Indian Journal of Science and Technology, *Vol 9(28), DOI: 10.17485/ijst/2016/v9i28/93968, July 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

On Gradient Descriptor Distribution-based Dimensionality Reduction Analysis: A Case of Recognizing Plastic Surgery Sample Faces

Chollette C. Olisah[1*], Peter Ogedebe[1] and Ghazali Sulong[2]

[1]Department of Computer Science, Baze University, Abuja, Nigeria; chollette.olisah@bazeuniversity.edu.ng, peter.ogedebe@bazeuniversity.edu.ng
[2]Faculty of Computing, Universiti Teknologi Malaysia (UTM), Skudai, Malaysia; gazali@utmspace.edu

Abstract

Objectives: To analyze the influence of the sparseness distribution characteristics of gradient-based descriptor data on reduction of high-dimensional data, this paper presents experimental analysis on learned samples of gradient descriptor data. **Method:** In order to draw valid inferences, a single gradient descriptor, the Edge based Gabor Magnitude (EGM) facial descriptor, is used. The descriptor data is learned using various linear subspace dimensionality reduction methods. The subspace models are the Principle Component Analysis plus Linear Discriminant Analysis (PCA plus LDA), supervised Locality Preserving Projection (sLPP) and Locality Sensitive Discriminant Analysis (LSDA) under the LGE and OLGE general framework (which in the present is used to aid the characterization of the data geometric properties). **Findings:** Using the plastic surgery data set, the following observations were made. The global based linear subspace model (PCA plus LDA) which do not require complex neighborhood assignment performs favorably well in relation to the graph embedding models. This may be due to the fact that it only works on the basis of class information. The LSDA is observed to be more affected by the nature of the descriptor data influenced by the complexity of plastic surgery because in all its identification rates, a below 60% is achieved. On the other hand, the sLPP show to be a best fit model for the sparse nature of the descriptor data. This can be attributed to its data preserving property by which it is able to preserve the local structures of a sparse data (gradient-based) and so outperformed the PCA plus LDA and most importantly, the LSDA. **Applications/ Improvements:** Understanding the *best fit model* for certain descriptor data is as important as optimizing recognition rates, an important observation for the face recognition research community.

Keywords: Data Distribution, Descriptor Data, Dimensionality Reduction, Face Recognition, Graph Embedding, Linear Subspace Learning

1. Introduction

The tasks of subspace models are to find and exploit the intricate low-dimensional structures in high-dimensional data[1]. The earliest methods are the Principal Component Analysis (PCA)[2] and Linear Discriminant Analysis (LDA)[3]. The PCA projects the principal components, usually described as the eigenvectors, linearly along the direction of maximal variance[2]. The PCA, when observed from the perspective of discrimination by[4] shows that it is a poor representation of the discriminative information of a data. The reason being that; data variance might be moving towards a non-discriminative direction. Though, its holistic information processing capability, which according to the research work in psychophysics[5], can play vital role in face discrimination, if properly adopted. The LDA finds application in classification tasks[6] due to its ability to maximize the separability criterion of between-class scatter in relation to the within-class scatter. Later on [7]introduced the two-stage PCA plus LDA dimensionality reduction approach so as to silence the weaknesses of the individual models while celebrating their strengths. The PCA plus LDA works by maximizing the between-class

*Author for correspondence

criterion and minimizing the within-class criterion of projected data. In the more recent years, research in dimensionality reduction moved from the holistic based methods to the non-linear (manifold) methods. For example, there is the Isometric Mapping (ISOMAP)[8], Maximum Variance Unfolding (MVU)[9] and Laplacian Eigenmaps (LE)[10]. However, in most cases they are overtaken by their linear counterparts such as Locality Preserving Projection (LPP)[11] and Locality Sensitive Discriminant Analysis (LSDA)[12]. In some instances, as observed by[13], for the data having discontinuous distribution due to noise and outliers, the linear counterparts outperform the non-linear methods. In order to extend the capabilities of the linear and non-linear subspace methods, in [14]proposed a general dimensionality reduction framework. The framework is of two categories namely; Linear Graph Embedding (LGE) and Orthogonal Linear Graph Embedding (OLGE). Despite the long existence of different subspace learning models, there is still a fundamental problem to be considered. The fundamental problem is about finding the discriminative elements of low-dimensional structures in high-dimensional data. After all, it is not all low-dimensional structures in high-dimensional data that are discriminative. This brings us to the following question: How does subspace learning model respond to the distribution of descriptor data? The distribution of facial information data is highly dependent on the facial descriptor. A facial descriptor can either be of intensity, texture or gradient domain which invariably means that the resulting data can be of different distribution. For instance, the gradient data contains more zero elements in its matrices than the data formed from the intensity or the texture-based descriptors. Though, it can generally be argued that the non-zero elements of the gradient information describe significant features of an object necessary for discrimination. At this junction, it will be interesting to note that the processing of the gradient data might vary across subspace learning models because it may be difficult to ascertain the low-dimensional structures of the high-dimensional data which are the discriminative information. This point was also raised by[15]. Therefore, there could be a way to work-around the discriminative elements of the gradient descriptor data, this we will discuss shortly, though it does not fall within the scope of this paper.

On the basis of the above stated, the following ideas can be established: It will be interesting to be able to define the distributions of data built on different categories of descriptor domain, that is, the intensity, texture or gradient. The outcome can go a long way to help in the design and development of subspace models that adapts to different descriptor data distributions. These stated points might potentially remedy the oversampling/over fitting problem caused by discontinuity, outlier or noise in data suffered by most subspace learning methods. More also, the outcome of the analysis might be of significance to removing anomalies[16] in the data or clustering significant data[17]. This is an open area that can be further investigated, but in this paper we considered analyzing some contemporary descriptors from the intensity, texture and gradient descriptor domains for describing facial images from the plastic surgery data set. We further show that the gradient-based descriptor is highly discriminative in comparison with the intensity and texture-based descriptors.

In order to demonstrate that different subspace methods respond differently to gradient data, we experiment using linear subspace methods and show that for increased discrimination, the subspace methods that best fits the gradient descriptor data is optimal. The use of the linear subspace models is based on the experimental observation in[13]. This paper extends the experimental analysis of the Edge-based Gabor Magnitude (EGM) feature[18]. However, our contribution is purely based on presenting knowledge that is critical to research in dimensionality reduction and essential to practical face recognition systems. How is that so? We will demonstrate through systematic experimentation the significance of gradient-based descriptors in comparison with descriptors from the intensity or texture domains as they tailor well to complexities, discontinuities, outlier or noise in sample data. More also, we investigate the influence sparseness property (the wild randomness in sample point's) of gradient based descriptor data, precisely on EGM, might have on dimensionality reduction process of linear subspace methods in retaining essential low-dimensional features for recognition. This investigation is carried out based on experimental analysis using some linear subspace methods to support or refute the stated assumption. And as well be able to ascertain the subspace method that best preserves the discriminative capabilities of the gradient descriptor data for the given sample data.

The rest of the paper is organized as follows. In Section 2, a brief introduction on the linear subspace methods adopted in the experiments and their respective graph embedding framework are presented. In Section 3 experiments were carried out on the publically available

real-world two-dimensional data set (i.e. the plastic surgery data set) and results of the face recognition experiment on the individual surgery procedures (commonly practiced in literature) are presented. In Section 4 we carried out further statistical analysis on the gradient descriptor data with respect to the subspace methods in order to support our hypothetical claim. Lastly, is the conclusion in Section 5.

2. Dimensionality Reduction via Linear Subspace Methods

The subspace models under consideration are the LSDA and sLPP employed under the LGE and OLGE general framework and PCA plus LDA.

Given the feature vector $x \in \Re^D$ from a set of training samples $x1, x2, \ldots, xn$ that belong to any one of the c classes. We assume that each class has an unknown distribution. The optimal interest is to be able to map the original high-dimensional data of the feature vector in D-dimensional space onto a D-dimensional space by a transformation function κ expressed as:

$$\kappa : \Re^D \rightarrow \Re^d \tag{8}$$

κ can be any of PCA plus LDA, LSDA or sLPP[11,19] dimensionality reduction methods, usually D is of much lower dimension than d, i.e, $d \ll D$. The reduced feature vector $y \in \Re^d$ is defined as:

$$y = W^{\mathrm{T}} x \tag{9}$$

The optimal objective of the subspace learning algorithms is to search W, a matrix representation in which all the *significant* observations are well retained. The process of finding such a matrix varies with the different objective function of different subspace learning models.

2.1 PCA plus LDA

The PCA[2] is often used for dimensionality reduction owing to the fact that it can preserve much useful information within a small dimensional space, but lacks the capability to solve a classification problem. LDA on the other hand utilizes the class information to maximize separation of data points of different classes while minimizing the within class feature points. Using κ -PCA, the most expressive features y are obtained by the following objective function $J(W)$ defined as[2]:

$$\mathrm{i} = 1, 2, \ldots, \mathrm{n}. \tag{10}$$

To further employ κ -LDA, we denote the mean values and grand means of the classes ci as Mi and M. The within-class scatter matrix $S\omega$ and the between-class scatter matrix Sb are defined as[3]:

$$S_{\mathrm{b}} = \sum_{i=1}^{c} P(\Omega_i)(M_i - M)(M_i - M)^T, \tag{11}$$

Where $P(\Omega i)$ is the probability of the ith class.

The LDA derives a projection matrix A that maximizes the Fisher's discriminant criterion:

$$J(W_{LDA}) = \arg\max_{(W)} \frac{\left| W S_{\mathrm{b}} W^T \right|}{\left| W S_{\omega} W^T \right|} \tag{12}$$

The Fisher's discriminant criterion is maximized when W consists of the eigenvectors of the matrix[3]:

$$S_{\omega}^{-1} S_{\mathrm{b}} W = W\Delta, \tag{13}$$

Where W *and* Δ are the eigenvector and eigenvalue matrices of $S_{\omega}^{-1} S_{\mathrm{b}}$, respectively. The two-stage (PCA plus LDA)[7] dimensionality reduction approach is employed to reduce feature dimension while maximizing the between-class criterion and minimizing the within-class criterion projected data points.

2.2 Graph Embedding based sLPP and LSDA

The sLPP[19] is a variant of the LPP[11] that uses the class label information to construct the new feature points through projecting a similarity graph that preserves the essential manifold structure of the data points. The LSDA takes into consideration the data manifold structure by constructing a single nearest neighbour graph, which further uses class label information to construct two graphs: The between-class graph and the within-class graph. These linear subspace learning methods (sLPP and LSDA) share common graph embedding formulation along with their linearization[13]. The graph embedding framework objective function $J(W)$ for LSDA and sLPP is defined as[1,13]:

$$J(W) = \arg\min_{(W)} \frac{\sum_{i \neq j} \left\| W^{\mathrm{T}} x_i - W^{\mathrm{T}} x_j \right\|^2 S_j}{\sum_{i \neq j} \left\| W^{\mathrm{T}} x_i - W^{\mathrm{T}} x_j \right\|^2 S_j^p}, \tag{14}$$

Where $S_{ij} = \begin{cases} 1, if\ \ x_i \in N(x_j)\ or\ \ x_j \in N(x_i) \\ 0, otherwise \end{cases}$, which is an adjacency matrix that describes the neighbourhood relationship, N of sample points x_i (ith sample point) and x_j (jth neighbour of the ith sample point). S_{ij}^{p} denotes the unconnected sample points.

This generalized graph embedding framework describes the LGE and OLGE. The only difference is that for the OLGE, an optimization function that curbs redundancy of projected data is employed[20]. This might come-off as a costly function for some data. While the sLPP and LSDA can be defined in a single graph embedding framework, they differ in their respective approach for computing the similarity graph. From the adjacency graph denoted by S_{ij} the objective function for which the similarity map is obtained with sLPP is optimized as follows[11,19]:

$$J(W_{sLPP}) = \arg\min_{(W)} \sum i \neq j \ (y_i - y_j)^2 S_{ij} \ , \qquad (15)$$

and for the LSDA two objective functions are derived individually for the within class $S_{w,ij}$ graph and between class $S_{b,ij}$ graphs[12]:

$$J(W_{LSDA})_b = \arg\min_{(W)} \sum i \neq j \ (y_i - y_j)^2 S_{w,ij} \qquad (16)$$

$$J(W_{LSDA})_w = \arg\max_{(W)} \sum i \neq j \ (y_i - y_j)^2 S_{b,ij} \qquad (17)$$

The above arguments in (14)-(17) translate as follows. For the {arg min}, if the sample pairs (*xi*, *xj*) are close and of the same class label then, the feature points (*yi*, *yj*) are as well close. For the maximum argument {arg max}, if the sample pairs (*xi*, *xj*) are close, but are of different class labels then, (*yi*, *yj*) are considered far apart from each other.

For a set of training samples *x1*, *x2*,…, *xn* that belong to any one of a training subject sets. Suppose the training sets are given as follows: $\{\{x_K^{(1)}\}, \{x_K^{(2)}\}, \{x_K^{(3)}\}, \cdots, \{x_K^{(n)}\}\} \subset \xi$ (Note that K is the number of samples per subject and n is the total number of subjects in the database), the descriptor features are obtained and learned for all the images in the set using the κ-linear subspace transformation models, which can be any of PCA plus LDA, sLPP-LGE, sLPP-OLGE, LSDA-LGE, or LSDA-OLGE. The reason for learning a set of samples, known as the training sets, is for the face recognition system to generalize well to an unknown sample (test image)[11], which is possible in the reduced space because, as earlier noted, feature vectors that are of large dimensions can hinder the classification processes. In the subsequent section, we will present the experiment scenario and results of face recognition carried out on publically available real-world two-dimensional data set (i.e. the plastic surgery data set).

3. Experiments

In this section, we conduct several identifications and verification experiments using the plastic surgery data set and in the manner that is informative.

3.1 Database and Experimental Setup

Two evaluation scenarios: Identification and verification on the plastic surgery data set[21] are presented. The plastic surgery data set consists of one image each of pre-surgery and post-surgery images of real people who have undergone plastic surgery. For the identification scenario, two types of experiments, namely "without subspace learning" and "with subspace learning" are performed. The first experiment is the case of "without subspace learning". The recognition accuracy of a number of facial descriptors categorized as texture domain (LBP-based descriptors[22] and Gabor descriptor[23]) and gradient domain (the EGM facial descriptor[18]) with no training processes (i.e., applying subspace learning methods), are evaluated. The second experiment presents the various results of EGM facial descriptor on employing "subspace learning methods" at varied number of dimensions. Since the intrinsic complexity and dimension of data are assumed to lie in low dimensional space[24] it will be interesting to investigate how different dimensionality reduction methods transform the EGM data across different plastic surgery procedures. The dimensionality reduction methods evaluated are the linear subspace methods. Asides the reports in literature that have determined the efficiency of linear subspace methods over the nonlinear subspace methods[13,25], the reason why only the linear subspace methods are considered in this experiment is because, as earlier noted, the subspace of data lying on low-dimension is said to be linear[24]. In both experiments, the evaluation is on the basis of different plastic surgery procedures because it will be interesting to observe how the data at different plastic surgery procedures influence the subspace learned EGM facial pattern recognition system. On the existing aesthetic

plastic surgery database, we generated one image each of the available pre-surgery and post-surgery images, that is, their mirror versions. Firstly, this way, the under-sample problem suffered by some subspace learning models, which require a number of images in the training set for good generalization is avoided. Secondly, the mirror image can be used to recognize a face image[26,27]. Thirdly, so that one is able to evaluate recognition performance for a situation where there exists a post-surgery image in the gallery set. Therefore, the total number of images per subject in the database becomes four. In the two experiments, we report the Cumulative Match Curve (CMC) scores using the settings: Gallery set contains 3 images per subject and the remaining image per subject is used as the probe images, but in the case of "with subspace learning", the training is performed on the gallery set. In the verification scenario we evaluate the one-to-one identity verification capability of the subspace learned EGM facial pattern recognition system. In this scenario, the main interest is to determine whether the claimed identity of a subject (pre-surgery face image in the gallery set) matches the currently enrolled identity of an individual (i.e., the probe, which is the post-surgery image). We report the verification scores of the subspace learned EGM facial pattern recognition system on employing different linear subspace dimensionality reduction methods via the Receiver Operating Characteristics (ROC) curves. The experimental evaluation is based on a one-to-one setting, that is, of the 4 images available per subject we used 2 pre-surgery images for training. We selected 1 pre-surgery image each from a subject used for training to make up the gallery set, while a single post-surgery image each of a subject made up the probe set.

The dataset configuration and the experimental parameters applied for different descriptors for all the experiments are summarized in the given tables under the following sub-headings.

3.1.1 Dataset Configuration

The aesthetic plastic surgery procedures are listed alongside the number of subjects in each of the surgery procedures in Table 1.

3.1.2 Parameter used for the Descriptors

We replicate the settings for each of the descriptors (EGM, LGBP, CLBP-M-S, CLBP-S, CLBP-M) used in the experiment according to their original usage in literature. These settings are provided in Table 2. The abbreviations woc, SQI and rgbGE stand for without correction, self-quotient image and rgb-gamma encoding technique.

3.2 Evaluation of Gradient Descriptor (EGM) with some State of the Art Descriptors.

First, we present the identification results of different descriptor-based face recognition methods on the high-dimensional representation of the face data with-

Table 1. The constituents of the plastic surgery database

Modification Category	Plastic surgery procedure	What changes	Number of subjects
Soft Tissue	Blepharoplasty	Skin texture, relative position and size of eyes	105
	Skin Peeling	Skin texture	74
	Rhytidectomy	Global skin texture and relative position and size of features	308
	Dermabrasion	Skin texture	32
	Others	Skin texture, relative position and size of some features	56
	Browlift	Skin texture around the forehead, relative position of the brow	60
Hard Tissue	Rhinoplasty	Relative position of nose tips and size of the nose	192
	Otoplasty	Relative position of ear *concha*	74
	Cheek and Chin	Skin texture around the chin / jaw, relative position and size of chin line.	21

Table 2. Some parameter settings for the different descriptors

Method	Illumination Normalization	Pixel (p) Neighbourhood
CLBP-M [22]	woc	*p*-8
CLBP-S [22]	woc	*p*-8
CLBP-S-M [22]	woc	*p*-8
LGBP [23]	self-quotient image (SQI)	*p*-8
EGM [18]	rgbGE [18]	*p*-8 (defined by Sobel mask)

out employing subspace learning methods or a training process. The facial descriptors under comparison are the LBP variants (CLBP-M-S, CLBP-M and CLBP-S), LGBP and EGM. We report the recognition rate on Rank basis, considering at most the Rank 10. Rank is the ratio of correctly recognized faces in the order they are scored. If the probe face image is correctly identified at each recognition process it is ranked highest, this is represented as Rank-1, while the subsequent Ranks from 2 and above means that there is some degree of freedom. The results of the various facial descriptors (without subspace learning) in the recognition of surgically altered face images are shown in Figure 1. From the figure we make the following observations.

The EGM descriptor is observed to be more robust against the noise, outlier or discontinuity that might be present in the plastic surgery data set, which is shown by its above 60% Rank-1 recognition rate in all the experiments than the LGBP and LBP variants. The identification accuracy of the variants of LBP descriptor (texture) is rather disappointing. They failed to reach a satisfactory recognition rate despite having features of lower dimension. Overall, the EGM facial descriptor shows to have the best Rank recognition performance. It achieved as high as 100% recognition rate for all the Ranks especially in the case of recognizing Brow lift surgery altered face images. We observe that the plastic surgery procedures that directly impacted on the nose, eye or the overall face which have been found in psychophysics and computer vision to contribute largely to face recognition task, were only correctly recognized within the range 50% to 60%. However, in those procedures (Blepharoplasty, Rhytidectomy and Rhinoplasty) the EGM showed to have achieved the top most recognition rates in comparison to other descriptors. We can deduce that the EGM would have performed better if not for the viewpoint changes of images in the data set. Since for every two images I and J under the same viewpoint, that is, the EGM-I and EGM-J (gradient-based features) must be parallel at every point where they are defined[28]. For all other procedures, which are basically skin textures changing procedures, the EGM is most insensitive to the alterations they make on the faces, followed by LGBP. We also observed that the CLBP-S performed surprisingly well from Rank 5 to 10 in the recognition of Blepharoplasty surgery altered face images, while CLBP-M-S performed better than each features apart for all the other plastic surgery procedures.

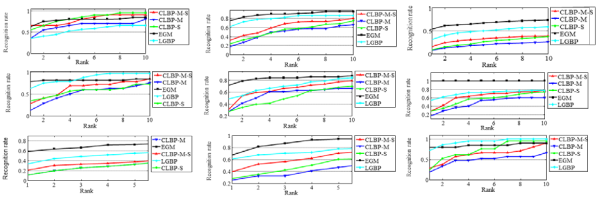


Figure 1. Identification result for CLBP-M-s, CLBP-M, CLBP-S, EGM and LGBP. Starting from left corner (coordinates 0, 1) to the lower right corner (coordinates 1, 0) are the blepharoplasty, skin peeling, rhytidectomy, dermabrasion, others, brow lift, rhinoplasty, otoplasty and cheek and chin aesthetic plastic surgery procedures, respectively.

3.3 Evaluation of EGM on Adopting different Subspace Learning Methods

In this second experiment, our interest is to observe how the subspace learning methods are capable of handling the sparseness property of the EGM data. We are referring to the ability of the subspace learning methods to retain a low-dimensional representation of the high-dimensional EGM data in such a way that the significant information of the high-dimensional data is preserved. Therefore, it will be interesting to observe how recognition rate transcends as the significant and discriminative information within the EGM are more retained in the reduced space. On this basis, we report the systematic analysis of the identification performances of EGM on employing different subspace learning methods. The subspace learning methods under comparison are the PCA plus LDA, sLPP and LSDA. The performances of sLPP and LSDA are obtained under a generalized LGE and OLGE framework at varied number of dimensions. We also report the recognition rate under the different plastic surgery procedures on Rank basis. We are considering the Rank 1 to 5 of which each of the graphs comprises of the result of varying dimensions of the subspace learning methods and represented by their various Ranks. We show in boldface in Table 3 the best performing subspace learning method for different plastic surgery procedures. The table summarizes the Rank 1 results corresponding to Figure 2. From the results of this experiment shown in Figure 2 we make the following observations.

According to the experimental results we see that for all the plastic surgery procedures experimented on, the LSDA did not improve the results of EGM in its high-dimensional space (that is, the case of the first experiment). The rec-

ognition performance of EGM from LSDA is surprisingly disappointing. This resulting outcome of LSDA cannot be likened to any report in literature on LSDA. Therefore, we infer that the poor performance of LSDA might be resulting from the nature of the data set. Also, the gradient domain being the building block of the EGM descriptor implies that the descriptor will conform to the sparseness property of gradient distribution, which is often characterized by the sparseness nature of the data distribution. We refer the reader to the following literatures[15,29,30–,32]. Therefore, attributing the failure of the LSDA to the sparseness characteristics of gradient-based data should not be out of place. The LSDA follows a local approach to the globality of LDA by constructing two graphs, a within-class graph and between-class graph, from one nearest neighbour graph[12]. These capabilities of the

LSDA should by no doubt make it perform better than PCA plus LDA, but this is not the case, it is rather the reverse. Thus, it is also by no means out of place to say that LSDA is unable to guarantee the global connectedness within the constructed between-class and within-class

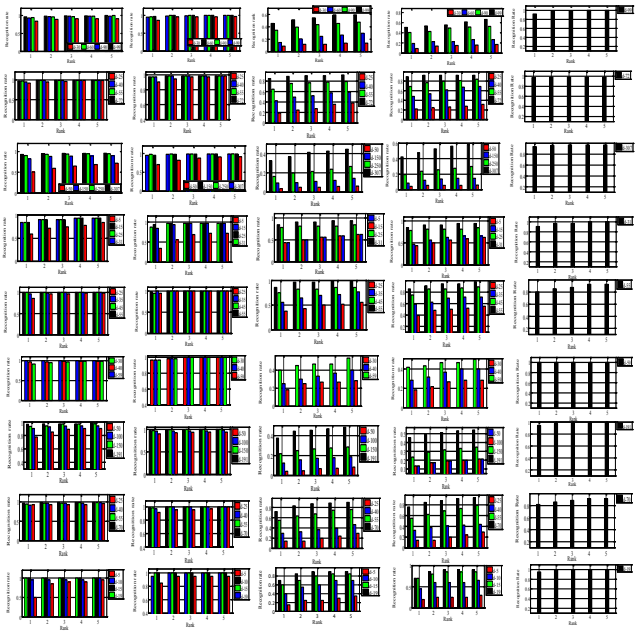


Figure 2. Identification results of various subspace learning (columns) from EGM with their respective dimensions for different plastic surgery procedures (rows). Row 1-9 (blepharoplasty, skin peeling, rhytidectomy dermabrasion, others, brow lift, rhinoplasty, otoplasty and cheek and chin, respectively) while column 1-5 (EGM-PCA+LDA, EGM-sLPP-LGE, EGM-sLPP-OLGE, EGM-LSDA-LGE and EGM-LSDA-OLGE, respectively).

graphs for this kind of data. If the globality is considered in constructing these graphs, it might remedy the said limitation of LSDA that is observed in this paper. From the experiment we see that the simple PCA plus LDA performed far better than the LSDA for all the plastic surgery data set cases we investigated. Since the PCA plus LDA is concerned with the projection of the within-class and between-class in relation to the global structure of the manifold, we can claim that it is somewhat insensitive to the underlying complexity of the data distribution and this worked in favour of the PCA plus LDA. In some of the experiments presented here, the PCA plus LDA performs comparably to sLPP despite its simplicity, e.g., skin peeling and brow lift data set cases. From the results, it is obvious that PCA (no loss of information[33]) complements LDA (classification capability[6]) and that is the reason why they both are a great merge for subspace learning.

Overall, the sLPP outperforms the PCA plus LDA and LSDA under the generalized framework; LGE and OLGE. We can attribute the performance of the sLPP to the following: sLPP follows a linear approximation to a nonlinear manifold learning method known as the Laplacian eigenmaps[34]. Unlike the LSDA, the sLPP uses a single graph and class labels to define the local neighbourhood information of the data points, which implies that connectedness of data points can exist. One major advantage of sLPP over other methods compared in this paper is that: 1. It best preserves the essential manifold structure[35] of the EGM within face space, and 2. The gradient-based data have been assumed to follow a Laplacian distribution[15], this best explains the reason for the better fit of the sLPP (note that this is an observation made from a performance point of view) to the essential manifold structure of EGM data and 3. It is a linear approximation of a nonlinear dimensionality reduction method, which implies that it is somewhat more suited for sample data with outliers than PCA plus LDA and LSDA. From the perspective of the generalized framework, the OLGE performed better than the LGE generalized framework. This is for the fact that the concept of orthogonality helps in better representation of the original data via curbing redundancy in projected data[19]. The role that orthogonality can play in subspace learning has been stated in[36], which is that the performance of subspace learning methods can be improved from the basis of orthogonality. We summarize the results of the experiment corresponding to Figure 2 on Rank 1 basis for each plastic surgery procedure in Table 3.

3.4 EGM Verification Results on Applying Different Subspace Learning Methods

The changes in facial appearance as a result of plastic surgery can range from mild to severe. After some plastic surgery procedures, a complete transformation of facial identity is expected, which is the reason why plastic surgery can be an avenue for criminals to conceal identity and remain elusive to face recognition systems. An example is the popular case of Andrew Moran a most wanted fugitive, who by means of plastic surgery evaded arrest[37]. It took the law enforcement four years to apprehend him. And it is likely that there are many others who have been evading arrest via the aid of plastic surgery. There has also been a case when as a result of the changes in facial appearance due to plastic surgery the customs officers at the China Hongqiao International airport could not relate the identity claim of a group of women to their passport identity in the system[38]. As reflected in the above incidents, the identities of the persons who undergo plastic surgery procedures have a high likelihood to tend towards a different person's identity than their actual identity. Therefore, the face verification experiment should also be a way of evaluating face recognition methods across different plastic surgery procedures. Figure 3 shows the ROC curves for different plastic surgery procedures,

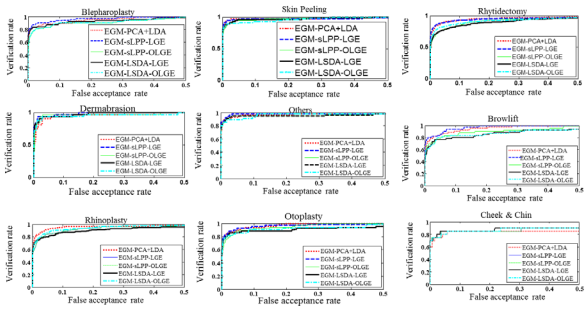


Figure 3. Verification results of subspace learning from EGM for various plastic surgery procedures. These are blepharoplasty, skin peeling, rhytidectomy dermabrasion, others, brow lift, rhinoplasty, otoplasty and cheek and chin, respectively. The subspace models are; EGM-PCA+LDA, EGM-sLPP-LGE, EGM-sLPP-OLGE, EGM-LSDA-LGE and EGM-LSDA-OLGE, respectively.

Table 3. Identification performances of subspace learning from EGM under different dimensions for various plastic surgery procedures

Method	Blepharoplasty (%) d-30 d-60 d-90 d-99	Skin-peeling (%) d-25 d-40 d-55 d-72	Rhytidectomy (%) d-50 d-150 d-250 d-307
EGM-PCA+LDA	92.00	97.30	88.20
EGM-sLPP-LGE	84.16 93.07 94.06 96.04	91.80 94.50 97.30 97.30	51. 51.00 82.80 90.60 93.20
EGM-sLPP-OLGE	86.14 95.05 96.04 97.03	90.40 97.30 97.30 97.30	70.50 94.80 97.70 94.80
EGM-LSDA-LGE	8.91 14.90 34.70 45.50	19.20 42.50 64.40 84.90	03.57 09.42 16.20 32.50
EGM-LSDA-OLGE	8.91 18.80 40.60 50.50	21.90 47.90 68.50 89.00	04.22 08.44 19.50 42.20
Method	**Dermabrasion (%) d-5 d-15 d-25 d-31**	**Others (%) d-25 d-35 d-45 d-55**	**Browlift (%) d-30 d-40 d-59**
EGM-PCA+LDA	90.60	80.40	100
EGM-sLPP-LGE	59.40 84.40 84.40 84.40	85.70 96.40 100 100	91.70 96.70 100
EGM-sLPP-OLGE	34.38 84.38 87.50 93.75	94.60 100 94.60 98.20	96.70 96.70 96.70
EGM-LSDA-LGE	43.80 43.80 78.10 84.10	37.50 55.40 75.00 85.70	18.30 28.30 41.70
EGM-LSDA-OLGE	43.80 46.90 78.10 84.40	39.30 58.90 75.00 85.70	18.30 25.00 40.00
Method	**Rhinoplasty (%) d-50 d-100 d-150 d-191**	**Otoplasty (%) d-25 d-40 d-55 d-70**	**Cheek&Chin (%) d-5 d-15 d-25 d-31**
EGM-PCA+LDA	94.30	81.70	80.40
EGM-sLPP-LGE	77.60 90.10 93.20 96.40	85.70 91.50 94.40 97.20	85.70 96.40 100 100
EGM-sLPP-OLGE	90.10 96.40 97.40 97.40	91.50 97.20 98.60 98.60	94.60 100 94.60 98.20
EGM-LSDA-LGE	4.17 12.50 21.90 37.50	14.10 29.60 54.90 71.80	37.50 55.40 75.00 85.70
EGM-LSDA-OLGE	13.50 13.50 23.40 45.30	14.10 32.40 54.90 76.10	39.30 58.90 75.00 85.70

namely blepharoplasty, skin peeling, rhytidectomy, dermabrasion, others (botox, liposhaving, etc.), browlift, rhinoplasty, otoplasty and cheek and chin surgeries. The ROC curve is the plots of the verification rate (which displays the probability that a correct identity is accepted) to the False Acceptance Rate (FAR) which displays the probability that an identity is incorrectly accepted. Given FAR at 0.01, 0.05 and 0.1, we summarize the verification results of the ROC curves in Table 4.

In this experiment we able to confirm the claim that LSDA is sensitive to factors affecting the data set because when presented with images of a person that are of the same angle (viewpoint) in the verification experiment, a tremendous difference in face recognition performance is observed for EGM with LSDA. Overall, the verification experiments for all the aesthetic plastic surgery procedures show impressive recognition ability of the subspace learned EGM facial pattern recognition system in matching successfully the pre-surgery image of a person to his/her post-surgery image.

Surprisingly, unlike in the previous experiment, the sLPP subspace learned EGM facial pattern recognition system under the LGE general framework performed better in comparison to the OLGE. This obviously can be attributed to the fact that there is only fewer data available for comparison. The reason is that the OLGE works best when the data available for comparison is huge. Also, despite the simplicity of the PCA plus LDA in comparison with the sLPP and LSDA under the generalized framework, the PCA plus LDA subspace learned EGM facial pattern recognition system performed favourably in comparison to the sLPP, but on the average the sLPP is still top best.

4. Statistical Analysis and Discussion

To demonstrate that the gradient descriptor data do not follow a normal distribution, we present a statistical analysis using a measure known as the Quantile-Quantile (Q-Q) plot. The Q-Q plot is a plot of the probability distributions of data points. Given in Figure 4 is the plot of a sample frontal face image whose intrinsic patterns are

Table 4. Verification performances of subspace learning from EGM for various plastic surgery procedures

Method	**Blepharoplasty (%)** **0.01 0.05 0.1**	**Skin-peeling (%)** **0.01 0.05 0.1**	**Rhytidectomy (%)** **0.01 0.05 0.1**
EGM-PCA+LDA	81.19 88.12 89.11	94.52 97.26 97.26	76.62 90.91 93.18
EGM-sLPP-LGE	82.18 90.10 93.07	91.78 95.89 97.26	80.84 89.94 93.51
EGM-sLPP-OLGE	74.26 82.18 90.10	83.56 94.52 94.52	64.29 79.87 85.39
EGM-LSDA-LGE	80.20 84.16 89.11	87.67 95.89 95.89	66.56 78.57 84.74
EGM-LSDA-OLGE	74.26 82.18 90.10	89.04 91.78 93.15	69.48 83.44 88.31
Method	**Dermabrasion (%)** **0.01 0.05 0.1**	**Others (%)** **0.01 0.05 0.1**	**Browlift (%)** **0.01 0.05 0.1**
EGM-PCA+LDA	75.00 90.63 96.88	89.29 92.86 96.43	73.33 85.00 90.00
EGM-sLPP-LGE	84.38 93.75 96.88	87.50 98.21 98.21	78.33 83.33 93.33
EGM-sLPP-OLGE	81.25 96.88 96.88	94.60 94.60 98.20	66.67 78.33 83.33
EGM-LSDA-LGE	84.38 93.75 93.75	85.71 89.31 90.15	63.33 76.67 80.00
EGM-LSDA-OLGE	78.13 93.75 93.75	55.40 75.00 85.70	61.67 80.00 83.33
Method	**Rhinoplasty (%)** **0.01 0.05 0.1**	**Otoplasty (%)** **0.01 0.05 0.1**	**Cheek&Chin (%)** **0.01 0.05 0.1**
EGM-PCA+LDA	80.73 92.71 95.83	81.69 90.14 94.37	70.00 80.00 85.00
EGM-sLPP-LGE	64.06 84.90 88.54	81.69 92.96 95.77	75.00 85.00 85.00
EGM-sLPP-OLGE	64.06 84.90 88.54	71.83 88.73 92.96	65.00 80.00 85.00
EGM-LSDA-LGE	72.92 80.21 87.50	83.10 85.92 88.73	65.00 80.00 85.00
EGM-LSDA-OLGE	70.83 86.98 90.63	77.46 85.92 91.55	75.00 80.00 85.00

defined using EGM, a gradient-based descriptor. Note that the best reading of the Q-Q plot is towards the positive, that is, the Right Hand Side (RHS), which is actually in the direction of significance.

The probability plot of the sample data is fitted over a standard normal quantile that is based on theoretical concepts of normal distributions. It can be clearly observed that the data points (indicated using the blue star points) for the EGM sample data deviated from the normal statistical distribution (indicated by the red line). However, carefully observing the distributions of the sample, it can be established that the distribution of the sample face image defined using EGM is heavy-tailed. Its distribution started deviating from the normal at point 0.93137 (x-axis) up until point 6.834 (y-axis).

To statistically investigate how the subspace learning methods respond to the distribution of the EGM data, the subspace learning methods introduced in this paper are used and are shown in Figure 5.

Using the linear subspace models, PCA plus LDA, sLPP and LSDA, the linear transform of the sample data of a face image defined by the EGM descriptor is demonstrated by means of the Q-Q probability plots. An obvious inference is drawn from Figure 5 (a-c) which is that sLPP best transforms the EGM data. The second

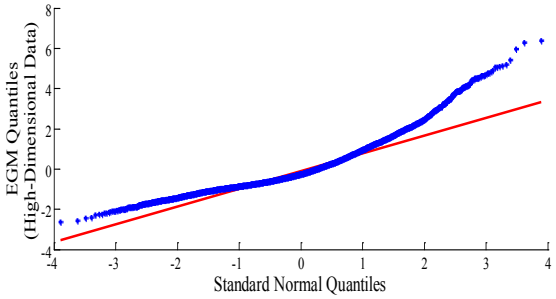
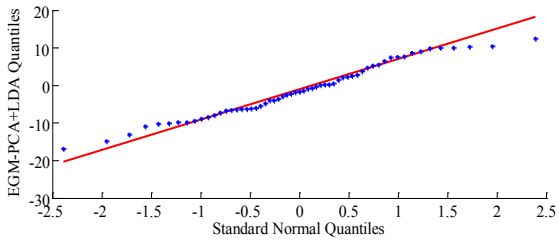
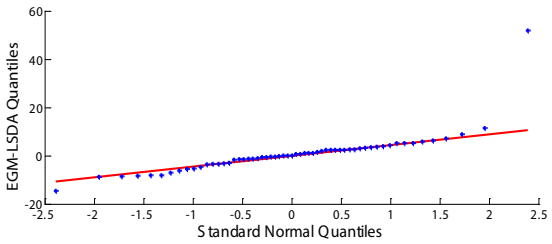


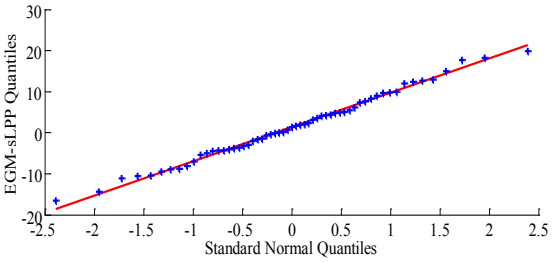
Figure 4. Statistical Q-Q plot of the distribution of sample face data. The high-dimensional data distribution of face sample defined using EGM descriptor.



(a)



(b)



(c)

Figure 5. Statistical Q-Q plot of the distribution of face data defined using subspace learning from EGM. **(a)** Subspace learning with PCA plus LDA. **(b)** Subspace learning with LSDA. **(c)** Subspace learning with sLPP.

best is the PCA plus LDA. The LSDA showed to have had one extreme outlier which is a clear indication that the EGM sample is heavy-tailed. The term heavy-tail means that there is a wild randomness in sample point's distribution. On removing the outlier, LSDA can be said to be more linearly distributed than PCA plus LDA. However, a statistical tool for such removal is required which can be addressed by defining the distributions of the data in order to proffer solutions. It can be said that a single data cannot alone define the behaviour of these methods, but we have just presented what seems like first-hand insight into the influence of a descriptors data distribution on dimensionality reduction methods. Further analysis can, however, be made in the future to observe from numerous sample data defined using EGM descriptor for different data sets known to be having discontinuity, outlier or noise.

5. Conclusion

Based on the need to represent effectively facial information, facial descriptors are used. We employed two categories of descriptors, namely texture and gradient domain descriptors. From the initial experiments, using various descriptors in each of the categories, we were able to observe that the gradient-based descriptor was more

superior to the texture-based descriptors with a good enough margin. But then the distribution of gradient-based descriptor data was statistically proven to be of heavy-tailed distribution due to the sparseness property of descriptors built on the concept gradients. For this reason, we carried out several systematic analyses, in this paper, for determining the linear subspace model that best fits the sparseness property of the distribution of the gradient descriptor data based on the facial image representation and recognition capabilities of the models. The performances of the linear subspace learning methods have been systematically evaluated and compared on real-world two-dimensional facial image data set. Our main findings are as follows.

Through our systematic analysis, it was evident that the resulting distribution of a descriptor face image data is significant to the discriminative ability of the descriptor on applying subspace learning methods for dimensionality reduction. We observed that the linear subspace learning methods used to achieve dimensionality reduction behaved differently than their usual known behaviour in literature. For instance, Locality Sensitive Discriminant Analysis (LSDA) fell short of expectation on top of the much added computational expenses. From the experimental results it was shown that the LSDA is highly sensitive to the distribution property of gradient-based data. Surprisingly, the Principle Component Analysis plus Linear Discriminant Analysis (PCA plus LDA) beat the LSDA in performance. We are somewhat speculating that it is the "no knowledge of the data distribution, but of the class information of the sample data" that contributes or is mainly responsible for the good performance of the PCA plus LDA. On the other hand, the supervised Locality Preserving Projection (sLPP) was observed to overcome the wild randomness (sparseness property) in the distribution of the gradient-based data which causes it to be heavy-tailed. A high recognition rate of the gradient-based descriptor was achieved using the sLPP for reducing the dimensionality of the data. It wasn't much of a surprise anyway that the sLPP improved the descriptor discriminative capability by 23.84% because the heavy-tailed distribution almost follows a Laplace distribution and the sLPP is a linear version of the Laplacian eigenmaps.

These findings do suggest that the distribution of data is a significant aspect of research in dimensionality reduction of a sample face image with features characterized by a descriptor, whose descriptor domain is either based on intensity, texture or gradient. Therefore, this paper may likely inspire various research efforts in dimensionality reduction to examine with careful consideration; the distribution of descriptor data in the design of a method that fits all distribution or has prior knowledge of the distribution of input data to the subspace learning model.

6. Competing Interests

The authors declare that they have no competing interests.

7. References

1. Yan S, Wang H, Liu J, Tang X, Huang TS. Misalignment-robust face recognition. IEEE Transactions on Image Processing. 2010 Mar; 19(4):1087–96.
2. Jolliffe IT. Principal Component Analysis. New York, NY: Springer-Verlag; 1986.
3. Swets D, Weng J. Using discriminant eigenfeatures for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1996 Aug; 18(8):831–6.
4. Turk M, Pentland A. Eigenfaces for recognition. Journal of Cognitive Neuroscience. 1991; 3(1):71–86.
5. Taubert J, Apthorp D, Aagten-Murphy D, Alais D. The role of holistic processing in face perception: Evidence from face inversion effect. Vision Research. 2011 Jun; 51(11):1273–8.
6. Tao D, Jin L. Discriminative information preservation for face recognition. Journal of Neurocomputing. 2012; 91(15):11–20.
7. Belhumeur P, Hespanha J, Kriegman D. Eigenfaces vs fisher faces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997; 19(7):711–20.
8. Tenenbaum J, Silva V, Langford J. A global geometric framework for nonlinear dimensionality reduction. Journal of Science. 2000; 290(5500):2319–23.
9. Weinberger KQ, Sha F, Saul LK. Learning a kernel matrix for nonlinear dimensionality reduction. ACM Proceedings of the 21st International Conference on Machine Learning; Banff, Canada. 2004 Jul 4-8. p. 1–10.
10. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. Journal of Neural Computation. 2003; 15(6):1373–96.
11. Niyogi X. Locality preserving projections. Proceedings of the Neural Information Processing Systems, arXiv; Vancouver, Canada. 2004 Dec 13-18. p. 153–60.
12. Cai D, He X, Zhou K, Han J, Bao H. Locality sensitive discriminant analysis. Proceedings of the 20th International Joint Conference on Artificial Intelligence. Hyderabad, India. 2007 Jan 6-12. p. 708–13.

13. Huang W, Yin H. On nonlinear dimensionality reduction for face recognition. Journal of Image and Vision Computing. 2012; 30(4):355–66.
14. Yan S, Xu D, Zhang B, Zhang HJ, Yang Q, Lin S. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007; 29(1):40–51.
15. Tzimiropoulos G, Zafeiriou S, Pantic M. Subspace learning from image gradient orientations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012; 34(12):2454–66.
16. Kumar SB, Subramanian K. Removing of anomalies in high dimensional data multi clustering structure. Indian Journal of Science and Technology. 2015 Nov; 8(32). DOI: 10.17485/ijst/2015/v8i32/87510.
17. Veeraiah D, Vasumathi D. Feature sub selection over high dimensional data based on classification models. Indian Journal of Science and Technology. 2016 Feb; 9(8). DOI: 10.17485/ijst/2016/v9i8/84166.
18. Chude-Olisah CC, Sulong G, Chude-Okonkwo U, Hashim S. Face recognition via edge-based Gabor feature representation for plastic surgery-altered images. EURASIP Journal on Advances in Signal Processing. 2014; 2014:102.
19. Zheng Z, Yang F, Tan W, Jia J, Yang J. Gabor feature-based face recognition using supervised locality preserving projection. Journal of Signal Processing. 2007; 87(10):2473–83.
20. Cheng M, Pun CM, Tang YY. Nonnegative class-specific entropy component analysis with adaptive step search criterion. Journal of Pattern Analysis and Applications. 2014; 17(1):113–27.
21. Singh R, Vatsa M, Bhatt H, Bharadwaj S, Noore A, Nooreyezdan S. Plastic Surgery: A new dimension to face recognition. IEEE Transactions on Information Forensics and Security. 2010; 5(3):441–8.
22. Guo Z, Zhang DA. Completed modeling of local binary pattern operator for texture classification. IEEE Transactions on Image Processing. 2010; 19(6):1657–63.
23. Zhang W, Shan S, Gao W, Chen X, Zhang H. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. Proceedings of the 10th IEEE International Conference on Computer Vision; Beijing. 2005 Oct 17-21. p. 786–91.
24. Ma Y, Niyogi P, Sapiro G, Vidal R. Dimensionality reduction via subspace and submanifold learning. IEEE Signal Processing Magazine. 2011; 28(2):14–126.
25. Yan J, Zhang B, Liu N, Yan S, Cheng Q, Fan W, Chen Z. Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing. IEEE Transactions on Knowledge and Data Engineering. 2006; 18(3):320–33.
26. Seo HJ, Milanfar P. Face verification using the lark representation. IEEE Transactions on Information Forensics and Security. 2011; 6(4):1275–86.
27. Etemad K, Chellappa R. Discriminant analysis for recognition of human face images. Journal of the Optical Society of America. 1997; 14(8):1724–33.
28. Chen Z, Liu C, Chang F, Han X, Wang K. Illumination processing in face recognition. International Journal of Pattern Recognition and Artificial Intelligence. 2014. p. 187–214.
29. Jia Y, Darrell T. Heavy-tailed distances for gradient based image descriptors. Proceedings of the Neural Information Processing Systems. arXiv; Granada, Spain. 2011. p. 397–405.
30. Kim HS, Schulze JP, Cone AC, Sosinsky GE, Martone ME. Multichannel transfer function with dimensionality reduction. Proceedings of SPIE 7530 on Visualization and Data Analysis; San Jose, California. 2010.
31. Nayak S, Sarkar S, Loeding B. Distribution-based dimensionality reduction applied to articulated motion recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009; 31(5):795–810.
32. Zheng Y, Grossman M, Awate SP, Gee JC. Automatic correction of intensity non-uniformity from sparseness of gradient distribution in medical images. Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention; Berlin Heidelberg: Springer; London, UK. 2009. p. 852–9.
33. Murali M. Principal component analysis based feature vector extraction. Indian Journal of Science and Technology. 2015 Dec; 8(35). DOI: 10.17485/ijst/2015/v8i35/77760.
34. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. Journal of Neural Computation. 2003; 15(6):1373–96.
35. He X, Yan S, Hu Y, Zhang HJ. Learning a locality preserving subspace for visual recognition. Proceedings of the 9th IEEE International Conference on Computer Vision; Nice, France. 2003 Oct 13-16. p. 385–92.
36. Cheng M, Fang B, Pun CM, Tang YY. Kernel view based discriminant approach or embedded feature extraction in high-dimensional space. Neurocomputing. 2011; 74(9):1478–84.
37. Moran A. Plastic surgery helped 'most wanted' fugitive evade arrest. 2015. Available from: http://www.telegraph.co.uk/news/uknews/crime/10054091/Andrew-Moranplastic-surgery-helped-most-wanted-fugitive-evade-arrest.html
38. South Korean Plastic Surgery Trips = Headaches for Customs Officers; 2011. Available from: http://shanghaiist.com/2009/08/04/south_korean_plastic_surgery_trips.php