Low-cost Automatic Inpainting for Artifact Suppression in Facial Images

André Sobiecki[1]*, Alexandru Telea*[1]*, Gilson Giraldi*[2]*, Luiz Antonio Pereira Neves*[3] *and Carlos Eduardo Thomaz*[4]

[1]Scientific Visualization and Computer Graphics, University of Groningen, Nijenborgh 9, Groningen,The Netherlands

[2]Department of Computing, National Laboratory of Scientific Computation, Petrópolis, Brazil

[3]Department of Computing, Paraná Federal University, Curitiba, Brazil

[4]Department of Artificial Intelligence Applied on Automation, University Center of FEI, São Bernardo do Campo, Brazil

{a.sobiecki, a.c.telea}@rug.nl, gilson@lncc.br, neves@ufpr.br, cet@fei.edu.br

Keywords: Image inpainting, Face reconstruction, Statistical Decision, Image Quality Index

Abstract: Facial images are often used in applications that need to recognize or identify persons. Many existing facial recognition tools have limitations with respect to facial image quality attributes such as resolution, face position, and artifacts present in the image. In this paper we describe a new low-cost framework for preprocessing low-quality facial images in order to render them suitable for automatic recognition. For this, we first detect artifacts based on the statistical difference between the target image and a set of pre-processed images in the database. Next, we eliminate artifacts by an inpainting method which combines information from the target image and similar images in our database. Our method has low computational cost and is simple to implement, which makes it attractive for usage in low-budget environments. We illustrate our method on several images taken from public surveillance databases, and compare our results with existing inpainting techniques.

1 INTRODUCTION

Facial recognition is the process of obtaining the identity of a person based on information obtained from facial appearance (Ayinde and Yang, 2002). Compared to other person identification methods such as biometric fingerprints, retina scans, and voice recognition, facial recognition has the advantage that it can be used in contexts where the collaboration of the person to be identified is not possible (Zhao et al., 2003; Hjelmas and Low, 2001). One such context is the identification of missing people based on existing photographs thereof.

In the last decades, automatic face recognition has received considerable attention from the scientific and commercial communities. However, several open issues still remain in this area. One such issue is that facial recognition tools are in general not efficient for poor quality facial images, *e.g.* in the presence of shadows, artifacts, and blurring (Zamani et al., 2008; Zhao et al., 2003; Castillo, 2006).

In some contexts, facial protographs are the only key to person identification. For example, the sites of Australian Federal Police (AFP, 2012), Federal Brazilian Government (FGB, 2012), and UK Missing People Centre (MPO, 2012) publish facial photographs of missing people. Often, such photographs are old, poorly digitized, and have artifacts such as folds, scratches, irregular luminance, molds, stamps, and written text, all of various sizes, shapes, texture, and color. Since the effectiveness of facial recognition methods depends on the quality of input images, it is of high importance that such images present the discriminant facial features (eyes, mouth, and nose) with minimal artifacts. Moreover, it is desirable that they all have a standard look, *e.g.* avoid outlier elements such as glasses, hair locks, highlights, and smiles.

We present here a computational framework for segmentation and automatic restoration of poor-quality facial images, based on a statistical model built from frontal facial images and inpainting techniques. The location of facial features is obtained by a mean image generated from a sample population of conforming facial images that provides privileged information about spatial location of the features. Salient outliers are found by statistically comparing the input image with this mean image. These outliers are next eliminated by using a modified inpainting technique which uses information from both the input image itself and the closest high-quality image in our image database. This produces images of sufficient quality for typical face recognition tasks. In

contrast to most existing digital restoration methods, we require no user input to mark regions in the image which has to be restored. Our method is simple to implement and can be run in real-time on low-cost PCs, which makes it attractive for utilization by government agencies in least developed countries.

2 RELATED WORKS

Many image inpainting techniques have been proposed in the last decade. Oliveira *et al.* (Oliveira et al., 2001) present a simple inpainting method which repeatedly convolves a 3×3 filter over the regions to inpaint. Although fast, this method yields significant blurring. Bertalmio *et al.* (Bertalmio et al., 2000; Bertalmio et al., 2001) estimate the local image smoothness, using the image Laplacian, and propagate this along isophote directions, estimated by the image gradient rotated 90 degrees and by Navier-Stokes equation. The Total Variational (TV) model (Chan and Shen, 2000a) uses an Euler-Lagrange equation coupled with anisotropic diffusion to keep the isophotes directions. Telea (Telea, 2004) inpaints an image by propagating the image information from the boundary towards the interior of the damaged area following a fast marching approach and a linear extrapolation of the image field outside the missing region. These methods give good results when the regions to restore are small.

To inpaint thicker regions, the Curvature-Driven Diffusion (CCD) method (Chan and Shen, 2000b) enhances the TV method to drive diffusion along isophote directions. Diffusion-based texture synthesis techniques have been proposed to achieve higher quality inpainting results (Bugeau and Bertalmio, 2009; Bugeau et al., 2009). Closer to our proposal, Li *et al.* (Li et al., 2010) present semantic inpainting, where the damaged image is restored based on texture synthesis using a most similar image from a given database. Perez *et al.* (Pérez et al., 2003) and Jeschke *et al.* (Jeschke et al., 2009) present seamless image cloning that uses a Poisson process to compute the seamless filling as well as a guidance vector field that incorporates prior knowledge of the damaged domain.

Image cloning gives very good results, but is relatively expensive in CPU and memory terms. Joshi *et al.* use example images, taken from a person's photo collection, to improve the luminance, blur, contrast, and shadows of that person's photograph (Joshi et al., 2010). Chou *et al.* also use the seamless image cloning of Perez *et al.* for the editing of facial features in digital photographs (Chou et al., 2012). However, in contrast to our goal of automatic removal of facial outlier artifacts, their goal was to assist users in previewing the results of explicitly chosen image modifications.

To achieve our goal of facial image preprocessing for supporting automatic missing persons identification in low-cost contexts, we need a method which is able to

- automatically detect the most salient artifacts in a low-quality facial photograph;
- remove these artifacts in a plausible way;
- work (nearly) automatically;
- require very low computational costs.

Next, we present our proposal in order to address these requirements.

3 PROPOSED METHOD

Figure 1 illustrates our computational pipeline for facial image restoration. We start by computing an image quality index, which classifies the input image as potentially benefiting from artifact removal or not (Sec. 3.1). If the image can be improved, we next identify, or segment, its artifacts using a statistical decision method where the image is compared to an existing database of facial images of various ethnicities (Sec. 3.2). The detected artifacts are next slightly enlarged using morphological dilation (Aptoula and Lefèvre, 2008) to remove small-scale spatial noise such as artifact boundary jaggies. Finally, we eliminate the artifacts by using a semantic inpainting, a variation of the method presented in (Li et al., 2010), which combines information from the input image and the image database (Sec. 3.3).

All our images (input and database) are normalized and equalized by the framework proposed in (Amaral and Thomaz, 2008). This is needed since existing facial image databases around the world contain images with several sizes, resolutions, and contrasts. The steps of our pipeline are explained next.

3.1 Image Quality Index

Objective image quality measures have an important role in image processing applications, such as compression, analysis, registration, restoration and enhancement. One such simple measure is the image quality index (Wang et al., 2004), which compares two images $\mathbf{x}$ and $\mathbf{y}$ based on luminance $l(\mathbf{x},\mathbf{y})$, contrast $c(\mathbf{x},\mathbf{y})$ and structure $s(\mathbf{x},\mathbf{y})$ comparison measures

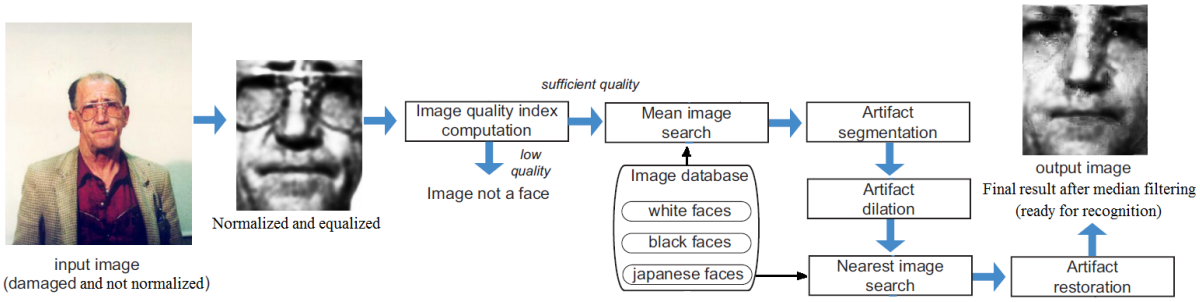


Figure 1: Pipeline of proposed method

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2}, \tag{1}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}, \tag{2}$$

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy}}{\sigma_x\sigma_y}. \tag{3}$$

where $\mathbf{x} = \{x_i | 1 \le i \le n\}$ and $\mathbf{y} = \{y_i | 1 \le i \le n\}$ are the two n-pixel images to compare. Here, μ_x, μ_y, σ_x^2, σ_y^2 and σ_{xy} are the mean, variance, and covariance of $\mathbf{x}$ and $\mathbf{y}$, respectively (Sellahewa and Jassim, 2010), *i.e.*

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)^2, \qquad \sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\mu_y)^2$$

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)$$

The quantities $l(\mathbf{x},\mathbf{y})$ and $c(\mathbf{x},\mathbf{y})$ range between 0 and 1, while $s(\mathbf{x},\mathbf{y})$ is between -1 and 1. For normalized and equalized (NEQ) images, as in our case, $l(\mathbf{x},\mathbf{y})$ and $c(\mathbf{x},\mathbf{y})$ have a maximum of 1 since NEQ images have very similar standard deviation and mean values. In contrast, covariance values σ_{xy} are distinct in NEQ images. Hence, $s(\mathbf{x},\mathbf{y})$ (Eqn. 3) is a good candidate for assessing the similarity of two NEQ images. Note that $s(\mathbf{x},\mathbf{y})$ does not give a direct descriptive representation of the image structures: It reflects the similarity between two image structures, where $s(\mathbf{x},\mathbf{y})$ equals one if and only if the structures of the two images $\mathbf{x}$ and $\mathbf{y}$ are exactly the same.

We use the image quality index s for two purposes. First, given an input image $\mathbf{x}$, we compare it with the average image $\bar{\mathbf{x}}$ of our pre-computed image database (see Sec. 3.2), using $s(\mathbf{x},\bar{\mathbf{x}})$. If $\mathbf{x}$ is too far away from $\bar{\mathbf{x}}$, then $\mathbf{x}$ is either not a face image, or an image we cannot improve; and so, we stop our pipeline. This is further detailed in the quality index discussion in Sec. 4. Otherwise, we determine the so-called outlier artifacts which differentiate $\mathbf{x}$ from $\bar{\mathbf{x}}$, and suppress them, as shown next.

3.2 Statistical Artifact Segmentation

Once we have chosen to improve the quality of an input image, we aim to *segment* those image parts, or artifacts, which deviate significantly from typical average images. These will be the target of our inpainting (Sec. 3.3).

For artifact segmentation, we use statistical decision methods based on inference theory (Spiegel and Stephens, 2008; Bussab and Morrettin, 2002), where samples in a database can generate privileged information. This makes it possible to discriminate artifacts which we next want to remove from regular image pixels. Given a database of N of NEQ-normalized facial frontal images $\mathbf{x}_i$, the mean image is given by:

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x_i}. \tag{4}$$

We compute mean images $\bar{\mathbf{x}}^{DS}$, $\bar{\mathbf{x}}^{LS}$, and $\bar{\mathbf{x}}^{JP}$ separately for images of people of three ethnicities (dark skin (*DS*), light skin (*LS*), and japanese (*JP*)). This avoids mixing up too different facial traits. Figure 2 shows the mean and standard deviation images for these three groups. For dark skin people, we use the FERRET database, (65 images) (P.J. Phillips and Rauss, 2000) and 14 images taken from the FEI database (Thomaz and Giraldi, 2010). For light skin people, we consider 100 images from the FEI database. For japanese people, we use the JAFFE database (100 images) (M.J. Lyons and Gyoba, 1998). These databases contain a variety of images for people of different races, ages, and appearances. Although all images are equalized, facial structure differences exist, *e.g.* larger nose and lips (*DS* images), shadows around the eyes (*LS* images), and significant mouth position variations (*JP* images).

We next determine the ethnicity of the input image $\mathbf{x}$ by finding its closest mean image $\bar{\mathbf{x}}_{min} \in \{\bar{\mathbf{x}}^{DS}, \bar{\mathbf{x}}^{LS}, \bar{\mathbf{x}}^{JP}\}$, in terms of Euclidean distance.

Finally, we use statistical segmentation to find the artifact regions to be removed. To illustrate this, let us suppose that the face samples are represented by a random field $\mathbf{x}$ that follows a normal distribution with mean $\bar{\mathbf{x}}_{min}$ and standard deviation σ.

Thus the distribution of the standardized variable is given by:

$$\mathbf{z} = \frac{\mathbf{x} - \bar{\mathbf{x}}_{min}}{\sigma}. \tag{5}$$

Here, σ is the standardized normal distribution of $\mathbf{x}$ with mean 0 and variance 1, computed similariy to $\bar{\mathbf{x}}$, *i.e.* separately for the white skin, dark skin, and japanese face databases.

Table 1 shows the values of per-pixel z scores and significance levels α. The set of z scores outside the range $[-2.58, 2.58]$ is called the critical region of the hypothesis or region of significance. This is the region of rejection of the hypothesis. The set of z scores inside the range$[-2.58, 2.58]$ is called the region of hypothesis acceptance or the non-significance region.

Table 1: Statistical Significance

Level of significance α	Values of z
10%	- 1.645 .. 1.645
5%	-1.96 .. 1.96
1%	- 2.58 .. 2.58
0.1%	-3.291 .. 3.291

In practice, we observed that a value of $\alpha \simeq 10\%$ gives a good selection of atypical face features. In other words, for each pixel z of the image $\mathbf{z}$, computed by Eqn. (5), we decide if it is an artifact pixel or not, based on the value of z and our threshold α (Tab. 1). The set of artifact pixels Ω are removed as described next.

3.3 Semantic Inpainting

Now that we know which pixels of our input image are likely to be artifacts, we show a way to remove them by using inpainting. As our approach combines the fast marching based inpainting (Telea, 2004) and Poisson image editing (Pérez et al., 2003) techniques, we first briefly outline relevant aspects of these techniques.

3.3.1 Inpainting using Fast Marching

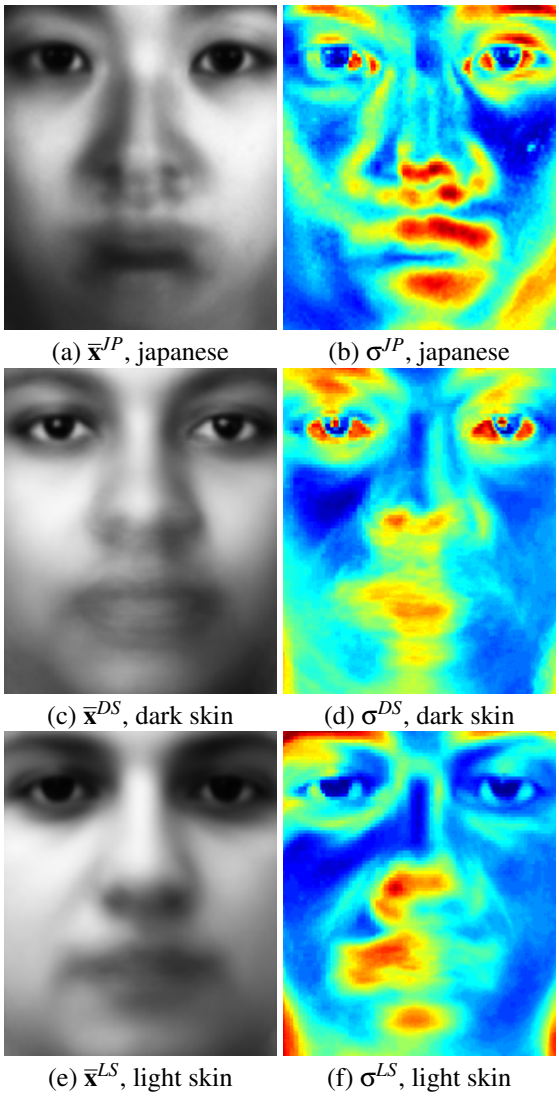
Inpainting technique based on fast marching method (FMM) considers a first order approximation $I_q(p)$ of the image at a point p situated on the boundary $\partial\Omega$ of the region to inpaint Ω

$$I_q(p) = I(q) + \nabla I(q) \cdot (p - q), \tag{6}$$

where q is a neighbor pixel of p located within a ball $B_\varepsilon(p)$ of small radius ε centered at p and $I(q)$ is the image value at q. Each point p is inpainted as function of all points $q \in B_\varepsilon(p)$, by summing the estimates of all points q, weighted by a weighting function $w(p,q)$:

$$I(p) = \frac{\sum_{q \in B_\varepsilon(p)} w(p,q) I_q(p)}{\sum_{q \in B_\varepsilon(p)} w(p,q)} \tag{7}$$

where the application-dependent weights $w(p,q)$ are normalized, *i.e.* $\sum_{q \in B} w(p,q) = 1$.



(a) $\bar{\mathbf{x}}^{JP}$, japanese (b) σ^{JP}, japanese

(c) $\bar{\mathbf{x}}^{DS}$, dark skin (d) σ^{DS}, dark skin

(e) $\bar{\mathbf{x}}^{LS}$, light skin (f) σ^{LS}, light skin

Figure 2: Mean images $\bar{\mathbf{x}}$ and color-coded standard deviation σ images for japanese, black skin, and light skin ethnicities.

The boundary $\partial\Omega$ is advanced towards the interior of Ω using the FMM (Sethian, 1999). While doing this, FMM implicitly computes the so-called distance transform $DT : \Omega \rightarrow \mathbb{R}^+$

$$DT(p \in \Omega) = \arg\min_{q \in \partial\Omega} \|p - q\|. \quad (8)$$

As the FMM algorithm visits all pixels of the damaged region the method gradually propagates gray-value information from outside Ω inwards by computing expression (7).

This inpainting method is fast and simple. However, for regions thicker than roughly 10..25 pixels, it creates an increased amount of blurring as we go farther from $\partial\Omega$ into Ω, given the smoothing nature of Eqn. 7. Since our facial artifacts are not guaranteed to be thin, this method is not optimal.

3.3.2 Poisson image editing

This image restoration method allows to achieve seamless filling of the damaged domain by using a source image data. It renders remarkably good results, even for thick target regions. The method is based on solving the equation:

$$\Delta I = div\ \mathbf{v} \quad \text{on } \Omega \quad (9)$$

with I known over the boundary $\partial\Omega$ and constrained to the values of $\mathbf{v}$, the so-called guidance field, that is derived from a source image. For instance, if we desires to seamlessly clone a source image I_{src} onto Ω, we can set $\mathbf{v} = \nabla I_{src}$, which effectively reduces Eqn. 9 to $\Delta I = \Delta I_{src}$ on Ω (Eqn. 10 in (Pérez et al., 2003)). Solving the Poisson problem (Eqn. 9), however, is expensive in terms of memory requirements, which is critical in our low-cost scenario (Farbman et al., 2009).

3.3.3 Proposed inpainting method

Both the FMM inpainting method and Poisson image editing basically use *only* the non-artifact areas of the input and near image to restore its artifacts.

However, we have additional information coming from our high-quality image database. Hence, our proposal is to combine the fast-and-simple inpainting method of (Telea, 2004) with the Poisson image editing (Pérez et al., 2003) augmented with information extracted from our image database (Fig. 3).

We start by applying FMM inpainting (Sec. 3.3.1) on the artifact region $\Omega \subset \mathbf{x}$ of our input image $\mathbf{x}$. We denote the result of this step, *i.e.* the image $\mathbf{x}$ where Ω has been inpainted, by I_{inp}. Next, we search in our face image database *DB* for the *closest* facial image I_{near}, *i.e.*

$$I_{near} = \arg\min_{\mathbf{y} \in DB} \|\mathbf{x} - \mathbf{y}\|. \quad (10)$$

As outlined in Sec. 3.3.1, when Ω is thick, FMM inpainting produces a blurred result I_{inp}. We correct this by mixing I_{inp} and I_{near} to yield:

$$I_2 = (1 - DT) I_{inp} + DT \cdot I_{near} \quad (11)$$

where DT is the distance transform of $\partial\Omega$ (Eqn. 8). This progressively mixes the inpainting result I_{inp} with the nearest image I_{near} in a progressive way. At the border $\partial\Omega$, we see the inpainted image, which smoothly extrapolates the non-artifact area out of the damaged region in the input image. In the middle of Ω the blurred effects observed in I_{inp} are attenuated by the high-quality image I_{near}.

Next, we calculate Laplacian $\Delta I_{near} = \frac{\partial^2 I_{near}}{\partial x^2} + \frac{\partial^2 I_{near}}{\partial y^2}$ of the closest image I_{near}, using central differences, and add this component to our restoration, *i.e.* compute

$$I_3 = I_2 + \Delta I_{near}. \quad (12)$$

This is a simplified form of the 'seamless cloning' presented in (Pérez et al., 2003) which just adds high-frequency features to I_2 through the Laplacian operation ΔI_{near}." In contrast to (Pérez et al., 2003), our approach is much cheaper, as it does not require solving a Poisson equation on Ω.

As a final step, we apply a 3-by-3 median filter on I_3 to yield the final reconstruction result. This further removes small-scale noise elements created by using the finite-difference Laplacian (Eqn. 12).

4 RESULTS

Our method can produce good artifact removal even in large and thick regions, due to the mix of inpainting (which preserves information specific to the input image) and prior information (which introduces information from a similar high-quality image from our image database). A key element is that our image database is composed by high quality images. Therefore, they are considered to be *good* for facial recognition purposes. Besides this information is suitable for both to extract the artifacts (Sec. 3.2) and to restore the corresponding damaged areas.

We tested our method on 79 frontal facial images provided by several public organizations: the Federal Brazilian Government (FGB, 2012), the Australian Federal Police (AFP, 2012), and the UK Missing People Organization (MPO, 2012). These images contain various levels of artifacts, *e.g.* annotations, scratches, glasses, obscuring facial hair, folds, and exposure problems. All input images were normalized and equalized as outlined in Sec. 3.

Segmentation: The Fig. 4 shows the results when

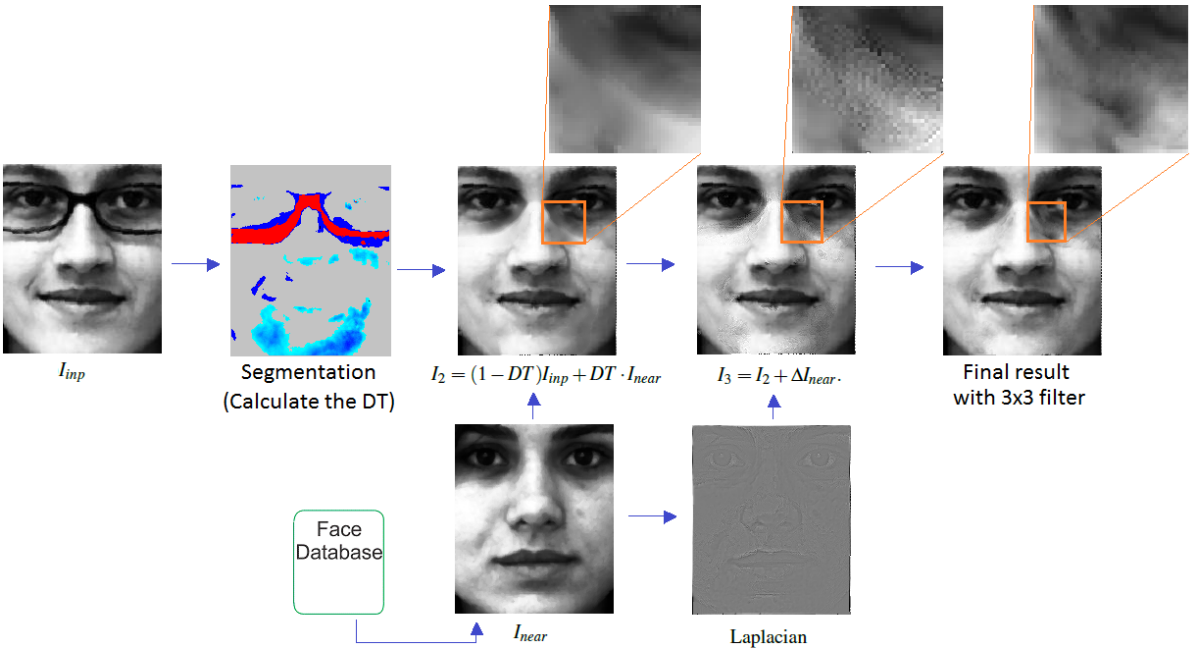


Figure 3: Diagram of the proposed inpainting method (see Sec. 3.3).

applying our segmentation approach for the images of collum (a) using the light-skin mean image (Fig. 4.(b)), dark-skin mean image (Fig. 4.(c)) and japanese mean image (Fig. 4.(d)). We observe that if significance level $\alpha \leq 1.0\%$ we can extract the artifacts (glasses) when using the light and dark-skin mean image images. Segmentation of artifacts works best and has similar perceived quality results for dark-skin and light-skin images (see Fig. 4). For several images, the mean of japanese people presents very good results. Although segmentation cannot select *all* details where an input image differs from its ethnic group's statistical average, it does a good job in selecting details deemed to be unsuitable for official person recognition photographs.

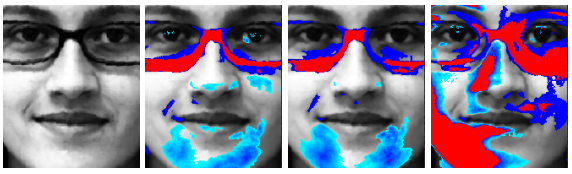
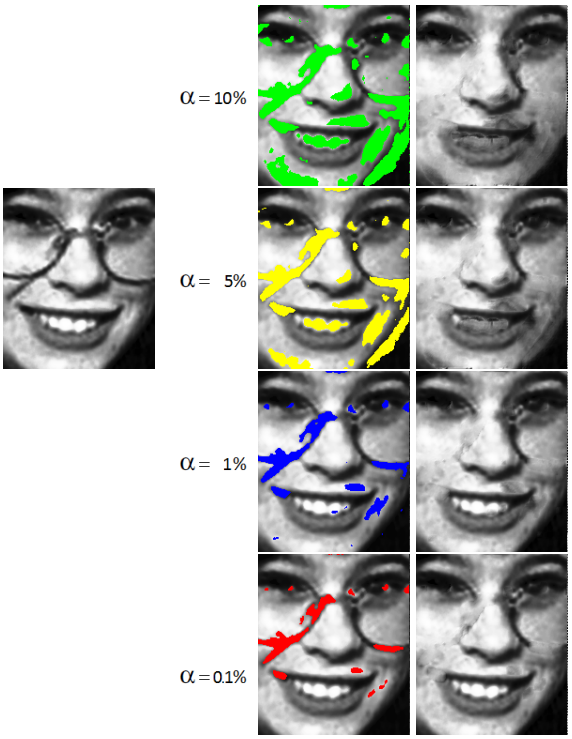


Figure 4: Segmentation comparison, with $\alpha = 0.1\%$ (red), $\alpha = 1\%$ (dark blue), and $\alpha = 10\%$ (light blue). (a) original image; (b) light skin mean image $\bar{\mathbf{x}}^{LS}$; (c) dark skin mean image $\bar{\mathbf{x}}^{DS}$; (d) japanese mean image $\bar{\mathbf{x}}^{JP}$

Significance level: Figure 5 shows results when changing the significance level α (Sec. 3.2). If we choose high values for α the method extracts the artifacts as well as other regions, as we can see on Figure 5 . Conversely, the choice of low α values, like in Figure 5, prevent such problem but is not efficient to segment the region of interest. The optimum value for α is application dependent and it depends on manual fine-tuning in general. For the test performed, the value $\alpha = 10\%$ has given good results in all cases. This fact can be verified in Figure 6 which shows the results using some values of α.

Artifacts: *Smile* is considered an artifact, as it does not comply with official standards for facial images concerning expression, illumination, and background (ANSI, 2004; ISO, 2004; Thakare and Thakare, 2012). Figure 7 presents an example of smile suppression. The bottom row shows a case where the smile has been well segmented and suppressed. The top row shows a less successful case: Part of the large smile (see Fig. 7 a, top) has not been captured by the segmentation, and as such, it leaked in the final reconstruction (Fig. 7 d, top). Also, note that our method can produce changes in facial expressions. For example, Fig. 7 (bottom row, d) shows a face where the eyes are a mix between the input image and nearest image, whereas the nose, and cheeks follow more closely the input image.

Figure 8 shows further results, and also compares them with three known inpainting techniques. In the top row, we see an image with a hair lock artifact. Segmenting this artifact is easy, as it is much darker than the mean light-skin image $\bar{\mathbf{x}}^{LS}$. Inpainting this artifact is also relatively easy, as it is not thick. The second row shows a face image which has a highlight, as



(a) (b) (c)

Figure 5: Segmentation results: (a) original image; (b) segmentation; (c) reconstruction.

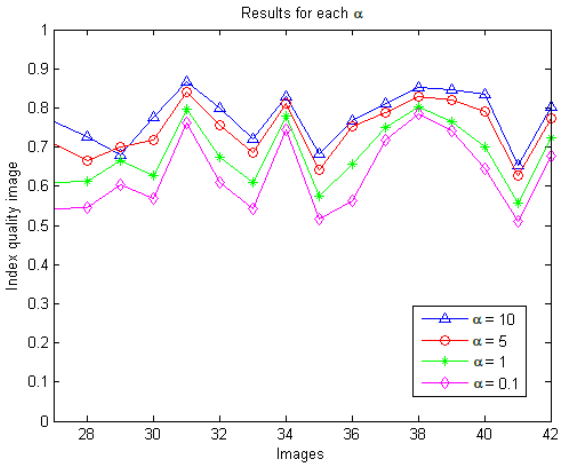


Figure 6: Results at each value of α.

well as a thin horizontal crease (most probably due to a photograph damage) on the cheeks. This artifact is more complicated to capture since its grayvalue field is similar with the mean image pixels nearby. However, the result (column *g*) shows that this artifact is also largely eliminated from the input image.

The third row of Figure 8 shows an image with glasses as an outlier artifact. Existing inpainting

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

Figure 7: Smile elimination: (a) input image; (b) nearest image; (c) segmentation; (d) proposed inpainting.

methods (columns d-f) succeed in eliminating this artifact less than our method (column g), because the eyes are reconstructed. The fifth row of Figure 8 shows a similar case. Here, the artifact region is thin, so all four tested inpainting methods produce eliminate the glasses equally well.

In the fourth row Figure 8 it is shown an image where a large shadow artifact appears under the nose, on the cheeks, and the mouth. As expected, standard inpainting cannot easily remove this problem, since the shadow slightly extends outside of the segmented region, on the base of the nose (dark area under the nose, Fig. 8 c, fourth row). By using the nearest image our method can be more efficient to remove the shadow while generating less spurious details. A similar effect is observed in Figure 8 (bottom row), where our method performs better for both to remove the cheek highlights and to reconstruct the opened eyes better than existing methods.

Quality index: Figure 9 shows the average structural image quality index (Equation 3) for the three ethnicities present in our face image database, the original input images for inpainting, and the results of the studied inpainting methods. Several observations follow. First, as expected, the face database images have a relatively high quality index (average $s \in [0.6, 0.8]$). Secondly, input images which are faces have a lower quality (average $s \sim 0.49$). This motivates our proposal to adjust such images to bring them in line with the quality level of the face database. In contrast, input images which are not faces have a much lower quality index (average $s \sim 0$). This allows us to decide whether an input image should be improved or deemed not improvable (because it is not a face): We choose a suitable threshold τ, in this case, $\tau = 0.1$. Input images with $s > \tau$ are likely faces, so they are further improved; images with $s < \tau$ are likely non-faces, and thus skipped from the process. Thirdly, we notice that the nearest image I_{near} used

(a) (b) (c) (d) (e) (f) (g)

Figure 8: Method results: (a) original image; (b) nearest image (not the same person, just similar); (c) segmentation, where green color indicates significance level $\alpha = 10\%$, yellow is 5%, blue is 1% and red is 0.1%; (d) inpainting (Oliveira et al., 2001); (e) inpainting (Bertalmio et al., 2001); (f) inpainting (Telea, 2004); (g) our method.

in our inpainting (Sec. 3.3.3) has an average high quality index, which justifies its explicit usage during inpainting. Finally, we see that our method yields, on average, results with a higher structural quality than the others studied inpainting methods.

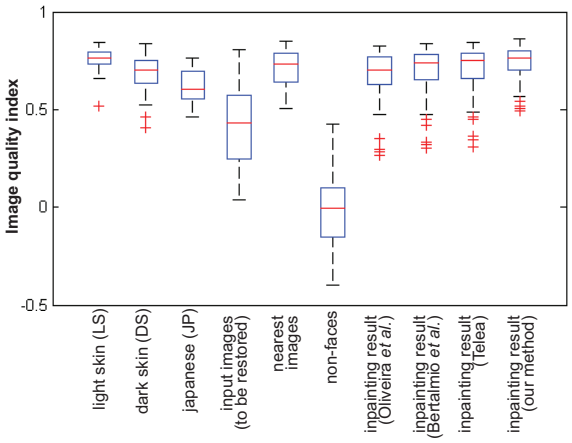


Figure 9: Structural image quality for the input images, face database, and result images.

Limitations: Despite of its capabilities, our method cannot handle facial images which deviate too much from the information provided by the predefined image database. Figure 10 shows such an example, which is the worst case we found in our tests: The input image (a) is too far away from *both* the average black-skin image (Fig. 2 c) and the nearest image (Fig. 10 b) to successfully remove the facial hair, nose pin, and large opaque glasses. However, we note that this type of outlier can be easily detectable by our approach: When the input image is too far away from the mean image, as mentioned above, our method reports that it cannot likely improve this image and stops.

Database: Selecting images that has good quality for the face database serves two purposes. First, this allows specifying what one considers to be an acceptable facial image in a given context (*e.g.* open eyes, no smiles, shadows, adorns, or imprints). Statistical characteristics of this collection are used to implicitly determine what are outliers, thus what has to be suppressed in a given input image. Secondly, features of the nearest image in this database are used in the restoration. Hence, if one wants to allow certain specific feature to persist in the restored images, this can be done by inserting images containing such features in the database.

Computational aspects: The proposed method is simple to implement. For an image of n pixels, its memory and computational complexities are $O(n\ log\sqrt{n})$ (Sethian, 1999) respectively, in contrast to more sophisticated inpainting algorithms (Bertalmio et al., 2001; Chan and Shen, 2000a; Chan and Shen, 2000b; Pérez et al., 2003; Joshi et al., 2010). This makes our method suitable for computers that have low capacity of memory.

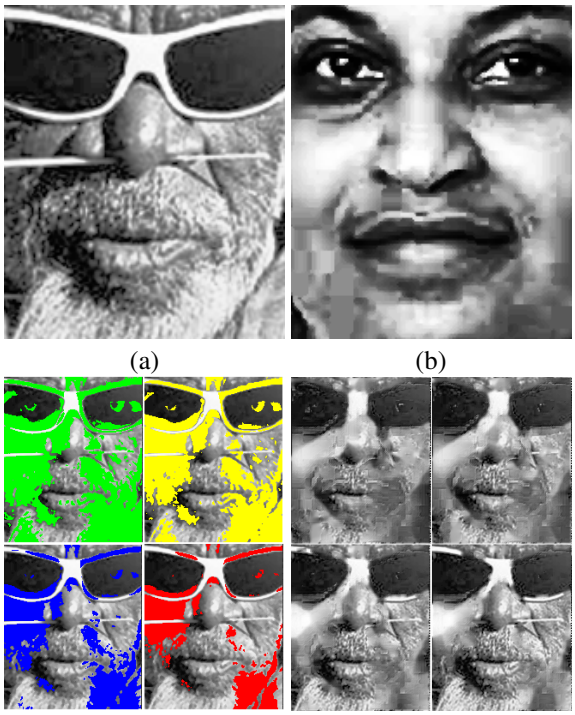


Figure 10: Challenging case for our method: (a) input image; (b) nearest image; (c) segmentation; (d) inpainting.

Validation: Although we have not validated the added value of our artifact suppresion method in a full face recognition pipeline, feedback from São Paulo law enforcement confirmed that artifact removal is indeed helpful in manual recognition tasks. Future work should be done to quantify the improvement in automatic face recognition tasks.

5 CONCLUSIONS

We have presented a computational framework for automatic inpainting of facial images. Given a database of facial images which are deemed to be of good quality for recognition tasks, our method automatically identifies outlier (artifact) regions, and reconstructs these by using a mix of information present in the input image and information from the provided image database. The proposed method is simple to implement and has low computational requirements, which makes it attractive for low-cost usage contexts

such as government agencies in least developed countries. We have tested our method using three real-world public image databases of missing people, and compared our restoration results with three popular methods used in image inpainting.

Potential improvements lie in the areas of more robust segmentation using artifact-specific quality metrics and using the k nearest images ($k \geq 1$) for inpainting and actual evaluation of inpainting in a real-world face recognition set-up.

REFERENCES

AFP (2012). Australian Federal Police, National Missing Persons Coordination Centre. http://www.missingpersons.gov.au.

Amaral, V. and Thomaz, C. (2008). Normalização espacial de imagens frontais de face. Dept. of Electrical Engineering, Univ. Center of FEI, Brazil. http://fei.edu.br/ cet/relatorio_tecnico_012008.pdf.

ANSI (2004). US standard for digital image formats for use with the facial biometric (INCITS 385).

Aptoula, E. and Lefèvre, S. (2008). A comparative study on multivariate mathematical morphology. *Pattern Recogn. Lett.*, 29(2):109–118.

Ayinde, O. and Yang, Y. (2002). Face recognition approach based on rank correlation of Gabor-filtered images. *Pattern Recogn.*, 35(6):1275–1289.

Bertalmio, M., Sapiro, G., and Bertozzi, A. (2001). Navier-stokes, fluid dynamics, and image and video inpainting. In *Proc. CVPR*, pages 355–362.

Bertalmio, M., Sapiro, G., Caselles, V., and Ballester, C. (2000). Image inpainting. In *Proc. ACM SIGGRAPH*, pages 417—424.

Bugeau, A. and Bertalmio, M. (2009). Combining texture synthesis and diffusion for image inpainting. In *Proc. VISAPP*, pages 26–33.

Bugeau, A., Bertalmio, M., Caselles, V., and Sapiro, G. (2009). A unifying framework for image inpainting. Technical report, Intitute for Math. and Applications (IMA). http://www.ima.umn.edu.

Bussab, W. and Morrettin, P. (2002). *Estátistica Básica*. Editora Saraiva, São Paulo, Brazil.

Castillo, O. (2006). Survey about facial image quality. Technical report, Fraunhofer IGD, TU Darmstadt.

Chan, T. and Shen, J. (2000a). Mathematical model for local deterministic inpaintings. In *Tech. Report CAM 00-11*. Image Processing Group, UCLA.

Chan, T. and Shen, J. (2000b). Non-texture inpainting by curvature driven diffusions. In *Tech. Report CAM 00-35*. Image Processing Group, UCLA.

Chou, J., Yang, C., and Gong, S. (2012). Face-off: Automatic alteration of facial features. *Multimedia Tools and Applications*, 56(3):569–596.

Farbman, Z., Hoffer, G., Lipman, Y., CohenOr, D., and Lischinski, D. (2009). Coordinates for instant image cloning. In *Proc. ACM SIGGRAPH*, pages 342–450.

FGB (2012). Federal Goverment of Brazil, Justice Ministry - Missing Persons. http://www.desaparecidos.mj.gov.br.

Hjelmas, E. and Low, B. (2001). Face detection: A survey. *CVIU*, 83(3):236–274.

ISO (2004). Final draft international standard of biometric data interchange formats (part 5, face image data, ISO–19794-5 FDIS).

Jeschke, S., Cline, D., and Wonka, P. (2009). A GPU laplacian solver for diffusion curves and poisson image editing. In *Proc. ACM SIGGRAPH Asia*, pages 1–8.

Joshi, N., Matusik, W., Adelson, E., Csail, M., and Kriegman, D. (2010). Personal photo enhancement using example images. In *Proc. ACM SIGGRAPH*, pages 1–15.

Li, H., Wang, S., Zhang, W., and Wu, M. (2010). Image inpainting based on scene transform and color transfer. *Pattern. Recogn. Lett.*, 31(7):582–592.

M.J. Lyons, S. Akamatsu, M. K. and Gyoba, J. (1998). Coding facial expressions with gabor wavelets. In *Proc. FG*, pages 200–204. IEEE.

MPO (2012). Missing People from United Kingdom. http://www.missingpeople.org.uk.

Oliveira, M. M., Bowen, B., Mckenna, R., and Chang, Y. (2001). Fast digital image inpainting. In *Proc. VIIP*, pages 261–266.

Pérez, P., Gangnet, M., and Blake, A. (2003). Poisson image editing. In *Proc. ACM SIGGRAPH*, pages 313–318.

P.J. Phillips, M. Hyeonjoon, S. R. and Rauss, P. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE TPAMI*, 22(10):1090–1104.

Sellahewa, H. and Jassim, S. (2010). Image-quality-based adaptive face recognition. *IEEE TIM*, 59(4):805–813.

Sethian, J. (1999). *Level Set Methods and Fast Marching Methods*. Cambridge Univ. Press, 2nd edition.

Spiegel, M. and Stephens, L. (2008). *Statistics*. Schaum's Outlines.

Telea, A. (2004). An image inpainting technique based on the fast marching method. *J. Graphics. Tools*, 9(1):23–34.

Thakare, N. and Thakare, V. (2012). Biometrics standards and face image format for data interchange - a review. *Int. J. of Advances in Engineering and Technology*, 2(1):385–392.

Thomaz, C. and Giraldi, G. (2010). A new ranking method for principal components analysis and its application to face image analysis. *Image Vision Comput.*, 28(6):902–913.

Wang, Z., Lu, L., and Bovik, A. (2004). Video quality assessment based on structural distortion measurement. *Sig. Proc.: Image Comm.*, 19(2):121–132.

Zamani, A., Awang, M., Omar, N., and Nazeer, S. (2008). Image quality assessments and restoration for face detection and recognition system images. In *Proc. AMS*, pages 505–510.

Zhao, W., Chellappa, R., Phillips, P., and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458.