemot

► **To cite this version:**

Reuben Farrugia, Christine Guillemot. Model and Dictionary guided Face Inpainting in the Wild. ACCV workshop on New Trends in Image Restoration and Enhancement, Nov 2016, Taipei, Taiwan. pp.17. hal-01388971

**HAL Id: hal-01388971**

**<https://inria.hal.science/hal-01388971>**

Submitted on 15 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Model and Dictionary guided Face Inpainting in the Wild

Reuben A. Farrugia, Christine Guillemot

University of Malta, INRIA Rennes-Bretagne-Atlantique

**Abstract.** This work presents a method that can be used to inpaint occluded facial regions with unconstrained pose and orientation. This approach first warps the facial region onto a reference model to synthesize a frontal view. A modified Robust Principal Component Analysis (RPCA) approach is then used to suppress warping errors. It then uses a novel local patch-based face inpainting algorithm which hallucinates missing pixels using a dictionary of face images which are pre-aligned to the same reference model. The hallucinated region is then warped back onto the original image to restore missing pixels.

Experimental results on synthetic occlusions demonstrate that the proposed face inpainting method has the best performance achieving PSNR gains of up to 0.74dB over the second-best method. Moreover, experiments on the COFW dataset and a number of real-world images show that the proposed method successfully restores occluded facial regions in the wild even for CCTV quality images.

## 1 Introduction

Human faces captured in unconstrained acquisition conditions are often partially occluded by other objects, long hair, beard and other accessories such as sunglasses and hats. Research in cognitive sciences has shown that the recognition ability of humans degrades when considering non-frontal [1] and partially occluded faces [2, 3]. Moreover, a recent study conducted by Klare *et al.* [4] revealed that state-of-the-art automated face recognition software still achieve low accuracies under these difficult conditions.

Image inpainting can be used to restore the occluded areas within an image and can be divided into three main categories: i) Diffusion-based, ii) Exemplar-based and iii) Global based methods. *Diffusion-based inpainting* [5] methods adopt smoothness priors via partial differential equations (PDEs) to diffuse local structures from the non-occluded to the occluded parts of an image. These methods are naturally well suited for completing small regions but inappropriate to conceal large areas. *Exemplar-based inpainting* methods [6] are local greedy methods that approximate the missing pixels within a patch using a combination of neighbouring patches with similar texture. These methods exploit the self similarity within an image and manage to better recover the missing texture. The third family of methods, often referred to as *Global inpainting* methods [7] introduce a global prior (*e.g.* low-rank, least squares or sparse) on the whole

image. These methods are known to work well when the missing areas are not too large. A complete review on image inpainting can be found in [8].

Face inpainting is a special case of inpainting where the image to be restored is known to be a face. Most face inpainting approaches found in the literature [9–15] adopt a dictionary of frontal faces as global prior to restore the occluded face images. Min and Dugelay [16] demonstrated that global methods are not suitable to conceal sparse occlusion and propose to use Fields of Experts [17] to inpaint sparse facial occlusions. However, the limitation of this approach (and all generic inpainting methods) is that it does not exploit the facial structure and thus fails to restore large occlusions. Moreover, none of the above mentioned methods are suited to restore non-frontal face images, which limits their application in practice. More recently, Burgos-Artizzu *et al.* [18] presented a face inpainting method which can be used to conceal non-frontal images. This method used a trained sub-dictionary per cluster to cater for different pose and expression variations. However, each sub-dictionary is of very large size and computationally intensive to train. Moreover, their training methodology does not guarantee that the pose of the face to be inpainted is well represented by one of the clusters.

The contribution of this paper is two-fold. We first present a *face-frontalization method*<sup>1</sup> that synthesizes the frontal pose of a face acquired in the wild. Unlike the method in [19], this work adopts a simple 2D warping method to get a first approximation of the frontal view and then applies a Robust Principal Component Analysis (RPCA) based method to suppress warping distortions. Unlike existing RPCA-based face denoising [20, 21] and alignment [22] methods, our method does not need multiple representations of the same test subject. In fact, our method considers that the training images (which do not include the test subject) to be samples on a manifold and tries to push the distorted frontal face on this manifold. We also present a novel *local patch-based face inpainting algorithm* which exploits the facial structure when computing the inpainting process. The proposed system is different from the work in [18] since it does not need to be trained a-priori and its application is not limited to the poses considered during the training-phase. It differs from existing face inpainting methods [9–16] since it is a local method which ensures that the reconstructed face images is locally smoother and in general more natural. It diverges from local patch-based image inpainting methods such as [6, 23–26] since the missing pixels are hallucinated using a dictionary of aligned faces. This work is somewhat related to face super-resolution [27, 28] which is not suitable for inpainting.

## 2 Related Work

### 2.1 Face Inpainting

Hwang and Lee [14] presented the first work on face inpainting, where they used morphable models to restore partially occluded faces. In their work, the

<sup>1</sup> Frontalization is a terminology recently introduced in [19] to refer to the process of synthesizing a frontal view of a person whose original pose is unconstrained.

authors approximate the deformation required for the non-occluded region, and then use the same deformations to synthesize the occluded region. Mo *et al.* [9] replaced the morphable model with a dictionary of frontal face images and derived the combination weights using non-negative least squares (NNLS) on the non-occluded part, which are then used to reconstruct the occluded face region. Several Principal Component Analysis (PCA) [10, 13, 15] based schemes have been proposed to extend the work of Mo *et al.*, where they try to restore a face image using a linear combination of eigenfaces.

The authors in [12] adopt a Bayesian formulation unifying the occlusion detection and recovery stages. The occlusion is detected using Graph Cuts while the inpainting is computed using Maximum a Posteriori (MAP) estimation. Another iterative strategy was presented by Zhou *et al.* [11] where Markov Random Fields (MRF) models were used to detect the support of the occluded region while sparse coding was used to approximate the reconstructed face. Min and Dugelay [16] employed RPCA to detect the sparse occlusions which are then inpainted using FoE.

A Pose and expression invariant face inpainting algorithm was recently presented in [18]. This method clusters more than 16,000 face images based on their landmark’s positions obtained using Robust Cascaded Pose Regression (RCPR) [29] and a dictionary is learned for each cluster using an online Sparse Coding method described in [30]. The landmarks and occlusion mask of the test face image are first extracted using RCPR. The landmarks are used to determine the closest cluster, and the corresponding dictionary, and the reconstructed face is computed using sparse coding. The training is performed off-line, but is very computationally intensive. Moreover, the pose variations considered in this method are limited by the poses represented by the clusters.

## 2.2 Overview of Exemplar based Inpainting

Exemplar based Inpainting methods exploit the self-similar property of natural images to propagate texture patterns from non-occluded pixels to fill holes. The seminal work of Criminisi [6] divides the input image in a known part  $\Phi$  and an unknown part  $\Omega$ . As shown in figure 1, the pixels in  $\Omega$  that are neighbours of  $\Phi$  form the front line, denoted by  $\delta\Omega$ . The target patch  $\Psi_p$  centred at  $p$  on  $\delta\Omega$  is selected using a patch order selection scheme. The patch  $\Psi_p$  has two non-overlapping parts: the known region  $\Psi_p^k = \Psi_p \cap \Phi$  and the unknown region  $\Psi_p^u = \Psi_p \cap \Omega$ . The missing pixels in  $\Psi_p$  are approximated using template matching.

There are several extensions to the Criminisi inpainting algorithm. Neighbour Embedding methods try to estimate the unknown pixels using a linear combination of several best matching patches, instead of simply copying the pixels of the best matching patch. The authors in [23] use non-local means to give higher weights to patches which are more similar to the known samples of the input patch. Sparse approximation was used instead in [24, 25] where they try to derive a least squares approximation of the known part under a constraint of sparsity on the weights. Several Neighbour Embedding methods were considered in [26].

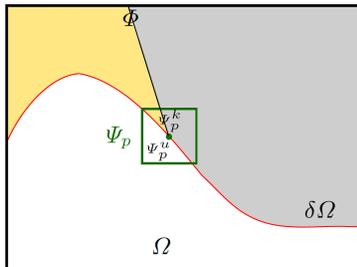


Fig. 1. Notation diagram of Criminisi's algorithm [6].

### 3 Proposed Face Inpainting in the Wild System

We introduce here an approach which allows pose and orientation independent inpainting of facial images. The schematic diagram of the proposed method is illustrated in figure 2, where in this example we are going to inpaint the cigar from the actor's mouth (marked in green). This method assumes that the landmark points (marked by red crosses) are available (either manually inserted or else detected using a state-of-the-art landmark detection algorithm such as [29, 31]). The face region  $\mathbf{F}$  is then segmented from the background region  $\mathbf{B}$  based on the peripheral landmark points marked with a dotted blue line. Only the occlusion on the face segment  $\mathbf{F}$  is concealed by the proposed system, since this contains the most important and discriminative information<sup>2</sup>. We will show in section 4.5 that the occluded region in  $\mathbf{B}$  can be restored using off the shelf inpainting algorithms since these areas are generally smooth and easier to conceal by exploiting the self-similar texture present in the neighbourhood of  $\Psi_p$ .

The standard Delaunay triangulation method is computed from the set of  $P$  landmark locations. Each landmark triplet from the face image  $\mathbf{F}$  forms a triangle, as shown in solid blue lines. A reference face shape model  $\mathbf{S}$  represents the shape of a frontal 2D-face whose  $P$  landmark points are normalized such that the distance between the eye centres (inter-eye distance) is of 40 pixels<sup>3</sup>. The frontal pose of the test subject  $\hat{\mathbf{F}}$  is synthesized using the proposed *frontalization* method, which employs a list of training face images aligned such that the inter-eye distance is set to 40 pixels (see section 3.1). The alignment of the training images can be done off-line since it is independent from the pose of the facial image being processed. The *Face inpainting process* then employs the same list of training face images to fill the occluded region  $\Omega$  (see section 3.2) which hallucinates the missing pixels to restore  $\hat{\mathbf{F}}^*$ .

<sup>2</sup> The accuracy of the segmentation process is dependent on the number of landmarks used. In this example, one can use more landmark points to segment the lower part of the face region (and possibly the entire occluded region) in segment  $\mathbf{F}$ .

<sup>3</sup> An inter-eye distance of 40 pixels is sufficient for identification. Nevertheless, this method is not affected by this resolution and higher (or lower) resolutions can be configured.

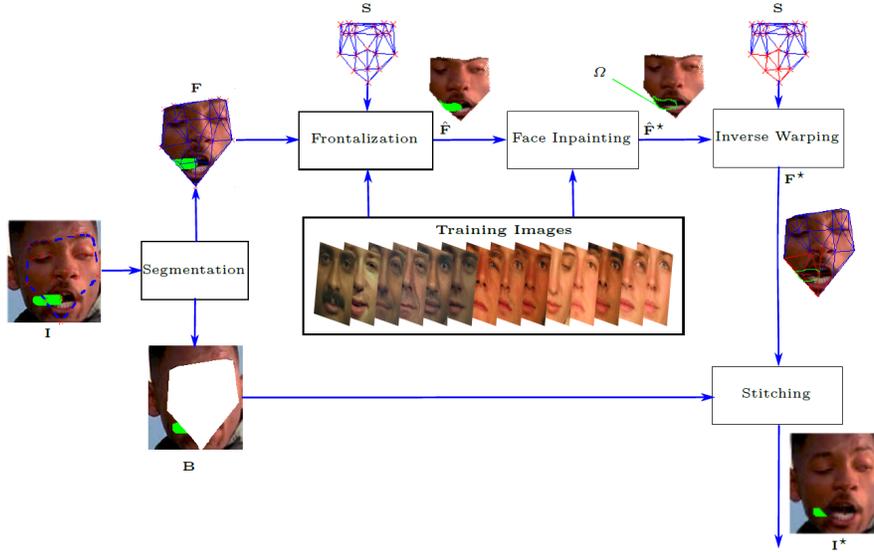


Fig. 2. Schematic diagram of the proposed Face Inpainting in the Wild method.

Only the recovered region  $\Omega$ , and therefore the corresponding triangles marked in red, is used to warp the inpainted region from the frontal view  $\hat{\mathbf{F}}^*$  onto the original image  $\mathbf{F}$  to fill its missing pixels. The resulting inpainted face image  $\mathbf{F}^*$  is made of the known part of  $\mathbf{F}$  and the warped inpainted region  $\Omega$  from the frontal view  $\hat{\mathbf{F}}^*$ . The inpainted face segment  $\hat{\mathbf{F}}^*$  and background images are then simply stitched together to form the restored image  $\mathbf{I}^*$ .

### 3.1 Face Frontalization

As mentioned above, the face image  $\mathbf{F}$  is partitioned into triangular regions using the Delaunay triangulation method. Similarly, the face shape model  $\mathbf{S}$  is partitioned into triangles. We use the 21 facial landmark points defined in [32] since it caters for both affine (rotation, translation and scaling) and more complicated 3-dimensional deformations from the frontal view. The first part of the face frontalization method proposed here is to transfer the pixels that reside within each triangle in  $\mathbf{F}$  onto the corresponding triangle in  $\mathbf{S}$  using piecewise affine transformation, to get the warped face  $\mathbf{F}_w$ . This warping process does not manage to recover all the pixels within  $\mathbf{F}_w$  because of scale and occlusions caused by the facial pose, and it therefore contains a number of missing pixels. Moreover,  $\mathbf{F}_w$  contains warping errors which can be significant when the face orientation deviates considerably from the frontal pose. The missing pixels can be recovered using the face inpainting algorithm described in section 3.2, while the suppression of warping distortions will be considered in the remaining part of this subsection.

We define  $\mathbf{s}_w$  to represent the support of known pixels in  $\mathbf{F}_w$  and denote a matrix  $\mathbf{M} \in \mathbb{R}^{|\mathbf{s}_w| \times N}$  where each column-vector corresponds to a vectorized representation of the  $N$  training images and the rows that are not contained in  $\mathbf{s}_w$  are removed. We also denote a vector  $\mathbf{f}_w = \hat{\mathbf{f}}_w + \mathbf{e}_w$  to represent the vectorized form of  $\mathbf{F}_w$ , where  $\mathbf{e}_w$  stands for the warping error and  $\hat{\mathbf{f}}_w$  is the noise-free version of the frontal face which is unknown. The matrix  $\mathbf{M}$  and vector  $\mathbf{f}_w$  are concatenated to form an augmented matrix  $\mathbf{M}_a$ , where without loss of generality  $\mathbf{f}_w$  is placed as the first column-vector in  $\mathbf{M}_a$ .

Several studies have shown that face images reside on a low-dimensional manifold [33, 34], and this property was extensively exploited in face analysis literature [35, 36]. This observation suggests that if we vectorize a number of aligned and undistorted face images and stack them into a matrix  $\mathbf{M}$ , then  $\mathbf{M}$  can be approximated by a low-rank matrix. However, given that  $\mathbf{f}_w$  is distorted by warping errors, this vector will be linearly independent from all the other column vectors of  $\mathbf{M}$ , which will inherently increase the rank of  $\mathbf{M}_a$ . We therefore decompose the augmented matrix  $\mathbf{M}_a$  using the Robust PCA (RPCA)[20]

$$\text{minimize } \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \text{ subject to } \mathbf{A} + \mathbf{E} = \mathbf{M}_a \quad (1)$$

where  $\|\mathbf{A}\|_*$  denotes the nuclear norm of  $\mathbf{A}$  that corresponds to the sum of its singular values,  $\|\mathbf{E}\|_1$  denotes the  $l_1$ -norm of  $\mathbf{E}$  and  $\lambda$  is a regularization parameter. The authors in [20] recommend to set the regularization parameter

$$\lambda = \frac{C}{\sqrt{\max(m^2, N)}} \quad (2)$$

where  $C$  is a positive constant and the resolution of the frontal image is  $m \times m$  ( $m = 67$  in this work). This problem is convex and can be solved in polynomial time. In this work we solve this problem using the inexact Augmented Lagrange Multiplier method [37].

The rationale behind this formulation is that the column-vectors in  $\mathbf{M}$  are points that reside on a low-dimensional manifold. The vector  $\hat{\mathbf{f}}_w$  is an undistorted face image, which is unknown, resides on the same manifold. However, the observed warped face image  $\mathbf{f}_w$  which contains warping errors does not reside on the manifold. By forcing the matrix  $\mathbf{A}$  to be low-rank we are pushing the warped face image  $\mathbf{f}_w$  on the low-dimensional manifold while the distortion caused by warping will go in the error matrix  $\mathbf{E}$ . The frontal denoised face image is then simply the first column-vector in  $\mathbf{A}$ , where the frontal face image  $\hat{\mathbf{F}}$  is derived by reshaping it to an  $m \times m$  image.

The use of RPCA was considered in the past for denoising [20, 21] and alignment [22]. However, opposed to our proposed frontalization method, these methods are applied on a set of face images of the same subjects, *i.e.* RPCA is applied directly on a matrix which consists of different representations of the same subject. On the other hand, our approach considers the training images to represent points on a low-dimensional manifold, where the subject being denoised is not present in the list of training images.

### 3.2 Patch based Inpainting of Frontalized Face Images

Figure 3 depicts the outline of the proposed face inpainting method. At each iteration, the patch  $\Psi_p$  with the highest priority is chosen for inpainting, where  $p$  represents the position of the patch. The priorities of each patch residing on the frontal line are assigned based on the product of three terms

$$P(p) = C(p)D(p)E(p) \quad (3)$$

where the confidence term  $C(p)$  is a measure of reliable information surrounding the pixel at position  $p$ , the data term  $D(p)$  reflects the presence of some structure perpendicular to the front-line, and the edge term  $E(p)$  that caters for structure which is non-orthogonal to the front line. More information and detail on how these terms are computed can be found in [26].

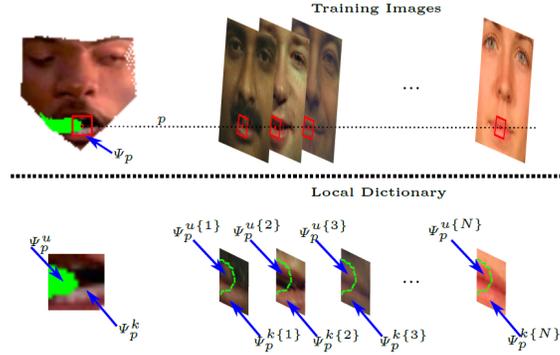


Fig. 3. Schematic diagram of the proposed Local Face Inpainting algorithm.

Given a training set of  $N$  facial images, we extract a patch at position  $p$  from each training image, where we denote  $\Psi_p^{\{i\}}$ ,  $i \in [1, N]$  to represent the collocated patch extracted from the  $i$ -th image. To simplify the notation, we divide the training patches  $\Psi_p^{\{i\}}$  into a known part  $\Psi_p^{k\{i\}} = \Psi_p^{\{i\}} \cap \Psi_p^k$  and an unknown part  $\Psi_p^{u\{i\}} = \Psi_p^{\{i\}} \cap \Psi_p^u$ . In other words,  $\Psi_p^{k\{i\}}$  ( $\Psi_p^{u\{i\}}$ ) corresponds to the collocated pixels from the  $i$ -th training image which are known (unknown) in  $\Psi_p$ .

The proposed method employs two dictionaries:  $\mathbf{D}_p^k$  which is used to derive the linear combination weights and  $\mathbf{D}_p^u$  which is used to approximate the missing pixels. The dictionary  $\mathbf{D}_p^k$  ( $\mathbf{D}_p^u$ ) is constructed by stacking the  $N$  vectorized representation of  $\Psi_p^{k\{i\}}$  ( $\Psi_p^{u\{i\}}$ ). The next step is to search for  $K$  neighbours that best represent the known part  $\Psi_p^k$  within  $\mathbf{D}_p^k$ . We define the support of the  $K$ -nearest neighbours by  $\mathbf{s}_p$  such that  $\mathbf{D}_p^k(\mathbf{s}_p)$  correspond to the  $K$  column-vectors which are closest in Euclidean space to the vectorized representation of  $\Psi_p^k$ . We then derive the combination weights  $\mathbf{w}^*$  that minimize the following constrained optimization problem

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\Psi_p^k - \mathbf{D}_p^k(\mathbf{s})\mathbf{w}\|_2^2 \text{ subject to } \sum_{i=1}^K w_i = 1 \quad (4)$$

This formulation corresponds to the locally linear embedding (LLE) introduced in [33] for data dimensionality reduction, which has an analytical solution. Note that the sum to unity constraint is important since this ensures that the reconstructed pixels preserve the original bit depth. Once the combination weight vector  $\mathbf{w}^*$  is found, the same weights are applied in the linear combination of the "unknown" pixels<sup>4</sup> in the dictionary  $\mathbf{D}_p^u(\mathbf{s}_p)$ , where the support  $\mathbf{s}_p$  is used to choose the  $K$  column vectors from  $\mathbf{D}_p^u$ . The unknown part  $\Psi_p^u$  is therefore approximated using

$$\Psi_p^{u*} = \mathbf{D}_p^u(\mathbf{s}_p)\mathbf{w}^* \quad (5)$$

This process is iterated until all occluded pixels are inpainted, and the final solution is denoted by  $\hat{\mathbf{F}}^*$ . The accuracy of this method is dependent on the accuracy of the frontalization method employed. In fact, if the frontalized image contains warping deformations, these will propagate within the missing region causing severe distortions in the resulting inpainted face.

We emphasize here that the LLE inpainting method in [26] approximates missing pixels using a combination of spatial neighboring patches with similar texture. However, it does not exploit the contextual information of the facial component being inpainted, and thus fails to reconstruct natural face images. The proposed method (see figure 3) finds collocated patches present within an aligned dictionary to find the neighboring patches suitable for reconstruction *i.e.* the right eye patches are used to inpaint the right eye of a test subject.

## 4 Experimental Results

The training images used in the following experiments are derived from the Color Feret dataset [38]. All these images are registered based on their eyes and mouth centre coordinates such that the inter-eye distance is of 40 pixels and the resolution is  $m \times m$  pixels, where in all experiments  $m = 67$ . For all the experiments, subjects wearing glasses were removed<sup>5</sup>, resulting in a total of  $N = 2081$  facial images to be used for training.

The frontalization method was tested using the LFW [39] dataset, while the inpainting process was evaluated using both AR [40] and the COFW [29] datasets. The *AR dataset* was used to evaluate the performance of the proposed face inpainting on synthetic occlusions where the ground-truth is available. For

<sup>4</sup> These pixels are known in the dictionary  $\mathbf{D}_p^u$  since the training images do not have occlusions. However, these pixels are collocated with the unknown pixels within the patch being inpainted  $\Psi_p$ .

<sup>5</sup> Subjects wearing glasses were removed since we want to use the training images to synthesize people without facial occlusions.

these experiments, one image per subject is considered, where subjects wearing glasses were removed resulting in a total of 84 test images. These images were aligned to the training images based on the eye and mouth center coordinates. The *COFW dataset* is a more challenging dataset which contains images with various poses and occlusions. However, this dataset cannot be used for objective evaluation since the ground truth (non-occluded face) is not available. This dataset is mainly used to qualitatively assess the performance of various face inpainting algorithms. The proposed system was also evaluated on a number of real-world images (see section 4.5) presenting more challenging scenarios.

Unless otherwise stated, we assume that the occluded region (called mask in the sequel) is manually marked, since the automated detection of occlusion is not in scope of this work. The automated detection of the occlusion mask introduces inaccuracies (see figure 8 for example) which will degrade the quality of the restored image. In order to be able to fairly assess the performance of the proposed approach with respect to existing methods, it was decided to use the same known occlusion mask. In addition, the landmark points are assumed to be manually defined by the user. The code is made available on-line <sup>6</sup>

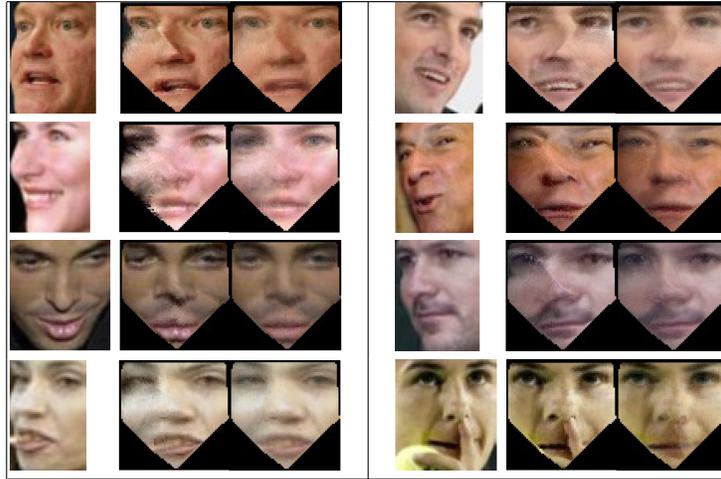
#### 4.1 Frontalization Analysis

Figure 4 compares the faces warped using piecewise affine warping (2<sup>nd</sup> column) and the final frontalized face where warping artefacts are suppressed using our modified RPCA method (3<sup>rd</sup> column). Pixels which are left empty in the frontal view are treated as holes and inpainted using the method described in section 3.2. In this experiment we consider a number of images from the LFW dataset [39], and set  $C = 2.3$  since it was found to give satisfactory performance. One can also notice that images that are warped using piecewise affine transformation without denoising contain significant warping artefacts that negatively affect the inpainting process. In contrast, it can be seen that the proposed RPCA denoising method manages to remove a substantial portion of the noise induced by warping. It can be noticed from the last row of images that our modified RPCA method treats occlusions as noise deviating the frontal face from the low-dimensional manifold and partially suppresses these occlusions automatically without having the occlusion mask.

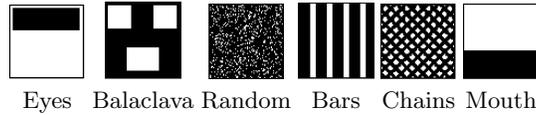
#### 4.2 Face Inpainting Parameter Selection

In this subsection we investigate the effect of two parameters, namely the number of neighbours  $K$  and the patch size, which affect the performance of the proposed face inpainting algorithm. For this experiment we considered six different masks: i) Both eyes are occluded (Eyes), ii) the face is covered by a balaclava (Balaclava), iii) 85% of the pixels are randomly occluded (Random), iv) the face is occluded by vertical bars (Bars), v) the face region is occluded by a chains link fence (Chains) and vi) the mouth region is covered (Mouth). Figure 5 show the masks

<sup>6</sup> Face Inpainting Demo: <https://goo.gl/ws3NG4>



**Fig. 4.** The first column from each cell shows the cropped images from the LFW dataset. The second column represents the piecewise affine warped face image and the fourth column were restored using our modified RPCA method.

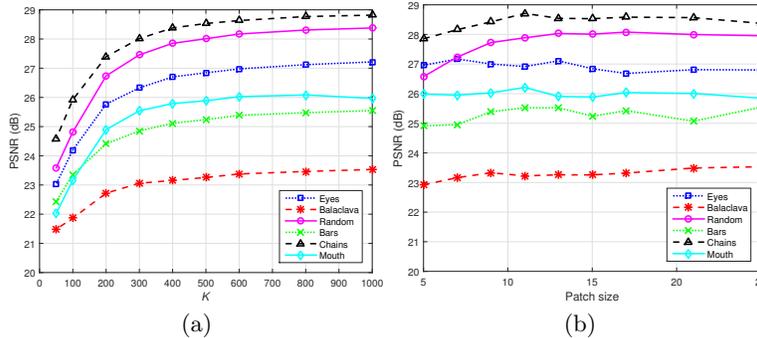


**Fig. 5.** Occlusion masks used for this experiment.

used for this experiment where the 84 test images from the AR dataset (described above) were used. The test images are already aligned with the training images and therefore the forward and inverse warping processes were disabled. Figure 6 depicts how the quality, measured in terms of PSNR, is affected by the number of neighbours  $K$  and the patch size. It can be seen from figure 6 a) (which was obtained by fixing the patch size to 15) that the PSNR increases steadily up till around  $K = 500$  beyond which no substantial gain in performance is observed. On the other hand, the results in figure 6 b) (which was obtained by fixing  $K = 500$ ) show that the performance is quite independent of the patch size. These results seem to generalize for all six masks considered. To get a compromise between quality and complexity we set  $K = 500$  and a patch-size of 15 for the following experiments.

### 4.3 Synthetic Occlusion

In this experiment we use the same set of 84 images from the AR dataset and masks discussed in section 4.2. Once again, given that the test images are aligned to the training images, the forward and inverse warping were disabled. We compare our proposed face inpainting method (described in section 3.2) with four



**Fig. 6.** Evaluating the performance of the proposed face inpainting method in section 3.2 as a function of a) number of candidates  $K$  and b) patch size.

Method	Eyes	Balaclava	Random	Bars	Chains	Mouth
Mo <i>et al.</i> [9]	25.19	22.30	22.98	23.51	23.56	25.81
Wang & Tao [10]	25.92	17.28	25.59	21.10	27.02	20.07
Zhou <i>et al.</i> [11]	26.37	22.91	25.48	24.67	26.86	24.25
Min & Dugelay [16]	20.22	14.10	27.38	18.62	27.79	14.13
Ours	<b>26.84</b>	<b>23.26</b>	<b>27.99</b>	<b>25.24</b>	<b>28.52</b>	<b>25.88</b>

**Table 1.** Quality analysis using the PSNR quality metric.

other baseline methods [9–11, 16]. The parameters used by these baseline methods were optimized using a methodology similar to the one described in section 4.2. Qualitative tests were conducted on frontal faces because all baseline methods [9–11, 16] were designed for frontal faces, and testing on non-frontal faces will penalize the baseline methods not our proposed framework.

The results in Table 1 show that the method of Zhou *et al.* [11] provides good performance, especially when the occlusion mask is not sparse. On the other hand, the method of Min & Dugelay [16] performs very well only when the error is sparse. Nevertheless, the proposed method achieves the best performance for all types of occlusions, achieving averaged PSNR gains of up to 0.74 dB relative to the second-best method.

#### 4.4 Real Occlusion

To assess the ability of the proposed method in the wild we used the testing images from the COFW dataset. Here, and in the following subsection, we adopt the full system described in section 3. Since the baseline methods against which we are comparing are not able to perform face inpainting in the wild, we have extended the reference methods to handle different poses by integrating them within our framework *i.e.* the test image is frontalized using our method described in section 3.1, the frontalized image is inpainted using either our approach (section 3.2)

or one of the baseline methods, and the reconstructed region is warped back to restore the missing pixels. Figure 7 shows some results with the COFW dataset, where it can be seen that the method of Min & Dugelay does not perform well on all these images. This is attributed to the fact that their method (like all exemplar-based inpainting methods including [26]) does not employ the facial structure when inpainting and therefore its application is limited to conceal very small occlusions. It can be seen that the images reconstructed using the methods in [10, 11] are generally noisy while those recovered by Mo *et al.* [9] provide discontinuities at the boundaries between the known and inpainted part of the face image. On the other hand, the images recovered using the proposed method are more natural and of higher quality. It must be noticed that the proposed method conceals regions which intersect with the face region  $\mathbf{F}$  but not the background region  $\mathbf{B}$ . Nevertheless, it will be shown in section 4.5 that the missing pixels in  $\mathbf{B}$  can be recovered using off the shelf inpainting methods.

Given that the code of Burgos-Artizzu *et al.* [18] was not released we could not compare it with our method directly. Nevertheless, we use the same images and occlusion mask adopted by the authors in [18]<sup>7</sup>. Figure 8 compares our proposed scheme with that of Burgos-Artizzu *et al.*. The face images reconstructed using their approach show visual discontinuities (see left image) and ringing artefacts (see right image). Opposed to this, our proposed method reconstructs smoother and visually more pleasing facial images.

#### 4.5 Real-World Examples

In this experiment we picked up a number of images from the Internet where the faces had a wider range of pose variations, quality (include CCTV quality images) and image resolutions (up to 1 Mega pixel) compared to those in the COFW dataset. The occlusion on the face region  $\mathbf{F}$  were inpainted using the proposed method while the occlusion on the non-facial part  $\mathbf{B}$  was concealed using the LLE based inpainting algorithm proposed in [26].

Figure 9 shows that the proposed method manages to reconstruct the occluded part of the face quite effectively, even when the pose is almost profile. This contrasts with the method in [18] where their approach is restricted to poses considered during training. Apart being successful for image editing (see figure 9 (a) - (d)), the proposed method was effective in reconstructing face images captured by CCTV (see figure 9 (f) - (h)) which are quite noisy and reveals the face which is appropriately camouflaged to reduce identification (see figure 9 (e)). The face inpainting process took between 5 - 11s on a standard machine (Intel Core i7 with 8GB of memory), using MATLAB.

## 5 Comments and Conclusion

In this paper we propose a novel face inpainting algorithm that can be used to restore occluded facial regions in the wild. This method introduces two innovative

<sup>7</sup> The images provided by Burgos-Artizzu *et al.* [18] were in grayscale, and therefore only results on grayscale images are presented here.

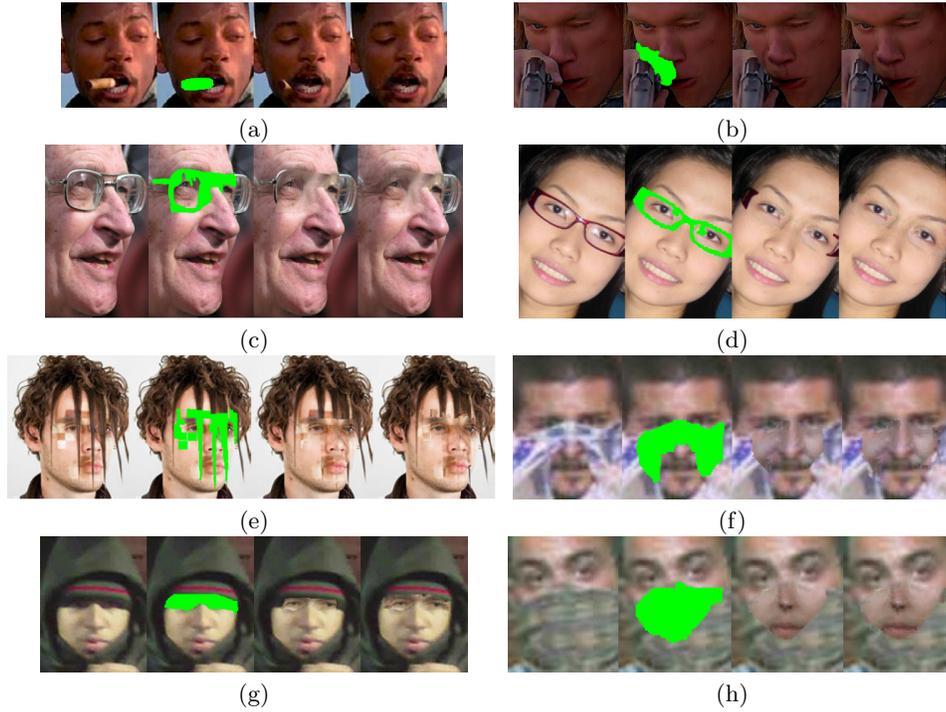


**Fig. 7.** Images from the COFW dataset which were recovered using different face inpainting algorithms best viewed by zooming onto the images.

components: i) it presents a frontalization method that synthesizes a frontal view of the face in an unconstrained pose and ii) it describes a local patch-based face inpainting algorithm. Experimental results show the importance of using our modified RPCA method to suppress warping artefacts introduced by the piecewise affine warping step, especially when the pose is almost profile. RPCA in our method differs from its classical use for denoising since it can denoise subjects not included in the dictionary. Experimental results also show that the proposed face inpainting in the wild system manages to inpaint different occlusions and reconstructs images of higher quality. The main advantages of the



**Fig. 8.** Original image (left), the results obtained using the Burgos-Artizzu *et al.* [18] (center) and the proposed face inpainting method (right).



**Fig. 9.** Original image (*left*), the original image with the region to be inpainted marked in pure green (*center-left*), the inpainted image to remove the occluded region intersecting the frontal face (*center-right*) and the final inpainted image where the background was inpainted using the LLE method in [26] (*right*).

proposed method are i) it can inpaint faces with a wider range of pose variations than existing methods, ii) its complexity is mainly dependent on the number of neighbours  $K$  and the size of the holes to be filled, iii) the restored images are generally locally smoother and of higher quality. This method was also evaluated on real-world images at different orientations, resolutions and quality, including CCTV quality images taken during riots in UK.

## References

1. Hancock, P.J., Bruce, V., Burton, A.: Recognition of unfamiliar faces. *Trends in Cognitive Sciences* **4** (2000) 330 – 337
2. Terry, R.L.: How wearing eyeglasses affects facial recognition. *Current Psychology* **12** (1993) 151–162
3. Yarmey, A.D.: Eyewitness recall and photo identification: A field experiment. *Psychology, Crime Law* **10** (2004) 53–68
4. Klare, B.F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., Grother, P., Mah, A., Burge, M., Jain, A.K.: Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2015) 1931–1939
5. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: *Annual Conference on Computer Graphics and Interactive Techniques*. (2000) 417–424
6. Criminisi, A., Perez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*. **13** (2004) 1200–1212
7. Liu, J., Musialski, P., Wonka, P., Ye, J.: Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** (2013) 208–220
8. Guillemot, C., Meur, O.L.: Image inpainting : Overview and recent advances. *IEEE Signal Processing Magazine* **31** (2014) 127–144
9. Mo, Z., Lewis, J., Neumann, U.: Face inpainting with local linear representations. In: *British Machine Vision Conference*. (2004) 37.1–37.10
10. Wang, Z.M., Tao, J.H.: Reconstruction of partially occluded face by fast recursive pca. In: *International Conference on Computational Intelligence and Security Workshops*. (2007) 304–307
11. Zhou, Z., Wagner, A., Mobahi, H., Wright, J., Ma, Y.: Face recognition with contiguous occlusion using markov random fields. In: *IEEE International Conference on Computer Vision*. (2009) 1050–1057
12. Lin, D., Tang, X.: Quality-driven face occlusion detection and recovery. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2007) 1–7
13. Hosoi, T., Nagashima, S., Kobayashi, K., Ito, K., Aoki, T.: Restoring occluded regions using fw-pca for face recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. (2012) 23–30
14. Hwang, B.W., Lee, S.W.: Reconstruction of partially damaged face images based on a morphable face model. *IEEE Transactions Pattern Analysis and Machine Intelligence* **25** (2003) 365–372
15. Liwicki, S., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Euler principal component analysis. *International Journal of Computer Vision* **101** (2012) 498–518
16. Min, R., Dugelay, J.L.: Inpainting of sparse occlusion in face recognition. In: *IEEE International Conference on Image Processing*. (2012) 1425–1428
17. Roth, S., Black, M.J.: Fields of experts. *International Journal of Computer Vision* **82** (2009) 205–229
18. Burgos-Artizzu, X.P., Zepeda, J., Clerc, F.L., Perez, P.: Pose and expression-coherent face recovery in the wild. In: *IEEE International Conference on Computer Vision Workshop*. (2015) 877–885
19. Hassner, T., Harel, S., Paz, E., Enbar, R.: Effective face frontalization in unconstrained images. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2015) 4295–4304

20. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? *Journal in ACM* **58** (2011) 11:1–11:37
21. Yuan, Z., Xie, X., Ma, X., Lam, K.M.: Color facial image denoising based on rpca and noisy pixel detection. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. (2013) 2449–2453
22. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions Pattern Anaysus and Machine Intelligence* **34** (2012) 2233–2246
23. Wong, A., Orchard, J.: A nonlocal-means approach to exemplar-based inpainting. In: *IEEE International Conference on Image Processing*. (2008) 2600–2603
24. Xu, Z., Sun, J.: Image inpainting by patch propagation using patch sparsity. *IEEE Transactions on Image Processing* **19** (2010) 1153–1165
25. Studer, C., Kuppinger, P., Pope, G., Bolcskei, H.: Recovery of sparsely corrupted signals. *IEEE Transactions on Information Theory* **58** (2012) 3115–3130
26. Guillemot, C., Turkan, M., Meur, O.L., Ebdelli, M.: Object removal and loss concealment using neighbor embedding methods. *Signal Processing: Image Communication* **28** (2013) 1405 – 1419
27. Ma, X., Zhang, J., Qi, C.: Position-based face hallucination method. In: *Proceedings of the IEEE International Conference on Multimedia and Expo*. (2009) 290–293
28. Jiang, J., Hu, R., Wang, Z., Han, Z.: Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning. *IEEE Transactions on Image Processing* **23** (2014) 4220–4231
29. Burgos-Artizzu, X.P., Perona, P., Dollár, P.: Robust face landmark estimation under occlusion. In: *IEEE International Conference on Computer Vision*. (2013) 1513–1520
30. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: *Annual International Conference on Machine Learning*. (2009) 689–696
31. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2012) 2879–2886
32. Koestinger, M., Wohlhart, P., Roth, P.M., Bischof, H.: Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In: *IEEE International Conference on Computer Vision*. (2011) 2144–2151
33. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290** (2000) 2323–2326
34. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** (2003) 1373–1396
35. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face recognition using laplacian-faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27** (2005) 328–340
36. Hu, C., Chang, Y., Feris, R., Turk, M.: Manifold based analysis of facial expression. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. (2004) 81–81
37. Lin, M., Chen, L., Wu, Y.M.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical report, University of Illinois (2009)
38. Phillips, Wechsler, H., Huang, J., Rauss, P.J.: The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing* **16** (1998) 295–306

39. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst (2007)
40. Martinez, A.M., Benavente, R.: The AR Face Database. Technical report, CVC (1998)