Implementation of Missing Data Imputation Schemes in Face Recognition Algorithm under Partial Occlusion

Justice Kwame Appati, Kofi Sarpong Adu-Manu, and Ebenezer Owusu

Department of Computer Science, University of Ghana, Legon, Accra, Ghana

Correspondence should be addressed to Ebenezer Owusu; ebeowusu@ug.edu.gh

Received 28 November 2021; Accepted 18 May 2022; Published 15 June 2022

Academic Editor: Rajesh Kaluri

Copyright © 2022 Justice Kwame Appati et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Face detection and recognition algorithms usually assume an image captured from a controlled environment. However, this is not always the case, especially in crowd control under surveillance or footage from a crime scene, where partial occlusions are unavoidable. Unfortunately, these occlusions have an adverse effect on the performance of these classical recognition algorithms. In this study, the performance of some selected data imputation schemes is evaluated on SVD/PCA frontal face recognition algorithm. The experiment was done on two datasets: Jaffe and MIT-CBCL, with immediate confirmation of the adverse effect of occlusion on the facial algorithm without implementing the imputation scheme. Further experimentation shows that IA is an ideal missing data imputation scheme that works best with the SVD/PCA facial recognition algorithm.

1. Introduction

In recent years, the research community has witnessed a significant breakthrough theoretically and in practice in using face recognition technologies [1,2]. The human face in the past has proven helpful in the development of many applications such as gesture and gender description [3], facial identification [4,5], age estimation and classification [6,7], and image forecasting and restoration [8]. Other social applications for monitoring, access control, banking systems, and forensic audit are made possible with face recognition though they pose peculiar challenges. Challenges such as scale, background, illumination conditions, occlusion, expression, and pose have been addressed to some extent by various proposed facial detection and recognition algorithms [9]. Despite the achievements of these proposed methods and systems, their level of accuracy does not match that of the human vision systems, which defines a gap of interest to the research community [10]. By augmenting the strength of methods such as local texture-based face representations, kernel-based feature extraction, robust illumination normalisation, multiple feature fusion, and distance transform based matching, the issue of illumination variation is mitigated [11]. Unfortunately, the adoption of the fusion strategy in [11] only helped resolve the issue of illumination while leaving the rest of the challenges mentioned earlier unattended. In the study of [12], shearlets’ robust features and edge detection capabilities were leveraged and further augmented to Local Binary Pattern (LBP) to handle the presence of heavy noise encountered by the face recognition system. From the study of [12], each face data is first divided into microblocks, with each block classified independently. The resultant contribution from each classified block is put together for improved performance. Next is the occlusion issue, which the study [13] attempts to address by proposing a Robust Principal Component Analysis (RPCA). Their study decomposed all training samples to obtain a low-rank and sparse content matrix where noise in the sparse content matrix was suppressed. However, to further increase the interclass information between the low-ranked matrices, the study of [14] proposed the need to decompose the training samples into a low-ranked matrix without occlusion and a sparse error matrix. The RPCA proposed by [13] was adopted by [14] in the low-ranked matrix. A subspace was obtained as an occlusion dictionary for face recognition in the process. The
study [15] proposed a framework for subspace learning by mapping both the train and test images onto a gradient face space. In effect, a robust face feature without occlusion was extracted to a large extent resulting in a reconstructed image with near-zero occlusion. The robustness of the proposed method is conditioned on the premise that the difference in two regions of occlusion of two completely different images approximately obeys the uniform distribution. Unfortunately, the uniform distribution assumption is not the case in practice. This observation makes the method unsuitable for recognising an arbitrarily occluded face image. However, in circumstances where facial recognition is to be carried out under a variational environment, the adoption of the fisherface becomes a valuable technique [16]. The technique leverages the Linear Discriminant Analysis (LDA) to optimally use the large interperson and small intraperson alteration to build a subspace like Principal Component Analysis (PCA). One major disadvantage of the technique is the implementation of the Euclidean in its data space resulting in its inability to scale well in a multimodally distributed face image set where data points are located in the nonlinear subspace [17]. Last but not least is the study by [18,19], where the Lambert reflecting function was proposed to construct Lambertian objects under varying illumination conditions with only nine (9) spherical harmonics. Finally is the work of [20] which introduced the concept of bilinear generative models to decompose orthogonal factors and further prove that a separable bilinear mapping exists between an input space and the lower-dimensional subspace. From this observation, the pose information and identity can be extracted explicitly, given that the parameters of the mappings can be determined accurately.

This study was motivated by the fact that partial occlusion remains an industrial challenge as research strives to propose an appropriate technique to augment the efforts of classical facial recognition algorithms that are well known to work best in a controlled environment. The inability of these classical facial recognition algorithms to function well under occlusion is the reason this study seeks to address. Our study used three imputation schemes as an image data recovery tool under an occluded environment to address that. The resultant image was further fed into the classical SVD/PCA facial recognition algorithm. Finally, the imputation schemes were evaluated, and the most robust scheme, which was found to be IA, was adopted as the best fit for the recognition system.

2. Material and Methods

2.1. Source of Data. This study used the Japanese Female Facial Expressions (Jaffe) [21] and the Massachusetts Institute of Technology (MIT-CBCL) [22] frontal face image dataset to evaluate the efficiency of data imputation schemes of the PCA/SVD face recognition algorithm. In the study, we used all the faces in Jaffe, which had 213 observations from ten unique individuals. We also used all the images in the MIT-CBCL-Synthetic folder and all the images in the MIT-CBCL-Test folder. In the MIT-CBCL-Synthetic folder, there are 3240 observations from ten unique individuals, while the MIT-CBCL-Test folder had 2000 observations from ten unique individuals. All 5453 face images were used for this study as training data. Samples of the Jaffe and MIT-CBCL dataset for the training phase are shown in Figure 1 and Figure 2, respectively.

The test dataset for this study was curated as an extension of the three data sources used for the training. The test data curation is as follows: First, a zero-patch of dimension $M \times N$ is defined reasonably enough to cover major geometric face features like the nose, cheek, mouth, eye, and eyebrow. Next, a percentage of this patch defined by the user is set to NaN at random, where the NaN represents the missing values. Finally, the patch with NaN is superimposed on the training dataset to obtain the “new” test dataset for this study. Samples of the test face as generated from Jaffe with a percentage of 5%–30% at an increment of 5% are shown in Figure 3. None of these images from the train or test dataset was resized in this study.

2.2. Data Preprocessing with Imputation Schemes. From Section 2.1, it is observed that the test dataset curated is incomplete and contains invalid data to establish a real case scenario. In practice, these incomplete data are made complete for further analysis to be carried out. For the purpose of this study, let T, bold lower-case character (e.g., x), and bold upper-case character (e.g., X) denote the transpose of a mathematical structure of a column vector and a data matrix, respectively. For any given data matrix, we shall let N and K denote the number of observations (thus, number of rows) and the number of variables (thus, number of columns), respectively. Finally, the jth observation of the data matrix will be denoted as $x_j = [x_{j1}, \ldots, x_{jK}]$ where j is the jth variable. Since the original data matrix comes with partial occlusion representing missing values, and with the assumption that the position of the observations is commutative, $x_j$ can be written as $x_j^T = [x_{j1}^T, \ldots, x_{jK}^T]$ where $x_{jK}^T$ is the variable with a missing value. At the same time, $x_{j}^T$ is the variable with available measurements. This further implies that the data matrix X can be represented as $X = [X^\#, X^*]$, and its centred covariance S as $S = X^TX/N - I$. Similarly, we can compute $X^*$ and $X^\#$’s centred covariance as $S^{**}$ and $S^{##}$. Suppose M is the missing data in X defined by equation (1) and $\odot$ the complement of M. In that case, the resultant matrix Z, after filling all missing values with zero, is defined as $Z = M \odot X$, where $\odot$ is the Hadamard elementwise product operator.

$$M = \begin{cases} 
1, & \text{x}_{ij} \text{ is missing}, \\
0, & \text{otherwise}.
\end{cases}$$

From these definitions and preambles, the following three subsections briefly summarise the imputation schemes used in this study. All these three methods presented are statistical methods for data correction.

2.2.1. Projection to the Model Plane (PMP). The “projection to the model plane” method at its inception was meant for Principal Component Analysis Model Exploitation (PCA-
ME) problem [23]. In such a problem, the PCA-ME assumes that complete data have been fitted with the PCA model, and all that is left to be done is the analysis of new observations having missing values. With this context in mind, the PMP method was jointly adapted with the regression-based methods as detailed in [24]. However, it defers slightly in how the imputation step is carried out. At the imputation phase, the PMP imputes at time $t$ the missing values using only the loadings matrix $P$ as defined by equation (2) instead of the covariance matrices.

$$y_{i,t}^* = p_i^T (p_i^{*T} p_i^{*})^{-1} p_i^{*T} y_{i,t}^{**}. \quad (2)$$

2.2.2. Nonlinear Iterative Partial Least Squares Regression (NIPALS). NIPALS is an iterative method that performs iterative regressions using only available data while ignoring the missing values. The modified version of the method adapts the NIPALS algorithm, and more details can be found in [25].

2.2.3. Data Augmentation (IA). Finally is data augmentation, an iterative method that imputes missing values based on predictions from PCA models. This method is also similar to the regression-based model. However, its imputation phase is substituted with an estimation from the loadings $P$ and scores of the PCA model as defined by equation

$$X_{i+1} = Z + M * (T_i P_i^T + 1_N m_i^T). \quad (3)$$

2.3. SVD Feature Extraction and PCA Reduction. The Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) algorithm are adopted in this study over other methods such as Linear Discriminant Analysis (LDA) since it is best suited to distinguish between different people. In contrast, methods like LDA are the best for a task like class separation. The goal of PCA is to reduce a complex dataset in dimension to a lower dimension by getting rid of extreme values that may be caused by noise, rotation, and
redundancy. Mathematically, the algorithm works as follows:

First, the image $X_i$ is vectorised and augmented columnwise in the form $X = [X^1 | X^2 | \cdots | X^N]$. The resultant augmented matrix is then averaged using equation

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X^i,$$  \hspace{1cm} (4)

where $N$ is the number of images. Next is the computation of the covariance matrix $C$ of the vectorised image matrix using equation

$$C = A^T A,$$ \hspace{1cm} (5)

where $A = (\xi^1, \xi^2, \ldots, \xi^N)$. The term $\xi^i$, which defines the mean centering of the $i$th image, is computed by subtracting the averaged image in equation (4) from the image vector $X^i$ as given in equation

$$\xi^i = X^i - \bar{X}.$$ \hspace{1cm} (6)

Given equation (5), the eigenface is constructed by first decomposing the covariance matrix $C$ using SVD. In the decomposition process, three different matrices $U$, $V$, and $\Xi$ are obtained where $U$ and $V$ are orthogonal matrices, and $\Xi$ is a diagonal matrix. Summing across the image the product of the $i$th column ($u^i$) of the orthogonal matrix $U$ and the $i$th mean centering image ($\xi^i$) in equation (6), we have the eigenface ($\eta$) defined as equation

$$\eta^i = \sum_{i=1}^{N} \xi^i u^i.$$ \hspace{1cm} (7)

Ordinarily, it is expected that the features extracted with SVD will have some redundant attributes that require the application of PCA based on the quality of the eigenvalue extracted from equation (8) where $\lambda^i$ is the eigenvalue and $\Gamma^i$ is the corresponding eigenvector.

$$C \Gamma^i = \lambda^i \Gamma^i.$$ \hspace{1cm} (8)

2.4. Research Implementation. This section of the study talks about the research design and how the design was implemented to achieve the results reported in Section 3. The computational environment for the study and the performance metric used to evaluate the $L_p$-Norm classifier is also discussed.

2.4.1. Research Design. The research design to be implemented in this study is shown in Figure 4. From the figure, we have four main blocks, with the first block being the preprocessing phase, followed by the Feature Extraction/Reduction Phase, Knowledge Curation, and finally, Performance Evaluation. Details of its implementation are discussed in Section 2.4.2.
2.4.2. Design Implementation and Experimentation. Test data is curated from the research design in Figure 4, as explained in Section 2.1. The test data in its corrupt form is filtered using some imputation schemes discussed in Section 2.2. Features are then extracted from the training dataset using SVD, while the relevant attributes are selected using PCA (see Section 2.3 for details). The cleaned test image is projected onto the extracted feature map to extract the required relevant features of the test image. The filtered features from the training dataset are stored as distilled knowledge. In contrast, the filtered features from the clean test data are evaluated against the distilled knowledge using the Lp-Norm (See Section 2.5) as a classifier. The matched test data are evaluated against the distilled knowledge using equation (12).

\[
\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\mu_x^2 + \mu_y^2 + c_1}\left(\frac{\sigma_x^2 + \sigma_y^2 + c_2}{\sigma_x + \sigma_y + c_1}\right)
\]

(9)

\[
\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=1}^{n-1} [I(i, j) - K(i, j)]^2,
\]

(10)

\[
\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I}{\text{MSE}}\right) - 10 \cdot \log_{10} (\text{MSE}),
\]

(11)

\[
\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}
\]

(12)

From equation (9), \(\mu_x\) and \(\mu_y\) are the means of \(x\) and \(y\), respectively. Similarly, \(\sigma_x\) and \(\sigma_y\) are the variances of \(x\) and \(y\), respectively. \(\sigma_{xy}\) is the covariance of \(x\) and \(y\), while \(c_1\) and \(c_2\) are two well-defined variables to help stabilise the division with a weak denominator. Also, in equation (12), \(TP\), \(TN\), \(FP\), and \(FN\) denote True Positive, True Negative, False Positive, and False Negative, respectively.

3. Results and Discussion

3.1. Experiment 1: Distortion Analysis of Imputation Schemes. In this setup, the objective is to assess the quality of restored images after applying the three imputation schemes. As shown in Table 1, it is evident that the data augmentation scheme is superior to PMP and NIPALS scheme across all the three datasets used and the three quality assessment metrics. The PMP scheme was also considered superior to NIPALS; however, further experimentation is required for a firm and a more valid decision on what scheme is appropriate.

2.5. Performance Evaluation. In evaluating the performance of the data imputation methods, four (4) metrics were used. These metrics are Structural Similarity Index Measure (SSIM), Peak Signal to Noise Ratio (PSNR), Mean Square Error (MSE), and accuracy (Acc) based on L2-Norm. The SSIM and PSNR were used for quality assessment of the recovered image, while the MSE was used to estimate the error incurred in recovering the image. Finally, accuracy was used to judge the performance of the face recognition model with respect to the occluded and recovered image. Mathematically, these metrics are defined in equation (9) to equation (12).

3.2. Experiment 2: Performance of Model without Imputation. This section further seeks to assess the performance of the SVD/PCA on occluded frontal face images before imputation. The level of occlusion for the model assessment was set to 5%
with an increment of 5% till 30% occlusion was obtained. From Table 2, it can be observed that the model adopted is more robust to 5% occlusion in all the three datasets selected. One interesting observation which required further investigation is the robustness of the recognition model at a varying distortion on the MIT-CBCL-Test dataset. This dataset recorded a perfect score of 100%. On the other hand, an apparent observation was noted for Jaffe and MIT-CBCL-Synthetic as the distortion rate increases; however, the reduction in performance is quite faster in the Jaffe dataset than that of the MIT-CBCL-Synthetic dataset. With only Experiment 2, it is fair to conclude that the model’s performance under occlusion is reasonably good. However, one must establish how factual this conclusion will be when the model is further “stressed” with a higher level of occlusion. Before then, it will be prudent at this point for one to establish if imputation schemes, as shown in Experiment 3, can help eliminate or, better still, reduce the effect of occlusion in the model.

3.3. Experiment 3: Effect of Imputation Schemes on Performance. In this section, Experiment 2 is repeated, taking into account the three imputation schemes discussed earlier. From Table 3, it is observed that the recognition model attained a perfect score of 100%, irrespective of the imputation methods used. With this observation, one may conclude that a “careful” selection of an imputation scheme is irrelevant as the ultimate goal is hinged on the ability of the recognition model to work efficiently well on the recovered image. Unfortunately, this study believes that this obvious conclusion might be too early to make, hence the need to test for the worst-case scenario. In this study, the worst-case scenario is associated with the instance where the cut-off region is wholly occluded, which is 100% distortion, hence the need for Experiment 4.

3.4. Experiment 4: Effect of Imputation Schemes on Performance at 100% Distortion. As the distortion of the facial image gets to 100%, the recognition algorithm turns out to be deficient when no imputation scheme is used. However, an accuracy of 92.25% was noted for the MIT-CBCL-Test dataset. Despite this observation, it is valid to conclude that the algorithm will generally fail at a higher percentage of occlusion. A sample of 100% partial occlusion from Jaffe, MIT-Synthetic, and MIT-Test is shown in Figure 5. In applying the imputation schemes prior to recognition, some significant increase in accuracy was achieved with a valid case for IA. PMP, on the contrary, worsen the recognition rate, as shown in Table 4. In this experiment, it is also prudent to conclude that all the imputation schemes do not work the same way as confirmed in Table 1. IA will be the best fit for the SVD/PCA facial recognition algorithm from this experiment. Another important observation that is a contradiction of thought in Experiment 1 is the NIPALS’ relatively poor quality test performance compared to PMP in all three datasets. Subject to that observation, one will expect a better performance in PMP over NIPALS, which is not the case in Experiment 5. With this said and the inconsistency between the performance of PMP and NIPALS, the study further proceeds to validate the complexity of IA against PMP, NIPALS, and other standard imputation schemes such as PCR, PLS, and TSR in Experiment 5.

3.5. Experiment 5: Average Unit Time Complexity of Imputation Schemes. From Experiment 1, it is evident that the area of an affected image characterised by a missing value measured in percentage directly affects the performance of the PCA/SVD facial algorithm. Resolving these missing values with various imputation schemes irrespective of the “marginal” distortion level per this study yielded a 100% accuracy indicating a total restoration of the corrupted image. However, in increasing the distortion rate to 100% on the affected reason, a different observation was noted in Experiment 3. This observation suggests another level of investigation of the imputation schemes employed in the study. In this section, we run each scheme on a single image taken from the Jaffe dataset (due to its inherent characteristics and interactions with the imputation scheme) at a varying percentage of missing values (i.e., 5%–30% at an increment of 5%). The CPU clock time was used as a metric for evaluation, and since the clock time can be affected negatively by CPU overhead, ten cycles were carried out per scheme to curb the effect. The average of the ten cycles is reported in Table 5 with a graphical presentation in Figure 6. From Figure 6, it is seen that though all the schemes could sufficiently impute missing values resulting in a perfect score, they do so at the expense of the computer resources. PCR recorded the highest cost, followed by PLS, NIPALS, TSR, PMP, and IA. It is also observed from Figure 6 that an increase in distortion rate leads to an increase in the cost of imputation, implying that one needs to be mindful in the

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>SSIM</th>
<th>PSNR</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaffe</td>
<td>PMP</td>
<td>90.88</td>
<td>26.88</td>
<td>133.46</td>
</tr>
<tr>
<td>MIT-CBCL-Synthetic</td>
<td>PMP</td>
<td>89.74</td>
<td>31.70</td>
<td>43.97</td>
</tr>
<tr>
<td>MIT-CBCL-Test</td>
<td>PMP</td>
<td>92.81</td>
<td>30.12</td>
<td>63.24</td>
</tr>
<tr>
<td>Jaffe</td>
<td>NIPALS</td>
<td>90.47</td>
<td>12.77</td>
<td>3440.0</td>
</tr>
<tr>
<td>MIT-CBCL-Synthetic</td>
<td>NIPALS</td>
<td>82.21</td>
<td>17.80</td>
<td>1078.2</td>
</tr>
<tr>
<td>MIT-CBCL-Test</td>
<td>NIPALS</td>
<td>90.73</td>
<td>19.81</td>
<td>679.54</td>
</tr>
<tr>
<td>Jaffe</td>
<td>IA</td>
<td>99.00</td>
<td>56.54</td>
<td>0.1443</td>
</tr>
<tr>
<td>MIT-CBCL-Synthetic</td>
<td>IA</td>
<td>99.68</td>
<td>55.64</td>
<td>0.1775</td>
</tr>
<tr>
<td>MIT-CBCL-Test</td>
<td>IA</td>
<td>99.85</td>
<td>58.34</td>
<td>0.0954</td>
</tr>
</tbody>
</table>
selection of an appropriate imputation scheme given a resource-constrained environment. For the purpose of this study, IA proved to be more efficient across all the four (4) experiments carried out in handling missing values and in a resource-constrained environment. It should be noted that Table 5 is added to this study to guide readers to reproduce the graph for further analysis.

3.6. Observation of Image Smoothing Operators. In this study, it is being observed that several works that perform facial recognition using PCA/SVD algorithms mostly perform image denoising prior to feature extraction operations. Assumptions in such studies are that a Gaussian distributed noise characterises the effects of the illumination on natural images or scenes. With this assumption made, Gaussian
filtering is carried out in the frequency domain using transforms such as Fast Fourier Transform (FFT) and Discrete Wavelet Transform (DWT). Though these assumptions may sound valid in their representations, this study noted the contrary using the same dataset (i.e., Jaffe and MIT-CBCL). It was found out that the PCA/SVD algorithm is already equipped with an inherent ability to scale well without the need for image smoothing operations, hence rendering that filtering operation insignificant in this study and attribution of extra cost in time complexity.

4. Conclusion

In this study, we treated partial occlusion as a missing value problem and validated the performance of three selected data imputation schemes. Five (5) overlapping experiments were carried out in the validation process, all attempting to select the best data imputation scheme that works well with the famous SVD/PCA facial recognition algorithm. Using three datasets with varying structure and context, it was concluded that IA works exceptionally well across all five experiments with SVD/PCA compared to NIPALS and PMP. Further study in this area of partial occlusion seeks to review the existence of other nonstatistical missing data imputation schemes and how they could be implemented in this domain for improved performance.

Data Availability

The data used for this study are available on request from the following sites MIT: http://cblc.mit.edu/software-datasets/heisele/facerecognition-database.html Jaffe: https://zenodo.org/record/3451524.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


